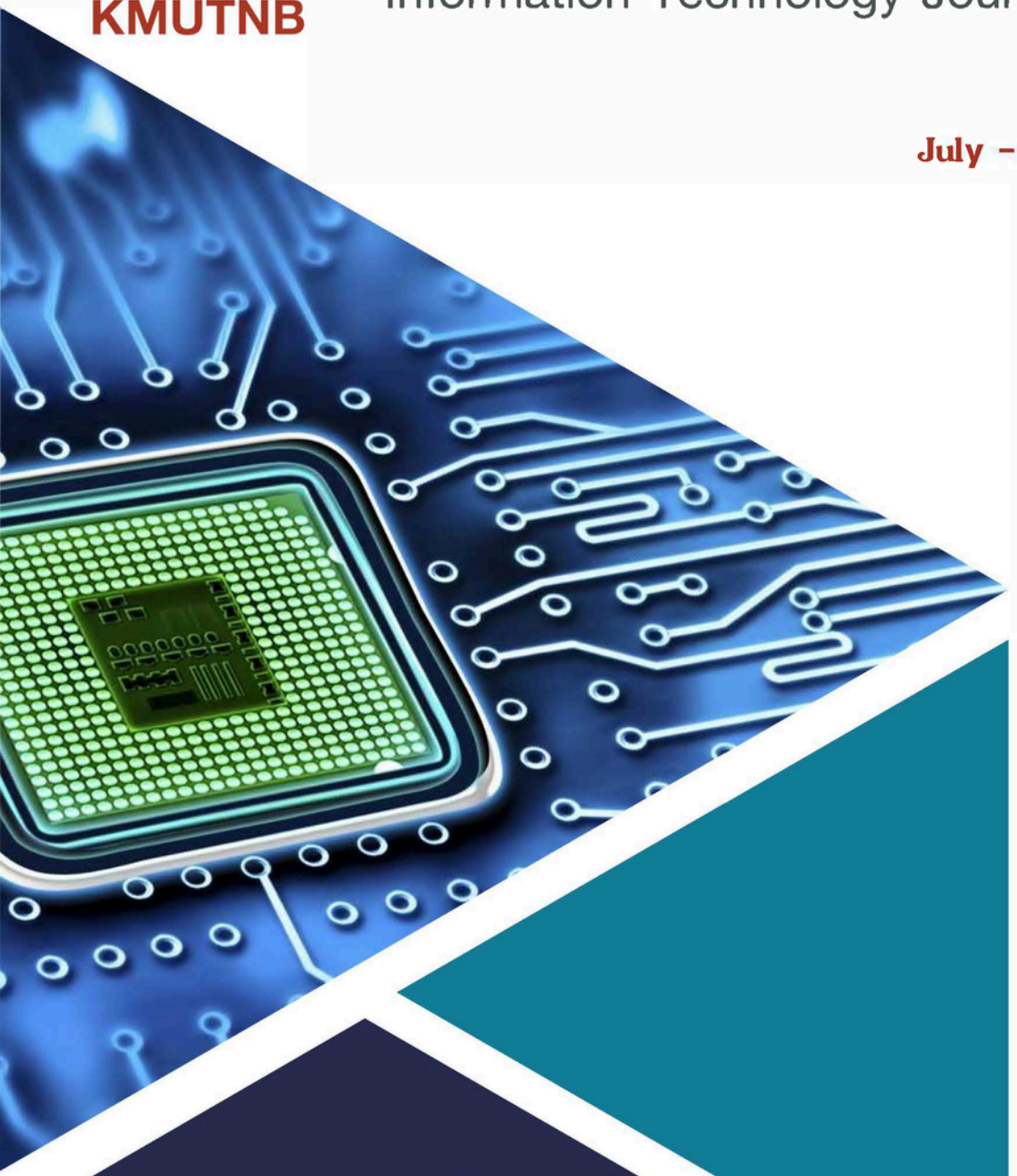
KMUTNB

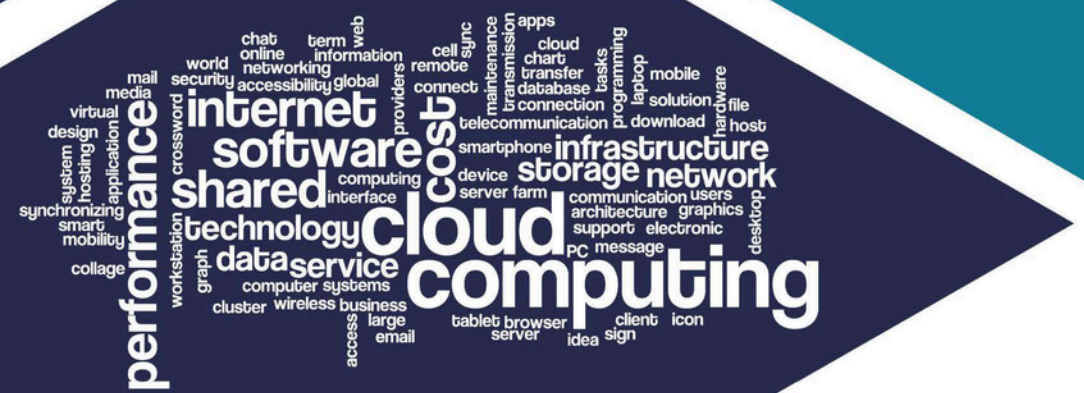
วารสารเทคโนโลยีสารสนเทศ มจพ.

Information Technology Journal KMUTNB

ITJ

Vol. 20 No. 2
July - December 2025





ISSN 1685-8573

TCI Group 1

วารสารเทคโนโลยีสารสนเทศ มจพ.
Information Technology Journal KMUTNB
(IT Journal KMUTNB)
ปีที่ 21 ฉบับที่ 2 เดือนกรกฎาคม - ธันวาคม 2568
Vol. 21, No. 2, July – December 2025
ISSN 1685-8573

รศ.ดร.อรุณี อินทรไพโรจน์	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
รศ.ดร.สุพจน์ นิตย์สุวัฒน์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
รศ.ดร.สุมิตรา นวลมีศรี	มหาวิทยาลัยราชภัฏสวนสุนันทา
รศ.ดร.พณณา ตั้งวรรณวิทย์	มหาวิทยาลัยราชภัฏเพชรบูรณ์
ผศ.ดร.อุไรวรรณ อินทร์แหยม	มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี
ผศ.ดร.เมธิญาณินฆ์ คำขาว	มหาวิทยาลัยเทคโนโลยีราชมงคลพระนคร
อ.ดร.อลิสา คงทน	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
ดร.ชูชาติ หฤไชยศักดิ์	ศูนย์อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC)

ฝ่ายประสานงานและฝ่ายจัดการ

ผศ.ดร.วัชรีวรรณ จิตต์สกุล	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
นางสาวพรพิมล ฝ้ายเทศ	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

กำหนดออก: ออกเผยแพร่ปีละ 2 ฉบับ

ฉบับที่ 1 เดือนมกราคม - มิถุนายน (บทความภาษาไทย)

ฉบับที่ 2 เดือนกรกฎาคม - ธันวาคม (บทความภาษาอังกฤษ)

กำหนดการรับและพิจารณาบทความ

รับพิจารณาบทความอย่างต่อเนื่อง

วัตถุประสงค์

1. เพื่อเผยแพร่ผลงานวิจัย/พัฒนาและผลการวิชาการด้านเทคโนโลยีสารสนเทศและวิทยาการคอมพิวเตอร์ ของคณาจารย์ เจ้าหน้าที่ และนักศึกษาของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือและสถาบันระดับอุดมศึกษาอื่น ๆ รวมทั้งนักวิจัย/พัฒนาจากหน่วยงานต่าง ๆ ทั้งภาครัฐและเอกชนภายในประเทศ
2. เพื่อเป็นศูนย์รวมแลกเปลี่ยนและกระจายองค์ความรู้ในนวัตกรรมใหม่ๆเกี่ยวกับเทคโนโลยีสารสนเทศ และวิทยาการคอมพิวเตอร์
3. เพื่อเป็นศูนย์รวมข่ายงานให้นักเทคโนโลยีสารสนเทศและวิทยาการคอมพิวเตอร์ได้เสนอแนวความคิดผลงานวิจัย/พัฒนาต่าง ๆ อันจะเป็นประโยชน์ต่อสถาบันและประเทศชาติ

เว็บไซต์

https://ph01.tci-thaijo.org/index.php/IT_Journal

อีเมล

itjournal@it.kmutnb.ac.th

เจ้าของ

คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
1518 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพฯ 10800
โทรศัพท์ 02-555-2726, 02-555-2704

บรรณาธิการแถลง

วารสารเทคโนโลยีสารสนเทศ มจพ. เป็นวารสารที่เผยแพร่ผลงานทางวิชาการ โดยรับบทความวิจัยและบทความวิชาการทั้งภาษาไทยและภาษาอังกฤษ เกี่ยวกับเทคโนโลยีสารสนเทศ ซึ่งบทความจะต้องได้รับการประเมินโดยผู้ทรงคุณวุฒิภายในหรือภายนอกมหาวิทยาลัย จำนวน 3 ท่าน โดยกระบวนการพิจารณานี้ผู้ทรงคุณวุฒิไม่ทราบข้อมูลของผู้ส่งบทความ (Double Blinded) และบทความจะต้องมีความคิดริเริ่มสร้างสรรค์ คุณค่าทางวิชาการ ความสมบูรณ์ของเนื้อหา และโครงสร้าง ภาษาที่ใช้ ความชัดเจนของสมมติฐาน/วัตถุประสงค์ ความชัดเจนของการนำเสนอและการจัดระเบียบบทความ ความถูกต้องทางวิชาการ การอภิปรายผล และการอ้างอิงที่ถูกต้องตามหลักวิชาการ ในนามของกองบรรณาธิการวารสารเทคโนโลยีสารสนเทศ มจพ. ขอขอบคุณทุกท่านที่ให้ความสนใจส่งบทความเพื่อพิจารณาในวารสารเทคโนโลยีสารสนเทศ มจพ. ปีที่ 21 ฉบับที่ 2 เดือนกรกฎาคม - ธันวาคม 2568 ในฉบับนี้กองบรรณาธิการได้คัดสรรบทความที่น่าสนใจมาเพื่อนำเสนอให้ผู้ที่สนใจเกี่ยวกับเทคโนโลยีสารสนเทศทุกท่าน

(ผู้ช่วยศาสตราจารย์ ดร.ศักดิ์ชาย ตั้งวรรณวิทย์)
บรรณาธิการวารสารเทคโนโลยีสารสนเทศ มจพ.

สารบัญ	หน้า

Application of Geographic Information System for Mapping Population Exposure to Flood Hazards in Thailand

Puvadol Doydee

Received: March 12, 2024
Revised: August 16, 2024
Accepted: August 28, 2024

* Corresponding Author: Puvadol Doydee, E-mail: puvadol.d@ku.th

DOI: 10.14416/j.it/2025.v2.001

Abstract

The assessment of population exposure to flood hazards in urban riverine areas is crucial to flood risk response and mitigation in Thailand. This study employed a free, open-source QGIS associated with various spatial datasets e.g. administrative boundaries, census data, built-up areas, and flood hazard. The objective was to estimate the population's exposure to flood hazards in Thailand. The analysis focused at the provincial level and estimated the population's exposure to a 25-year flood event. The findings revealed that the percentage of the Thai population exposed to riverine flood hazards ranged from zero to 99.86% and was categorized into five severity levels. Approximately 18.10 million Thai people (25.83%) dwell along rivers that are highly vulnerable to riverine floods. Nakhon Pathom province was the first highest risk of its population being exposed to riverine floods specifically nearby the Tha Cheen River. Concurrently, there were 8 provinces namely; 1) Nonthaburi, 2) Sing Buri, 3) Phra Nakhon Si Ayutthaya, 4) Samut Songkhram, 5) Ang Thong, 6) Pathum Thani, 7) Bangkok and 8) Samut Sakhon were determined as having the highest vulnerabilities to riverine floods, while Phangnga, Krabi and Phuket showed the lowest vulnerabilities. The findings of this study provide valuable insights for policymakers to facilitate preparedness and improve effective strategies to mitigate the flood hazards.

Keywords: Spatial Data Analysis, Open-Source QGIS, Flood Hazards, Population Exposure, Thailand.

1. Introduction

Floods are a natural phenomenon with the potential for devastating loss of life, social and economic disruption, outbreaks of infectious diseases, water and food insecurity, and adverse effects on mental health [1], [2]. Floods and hydrometeorological hazards are the most frequent causes of devastation globally [2], [3], [4]. Floods are among the most catastrophic natural disasters [1], [3], [5]. The rapidity of the havoc floods can cause underscores the serious threat that floods pose to infrastructure, communication systems, crops, livestock, and natural ecosystems [6], [7]. For example, Thailand's devastating 2011 flood shown resulted in numerous casualties which affected millions of people, caused extensive economic loss [8], [9].

Thailand has been highly exposed to and vulnerable to flood risks due to its riverine geography. These phenomena are occurring more frequently due to climate change and the increase in impervious surface area. Thailand is in a region prone to an annual monsoon season. During its rainy season from May to October, the country experiences upwards of 88% of its yearly rainfall [10], [11]. Particularly, riverine floods pose a major threat to agricultural production and urban areas located near principal rivers. However, various climatic and non-climatic dynamics give rise to different types of floods such as riverine, flash, urban, glacial-lake outburst, and coastal floods. This study focuses on riverine floods, as the latest spatial dataset of excessive rainfall, which caused main rivers to exceed their capacity.

** Department of Agriculture and Resources, Faculty of Natural Resources and Agro-Industry, Kasetsart University Chalermphrakiat Sakon Nakhon Province Campus.*

Mapping population exposure flood hazards is an essential tool for disaster risk management, especially in the issue of climate change [12], [13]. It helps identify location and situation of population exposure and provides a better understanding of spatial characteristics, which have not yet been studied before. In This study, Geographical Information Systems (GIS) have been applied to map population exposure to flood phenomena with the use of a free, open-source QGIS toolkit. Again, population exposure for this paper refers to the elements of a community affected by a flood such as, population, shelter, amenability and support services [14], [15]. The extent of flood hazards, population and their built environment were chosen to characterize population exposure [16].

2. Theoretical background and related researches

2.1 GIS data models

Data model define how real-world of land use and land cover features are represented in a GIS [17], [18]. Geospatial data are basically represented by two main structures; vector and raster datasets. The representation of real-world spatial information in vector and raster format is shown in Figure 1. The vector data model consists of points, lines, and polygons created using beginning and ending nodes and intervention vertices, each with detailed x,y coordinate information. The raster data model stores the spatial information in a user-defined grid where every pixel has a unique geographic location and attribute value.

2.1.1 Vector Data

Vector maps represent the most common form of thematic maps e.g. road atlases, GPS navigation devices, and Internet mapping engines display vector maps. The vector data structure uses points, lines, and polygons to represent spatial features. Points object are assigned using simple x, y coordinates; lines consist of connected x,y coordinates called nodes or vertices,; and polygons (areas) are simply enclosed lines where the beginning and ending nodes have the same coordinate value (Figure 1b)

2.1.2 Raster Data

Raster data have a more simple data structure than vector data due to raster data are represented in space by an array or grid of cell (Figure 1c). A raster layer contains cells arranged in rows and columns. Each cell contains a value or digital number that describes the phenomena being examined. The cell in the array are commonly called pixel, and they are usually square (i.e. the sides of a pixel have the same dimension) and in remote sensed raster data called spatial resolution. Unlike the vector data model, the raster data model has remained relatively unchanged since its inception.

Raster data area especially useful to describe spatial phenomena that vary continuously across the landscape. Such phenomena include elevation above sea level, precipitation [10], [11], temperature, etc. [12], [13].

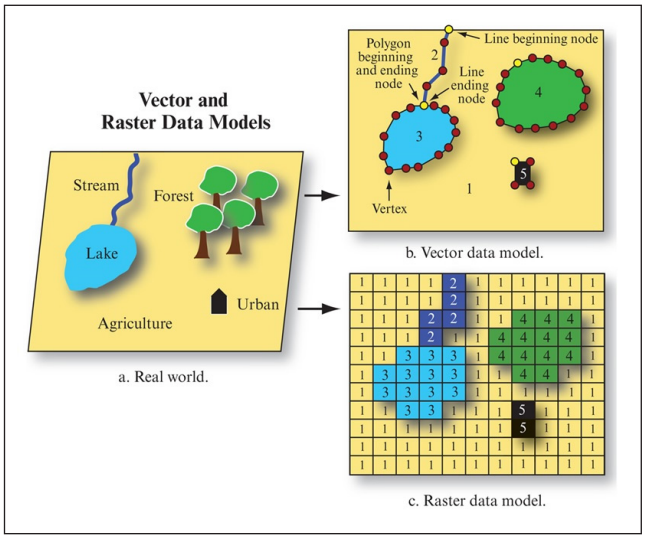


Figure 1. *Features in the real world can be represented using vector and raster data model. a) The real world. b) The real world portrayed in vector format. c) The real world portrayed in raster format [18].*

2.2 QGIS for mapping population exposure to flood hazards

This study used QGIS, a freely, open-source geographic information system software, to map population exposure to flood hazards in Thailand. QGIS facilitated the creation of flood exposure maps by identifying areas susceptible to flooding and the population residing in those areas.

The percentage of the population exposed to flood hazards for each province was calculated using the total population of the assessed area. The spatial data flow and the geospatial analysis conducted using the QGIS was illustrated (Figure 2). The flowchart showed the sequential steps involved in the processing and analysis of spatial data, highlighting the key stages and relationships within GIS framework.

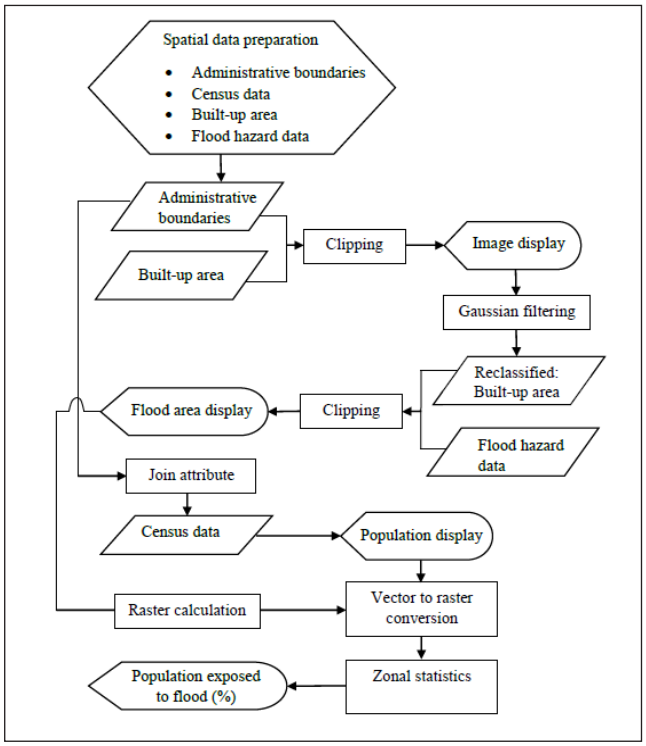


Figure 2. *Flowchart of spatial data preparation and geospatial techniques and analysis using QGIS.*

A Gaussian filter, which assigned weight to the pixel based on their spatial distribution, was utilized to estimate population density [19]. The population density map depicted higher density in central pixels compared to those in the corners of the grid. The population within each administrative boundary was calculated by multiplying the estimated population density with the census data. Riverine flood analysis within each administrative boundary was performed using raster data model of flood hazard.

The percentage of population that was exposed to flood hazards was determined through a comprehensive analysis applying open-sources datasets such as the Humanitarian Data Exchange, Copernicus Global Land Cover, the United Nations Environment Programme Global Resource Information Database (UNEP-GRID) associated with various geospatial techniques. The land cover data with spatial resolution of 100 m x 100 m was classified into two categories namely, built-up areas and non-built-up areas. Population density pixels were derived from administrative boundaries and census data, while considering their proximity to accumulated built-up areas. Pixels with higher population density were assigned greater weight in the population density calculations compared to those with lower population densities.

In addition, there are various papers that using QGIS techniques for monitoring and mapping flood disaster and related researches as follows:

Ariyani et al. [20] to assess flood risk in the Bangkok and Masjid watersheds at Riau, Indonesia. Flood hazard map was arranged by slope, land cover, elevation, rainfall, buffer zone, and soil type, which is done with the help of QGIS. The tool integrated with satellite spatial data and classified the vulnerability level being very high, high, medium, low, and very low.

Musunura and Marshall [21] calculated percentage of the population exposed to flood hazard using QGIS case study in Lao People's Democratic Republic. Various open source dataset such as Copernicus Global Land Cover data, Census data and Administrative Boundaries data and Flood Hazard data were imported to QGIS to construct the flood disaster thematic maps.

Samsudin et al. [22] evaluated the impact of dam break scenario under probable maximum precipitation (PMF) condition at Puah hydropower dam using QGIS and open source data for emergency preparedness and early warning systems focusing on rescue work if disaster occurs.

Renjith et al. [23] applied QGIS for delineating the flood prone areas and to create a flood risk map for the Pathanamthitta District of Kerala.

3. Research Methodology

3.1 Spatial data preparation

This study conducted a provincial based mapping of population exposure to flood disasters in Thailand utilizing publicly available datasets. The GIS datasets were obtained

from global spatial datasets such as the Humanitarian data exchange and Copernicus global land service [24]. The land cover data provided pixel-level information with a spatial resolution of 100 m x 100 m, representing land areas of 10,000 square meters of each pixel. A global risk data platform was used in conjunction with various geospatial techniques [25].

The administrative boundary with population attribute dataset of Thailand, latest version on January 27, 2022, was incorporated to obtain spatial and geodatabase. Census data in the form of a *.CSV file (Excel) was utilized with the attribute data most updated on April 20, 2021. This information was derived from the sources as indicated in Table 1. It enabled the calculation of the total population within administrative boundaries and facilitated an assessment of population exposure to flood hazards associated with the latest built-up area data in year of 2019.

The percentage of the population exposed to flood hazards was estimated using flood hazard data from UNEP GRID. The flood hazard assessment employed a probabilistic approach, modeling riverine floods from major river basins in Thailand.

This study adopted a 25-year return period for flood hazards [26], [27] due to in community and urban planning, a 25-year return period is often used as a standard for designing flood protection measures. It represents a compromise between frequent smaller floods and very rare, extreme events. This approach helps in creating flood resilience systems that can handle reasonably expected flooding without being excessively conservative or costly.

Hazard maps were developed at a spatial resolution of 1 km x 1 km and were validated against satellite flood footprints from various sources, including the Dartmouth Flood Observatory (DFO) and The United Nations Satellite Centre (UNOSAT) flood portal. These maps coincided with significant events of the 2011 Thailand flood.

3.2 GIS data processing

QGIS was employed for viewing, editing, and analysis of geospatial data [28]. The administrative boundary data were imported into QGIS and analyzed at the provincial level.

Table 1. *Selected GIS spatial open source dataset and related hyperlinks.*

Spatial data	Name	Open Source Dataset
Administrative boundaries	Humanitarian Data Exchange	https://data.humdata.org/dataset/cod-ab-tha
Census data	Humanitarian Data Exchange	https://data.humdata.org/dataset/cod-ps-tha
Built-up area	Copernicus Global Land Service	https://lcviewer.vito.be/download
Flood hazard	Global Risk Data Platform	https://preview.grid.unep.ch/

This spatial data was represented as vector data model consisting of polygons representing GIS features with the WGS84 map projection that encompassed the entire terrestrial area of Thailand.

The built-up area data was shown as a raster data model, comprised a grid of cells or pixels with assigned values representing information. Built-up areas were assigned a digital number (DN) value of 100 (bright color), while non-built-up areas were assigned a DN value of 0 (dark color). These datasets were overlaid with the Thailand administrative boundaries. Geoprocessing techniques such as merge and clip operations were used to combine four raster files into a single built-up area file, then clipped to fit within the boundaries of Thailand. To estimate the population density for each provincial area, a Gaussian filter was applied to the built-up area layer. This filter facilitated the transformation of pixel population density, taking into consideration the characteristics of each area. Population density tends to be higher in built-up areas, such as cities or urban areas with limited land availability, compared to non-built-up areas or rural zones with more abundant land.

The accumulated built-up areas with the highest population density were situated at the center of the grid, while areas farther away from the administrative boundaries were assigned lower values. The built-up layer was reclassified into two categories namely, 1) built-up areas with a DN value of 1 and 2) non-built-up areas with a DN value of 0.

3.3 Flood hazard mapping

The flood inundation layer was imported into GIS and clipped with the boundary of the built-up area using geo-processing

techniques. The census dataset was layered and integrated into the Thailand administrative boundaries. An attribute join was performed using a primary key to integrate the population data from the census spatial database into the attribute of the administrative boundaries.

To distribute the census data within the estimated pixels, either rasterization or vector to raster conversion techniques were employed, resulting in the generation of a population raster. The raster calculator function was utilized to generate the flood hazard calculation layer.

Zonal statistics analysis was conducted to generate the population exposed to the flood hazard map. This analysis determined the percentage of the population exposed to flood hazards for each administrative boundary, utilizing the total population of the assessed area.

4. Results and Discussions

4.1 Population exposed to flood hazards

In 2022, Thailand's population was estimated to be 70.08 million people [29]. In this study, as latest (2022) flood open source dataset available, it was found that in Thailand around 19 million people (27.74% of the population) located in nine of the seventy seven provinces were exposed to a 25-year flood event. This exposure level was classified as extremely high in severity level (SL), ranging from 79.17% to 99.86%. Among these provinces, Nakhon Pathom had the highest risk with 99.86% (Figure 3a) of its population being exposed to river floods specifically, the Tha Cheen River [30]. Though this seems to align with the newspaper and television accounts of the 2022 Nakhon Pathom river flood, no academic studies, in either Thai or English, have been published which analyzed the impacts of the flood that would either confirm or refute this.

Rivers, streams and riverine communities in Thailand, as they are worldwide, are prone to flooding and adversely impacted due to heavy precipitation or tropical storms [31], [32], [13]. These areas are more susceptible to flood hazards, and it is observed that the exposed population tends to reside in flood-prone zones [6], [7], [21].

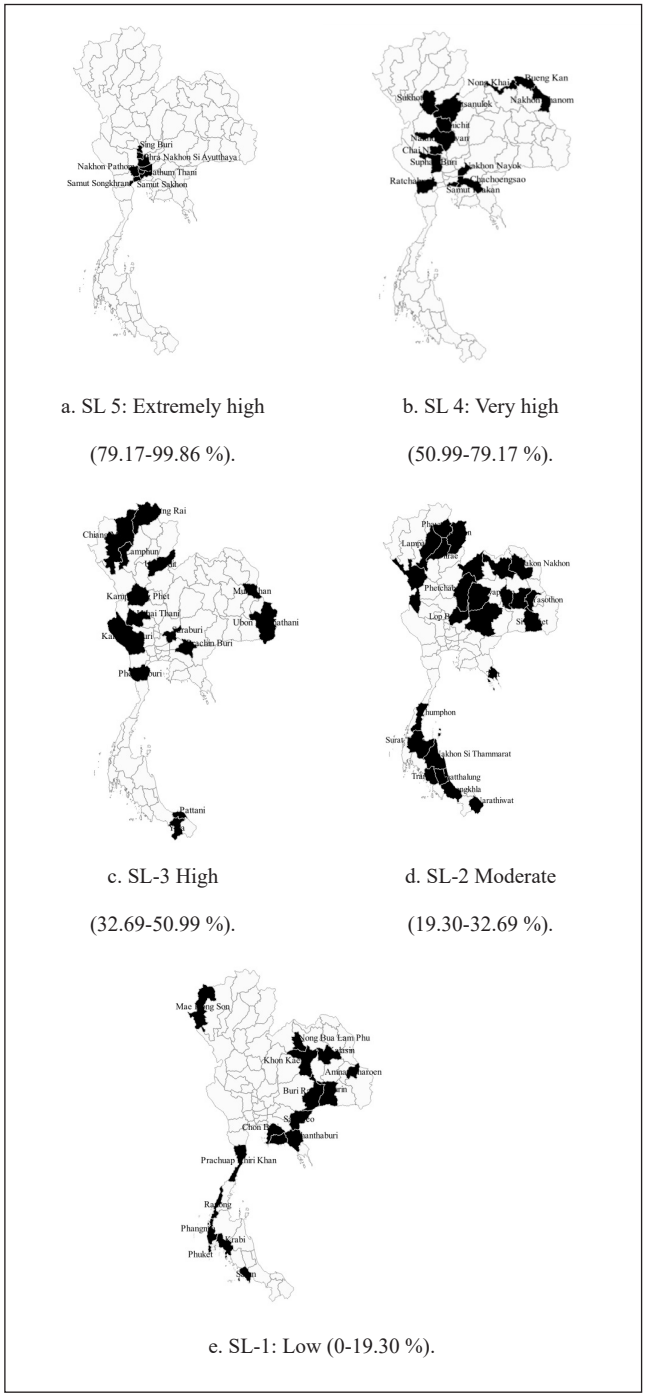


Figure 3. *Maps display flood severity level (SL) in Thailand at the provincial level. a. SL-5: Extremely high (79.17-99.86%). b. SL-4: Very high (50.99-79.17%). c. SL-3: High (32.69-50.99%). d. SL-2: Moderate (19.30-32.69%). e. SL-1: Low (0-19.30%).*

In Thailand, these nine provinces, Nakhon Pathom, Nonthaburi, Sing Buri, Phra Nakhon Si Ayutthaya, Samut Songkhram, Ang Thong, Pathum Thani, Bangkok and Samut Sakhon exhibited an SL-5 flood hazard exposure level classified as "Extremely high," ranging from 79.17% to 99.86% (Figure 3a,

Figure 4). Additionally, these thirteen provinces, Nong Khai, Chai Nat, Phichit, Nakhon Phanom, Sukhothai, Chachoengsao, Bueng Kan, Phitsanulok, Suphan Buri, Ratchaburi, Nakhon Sawan, Samut Prakan and Nakhon Nayok were classified as SL-4 (Very high) with exposure ranging from 50.99% to 79.17% (Figure 3b, Figure 4).

Fourteen provinces, Uttaradit, Kanchanaburi, Kamphaeng Phet, Mukdahan, Prachin Buri, Pattani, Phetchaburi, Lamphun, Chiang Mai, Chiang Rai, Ubon Ratchathani, Yala, Saraburi and Uthai Thani were designated as SL-3 (High) with exposure levels ranging from 32.69% to 50.99% (Figure 3c, Figure 4).

Twenty-four provinces, Surat Thani, Lampang, Narathiwat, Trang, Yasothon, Phayao, Roi Et, Phrae, Tak, Songkhla, Chumphon, Chaiyaphum, Lop Buri, Udon Thani, Trat, Nakhon Si Thammarat, Maha Sarakham, Loei, Nakhon Ratchasima, Nan, Phatthalung, Sakon Nakhon, Si Sa Ket and Phetchabun were categorized as SL-2 (Moderate) with exposure levels ranging from 19.30% to 32.69% (Figure 3d, Figure 4).

Finally, seventeen of Thailand's provinces exhibited an SL-1 (Low) flood hazard exposure, ranging from 0% to 19.30%. These were Nong Bua Lam Phu, Chanthaburi, Khon Kaen, Mae Hong Son, Kalasin, Surin, Rayong, Satun, Sa Kaeo, Prachuap Khiri Khan, Buri Ram, Ranong, Chon Buri, Amnat Charoen, Phangnga, Krabi and Phuket (Figure 3e, Figure 4). In addition, Phuket is not significantly affected by riverine floods, but still be susceptible to coastal floods or tsunamis [33], [34].

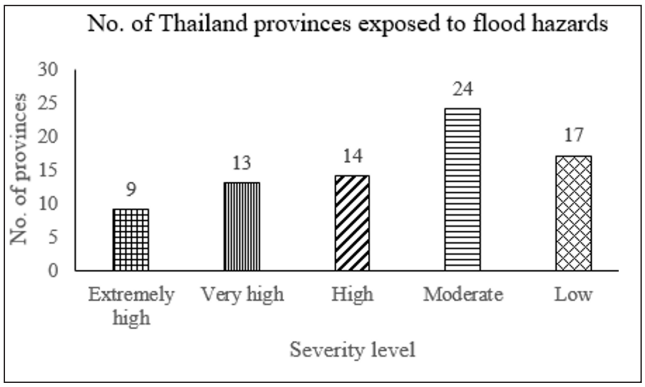


Figure 4. *Number of provinces exposed to flood hazards.*

Population exposure was categorized into five distinct severity level, ranging from low to extremely high [35]. These levels of exposure are visually represented through a color scheme, wherein variations in color intensity signify varying degrees of vulnerability. Specifically, darker shades correspond to higher levels of exposure to riverine flood hazards and shown in Figure 5.

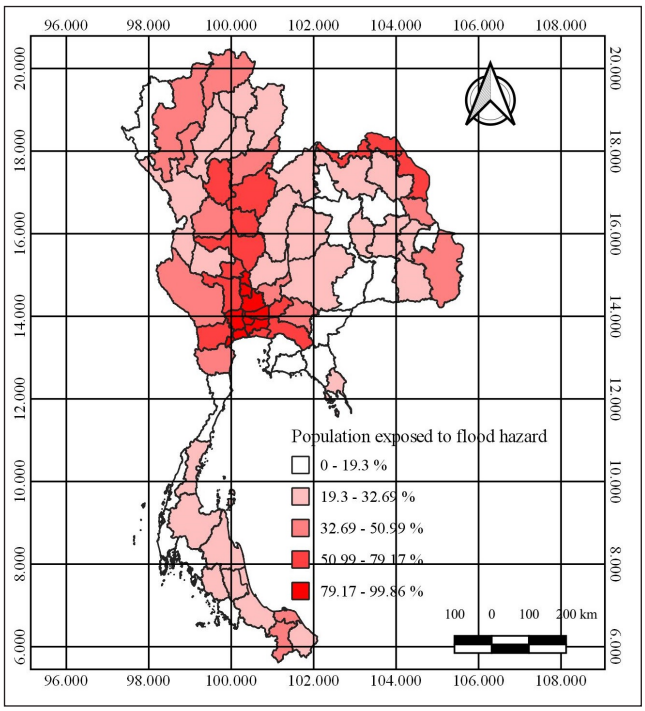


Figure 5. *Map showing the percentage of population exposure to flood hazards in Thailand.*

4.2 Flood risks and impacts

Annually, global flooding affects approximately 350 million people [36]. Extreme flood events, such as the 2011 floods, have impacted over 13 million people in Thailand [37]. Urbanization and the expansion of impervious infrastructure contribute to increased flood severity level. In Thailand, agricultural lands with high chemical fertilizer usage can contaminate nearby water bodies during floods. Wu et al. [38] reported the role of precipitation and flood events in transporting microplastics, which pose risks to freshwater ecosystems and the food chain.

Thus, there is a need for a holistic approach to river management by considering both flood prevention and the preservation of river habitats [39]. Floods have both positive and negative impacts, such as creating fish habitats and increasing groundwater depth [40], but also causing landslides, crop and livestock losses, and waterway pollution [41] - [44].

Floods not only cause physical damage but also have mental and social implications, particularly in challenging circumstances like the COVID-19 pandemic [45]. In the recovery stage, it is essential for local governments to consider flood insurance payouts to support affected farmers [46]. Flood hazard maps can aid local governments, national and international organizations in reducing flood risk, obtaining financial support and the planning of flood shelters [47].

Recent research suggests that global extreme precipitation and changes in land use contribute significantly to flood damage. Despite data challenges, global datasets have proven valuable in hydrological and climate impact studies [31]. Remote sensing technique was importance of measuring and assessing flood impacts for mitigation practices [48]. Given the significance of flooding in Thailand, a community-based approach is crucial for effective flood management and preparedness, involving all stakeholders in the planning and implementation of flood mitigation strategies to secure national infrastructure in the future [7].

The limitation of the study is the reliance on spatial datasets and free, open-source software, which may not always provide the most up-to-date or comprehensive data. Therefore, it is important to consider commercial spatial datasets and ArcGIS software. This could affect the accuracy of exposure estimates and flood risk categorization. Furthermore, the study's focus on a specific flood return period (25 years) might not account for more extreme or less frequent events, potentially underestimating risk in some areas. The variation in flood exposure across provinces highlights the need for more granular data and localized analysis to enhance the precision of risk assessments. Despite these limitations, the findings offer valuable insights for policymakers, emphasizing the need for targeted flood preparedness and mitigation strategies tailored to the specific vulnerabilities of different provinces. This study provides a foundation for improving flood risk management by identifying high-risk areas and informing more effective flood response strategies.

5. Conclusions

The findings of this study offer a benefit of information technology (IT) such GIS approach to map population exposure to flood risk, serving as a valuable reference for national-level flood risk management, prevention and disaster reduction. Floods pose threats to food security, nutrition, and the livelihoods of vulnerable communities leading to reduced food production.

In conclusion, GIS proves to be a robust tool for mapping population exposure to flood hazards in Thailand. By overlaying flood hazard and population distribution data on a geographic base map. GIS enables the identification of high-risk flood areas and vulnerable populations. The resulting maps inform decision-makers and aid in the development of mitigation strategies to minimize flood impacts on the population. These study findings facilitate a comprehensive understanding of flood risks in Thailand and support the implementation of measures to mitigate its effects. The utilization of GIS for mapping population exposure to flood hazards represents a crucial advancement in ensuring the safety and well-being of Thailand's population.

Legislators and the officers in charge should conduct public awareness campaigns regarding community preparedness and education related to organized early warning systems, emergency response, and improving flood defenses. Nature-based solutions, such as wetland restoration, reforestation, and creating floodplains to absorb excess water, are recommended for incorporation in high-risk communities.

6. Acknowledgement

We express our gratitude to the Office of Chalermphrakiat Sakon Nakhon Province Campus for generously providing the necessary IT facilities for this research. Additionally, we would like to acknowledge the support received from the Advanced Institute on Knowledge-Based Actions for Disaster Risk Reduction (AI-KBA), Taipei, Taiwan, and the International Council for Science, which partially funded this study.

7. References

[1] G. Singh and A. Pandey. "Flash flood vulnerability assessment and zonation through an integrated approach in the Upper Ganga Basin of the Northwest Himalayan region in Uttarakhand." *International Journal of Disaster Risk Reduction,* Vol. 66, pp. 1-18, December, 2021.

[2] Intergovernmental Panel on Climate Change (IPCC). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge and New York, 2012.

[3] United Nations International Strategy for Disaster Reduction (UNISDR). *Global assessment report on disaster risk reduction 2015.* United Nations Office for Disaster Risk Reduction. Geneva, Switzerland, 2015.

[4] J. Sahani, P. Kumar, S. Debel, C. Spyrou, M. Loupis, L. Aragão, F. Porcu, M.A.R. Shah, and S.D. Sabatino. "Hydro-meteorological risk assessment methods and management by nature-based solutions." *Science of the Total Environment,* Vol. 696, pp. 1-17, December, 2019.

[5] A. Goffi, D. Stroppiana, P.A. Brivio, G. Bordogna, and M. Boschetti. "Towards an automated approach to map flooded areas from Sentinel-2 MSI data and soft integration of water spectral features." *International Journal of Applied Earth Observation and Geoinformation,* Vol. 84, pp. 1-14, February, 2020.

[6] B. K. Osei, I. Ahenkorah, A. Ewusi, and E. B. Fiadonu. "Assessment of flood prone zones in the Tarkwa mining area of Ghana using a GIS-based approach." *Environmental Challenges,* Vol. 3, pp. 1-12, April, 2021.

[7] I. Pal, P. Doydee, T. Utarasakul, P. Jaikaew, K. A. B. Razak, G. Fernandez, T. Huang, and C. S. Chen. "System approach for flood vulnerability and community resilience assessment at the local level - a case study of Sakon Nakhon province, Thailand." *Kasetsart Journal of Social Sciences,* Vol. 42, No. 1, pp. 107-116, January-March, 2021.

[8] United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP). *Overview of Natural Disasters and Their Impacts in Asia and the Pacific, 1970-2014.* Economic and Social Commission for Asia and the Pacific, Bangkok, Thailand, 2015.

[9] H. H. Loc, E. Park, D. Chitwatkulsiri, J. Lim, S. H. Yun, L. Maneechot, and D. M. Phuong. "Local rainfall or river overflow? Re-evaluating the cause of the Great 2011 Thailand flood." *Journal of Hydrology,* Vol. 589, pp. 1-9, October, 2020.

[10] H. Stoklosa, C. J. Burns, A. Karan, M. Lyman, N. Morley, R. Tadee, and E. Goodwin. "Mitigating trafficking of migrants and children through disaster risk reduction: Insights from the Thailand flood." *International Journal of Disaster Risk Reduction,* Vol. 60, pp. 1-14, June, 2021.

[11] W. Pratoomchai, C. Ekkawatpanit, N. Yoobanpot, K. T. Lee. "A Dilemma between Flood and Drought Management: Case Study of the Upper Chao Phraya Flood-Prone Area in Thailand." *Water,* Vol. 14, No. 24, pp. 1-15, December, 2022.

[12] S. F. Balica, I. Popescu, L. Beevers, and N. G. Wright. "Parametric and physically based modelling techniques for flood risk and vulnerability assessment: a comparison." *Environmental Modelling & Software,* Vol. 41, pp. 84-92, March, 2013.

[13] G. M. Membele, M. Naidu, and O. Mutanga. "Examining flood vulnerability mapping approaches in developing countries: A scoping review." *International Journal of Disaster Risk Reduction,* Vol. 69, pp. 1-25, February, 2022.

[14] H. de Moel, J. C. J. H. Aerts, and E. Koomen. "Development of flood exposure in the Netherlands during the 20th and 21st century." *Global Environmental Change,* Vol. 21, No. 2, pp. 620-627, May, 2011.

[15] E. E. Koks, B. Jongman, T. G. Husby, and W. J. W. Botzen. "Combining hazard, exposure and social vulnerability to provide lessons for flood risk management." *Environmental Science & Policy,* Vol. 47, pp. 42-52, March 2015.

[16] E. Tate, M. A. Rahman, C. T. Emrich, and C. C. Sampson.

"Flood exposure and social vulnerability in the United States." *Natural Hazards,* Vol. 106, pp. 435-457, January, 2021.

[17] P. V. Bolstad and J. L. Smith. "Errors in GIS assessing spatial data accuracy." *Journal of Forestry,* Vol. 90, pp. 21-29, 1992.

[18] J. R. Jensen and R. R. Jensen. *Introductory Geographic Information System.* Pearson Education, Inc, Lake Ave., Glenview, IL.U.S.A., 2013.

[19] Y. Mei, Z. Gui, J. Wu, D. Peng, R. Li, H. Wu, and Z. Wei. "Population spatialization with pixel-level attribute grading by considering scale mismatch issue in regression modeling." *Geo-spatial Information Science,* Vol. 25, No. 3, pp. 365-382, July, 2022.

[20] D. Ariyani, A. K. Balqis, D. Abdaa, R. N. Arini, A. P. Dewi, and S. P. KT. "Flood hazard mapping using QGIS spatial analysis in Bangko and Masjid watershed at Riau, Indonesia." *Journal Pengelolaan Sumberdaya Alam dan Lingkungan (JPSL),* Vol. 13, No. 3, pp. 362-371, 2023.

[21] A. Musunuru and A. Marshall, *Mapping population exposure to flood hazards: step by step guide on the use of QGIS.* Available Online at https://repository.unescap.org/handle/20.500.12870/3898, accessed on 6 March 2023.

[22] S. H. Samsudin, N. A. Zuhaily, A. Setu, R. S. Muda, and M. F. M. Amin. "Dam break flood mapping and analysis using Open Source GIS Tool and Data." *IOP Conference Series: Earth and Environmental Science,* Vol. 1167, pp. 1-9, 2023.

[23] A. Renjith, A. Jossy, A. Mary-S, A. Xaviour, and A. Baby. "Flood mapping in Pathanamthitta district using QGIS software." *International Journal of Engineering Research & Technology (IJERT),* Vol. 10, No. 06, pp. 136-139, 2022.

[24] M. Buchhorn, B. Smets, L. Bertels, B. D. Roo, M. Lesiv, N. E. Tsendbazar, M. Herold, and S. Fritz. "Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2019: Globe (V3.0.1) [Data set]." *Zenodo,* 2020.

[25] P. Bolstad. *GIS Fundamentals: A First Text on Geographic Information Systems.* 5th ed, Eider Press, White Bear Lake, Minnesota; 2016.

[26] Royal Irrigation Department (RID). *Riverine flood prevention and mitigation plan (rainy season) 2022.* Bureau of Water Management and Hydrology, Royal Irrigation Department, Ministry of Agriculture and Cooperatives, Bangkok, Thailand, 2022.

[27] Q. Din, A. Rashid, A. Rahim, and I. Ullah. "Flood risk assessment of the population in Afghanistan: A spatial analysis of hazard, exposure, and vulnerability." *Natural Hazards Research,* Vol. 4, No. 1, pp. 46-55, March, 2024.

[28] QGIS Development Team, *QGIS Geographic Information System. Open Source Geospatial Foundation Project.* Available Online at http://qgis.osgeo.org, accessed on 22 July 2022.

[29] Statista, *Total population in Thailand from 2018 to 2028.* Available Online at https://www.statista.com/statistics/331 889/total-population-of-thailand, accessed on 12 November 2023.

[30] S. Rantasewee, P. Teerapunyapong, A. Rittima, K. Surakit, Y. Phankamolsil, A. S. Tabucanon, W. Sawangphol, J. Kraisangka, and Y. Talaluxmana. "Impacts of the 2011 Thailand flood on groundwater recharge potential in flood retention area in the middle reach of Tha Chin River." *Engineering Access,* Vol. 8, No. 2, pp. 186-191, July-December, 2022.

[31] N. S. Fernando, S. Shrestha, S. KC, and S. Mohanasundaram. "Investigating major causes of extreme floods using global datasets: A case of Nepal, USA & Thailand." *Progress in Disaster Science,* Vol. 13, pp. 1-14, January, 2022.

[32] H. Yolina, Rojali, and E. Irwansyah. "Development of flood-prone area classification program using linear classifier method based on geomorphic flood index and land cover." *Procedia Computer Science,* Vol. 216, pp. 396-405, 2023.

[33] K. Pakoksung, P. Latcharote, P. Suttinon, P. Bunditsakulchai, A. Suppasri, and F. Imamura. "The probability of community-scale building damage and economic loss in Thailand increased after the 2004 Indian Ocean tsunami." *International Journal of Disaster Risk Reduction,* Vol. 79, pp. 1-22, September, 2022.

[34] Z. Zhang, A. B. Kennedy, and J. P. Moris. "Tsunami wave loading on a structural array behind a partial wall." *Coastal Engineering,* Vol. 179, pp. 1-12, January, 2023.

[35] L. A. Garcia, M. L. Martinez-Chenoll, I. Escuder-Bueno, and A. Serrano-Lombillo. "Assessing the impact of uncertainty on flood risk estimates with reliability analysis using 1-D and 2-D hydraulic models." *Hydrology and Earth System Sciences,* Vol. 16, No. 7, pp. 1895-1914, July, 2012.

[36] S. Jonkman. "Global perspectives on loss of human life caused by floods." *Natural Hazards,* Vol. 34, No. 2, pp. 151-175, February, 2005.

[37] M. Haraguchi and U. Lall. "Flood risks and impacts: A case study of Thailand's floods in 2011 and research questions for supply chain decision making." *International Journal of Disaster Risk Reduction,* Vol. 14, Part 3, pp. 256-272, December 2015.

[38] J. Wu, Z. Jiang, Y. Liu, X. Zhao, Y. Liang, W. Lu, and J. Song. "Microplastic contamination assessment in water and economic fishes in different trophic guilds from an urban water supply reservoir after flooding." *Journal of Environmental Management,* Vol. 299, pp. 1-11, December, 2021.

[39] H. J. Hung, W. C. Lo, C. N. Chen, and C. H. Tsai. "Fish's habitat area and habitat transition in a river under ordinary and flood flow." *Ecological Engineering,* Vol. 179, pp. 1-18, June, 2022.

[40] X. Wang, G. Zhang, and Y. J. Xu. "Impacts of the 2013 extreme flood in Northeast China on regional groundwater depth and quality." *Water,* Vol. 7, No. 8, pp. 4575-4592, August, 2015.

[41] Y. Hong, R. F. Adler, A. Negri, and G.J. Huffman. "Flood and landslide applications of near real- time satellite rainfall products." *Natural Hazards,* Vol. 43, pp. 285-294, March 2007.

[42] A. Rahman and A. Khan. "Analysis of flood causes and associated socio- economic damages in the Hindukush region." *Natural Hazards,* Vol. 59, No. 3, pp. 1239-1260, December, 2011.

[43] F. L. Ogden, N. R. Pradhan, C. W. Downer, and J. A. Zahner. "Relative importance of impervious area, drainage density, width function, and subsurface storm drainage on flood runoff from an urbanized catchment." *Water Resources Research,* Vol. 47, No. 12, pp. 1-12, December, 2011.

[44] L. Hubbard, D. W. Kolpin, S. J. Kalkhoff, and D. M. Robertson. "Nutrient and sediment concentrations and corresponding loads during the historic June 2008 flooding in Eastern Iowa." *Journal of Environmental Quality,* Vol. 40, No. 1, pp. 166-175. January-February, 2011.

[45] J. Park, M. Son, and C. Park. "Natural disasters and deterrence of economic innovation: a case of temporary job losses by hurricane Sandy." *Journal of Open Innovation: Technology, Market, and Complexity,* Vol. 3, No. 1, pp. 1-23, March 2017.

[46] United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP). *The Disaster Riskscape Across Asia-Pacific.* United Nations Economic and Social Commission for Asia and the Pacific. Bangkok, Thailand, 2019.

[47] K. Uddin and M. A. Matin. "Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology." *Progress in Disaster Science,* Vol. 11, pp. 1-13, October, 2021.

[48] E. S. Hermas, A. Gaber, and M. Bastawesy. "Application of remote sensing and GIS for assessing and proposing mitigation measures in flood-affected urban areas, Egypt" *The Egyptian Egyptian Journal of Remote Sensing and Space Science,* Vol. 24, No. 1, pp. 119-130, February 2021.

A Comparison of Classification Methods of Hypothyroid Disease Prediction

Kulchaya Pongsawaeng*, Ausron Binmaduereh*, Panuphong Jenrotphondet*, and Orasa Patsadu*

Received: June 19, 2024
Revised: September 18, 2024
Accepted: September 23, 2024

* Corresponding Author: Orasa Patsadu, E-mail: orasa.p@mail.rmutk.ac.th

DOI: 10.14416/j.it/2025.v2.002

Abstract

This paper proposes a comparison of classification methods of hypothyroid disease prediction using data mining techniques. A dataset from the UCI repository with the thyroid disease dataset is used to prepare data with missing value handling, imbalance class handling, and suitable attribute selection. Then, the dataset is used to build the model by comparing the performance of classification methods such as Multilayer Perceptron, Support Vector Machine, and Decision Tree. The result shows that the Decision Tree achieves high performance with an accuracy of 99.61%, which is higher than the Multilayer Perceptron and Support Vector Machine with an accuracy of 96.46 % and 92.93%, respectively. In addition, we compared the result with state-of-the-art, which uses a similar technique to our proposed method. The result shows that our proposed method also outperforms previous research. Therefore, we decided to use Decision tree model for the prototype system development in hypothyroid disease prediction to support physicians' decision-making for diagnosis and treatment. Furthermore, this paper proposes data visualization to help users for primary risk assessment of a chance of hypothyroid disease to acknowledge risk before deciding to meet physicians using demographic information. Therefore, it will reduce the cost of medical and death rates.

Keywords: Hypothyroid Disease, Data Mining, Comparison of Classification Methods, Healthcare, Decision Support System.

1. Introduction

Thyroid disease is abnormal in the thyroid gland. Thyroid patients tend to increase, especially with hypothyroid disease. Most patients are women, who have an opportunity of thyroid disease more than men, especially females older than 40 years. Hypothyroid disease is divided into 2 states such as primary hypothyroidism and secondary hypothyroidism. The symptom indicates hypothyroid disease such as tiredness, pains, and aches, and gain weight. In addition, side effects from using certain drugs affect to hormone production of the thyroid gland such as cardiovascular drugs, mental conditions, and cancer disease. In addition, thyroidectomy affects to stop thyroid hormone production [1], [2]. Several researchers use classification techniques to solve medical [3] - [5]. In particular, research proposes a method for hypothyroid disease prediction using the classification method. Several classification techniques are used to build models for hypothyroid disease prediction, and each research focuses on model building [6] - [10]. According to literature reviews, most of the research also lacks a decision support system for primary risk self-assessment using data visualization. It became the motivation for our research. Therefore, our hypothesis is to propose a high-performance model for hypothyroid disease prediction and show data visualization for decision support in treatment.

This paper proposes a comparison of classification methods of hypothyroid disease prediction. The thyroid disease dataset from UCI is used to prepare data for model building by missing value handling, imbalance class handling, and appropriated attribute selection. Then, data is used to build the model by comparing the performance of classification methods such as Multilayer Perceptron, Support Vector Machine, and Decision Tree. In addition, our research proposes data visualization to help users for primary risk self-assessment to acknowledge primary risk before deciding to meet physicians

* Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Krungthep.

using demographic information to reduce medical costs and death rate.

The rest section of our research is section 2, which is related works. Section 3 is the research methodology. Section 4 is the results and case study. The last section is the conclusion.

2. Related Works

"Hypothyroidism is thyroid hormone production less than what is usual in the body, which causes hormone deficiency" [2], [11]. Several researchers proposed methods for hypothyroid disease detection. Olalekan, et al. [6] proposed a method for hypothyroid disease prediction using Ensemble learning with Bagging. There is comparative Ensemble learning by Bagging with Decision Tree and Bagging with SimpleCart. The result shows that the method for hypothyroid disease prediction using Bagging with Decision Tree achieves an accuracy of 99.60%. The accuracy of hypothyroid disease prediction using Bagging with SimpleCart is 99.55%.

Guleria, et al. [7] proposed a method to predict hypothyroidism using the classification method. Decision Tree, Random Forest, Naïve Bayes Multiclass classifier, and Deep Learning (ANN) are used to build the model. The result shows that Decision Tree and Random Forest achieve hypothyroidism prediction with an accuracy of 99.5758% and 99.3107%, respectively.

Naeem, et al. [8] proposed hypothyroidism disease prediction using the classification method. K-Nearest Neighbor, Naïve Bayes, and Support Vector Machine are used to build a model for hypothyroidism disease prediction. The result shows that Support Vector Machine achieves an accuracy of 84.72% when compared to other classifiers.

Kumar, et al. [9] proposed hypothyroid prediction using the classification method. Decision Tree and Neural Networks are used to build models for hypothyroid prediction. The result shows J48 can predict hypothyroid disease with an accuracy of 99.58%.

Gothane [10] proposed a method to predict hypothyroidism using the classification method. ZeroR classifier is used to build a model for hypothyroidism prediction. The result shows that this model can predict hypothyroidism with an accuracy of 92.2853 %.

Hongboonmee and Trepanichkul [12] proposed a comparative performance of classification to analyze risk factors that affect hypothyroid disease prediction with data mining techniques. There are 3 techniques: Artificial Neural Network, Decision Tree, and Naïve Bayes with a dataset from Phitsanulok Hospital. The result shows that Artificial Neural Network achieves an accuracy of 82.97%.

Parimala and Vadivu [13] presented a method to predict thyroid disease using Support Vector Machine and K-Nearest Neighbor. This model can predict accuracy.

Lim, et al. [14] proposed a method to detect thyroid disease using the classification method. The dataset is used to prepare data using featurewiz to select features. Decision Tree, K-Nearest Neighbor, Logistic Regression, Naïve Bayes, Support Vector Machine, and Ensemble machine learning algorithms (Random Forest and Extreme Gradient Boost) are used to build the model. The proposed model can detect thyroid disease with an accuracy of 99.45%.

Alyas, et al. [15] presented a method for thyroid disease prediction using Decision Tree, Random Forest, K-Nearest Neighbor, and Artificial Neural Networks. The result shows that Random Forest can predict thyroid disease with an accuracy of 94.8%.

Kurnaz, et al. [16] proposed a method for thyroid disease prediction using the classification method. The dataset is prepared using the Non-Sorting Genetic algorithm to select attributes for model building. Decision Tree, K-Nearest Neighbor, and Support Vector Machine are used to build a model for thyroid disease prediction. The result shows that this model can accurately detect thyroid disease.

Chaganti, et al. [17] presented a method for thyroid disease prediction using the classification method. The dataset is used to prepare for attribute selection and build the model using RF, GBM, ADA, LR, and SVM. The result shows that RF archives

thyroid disease prediction with an accuracy of 99% when compared to other methods.

Zhang, et al. [18] proposed a framework for comprehensively understanding thyroid disease using Association rule. The dataset is used to build a model with Apriori and the FP-Tree algorithms. The result shows that this model can create rules to detect thyroid disease.

3. Research Methodology

The model-building process consists of 3 processes to predict hypothyroid disease, as shown in Figure 1.

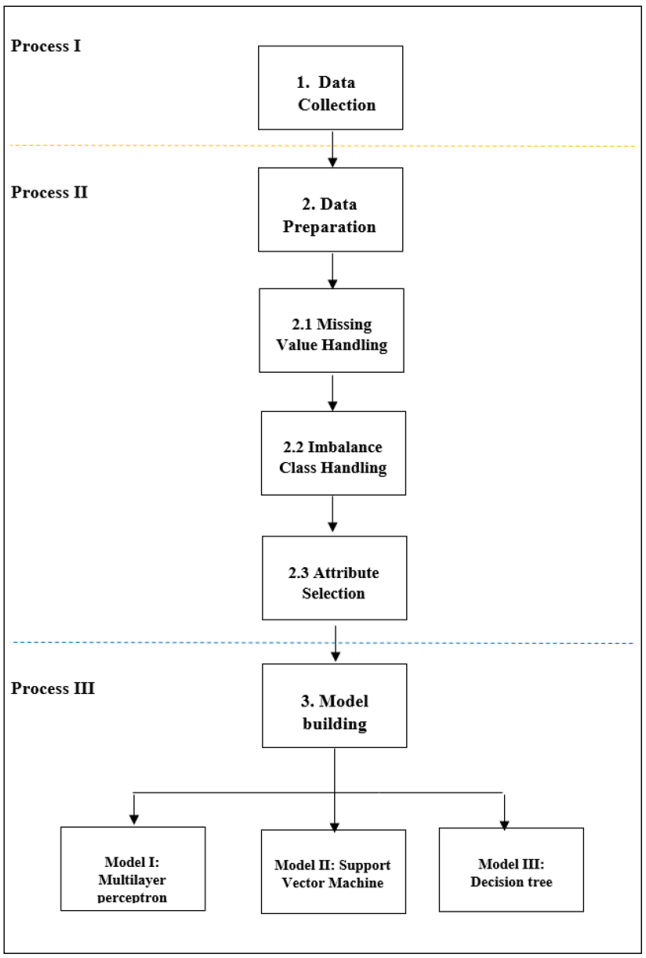


Figure 1. *The process of model building.*

From Figure 1, the model-building process consists of 3 processes: data collection, data preparation, and model-building, as explained in the next step.

3.1 Data Collection

This research uses a thyroid disease dataset, which is open data from the UCI repository. Thyroid disease data set collects from Garavan Institute in Sydney, Australia [19] using data of "allhypo". There are 30 attributes (29 attributes and 1 target). There is a total of 2,800 rows [20].

3.2 Data Preparation

Once we get data, we prepare data, which consists of 3 steps as follows.

1) Missing Value Handling

From the dataset, we found that data has numerous missing values. We handle missing values using the mode for nominal and compute the mean for numeric [21]. In addition, we removed 2 attributes: query hyperthyroid TBG because it does not data 100% and referral source. Therefore, there are 27 attributes from 29 attributes (2800 rows).

2) Imbalance Class Handling

When we handle missing values, we handle imbalanced classes. We found that the target class consists of class 0 = 2580 and class 1 = 220. Therefore, we handle the imbalance class using SMOTE [21], which increases data by random from the original dataset to distribute data of small class sizes and reduce the overfitting of data. There are 2800 rows (equal number of classes) for model building.

3) Attribute Selection

Once, we handle the imbalance class. We have set experiments to select suitable attributes for model building. There are 2 methods for comparing the model performance by considering accuracy and processing time. The first method is all attribute selection. The second method is attribute selection using ClassifierSubsetEval with BestFirst [21]. From our experiment, the result shows that suitable attribute for model building is shown in Table 1.

Table 1. *The comparison results of suitable attribute selection for model building.*

Method	Accuracy	Processing time
All attributes (27 attributes)	99.43 %	7.72 seconds
our experiment result (16 attributes)	99.61 %*	0.02 seconds*

*high accuracy and less processing time

From Table 1, our experiment result of suitable attribute selection for model building consists of 16 attributes such as on thyroxine, query on thyroxine, sick, pregnant, thyroid surgery, I131 treatment, lithium, goiter, tumor, TSH measured, TSH, T3, TT4 measured, T4U, FTI measured, and TBG measured. These attributes have high accuracy and take less time to process when compared to all attribute selections. Therefore, we use these attributes for model building.

3.3 Model Building

After, we prepared the data. We build a model to predict hypothyroid disease by comparison of classification methods such as Multilayer Perceptron, Support Vector Machine, and Decision Tree [21], [22]. These techniques were selected to build the model because they are easy to interpretability, robust, fast to process, suitable to complex problems, and high accuracy. Therefore, we use these techniques for model building. Each classifier sets the parameters as follows.

For Multilayer Perceptron, we have experimented to define the best parameter for high-performance model building. We set learning rate = 0.3, momentum = 0.2, hidden layer node = 9, input node = 16, and output node = 2 (yes, no), as shown in Figure 2.

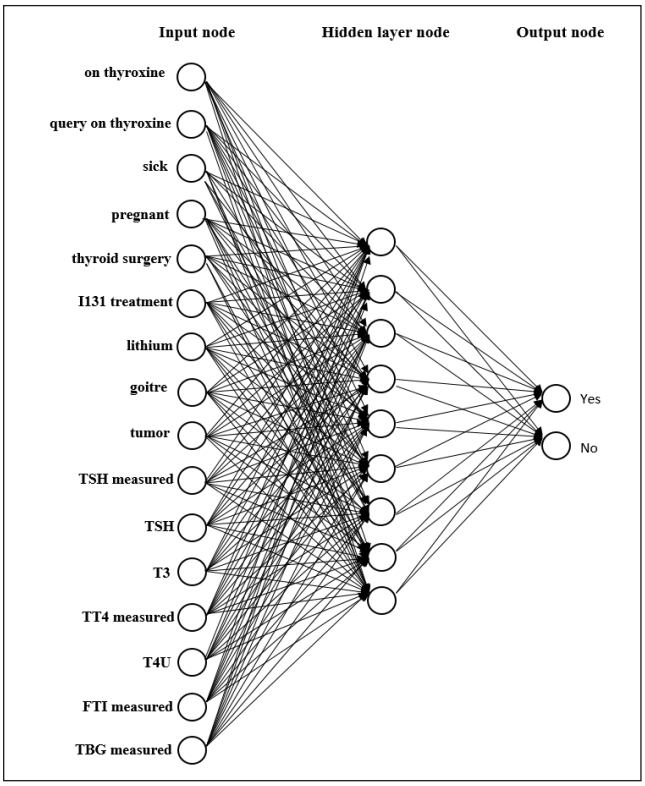


Figure 2. *Model for hypothyroid disease prediction using MLP.*

For the Support Vector Machine, we have experimented in several cases. The result shows that the high-performance model is a kernel function set (polynomial kernel).

For Decision Tree, we have experimented to build the model. The result shows that the suitable parameter is J48, confidence factor = 0.001, and seed = 10.

3.4 Model Performance Measurement

For performance measurement of the model, we use a confusion matrix (Table 2) to compute measurements such as Accuracy, Precision, Recall, and F-measure [21], [22] as shown in Equation 1-4.

Table 2. *Confusion matrix [21].*

Actual result / Prediction result	True	False
True	TP	FP
False	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP+FN+TN+FP} \quad (3)$$

$$\text{F-measure} = \frac{(\text{Precision x Recall}) \text{ x } 2}{(\text{Precision+Recall})} \quad (4)$$

4. Results and Case Study

4.1 Experimental Result and Discussions

After model building for hypothyroid disease prediction using data mining techniques, we compared model performance with 3 techniques: Multilayer Perceptron, Support Vector Machine, and Decision Tree using a dataset from sample data of a total of 2,800 rows (16 attributes), which is described in section 3.2. The dataset is divided into a training dataset and a testing dataset with 5-fold cross-validation. In the model processing, we use an Intel core i5 Central Processing Unit (CPU) @3.10GHz and 4 GB RAM processing platform. The result is shown in Table 3.

Table 3. *The result of model performance measurement with our proposed model.*

	Precision	Recall	F-measure	Accuracy
Multilayer Perceptron	0.963	0.965	0.962	96.46 %
Support Vector Machine	0.929	0.929	0.903	92.93%
Decision Tree	0.996	0.996	0.996	99.61 %*

*high accuracy

From Table 3, Decision Tree model achieves an accuracy of 99.61% and uses a processing time of 0.02 seconds when compared to other models. Multilayer Perceptron and Support Vector Machine have accuracy of 96.46% and 92.93%, respectively. The result of the model performance comparison shows that Decision Tree can effectively predict in several cases such as on antithyroid medication and thyroxine. This model can detect accuracy. Therefore, we selected Decision Tree model to develop a prototype system for hypothyroid disease prediction because it has high accuracy when compared to other models and conforms to our hypothesis.

In addition, we compared our proposed model and state-of-the-art, which uses similar techniques to our proposed method. Guleria, et al. [7] use a public dataset of hypothyroidism. The dataset is extracted feature to select appropriate attributes for model building. Decision Tree, Random Forest, Naïve Bayes Multiclass classifier, and Deep Learning based on ANN are used to build the model. Also, Kumar, et al. [9] use a dataset from the UCI repository. The dataset's features are extracted using the dimensionality reduction technique. Appropriated attributes consist of 12 attributes such as On thyroxine, Pregnant, TSH measured, TSH, Goitre, T3, TT4 measured, TT4, Query hypothyroid, Thyroid surgery, and FT1. Decision Tree and Neural Networks are used to build models. Lim, et al. [14] use a dataset (demographic data, the characteristics, the mediations and treatments, different types of hormoneindexes, referral sources) from Garavan Institute and J. Ross Quinlan, New South Wales Institute, Sydney, Australia. The dataset is extracted features using Featurewiz. Decision Tree, K-Nearest Neighbor, Logistic Regression, Naïve Bayes, Support Vector Classifier, and ensemble machine learning algorithms (Random Forest and Extreme Gradient Boost) are used to build the model. The resulting model performance comparison is shown in Table 4.

Table 4. *The result of model performance comparison.*

	Guleria, et al. [7]	Kumar, et al. [9]	Lim, et al. [14]	Our proposed model
Accuracy	99.5758%	99.58%	99.45%	99.61 %*
Precision	NA	NA	NA	0.996
Recall	0.996	NA	NA	0.996
F-measure	NA	NA	0.99	0.996

*high accuracy

From Table 4, the result shows that our proposed model is also high performance when compared to previous research. In conclusion, our proposed model can be applied to predict hypothyroid disease, which is beneficial for physicians to support decisions in diagnosis for further treatment.

4.2 Case Study

In this section, we demonstrate a case study to show the process of our prototype system. We randomly selected the test dataset in 1 case, as shown in Figure 3.

From Figure 3, we test data in our prototype system to predict hypothyroid disease. When the system is processed, the result shows that this case study has a chance of hypothyroid disease, as shown answer "Yes". Therefore, the user can acknowledge for preliminary risk estimation.

To make it easier to understand for users and physicians, we proposed data visualization to present demographic information, as shown in Figure 4.

From Figure 4, data visualization shows demographic information to support the decision of preliminary risk estimation of a chance of hypothyroid disease for the user before deciding to meet a physician to reduce the cost of medical and receive timely treatment. For example, the user is female and age 65 years. She takes thyroxine, is pregnant, and queries hypothyroid. From these information considerations, the user has a preliminary risk of a chance of hypothyroid disease. Therefore, the user should hurry up to meet the physician for further diagnosis.

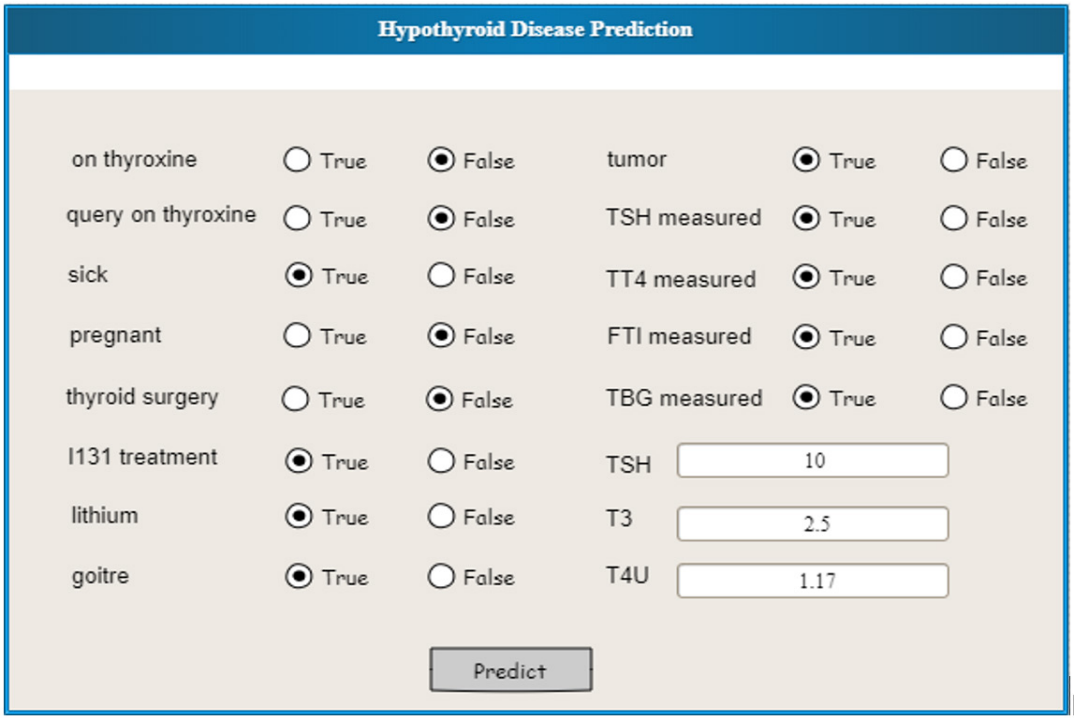


Figure 3. *Case study of hypothyroid disease prediction.*

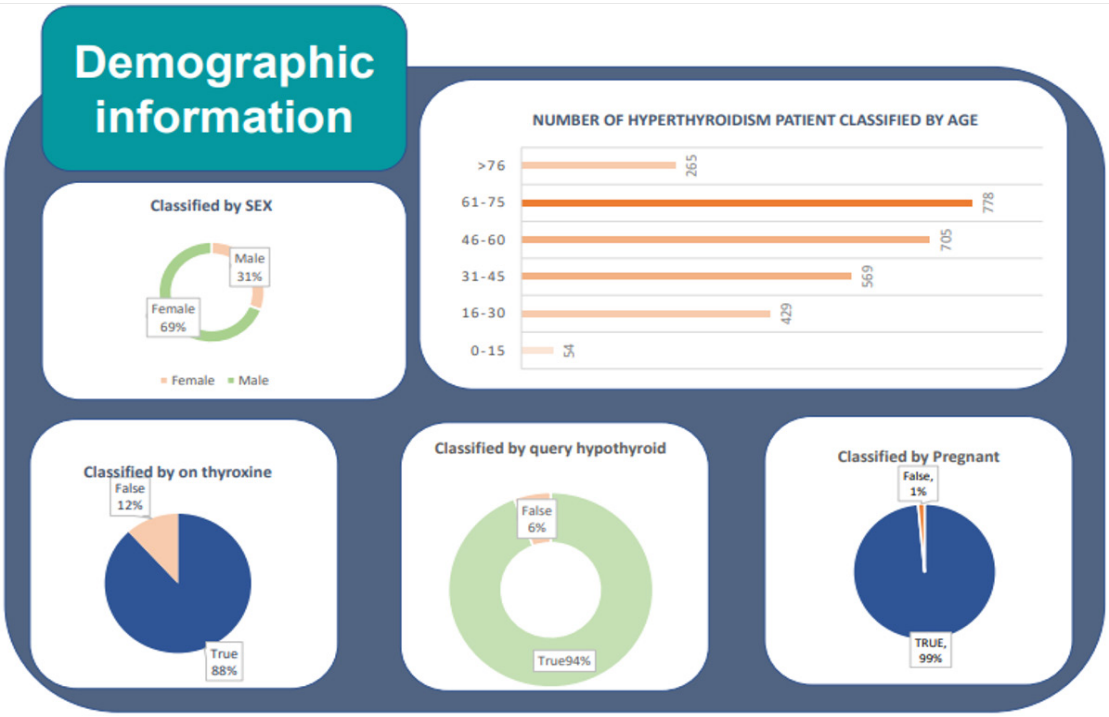


Figure 4. *Data visualization for decision support to detect hypothyroid disease.*

5. Conclusion

In this paper, we proposed a model for hypothyroid disease prediction using Multilayer Perceptron, Support Vector Machine, and Decision Tree based on a thyroid disease dataset. Firstly, the dataset is prepared to handle missing values. Secondly, the dataset deals with the imbalanced class. Thirdly, the dataset is selected as a suitable attribute for model building. Lastly, Multilayer Perceptron, Support Vector Machine, and Decision Tree are used to build models. From our experiment, Decision Tree has a high accuracy of 99.61% and uses a processing time of 0.02 seconds compared to other classifications. Multilayer Perceptron has an accuracy of 96.46%, and Support Vector Machine has an accuracy of 92.93%. In addition, we compared our proposed model to previous research. Guleria, et al. [7], Kumar, et al. [9], and Lim, et al. [14] use a similar technique to our proposed model. The result shows that Decision Tree also has high accuracy. Therefore, we use Decision Tree to develop a prototype system for hypothyroid disease prediction to support the decision of physicians for diagnosis and treatment to reduce the death rate. Moreover, we proposed data visualization for users to estimate the preliminary risk of a chance of hypothyroid disease before deciding to meet a physician to reduce costs and receive timely treatment.

In the future, we plan to develop the system for real use to reduce the cost and risk of a chance of hypothyroid disease in the severe stage.

6. References

[1] National Institute of Diabetes and Digestive and Kidney Diseases, *Hypothyroidism (Underactive Thyroid)*. Available Online at https://www.niddk.nih.gov/health-information/endocrine-diseases/hypothyroidism, accessed on 27 April 2024.

[2] O. M. Ahmed and R. G. Ahmed. "Hypothyroidism." *In Tech Open Access Publisher,* Chapter 1, pp. 1-20, 2015.

[3] O. Patsadu, K. Thepmee, K. Phuengliam, and M. Sirimongkol. "Abnormal Gait Pattern Recognition of Stroke Patient in Initial Stage Using Smartphone and Hybrid Classification Methods." *Information Technology Journal,* Vol. 18, No. 2, pp. 21-33, July-December, 2022.

[4] O. Chunhapran, D. Noolek, P. Labcharoenwongs, and T. Yampaka. "Multi-View Combination using Mutual Information and 3-D Euclidean Distance for Breast Cancer Classification." *Information Technology Journal,* Vol. 18, No. 2, pp. 44-54, July-December, 2022.

[5] N. Nonsiri, R. Manassila, and K. Somkanta. "Data Classifying to Diagnose Diabetes Risk Using Data Mining Techniques." *The Journal of King Mongkut's University of Technology North Bangkok,* Vol. 33, No. 2, pp. 538-547, April-June, 2023.

[6] J. A. Olalekan, F. Ogwueleka, and P. O. Odion. "Effective and Accurate Bootstrap Aggregating (Bagging) Ensemble Algorithm Model for Prediction and Classification of Hypothyroid Disease." *International Journal of Computer Applications,* Vol. 176, No. 39, pp. 40-48, July, 2020.

[7] K. Guleria, S. Sharam, S. Kumar, and S. Tiwari. "Early Prediction of Hypothyriodism and Multiclass Classification Using Predictive Machine Learning and Deep Learning." *Measurement: Sensors,* Vol. 24, pp. 1-7, December, 2022.

[8] A. B. Naeem, B. Senapati, A.S. Chauhan, M. Makhija, A. Singh, M. Gupta, P.K. Tiwari, and W. M. F. Abdel-Rehim. "Hypothyroidism Disease Diagnosis by Using Machine Learning Algorithms." *International Journal of Intelligent System and Applications in Engineering,* Vol. 11, No. 3, pp. 368-373, July, 2023.

[9] R. P. R. Kumar, M. S. Lakshmi, B. S. Ashwak, K. Rajeshwari, and S. Zaid. "Thyroid Disease Classification using Machine Learning Algorithms." *E3S Web of Conferences,* Vol. 391, pp. 1-7, June, 2023.

[10] S. Gothane. "Data Mining Classification on Hypo Thyroids Detection: Association Women Outnumber Men."

International Journal of Recent Technology and Engineering, Vol. 8, No. 6, pp. 601-604, March, 2020.

[11] L. Chake, S. Razvi, I. M. Bensenor, F. Azizi, E. N. Pearce, and R. P. Peeters. "Hypothyroidism." *Nat Rev Dis Primers,* Vol. 8, No. 1, pp. 1-17, June, 2022.

[12] N. Hongboonmee and P. Trepanichkul. "Comparison of Data Classification Efficiency to Analyze Risk Factors that Affect the Occurrence of Hyperthyroid Using Data Mining Techniques." *Journal of Information Science and Technology,* Vol. 9, No. 1, pp. 41-51, January-June, 2019.

[13] S. Parimala and P. S. Vadivu. "Optimizing Thyroid Stage Classification System Using Enhanced Data Mining Algorithms." *Proceedings of 2nd International Conference on Mathematical Techniques and Applications,* India, 2022.

[14] S.T. Lim, K.W. Khaw, X. Chew, and W.C. Yeong. "Prediction of Thyroid Disease Using Machine Learning Approaches and Featurewiz Selection." *Journal of Telecommunication, Electronic and Computer Engineering,* Vol. 15, No. 3, pp. 9-16, July-September, 2023.

[15] T. Alyas, M. Hamid, K. Alissa, T. Faiz, N. Tabassum, and A. Ahmad. "Empirical Method for Thyroid Disease Classification Using a Machine Learning Approach." *BioMed Research International,* Vol. 2022, pp. 1-10, June, 2022.

[16] S. Kurnaz, M.S. Mohammed, and S. J. Mohammed. "A High Efficiency Thyroid Disorders Prediction System with Non-Dominated Sorting Genetic Algorithm NSGA-II as a Feature Selection Algorithm." *Proceedings of 2020 International Conference for Emerging Technology,* India, pp. 1-6, 2020.

[17] R. Chaganti, F. Rustam, I.D.L.T. Díez, J.L.V. Mazón, C.L., and I. Ashraf. "Thyroid Disease Prediction Using Selective Features and Machine Learning Techniques." *Cancers,* Vol. 14, No. 16, pp. 1-23, August, 2022.

[18] X. Zhang, V.C.S. Lee, and J.C. Lee. "Unveiling Thyroid Disease Associations: An Exceptionality-Based Data Mining Technique." *Endocrines,* Vol. 4, pp. 558-572, July, 2023.

[19] Q. Ross, *Thyroid Disease. UCI Machine Learning Repository.* Available Online at https://doi.org/10.24432/C5D010, accessed on 2 March 2024.

[20] UCI Machine Learning Repository, *Thyroid disease.* Available Online at https://archive.ics.uci.edu/dataset/102/thyroid+disease, accessed on 2 March 2024.

[21] J. Han and M. Kamber. *Data Mining Concepts and Techniques.* Morgan Kaufmann Publishers, Third Edition, 2012.

[22] P. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining.* Pearson Education, Limited, International Edition, 2014.

Chatbot Application for Learning Computer Laws Using Artificial Intelligence

Sukuma Uamcharoen

Received: April 19, 2024
Revised: September 19, 2024
Accepted: September 23, 2024

* Corresponding Author: Sukuma Uamcharoen, E-mail: sukuma.uam@mail.pbru.ac.th

DOI: 10.14416/j.it/2025.v2.003

Abstract

This research aimed to develop and evaluate a "Chatbot Application for Learning Computer Laws using Artificial Intelligence." The study involved assessing user attitudes towards this application, utilizing questions and answers derived from pertinent laws in Thailand, including the "Computer-Related Crime Act," the "Copyright Act," "Thailand's Personal Data Protection Act," the "Cybersecurity Act," and the "Patent Act." In an evaluation of the Chatbot AI performance with a sample group of 100 evaluators, the following metrics were observed: Accuracy = 0.98, Precision = 1.00, Recall = 0.97, and F1-Score = 0.98. The research outcomes encompassed the successful development of the chatbot application and the summary results of the performance from a sample data set of the Chatbot Application for Learning Computer Law using Artificial Intelligence. The sample size, comprising 244 undergraduate students from Western Rajabhat Universities, was determined using the Taro Yamane table. The universities included Kanchanaburi Rajabhat University, Nakhon Pathom Rajabhat University, Phetchaburi Rajabhat University, and Muban ChomBueng Rajabhat University. The assessment revealed that the overall of performance from a sample data set was at the highest level ($\bar{X}$ = 4.59, SD. = 0.09).

Keywords: Chatbot, Dialog flow, Artificial Intelligence.

1. Introduction

Computers in Thailand are mostly now connected to the Internet, which constructs massive online social networks in the country that follow with the increase in computer crimes as well. Therefore, the Thailand Computer Crimes Act 2007 was enacted to deal with the problems of the promising increase in computer crimes. The Computer Crimes Act (No. 2) B.E. 2560 (2017) amendment is, therefore, a specific law applicable to Internet or computer-related offenses. It is a law that aims to prevent and suppress crimes related to computers and electronic data in the country [1].

The Thailand computer laws are the Copyright Act B.E. 2537 (1994) amended by Copyright Act (No. 2) B.E. 2558 (2015), Copyright Act (No.3) B.E. 2558 (2015) and Copyright Act (No.4) B.E. 2561 (2018), Thailand's Personal Data Protection Act B.E. 2562 (2019) and Cybersecurity Act B.E. 2562 (2019), Patent Act B.E. 2522 (1979) as amended by the Patent Act (No.2) B.E 2535 (1992) and the Patent Act (No.3) B.E 2542 (1999).

Based on the researcher's experience of teaching law and ethics for computer professionals and law and ethics for the digital workforce for over 10 years, it has been found that students have limited knowledge about digital laws, as evidenced by past exam scores. Therefore, having a chatbot facilitates immediate interactive learning and motivates students in various ways such as through interactive and conversational methods. This allows for interactions at any time and place. Students who engage in self-directed learning based on their interests align with the findings of [2], They identified key reasons for using chatbot programs to develop a question-and-answer system on digital law issues: 1) The ability to communicate with users at any time, 2) Minimal time required for user communication, 3) Accurate and relevant responses

* *Department of Digital Innovation, Faculty of Information Technology, Phetchaburi Rajabhat University.*

to user queries, and 4) The capability to develop chatbots that operate on web applications, allowing users to access the chatbot through web applications without needing to download and install an app on their smartphones. The researcher has developed a chatbot system for addressing questions related to digital law and information. This enables users to study laws more conveniently and quickly. This approach facilitates access for individuals seeking knowledge about digital and related laws.

2. Theoretical background and related research

2.1 Artificial Intelligence

Artificial intelligence is related to thinking and reasoning processes, with humans as the prototype. It branches into several fields, such as 1) Natural Language Processing (NLP), which deals with communication using human languages, such as text, letters, and speech. Current examples include various chatbots.

NLP's primary goal is to enable computers to understand and accurately follow the natural language communication of humans. The principles of NLP used for the Thai language have specific characteristics unique to Thai. The key principles of NLP for Thai are as follows:

Tokenization and Word Segmentation: This involves breaking down Thai text into meaningful units. Unlike some other languages, Thai does not have spaces between words. For example, words such as deception, collection, entry, giving, message, imitations of rights, personal data, necessity, consent, and independence are segmented.

Part-of-Speech Tagging (POS): This involves identifying the grammatical categories of words in Thai based on their context, such as nouns and verbs.

Named Entity Recognition (NER): This involves identifying and categorizing entities, such as copyright or section 5, in the natural language of Thai.

Dependency Parsing: This involves analyzing the grammatical structure of Thai sentences to find the relationships between words in a sentence.

Sentiment Analysis: This involves analyzing the sentiment in the Thai natural language, such as determining whether a text is positive, negative, or neutral.

Text Classification: This involves categorizing text into different types or labels, such as Acts or sections.

Speech Recognition: This involves converting spoken Thai language into text for searching purposes, such as recognizing the phrase "คอมพิวเตอร์ มาตรา 8" (computer section 8).

2) Computer Vision, which enables computers to perceive various environments from images, such as recognizing human faces. 3) Robotics, which involves making machines perform various tasks accurately and intelligently. And 4) Expert Systems, which simulate experts in specific fields using computer programs to perform tasks or assist human experts in decision-making.

The National Electronics and Computer Technology Center (NECTEC) developed the AI FOR THAI platform to serve as a digital infrastructure for the Thai population. This platform, integrated with NECTEC's Artificial Intelligence (AI) research, is offered as API services for users and developers to create and enhance applications that benefit both businesses and society. The AI FOR THAI platform comprises 11 services, including Basic NLP, Tag Suggestion, Machine Translation, Sentiment Analysis, Character Recognition, Object Recognition, Face Analytics, Persons & Activity Analytics, Speech to Text, Text to Speech, and Chatbot [3].

2.2 Literature review

Chatbot is a computer application developed using artificial intelligence (AI) and natural language processing (NLP) technology. It has an interface that users can query for an answer in a specific area of knowledge. A Chatbot can be either a voice-enabled or texting-enabled system. AI will recognize keywords in speech or sentences, and NLP will process to construct the meaning of the query and search for the optimized answer from the pre-defined databases or text files, which are subsets of a knowledge-based system. The answer to a query can be in voice, text, or figure. There were lots of research

topics on Chatbot and lots of applications have been developed such as the "College Enquiry Chatbot" [4], the "Designing a Chatbot that Simulates an Historical Figure" [5], and the "Chatbot-A Java Based Intelligent Conversational Agent" [6]. Chatbots can be used in many disciplines. For example, Chatbot in business, Chatbot in education, Chatbot in financial services, Chatbot in law.

Currently, Thailand has several digital laws, also known as Cyberlaws, such as the Copyright Act, which grants exclusive rights to creators for their works. The Computer-Related Crime Act highlights the importance of computer systems in business and daily life, addressing actions that disrupt computer operations. Cybersecurity laws focus on measures to prevent and mitigate cyber threats that impact national security, economic stability, military security, and public order. The Patent Act protects inventions and designs through legal documentation. The Personal Data Protection Act (PDPA) safeguards any information that can identify an individual, directly or indirectly, excluding data about deceased persons.

The content of these laws is extensive and written in formal legal language, making it difficult to understand and access the essential points. Users often need to search government or legal websites to find relevant information. This challenge led to the development of a web application to facilitate asking and answering important questions about digital laws through a chatbot. This web application provides easy access to knowledge about digital laws at any time.

The chatbot's efficiency, effectiveness, and user satisfaction demonstrate its value in providing immediate Q&A services on legal issues related to digital innovations. [2]

This research presents a prototype of a recommendation and guidance system for legal processes in handling computer crime cases. The system utilizes a chatbot on the LINE application to answer questions and provide guidance to users and police officers on matters related to computer crime. The information is based on the Computer-Related Crime Act. The researcher focuses on the practical training of police cadets.

This research shows that police cadets can effectively use the chatbot system to assist and provide guidance in handling computer crime cases quickly. [7]

Thailand Computer law first be announced in 2007, called "Computer-Related Crime Act B.E. 2550," which was amended by the "Computer-Related Crime Act (No. 2) B.E. 2560" in 2017, ten years later. In this Act, there are several sections describing the terms used, responsibility and control for the execution of this Act, the authority to issue a Ministerial Rule for the execution of this Act, the Computer-related offenses, and the Competent officials. Noticed that the two Acts had already been published for 15 years and 5 years, accordingly, which were relatively new when compared to the other laws. So, the awareness and knowledge of the content of these Acts among the Thai citizens and concerned parties are limited. The students in computer-related disciplines in the university who will pursue the computer profession and be the workforce in the computer industry in the country should have proficiently understood all contents of the two Acts to protect. The two Acts 's purposes are to govern cyber security and computer system protection. Later, in 2019, the law that governs data protection was announced as "Thailand's Personal Data Protection (PDPA) Act BE 2562" with a one-year grace period. This law has just been officially affected in the year 2021. The content in the law's sections is new to everyone. The current tools of study for these Acts are content reading tools which are books, websites, and social media. The searching tool in electronic form has only a Google search, which provides a hundred or thousand answers that relate to the keywords provided, which might not be specific enough to the questions that learners would like to know or learn.

As described earlier, Chatbot is an efficient learning tool that can be used in education for learning subjects by querying the question learners would like to know. The AI and NLP are the background mechanisms of the application, which will look up the answer from the databases or files system and respond to the learner an answer through an application's interface interactively. This paper then proposes the development of a Chatbot for learning Thai computer law and PDPA law

to increase efficiency and overcome those existing pain points in Thai computer law and PDPA law learning, as indicated before.

3. Research Methodology

This research has endorsed ethical considerations in humans. COA No. 031/2022 from Kanchanaburi Rajabhat University Research Ethics Committee.

The design of a conceptual framework of this research is shown in Figure 1.

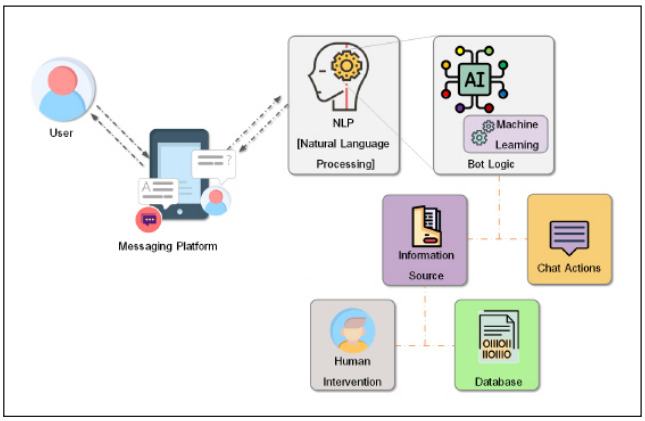


Figure 1. *A Conceptual Framework of Chatbot Application for Learning Computer Laws by Using Artificial Intelligence.*

From Figure 1, the chatbot application is a platform used to exchange messages between human and computer programs, analyzing the user's needs based on the complexity of the language and being able to interact in real time. A section, called Natural Language Processing (NLP) is a natural language processing technology that gives computer programs the ability to understand human language as spoken or written. Artificial Intelligence (AI) works with machine learning (ML) to help analyze and recognize user needs. Based on natural language processing (NLP), chatbot applications can initiate user interactions with Chat Action, which is a collection of conversations with users by retrieving real-time data via an information source, which can be either extracted from the stored database or file system through the API or passed on to human intervention. The research methodology consists of 2 parts:

3.1 Build the "Chatbot Application to Learn Computer Law using Artificial Intelligence" by preparing the dataset for the training process using the keywords derived from the contents of the "Computer Crime Act" as the input data. The output data is the information after training. After that, design a chatbot application from the process of preparing the data used for training to be the input sentences as input data. Output data is the complete sentence that will be the initial answer to users. The steps were as follows:

In this research, Dialogflow was applied to the study of computer law. Dialogflow integrates techniques that understand user inputs and respond accurately. It comprises Machine Learning and statistical machine translation to analyze training phrases (train set) and user questions (test set). This analysis helps understand patterns and relationships between words and phrases. From this analysis, Dialogflow creates an internal model, which becomes an algorithm that updates with each change. It uses Rule-based grammar matching for easy training and ML matching for creating statistical models from training data, which includes entities, sentiment, moderation, and categories.

The data preparation process in this research involves splitting the data into two sets: the training set for training the chatbot model to learn language patterns, conversational behavior, and response methods, and the Test Set for evaluating the trained model's performance by comparing its responses to the correct answers. The steps for preparing the Train Set and Test Set are as follows:

1) Define the Scope of Data: Define the language that the chatbot will understand and respond to in Thai natural language, set the domain, and define the topics the chatbot can ask and answer questions about, such as the Personal Data Protection Act, the Computer-Related Crime Act, and the Copyright Act.

2) Data Collection: Select information from the publication of Acts by copying the essential content within the Acts to form the answer set.

3) Data Cleaning: Remove duplicate data, and format the data into an online database format so that the program

can retrieve answers and make corrections if necessary.

4) Data Splitting into Train Set and Test Set: The research uses methods to split the data into Train Set and Test Set, creating question sets from analyzing the content of the Acts, such as from the analysis of events.

5) Data Acquisition Method: Use augmentation to modify existing data to synthesize key terms for constructing question sentences. Utilize Natural Language AI (Figure 2) to synthesize key terms related to the Personal Data Protection Act, B.E. 2562, Section 19, including law, case, committee, request, withdrawal, deception, collection, entry, giving, message, limitation of rights, personal data, committee, necessity, consent, independence, provision, service, benefit, impact, controller, Act, language, electronic system, purpose, method, condition, contract, part, book, chapter, condition, owner, etc.

3.1.1 Users enter a text message in the chat field in the application or choose to use speech to convert to text in the application. The message can be a search term.

3.1.2 When the chatbot application receives the content of the message, it will be forwarded to the back-end service with a chatbot agent that uses artificial intelligence (AI) and natural language processing (NLP) to interpret the content and recognize the needs of users.

3.1.3 After the automatic conversation dialog, the system searches and gets the right answer. The response will be sent to the users through the chat box in the same application. A new cycle then begins through conversations between the users and the chatbot application.

The principles of operation of the chatbot application are in Figure 3 from the chat channel in the LINE Chat application, receive messages from the user in the Client section, and then connect through the Line messaging API channel, which is the transmission between the Dialogflow automated chat system in the backend management system. with the LINE Chat application platform.

The method for communicating is part of the blackened management system within the LINE Chat application platform; LINE Chat is handled via message notification as a standard

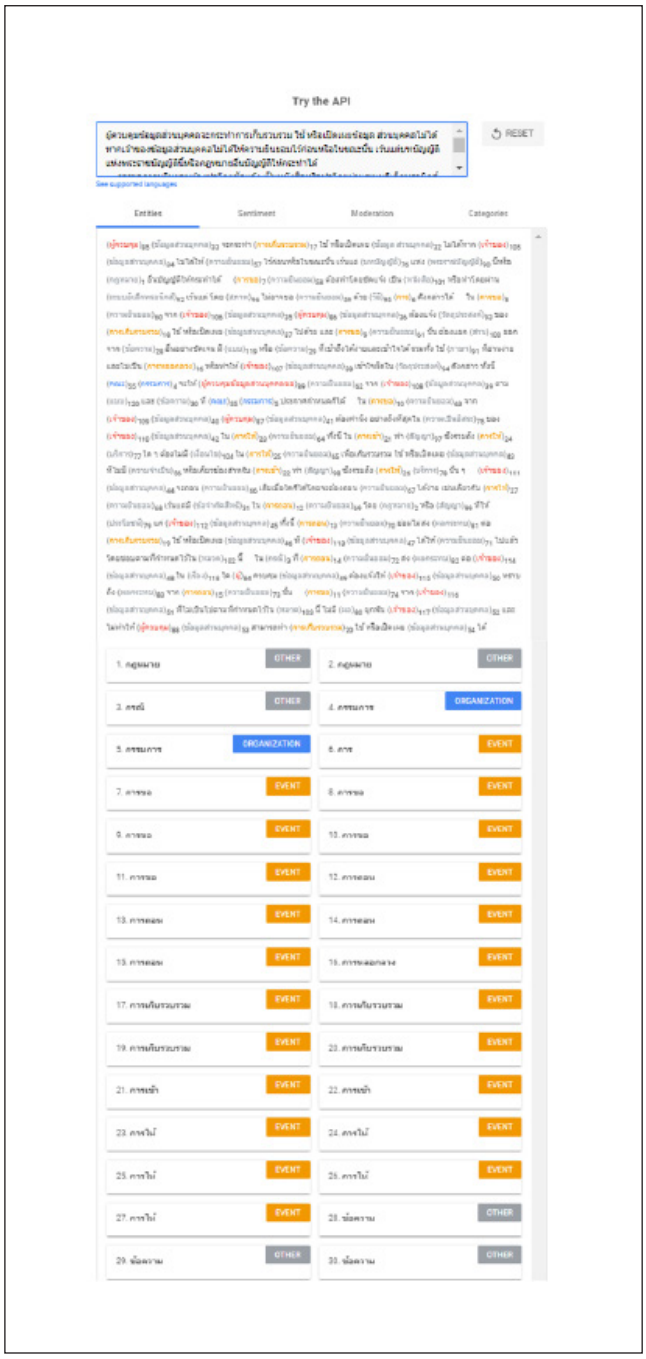


Figure 2. *Train Set and Test Set*

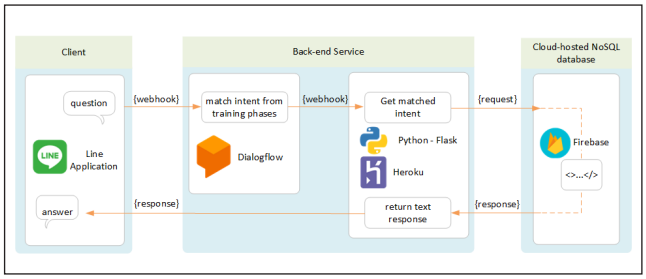


Figure 3. *The principle of the operation of the chatbot application.*

LINE Server provided, known as a webhook. A webhook is an Application Programming Interface or a type of API that operates on events rather than on requests.

In a normal API connection channel, the default client-side application sends a request to the server or service provider. To receive a response and to get information that is up to date at all times in real-time in 4.

The application has to keep sending the same words without knowing when the data will be changed. Then the server side has to handle a large number of requests. While most of the data has not been changed at all, it's a waste of resources and this is a normal API limitation.

Webhooks are sometimes referred to as "reverse APIs" instead of functioning as the usual API-style connections. It functions based on the events in which the dataset changes will be tracked by the client-side application, the real-time dataset will be sent to the client-side application as needed. Then the server can send more data than it receives. In short, the Webhooks are an API service that allows data to be transmitted as soon as a specific event occurs. Then, when the received message arrives in the dialogue automation, the Dialogflow will be responsible for understanding the intent of the received message. And compared to the message thread that has already been prepared in Figure 4.

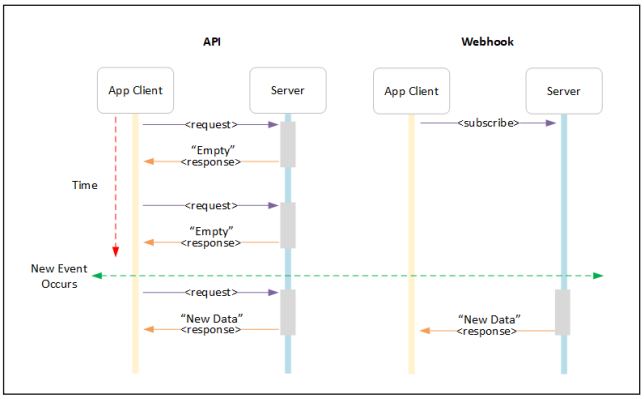


Figure 4. *A comparison of the normal API connection and the Webhook connection.*

3.2 Assessed user attitudes toward the usage of the "Chatbot Application for Learning Computer Law using Artificial Intelligence"

The population was 627 undergraduate students who enrolled in the Department of Computers from the Western Group Rajabhat University. The researcher received the sample size of 244 students by using Taro Yamane's table at a 95% confidence level, with a tolerance of ± 5. The sample was selected by purposive sampling from the undergraduate students. The researchers developed a questionnaire as the user attitude assessment tool of the "Chatbot Application for Learning Computer Law using Artificial Intelligence" and as the data collection tool from a sample of 244 students. The IOC test was reviewed by three experts. This test consists of 26 questions. The attitude test consisted of 11 items consisting of 4 items of content, 5 items of usability, 3 items of efficiency, and 3 items of effectiveness. Five levels of assessment criteria for attitude, content, usability, efficiency, and effectiveness were as follows:

The average is 4.51 - 5.00 = the highest level.

The average is 3.51 - 4.50 = high level

The average is 2.51 - 3.50 = moderate level.

The average is 1.51 - 2.50 = low level.

The average is 1.01 - 1.50 = the lowest level.

4. Result and Discussion

The researcher presented the findings from this research which were classified according to research objectives as follows:

4.1 Computer Laws Chatbot Application

The author developed the "Chatbot Application for Learning Computer Laws by using Artificial Intelligence" by using AI and NLP as a core technology and Line application as the users' interface, as shown in Figure 5 and Figure 6.

4.2 Performance Evaluation

The author evaluates the Chatbot/ AI performance with measure metrics.

1) Accuracy = (TP+TN)/ (TP+TN +FP+FN)

= (67+31)/(67+31+0+2)

= 98/100

= 0.98

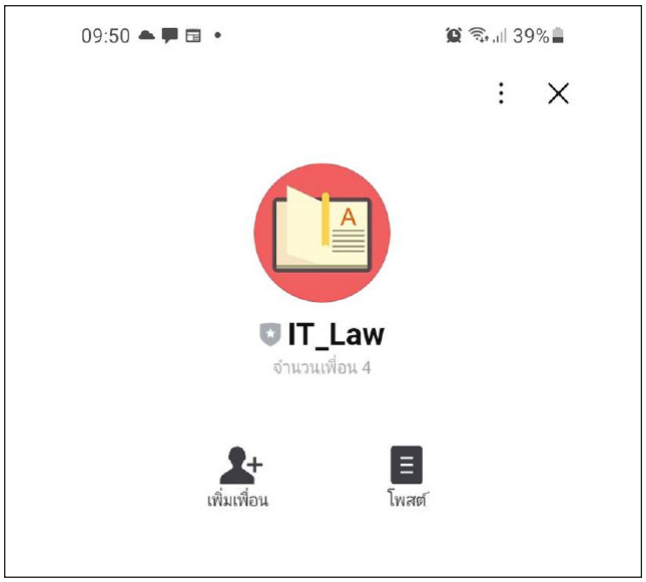


Figure 5. *The "Chatbot Application for Learning Computer Laws by Using Artificial Intelligence".*

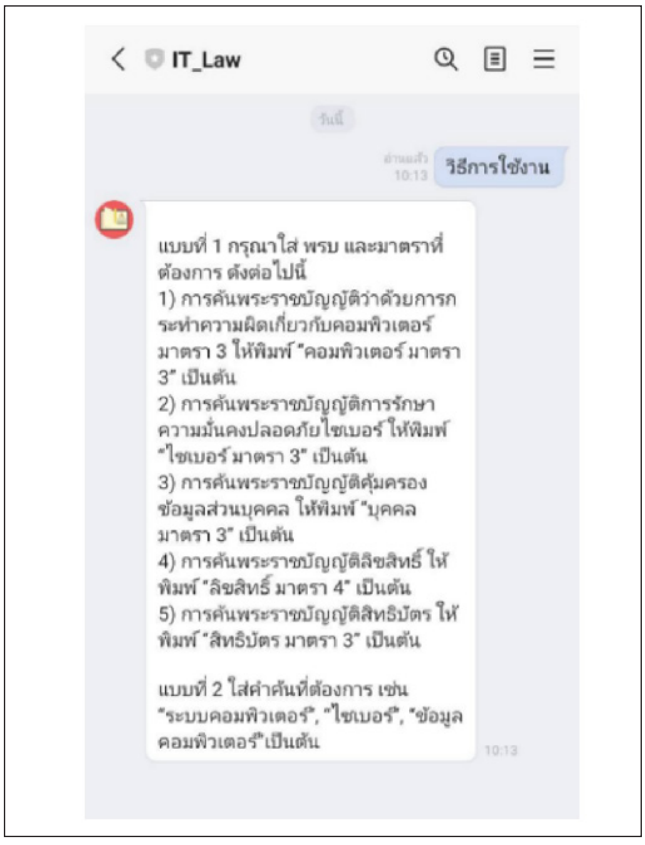


Figure 6. *A guideline for using the "Chatbot Application to Learn Computer Laws using Artificial Intelligence"*

2) Precision = TP/(TP+FP)
= 67/(67+0)
= 1.00

3) Recall = TP/(TP+FN)
= 67/67+2
= 67/69
= 0.97

4) F1-Score

F1 = 2*[(precision*recall)/(precision +recall)]
= 2*[(1*0.97)/(1+0.97)]
= 2*[0.97/1.97]
= 0.98

Remark:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

4.3 User Satisfaction

Users tried out the "Chatbot Application to Learn Computer Laws using Artificial Intelligence" that was developed and assessed users' attitudes toward the experiences in using the "Chatbot Application to Learn Computer Laws using Artificial Intelligence" after the trial. The results of the assessment are shown in Table 1.

The assessment results of the Western Rajabhat Universities. The assessment results of the students of the "Computer Education" at the "Muban Chombueng Rajabhat University" found that the "satisfaction as a sample data" domain was at the highest level ($\bar{X}$ = 4.70, SD. = 0.47), the "content" domain was at the highest level ($\bar{X}$ = 4.74, SD. = 0.45), the "usability" domain was at the highest level ($\bar{X}$ = 4.71, SD. = 0.46), the " efficiency " domain was at the highest level ($\bar{X}$ = 4.64, SD. = 0.49), the " effectiveness " domain was at the highest level ($\bar{X}$ = 4.69, SD. = 0.46) and the overall of performance from a sample data set was at the highest level ($\bar{X}$ = 4.59, SD. = 0.09).

Table 3. *The summary results of the performance from a sample data set of the "Chatbot Application for Learning Computer Law using Artificial Intelligence".*

University / Domain	Kanchanaburi Rajabhat University		Nakhon Pathom Rajabhat University		Phetchaburi Rajabhat University		Muban Chombueng Rajabhat University		Total	
	$\bar{X}$	SD.	$\bar{X}$	SD.	$\bar{X}$	SD.	$\bar{X}$	SD.	$\bar{X}$	SD.
The satisfaction as a sample data	4.57	0.57	4.46	0.58	4.54	0.60	4.70	0.47	4.57	0.09
The content	4.64	0.53	4.44	0.59	4.62	0.59	4.74	0.45	4.61	0.11
The usability	4.62	0.58	4.44	0.63	4.69	0.49	4.71	0.46	4.62	0.11
The efficiency	4.56	0.56	4.46	0.60	4.56	0.64	4.64	0.49	4.56	0.06
The effectiveness	4.65	0.59	4.43	0.58	4.54	0.59	4.69	0.46	4.58	0.10
Total Average:	4.61	0.56	4.45	0.60	4.59	0.58	4.70	0.46	4.59	0.09

"Chatbot Application for Learning Computer Law using Artificial Intelligence" in this research uses AI technology to automate chat responses with Dialogflow and a Line Application as a tool to help learn effectively. Students can find answers quickly anytime, anywhere. This research corresponds to the research on an "Application of Artificial Intelligence Chatbot for Learning, a case study of the periodic table with content and detailed properties of various elements". These data were used to create an artificial intelligence system that responds automatically to chats using the Dialogflow program and connects to Line Application [8]. It also corresponds to research on the "Smart Village Management System via Smart Phone with Line API" [9], which is a system developed by using the Line API (Application Programming Interface), which causes users to be able to use the system without having to install other applications. It is convenient for users who are normally familiar with the use of Line, with a chatbot to automatically answer the questions, which makes it as efficient as possible.

A Chatbot Application for Learning Computer Law using Artificial Intelligence" used the Dialog Flow and a Line application to process questions and get answers for users. This is in line with the research on the "Tourism Promotion Information System via Line Chatbot Application System in Phitsanulok Province" which has the assessment for the efficiency and the effectiveness of the system at a high level [10] and in line with the "Developing an Information System for Learning Computer Network Vocabulary through the Line Chatbot Application" which the efficiency and effectiveness of the Information System through the Line Chatbot is at the highest level [11].

"A Chatbot Application for Learning Computer Law using Artificial Intelligence" has the assessment on its efficiency and effectiveness which the findings are consistent with the findings of the research on "Studied and Developed Chatbots via Web Applications for Q&A on Digital Legal Issues and Related Issues via Chatbots" [2]. An application allows users to be able to query the chatbots through a web application that is always connected to the Internet. An assessment of the research study found that an application has a high level of efficiency and effectiveness. The research on "The Development of the Supporting Data Providing System for the Khorat Geopark Attraction using Digital Technology" [12] used machine learning (ML) and natural language processing (NLP) as a background mechanism to process the chat bot's response and used a responsive web page as a data presentation model. The performance evaluation of the system was also at a high level.

5. Conclusions

Chatbot Application for Learning Computer Laws by using Artificial Intelligence has the main objectives: 1) develop a chatbot application for learning computer laws by using artificial intelligence and 2) assess the attitudes of users of a Chatbot Application for Learning Computer Laws by using Artificial Intelligence. It was found that 1) the Chatbot Application for Learning Computer Laws by using Artificial Intelligence was successfully developed and function as design, 2) the overall results of the assessment of the user attitudes of the Chatbot Application for Learning Computer Laws by using Artificial Intelligence of the students of Computer Education/ Technology at the Kanchanaburi Rajabhat University, the students of the Computer Education at the Muban Chombueng Rajabhat University, and the students of the Computer at the Phetchaburi Rajabhat University were at the highest level. And the assessment of the attitudes of the users of a Chatbot Application for Learning Computer Laws by using Artificial Intelligence of the students of Computer Education at the Nakhon Pathom Rajabhat University was at a high level.

6. Acknowledgments

Thank you, Phetchaburi Rajabhat University, for providing funding for research.

7. References

[1] W. Meethum and B. Khwayota. "Enforce Laws in Accordance with Statute of Computer Crime Act Criminal Law (NO.2) B.E. 2560." *Paññа Panithan Journal,* Vol.6, No.1, pp. 135-139, January-June, 2021.

[2] C. Tantikanedee and V. Chooprayoon. "A Smart Web Application for Cyberlaw Questions and Answers via a Chatbot Program." *TLA Research Journal,* Vol. 14, No. 2, pp. 36-53, July-December, 2021.

[3] C. Tapsai, P. Meesad, and H. Unger. "An Overview on the Development of Thai Natural Language Processing." *Information Technology Journal,* Vol. 15, No. 2, pp. 45-52, July-December, 2019.

[4] P. Slave, et al., "College Enquiry Chat Bot." *International Journal on Recent and Innovation Trends in Computing and Communication,* Vol. 5, No. 3, pp. 463-466, 2017.

[5] E. Haller and T. Rebedea. "Designing a Chat-bot that Simulates an Historical Figure." *2013 19th International Conference on Control Systems and Computer Science,* Romania, pp. 582-589, 2013.

[6] L.S. Chetan Rao, et al. "Chatbot-a Java Based Intelligent Conversational Agent." *International Research Journal of Engineering and Technology (IRJET),* Vol. 4, No. 4, pp. 3575-3578, 2017.

[7] S. Pan-aon, A. Chompoonuch, and W. Keardsri. "A Prototype System for Advising and Guiding The Legal." *Journal of Criminology and Forensic Science,* Vol. 8, No. 1, pp. 114-126, 2022.

[8] P. Tangkraingkij. "Application of Artificial Intelligence Chatbot for Learning." *Royal Thai Air Force Medical Gazette,* Vol. 66, No. 2, pp. 64-73, May-August, 2020.

[9] K. Mitsophonsiri, S. Thangsathityangkui, and S. Ongsuwan. "Smart Village Management System via Smart Phone with Line API." *Kasem Bundit Engineering Journal,* Vol. 10, No. 2, pp. 165-194, May-August, 2020.

[10] J. Sartmunee, P. Soonthonrot, K. Thongfak, and P. Kaewta. "The development of Information System for Tourism Promotion by Application LINE Chatbot in Phitsanulok." *Journal of Accountancy and Management,* Vol. 13, No. 1, pp. 100-111, January-March, 2021.

[11] J. Chitiyaphol and P. Dornpinij. "Developing an Information System for Learning Computer Network Vocabulary through the Line Chatbot Application." *Interdisciplinary Academic and Research Journal,* Vol. 2, No. 4, pp. 607-618, July-August, 2022.

[12] T. Sungsri, T. Sungsri, U. Ua-apisitwong, S. Krompho, J. Panomjerasawat, K. Chamnapon, M. Thongdee, and A. Dermsantear. "The Development of the Supporting Data Providing System for the Khorat Geopark Attraction using Digital Technology." *Science and Technology Research Journal Nakhon Ratchasima Rajabhat University,* Vol. 6, No. 2, pp. 19-28, July-December, 2021.

Utilize Novel Algorithms to Acquire, Analyze, and Extract Data from TikTok Discover Page and Education-Related Topics

Jincheng Zhang* and Thada Jantakoon*

Received: March 26, 2024
Revised: October 1, 2024
Accepted: October 11, 2024

* Corresponding Author: Jincheng Zhang, E-mail: zjc1639834588@gmail.com

DOI: 10.14416/j.it/2025.v2.004

Abstract

Due to the swift advancement of research and technology, particularly in the fields of computer science and data science, individuals are progressively employing these technologies, along with others, in the realm of education. This study encompasses the development, creation, and utilization of a comprehensive range of techniques, spanning from data collecting to data analysis and mining. It introduces a novel algorithm and methodology for acquiring and refining data, as well as three innovative algorithms for data analysis and exploration. This project collects data on the topics featured on the TikTok Discover page for the purpose of doing data analysis and data mining. The research methodologies employed in this work encompass empirical research, experimental verification, algorithm design and optimization, system design, and implementation. Our study examined and extracted educational content from TikTok Discover pages. We studied the popularity of this data from various perspectives and levels. This allows users to easily and efficiently locate the specific information they are interested in for further investigation. Analysis, sentiment analysis, and potential anomalous data were discovered. The analysis and extraction of this data offer educational practitioners' significant insights that can be utilized to inform and direct educational practice.

Keywords: Novel Algorithms, Tiktok Discover Page, Data Analysis, Data Mining, Education-Related Topics, Education ICT.

1. Introduction

Individuals are progressively retrieving information from the Internet with the intention of doing data analysis and mining, as well as extracting important insights and knowledge for educational reasons [1], [2]. This can be attributed to the exponential expansion of scientific advancements and technological breakthroughs, namely in the realm of computer science. Data analysis and data mining technologies have the potential to greatly improve education by enhancing instruction quality, identifying weaknesses in the current educational system, improving instructional content design, and increasing student interest and efficiency [3], [4].

As a website with huge influence in the world, TikTok has a large number of users and data, including a large amount of valuable data waiting for people to mine and develop.

Computer science and data science have experienced significant advancements in recent years, leading to their extensive use across many domains. Within the education sector, there is a growing trend of utilizing computer science and data science technology to enhance and improve the industry. These encompass individuals' comprehensive and thorough utilization of data mining technology in the field of education. These strategies involve utilizing data mining to accomplish various tasks such as predicting students' academic performance [5], identifying tailored learning for students [6], evaluating existing courses [7], and predicting the likelihood of students dropping out [8].

This study devised and introduced a novel algorithm

* Faculty of Science and Technology, Rajabhat Maha Sarakham University

and methodology for efficiently and accurately acquiring data pertaining to the content featured on the TikTok Discover page. This technique and method are the first of their kind in the world, designed to efficiently and quickly gather data specifically related to the content found on the TikTok Discover page. Indeed, this novel technology and methodology can be applied to gather data from several other websites as well. This study is the first in the world to analyze and mine data from TikTok's discovery page topics. We developed this novel algorithm and methodology to get data from the Internet due to the presence of intricate anti-crawler obstacles encountered when individuals attempt to extract data from certain websites, such as TikTok, using crawlers. We attempted to employ a web crawler to extract data from the TikTok website, however we faced numerous intricate anti-crawler obstacles on TikTok. Typically, the software encounters errors after the crawler completes the data crawling process. Our algorithm offers a highly effective, straightforward, and efficient solution to problems.

In this study, we devised and suggested three novel algorithms for doing data analysis and data mining. The three methods are "Keyword Distance Weighted Frequency," "Keyword Distance Weighted Frequency-Inverse Document Frequency," and "Keyword Distance Weighted Frequency-Emotion Analysis Frequency." These three novel methods can be utilized for data analysis and data mining across several domains. We conducted a comprehensive analysis and data extraction of educational content on TikTok. Discover: doing comprehensive analysis of the prevalent data from many perspectives and levels, enabling users to effortlessly and effectively locate the pertinent facts of their interest for further investigation. Analysis, sentiment analysis, and potential anomalous data were detected. The analysis and mining of this data offer educational practitioners' significant insights that can be utilized to inform educational practice. This study develops new algorithms for general data mining purposes and conducts experiments in the education domain.

Recently, research in educational data mining has increasingly focused on integrating social media data, including platforms such as TikTok, to better understand educational trends and behaviors. Research has highlighted how data extraction and analysis algorithms can effectively process large amounts of content from such platforms to gain insights into user preferences, learning patterns, and popular educational topics. Our work follows this trend by integrating data from TikTok's Discover page, using novel algorithms to extract and analyze education-related content. This approach provides educators and researchers with actionable insights that enhance decision-making in educational technology.

1.1 Research framework and structure

The scope of this research is comprehensive and detailed, mainly including the following aspects:

1.1.1 Use the Python programming language to obtain and filter data on TikTok Discover page themes.

1.1.2 Apply data analysis and data mining technology, and use 11 codes to analyze the collected data on TikTok Discover page topics in different aspects and levels. This includes the use of 3 new algorithms we invented and proposed for data analysis and data mining.

1.1.3 Explain the results of data analysis and data mining so that they can be applied in educational practice.

1.1.4 Research Focus and Scope: Data mining algorithms can be used in education for personalized learning, grade prediction, and sentiment analysis. Clustering algorithms help identify students' learning styles, regression analysis predicts grades, and sentiment analysis processes feedback to improve courses. These applications improve learning outcomes, optimize teaching management, and closely integrate data mining with educational practice. In this study, we pay special attention to how to enhance the analytical capabilities of educational content through effective algorithm design. We invented and proposed keyword distance weighted frequency (KDWF), which enhances TF-IDF and sentiment analysis by considering the distance between keywords and surrounding words. Unlike traditional TF-IDF, which only focuses on word frequency, KDWF gives higher weights to words with close distances,

thereby more accurately reflecting the text context relationship. In sentiment analysis, KDWF can identify words related to the sentiment tendency of keywords and improve the accuracy of sentiment recognition. This method has broad application potential in social media, market analysis, and data mining in the field of education.

1.2 Benefits and beneficiaries related to this research

Through this research we expect to gain the following benefits:

1.2.1 Provide a complete set of methods from data acquisition to data analysis and data mining.

1.2.2 We invented and proposed an algorithm and method to solve the anti-crawler problem to obtain data simply and reliably.

1.2.3 We invented and proposed 3 algorithms for data analysis and mining. It has wide application value.

1.2.4 We have explained the results of data analysis and data mining on the topic data of education-related discovery pages on TikTok, which can be used by people in educational practice. Such as hot words, hot phrases, hot topics, related phrase data classification, sentiment analysis, etc. Provide reference for educational practice.

In general, people can use the relevant content in this study for data acquisition, data analysis and data mining on any topic that people are interested in. People can use the relevant content in this study for research or commercial realization, etc. Beneficiaries include teachers, educational managers, managers in other fields, researchers in any field, managers in any field, and business people in any field, etc.

2. Related Theories and Research

This study uses some new algorithms to obtain, analyze and mine education-related data on TikTok. Among them, we invented and proposed a new algorithm and method to quickly and reliably obtain data on TikTok Discover page topics. Then we used 11 codes to conduct data analysis and data mining from different aspects and levels, including using 3 new algorithms that we invented and proposed for data analysis and data mining. They are "Keyword Distance Weighted Frequency", "Keyword Distance Weighted Frequency-Inverse Document Frequency" and "Keyword Distance Weighted Frequency-Emotion Analysis Frequency". Here we will introduce current theory and previous related research relevant to our study.

2.1 Theoretical Framework

Google is currently the world's largest search engine. A large amount of valuable data can be searched, providing users with massive data [9], [10]. Google uses crawlers to obtain data and then saves the obtained data to Google's servers. Then when the user searches on Google, Google provides the corresponding data to the user through the corresponding algorithm [11], [12]. In this study we use Google's advanced search function [9]. For example, we search intitle:education intext:the OR intitle:the site:https://www.tiktok.com/discover/ on Google to obtain data. It means that it is designed to retrieve two types of content from TikTok's Discover page: one is content that contains both "education" and "among" in the title, or content that contains "among" in the text and "education" in the title. First we prepared a word list of 500 common English words. Then use Python code to replace the keyword "the" in intitle:education intext:the OR intitle:the site:https://www.tiktok.com/discover/. Generates 500 search terms for advanced searches on Google. Then we search on Google using advanced search terms like intitle:education intext:the OR intitle:the site:https://www.tiktok.com/discover/. Then directly copy and paste these Google search results into a world file. Then we quickly extracted the topic name data of each TikTok discovery page topic through Python code, and used some filtering conditions to perform some corresponding screening. In this way we get the corresponding data. We then used Python code to quickly filter the collected data, such as deleting duplicate text and extracting relevant text. The restrictions imposed by the Google search engine, such as human-machine verification, can be easily resolved manually.

Searching and collecting a moderate amount of data on Google is reasonable and legal, but it still needs to follow legal

and ethical guidelines. According to the Digital Millennium Copyright Act (DMCA) and the General Data Protection Regulation (GDPR), when using public data, it is necessary to ensure that copyrights are not violated and personal privacy is respected.

Collecting data without the user's explicit consent may bring ethical issues. GDPR requires user consent when processing personal data and transparent notification of the purpose of data use. Therefore, data anonymization is crucial to avoid leaking personally identifiable information.

2.2 Key Concepts and Definitions

Google Advanced Search can filter Google's search results more precisely to find search results that better meet user requirements. And we use Python to quickly generate search terms for Google advanced search and use Python to quickly extract and filter out the data we need in the world file. Document data can be processed quickly and efficiently using Python code [13]. Data analysis and data mining techniques have a relatively long history, and are currently very popular and relevant technologies are developing rapidly [14]. Data analysis and data mining technology are currently widely used in various industries. Using these technologies can better understand the current situation and enable people to make better decisions about their businesses [15], [16].

2.3 Summarize

At present, digital technology is developing rapidly and is practical and powerful. All walks of life are using digital technology to upgrade. Digital technology brings many new opportunities and unlimited possibilities to all walks of life [17]. Several technologies invented and proposed in this study have strong practicality in the current environment. The method we proposed to obtain data can well avoid the anti-crawler problems that would exist if data were obtained through crawlers. And the targeted data people need can be obtained quickly and efficiently. Users can also replace the Google advanced search terms intitle:education intext:the OR intitle: the site:https://www.tiktok.com/discover/ with intitle:education intext:of OR intitle:of intext:"20.7M views" site: https://www.tiktok.com/discover/, etc., so as to be able to search for more detailed and more accurate data. Of course, this method and algorithm can also be used to collect data outside of Google. In this study we collect data on Google. And we proposed three new data analysis and data mining algorithms. These three new algorithms are very suitable for use in data analysis and data mining of data with keywords.

3. Research methods

3.1 Literature Review and Theoretical Analysis

In this study, we conducted a systematic literature review and theoretical analysis, including Google advanced search, using Python code to quickly and efficiently process document data, data analysis and data mining technology, etc. These related theoretical analyzes provide effective support for this research.

3.2 Interdisciplinary Research and Application

We applied it across disciplines in the field of education through some of the new algorithms we invented and proposed. We obtained, analyzed and mined education-related data on TikTok, conducted data analysis and data mining through 11 codes, and obtained information on education. Relevant and useful insights and knowledge. Such as hot words, hot phrases, hot topics, related phrase data classification, sentiment analysis, etc. Provide reference for educational practice. Make the education field more efficient through digital technology through an interdisciplinary approach, etc.

3.3 Empirical Research and Experimental Verification

Through experiments and empirical analysis, the effectiveness and reliability of the crawler and data mining methods in this scenario were verified. The experimental results were analyzed and discussed, and case analysis and conclusions were proposed, which provided important reference and guidance for subsequent research and practice.

3.4 System Design and Implementation

In this study, we used multiple new algorithms and methods we invented and proposed to obtain, analyze and mine education-related data on TikTok, and built a complete system to apply it in real application scenarios.

3.5 Algorithm Design and Optimization

In this study, we propose a new method of obtaining data on the Internet that is fast, simple, and reliable, which can well avoid the anti-crawler problems that would exist if data were obtained through crawlers. And the targeted data people need can be obtained quickly and efficiently. And we proposed 3 new data analysis and data mining algorithms. These 3 new algorithms are very suitable for use in data analysis and data mining of data with keywords.

3.6 Applications

In this study, we use multiple new algorithms and methods we invented and proposed to obtain, analyze and mine education-related data on TikTok, and apply them in real application scenarios. Provide reference and reference for people to use the multiple new algorithms and methods we invented and proposed in real application scenarios.

3.7 Dataset Construction and Annotation

We created a data set to acquire, analyze and mine education-related data on TikTok, including acquired data, filtered data, analyzed data, etc. Ensure the reliability of our research.

3.8 Comparison And Effectiveness Verification Of Keyword Distance Weighted Frequency (KDWF) With Existing Methods

In order to explore the method keyword distance weighted frequency (KDWF) and its advantages over existing methods, this section will analyze from multiple perspectives.

3.8.1 Comparison of KDWF with existing methods

When comparing KDWF with existing methods, we use the following non-traditional indicators for a more comprehensive evaluation:

Feature extraction ability:

KDWF: Combining the frequency of keywords with their distance in the text, KDWF can effectively capture contextual information and highlight the importance of keywords in different contexts. This method can better identify relevance and semantic relationships.

Existing methods: Many existing methods rely on simple frequency statistics and fail to fully utilize the distance relationship between words, which may lead to the neglect of some important information.

Information gain:

KDWF: By introducing distance weights, KDWF can better evaluate the information gain of keywords in the text. The relative position between words can reveal the hierarchy and relevance of the topic.

Existing methods: Traditional methods often only consider the number of occurrences of words, resulting in insufficient evaluation of information gain.

Robustness:

KDWF: This method is highly resistant to noise and redundant information in the text because it can distinguish the distance between important keywords and irrelevant words and reduce the impact of irrelevant information.

Existing methods: Many traditional methods perform poorly when faced with noisy data and cannot effectively extract useful features.

3.8.2 Verification of the effectiveness of KDWF

To verify the effectiveness of KDWF, we adopt a case analysis strategy:

By analyzing specific text instances, we show the keyword extraction results of KDWF and compare them with the results of existing methods. The advantages of KDWF in capturing contextual information and keyword importance can be illustrated by specific examples.

Through the above methods, we hope to be able to comprehensively evaluate the performance of KDWF and prove its advantages over existing methods in theory and practice. This will provide a solid foundation for future research and applications.

4. Collect, Analyze, and Mine Data

4.1 A New Method and Algorithm for Collecting Data

This research invents and proposes a new algorithm and method to obtain TikTok's discovery page topic data simply, quickly and efficiently. This is the world's first simple, fast and efficient algorithm and method to obtain data on the topic

of TikTok's discovery page. Of course, this new algorithm and method can also be used to collect data from other websites. And it is the world's first study to conduct data analysis and data mining on TikTok's discovery page topic data.

In this study we used Google's advanced search capabilities. For example, we search intitle:education intext:the OR intitle: the site:https://www.tiktok.com/discover/ on Google to obtain data. It means that it is designed to retrieve two types of content from TikTok's Discover page: one is content that contains both "education" and "among" in the title, or content that contains "among" in the text and "education" in the title.

For example, search for "intitle:education intext: the OR intitle:the site:https://www.tiktok.com/discover/" on Google Then you will get the following results:

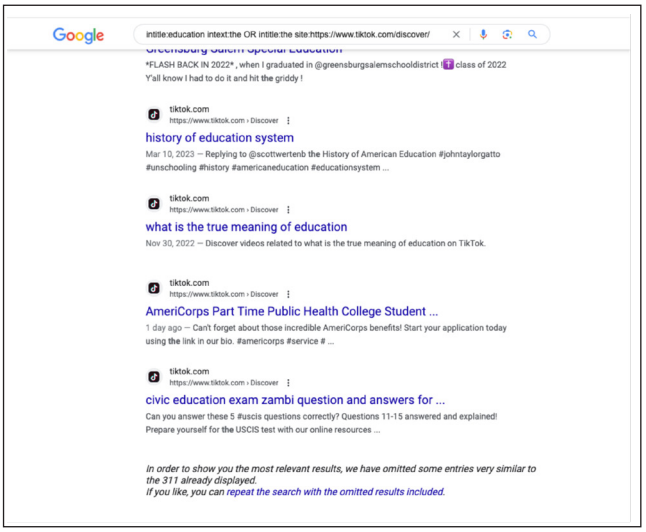


Figure 1. *We search on Google for "intitle:education intext: the OR intitle:the site:https://www.tiktok.com/discover/".*

First we prepared a word list of 500 common English words (500 common English words come from this website: https://www.smart-words.org/500-most-commonly-used-english-words.html) . (You can also use other word or phrase lists such as larger, smaller, other languages, keywords and strongly related word or phrase lists, etc.) Then replace intitle:education intext:the OR intitle with the Python code The keyword "the" in :the site:https://www.tiktok.com/discover/ generates 500 search terms for advanced search on Google. The code used is as follows (In this article, all Python code will use pseudo code.):

```
Import pandas as pd
Read CSV file into DataFrame (df)
Define text template with placeholder for word
Initialize empty list (replaced_texts)
For each word in df:
    Format template with word
    Append formatted text to replaced_texts
Create DataFrame (result_df) from replaced_texts
Save result_df to new CSV file
Print completion message
```

Then we search on Google using advanced search terms like intitle:education intext:the OR intitle:the site:https://www.tiktok.com/discover/. Then directly copy and paste these Google search results into a world file. (Of course, you can also use a crawler to automatically search on Google and then directly copy and paste these Google search results into a world file.) In this study, we did not log in to the Google account in guest mode when searching for data on Google. Use Google browser, and make the following settings for Google browser, set Display language to English, set Results language filter to English, set Results region to United States, and set Search customization to of. When we collect data and encounter "repeat the search with the omitted results included" shown at the bottom of the page in Figure 3, we click "repeat the search with the omitted results included". You can set the Google browser according to your needs to obtain popular search results in different countries, regions, languages, etc.

Then we used Python code to quickly extract the topic name data of each TikTok discovery page topic in the world file, and used some filtering conditions to perform some corresponding screening. In this way we get the corresponding data. We used this method to collect more than 10,000 rows of unique topic name data on TikTok's discovery page topics. Use Python code to quickly extract the topic name data of each TikTok discovery page topic. The code used is as follows:

```
Import libraries
```

Define function to check Chinese characters

Set folder and CSV paths

Initialize written_lines set

Open CSV for writing

For each .docx file:

Read document

For each paragraph:

If font size is 20pt and contains "education":

Truncate text at '|'

If valid line:

Write line to CSV

Add line to written_lines

Print message

Some of the data we obtained are as follows:

Table 1. *Some examples of data we obtained.*

text_column
Gerry Brooks Video in A Free Education
Education Alive School
john spencer education

Users can also replace the Google advanced search terms intitle:education intext:the OR intitle:the site:https://www.tiktok.com/discover/ with intitle:education intext:of OR intitle:of intext" 20.7M views " site: https://www.tiktok.com/discover/, etc., so as to be able to search for more detailed, more and more accurate data. Of course, this method and algorithm can also be used to collect data in search engines other than Google.

data analysis:

After we used our methods and algorithms to obtain data on the topic of the TikTok Discover page, we used the following 11 codes to analyze the collected data from different levels or aspects.

4.2 1st Code for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Read CSV and extract text column

Clean text (remove non-ASCII)

Compute TF-IDF values

Get total heat for each word

Convert to DataFrame, transpose, and rename columns

Save DataFrame to CSV

Print completion message

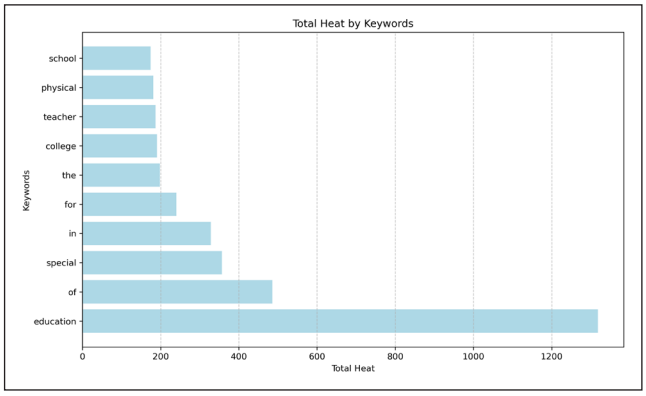


Figure 2. *The top 10 data output by the 1nd code.*

Analysis results:

We run this code [18]. In the output "File output by code 1.csv" file, there are a total of 9581 words and the "Total Heat" values corresponding to these 9581 words.

The top 10 are: education (1318.568818), of (485.6832716), special (356.6052461), in (328.508117), for (240.6092889), the (197.9038619), college (190.5411685), teacher (186.909698), physical (180.9941169), school (174.5185412).

Taking the ninth word "physical" as an example, we can predict that people may have relatively high interest or demand in knowledge related to "physical" or in the subject of physical. When we search and compare "biology" and "math" in the "File output by code 1.csv" file. Among them, "biology" is ranked 1106th among all words, and its "Total Heat" value is "4.824677856". Among them, "math" is ranked 208th among all words, and its "Total Heat" value is "19.94714983". Since the "Total Heat" value of "math" is greater than the "Total Heat" value of "biology". Therefore, we can predict that people may have higher interest or demand in knowledge related to "math" or corresponding subjects than to knowledge related to "biology" or related subjects.

4.3 Second Code for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Read CSV and select text column
Load BERT tokenizer and model

Compute BERT embeddings for each text
Calculate similarity matrix and heat scores
Add scores to DataFrame

Sort and reset index
Save DataFrame to CSV

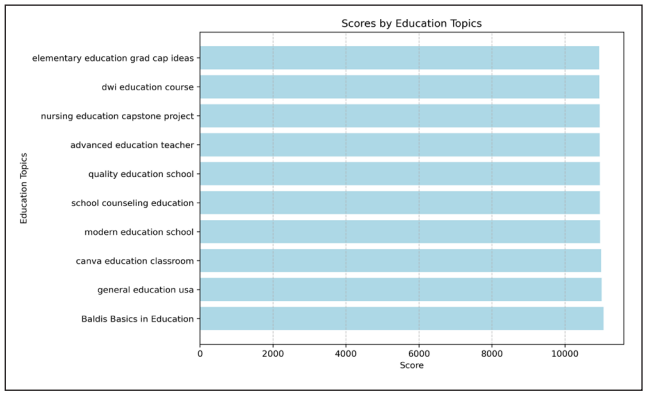


Figure 3. *The top 10 data output by the 2nd code.*

Analysis results:

We run this code [19]. In the "File output by code 2.csv" file it outputs, there are more than 16,000 rows of data and corresponding "score" values.

The top 10 are: Baldis Basics in Education (11058.722), general education usa (11003.078), canvas education classroom (10990.033), modern education school (10963.553), school counseling education (10956.008), quality education school (10955.442), advanced education teacher (10954.297), nursing education capstone project (10953.441), dwi education course (10944.557), elementary education grad cap ideas (10940.563).

Taking the data "quality education school" with the sixth-ranked "score" value as an example, we can predict that people value quality education, and people prefer schools that provide quality education.

4.4 3rd Code for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Read CSV and select text column
Split data into training and testing sets

Vectorize text data using TF-IDF
Train Naive Bayes classifier

Get predicted probabilities for test set
Rank texts by probabilities

Save ranking results to CSV

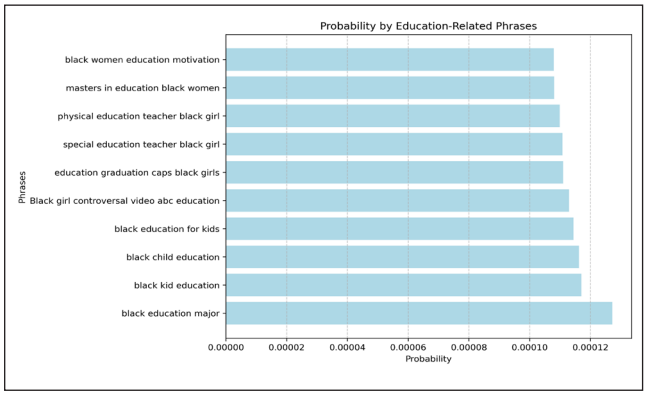


Figure 4. *The top 10 data output by the 3nd code.*

Analysis results:

We run this code [20]. In the "File output by code 3.csv" file it outputs, there are more than 3,000 rows of data and corresponding "Probability" values. The following are the first 3 rows of data sorted from high to low according to the "Probability" value:

Table 2. *The first 3 rows of data in the "File output by code 3.csv" file sorted from high to low according to the "Probability" value.*

Text	Probability
black education major	0.000127205
black kid education	0.000117005
black child education	0.000116268

The top 10 are: black education major (0.000127205), black kid education (0.000117005), black child education (0.000116268), black education for kids (0.000114405), Black girl controversal video abc education (0.000112992), education graduation caps black girls (0.000111029), special education teacher black girl (0.000110804), physical education teacher black girl (0.00010988), masters in education black women (0.000108051), black women education motivation (0.000107964).

From the first 3 rows of data, we can predict that educational content related to black will be very popular or concerned.

4.5 4rd Code for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Download stopwords
Read CSV file

Preprocess text data
Create dictionary and corpus

Run LDA model
Save model

Generate and save ranked topics

Analysis results:

We run this code [21], [22]. In the "File output by code 4.csv" file it outputs, there are a total of 10 subject data. The following is the data of the first topic:

Based on this topic data, we can predict that people may have a relatively high interest in information and content related to "major" in education.

4.6 Code 5 for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Table 3. *Data of the first topic in the "File output by code 4.csv" file.*

Topic	Words
0	department: 0.0837642252445221, major: 0.08170474320650101, system: 0.06884553283452988, higher: 0.0479467436671257, grade: 0.04165898263454437, program: 0.026927508413791656, institute: 0.024929115548729897, first: 0.02261144109070301, general: 0.01799621991813183, american: 0.016757341101765633

Read CSV file
Select text column

Preprocess text and extract features using TF-IDF
Apply KMeans clustering

Add cluster labels to dataframe
Save labeled dataframe to CSV

Reduce dimensions with t-SNE
Visualize clusters

Exit program

Analysis results:

We run this code [23], [24]. We use codes for classification. In the "File output by code 5.csv" file it outputs, the data is divided into 20 categories. You can find data with relatively high similarity in the same category for data analysis and data mining.

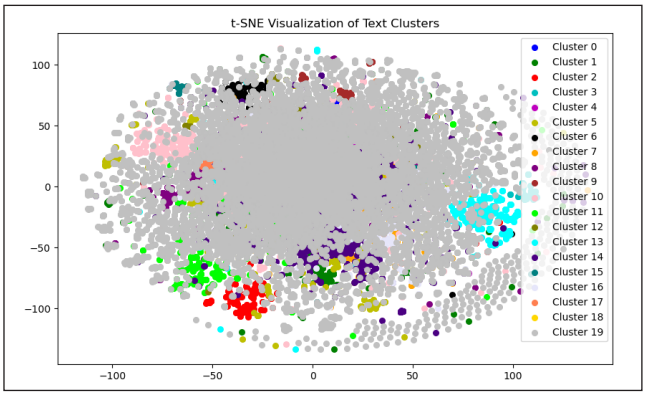


Figure 5. *Pictures of classification results.*

4.7 Code 6 for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Download models

Define functions for:
- Text processing
- Aspect extraction
- Sentiment determination

Read CSV, create pipeline, initialize counters

For each text:
Analyze, count, and print results

Save results to CSV

Calculate and print ratios
Analysis results:

We run this code [25]. The output is: Sentiment Ratios: {'positive': 0.07309252599691757, 'negative': 0.026165211075484064, 'neutral': 0.9007422629275984}. This result shows that people's attitude towards education is neutral to positive. And people can analyze changes in people's attitudes towards education in different time periods by collecting data in different time periods.

4.8 Code 7 for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Read CSV

Define functions:
- Build index
- Cosine similarity
- Term frequency

Main:
Create documents
Build index

For each document:
Calculate similarity

Sort and save results
Print similar documents
Analysis results:

We run this code [26]. We use the code to find the data we collected that are more similar to the text 'information and communication technology for education'. Of course, you can also search using other topics that interest you. In the "File output by code 7.csv" output file, we rank all the collected data according to the similarity with the text 'information and communication technology for education' from high to low. Here are the first 3 rows of data sorted by "Similarity" value from high to low:

Table 4. *The first 3 rows of data in the "File output by code 7.csv" file sorted from high to low according to the "Similarity" value.*

Document	Similarity
technology and education	0.707106781
technology education	0.577350269
communication education	0.577350269

You can use this method to quickly find the topic name data of the TikTok Discover page related to the topic you are interested in. Or used to analyze popular search words or popular search phrases, etc.

4.9 8th Code for Data Analysis and Data Mining:

The code used is shown below:

Import libraries

Read CSV and select text data

```
Convert text to bag-of-words

Define custom dataset class

Define Generator model

Define Discriminator model

Define GANomaly model:
    - Initialize generator and discriminator
    - Define training method

Main:
    Create dataset and data loader
    Initialize GANomaly model
    Train model

  Detect anomalies on test data

    Create results DataFrame
    Save results to CSV
    Analysis results:
```

We run this code [27], [28]. We used the code to find out and found 207 data that were marked as anomalies. This "anomaly" may mean that these sentences are different from other sentences, and we may have some unexpected gains when we analyze and mine these different data. The following is some of the data marked as anomalies:

Table 5. *Some of the data marked as anomalies.*

Text	Anomaly
What are you gonna even do with an education degree	TRUE
What are you going to do with a masters in education	TRUE
When to Apply for Uif for Education Assistants	TRUE

4.10 Code 9 for Data Analysis and Data Mining:

The code used is shown below:

```
Import libraries

Define weights for words

Read CSV and select text column

Set keyword

Initialize word scores

For each text in text data:
    Split text into words
    For each word in words:
        If word is keyword:
            For j from 1 to 10:
                If left word exists:
                    Update score for left word
                If right word exists:
                    Update score for right word

Create DataFrame from word scores
Round scores to one decimal

Save DataFrame to CSV
Print output file message
    Analysis results:
```

Here we invented and used the "Keyword Distance Weighted Frequency" algorithm for data analysis and mining. The "Keyword Distance Weighted Frequency" (KDWF) algorithm calculates the importance of words surrounding a keyword by assigning weights based on their proximity to the keyword in a sentence. Here's the general formula for this algorithm:

Weight calculation for words around a keyword:

Weight (wi, k) = max (0,1−0.1×Distance (wi, k))

wi is a word at position i,

k is the keyword,

Distance (wi, k) is the number of words between wi and k (to the left or right),

The weight decreases by 0.1 for each additional word distance from the keyword. (You can use different values according to different requirements.)

Score calculation for each word across multiple sentences: For a word www across multiple sentences S1, S2, ..., Sn containing the keyword k:

$$\textit{Total Score (w)} = \sum_{j=1}^{n} Weight\,(w, k\ in\ S_j)$$

where:

n is the number of sentences containing the keyword k,

Weight (w, k in Sj) is the weight of word w in the j^{th} sentence.

Handling multiple keywords: If a sentence contains multiple occurrences of the keyword k, the score for a word is the sum of its weights relative to each occurrence:

$$\textit{Total Score (w)} = \sum_{l=1}^{m} Weight\,(w, k_l)$$

where m is the number of keyword occurrences, and kl is the l^{th} occurrence of the keyword in the sentence.

In summary, the score of a word is determined by its proximity to the keyword, with closer words receiving higher weights. The total score of a word across sentences is the sum of its weighted scores.

The pseudo code of "Keyword Distance Weighted Frequency" (KDWF) is as follows:

```
Input: text_data, keyword, weights

Initialize: word_scores = empty dictionary

For each text in text_data:
words = text.lower().split()

For i, word in enumerate(words):
If word == keyword:
For j from 1 to 10:
If i - j >= 0:
left_word = words[i - j]
If left_word != keyword:
word_scores[left_word] += weights[j]

If i + j < len(words):
right_word = words[i + j]
If right_word != keyword:
word_scores[right_word] += weights[j]

Convert word_scores to DataFrame
Round scores
Save results to CSV
```

Because in a sentence, the words closest to a key word usually play an important role in the understanding and context of the key word. This is because adjacent words may provide more information, modification, or context about the keyword [29], [30].

For example, education in different lines of text such as "What Are You Gonna Do with Education", "What are you gonna even do with an education degree", "what to do with a masters degree in education cry" is counted as a keyword, The weight of the nearest word to the left and right of a keyword word appearing once is calculated as 1, and the weight of the second nearest word to the left and right of a keyword word to appear once is calculated as 0.9. The weight of the third word appearing once to the left and right of the keyword word is calculated as 0.8. The weight of the 10th word appearing once to the right and left of the keyword word is calculated as 0.1. But the word education is not counted. For example, if you use the three lines "What Are You Gonna Do with Education", "What are you gonna even do with an education degree" and "what to do with a masters degree in education cry", then in "What Are You Gonna Do with an education degree"? The score for "What" in "Do with Education" is 0.5. The score for "What" in "What are you gonna even do with an education degree" is 0.3. In "what to do

with a masters degree in education cry" the score of "What" is 0.3. The total score of "What" among these three pieces of data is 0.5+0.3+0.3 = 1.1.

In "What Are You Gonna Do with Education" the score for "are" is 0.6. The score for "are" in "What are you gonna even do with an education degree" is 0.4. There is no word "are" in "what to do with a masters degree in education cry", so the score is 0. The total score of "are" in these three pieces of data is 0.6+0.4+0 = 1.

There is no word "degree" in "What Are You Gonna Do with Education", so the score is 0. The score for "degree" in "What are you gonna even do with an education degree" is 1. There is no word "degree" in "what to do with a masters degree in education cry", so the score is 0. The total score of "degree" in these three pieces of data is 0+1+0 = 1.

In this way, the scores of all the words in the three lines of data "What Are You Gonna Do with Education", "What are you gonna even do with an education degree" and "what to do with a masters degree in education cry" are calculated in sequence.

If the input piece of data contains multiple keywords "education", in this case, we will consider the words surrounding each "education" keyword and accumulate the scores of each keyword. For example, in "what to do with a masters degree in education cry education", the score of "in" in the first "education" is 1, and the score of "in" in the second "education" is 0.8, then in "what The total score of "in" in "to do with a masters degree in education cry education" is: 1+0.8 = 1.8.

This algorithm is called "Keyword Distance Weighted Frequency".

Let's run this code. In the output "File output by code 9.csv" file, there are a total of more than 9,000 words and their corresponding "Score" values.

Taking the 10th word "minecraft" sorted from high to low according to the "Score" value as an example, we can predict that people may have relatively high interest or demand for "minecraft1"-related content, and can use "minecraft1"-related content in education. When we search and compare "art" and "science" in the "File output by code 1.csv" file. Among them, "art" is ranked 156th among all words, and its "Total Heat" value is "44.9". Among them, "science" is ranked 125th among all words, and its "Total Heat" value is "53.5". Since the "Score" value of "science" is greater than the "Score" value of "art". Therefore, we can predict that people may have higher interest or demand in knowledge related to "science" or corresponding subjects than to knowledge related to "art" or corresponding subjects.

4.11 Code 10 for Data Analysis and Data Mining:

The code used is shown below:

Import library

Read CSV file

Extract Score column

Calculate sum of Score column

Calculate result as Score divided by sum

Insert result into new column in DataFrame

Save DataFrame to new CSV file

Print success message

Import libraries

Read CSV file

Select text column for analysis

Initialize CountVectorizer

Transform text to word count vector

Initialize TfidfTransformer

Fit transformer to word count vector

Get list of words

Get IDF values

Create DataFrame with words and IDF values

Save DataFrame to CSV file

Print success message

Import pandas

Read first CSV file

Read second CSV file

Merge both DataFrames on 'Word'

Calculate product of specific columns

Save merged DataFrame to CSV file

Print success message

Analysis results:

The "Keyword Distance Weighted Frequency" algorithm we invented and proposed can be used alone or in combination with other algorithms. For example, here we use the "Keyword Distance Weighted Frequency" algorithm in combination with the "Inverse Document Frequency". A new algorithm "Keyword Distance Weighted Frequency-Inverse Document Frequency" is formed. Inverse Document Frequency measures the importance of words in the entire document collection. This algorithm combines the advantages of Keyword Distance Weighted Frequency and Inverse Document Frequency.

Let's run this code. In the output "File output by code 10.csv" file, there are a total of more than 9,000 words and their corresponding "Keyword Distance Weighted Frequency-Inverse Document Frequency" values, etc.

Taking the data "2023" with the 8th ranked "Keyword Distance Weighted Frequency-Inverse Document Frequency" value as an example, we can predict that the data we collect contains a large amount of data related to 2023. Taking the 7th ranked data "bad" as an example, we can predict that some people may be dissatisfied with educational or education-related content.

Comparison with Existing Methods

Although the KDWF-IDF (Keyword Distance Weighted Frequency-Inverse Document Frequency) algorithm and its variants are regarded as novel feature representation methods, a rigorous quantitative comparison with traditional text processing algorithms is crucial. To evaluate the performance of KDWF-IDF compared to classical algorithms, this study compares it with the traditional TF-IDF algorithm. In the process, cosine similarity is used to quantify the similarity between the text feature vectors generated by the two algorithms.

The cosine similarity calculated by the feature vectors of KDWF-IDF and TF-IDF is 0.9978, indicating that the two algorithms have extremely high consistency in text feature representation. A cosine similarity close to 1 means that the feature vectors generated by KDWF-IDF and TF-IDF are almost exactly the same in direction. This result shows that the KDWF-IDF method is basically consistent with TF-IDF in maintaining the core information of text features, while introducing a mechanism based on keyword distance weighting, which may provide additional fine-tuning capabilities in specific application scenarios.

Although the cosine similarity results show a high similarity between KDWF-IDF and TF-IDF, KDWF-IDF can still improve the flexibility and accuracy of feature expression in specific situations through its unique weighting mechanism. Therefore, KDWF-IDF not only retains the advantages of the traditional TF-IDF algorithm, but also has the potential to further improve performance in text analysis tasks. This comparison provides quantitative evidence for the effectiveness of KDWF-IDF, proves the robustness of the algorithm on traditional benchmarks such as TF-IDF, and lays the foundation for further research on its advantages in practical applications.

4.12 11th Code for Data Analysis and Data Mining:

The code used is shown below:

Import pandas and Counter

Read CSV file

Convert DataFrame to list of tuples

Count occurrences of rows

Add count to DataFrame

Drop duplicate rows

Save updated DataFrame to CSV file

Import pandas

Read CSV file

Filter out rows with 'education' in 'aspect' column

Extract 'Count' column

Calculate sum of 'Count'

Normalize 'Count' values

Insert normalized results into DataFrame

Save updated DataFrame to new CSV file

Import pandas

Read first CSV file

Read second CSV file

Merge DataFrames on the specified columns

Multiply specific columns and store result in a new column

Save the modified DataFrame to a new CSV file

Import pandas

Read CSV file

Select sentiment and frequency columns

Initialize counters and total frequency

For each row in sentiment column:

Add frequency to total

Update sentiment counters

Calculate total count of sentiments

Calculate percentages for each sentiment

Print total frequency and sentiment sums

Print sentiment percentages

Analysis results:

Here we use the "Keyword Distance Weighted Frequency" algorithm we invented and proposed in conjunction with sentiment analysis. A new algorithm "Keyword Distance Weighted Frequency-Emotion Analysis Frequency" is formed. This algorithm can perform more accurate sentiment analysis when there are keywords.

Let's run this code. The result is Total Sum: 0.0014664594825907539, Sum of Negative Sentiment: 9.278729754058212e-05, Sum of Neutral Sentiment: 0.0007849887007069968, Sum of Positive Sentiment: 0.0005886834843431 The proportion of various emotional results calculated by 047 is: Percentage of Negative Sentiment: 0.0632730045678874, Percentage of Neutral Sentiment: 0.5352951854627558, Percentage of Positive Sentiment: 0.40143180996935685.

From this we can know that people's emotions towards education are positive. And the Percentage of Positive Sentiment is as high as: 0.40143180996935685.

And we can use two or more of the above 11 codes to comprehensively perform data analysis and data mining. For example, we can first use the 5th code or the 7th code to find the topic data strip that interests you, and then use the 2nd code or the 3rd code to check the popularity of the topic data strip that interests you, etc.

The insights gleaned from our data analysis of TikTok's discovery page topics reveal significant trends that can be instrumental in shaping educational practices and policies. For instance, the prominence of terms related to "quality education" suggests an increasing public interest in high-standard educational environments. This insight could guide educators and administrators in developing programs that emphasize quality and effectiveness, ultimately improving student outcomes.

Moreover, our findings related to the popularity of certain subjects, such as "math" over "biology," indicate where educational resources might be allocated for greater impact. Educational institutions could leverage this data to enhance their curriculum offerings, prioritize professional development for educators in high-demand subjects, and create targeted marketing strategies to attract students.

Furthermore, the neutral-to-positive sentiment observed in attitudes toward education highlights an opportunity for educational stakeholders to cultivate this positive perception. Engaging with communities through social media platforms like TikTok can foster an interactive learning environment, encouraging student participation and community support.

Additionally, the versatility of our proposed data collection method allows for its application beyond TikTok, making it

a valuable tool for researchers and educators in various fields. By adapting this methodology to different platforms and content types, we can continuously refine our understanding of public interest in education and related topics.

Ultimately, the insights derived from this study underscore the importance of data-driven decision-making in education. By systematically analyzing trending topics and public sentiment, educators and policymakers can make informed choices that enhance learning experiences and address the evolving needs of students and communities.

5. Results and analysis

In this section, we introduce the relevant results and analysis of our use of some new algorithms to obtain, analyze and mine the topic data of education-related pages on TikTok Discover. We used a new algorithm and method we invented to quickly and reliably obtain data on TikTok Discover page topics, and used 11 codes to conduct data analysis and data mining. This includes using 3 new algorithms we invented and proposed for data analysis and data mining. They are "Keyword Distance Weighted Frequency", "Keyword Distance Weighted Frequency-Inverse Document Frequency" and "Keyword Distance Weighted Frequency-Emotion Analysis Frequency".

We employed 11 codes to perform comprehensive data analysis and mining on various aspects and levels of the acquired data. These encompass the following and additional items:

5.1 Analyze popular data from different aspects and levels:

This includes finding popular search words, popular search phrases, popular data bars, popular topics, and more in the data. These include using codes 1, 2, 3, 4, 9, 10, etc. Including using the "Keyword Distance Weighted Frequency" and "Keyword Distance Weighted Frequency-Inverse Document Frequency" algorithms we invented and proposed for data analysis and mining.

5.2 Allow users to find relevant data they are interested in simply and efficiently:

This involves utilizing classification and search data to enable consumers to easily and effectively locate pertinent information that aligns with their interests, including the utilization of codes such as 5 and 7.

5.3 Emotion analysis:

We employ sentiment analysis to assess the sentiment expressed in the gathered data. After careful analysis, we have determined that individuals generally hold favorable sentiments towards education. These include utilizing codes 6, 11, and so on. It encompasses the utilization of our self-developed "Keyword Distance Weighted Frequency-Emotion Analysis Frequency" technique for data analysis and mining.

5.4 Find possible anomalous data:

We examine the gathered data to detect potential irregularities. Individuals have the ability to observe and examine this potential irregularity data in order to identify any distinct patterns or deviations. These include utilizing code 8, among other methods.

6. Conclusions and future directions

In this section, we will provide a concise overview of the research and propose potential future research areas that involve the utilization of novel algorithms for acquiring, analyzing, and extracting education-related page topic data from TikTok Discover.

6.1 Summary of Findings

We conducted a comprehensive analysis and extraction of educational content on TikTok Discover. By examining popular data from many perspectives and levels, our research aimed to facilitate users in efficiently locating important information for future investigation. Analysis, sentiment analysis, and potential anomalous data were detected. The analysis and extraction of this data offer educational practitioners' significant insights that can be utilized to direct educational practice. We present novel algorithms and methodologies for the collection, analysis, and extraction of data.

6.2 A Complete Set of Methods from Data Collection to Data Analysis and Mining

In this study, we developed and used a complete set of

methods from data collection to data analysis and mining to study the acquisition, analysis and mining of education-related page topic data on TikTok Discover. This complete set of data collected Data analysis and mining methods can not only be used for education-related research, this complete set of methods and processes can be used in almost every industry. People can use this complete set of methods and processes to use data for research or commercial monetization.

6.3 A New Algorithm and Method for Obtaining Data

This research invented and proposed a new algorithm and method that can quickly and reliably obtain data on the theme of the TikTok Discover page. This is the world's first simple, fast, and efficient algorithm and method to obtain data on the theme of the TikTok Discover page. Of course, this new algorithm and method can also be used to collect data from other websites. The reason why we invented this new algorithm and method to obtain data on the Internet is because if people want to obtain data from some websites such as TikTok through crawlers, they may encounter many complex anti-crawler problems. The algorithm and method we invented can solve this problem very well, simple and efficient.

6.4 3 New Algorithms for Data Analysis and Data Mining

In this study, we invented and proposed three new algorithms for data analysis and data mining. They are "Keyword Distance Weighted Frequency", "Keyword Distance Weighted Frequency-Inverse Document Frequency" and "Keyword Distance Weighted Frequency-Emotion Analysis Frequency". People can use these three new algorithms for data analysis and data mining in various fields. These three new algorithms are particularly suitable for data analysis and data mining in data with keywords.

6.5 future Direction

In future research, we can collect this complete set of data into data analysis and mining methods, this new algorithm and method for obtaining data, and these three new algorithms for data analysis and data mining. Research in more detailed education-related directions. Such as "Computer Science Education", "Science Education", "Application of Artificial Intelligence in Education", "Science Education" and other more detailed research directions related to education. This complete set of data collection to data analysis and mining methods and these 3 new algorithms for data analysis and data mining can also be used in other fields such as "art", "entertainment", "artificial intelligence", "Science" and other research directions in other fields. And we can collect data in different time periods, conduct data analysis and data mining in certain fields in different time periods, and compare changes in different time periods.

The multiple algorithms invented and proposed in this study can be applied not only to the field of education but also to many other fields such as medicine, entertainment, art, and culture, agriculture, manufacturing, etc.

In future work, people can test on larger data sets to improve the accuracy of the algorithm. In addition, if a large amount of relevant data of a website is obtained through Google, there may be potential ethical and legal issues, which is also worthy of attention.

In future research directions, in addition to continuing to deepen research in the field of education, we can also apply this method and algorithm to the following specific directions:

Detailed research on educational topics: Future research can focus on more specific directions such as "user behavior analysis of online education platforms", "the application effect of virtual reality in education", and "differences in educational needs under different cultural backgrounds". This will help improve the pertinence and efficiency of educational content.

Cross-domain application: The application of algorithms should not be limited to the field of education, but can also be extended to fields such as healthcare (such as sentiment analysis of patient health data), art and entertainment (such as popular trend prediction), and agriculture (such as analysis of market demand for agricultural products). This will greatly improve the accuracy and operability of data-driven decision-making.

In terms of potential applications, this set of methods and algorithms has a wide range of application potential and can be used by various industries for data-driven tasks such

as market analysis, user behavior prediction, and trend discovery. These tools are not only suitable for academic research, but also can provide business insights for enterprises, thereby improving the accuracy of business decisions.

As for research limitations, our current work is mainly based on educational topic data from the TikTok Discover page, and may face limitations of different data sets and platforms in broader applications. In addition, algorithm performance may encounter bottlenecks when processing larger-scale data, especially in the analysis and mining of real-time data. Therefore, future research needs to focus on how to optimize algorithm performance on larger data sets while ensuring that data acquisition and processing comply with relevant ethical and legal regulations.

7. References

[1] R. S. J. D. Baker and K. Yacef. "The state of educational data mining in 2009: A review and future visions." *Journal of Educational Data Mining,* Vol. 1, No. 1, pp. 3-17, 2009.

[2] G. Siemens and R. S. J. d. Baker. "Learning analytics and educational data mining: towards communication and collaboration." *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* Vancouver, Canada, pp. 252-254, 2012.

[3] R. S. J. d. Baker. "Data mining for education." *In: McGaw, B., Baker, E., Peterson, P. (eds.) International Encyclopedia of Education,* 3rd edn., Vol. 7, Elsevier, Oxford, pp. 112-118, 2010.

[4] P. Baepler and C. J. Murdoch. "Academic analytics and data mining in higher education." *International Journal for the Scholarship of Teaching & Learning,* Vol. 4, No. 2, 2010.

[5] R. S.Baker, A. T. Corbett, and K. R. Koedinger. "Detecting student misuse of intelligent tutoring systems." *Proceedings of 7th International Conference (ITS2004),* Maceió, Alagoas, Brazil, pp. 531-540, 2004.

[6] P. Long and G. Siemens. "Penetrating the Fog: Analytics in Learning and Education." *EDUCAUSE Review,* Vol. 46, No. 5, pp. 30-40, 2011.

[7] K. E.Arnold and M. D. Pistilli. "Course Signals at Purdue: Using Learning Analytics to Increase Student Success." *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge,* pp. 267-270, 2012.

[8] D. Gašević, S. Dawson, and G. Siemens. "Let's Not Forget: Learning Analytics Are About Learning." *TechTrends,* Vol. 59, No. 1, pp. 64-71, 2015.

[9] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. "Analysis of a Very Large Web Search Engine Query Log." *ACM SIGIR Forum,* Vol. 33, No. 1, pp. 6-12, 1999.

[10] R. Jones and K. L. Klinkner. "Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs." *Proceedings of the 17th ACM conference on Information and knowledge management,* pp. 699-708, 2008.

[11] T. Y. Liu. "Learning to Rank for Information Retrieval." *Foundations and Trends® in Information Retrieval,* Vol. 3, No. 3, pp. 225-331, 2009.

[12] S. Brin and L. Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems,* Vol. 30, pp. 107-117, 1998.

[13] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* O'Reilly Media, Inc., 2009.

[14] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques.* Third Edition, Morgan Kaufmann Publishers is an imprint of Elsevier, 2011.

[15] H. Chen, R. H. Chiang, and V. C. Storey. "Business intelligence and analytics: From big data to big impact." *MIS Quarterly,* Vol. 36, No. 4, pp. 1165-1188, 2012.

[16] U. Fayyad, G. P. Shapiro, and P. Smyth. "From data mining to knowledge discovery in databases." *AI Magazine,* Vol. 17, No. 3, pp. 37-54, 1996.

[17] E. Brynjolfsson and A. McAfee. *The Second Machine*

Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company, New York, London, 2014.

[18] K. S. Jones. "A statistical interpretation of term specificity and its application in retrieval." *Journal of Documentation,* Vol. 28, No. 1, pp. 11-21, 1972.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. "Attention is All You Need." *Advances in Neural Information Processing Systems,* pp. 5998-6008, 2017.

[20] T. Bayes. "An essay towards solving a problem in the doctrine of chances." *Biometrika,* Vol. 45, No. 3-4, pp. 296-315, 1958.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan. "Latent Dirichlet Allocation." *Journal of Machine Learning Research,* Vol. 3, pp. 993-1022, 2003.

[22] M. Hoffman, F. Bach, and D. Blei. "Online learning for latent dirichlet allocation." *Advances in Neural Information Processing Systems,* Vol. 23, pp. 1-9, 2010.

[23] J. MacQueen. "Some methods for classification and analysis of multivariate observations." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, California, USA: University of California Press, pp. 281-297, 1967.

[24] L. v. d. Maaten and G. Hinton. "Visualizing Data using t-SNE." *Journal of Machine Learning Research,* Vol. 9, pp. 2579-2605, November, 2008.

[25] W. Xue and T. Li. "Aspect Based Sentiment Analysis with Gated Convolutional Networks." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics,* Melbourne, Australia, pp. 2514-2523, 2018.

[26] H. E. Williams, J. Zobel, and D. Bahle. "Fast Phrase Querying With Combined Indexes." *ACM Transactions on Information Systems,* Vol. 22, No. 4, pp. 573-594, 2004. doi.org/10.1145/1028099.1028102.

[27] V. Chandola, A. Banerjee, and V. Kumar. "Anomaly detection: A survey." *ACM Computing Surveys (CSUR),* Vol. 41, No. 3, pp. 1-58, 2009.

[28] I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative adversarial nets." *Advances in Neural Information Processing Systems,* Vol. 27, 2014.

[29] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. "Indexing by latent semantic analysis." *Journal of the American Society for Information Science,* Vol. 41, No. 6, pp. 391-407, 1990.

[30] C. D. Manning, P. Raghavan, and H. Schütze. *An Introduction to Information Retrieval.* Cambridge University Press, Cambridge, 2009.

Safeguarding Skies: Airport Cybersecurity in the Digital Age

Suphannee Sivakorn*, Nuttaya Rujiratanapat*, Yotsapat Ruangpaisarn*,
Chanond Duangpayap* and Sakulchai Saramat*

Received: March 6, 2024
Revised: November 16, 2024
Accepted: November 25, 2024

* Corresponding Author: Suphannee Sivakorn, E-mail: suphannee_si@rmutto.ac.th

DOI: 10.14416/j.it/2025.v2.005

Abstract

The aviation industry faces significant vulnerabilities from both physical and cybersecurity threats, highlighting the urgent need for enhanced cybersecurity measures amid increasingly sophisticated attacks. This paper systematically reviews emerging threats at airports, analyzing real-world incidents and relevant literature while mapping risks to the MITRE ATT&CK Matrix, a widely recognized knowledge base for categorizing cyberattack tactics, techniques, and procedures. This is the first to apply the MITRE Matrix to airport security risks, offering a novel approach to understanding and mitigating these challenges. Building on this analysis, the paper advocates for modern cybersecurity defense models, emphasizing Cybersecurity Frameworks and Zero Trust Architecture, as well as critical measures for supply chain risk management and strategies to mitigate ransomware and DoS attacks. Our analysis provides insights into vulnerabilities and actionable recommendations, serving as a comprehensive guide for aviation stakeholders to strengthen defenses against evolving cybersecurity threats.

Keywords: Airport Cybersecurity, Aviation Cybersecurity, Cyber Threats in Aviation, Critical Infrastructure, Smart Airport.

1. Introduction

In the physical domain, airport security entails the screening of passengers, baggage, cargo, and the fortification of secure areas within the airport premises. Airport authorities undertake substantial efforts to prevent unlawful interference and ensure their security practices meet current standards. However, in the realm of technology, cybersecurity often receives inadequate attention than physical security, as evidenced by a 2017 survey of the top major airports in Europe and the U.S., wherein only 59% of respondents claimed to have an effective cybersecurity policy [1]. This oversight is concerning, especially given a 530% increase in cyberattacks within the aviation industry from 2019 to 2022 [2]. Recent initiatives by authorities, including the Transportation Security Administration (TSA) and the International Air Transport Association (IATA), emphasize the need for enhanced cybersecurity measures, mandating proactive steps to mitigate cyber threats [3], [4].

In an effort to strengthen the cybersecurity posture of the aviation industry, this study comprehensively explores existing literature and recent data on airport cybersecurity. We examine airport technologies, security concerns, and recent cyber incidents in Section 2, and outline our research methodology in Section 3. Section 4 presents a systematic literature review of airport cybersecurity from the past five years, and Section 5 categorizes the current landscape and identifies relevant risks. We map these risks to the MITRE ATT&CK Matrix for Enterprise, a widely recognized cybersecurity knowledge base for developing effective security strategies in Section 6. Section 7 presents modern cybersecurity defense strategies, advocating for Cybersecurity Frameworks and Zero Trust Architecture and highlights essential measures for mitigating supply chain risks, ransomware, and denial-of-service (DoS) attack. Section 8 discusses key challenges

* Department of Computer Science, Faculty of Science and Technology, Rajamangala University of Technology Tawan-ok

and outlines future research directions in this field, with the conclusion presented in Section 9.

The major contributions of this paper are as follows:

- We conducted an extensive analysis of recent cyberattacks and a literature review from the past five years, categorizing key cybersecurity risks into nine distinct areas to clarify the challenges faced by modern airport operations.
- We correlate these identified risks with the MITRE ATT&CK Matrix for Enterprise, making this paper the first to map airport security risks to the Matrix. This serves as a valuable tool for exploring each risk through practical defenses and best practices outlined in the MITRE knowledge base.
- Based on our analysis, we advocate for adopting modern security practices, including Cybersecurity Frameworks and Zero Trust Architecture, along with practical measures to defend against evolving airport cyber threats.

2. Background

Understanding airport technologies, threat actors, and recent high-profile incidents highlights the need for stronger security measures. This section examines airport technology, cybersecurity concerns, the landscape of cyber threat actors, and notable attack incidents.

2.1 Airport Technology and Security Concerns

Airport operations have transformed significantly to support the global aviation industry growth, leading to advancements in technology aimed at enhancing efficiency and service. The evolution of airport technology is delineated into four stages: Airport 1.0 - 4.0.

Airport 1.0 primarily focuses on ensuring the physical safety of operations, with no security concerns [5].

Airport 2.0 incorporates technologies for collaboration technologies such as IP telephony, broadband, and Wi-Fi [6]. Following the events of 9/11, the TSA was established to oversee transport security [7].

Airport 3.0 or "Smart Airport" integrates the Internet of Things (IoT), Artificial Intelligence, smart sensors to enhance passenger experience [8].

Airport 4.0 emphasizes the use of technologies to support airport operations and enhance passenger experiences, with a focus on data analytics as a core capability [9], [10].

While these airport advancements offer notable benefits, they are susceptible to interference and malicious modification without proper deployment.

2.2 Airport Cyber Threat Actors

Incidents of air terrorism have led adversaries to adapt their tactics in both physical and cyber domains [7]. This section outlines four types of cyber threat actors:

1: Advanced Persistent Threat (APT): Organized groups or foreign governments motivated by political or economic goals. They often target critical infrastructure, including airports [11] - [16].

2: Cybercrime: Attackers in this category target systems for valuable and sensitive information from passengers and airport employees [17] - [21], particularly those that are internet-facing or publicly accessible [17].

3: Peer Group Service Disruption: Hackers motivated by political agendas or beliefs whose focus is on service disruptions rather than data theft and financial gain [22] - [30].

4: Insider Threats: Risks that originate from within the organization, typically associated with current or former members of the organization, it may also arise from third parties such as contractors and temporary workers.

2.3 Recent Notable Cybersecurity Incidents

The urgency of cybersecurity in airports has become apparent through numerous studies [1], [5], [8]. From 2022 to 2024, various notable incidents have highlighted cyberattacks affecting airport operations and public perception. Table 1 details and categorizes these incidents by attack type, including Denial-of-Service, Ransomware, Vulnerability Exploitation, and Phishing. This analysis will assist in identifying cybersecurity risks and associated attack vectors for airports in Section 5.

Denial-of-Service (DoS). Recently, several major U.S. airports were targeted by coordinated DoS attacks [29], which overloaded airport servers. Similar incidents have occurred at various airports worldwide [23] - [31]. In some cases, attackers have demanded cryptocurrency payments to stop the attacks, exploiting the difficulty of tracing such transactions [32].

Ransomware. The airport industry has experienced a significant increase in ransomware attacks, primarily due to system vulnerabilities and phishing attempts [33] - [35], [38]. In 2024, notable incidents led to delays in passenger processing and flight schedules [34], [35], while others resulted in the leakage of sensitive data [33].

Vulnerability Exploitation poses significant risks for airports, which rely on variety of software applications for their operations, ranging from flight scheduling, air traffic control, baggage handling, and security systems [1], [5], [8]. This diversity broadens the attack surface, introducing potential vulnerabilities and inadequate security practices from vendor [17], [19], [20] - [22], [25], [36], [37].

Phishing. Although no new incidents have been disclosed recently, phishing remains a significant threat with airport employees and customers vulnerable to scams [38]. During a recent global outage linked to CrowdStrike [39], opportunistic hackers exploited the situation by sending fake information to scam IT personnel [40].

3. Research Methodology

This study synthesizes recent airport cybersecurity incidents, literature, insights from online sources, and an examination of various cybersecurity standards and policies. Our goal is to identify and delineate the prevailing cybersecurity threats and risks, categorizing them in alignment with the MITRE ATT&CK Matrix. By systematically mapping these risks to the Matrix, we provide a strategic approach for mitigating cybersecurity risks and applying effective defense techniques based on current best practices.

4. Literature Review

We conduct a comprehensive literature review by searching academic databases, including Google Scholar, ResearchGate, Scopus, and Web of Science, using the following keywords: "airport AND cybersecurity", "aviation AND cybersecurity", "airport AND information security", "airport AND IT security", "smart airport", and "airport AND cyber risk". Our focus was on peer-reviewed studies from the last five years addressing the impacts of cybersecurity on modern airports, challenges, and risks, while excluding studies on airport physical security or unrelated aviation topics. In total, we reviewed 31 publications, categorizing them into eight primary areas: (1) Critical Infrastructure, (2) IoT, Smart Devices, and AI Technology, (3) Supply Chain, (4) Cybersecurity Awareness, (5) Risk and Threat Analysis, (6) Standards and Regulations, (7) Cybersecurity Framework, and (8) Case Studies and Surveys. Table 2 presents the number of publications in each category and highlights specific airport security risks where applicable.

Critical Infrastructure. This category focuses on the critical infrastructures of airports, such as communication protocols, Air Traffic Management (ATM), and surveillance technologies [41] - [47]. These studies analyze vulnerabilities and mitigations, with examples including man-in-the-middle attacks between aircraft and ground control [42], the lack of encryption in the Automatic Dependent Surveillance-Broadcast (ADS-B) [43], [45], and the security concerns related to digitization of the Traffic Collision Avoidance System (TCAS) [44]. These findings underscore the need to address cybersecurity risks in airports. To this end, we associate these publications related to airport security risks as follows: (1) Insecure Network Architecture, (2) Malware and Ransomware (3) Data Breach and (4) DoS.

IoT, Smart Devices, and AI Technology. Research here addresses cybersecurity risks from IoT devices and AI technologies used in airport operations [1], [5], [8], [48] - [50]. Consequently, we associate these publications with specific risks, including: (1) Public-facing Access, (2) Insecure Network Architecture, (3) Internet-facing Applications and Services,

Table 1. *Summary of Publicly Disclosed Notable Cybersecurity Incidents at Airports (2022-2024).*

Attack Incident	Attack Incident Summary	Year	Attack Technique	Impact on Airport Services			Threat Actor Type*
				Operational Disruption	Website or Application	Data Leakage	
Seattle Airport [33]	The Port of Seattle confirmed that a ransomware attack caused significant outages at Seattle-Tacoma International Airport, affecting services like Wi-Fi, check-in kiosks, and passenger displays. The attack also resulted in some data being stolen and encrypted	2024	Ransomware	•		•	2
Pau-Pyre´ne´es Airport [34]	Pau-Pyre´ne´es Airport was hit by a ransomware attack from the MONTI group, which exfiltrated sensitive data and published it on the dark web.	2024	Ransomware			•	2
Croatia's Split Airport [35]	Split Airport in Croatia experienced a ransomware attack that resulted in flight cancellations and delays. The incident has been linked to the Akira group, which is associated with the Russian-based Conti group.	2024	Ransomware	•			2
Los Angeles International Airport [36]	A hacker group, IntelBroker, exploited the airport's CRM system vulnerability, accessing a database with sensitive information (e.g., private plane owners' full names, emails, CPA numbers)	2024	Vulnerability Exploitation			•	2
Copenhagen Airport [31]	The airport website was taken offline. Passengers were advised to use their smartphones as an alternative to receive updates on their flights.	2024	DoS		•		unknown
Beirut International Airport [17]	Hackers displayed a message on screens at the airport threatening to bomb the airport.	2024	unknown	•			3
Long Beach Airport [18]	Part of Long Beach City system cyberattack. The website was taken offline.	2023	Ransomware		•		2
Cairo International Airport [23]	The airport website and mobile application were taken down. Anonymous Collective hacker group took credit for the attack.	2023	DoS		•		3
Czech and Prague Airport [24]	The airport website was taken offline.	2023	DoS		•		3
Quere´taro Intercontinental Airport [19]	LockBit ransomware hacker group took credit for the attack, threatening to leak data. The airport claimed that stolen data was in the public domain.	2023	Ransomware			•	2
Montreal-Trudeau International Airport [25]	Border checkpoint outages e.g., check-in kiosks and electronic gates caused significant delays. A hacker group, NoName057(16) claimed responsibility.	2023	DoS	•			3
Charles de Gaulle Airport [26]	The airport website was taken offline. Cybercriminal, Dark Storm, claimed responsibility.	2023	DoS		•		3
UK Airports [27]	The airport website was taken offline. UserSec hacker group claimed the responsibility.	2023	DoS		•		3
Kenya Airports Authority [20]	Data breach incident. Attackers released data including procurement plans, physical plans, site surveys, invoices and receipts.	2023	unknown			•	2
German Airports [28]	Several German airports' websites were taken offline.	2023	DoS		•		3
US Major Airports [29]	Coordinated DoS attacks targeted several major US airports. A hacker group, Killnet claimed responsibility.	2022	DoS		•		3
Brazil Airports [37]	Rio de Janeiro airport's electronic displays were hacked to show pornographic movies instead of ads and flight info.	2022	Vulnerability Exploitation	•			unknown
Italian Airports [30]	Coordinated DoS attacks targeted several Italian airports. A hacker group, Killnet claimed responsibility.	2022	DoS		•		3
Swissport at Zurich Airport [21]	Airport ground services and air cargo, Swissport, were hit with a ransomware attack causing Zurich Airport operation disruptions.	2022	Ransomware	•			2

The sign ● indicates the impact on airport services from the attack.

*The Threat Actor Type number delineates the category of cyber threat actors in Section 2.2

(4) Malware and Ransomware, (5) Data Breach, and (6) DoS as these threats exploits internet connectivity used by IoT and smart devices. Furthermore, vulnerabilities in these products can lead to supply chain attacks [51].

Supply Chain and Third Party. This category examines cybersecurity vulnerabilities arising from supply chain and third-party partnerships. For example, Hann (2020) emphasized the complex socio-technical landscape of the ATM System [47], emphasizing the need for attention to sectors critical to airport operations, particularly in the context of digital cyber warfare [51], as discussed.

Cybersecurity Awareness. This category investigates the effectiveness of cybersecurity awareness training within airport environments. While numerous studies have highlighted the significance of cybersecurity awareness training [1], [5], [8]. However, only one recent publication by Sabillon et al. (2023) [52] has thoroughly examined this topic. We categorize these publications under the following risks: (1) Social Engineering, (2) Insider Threats, and (3) Data Breach, as these risks often arise from human [1], [5], [8], [52] - [54].

Risk and Threat Analysis. This research category conducts literature reviews to identify risks and threats affecting airports. Studies provide insights into threats and recommend improvements for threat detection and response [41], [43], [49], [55] - [64]. Numerous works study airport cybersecurity incidents [55] - [60], including threat actor typologies [58], [61] and associated risks and threats in relation to ICAO (International Civil Aviation Organization) standards [55], which encompass the entire spectrum of airport security risks.

Standards and Regulations. Publications in this category review existing cybersecurity standards and regulations pertinent to the aviation sectors [63] - [66]. They study challenges and gaps in airport cybersecurity policies posed by rapid technological development and call for international cooperation and standardized policies, which currently remain insufficient [64].

Cybersecurity Framework. This category investigates frameworks tailored for airports, focusing on models that systematically manage risks and enhance resilience. While many studies agree the necessity of these frameworks [1], [5], [8], [41], [55], [67], [68] for example, adopting the National Institute of Standards and Technology's (NIST) Cybersecurity Framework to comply with ICAO standards [55], only few recent publications [55], [67], [68] provide actionable details. Nevertheless these frameworks often lack comprehensive insights into adversary behavior, which are essential for identifying and responding to threats throughout an attack's lifecycle. Further details will be provided in Section 6.

Case Study and Survey. This category focuses on research examining the cybersecurity posture of specific airports. Publications may present case studies based on geography [60], [62], [63], or specific events [69], [70] like the COVID-19 pandemic [70] to gather insights on cybersecurity practices and challenges.

5. Airport Security Risks

This section provides detailed exploration of cybersecurity risks associated with attack vectors (Section 2.3) and those identified in our literature review (Section 4).

5.1 Public-facing Accesses

BYOD. The practice of Bring-Your-Own-Device (BYOD) commonly raises concern due to the exposure of organizations to vulnerabilities [71], [72]. However, in airport settings, passengers commonly use their personal devices. The diversity of connected devices in this context complicates device management [73], heightening the security risk when these devices connect to airport networks.

Public Access Services such as Wi-Fi access, check-in kiosks, and charging stations enhance the passenger experience [6], but they also increase risks by leaving users vulnerable to cyberattacks [74] - [76], particularly if proper network segmentation is not implemented.

Man-in-the-Middle (MITM) attacks are prevalent on public Wi-Fi, where cybercriminals eavesdrop using network snooping and sniffing tools to steal sensitive information [75], [77]–[79]. Despite this, many critical websites continue to serve content over unencrypted connections [79] - [81].

[1]NIST Cybersecurity Framework: https://www.nist.gov/cyberframework

Table 2. *Airport Cybersecurity Publications by Category and Associated Security Risks.*

Publication Category	List of Publications	Number of Publications	Associated Security Risks (when applicable)
Critical Infrastructure	[41] - [47]	9	Insecure Network Architecture Malware and Ransomware Data Breach DoS
IoT, Smart Devices and AI Technology	[1], [5], [8], [48] - [50]	6	Public-facing Access Insecure Network Architecture Internet-facing Applications and Services Malware and Ransomware Data Breach Supply Chain and Third Party DoS
Supply Chain and Third Party	[47]	1	Supply Chain and Third Party
Cybersecurity Awareness	[52]	1	Social Engineering Insider Threats Data Breach
Risk and Threat Analysis	[41], [43], [49], [55] - [64]	13	Public-facing Access Insecure Network Architecture Internet-facing Applications and Services Social Engineering Malware and Ransomware Data Breach Supply Chain and Third Party Insider Threat DoS
Standard and Regulation	[63] - [66]	4	(inapplicable)
Cybersecurity Framework	[55], [67], [68]	3	(inapplicable)
Case Study and Survey	[60], [62], [63], [69], [70]	5	(inapplicable)
	Total	**31**	Public-facing Access Insecure Network Architecture Internet-facing Applications and Services Social Engineering Malware and Ransomware Data Breach Supply Chain and Third Party Insider Threat DoS

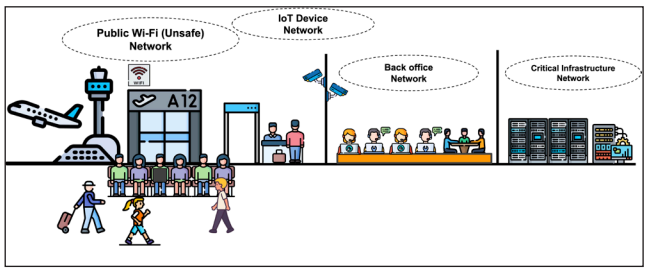


Figure 1. *Basic Network Segmentation for Airport Security.*

Malware from Untrusted Devices. In this attack, bad actors aim to inject malicious payload onto Wi-Fi users' devices [82], [83]. Adversaries may target vulnerabilities on popular devices, e.g., iOS [84]. Public charging stations also pose risks, known as "juice jacking", where attackers use these stations to spread malware and extract data from smartphones [85].

Malicious Hotspots. A malicious hotspot, also known as a "rogue access point" poses a significant threat to public Wi-Fi users. The attacker sets up a wireless access point with an identical SSID to deceive users, making users vulnerable to MITM or other network attacks [86] - [88].

5.2 Insecure Network Architecture

An insecure network architecture may allow attackers to gain access to internal systems and move laterally across organization assets. Effective network segmentation is a key component, enabling administrators to manage network interactions more securely by implementing security policies, with varying levels of security and trust assigned to different applications [89]. Figure 1 illustrates a basic example of network segmentation applicable to an airport, where each segment

requires specific measures and is separated based on the different entities and stakeholders involved, which can be described as follow:

Public Wi-Fi. To prevent potential malicious activities spreading to other airport entities such as CCTV systems [1] and malware incidents at Vienna Airport [90], it should be completely segregated from other airport networks.

IoT Devices. IoT devices often rely on vendor or third-party-based solutions, making them vulnerable to third-party security risks (Section 5.7). Consequently, it is recommended to isolate them from other networks, particularly critical networks.

Back Office is responsible for the administration, operations and logistics of the airport. Given the human-centric nature of these operations, this network is prone to risks such as phishing, social engineering, and other human errors. This network should be kept separated for added security.

Critical Infrastructures includes crucial assets for the airport operations. Access to this network should be restricted from the public network, robust authentication and encryption measures must be implemented, as highlighted in several studies discussed in Section 4.

5.3 Internet-facing Applications and Services

Security Vulnerabilities. Adversaries often exploit weaknesses in internet-facing applications such as airport websites, and mobile applications. These vulnerabilities can arise from software bugs, design flaws, or unpatched vulnerabilities, as discussed in Section 2.3 and Section 4.

Weak authentication practices in internet-facing applications pose significant risks for airports, potentially leading to unauthorized access to critical systems. Additionally, remote access for employees can further complicate security; if authentication credentials are weak, attackers may gain broader access to internal networks [91].

5.4 Social Engineering

Phishing. Social engineering attacks exploit human vulnerabilities, with phishing being a significant concern. In 2013, over 75 U.S. airports reported incidents involving phishing emails designed to deceive users into disclosing financial information [61]. Some of these attacks predominantly target employees with privileged access to critical systems [8].

5.5 Malware and Ransomware

Ransomware incidents often lead to airport operational disruptions and passenger experience [18], [19], [21], [33] - [35]. Additionally, malware attacks may lead to data breaches, exposing sensitive information such as passenger records, payment details, and employee credentials [31], [32]. Such breaches jeopardize privacy and can incur financial costs for airports, including remediations and regulatory fines.

5.6 Data Breach

Data breaches often results in the unauthorized access, disclosure, or theft of sensitive information [19], [20], [33], [34], [36]. Additionally, breaches of sensitive operational information can undermine airport operations and lead to security vulnerabilities [20]. In some cases, attackers may exfiltrate data and demand ransom for its return or for the decryption of compromised systems [32].

5.7 Supply Chain and Third-Party

Security Vulnerabilities in systems can allow attackers to gain unauthorized access. These weaknesses may stem from known or unknown third-party software and hardware bugs, and misconfigurations [92]. Zero-day vulnerabilities pose particular risks, as attacks can occur before developers issue patches. Concerns about IoT and vendor solution vulnerabilities are amplified by the potential for threat actors to compromise not only the affected device but also other network assets [93].

No Security Update Mechanism. Many solutions, particularly IoT devices, may lack a security update mechanism, leaving them vulnerable even after patches have been released [94].

No Common Standards and Specifications. The lack of universally accepted standards for IoT device development leads to inconsistent implementations and design choices, which negatively affect security. Users must manage multiple technologies to effectively support these devices [5].

Supply Chain Compromise involves manipulating products

before they reaches consumer, creating vulnerabilities in critical systems. For example, compromised chips or drivers in smart devices at airports can expose systems to attack [95]. High-profile incidents, such as the SolarWinds hack, affected over 18,000 networks globally [96]. Additionally, a recent incident in Lebanon further illustrates the dangers, where devices were reportedly manipulated for digital warfare [51].

No Physical Hardening. IoT devices are often deployed in various locations throughout the airport, making them vulnerable to tampering during unattended operations. Physical access can result in theft and unauthorized access to internal circuits and overwriting changes [1].

5.8 Insider Threat

An insider threat is a security risk posed by individuals who misuse their access or privileged accounts. A malicious insider, often referred to as a "Turncloak," intentionally abuses legitimate access to steal sensitive information or manipulate critical aviation systems. Mitigating this threat involves adhering to information security management standards and guidelines [97].

5.9 Denial-of-Service

As outlined in Section 2, DoS attacks on airport websites are significant threats to the aviation sector, with recent treads showing demands ransom payments to stop these attacks, aided by the anonymity of cryptocurrencies [32].

6. Airport Security Risks and MITRE ATT&CK Matrix

This section provides a comprehensive analysis of the security risks faced by airports, categorizing these risks in alignment with the MITRE ATT&CK Matrix for Enterprise (or MITRE Matrix) [98]. This widely adopted cybersecurity knowledge base outlines the tactics, techniques and procedures (TTPs) utilized globally for threat analysis and security defenses. Notably, this paper is the first to propose applying the MITRE Matrix to bolster the cybersecurity posture of airports. We correlate all identified airport security risks—derived from cybersecurity incidents and a systematic literature review—with MITRE techniques to mitigate cyberattacks arising from these identified risks.

6.1. MITRE Matrix: TTPs

The MITRE Matrix categorizes attacker tactics and techniques. Each tactic represents a high-level goal, while the techniques describe the specific methods employed to achieve that goal, both indexed for easy reference. Each technique includes (1) procedures based on real-world incidents, (2) mitigations with actionable defense recommendations such as configurations and tools, and (3) detection strategies and recommendations for identifying the attacks. We believe that this comprehensive information enables airport security personnel to effectively implement strategies to mitigate identified risks.

6.2 Airport Security Risks with the MITRE Matrix

The MITRE Matrix is a valuable tool for identifying and mapping airport security risks related to potential attacks. Given the complexity of vulnerabilities, some risks may align with multiple MITRE techniques. This paper focuses on two key tactics: Initial Access (TA001) and Impact (TA0040). Initial Access is fundamental as it represents the first step for adversaries to gain entry into protected systems. The Impact tactic, particularly T1498 (Network Denial of Service), is emphasized due to its prevalence due to its frequency in recent incidents discussed in Section 2.3.

Table 3 provides an overview of categorized airport security risks and their associated MITRE techniques, listing all ten Initial Access techniques and one Impact technique (retrieved September 2024). Each technique is identified and referenced by an ID (e.g., T1189, T1190).

6.3 MITRE Initial Access Techniques (TA0001)

Initial Access is a critical phase in the cyber kill chain, representing the methods adversaries use to gain entry into target systems. Below are the relevant techniques from the MITRE Matrix associated with Initial Access and the corresponding airport security risks:

Public-facing Access (T1659, T1190, T1200): Techniques such as Content Injection (T1659) allow attackers to insert malicious content into network traffic, often through public Wi-Fi. Exploiting vulnerabilities in public-facing applications (T1190),

[2] MITRE Matrix Initial Access Tactic: https://attack.mitre.org/tactics/TA0001/

[3] MITRE Matrix Impact Tactic: https://attack.mitre.org/tactics/TA0040/

Table 3. *Summary of Airport Security Risks Linked to MITRE Matrix Tactics and Techniques.*

Airport Security Risk	Initial Access: TA0001										Impact: TA0040
	T1659	T1189	T1190	T1133	T1200	T1566	T1091	T1195	T1199	T1078	T1498
1. Public-facing Accesses	•		•		•						
2. Insecure Network Architecture	•		•	•							
3. Internet-facing Applications and Services			•	•							
4. Social Engineering Attacks	•	•				•	•			•	
5. Malware and Ransomware	•	•	•	•	•	•	•	•	•	•	
6. Data Breach	•	•	•	•	•	•	•	•	•	•	
7. Supply Chain and Third Party				•	•			•	•		
8. Insider Threats							•		•	•	
9. DoS											•

The sign ● indicates that the airport security risk shown can be categorized according to the specific MITRE Matrix technique.

Associated MITRE Initial Access Tactic (TA0001)

ID	Technique
T1659	Content Injection
T1189	Drive-by Compromise
T1190	Exploit Public-Facing Application
T1133	External Remote Services
T1200	Hardware Additions
T1566	Phishing
T1091	Replication Through Removable Media
T1195	Supply Chain Compromised
T1199	Trusted Relationship
T1078	Valid Accounts

Associated MITRE Impact Tactic (TA0040)

ID	Technique
T1498	Network Denial of Service

such as kiosks and charging stations, can provide unauthorized access due to unpatched vulnerabilities or misconfigurations. This may be coupled with Hardware Additions (T1200), where attackers exploit exposed ports to introduce unauthorized devices [99].

Insecure Network Architecture (T1659, T1190, T1133): Techniques such as Content Injection (T1659) and Exploiting Public-facing Applications (T1190) become more dangerous in poorly secured environments, allowing attackers to gain initial access and subsequently move laterally. Additionally, External Remote Services (T1133) may compromise internal network by enabling unauthorized access through insecure remote connections.

Internet-facing Applications and Services (T1190, T1133): The presence of internet-facing applications and services exposes airports to significant cybersecurity risks via techniques such as Exploiting Public-facing Applications (T1190) and External Remote Services (T1133). Attackers can target vulnerabilities within publicly accessible systems—like online booking websites and service APIs—to gain unauthorized access to sensitive assets. Insecure remote services can also create entry points for attackers, allowing them to gain unauthorized access to internal systems through airport VPNs [100].

Social Engineering (T1659, T1189, T1566, T1091, T1078): Techniques such as Content Injection (T1659), Drive-by Compromise (T1189), and Phishing (T1566) are utilized to manipulate victims. The use of insecure removable media (T1091) may allow untrusted devices to introduce malware to critical systems [101]. Technique like Valid Accounts (T1078) may enable attackers to gain access via stolen account.

Malware, Ransomware and Data Breach (T1659, T1189, T1190, T1133, T1200, T1566, T1091, T1195, T1199, T1078): Malware, ransomware, and data breaches exploit various techniques within airport systems. Initial Access tactics, such as Content Injection (T1659) and Exploiting

Public-facing Applications (T1190) enable attackers to infiltrate via public interfaces. Techniques like Drive-by Compromise (T1189), Phishing (T1566), and the use of insecure removable media (T1091) increase the risk of malware and ransomware, ultimately leading to data breaches. External Remote Services (T1133) and Valid Accounts (T1078) allow attackers to leverage stolen credentials to penetrate into the networks, facilitating ransomware and data theft. Risks from Supply Chain Compromise (T1195) and Trusted Relationship (T1199) may introduce vulnerabilities into overall security.

Supply Chain and Third Party (T1195, T1199): With multiple party involved, techniques such as Supply Chain Compromise (T1195) and Trusted Relationship (T1199) presents significant threat to airports, allowing attackers to exploit vulnerabilities without raising immediate suspicion.

Insider Threat (T1091, T1199, T1078): Techniques such as Insecure Removable Media (T1091) can enable employees to introduce malware into the system. Exploitation of Trusted Relationships (T1199) may allow insiders to manipulate external connections, leading to unauthorized sharing of sensitive information. Additionally, Valid Accounts (T1078) could allow insiders to misuse their credentials.

6.4 MITRE Impact (TA0040) for Denial-of-Service

Denial-of-Service (T1498): According to the MITRE framework, DoS attacks fall under the Impact tactic, specifically technique ID T1498. This can be executed through methods such as direct network floods (T1498.001) or reflection amplification (T1498.002).

7. Modern Defenses for Airport Security

In this section, we outline modern security defense strategies and best practices specifically designed for airport. The key focus areas include the adoption of Cybersecurity Frameworks and the implementation of Zero Trust Architecture, along with additional defenses addressing the identified risks in previous sections.

7.1 Cybersecurity Frameworks and Requirements

Numerous studies emphasize the important of adopting cybersecurity frameworks to enhance airport security [1], [5], [8], [41], [55], [67], [68]. Frameworks, such as NIST Cybersecurity Framework, can help identify weaknesses and facilitate development of security objectives. The Civil Air Navigation Services Organization (CANSO) has proposed guidelines to elevate security levels through these frameworks [102], and several airports, including Airports of Thailand, have already implemented such policies [103].

Recently, the TSA issued new cybersecurity requirements, mandating all U.S. airports and aircraft operators to develop cybersecurity policies [104], [105]. Additionally, ICAO has created Standards and Recommended Practices [106], [107] that urge airports to implement cybersecurity risk management frameworks and collaborate to advance ICAO's cybersecurity framework.

7.2 Zero Trust Architecture

Zero Trust is a modern cybersecurity framework that prioritizes verifying and protecting all entities based on the principle of least privilege. It involves capturing and analyzing logs for effective threat response, acknowledging that internal threats may stem from untrusted devices and personnel. The following discussion highlights the benefits of implementing Zero Trust architecture in airport security.

Visibility for Subsystems. The initial phase of an integrated Zero Trust architecture involves identifying organizational assets, their value, stakeholders, and connectivity. This foundational step prioritizes Zero Trust security policies, including secure access, the principle of least privilege, and enhanced visibility into subsystems.

Network Segmentation. Network segmentation is a foundational element of the Zero Trust architecture, allowing network administrators to enforce the principle of least privilege. For instance, the IoT network is isolated from other segments to prevent unauthorized parties from exploiting IoT vulnerabilities and mitigate the risk of lateral movement within the airport's infrastructure.

Demilitarized Zone. Internet-based services are prime targets for cyberattacks, making it essential to configure a separate network segment called a "Demilitarized Zone" (DMZ). The DMZ acts as a controlled gateway between the internal network and the internet, enforcing strict connectivity rules through firewalls and packet filtering.

The External Policy Enforcement Point (PEP) in Figure 2 filters malicious internet traffic, while the Internal PEP manages the traffic between DMZ servers and the internal network. This layered security approach ensures that external services remain isolated from internal systems.

7.3 Security Awareness Training. The aviation sector may soon face regulatory mandates requiring security awareness training for employees [4], [7]. To effectively implement such programs, organizations actively engage in the development process, providing continuous feedback to enhance the training effectiveness.

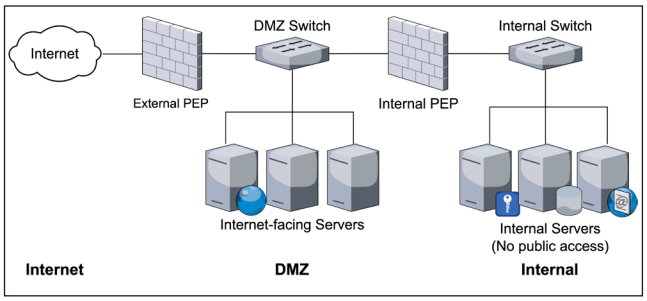


Figure 2. *DMZ subnet that separates an enterprise internal network from other untrusted networks e.g., the Internet.*

7.4 Malware, Ransomware and Data Breach

In addition to previously discussed security measures, airports should implement targeted strategies to detect malware and ransomware. The following mitigation strategies can strengthen an airport's cybersecurity posture:

Endpoint Detection and Response (EDR). To bolster cybersecurity, airports should implement advanced anomaly detection and monitoring to swiftly identify unusual patterns indicative of ransomware. EDR solutions provide real-time analysis of host activities, enabling rapid detection of malicious behaviors [108].

Data Backup and Disaster Recovery. It is imperative to implement comprehensive IT disaster recovery plans that encompass regular off-site data backups. These backups must be secured against common threats, such as ransomware that seeks to compromise backup files (T1486) [109]. Routine testing of backup restoration procedures is crucial to ensure their usability in the event of an incident.

Data Breach Prevention involves proactive monitoring of data for irregular patterns—such as unexpected changes in data size, unusual timestamps, or unauthorized access attempts—is essential for early detection of potential breaches. Strong authentication and encryption for data access further safeguard sensitive information from unauthorized users and ensure compliance with relevant regulatory data security and privacy requirements.

7.5 Supply Chain and Third Party Risk Management

Due Diligence in Supply Chain Management. Any third-party solutions integrated into airport systems must be treated as part of the airport's threat landscape. A rigorous selection process should assess adaptability, security features, secure update mechanisms, and support systems to effectively manage security liabilities and reduce the risk of unforeseen incidents.

Physical Access Restrictions and Tamper Proofing. IoT devices deployed throughout airports are susceptible to physical access attacks, where criminals may steal them for unauthorized entry. To mitigate this risk, various tamper-proof techniques can be applied [110], [111]. For instance, devices can be housed in secure or tamper-resistant cases and disabling or factory resetting.

7.6 Insider Threat Mitigation

Mitigating insider threats is challenging, as they often bypass traditional security measures. Prevention relies on the principle of least privilege, restricting user access to essential functions, along with monitoring for anomalous behaviors. Implementing a Zero Trust Architecture strengthens security by requiring identity verification for every access request and ensuring all access is logged and analyzed.

7.7 Denial-of-Service Mitigation

Cloud or Hybrid DoS Scrubbing Platforms enhance security by redirecting traffic through specialized infrastructure

that filters out malicious traffic before it reaches the airport's network [112], [113]. Incorporating redundancy and failover mechanisms is also essential, as it improves resilience and minimizes downtime, ensuring essential services remain operational during an attack.

7.8 Collaborative Threat Intelligence Sharing

Numerous studies underscore the importance of information sharing within the industry [5], [61], [64], [114]. A real-world example demonstrates the effectiveness: during a spear phishing campaign targeting airport executives, emails containing malware were detected. Through collaborative efforts, the attack was neutralized across the sector [115]. By adopting a multi-faceted approach, including collaborative threat intelligence sharing, airports can enhance their defenses against evolving cyber threats.

8. Challenges and Future Research Directions

While a range of defenses have been detailed, significant challenges remain that must be addressed in order to further enhance airport cybersecurity

Evolving Threat Landscape. Cyber threats are continuously evolving and becoming more sophisticated, and diversified. This necessitates ongoing research and airport adaptability to counter new attack vectors.

Resource Constraints. Smaller airports often face significant resource limitations e.g., budget and personnel, hindering the implementation of cybersecurity measures.

Integration of Legacy Systems. Integrating modern security measures with outdated systems presents significant challenges and often requires substantial investment.

Future Research Directions. Our literature review reveals a pressing need for further research in key areas: (1) Supply Chain and Third-Party Risks, (2) Cybersecurity Awareness, and (3) Development of Airport-Specific Cybersecurity Frameworks. The limited publications in these domains highlight the unique challenges airports face.

Additionally, future research should investigate the integration of advanced technologies like Machine Learning, AI, and Generative AI, focusing on their effectiveness in enhancing airport operations while mitigating potential vulnerabilities. While existing studies have started to address these gaps, ongoing research is essential to keep up with emerging threats and solutions.

9. Conclusion

This paper explores the critical area of airport cybersecurity, highlighting the seriousness of emerging threats in this domain. Through insights gained from recent real-world incidents and a systematic literature review, we conducted a comprehensive analysis and categorized major cybersecurity risks confronting airports, aligned with the MITRE ATT&CK Matrix, providing a valuable framework for exploring practical defenses and best practices articulated in the MITRE knowledge base.

In conclusion, we advocate for the adoption of modern security policies, including robust Cybersecurity Frameworks and Zero Trust Architecture, alongside critical security measures. This study aims to enhance the aviation industry's understanding of the current threat landscape and provide a foundation for enhancing cybersecurity defense and resilience.

10. References

[1] G. Lykou, A. Anagnostopoulou, and D. Gritzalis. "Smart Airport Cybersecurity: Threat Mitigation and Cyber Resilience Controls." *Sensors,* Vol. 19, No. 1, 2019.

[2] EUROCONTROL, *Aviation under Attack from a Wave of Cybercrime.* Available Online at https://www.eurocontrol.int/publication/eurocontrol-think-paper-12-aviation-under-attack-wave-cyber crime, accessed on 1 February 2024.

[3] Business Insurance, *US to add cybersecurityrequirements for critical aviation systems.* Available Online at https://www.businessinsurance.com/article/20221012/NEWS06/912353045/US-to-add-cybersecurity-requirements-for-critical-aviation-systems, accessed on 1 February 2024.

[4] IATA, Compilation of Cyber Security Regulations, *Standards, and Guidance Applicable to Civil Aviation Edition 3.0.* Available Online at https://www.iata.org/contentassets/4c51b00fb25e4b60b38376a935e278b/compilation-of-cyber-regulations-standards-and-guidance3.0.pdf, accessed on 1 February 2024.

[5] N. Koroniotis, N. Moustafa, F. Schiliro, P. Gauravaram, and H. Janicke. "A Holistic Review of Cybersecurity and Reliability Perspectives in Smart Airports." *IEEE Access*, Vol. 8, pp. 209802-209834, 2020.

[6] A. Fattah, H. Lock, W. Buller, and S. Kirby. *Smart Airports: Transforming Passenger Experience to Thrive in the New Economy.* Available Online at https://www.cisco.com/c/dam/en_us/about/ac79/docs/pov/Passenger_Exp_POV_0720aFINAL.pdf, accessed on 1 February 2024.

[7] TSA, *20 years after 9/11: The state of the transportation security administration.* Available Online at https://shorturl.at/1lsGo, accessed on 1 February 2024.

[8] G. Lykou, A. Anagnostopoulou, and D. Gritzalis. "Implementing Cyber Security Measures in Airports to Improve Cyber Resilience." *Proceedings of the 2018 Global Internet of Things Summit,* pp. 1-6, 2018.

[9] J. H. Tan and T. Masood. "Adoption of Industry 4.0 Technologies in Airports - A Systematic Literature Review." *ArXiv,* pp. 1-25, 2021.

[10] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and E. S. Gonzalez. "Understanding the Adoption of Industry 4.0 Technologies in Improving Environmental Sustainability." *Sustainable Operations and Computers,* Vol. 3, pp. 203 - 217, 2022.

[11] K. Gopalakrishnan, M. Govindarasu, D. Jacobson, and B. M. Phares. "Cyber Security for Airports." *International Journal for Traffic and Transport Engineering (IJTTE),* Vol. 3, No. 4, pp. 365-376, 2013.

[12] Bloomberg, *Colonial Pipeline Cyber Attack: Hackers Used Compromised Password.* Available Online at https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password, accessed on 1 February 2024.

[13] CNN, *Ransomware attack hits New Jersey county.* Available Online at https://www.cnn.com/2022/05/26/politics/new-jersey-somerset-county-ransomware-attack, accessed on 1August 2023.

[14] Threatpost, *N.J.'s Largest Hospital System Pays Up in Ransomware Attack.* Available Online at https://threatpost.com/ransomware-attack-new-jersey, accessed on 1 February 2024.

[15] Mandiant, *Advanced Persistent Threats (APTs) -- Threat Actors & Groups.* Available Online at https://www.mandiant. com/resources/insights/apt-groupsm, accessed on 1 June 2023.

[16] ZDNET, *Russian state hackers behind San Francisco Airport Hack.* Available Online at https://www.zdnet.com/article/russian-state-hackers-behind-san-francisco-airport-hack/, accessed on 1 February 2024.

[17] Security Affairs, *A Cyber Attack Hit The Beirut International Airport.* Available Online at https://securityaffairs.com/157079/hacking/cyber-attack-hit-beirut-international-airport.html, accessed on 1 February 2024.

[18] Homeland Security Today, *Long Beach Airport's Website Taken Down By Cyber Attack.* Available Online at https://www.hstoday.us/subject-matter-areas/transportation/long-beach-airports-website-taken-down-by-cyber-attack/, accessed on 1 June 2023.

[19] The Record, *Major Mexican airport confirms experts are working to address cyberattack.* Available Online at https://therecord.media/queretaro-international-airport-mexico-cyberattack, accessed on 1 February 2024.

[20] NTV, *KAA confirms data breach, says no sensitive data leaked.* Available Online at https://ntvkenya.co.ke/news/kaa-confirms-data-breach-says-no-sensitive-data-leaked/, accessed on 1 May 2023.

[21] Airport Technology, *Ransomware attack on Swissport*

causes delay at Zurich Airport. Available Online at https://www.airport-technology.com/news/ransomware-attack-swissport-zurich-airport/, accessed on 1 February 2024.

[22] Security Affairs, *A Cyber Attack Hit The Beirut International Airport.* Available Online at https://securityaffairs.com/157079/hacking/cyber-attack-hit-beirut-international-airport.html, accessed on 1 February 2024.

[23] The Cyber Express, *DDoS Cyberattack Hits Cairo International Airport: Anonymous Collective Claims Responsibility.* Available Online at https://thecyberexpress.com/cairo-international-airport-cyberattack, accessed on 1 February 2024.

[24] Czech Police, *Interior Ministry, Airport Websites Come Under Cyber Attack.* Available Online at https://brnodaily.com/2023/10/24/news/czech-police-interior-ministry-airport-websites-come-under-cyber-attack/, accessed on 1 February 2024.

[25] The Record, *Canada blames border checkpoint outages on cyberattack.* Available Online at https://therecord.media/canada-border-checkpoint-outages-ddos-attack-russia, accessed on 1 May 2023.

[26] Tech Monitor, *Charles de Gaulle Airport website offline after suspected '#OpFrance' DDoS cyberattack.* Available Online at https://techmonitor.ai/technology/cybersecurity/opfrance-cyberattack-charles-de-gaulle-airport, accessed on 1 June 2023.

[27] Mirror, *UK airports targeted by coordinated Russia cyberattack groups.* Available Online at https://www.mirror.co.uk/travel/news/uk-airports-targeted-coordinated-russia-30504938, accessed on 1 February 2024.

[28] Information Week, *The DDoS Attack on German Airport Websites and What IT Leaders Can Learn.* Available Online at https://shorturl.at/7ACXL, accessed on 1 February 2024.

[29] The Associated Press, *Denial-of-service attacks knock US airport websites offline.* Available Online at https://apnews.com/article/technology-business-atlanta-680cf93f7eb0300127448c35299ad66e, accessed on 1 February 2024.

[30] Euractiv, *Italy target of major Russia-linked cyberattack, again.* Available Online at https://shorturl.at/Kvn9C, accessed on 1 February 2024.

[31] CyberMaterial, *Cyberattack Hit Copenhagen Airport.* Available Online at https://cybermaterial.com/cyberattack-hit-copenhagen-airport/, accessed on 1 February 2024.

[32] The Wall Street Journal, *Why Hackers Use Bitcoin and Why It Is So Difficult to Trace.* Available Online at https://www.wsj.com/articles/why-hackers-use-bitcoin-and-why-it-is-so-difficult-to-trace-11594931595, accessed on 1 February 2024.

[33] SecurityWeek, *Data Stolen in Ransomware Attack That Hit Seattle Airport.* Available Online at https://www.securityweek.com/data-stolen-in-ransomware-attack-that-hit-seattle-airport, accessed on 1 February 2024.

[34] Halcyon Tech, *Monti Ransomware Attack on Aéroport de Pau.* Available Online at https://ransomwareattacks.halcyon.ai/attacks/monti-ransomware-attack-on-aeroport-de-pau, accessed on 1 February 2024.

[35] BARRON'S, *Cyberattack Hits Croatia's Split Airport.* Available Online at https://www.barrons.com/news/cyberattack-hits-croatia-s-split-airport-dac3d776, accessed on 1 February 2024.

[36] HACKREAD, *Hackers Leak 2.5M Private Plane Owners' Data Linked to LA Intl. Airport Breach.* Available Online at https://hackread.com/hackers-leak-private-plane-owners-data-la-airport-breach/, accessed on 1 August 2024.

[37] AP, *Hacked Brazil Airport Screens Show Porn to Travelers.* Available Online at https://apnews.com/article/entertainment-caribbean-brazil-c0842e915c403c41830 6433cdfc406a6, accessed on 1 February 2024.

[38] KTSM.com, *FBI warns cyber criminals are spoofing airport websites and Wi-Fi.* Available Online at https://shorturl.at/vJRE7, accessed on 1 February 2024.

[39] The New York Times, *Stranded in the CrowdStrike Meltdown: 'No Hotel, No Food, No Assistance'.* Available Online at https://www.nytimes.com/2024/09/13/travel/crowdstrike-outage-delta-airlines.html, accessed on 1 February 2024.

[40] BCC, *Scam warning as fake emails and websites target users after outage. BBC.* Available Online at https://www.bbc.com/news/articles/cq5xy12pynyo, accessed on 1 February 2024.

[41] G. Dave, G. Choudhary, V. Sihag, I. You, and K.-K. R. Choo. "Cyber Security Challenges in Aviation Communication, Navigation, and Surveillance." *Computers & Security,* Vol. 112, 2022.

[42] E. Andreev and D. Dimitrov. "Analysis of Cyber Vulnerabilities in Civil Aviation and Recommendations for Their Mitigation." *Aeronautical Research and Development,* Vol. 1, pp. 90-99, 2022.

[43] A. Elmarady and K. Rahouma. "Studying Cybersecurity in Civil Aviation, including Developing and Applying Aviation Cybersecurity Risk Assessment." *IEEE Access,* Vol. 4, 2016.

[44] M. L. Salgado and M. S. de Sousa. "Cybersecurity in Aviation: The STPASEC Method Applied to the TCAS Security." *2021 10th Latin-American Symposium on Dependable Computing (LADC),* Florianópolis, Brazil, pp. 1-10, 2021.

[45] S. Khandker, H. Turtiainen, A. Costin, and T. Hämäläinen. "Cybersecurity Attacks on Software Logic and Error Handling within ADS-B Implementations: Systematic Testing of Resilience and Countermeasures." *IEEE Transactions on Aerospace and Electronic Systems,* Vol. 58, No. 4, pp. 2702-2719, 2022.

[46] A. A. Alsulami and S. Zein-Sabatto. "Resilient Cyber-security Approach for Aviation Cyber-physical Systems Protection against Sensor Spoofing Attacks." *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC),* NV, USA, pp. 565-571, 2021.

[47] J. Haan. "Specific Air Traffic Management Cybersecurity Challenges: Architecture and Supply Chain." *ICSEW'20: Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops,* New York, NY, USA, pp. 245-249, 2020.

[48] C. Aranzazu-Suescun, L. F. Zapata-Rivera, O. G.-M. Saenz, and J. M. Christensen. "Securing IoT Surveillance Airport Infrastructure." *Proceedings of the 2024 International Conference on Smart Applications, Communications and Networking,* Harrisonburg, VA, USA, pp. 1-7, 2024.

[49] E. Pik. "Airport Security: The Impact of AI on Safety, Efficiency, and the Passenger Experience." *Journal of Transportation Security,* Vol. 17, No. 1, December, 2024.

[50] D. Shevchuk and I. Steniakin. "A Holistic Approach to Ensuring Safety and Cybersecurity in the Use of Intelligent Technologies in Air Transport." *Electronics and Control Systems,* Vol. 1, No. 75, pp. 97-101, 2023.

[51] The Wall Street Journal, *Pager Attacks in Lebanon 'Weaponize' Supply Chains.* Available Online at https://www.wsj.com/articles/pager-attacks-in-lebanon-weaponize-supply-chains-60722390, accessed on 1 February 2024.

[52] R. Sabillon and J.R.B. Higuera. "The Importance of Cybersecurity Awareness Training in the Aviation Industry for Early Detection of Cyberthreats and Vulnerabilities." *International Conference on Human-Computer Interaction,* pp. 461-479, 2023.

[53] S. Chockalingam, E. Nystad, and C. Esnoul. "Capability Maturity Models for Targeted Cyber Security Training." *Proceedings of the International Conference on Human-Computer Interaction,* pp. 576-590, 2023.

[54] The Hacker News, *Why Human Error is #1 Cyber Security Threat to Businesses in 2021.* Available Online at https://thehackernews.com/2021/02/why-human-error-is-1-cyber-security.html, accessed on 1 February 2024.

[55] P. Stastny and A. M. Stoica. "Protecting Aviation Safety Against Cybersecurity Threats." *IOP Conference Series: Materials Science and Engineering,* Vol. 1226, No. 1, pp. 12-25, 2022.

[56] H. Saada, R. Orizio, and S. Sebastio. "Modeling and Conducting Security Risk Assessment of Smart Airport Infrastructures with SECRAM." *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security,* No. 59, pp. 1-7, 2024.

[57] T. Jeeradist. "Flight Delays and Cancellations Due to Airport Technology Network Disruptions Worldwide." *KBU Journal of Aviation Management: KBUJAM,* Vol. 2, No. 1, pp. 51-60, 2024.

[58] L. Florido-Benítez. "The Types of Hackers and Cyberattacks in the Aviation Industry." *Journal of Transportation Security,* Vol. 17, No. 13, 2024.

[59] H. Su and W. Pan. "Using Digital Twins to Integrate Cyber Security with Physical Security at Smart Airports." *Interdisciplinary Journal of Engineering and Environmental Sciences,* Vol. 10, No. 1, pp. 38-45, January-March, 2023.

[60] S. Samuri, M.F.A. Khir, Z.M. Amin, and M. F. N. Mohammad. "Cybersecurity Maturity Framework for International Airports in Malaysia: A Systematic Literature Review (SLR)." *Journal of Information and Knowledge Management (JIKM),* Vol. 2, pp. 156-167, 2023.

[61] E. Ukwandu, M. Ben-Farah, H. Hindy, M. Bures, R. Atkinson, C. Tachtatzis, I. Andonovid, and X. Bellekens. "Cybersecurity Challenges in Aviation Industry: A Review of Current and Future Trends." *Information,* Vol. 13, No. 3, 2022.

[62] L.F. Benítez. "Identifying Cyber Security Risks in Spanish Airports." *Cyber Security: A Peer-Reviewed Journal,* Vol. 4, No. 3, pp. 267-291, 2021.

[63] M. Karpiuk and M. Kelemen. "Cybersecurity in Civil Aviation in Poland and Slovakia." *Cybersecurity and Law,* Vol. 8, No. 2, pp. 70-83, 2022.

[64] C. Nobles, D. Burrell, and T. Waller. "The Need for a Global Aviation Cybersecurity Defense Policy" *Land Forces Academy Review,* Vol. 27, No. 1, pp. 19-26, March 2022.

[65] V. Filinovych and Z. Hu. "Aviation and the Cybersecurity Threats." *Proceedings of the International Conference on Business, Accounting, Management, Banking, Economic Security and Legal Regulation Research (BAMBEL 2021),* pp. 120-126, 2021.

[66] M. Klenka. "Aviation Cyber Security: Legal Aspects of Cyber Threats." *Journal of Transportation Security,* Vol. 14, No. 3, pp. 177-195, December, 2021.

[67] S. Adhikari and S. Mirchandani. "Integrating Risk Assessment Modeling with Aviation Cybersecurity Framework." *AIAA AVIATION 2020 FORUM,* pp. 29-32, 2020.

[68] S. Adhikari. "An Analysis of AIAA Aviation Cybersecurity Framework in Relation to NIST, COBIT and DHS Frameworks." *AIAA AVIATION 2020 FORUM,* pp. 2930, 2020.

[69] B. Kotkova. "Information Systems and Technologies for the Safe Operation of Airports." *Proceedings of the 26th International Conference on Circuits, Systems, Communications and Computers,* IEEE, pp. 161-166, 2022.

[70] R. A. Ramadan, B. W. Aboshosha, J. S. Alshudukhi, A. J. Alzahrani, A. El-Sayed, and M. M. Dessouky. "Cybersecurity and Countermeasures at the Time of Pandemic." *Journal of Advanced Transportation,* Vol. 2021, No. 1, 2021.

[71] Trend Micro, *The case for making BYOD safe-security news.* Available Online at https://www.trendmicro.com/vinfo/us/security/news/internet-

of-things/the-case-for-making-byod-safe, accessed on 1 February 2024.

[72] C. Brook, *The Ultimate Guide to BYOD Security: Definition & More.* Available Online at https://digitalguardian.com/blog/ultimate-guide-byod-security-overcoming-challenges-creating-effective-policies-and-mitigating, accessed on 1 June 2023

[73] A. Bridgwater, *How mobile device management is taking on the BYOD challenge.* Available Online at https://www.theregister.com/2014/11/08/mobile working/, accessed on 1 June 2023.

[74] Y. Joshi, D. Das, and S. Saha. "Mitigating Man in the Middle Attack Over Secure Sockets Layer." *Proceedings of the 2009 IEEE International Conference on Internet Multimedia Services Architecture and Applications (IMSAA),* pp. 1-5, 2009.

[75] M. Marlinspike, *New Tricks for Defeating SSL in Practice. Black Hat USA,* Available Online at https://www.blackhat.com/presentations/bh-dc-09/Marlinspike/BlackHat-DC-09-Marlinspike-Defeating-SSL.pdf, accessed on 1 June 2023.

[76] Norton, *Public Wi-Fi: An Ultimate Guide on the Risks + How to Stay Safe.* Available Online at https://us.norton.com/blog/privacy/public-wifi, accessed on 1 February 2024.

[77] S. Englehardt, D. Reisman, C. Eubank, et al. "Cookies that Give You Away: The Surveillance Implications of Web Tracking." *Proceedings of the 24th International Conference on World Wide Web,* Florence, Italy, pp. 289-299, 2015.

[78] X. Zheng, J. Jiang, J. Liang, et al. "Cookies Lack Integrity: Real-World Implications." *Proceedings of the 24th USENIX Security Symposium,* Washington, D.C, pp. 707-721, 2015.

[79] S. Sivakorn, I. Polakis, and A. D. Keromytis. "The Cracked Cookie Jar: HTTPS Cookie Hijacking and the Exposure of Private Information." *Proceedings of the 2016 IEEE Symposium on Security and Privacy,* pp. 724-742, 2016.

[80] S. Sivakorn, A. D. Keromytis, and J. Polakis. "That's the Way the Cookie Crumbles: Evaluating HTTPS Enforcing Mechanisms." *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society,* 2016.

[81] S. Sivakorn, P. Sirawongphatsara, and N. Rujiratanapat. "Web Encryption Analysis of Internet Banking Websites in Thailand." *Proceedings of the 17th International Joint Conference on Computer Science and Software Engineering,* pp. 139-144, 2020.

[82] M. Kacic, P. Hanacek, M. Henzl, and P. Jurnecka. "Malware Injection in Wireless Networks." *Proceedings of the 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems,* Vol. 01, pp. 483-487. 2013.

[83] A. Zimba, Z. Wang, and M. Mulenga. "Cryptojacking Injection: A Paradigm Shift to Cryptocurrency-based Web-centric Internet Attacks." *Journal of Organizational Computing and Electronic Commerce,* Vol. 29, pp. 40-59, 2019.

[84] J. Spaulding, A. Krauss, and A. Srinivasan. "Exploring an Open WiFi Detection Vulnerability as a Malware Attack Vector on iOS Devices." *Proceedings of the 7th International Conference on Malicious and Unwanted Software,* pp. 87-93, 2012.

[85] Krebson Security, *Why is 'Juice Jacking' Suddenly Back in the News?.* Available Online at https://krebsonsecurity.com/2023/04/why-is-juice-jacking-suddenly-back-in-the-news/, accessed on 1 February 2024.

[86] Kaspersky, *What is an Evil Twin Attack? Evil Twin Wi-Fi Explained.* Available Online at https://www.kaspersky.com/resource-center/preemptive-safety/evil-twin-attacks, accessed on 1 February 2024.

[87] V. Roth, W. Polak, E. G. Rieffel, and T. Turner. "Simple and Effective Defense Against Evil Twin Access Points." *Wireless Network Security,* pp. 220-235, 2008.

[88] H. Gonzales, K. Bauer, J. Lindqvist, D. McCoy, and D. Sicker. "Practical Defenses for Evil Twin

Attacks in 802.11." *Proceedings of the 2010 IEEE Global Telecommunications Conference GLOBECOM 2010,* pp. 1-6, 2010.

[89] Palo Alto Networks, *What Is Network Segmentation?*. Available Online at https://shorturl.at/rCIDK, accessed on 1 June 2023.

[90] The Local Austria, *Turkish suspect identified in Vienna airport cyberattack.* Available Online at https://www.thelocal.at/20170228/suspect-identified-in-vienna, accessed on 1 June 2023

[91] H. Abbas, N. Emmanuel, M. F. Amjad, et al. "Security Assessment and Evaluation of VPNs: A Comprehensive Survey." *ACM Computing Surveys,* Vol. 55, No. 13s, pp.1-47, 2023.

[92] CISA, *2021 Top Routinely Exploited Vulnerabilities.* Available Online at https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-117a, accessed on 1 June 2023.

[93] TechTarget, *The Mirai IoT Botnet holds strong in 2020.* Available Online at https://shorturl.at/GMUA8, accessed on 1 June 2023.

[94] Keyfactor, *Top 10 IoT Vulnerabilities in Your Devices.* Available Online at https://www.keyfactor.com/blog/top-10-iot-vulnerabilities, accessed on 1 June 2023.

[95] Bloomberg, *China Used a Tiny Chip in a Hack That Infiltrated U.S. Companies*. Available Online at https://shorturl.at/tMHdW, accessed on 1 February 2024.

[96] WIRED, *Hacker Lexicon: What Is a Supply Chain Attack?*. Available Online at https://www.wired.com/story/hacker-lexicon-what-is-a-supply-chain-attack/, accessed on 1 June 2023.

[97] M. Theoharidou, S. Kokolakis, M. Karyda, and E. A. Kiountouzis. "The insider threat to information systems and the effectiveness of ISO17799." *Computers & Security,* Vol. 24, pp. 472-484, 2005.

[98] MITRE. *MITRE ATT&ACK Matrix for Enterprise.* Available Online at https://attack.mitre.org/, accessed on 1 September 2024.

[99] SecureList, *DarkVishnya: Banks attacked through direct connection to local network.* Available Online at https://securelist.com/darkvishnya/89169/, accessed on 1 February 2024.

[100] SecureWorks, *Gold Sahara.* Available Online at https://www.secureworks.com/research/threat-profiles/gold-sahara, accessed on 1 February 2024.

[101] Google, *Turla: A Galaxy of Opportunity.* Available Online at https://cloud.google.com/blog/topics/threat-intelligence/turla-galaxy-opportunity, accessed on 1 February 2024.

[102] CANSO, *CANSO Standard of Excellence in Cybersecurity.* Available Online at https://canso.org/publication/canso-standard-of-excellence-in-cybersecurity/, accessed on 1 February 2024.

[103] Airports of Thailand (AOT), *AOT ICT Security Policy, AOT Cyber Security Policy and AOT Personal Data Protection Policy.* Available Online at https://corporate.airportthai.co.th/th/cybersecurity-th/, accessed on 1 February 2024.

[104] TSA, *TSA issues new cybersecurity requirements for airport and aircraft operators.* Available Online at https://shorturl.at/eXlvZ, accessed on 1 June 2023.

[105] T. Szuba. *Safeguarding Your Technology: Practical Guidelines for Electronic Education Information Security.* National Center for Education Statistics, 1998.

[106] ICAO, *Annex 17 - Aviation Security, ICAO - International Standards and Recommended Practices.* Available Online at https://shorturl.at/JQHkr, accessed on 1 February 2024.

[107] ICAO, *AVIATION CYBERSECURITY. (2022),* Available Online at https://www.icao.int/aviation cybersecurity/Pages/default.aspx, accessed on 1 February 2024.

[108] S.-J. Lee, H.Y. Shim, Y.R. Lee, T.R. Park, S.H. Park, and I.G. Lee. "Study on Systematic Ransomware Detection Techniques." *Proceedings of the 24th International Conference on Advanced Communication Technology (ICACT),* pp. 297-301. 2022

[109] MITRE, *MITRE ATT&CK Matrix - Data Encrypted for Impact.* Available Online at https://attack.mitre.org/techniques/T1486/, accessed on 1 February 2024.

[110] J. Katz. "Universally Composable Multi-party Computation using Tamper-proof Hardware." *Advances in Cryptology – EUROCRYPT 2007,* pp. 115-128, 2007.

[111] R. Gennaro, A. Lysyanskaya, T. Malkin, S. Micali, and T. Rabin. "Algorithmic Tamper-proof (ATP) Security: Theoretical Foundations for Security Against Hardware Tampering." *Theory of Cryptography Conference TCC 2004,* 2004.

[112] MITRE, *MITRE ATT&CK Matrix - Network Denial of Service.* Available Online at https://attack.mitre.org/techniques/T1498/, accessed on 1 February 2024.

[113] WIRED, *GitHub Survived the Biggest DDoS Attack Ever Recorded.* Available Online at https://www.wired.com/story/github-ddos-memcached/, accessed on 1 February 2024.

[114] I. A. Shah, N. Jhanjhi, and S. Brohi. "Cybersecurity Issues and Challenges in Civil Aviation Security." *Cybersecurity in the Transportation Industry,* pp. 1-23, 2024.

[115] D. S. Turetsky, B. H. Nussbaum, and U. Tatar. *Success Stories in Cybersecurity Information Sharing.* The College of Emergency Preparedness, Homeland Security and Cybersecurity University at Albany, 2020.

Information Technology Journal

KMUTNB

คณะเทคโนโลยีสารสนเทศและนวัตกรรมดิจิทัล
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
1518 ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพฯ 10800
Tel: 02-555-2726
Website: http://www2.it.kmutnb.ac.th/journal