

การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน*

Graduation Forecasting using Support Vector Machine

พรณิภา บุตรเอก**

สุรเดช บุญลือ***

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน ชุดข้อมูลกลุ่มตัวอย่างที่นำมาใช้ทดสอบเป็นข้อมูลของนักศึกษาระดับปริญญาตรีหลักสูตร 4 ปี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยนอร์ทกรุงเทพ ระหว่างปีการศึกษา 2547-2551 จำนวนทั้งสิ้น 138 ระเบียบ 18 คุณลักษณะ ชุดข้อมูลที่ได้จะนำส่งเข้าตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา ซึ่งจะใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) มาทำการสร้างตัวแบบพยากรณ์และเปรียบเทียบเพื่อคัดเลือกตัวแบบพยากรณ์ที่มีประสิทธิภาพสูงสุด โดยการทดสอบประสิทธิภาพตัวแบบจะใช้วิธีไขว้พบ 15 ส่วน ผลการทดสอบ พบว่า ตัวแบบการพยากรณ์ที่ใช้เทคนิค SVM-PK ให้ความแม่นยำในการพยากรณ์สูงที่สุดที่ร้อยละ 89.13 โดยมีผลการประเมินค่าความแม่นยำตรงกับ 0.878 ค่าความระลึกลับเท่ากับ 0.891 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.860 นอกจากนี้ยังให้ค่า RMSE น้อยที่สุดเท่ากับ 0.3297 โดยที่ตัวแบบพยากรณ์ซัพพอร์ตเวกเตอร์แมชชีนที่ใช้เคอร์เนลฟังก์ชันแบบ SVM-NP SVM-RBF และ SVM-PUK ให้ค่าความแม่นยำในการพยากรณ์เท่ากันที่ร้อยละ 87.68 ส่วนตัวแบบพยากรณ์ C4.5 และ BP-ANN ให้ค่าความแม่นยำร้อยละ 83.33 และ 84.36 ตามลำดับ สรุปได้ว่าตัวแบบพยากรณ์ SVM-PK เป็นตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

คำสำคัญ: การพยากรณ์โอกาสสำเร็จการศึกษา ซัพพอร์ตเวกเตอร์แมชชีน

*เพื่อศึกษาการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน ประเมินและเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาระหว่างเทคนิคซัพพอร์ตเวกเตอร์แมชชีน เทคนิคต้นไม้ตัดสินใจ และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ เพื่อหาตัวแบบพยากรณ์ที่เหมาะสมที่สุดในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

**สาขาวิชาเทคโนโลยีสารสนเทศ บัณฑิตวิทยาลัย มหาวิทยาลัยนอร์ทกรุงเทพ E-mail: khoonjahao@gmail.com

***ผู้ช่วยศาสตราจารย์ ประจำสาขาวิชาวิทยาการคอมพิวเตอร์ บัณฑิตวิทยาลัย มหาวิทยาลัยนอร์ทกรุงเทพ

Abstract

The purpose of this paper is to present Support Vector Machine techniques for graduation forecasting. This dataset is composed of 4-year undergraduate in computer science course, Faculty of Information Technology, North Bangkok University, during an academic year 2547-2551, a total of 138 records, 18 features. The dataset will be transferred to a forecast model of graduation by using Support Vector Machine (SVM), Decision Tree (C4.5), and Multi-Layer Perceptron (BP-ANN) techniques. These models are chosen once with high performance of forecast. In performance tests the model will use 15 Fold Cross-Validation. A result of comparison forecast model using the SVM-PK techniques had the highest accuracy in forecasting: 89.13 percent by the Precision equaled 0.878, Recall equaled and 0.891 and F-Measure equaled 0.860 and RMSE got the smallest value equaled 0.3297. The models for SVM-NP, SVM-RBF, and SVM-PUK had an accuracy in forecast equaled 87.68 percent, for C4.5 and BP-ANN forecast had an accuracy of 83.33 percent and 84.36, respectively. Concluded that the forecasting model using SVM-PK is most appropriate for graduation forecasting of graduating students.

Keyword: Graduation Forecast, Support Vector Machine

บทนำ

การศึกษาในระดับอุดมศึกษาโดยเฉพาะปริญญาตรีเป็นการศึกษาที่มีต้นทุนสูง ไม่ว่าจะเป็นต้นทุนทางด้านเวลาหรือค่าใช้จ่ายที่ขึ้นอยู่กับสาขาวิชาที่เลือกเรียน นอกจากนี้ยังรวมถึงความตั้งใจ ความหวังของนักศึกษาและครอบครัวด้วย ดังนั้นหากนักศึกษาไม่มีความพร้อมและไม่มีความเข้าใจในสาขาวิชาที่ตนเองเลือกเรียนก็อาจจะทำให้นักศึกษาไม่สำเร็จการศึกษาตามระยะเวลาที่กำหนดได้ หากเป็นเช่นนั้นแล้วจะทำให้ต้นทุนทั้งหมดถูกใช้ไปอย่างเปล่าประโยชน์ นอกจากนี้อาจทำให้เกิดปัญหาต่างๆ ตามมา ไม่ว่าจะเป็นปัญหาทางด้านจิตใจ ปัญหาทางด้านครอบครัวหรือแม้กระทั่งปัญหาทางด้านสังคม ซึ่งปัญหาเหล่านี้จะลดลงหรือหมดไปหากนักศึกษาเลือกเรียนสาขาได้ตรงกับความสามารถของตนเอง หรือได้รับคำปรึกษาจากอาจารย์ที่ปรึกษาอย่างทันทั่วถึง

สถาบันการศึกษาทั้งภาครัฐและเอกชนส่วนใหญ่มีการจัดเก็บข้อมูลต่างๆ ของนักศึกษาไว้เป็นจำนวนมาก ข้อมูลเหล่านี้จะถูกนำมาใช้ประโยชน์เฉพาะตอนที่นักศึกษากำลังศึกษาอยู่เท่านั้น แต่เมื่อจบการศึกษาไปแล้วข้อมูลเหล่านี้กลับไม่ได้ถูกนำมาใช้ให้เกิดประโยชน์อีกเลย ดังนั้นนักวิจัยจึงนำข้อมูลเหล่านี้มาทำการศึกษาเกี่ยวกับการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาที่กำลังศึกษาอยู่ โดยมีการนำเทคนิคเหมืองข้อมูล (Data Mining) เข้ามาใช้ ซึ่งจะนำข้อมูลของนักศึกษาที่ถูกจัดเก็บไว้เป็นจำนวนมากมาสร้างตัวแบบพยากรณ์และเมื่อได้ตัวแบบที่น่าเชื่อถือแล้วก็สามารถนำข้อมูลของนักศึกษาที่เข้ามาใหม่มาทำนายได้ว่านักศึกษามีโอกาสสำเร็จการศึกษาในสาขาที่ตนเองเลือกเรียนได้หรือไม่ จากการทบทวนงานวิจัยพบว่ามีนักวิจัยได้นำเทคนิคเหมืองข้อมูลมาใช้ในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา ดังเช่นงานวิจัยของ ทิพย์สุดา วงศ์คำดี (2549) ที่ได้นำเสนอเกี่ยวกับการเปรียบเทียบแบบจำลองการพยากรณ์ผลสำเร็จการศึกษาของนักศึกษา โดยใช้วิธีการ

วิเคราะห์จำแนกประเภทเทคนิคโครงข่ายประสาทเทียมและเทคนิคโครงข่ายประสาทเทียมที่มีการคัดเลือกปัจจัยโดยใช้วิธีการทางสถิติ ผลการวิจัยพบว่า โครงข่ายประสาทเทียมแบบแพร่กระจายย้อนกลับที่มีการคัดเลือกปัจจัยโดยใช้วิธีการทางสถิติ มีประสิทธิภาพการพยากรณ์ดีกว่าโครงข่ายประสาทเทียมที่มีนำเข้าปัจจัยทั้งหมดและการวิเคราะห์จำแนกประเภท ด้วยค่าเฉลี่ยความถูกต้องของการจำแนกกลุ่มร้อยละ 94.83 93.54 และร้อยละ 81.5 ตามลำดับ นอกจากนี้ บุชรา ลัมพิพัฒนางกูร (2549) ได้นำเสนอการประยุกต์ใช้วิธีการพยากรณ์โอกาสในการสำเร็จการศึกษาโดยใช้เทคนิคจีเนติกอัลกอริทึมเพื่อเลือกลักษณะที่เหมาะสมก่อนนำไปใช้ในการพยากรณ์ด้วยโครงข่ายประสาท (GA/ANN) เปรียบเทียบกับโครงข่ายประสาทแบบทั่วไป (ANN) จากการทดสอบพบว่า GA/ANN สามารถลดจำนวนคุณลักษณะที่ใช้ในโมเดลได้อย่างชัดเจน โดย SGA/ANN สามารถลดจำนวนคุณลักษณะลงร้อยละ 42 และ MGA/ANN สามารถลดจำนวนคุณลักษณะลงร้อยละ 45 และเทคนิคที่แนะนำนั้นยังให้ประสิทธิภาพการทำนายที่ใกล้เคียงกับการใช้โครงข่ายประสาทเพียงอย่างเดียว เมื่อทำการเปรียบเทียบระหว่างเทคนิค ANN กับ SGA/ANN และเทคนิค ANN กับ MGA/ANN โดยใช้ t-Test Paired Two Sample for Means มีค่า P-Value เท่ากับ 0.64 และ 0.98 ตามลำดับ และเมตตา คงคากุล (2550) ได้นำเสนอเทคนิคการพยากรณ์ผลสัมฤทธิ์ทางการศึกษาของนักศึกษาจากพฤติกรรมการใช้อินเทอร์เน็ตโดยวิธีโครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer Perceptron) โดยใช้เทคนิควิธีการสอนด้วยอัลกอริทึมของ Levenberg-Marquardt มาประยุกต์ใช้ในการพยากรณ์ ผลจากการศึกษาพบว่าโครงข่ายประสาทเทียมให้ค่าประสิทธิภาพการพยากรณ์เท่ากับร้อยละ 58 จากงานวิจัยที่กล่าวมาจะเห็นได้ว่านักวิจัยส่วนใหญ่เลือกใช้เทคนิคโครงข่ายประสาทเทียมในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา แต่จากการทบทวนงานวิจัยที่หลากหลายพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความแม่นยำในการพยากรณ์มากกว่าเทคนิคโครงข่ายประสาทเทียม เช่น งานวิจัยของ สุรเดช บุญลือ และคณะ (2553) ได้นำเสนอเกี่ยวกับการทำนายโรคพาร์กินสันโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน โดยทำการเปรียบเทียบระหว่างเทคนิคซัพพอร์ตเวกเตอร์แมชชีน โครงข่ายประสาทเทียมแพร่กลับ และเทคนิคต้นไม้ตัดสินใจในการทำนายโรคพาร์กินสัน ผลจากการศึกษาพบว่า ตัวแบบทำนายที่ใช้ซัพพอร์ตเวกเตอร์แมชชีน มีความแม่นยำมากที่สุดร้อยละ 92.19 ขณะที่ตัวแบบทำนายที่ใช้โครงข่ายประสาทเทียมแพร่กลับ และเทคนิคต้นไม้ตัดสินใจมีความแม่นยำร้อยละ 90.10 และ 88.02 ตามลำดับ นอกจากนี้ ภรณ์ยา อามฤตรัตน์ และคณะ (2552) ได้นำเสนอการเปรียบเทียบประสิทธิภาพการคัดเลือกและจำแนกข้อมูลด้วยวิธีการทางโครงข่ายประสาทเทียม โดยเปรียบเทียบวิธีการจำแนกข้อมูลแบบโครงข่ายประสาทเทียมแบบ Multi-Layer Perceptron และ ซัพพอร์ตเวกเตอร์แมชชีน

ผลจากการศึกษา พบว่า การใช้แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนในการจำแนกข้อมูลจะมีประสิทธิภาพดีกว่าการใช้แบบจำลองโครงข่ายประสาทเทียมแบบ Multi-Layer Perceptron นิเวศ จิระวิจิต และนรินทร์ พนาวาส (2554) ได้นำเสนอวิธีการจำแนกความคิดเห็นโดยใช้เทคนิคการเรียนรู้ของเครื่อง โดยทำการทดสอบประสิทธิภาพการจำแนกด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ เนอ์ฟเบย์ และเคเนียร์เนสเนเบอร์ ผลจากการศึกษา พบว่า การเรียนรู้ด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนให้ประสิทธิภาพในการจำแนกที่ดีที่สุด โดยให้ค่าความถูกต้องร้อยละ 86.30 และ นางเยาว์ ไนอรุณ และพรรณี สิทธิเดช (2555) ได้นำเสนอการใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนและโครงข่ายประสาทเทียมแบบแพร่กลับสำหรับจำแนกผู้ป่วยโรคหัวใจขาดเลือดและโรคหัวใจรูปร่างอื่นออกจากกันให้ชัดเจน เนื่องจากสองโรคนี้มีลักษณะอาการคล้ายกันทำให้แพทย์วินิจฉัยโรคได้ยาก ผลจากการศึกษา พบว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนได้ค่าความถูกต้องร้อยละ

96.46 ส่วนเทคนิคโครงข่ายประสาทเทียมแบบแพร่กลับได้ค่าความถูกต้องร้อยละ 88.21 ซึ่งสรุปได้ว่าเทคนิคซอฟต์แวร์เวกเตอร์แมชชีนเป็นเทคนิคที่เหมาะสมในการจำแนกผู้ป่วยโรคหัวใจขาดเลือดออกจากโรคหัวใจรูปแบบอื่น

งานวิจัยนี้จึงมีแนวคิดที่จะศึกษาการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้ซอฟต์แวร์เวกเตอร์แมชชีน รวมทั้งมีการเปรียบเทียบและคัดเลือกอัลกอริทึมการเรียนรู้ของเครื่องที่จะนำมาประยุกต์ใช้ในการสร้างตัวแบบการพยากรณ์ที่มีประสิทธิภาพสูงที่สุดในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

วิธีการวิจัย

วิธีการวิจัยครั้งนี้แบ่งออกเป็น 3 ขั้นตอน ประกอบด้วย 1) การจัดเตรียมข้อมูล 2) การสร้างตัวแบบพยากรณ์ และ 3) การวัดประสิทธิภาพของตัวแบบพยากรณ์

การจัดเตรียมข้อมูล

ข้อมูลกลุ่มตัวอย่างที่นำมาใช้ทดสอบเป็นข้อมูลระเบียบประวัติของนักศึกษาระดับปริญญาตรีหลักสูตร 4 ปี สาขาวิชาวิทยาการคอมพิวเตอร์ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยนอร์ทกรุงเทพ ระหว่างปีการศึกษา 2547-2551 จำนวนทั้งสิ้น 138 ระเบียบ ประกอบด้วยข้อมูลประวัติส่วนตัวและข้อมูลผลการเรียนของนักศึกษา ซึ่งแบ่งเป็นข้อมูลนำเข้าจำนวน 18 คุณลักษณะ (แสดงดังตารางที่ 1) และข้อมูลนำออก 1 คุณลักษณะ คือ สถานะสำเร็จการศึกษาของนักศึกษา (สำเร็จการศึกษาตามระยะเวลาที่กำหนด/ไม่สำเร็จการศึกษาตามระยะเวลาที่กำหนด) โดยข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลของนักศึกษาระดับชั้นปีที่ 1 เท่านั้น (3 ภาคเรียน)

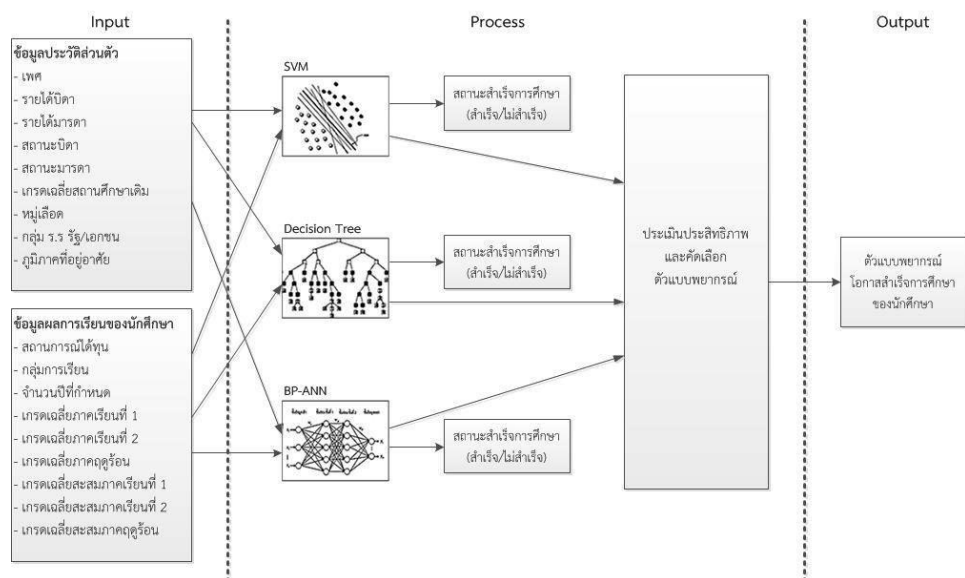
ตารางที่ 1 ข้อมูลนำเข้าที่ใช้ในการพยากรณ์

ข้อมูลประวัติส่วนตัว	ข้อมูลผลการเรียนของนักศึกษา
เพศ	สถานะการได้รับทุน
รายได้บิดา	กลุ่มการเรียน
รายได้มารดา	จำนวนปีที่กำหนด
สถานะบิดา	เกรดเฉลี่ยภาคเรียนที่ 1
สถานะมารดา	เกรดเฉลี่ยภาคเรียนที่ 2
เกรดเฉลี่ยสถานศึกษาเดิม	เกรดเฉลี่ยภาคฤดูร้อน
หมู่เลือด	เกรดเฉลี่ยสะสมภาคเรียนที่ 1
กลุ่ม ร.ร รัฐ/เอกชน	เกรดเฉลี่ยสะสมภาคเรียนที่ 2
ภูมิภาคที่อยู่อาศัย	เกรดเฉลี่ยสะสมภาคฤดูร้อน

หลังจากเก็บรวบรวมข้อมูลของนักศึกษาแต่ละคนได้แล้ว ก็จะนำข้อมูลที่ได้มาทำการตรวจสอบความสมบูรณ์ และทำการแทนค่าข้อมูลเพื่อใช้เป็นคุณลักษณะในการสร้างและทดสอบตัวแบบการพยากรณ์ต่อไป

การสร้างตัวแบบพยากรณ์

เป็นขั้นตอนของการออกแบบและสร้างตัวแบบพยากรณ์ทั้ง 3 ตัวแบบ หลังจากนั้นจะมีการคัดเลือกตัวแบบพยากรณ์ที่มีประสิทธิภาพความแม่นยำสูงสุด ระหว่างตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) โดยใช้คอร์เนลฟังก์ชันแบบ Normalise Polynomial Kernel (SVM-NP) Polynomial Kernel (SVM-PK) Radial Basis Function Kernel (SVM-RBF) และ Pearson VII function-based universal kernel (SVM-PUK) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยใช้อัลกอริทึมแบบ C4.5 และเทคนิคโครงข่ายประสาทเทียมโดยใช้อัลกอริทึมแบบย้อนกลับ (BackPropagation Neuron Network: BP-ANN) เพื่อนำตัวแบบที่ได้ไปใช้ในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาต่อไป ซึ่งข้อมูลที่น่าเข้าฝึกสอนและทดสอบจะมี 18 คุณลักษณะ ส่วนข้อมูลนำออกจะเป็นสถานะสำเร็จการศึกษาของนักศึกษา (สำเร็จการศึกษาตามระยะเวลาที่กำหนด/ไม่สำเร็จการศึกษาตามระยะเวลาที่กำหนด) ดังภาพที่ 1



ภาพที่ 1 กระบวนการสร้าง ประเมิน และคัดเลือกตัวแบบพยากรณ์

จากภาพที่ 1 เป็นกระบวนการสร้าง ประเมิน และคัดเลือกตัวแบบพยากรณ์ โดยในขั้นตอนแรกเป็นการนำข้อมูลของนักศึกษาส่งเข้าไปยังตัวแบบการพยากรณ์ทั้ง 3 ตัวแบบ ได้แก่ ตัวแบบซัพพอร์ตเวกเตอร์แมชชีน ตัวแบบต้นไม้ตัดสินใจ และตัวแบบโครงข่ายประสาทเทียม ซึ่งจะทำการฝึกสอนและทดสอบเพื่อเปรียบเทียบค่าความแม่นยำว่าตัวแบบใดจะสามารถพยากรณ์ได้แม่นยำที่สุด

การวัดค่าประสิทธิภาพของตัวแบบพยากรณ์

การวิจัยครั้งนี้ผู้วิจัยได้ทำการวัดประสิทธิภาพตัวแบบการพยากรณ์โดยใช้เกณฑ์การวัดประสิทธิภาพของตัวแบบรู้จำด้วยวิธี Predictive Modeling (Brame, 2007) ซึ่งประกอบด้วยค่าความแม่นยำตรง (Precision) ค่าความระลึก (Recall) ค่าประสิทธิภาพโดยรวม (F-Measure) และค่าความแม่นยำ (Accuracy) ดังสมการ (1) (2) (3) และ (4) ตามลำดับ

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

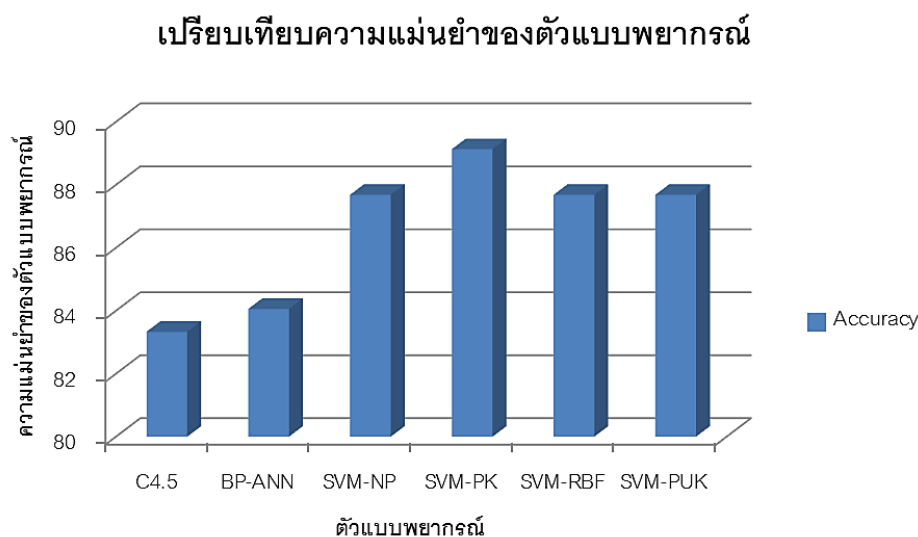
$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F - \text{Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

ผลการวิจัย

จากการนำข้อมูลประวัติของนักศึกษาจำนวน 138 ระเบียน ไปทดสอบกับตัวแบบพยากรณ์ 3 ตัวแบบที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) ด้วยวิธีไขว้ทบ 15 ส่วน (15 Fold Cross-Validation) พบว่า ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนให้ค่าความแม่นยำในการพยากรณ์สูงกว่าตัวแบบที่ใช้เทคนิคต้นไม้ตัดสินใจและเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ โดยที่ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์ที่มีการเลือกเคอร์เนลฟังก์ชันที่ใช้แบบ SVM-PK ให้ค่าความแม่นยำในการพยากรณ์สูงที่สุด มีค่าเท่ากับร้อยละ 89.13 ส่วนเคอร์เนลฟังก์ชันที่เหลือ ได้แก่ SVM-NP SVM-RBF และ SVM-PUK นั้น ให้ค่าความแม่นยำในการพยากรณ์เท่ากันเท่ากับร้อยละ 87.68 อีกทั้งตัวแบบพยากรณ์ C4.5 และ BP-ANN ให้ค่าความแม่นยำร้อยละ 83.33 และ 84.06 ตามลำดับ ดังภาพที่ 2



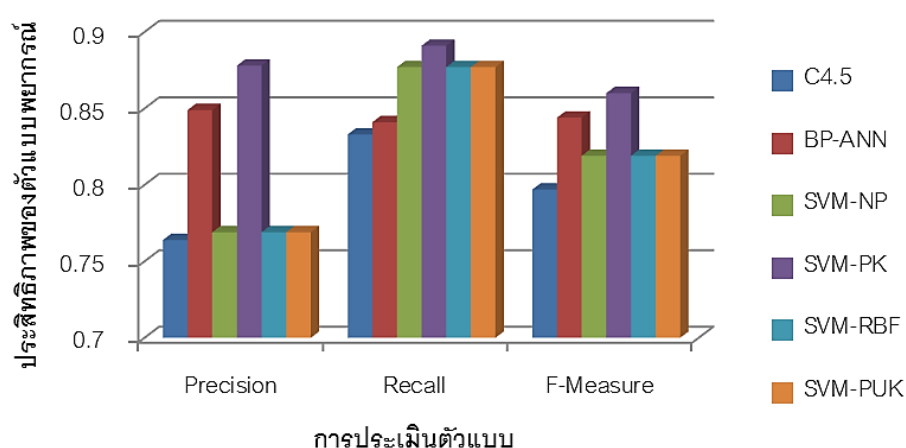
ภาพที่ 2 กราฟเปรียบเทียบค่าความแม่นยำของตัวแบบพยากรณ์ (Accuracy)

ผลการทดสอบประสิทธิภาพของตัวแบบพยากรณ์โดยใช้ค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวม พบว่า ตัวแบบพยากรณ์ SVM-PK ให้ประสิทธิภาพในการพยากรณ์ที่ดีที่สุด ดังภาพที่ 3 โดยให้ค่าความแม่นยำเท่ากับ 0.878 ค่าความระลึกเท่ากับ 0.891 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.860 ตัวแบบพยากรณ์ C4.5 ให้ค่าความแม่นยำเท่ากับ 0.764 ค่าความระลึกเท่ากับ 0.833 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.797 ตัวแบบพยากรณ์ BP-ANN ให้ค่าความแม่นยำเท่ากับ 0.849 ค่าความระลึกเท่ากับ 0.841 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.844 ส่วนตัวแบบพยากรณ์ SVM-NP SVM-RBF และ SVM-PUK ให้ประสิทธิภาพในการพยากรณ์เท่ากัน โดยให้ค่าความแม่นยำเท่ากับ 0.769 ค่าความระลึกเท่ากับ 0.877 และค่าประสิทธิภาพโดยรวมเท่ากับ 0.819 ดังตารางที่ 2

ตารางที่ 2 ผลการเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์

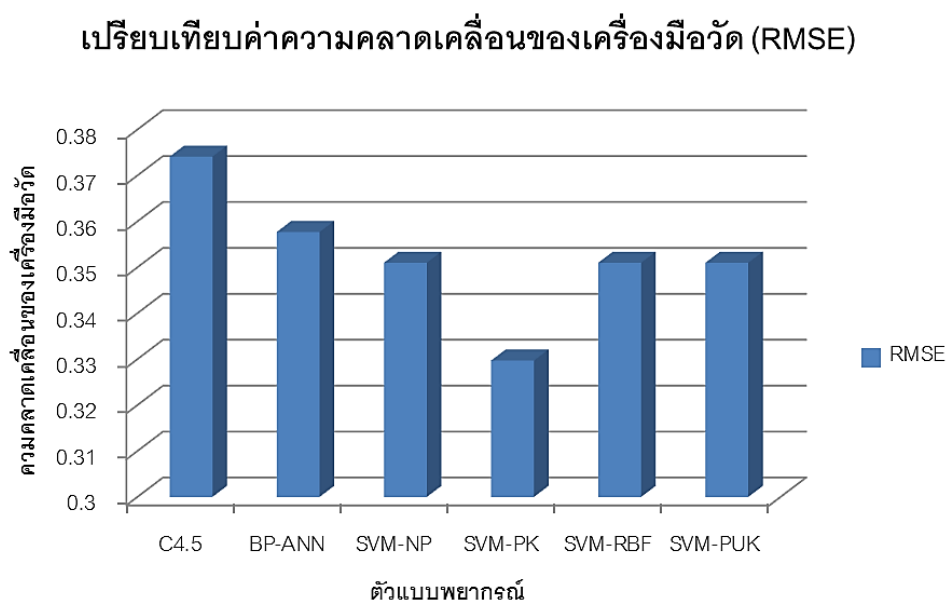
ตัวแบบพยากรณ์	Precision	Recall	F-Measure
C4.5	0.764	0.833	0.797
BP-ANN	0.849	0.841	0.844
SVM-NP	0.769	0.877	0.819
SVM-PK	0.878	0.891	0.860
SVM-RBF	0.769	0.877	0.819
SVM-PUK	0.769	0.877	0.819

เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์



ภาพที่ 3 เปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

นอกจากนี้ เมื่อทำการเปรียบเทียบค่าความคลาดเคลื่อนของเครื่องมือวัด (Root Mean Square Error: RMSE) พบว่า ตัวแบบพยากรณ์ SVM-PK ให้ค่าความคลาดเคลื่อนน้อยที่สุดเท่ากับ 0.3297 ตัวแบบพยากรณ์ BP-ANN เท่ากับ 0.3577 ตัวแบบพยากรณ์ SVM-NP SVM-RBF และ SVM-PUK ให้ค่าความคลาดเคลื่อนเท่ากันเท่ากับ 0.351 และตัวแบบพยากรณ์ C4.5 ให้ค่าความคลาดเคลื่อนสูงที่สุดเท่ากับ 0.3741 ดังภาพที่ 4



ภาพที่ 4 เปรียบเทียบความคลาดเคลื่อนของเครื่องมือวัด (RMSE)

ดังนั้นตัวแบบพยากรณ์ที่เหมาะสมที่สุดในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา คือ ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีน โดยจะใช้เคอร์เนลฟังก์ชันแบบ Polynomial Kernel (SVM-PK) เนื่องจาก SVM-PK นั้นให้ประสิทธิภาพและค่าความแม่นยำในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาดีที่สุด

สรุปและอภิปรายผล

งานวิจัยนี้ได้นำความรู้ทางด้านวิทยาการคอมพิวเตอร์มาประยุกต์ใช้กับการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา โดยมีการนำอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning) ที่ประกอบด้วยเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (SVM) เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (BP-ANN) มาทำการสร้างตัวแบบพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา จากนั้นจึงทำการเปรียบเทียบหาประสิทธิภาพความแม่นยำในการพยากรณ์ เพื่อหาตัวแบบพยากรณ์ที่ให้ความแม่นยำสูงสุดที่เหมาะสมกับการนำไปใช้ในการพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษา

ผลการทดสอบพบว่าตัวแบบการพยากรณ์ที่ใช้เทคนิค SVM-PK ให้ผลการประเมินค่าความแม่นยำ ค่าความระลึก และค่าประสิทธิภาพโดยรวม มากที่สุดเท่ากับ 0.878 0.891 และ 0.860 ตามลำดับ โดยให้ค่าความแม่นยำในการพยากรณ์ที่ร้อยละ 89.13

จากผลการวิจัยที่พบว่า ตัวแบบพยากรณ์ที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีค่าความแม่นยำในการพยากรณ์มากกว่าทั้งเทคนิคต้นไม้ตัดสินใจและเทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ อาจเนื่องจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีนมีพฤติกรรมที่จะแยกแยะข้อมูล โดยใช้สมการระนาบหลายมิติโดยจะพยายามหาจุดข้อมูลที่ให้ได้สมการระนาบหลายมิติที่ใช้แบ่งแยกดีที่สุด (Optimal Hyperplane) มีความถูกต้องที่สุด โดยพิจารณาจากระยะห่าง (Margin) ระหว่างคลาส ซึ่งเส้นระนาบที่ดีที่สุดนี้จะสามารถจำแนกกลุ่มนักศึกษาที่สำเร็จการศึกษาและกลุ่มที่ไม่สำเร็จการศึกษาได้อย่างมีประสิทธิภาพซึ่งสอดคล้องกับผลการวิจัยของ นิเวศ จิระวิจิตชัย และนรินทร์ พนาวาส (2554) ได้ศึกษาการจำแนกความคิดเห็นโดยใช้เทคนิคการเรียนรู้ของเครื่อง และทำเปรียบเทียบประสิทธิภาพการจำแนกด้วยอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีน ต้นไม้ตัดสินใจ เนอฟเบย์ และเคเนียร์สเนเบอร์ ผลการวิจัยพบว่า อัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนให้ประสิทธิภาพในการจำแนกที่ดีที่สุดโดยให้ค่าความถูกต้องเท่ากับร้อยละ 86.30 นอกจากนี้ยังสอดคล้องกับงานวิจัยของ ภรณ์ยา อามฤตรัตน์ และพยุ่ง มีสัจ (2553) ได้ศึกษาเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางโครงข่ายประสาทเทียม พบว่าซัพพอร์ตเวกเตอร์แมชชีนที่ใช้เคอร์เนลฟังก์ชันแบบ Polynomail มีค่าความถูกต้องของการพยากรณ์สูงกว่าเคอร์เนลฟังก์ชันแบบ RBF ที่ร้อยละ 95.94 และ 74.37 ตามลำดับ

เอกสารอ้างอิง

ภาษาไทย

ทิพย์สุดา วงศ์คำดี. (2549). “การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้โครงข่ายประสาทเทียม.”

การศึกษาระดับปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น.

นงเยาว์ ไนอรุณ และพรณี สิทธิเดช (2555). “การใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนและโครงข่ายประสาทเทียมแบบแพร่กลับสำหรับการจำแนกผู้ป่วยโรคหัวใจขาดเลือดและโรคหัวใจรูปแบบอื่น”.

การประชุมวิชาการระดับชาติ “วิทยาศาสตร์วิจัย” ครั้งที่ 4, (มีนาคม): 69-74.

นิเวศ จิระวิจิตชัย และนรินทร์ พนาวาส. (2554). “การจำแนกความคิดเห็นโดยใช้เทคนิคการเรียนรู้ของเครื่อง.”

บัณฑิตวิทยาลัย มหาวิทยาลัยศรีปทุม วิทยาเขตชลบุรี.

บุษรา ลิมพิพัฒนางกูร. (2549). “การพยากรณ์โอกาสสำเร็จการศึกษาของนักศึกษาโดยใช้จูนิตอัลกอริทึม

แบบหลายวัตถุประสงค์ เปรียบเทียบกับวิธีทางระบบโครงข่ายประสาทเทียม: กรณีศึกษานักศึกษา

ภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต

วิทยาการคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ภรณ์ยา อามฤตรัตน์ และพยุ่ง มีสัจ. (2553). “การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่ม ข้อมูลของ

โรคลมร้อน ด้วยวิธี การทางโครงข่ายประสาทเทียม.” **National Conference on Computing and Information Technology ครั้งที่ 6**, (มิถุนายน): 116-121.

ภรณ์ยา อัมฤตรัตน์, เดช ธรรมศิริ, วาทีนี น้อยเพียร, ภัทรารุณี แสงศิริ, ณรงค์ โพธิ์ และพวง มีสัจ. (2552).

“การเปรียบเทียบประสิทธิภาพการคัดเลือกและจำแนกข้อมูลด้วยวิธีการทางเครือข่ายประสาทเทียม.” **National Conference on Computing and Information Technology ครั้งที่ 5**, (พฤษภาคม): 131 - 138.

เมตตา คงคากุล. (2550). “การพยากรณ์ผลสัมฤทธิ์ทางการเรียน จากพฤติกรรมการใช้อินเทอร์เน็ตโดยวิธีโครงข่ายประสาทเทียม กรณีศึกษา: นักศึกษาปริญญาตรี มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึง.” สารนิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ บัณฑิตวิทยาลัย มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

สุรเดช บุญลือ จักรกริช เคล้าละม่อม และแสงนภา วันเพ็ญ. (2553). “การทำนายโรคพาร์กินสันโดยใช้ ซัพพอร์ตเวกเตอร์แมชชีน.” **National Conference on Computer Information Technologies**. (มกราคม): 7-12.

ภาษาต่างประเทศ

Max Brame. “Principles of Data Mining.” (2007). **Springer-Verlag London Limited**, pp 173-176.