

The Development of a Predication Model for Academic Achievement of Students During a Learning Process by Using Data Mining Techniques

Werayut Pimpaporn^{1,*}

¹ Faculty of Science, Kasetsart University Sriracha Campus

Received: 18 December 2017

Revised: 2 January 2018

Accepted: 23 March 2018

ABSTRACT

The purpose of this research was to develop a prediction model for educational achievement of students during the study by using mining techniques. The data were obtained from academic transcripts of graduate students in the Information Technology Program and Computer Science Program. The data included the grade levels of all courses and the Grade Point Average (GPA.). The data were divided into 7 sub-series based on the semesters. The model was developed with the use of four data mining techniques, including the multiple linear regression, simple linear regression, multilayer perceptron, and support vector machine for regression. The results showed that the prediction model of both programs had more accuracy to predict the study's result when the students studied in higher year levels. The RMSE of the study's result when the students graduated were 0.09 to 0.28, and the RAE were at 18.40 - 60.20 levels.

Keywords : Prediction, Educational achievement, Study in program, Data mining techniques

* Corresponding Author; Email: werayut.p@ku.th

การพัฒนาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียน ระหว่างการศึกษในหลักสูตร โดยเทคนิควิธีเหมืองข้อมูล

วีระยุทธ พิมพ์ภรณ์^{1,*}

¹คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาดำเนินใช้เทคนิคเหมืองข้อมูล โดยใช้ข้อมูลประวัติผลการศึกษาของผู้เรียนในระดับปริญญาตรีหลักสูตรเทคโนโลยีสารสนเทศและวิทยาการคอมพิวเตอร์ที่สำเร็จการศึกษาแล้ว ประกอบไปด้วยข้อมูล ผลการเรียนในทุกรายวิชา และเกรดเฉลี่ยที่ผู้เรียนได้เมื่อจบการศึกษา ข้อมูลแบ่งออกเป็นหลักสูตรละ 7 ชุดข้อมูลย่อยตามลำดับภาคการศึกษา แบบจำลองพัฒนาขึ้นด้วยเทคนิควิธีเหมืองข้อมูล 4 เทคนิค ได้แก่ การถดถอยเชิงเส้นพหุ การถดถอยเชิงเส้นอย่างง่าย แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอน หลายชั้น และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ผลการวิจัยพบว่าทั้ง 2 หลักสูตร แบบจำลองสำหรับการพยากรณ์มีความถูกต้องในการพยากรณ์เพิ่มขึ้นเมื่อผู้เรียนศึกษาในหลักสูตรชั้นปีที่สูงขึ้น โดยมีค่าความคลาดเคลื่อน (RMSE) ของผลการเรียนเมื่อสิ้นสุดการศึกษาในระหว่าง 0.09 – 0.28 คิดเป็นร้อยละค่าเฉลี่ยความผิดพลาด (RAE) ที่ระดับ 18.40 – 60.20

คำสำคัญ : การพยากรณ์ ความสำเร็จทางการศึกษา การศึกษาในหลักสูตร เทคนิควิธีเหมืองข้อมูล

* ผู้ประสานงานหลัก; อีเมล: werayut.p@ku.th

ความเป็นมาและความสำคัญของปัญหา

สถาบันอุดมศึกษามีหน้าที่หลักในการให้บริการวิชาการด้านต่างๆ การพัฒนาองค์กรให้ก้าวไปสู่ องค์กรแห่งความรู้ (Knowledge Organization) จึงถือเป็นภารกิจสำคัญประการหนึ่งของสถาบันอุดมศึกษา การนำแนวคิดในการพัฒนากระบวนการจัดการเรียนการสอนที่มุ่งเน้นไปที่ประสิทธิภาพ ถือเป็นการเสริมสร้างความเข้มแข็งด้านการบริหารจัดการการศึกษา ปัจจุบันสภาพแวดล้อมทางด้านเทคโนโลยีการศึกษาได้มีการปรับเปลี่ยนรูปแบบและมีการพัฒนาอย่างต่อเนื่อง เพื่อเป็นการตอบสนองต่อเทคโนโลยีการสื่อสารและโทรคมนาคมที่มีความก้าวหน้าอย่างรวดเร็ว การนำเทคโนโลยีทางการศึกษาที่ทันสมัยมาประยุกต์ใช้กับการบริหารจัดการการเรียนการสอนในสถาบันอุดมศึกษาถือได้ว่าเป็นการสร้างความสำเร็จได้เปรียบทางการแข่งขันของสถาบันอุดมศึกษาได้เป็นอย่างดี

การใช้เทคโนโลยีสารสนเทศมีความสำคัญกับทุกหน่วยงาน โดยเฉพาะอย่างยิ่งงานด้านการศึกษา สถาบันการศึกษาโดยส่วนใหญ่มีการจัดเก็บข้อมูลต่างๆ ของนักศึกษา เช่น ข้อมูลนักศึกษา ข้อมูลประวัติการศึกษา ข้อมูลเกรด เป็นต้น ทั้งนี้เพื่อใช้ในการประมวลผลข้อมูลตามกระบวนการจัดการศึกษา ดังนั้นเมื่อสถาบันอุดมศึกษาต้องการเพิ่มขีดความสามารถในการบริหารจัดการการเรียนการสอนอย่างมีประสิทธิภาพจึงจำเป็นต้องนำเทคโนโลยีสารสนเทศที่ทันสมัยมาประยุกต์ใช้ในหน่วยงาน

งานวิจัยฉบับนี้มุ่งเน้นการพัฒนาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาในหลักสูตรโดยเทคนิควิธีเหมืองข้อมูล เพื่อให้ผู้ที่มีหน้าที่บริหารจัดการหลักสูตรทราบถึงแนวโน้มและคาดคะเนผลที่จะเกิดขึ้นกับผู้เรียนในอนาคต ทั้งนี้เพื่อเป็นการพัฒนากลไกการบริหารจัดการหลักสูตรให้มีประสิทธิภาพมากยิ่งขึ้น โดยแบบจำลองจะถูกพิจารณาจากประวัติการเรียนของผู้เรียนที่สำเร็จการศึกษาใน 2 หลักสูตร ได้แก่ หลักสูตรเทคโนโลยีสารสนเทศ และหลักสูตรวิทยาการคอมพิวเตอร์ ซึ่งมีการจัดการเรียนการสอนในภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา สำหรับงานวิจัยนี้ผู้วิจัยเลือกใช้เทคนิควิธีการสร้างแบบจำลอง ด้วยเทคนิคเหมืองข้อมูล (Data Mining techniques) 4 เทคนิค ได้แก่ การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) โดยผลลัพธ์ของการประมวลผลข้อมูลจะอยู่ในรูปของแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียน

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาในหลักสูตร ด้วยเทคนิคเหมืองข้อมูล
2. เพื่อเปรียบเทียบประสิทธิภาพ และเลือกเทคนิควิธีที่เหมาะสม ในการวิเคราะห์ข้อมูลการพยากรณ์ ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาในหลักสูตร

เอกสารและงานวิจัยที่เกี่ยวข้อง

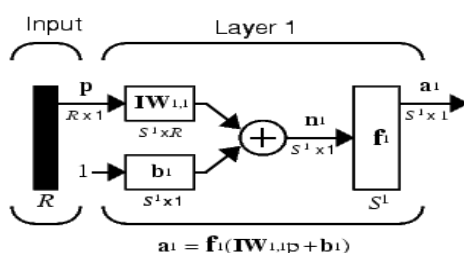
เหมืองข้อมูล (Data Mining) คือกระบวนการที่ใช้หลักการทางคณิตศาสตร์ สถิติ เพื่อทำให้คอมพิวเตอร์สามารถเรียนรู้และจดจำข้อมูลต่างๆ เพื่อใช้ในการค้นหารูปแบบ (Pattern) กฎเกณฑ์ (Rule) ของความสัมพันธ์ใน

ชุดข้อมูล (Data Set) เพื่อสกัดข้อมูลซึ่งอยู่ในรูปของสารสนเทศ (Information) เพื่อใช้ในกระบวนการสร้างองค์ความรู้ และค้นหาองค์ความรู้ที่อยู่ในข้อมูล (Roiger, 2017; Witten, Frank et al., 2016) ทั้งนี้เพื่อนำผลการวิเคราะห์ข้อมูลมาใช้ในกระบวนการตัดสินใจ มุ่งเน้นการหาค่าที่จะเกิดขึ้นในอนาคต (Prospective) เหมือนข้อมูลจึงถือเป็นรูปแบบการวิเคราะห์ข้อมูลเชิงลึก ผลลัพธ์ของกระบวนการจะขึ้นอยู่กับการกำหนดวัตถุประสงค์และผลลัพธ์ที่ต้องการ ในการใช้เหมือนข้อมูลสำหรับวิเคราะห์หาองค์ความรู้จึงจำเป็นต้องใช้วิธีที่หลากหลาย เพื่อให้ได้ผลลัพธ์ในการวิเคราะห์ที่ดีที่สุด โดยขั้นตอนเริ่มต้นของกระทำเหมือนข้อมูลคือ การจัดเก็บข้อมูล และตีความข้อมูล เพื่อค้นหาองค์ความรู้ที่ซ่อนอยู่ อาจอยู่ในลักษณะของรูปแบบ (Pattern) หรือความสัมพันธ์ (Relation) ซึ่งไม่ปรากฏให้เห็นโดยตรงเมื่อพิจารณาจากข้อมูล ดังนั้น การทำเหมือนข้อมูลจึงมีความเกี่ยวข้องกับการสืบค้น (Search) กระบวนการสกัดข้อมูล (Data Extraction) กระบวนการเข้าถึง (Data Access) และกระบวนการสร้างความสัมพันธ์ (Relation) จากชุดข้อมูล โดยมีจุดมุ่งหมายเพื่อนำผลลัพธ์ดังกล่าวไปใช้ประโยชน์

การวิเคราะห์การถดถอย (Linear Regression) เป็นวิธีการวิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ ซึ่งเทคนิคนี้จะอาศัยความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรเพื่อใช้ในการทำนาย โดยลักษณะของสมการ Linear regression (Harrell Jr, 2015) จะแสดงให้เห็นถึงค่าความสัมพันธ์ของตัวแปรตามกับตัวแปรอิสระในลักษณะของข้อมูลเชิงปริมาณ โดยสามารถเขียนได้ดังนี้

$$X\beta = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (1)$$

โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron : MLP) โครงข่ายประสาทเทียมแบบหลายชั้น (Jananii, 2012) ที่มีลักษณะการเชื่อมต่อแบบโยงไปข้างหน้าแบบทั่วถึง (Fully connected feed-forward nets) สามารถมีจำนวนชั้นตั้งแต่หนึ่งชั้นขึ้นไป (Frias-Martinez et al., 2006) ถูกพัฒนาขึ้นมาเพื่อแก้จุดอ่อนของโครงข่ายประสาทเทียมแบบชั้นเดียว (Single layer perceptron) ให้ความสามารถในการคำนวณที่สูงขึ้น โดยโครงข่ายประสาทเทียมแบบหลายชั้นจะประกอบไปด้วย ชั้นข้อมูลเข้า (Input layer) ตามจำนวนแอตริบิวต์ของชุดข้อมูล ชั้นซ่อน (Hidden Layer) ความเหมาะสมของการกำหนดขึ้นอยู่กับการทดสอบประสิทธิภาพ และชั้นผลลัพธ์ (Output Layer)



ภาพที่ 1 โครงสร้างประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multi-Layer Perceptron : MLP)

ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) ซัพพอร์ตเวกเตอร์แมชชีน หรือ SVM เป็นอัลกอริทึมที่ใช้ในการจำแนกกลุ่มข้อมูล ด้วยวิธีการหาระนาบการตัดสินใจ (Decision Hyperplane) หรือไฮเปอร์เพลนที่เหมาะสม สำหรับการแบ่งข้อมูล 2 ส่วนจากกัน SVM ยังถูกนำมาประยุกต์ใช้ในการสร้างสมการในประมาณการค่าฟังก์ชันเชิงเส้น $f(x)$ ที่ใช้แทนระนาบตัดสินใจ ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย หรือ SVR จะเป็นการนำข้อมูลปัจจุบันและข้อมูลในอดีตจำนวนหนึ่งมาทำการเรียนรู้ (Training) เพื่อให้

ทราบถึงรูปแบบสำหรับคาดการณ์ผลซึ่งจะเกิดขึ้นในอนาคต ในงานวิจัยนี้ผู้วิจัยเลือกใช้วิธีสร้างสมการในประมาณการค่าฟังก์ชันเชิงเส้นด้วยวิธี Sequential Minimal Optimization for SVM Regression (SMOreg) (Shevade et al., 2000; Smola & Schölkopf, 2004)

งานวิจัยที่เกี่ยวข้องกับการพัฒนาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาด้วยเทคนิคเหมืองข้อมูล พบว่า โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา (Vilailuck et al., 2015) ได้ใช้ข้อมูลนักเรียนระดับมัธยมศึกษาปีการศึกษา 2548 ถึง 2556 มาพัฒนาคัลลิ่งข้อมูลและสร้างตัวแบบในการพยากรณ์ผลการเรียนรู้ด้วยเทคนิควิธีโครงข่ายประสาทเทียมแบบมัลติเลเยอร์ โดยแบบจำลองมีความถูกต้องในการจำแนกร้อยละ 92.59 จากผลการวิจัยพบว่าปัจจัยที่ส่งผลกระทบต่อผลการเรียนได้แก่ แผนการเรียน ผลการเรียนเฉลี่ยวิทยาศาสตร์ ผลการเรียนเฉลี่ยสังคมศึกษา ผลการเรียนเฉลี่ยภาษาอังกฤษ และผลการเรียนเฉลี่ยระดับมัธยมศึกษาชั้นปีที่ 3 ซึ่งส่งผลกระทบต่อผลการเรียนในระดับมัธยมศึกษาปีที่ 4 โดยผลการวิจัยสามารถนำไปใช้ในการแนะนำด้านการวางแผนการจัดการเรียนการสอนได้

การประยุกต์ใช้เทคนิคเหมืองข้อมูลเพื่อใช้ในการจัดการผู้เรียนในระหว่างการเรียนรู้โดยมีการสร้างแบบจำลองเพื่อทำนายผู้เรียนที่มีโอกาสออกกลางคัน (Dropouts) (Yadav et al., 2012) ซึ่งถือเป็นปัญหาสำคัญที่เกิดขึ้นในระหว่างการจัดการเรียนการสอน โดยการวิจัยได้มีการพัฒนาแบบจำลองการพยากรณ์ด้วยเทคนิควิธีการจำแนกข้อมูล (Classification) โดยสร้างแบบจำลองต้นไม้ตัดสินใจ (Decision Tree Classifier) เพื่อใช้สำหรับพยากรณ์ประสิทธิภาพในการเรียนของผู้เรียน ทั้งนี้เพื่อสามารถระบุผู้เรียนที่มีประสิทธิภาพในการเรียนอยู่ในระดับต่ำและมีโอกาสออกจากการเรียนกลางคัน เพื่อให้อาจารย์ที่ปรึกษาสามารถให้คำแนะนำในการเรียนได้อย่างทันท่วงที

วิธีดำเนินการวิจัย

ศึกษาปัญหาและวิเคราะห์ข้อมูล

ขั้นตอนแรกผู้วิจัยทำการรวบรวมข้อมูลประวัติผลการศึกษาของผู้เรียนในระดับปริญญาตรี ที่ศึกษาในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ และหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ จัดการเรียนการสอนภายในภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์ ศรีราชามหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา โดยชุดข้อมูลที่ใช้ในกระบวนการสร้างแบบจำลองการพยากรณ์ประกอบไปด้วยข้อมูลผลการเรียนของผู้เรียนที่สำเร็จการศึกษาในทุกสาขาวิชา และเกรดเฉลี่ยที่ผู้เรียนได้เมื่อจบการศึกษาโดยค่าดังกล่าวจะถูกนำมาเป็นตัวแปรที่ต้องการพยากรณ์ในการวิจัยนี้

เมื่อพิจารณาจากความคาดหวังของการจัดการศึกษาในปัจจุบันพบว่าปัจจัยหนึ่งที่มีความสำคัญ เป็นตัวแปรที่ชี้วัดความสำเร็จของการบริหารจัดการหลักสูตรคือ อัตราความสำเร็จในการเรียนของผู้เรียนในหลักสูตร กล่าวได้ว่าผู้บริหารจัดการหลักสูตรโดยทั่วไปคาดหวังให้ผู้เรียนสามารถสำเร็จการเรียนในหลักสูตรที่ผู้เรียนศึกษาอยู่และมีผลการเรียนที่ดีหรือระดับเกรดที่สูง จากเหตุผลดังกล่าวพบว่าคำถามที่สำคัญซึ่งเกิดขึ้นระหว่างที่ผู้เรียนอยู่ในระหว่างการศึกษาในหลักสูตรคือ ข้อมูลผลการเรียนของผู้เรียนที่ผ่านมาเมื่อนำข้อมูลดังกล่าวมาคาดคะเนผลการเรียนเมื่อสิ้นสุดการเรียนของผู้เรียน จะมีค่าระดับเกรดอยู่ในระดับใด เช่น มีโอกาสได้เกียรตินิยมหรือไม่ ผ่านเกณฑ์การจบหรือไม่ เป็นต้น

การสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาในหลักสูตร ผู้วิจัยเลือกใช้ข้อมูลของผู้สำเร็จการศึกษาใน 2 หลักสูตร โดยมีรายละเอียดดังตาราง

ตารางที่ 1 จำนวนผู้สำเร็จการศึกษา

หลักสูตร		ปีที่เข้าศึกษา						รวม
		2549	2550	2551	2552	2553	2554	
วิทยาการคอมพิวเตอร์	S05	53	64	57	84	79	66	403
เทคโนโลยีสารสนเทศ	S06	20	35	69	54	40	26	244
วิทยาการคอมพิวเตอร์	S09		54	67	101	73	76	371
เทคโนโลยีสารสนเทศ	S10		25	65	71	43	41	245
รวม		73	178	258	310	235	209	1,263

จากตารางที่ 1 แสดงข้อมูลผู้สำเร็จการศึกษาในหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ ในผู้เรียนภาคปกติ (S05) และนิสิตภาคพิเศษ (S06) หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ในผู้เรียนภาคปกติ (S09) และนิสิตภาคพิเศษ (S10) ทั้งหมด 1,263 คน โดยข้อมูลผลการเรียนในทุกรายวิชาของผู้สำเร็จการศึกษาจะถูกนำมาใช้ในกระบวนการวิเคราะห์เพื่อสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียน

การจัดเตรียมข้อมูล

กระบวนการสร้างแบบจำลองในการวิจัยสิ่งที่เป็นประการหนึ่งคือการคัดเลือกตัวแปรที่มีความสัมพันธ์กับความสำเร็จทางเรียนของผู้เรียน ในการวิจัยนี้ผู้วิจัยทำการสร้างแบบจำลองด้วยเทคนิควิธีเหมืองข้อมูล โดยอาศัยระดับผลสัมฤทธิ์ทางการเรียนในแต่ละรายวิชา ทั้งนี้ลักษณะของผลสัมฤทธิ์ทางการเรียนคือค่าที่วัดได้จาก ความรู้ ความเข้าใจ และความสามารถของผู้เรียนที่เกิดจากการเรียนการสอนซึ่งวัดผลได้จากการทดสอบ โดยการนำผลไปพิจารณาเทียบกับเกณฑ์ที่ได้กำหนดไว้ เพื่อให้สามารถแยกผู้เรียนออกเป็นกลุ่มๆ ได้ตามระดับความรู้ โดยมาตรฐานระดับความรู้ของสถาบันอุดมศึกษาในประเทศไทย ใช้ระดับผลสัมฤทธิ์ทางการเรียนแยกออกเป็น 8 ระดับ คือ A (Excellent), B+ (Very Good), B (Good), C (Average) จนถึงระดับ F โดยข้อมูลดังกล่าวจะถูกนำมาเป็นชุดข้อมูลในการวิเคราะห์ประสิทธิภาพของอัลกอริทึมในการสร้างแบบจำลอง

สำหรับกระบวนการในการจัดเตรียมชุดข้อมูล (Data Set) ที่ใช้ในการสร้างแบบจำลอง ผู้วิจัยมีกระบวนการในการเลือกรายวิชาที่ใช้เป็นตัวแปรในการพยากรณ์โดยแบ่งรายวิชาออกเป็น 8 ชุดข้อมูลย่อย โดยเลือกรายวิชาในแต่ละภาคการศึกษาตามแผนการเรียนของผู้เรียนในแต่ละหลักสูตร ซึ่งประกอบไปด้วย หมวดวิชาศึกษาทั่วไป หมวดวิชาเฉพาะ ทั้งนี้ไม่รวมหมวดวิชาเลือกเสรีซึ่งผู้เรียนไม่ได้ลงทะเบียนเรียนทุกคน เพื่อไม่ให้เกิดข้อมูลที่มีค่าสูญหาย (Missing Value) ในชุดข้อมูล ซึ่งจะส่งผลกระทบต่อประสิทธิภาพของแบบจำลองโดยตรง สำหรับรายวิชาที่ผู้วิจัยเลือกใช้ทั้ง 2 หลักสูตรเป็นรายวิชาที่มีการจัดการเรียนการสอนตามแผนการเรียนของผู้เรียน ตลอด 7 ภาคการศึกษา โดยมีจำนวนรายวิชาดังตาราง

ตารางที่ 2 จำนวนรายวิชาที่ใช้สร้างแบบจำลองการพยากรณ์

หลักสูตร		จำนวนรายวิชา ในหลักสูตร	จำนวนรายวิชา สำหรับการสร้างแบบจำลอง							ร้อยละ จำนวน รายวิชา
			ภาคการศึกษา/ชั้นปีที่เรียน							
			1/1	2/1	1/2	2/2	1/3	2/3	1-2/4	
เทคโนโลยี	S05	135	6	8	8	6	7	8	7	37%
สารสนเทศ	S06	130								38%
วิทยาการ	S09	129	7	7	5	8	6	4	2	29%
คอมพิวเตอร์	S10	124								30%

จากตารางแสดงรายวิชาที่ถูกคัดเลือกมาถือเป็นปัจจัยในการสร้างแบบจำลองในการพยากรณ์ความสำเร็จทางการเรียนของผู้เรียนระหว่างการเรียนในหลักสูตร เมื่อพิจารณาจากจำนวนปัจจัยที่ใช้ในการสร้างแบบจำลองในหลักสูตรเทคโนโลยีสารสนเทศทั้งภาคปกติและภาคพิเศษ พบว่า มีรายวิชาที่มีผู้เรียนลงทะเบียนเรียนครบทั้งหมดของหลักสูตรคิดเป็น 37% และ 38% ตามลำดับ สำหรับหลักสูตรวิทยาการคอมพิวเตอร์ทั้งภาคปกติ และภาคพิเศษ มีรายวิชาที่มีผู้ลงทะเบียนเรียนครบทั้งหมดคิดเป็น 29% และ 30% ตามลำดับ กล่าวได้ว่ารายวิชาที่คัดเลือกถือเป็นรายวิชาที่จะเป็นตัวแทนของรายวิชาในหลักสูตรซึ่งจะถูกนำไปสร้างแบบจำลองในการพยากรณ์ต่อไป

การสร้างตัวแบบและการวัดประสิทธิภาพ

งานวิจัยนี้มุ่งเน้นการสร้างแบบจำลองการพยากรณ์ที่มีประสิทธิภาพและแสดงให้เห็นถึงความสัมพันธ์ของตัวแปรแต่ละตัวกับความสำเร็จในการเรียน ผู้วิจัยเลือกใช้โปรแกรม WEKA version 3.6.7 ทำการวิเคราะห์ข้อมูลและสร้างแบบจำลอง สำหรับการทดลองผู้วิจัยทำการแบ่งชุดข้อมูล (Data set) เพื่อใช้ในการสร้างแบบจำลองออกเป็น 7 ชุดข้อมูลสำหรับแต่ละหลักสูตร ดังตารางที่ 2 เพื่อค้นหาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการเรียนในหลักสูตร โดยเทคนิควิธีเหมืองข้อมูล สำหรับวิธีการสร้างแบบจำลองผู้วิจัยเลือกเทคนิควิธี 4 เทคนิค ได้แก่ การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) โดยผลลัพธ์ที่ได้จากเทคนิควิธีทั้ง 4 จะถูกนำมาคัดเลือกเทคนิควิธีที่มีค่าประสิทธิภาพสูงที่สุด สำหรับการกำหนดดัชนีวัดประสิทธิภาพของแบบจำลองแต่ละแบบ ผู้วิจัยเลือกใช้เทคนิคการวัดความคลาดเคลื่อน (Error of Measurement) 2 วิธี ได้แก่ Root Mean Squared Error (RMSE) และ Relative Absolute Error (RAE) เพื่อแสดงการเปรียบเทียบประสิทธิภาพของแบบจำลอง

ผลการวิจัย

จากผลการทดลอง ผู้วิจัยได้ทำการวัดประสิทธิภาพการสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการเรียนในหลักสูตร ด้วยโปรแกรม WEKA ตามวิธีการดำเนินงานวิจัย และคำนวณค่า Root Mean Squared Error (RMSE) และ Relative Absolute Error (RAE) โดยใช้อัลกอริทึมทั้ง 4 อัลกอริทึม ผลการทดสอบสามารถแบ่งออกเป็น 2 ส่วนดังนี้

ผลการวัดประสิทธิภาพการสร้างแบบจำลองหลักสูตรเทคโนโลยีสารสนเทศ

จากชุดข้อมูลที่ผู้วิจัยเลือกใช้ในการสร้างแบบจำลองจำนวน 7 ชุดข้อมูล โดยแบ่งชุดข้อมูลตามภาคการศึกษาของผู้เรียนซึ่งผู้เรียนจะอยู่ในกระบวนการเรียนการสอนทั้งหมด 8 ภาคการศึกษา สำหรับภาคการศึกษาที่ 7 และ ภาคการศึกษาที่ 8 ซึ่งผู้เรียนอยู่ในชั้นปีที่ 4 เนื่องจากเป็นภาคการศึกษาที่กำหนดให้ผู้เรียนเลือกรายวิชาในกลุ่มวิชาเฉพาะเลือกซึ่งผู้เรียนแต่ละคนมีการเรียนในรายวิชาที่แตกต่างกัน จึงทำการรวมรายวิชาในภาคการศึกษาที่ 7 และ 8 เป็น 1 กลุ่มตัวแปร นำชุดข้อมูลที่จัดแบ่งมาสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษในหลักสูตร และวิเคราะห์ประสิทธิภาพที่ได้จากเทคนิควิธีเหมือนข้อมูลทั้ง 4 วิธี ได้แก่ การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression)

โดยใช้วิธีการเปรียบเทียบความคลาดเคลื่อนด้วยวิธีการ RMSE และ RAE จากผลการทดลองสามารถแสดงได้ดังตารางที่ 3

ตารางที่ 3 รายละเอียดการเปรียบเทียบประสิทธิภาพการสร้างแบบจำลองหลักสูตรเทคโนโลยีสารสนเทศ

ชุดข้อมูล			การวัดความคลาดเคลื่อน (Error of Measurement)							
ลำดับ	ภาคการเรียน (เทอม/ปี)	จำนวน รายวิชา	MLR		LR		MLP		SMOreg	
			RMSE	%RAE	RMSE	%RAE	RMSE	%RAE	RMSE	%RAE
IT#1	1/1	6	0.26	59.18	0.34	76.79	0.29	65.66	0.26	59.86
IT#2	1/1 – 1/2	14	0.21	47.99	0.33	73.77	0.27	59.92	0.21	48.77
IT#3	1/1 – 2/1	22	0.18	39.92	0.31	69.95	0.29	62.23	0.18	40.14
IT#4	1/1 – 2/2	28	0.14	30.33	0.30	67.06	0.26	53.75	0.14	30.20
IT#5	1/1 – 1/3	35	0.11	23.47	0.28	61.39	0.20	41.80	0.11	23.38
IT#6	1/1 – 2/3	43	0.09	20.49	0.28	62.15	0.17	36.13	0.10	20.73
IT#7	1/1 – ปี 4	50	0.09	18.68	0.28	62.15	0.16	34.29	0.09	18.04

จากผลการทดลองในตารางที่ 3 แสดงให้เห็นถึงประสิทธิภาพของแบบจำลองในการพยากรณ์ที่ใช้ในการประมาณเกรดเมื่อจบการศึกษาของผู้เรียนของหลักสูตรเทคโนโลยีสารสนเทศในแต่ละภาคการศึกษา ซึ่งได้จากเทคนิควิธีทางเหมือนข้อมูลทั้ง 4 แบบ เมื่อพิจารณาจากผลการทดลองพบว่า เมื่อผู้เรียนลงทะเบียนเรียนในภาคการศึกษาที่ 1 ปีการศึกษาที่ 1 (IT#1) และได้ผลการเรียนใน 6 รายวิชา แบบจำลองการพยากรณ์ที่สร้างขึ้นโดยเทคนิควิธีการถดถอยเชิงเส้นพหุ (Multiple Linear Regression : MLR) มีประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุดโดยวัดค่าความคลาดเคลื่อน RMSE = 0.26 และ RAE = 59.18% เมื่อผู้เรียนเรียนจบปีการศึกษาที่ 1 (IT#2) และเลือกรายวิชาจำนวน 14 รายวิชา ผลการสร้างแบบจำลองพบว่าเทคนิควิธีการถดถอยเชิงเส้นพหุ (Multiple Linear Regression : MLR) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุดโดยวัดค่าความคลาดเคลื่อน RMSE = 0.21 และ RAE = 47.99% และเมื่อผู้เรียนผ่านการเรียนในชั้นปีที่ 2 (IT#4) และเลือกรายวิชาจำนวน 28 รายวิชา ผลการสร้างแบบจำลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการ

ถดถอย (Support Vector Machine for Regression : SMOReg) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุดโดยวัดค่าความคลาดเคลื่อน RMSE = 0.14 และ RAE = 30.20% และเมื่อนิสิตเรียนจบปีการศึกษาที่ 3 (IT#6) และ ปีการศึกษาที่ 4 (IT#7) พบว่าเทคนิควิธีการถดถอยเชิงเส้นพหุ (Multiple Linear Regression : MLR) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุด โดยมีค่า RMSE = 0.11 RAE = 23.38% และ RMSE = 0.09 และ RAE = 18.04% ตามลำดับ

ผลการวัดประสิทธิภาพการสร้างแบบจำลองหลักสูตรวิทยาการคอมพิวเตอร์

สำหรับการสร้างแบบจำลองการพยากรณ์ในหลักสูตรวิทยาการคอมพิวเตอร์ ผู้วิจัยเลือกใช้ชุดข้อมูลในการสร้างแบบจำลองจำนวน 7 ชุดข้อมูล โดยแบ่งชุดข้อมูลตามภาคการศึกษาของผู้เรียนซึ่งผู้เรียนจะอยู่ในกระบวนการเรียนการสอนทั้งหมด 8 ภาคการศึกษา สำหรับภาคการศึกษาที่ 7 และ ภาคการศึกษาที่ 8 ซึ่งผู้เรียนอยู่ในชั้นปีที่ 4 เนื่องจากเป็นภาคการศึกษาที่กำหนดให้ผู้เรียนเลือกรายวิชาในกลุ่มวิชาเฉพาะเลือกซึ่งผู้เรียนแต่ละคนมีการเรียนในรายวิชาที่แตกต่างกัน จึงทำการรวมรายวิชาในภาคการศึกษาที่ 7 และ 8 เป็น 1 กลุ่มตัวแปร จากผลการทดลองสามารถแสดงได้ดังตารางที่ 4

ตารางที่ 4 รายละเอียดการเปรียบเทียบประสิทธิภาพการสร้างแบบจำลองหลักสูตรวิทยาการคอมพิวเตอร์

ชุดข้อมูล			การวัดความคลาดเคลื่อน (Error of Measurement)							
ลำดับ	ภาคการเรียน (เทอม/ปี)	จำนวน รายวิชา	MLR		LR		MLP		SMOReg	
			RMSE	%RAE	RMSE	%RAE	RMSE	%RAE	RMSE	%RAE
CS#1	1/1	7	0.28	60.09	0.35	74.36	0.32	68.18	0.29	60.02
CS#2	1/1 – 1/2	14	0.23	47.65	0.36	77.59	0.32	64.90	0.23	47.11
CS#3	1/1 – 2/1	19	0.17	35.48	0.36	75.20	0.28	55.12	0.17	35.94
CS#4	1/1 – 2/2	27	0.13	26.92	0.29	61.66	0.25	49.64	0.13	26.53
CS#5	1/1 – 1/3	33	0.11	21.62	0.30	62.78	0.20	40.94	0.11	21.55
CS#6	1/1 – 2/3	37	0.09	18.85	0.30	62.78	0.17	34.69	0.09	18.84
CS#7	1/1 – ปี 4	39	0.09	18.40	0.30	62.78	0.16	32.34	0.09	18.68

จากผลการทดลองในตารางที่ 4 แสดงให้เห็นถึงประสิทธิภาพของแบบจำลองในการพยากรณ์ที่ใช้ในการประมาณเกรดเมื่อจบการศึกษาของผู้เรียนของหลักสูตรเทคโนโลยีสารสนเทศในแต่ละภาคการศึกษา ซึ่งได้จากเทคนิควิธีทางเหมืองข้อมูลทั้ง 4 แบบ พิจารณาจากผลการทดลอง เมื่อผู้เรียนเรียนจบปีการศึกษาที่ 1 (CS#2) และเลือกรายวิชาจำนวน 14 รายวิชา ผลการสร้างแบบจำลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression : SMOReg) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุดโดยวัดค่าความคลาดเคลื่อน RMSE = 0.23 และ RAE = 47.11% และเมื่อผู้เรียนผ่านการเรียนในชั้นปีที่ 2 (CS#4) และเลือกรายวิชาจำนวน 27 รายวิชา ผลการสร้างแบบจำลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression : SMOReg) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงสุดโดยวัด ค่าความคลาดเคลื่อน RMSE = 0.13 และ RAE = 26.53% และเมื่อนิสิตเรียนจบปีการศึกษาที่ 3 (CS#6) ผลการสร้างแบบจำลองพบว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับ

การถดถอย (Support Vector Machine for Regression : SMOReg) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงที่สุดโดยวัดค่าความคลาดเคลื่อน $RMSE = 0.09$ และ $RAE = 18.84\%$ ในปีการศึกษาที่ 4 (CS#7) พบว่าเทคนิควิธีการถดถอยเชิงเส้นพหุ (Multiple Linear Regression : MLR) ประสิทธิภาพในการพยากรณ์ผลการเรียนเมื่อผู้เรียนสำเร็จการศึกษาสูงที่สุด โดยมีค่า $RMSE = 0.09$ $RAE = 18.40\%$

สรุปและอภิปรายผล

งานวิจัยนี้มีวัตถุประสงค์ในการสร้างแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษา โดยใช้เทคนิคเหมืองข้อมูล ซึ่งมีการนำวิธีการสร้างแบบจำลอง 4 เทคนิค ได้แก่ การถดถอยเชิงเส้นพหุ (Multiple Linear Regression) การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) โดยชุดข้อมูล (Data Set) ที่นำมาใช้ในการวิจัยครั้งนี้เป็นผลการศึกษาในระหว่างที่ผู้เรียนศึกษาอยู่ในหลักสูตร โดยมีจำนวนผู้เรียนที่สำเร็จการศึกษาแล้วทั้งสิ้น 1,263 คน ผลลัพธ์ที่ได้จากการสร้างแบบจำลองจะถูกนำไปประเมินผลประสิทธิภาพแบบจำลองโดยใช้การวัดความคลาดเคลื่อน (Error of Measurement) 2 วิธีได้แก่ Root Mean Squared Error (RMSE) และ Relative Absolute Error (RAE) เพื่อใช้เป็นตัวชี้ในการเลือกตัวแบบที่เหมาะสมสำหรับพยากรณ์ความสำเร็จของผู้เรียนในแต่ละภาคการศึกษาของผู้เรียน ผลการวิจัยแสดงให้เห็นว่า แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีความเหมาะสมในการพยากรณ์ความสำเร็จของผู้เรียนเมื่อเรียนอยู่ในภาคการศึกษาที่ 1 2 3 และ 6 สำหรับแบบจำลองการพยากรณ์ที่ได้จากอัลกอริทึมซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยมีความเหมาะสมในการพยากรณ์ความสำเร็จของผู้เรียนเมื่อเรียนอยู่ในภาคการศึกษาที่ 4 5 และ 7

เมื่อเปรียบเทียบประสิทธิภาพของแบบจำลองพบว่า ผู้เรียนในหลักสูตรเทคโนโลยีสารสนเทศและวิทยาการคอมพิวเตอร์จะมีค่าความถูกต้องในการพยากรณ์เพิ่มขึ้นเมื่อผู้เรียนศึกษาในหลักสูตรชั้นปีที่สูงขึ้น และเมื่อพิจารณาจากผลการทดลองสามารถพบว่าในชั้นปีที่ 2 ภาคการศึกษาที่ 2 ผลการพยากรณ์มีค่าความผิดพลาดร้อยละ 30.20 ในหลักสูตรเทคโนโลยีสารสนเทศ และร้อยละ 26.53 ในหลักสูตรวิทยาการคอมพิวเตอร์ จากค่าระดับความผิดพลาดดังกล่าวสะท้อนให้เห็นได้ว่าชั้นปีที่ 2 มีความสำคัญต่อการกำหนดผลสัมฤทธิ์ทางการเรียนของผู้เรียนในหลักสูตรเมื่อศึกษาจนจบการศึกษา ทั้งนี้สอดคล้องกับการออกแบบหลักสูตรในด้านการจัดการรายวิชาเฉพาะบังคับ และวิชาเฉพาะเลือกซึ่งเป็นรายวิชาที่สอดคล้องกับความถนัดและความสนใจของผู้เรียน

จากผลการวิจัยพบว่าสาเหตุสำคัญที่ส่งผลต่อประสิทธิภาพของแบบจำลอง ได้แก่ การเลือกใช้เทคนิควิธีการสร้างแบบจำลองซึ่งถือเป็นส่วนสำคัญตามแนวทางในการค้นพบความรู้แฝงในฐานข้อมูล และเมื่อพิจารณาถึงรายวิชาที่ถูกเลือกมาใช้ในการสร้างแบบจำลองในแต่ละภาคการศึกษาจะเห็นได้ว่า รายวิชาที่เปลี่ยนไปในแต่ละภาคการศึกษาคือจำนวนตัวแปร (Feature) ที่เป็นองค์ประกอบของชุดข้อมูลในการสร้างแบบจำลอง การเพิ่มขึ้นของตัวแปรที่เลือกในแต่ละภาคการศึกษาส่งผลต่อประสิทธิภาพที่แตกต่างกันในแต่ละเทคนิควิธีที่ใช้สำหรับการสร้างแบบจำลอง (Guyon & Elisseeff, 2003; Han, 2012; Ladha & Deepa, 2011) ซึ่งสอดคล้องกับผลการทดลองที่มีค่าความคลาดเคลื่อนแตกต่างกันในแต่ละเทคนิควิธี

ผลจากการพัฒนาแบบจำลองการพยากรณ์ความสำเร็จทางการศึกษาของผู้เรียนระหว่างการศึกษาในหลักสูตร สามารถนำไปเป็นแนวทางในการพัฒนาวิธีการใช้ประโยชน์จากชุดข้อมูลที่มีอยู่เดิมในองค์กรเพื่อสร้าง

สารสนเทศใหม่ รวมทั้งใช้ประโยชน์ในด้านการวางแผนการจัดการเรียนการสอน แบ่งได้เป็น 2 ส่วนได้แก่ ส่วนของผู้เรียนในหลักสูตรสามารถนำข้อมูลผลการศึกษาของตนเองที่ได้รับจากการเรียนในหลักสูตรมาใช้เป็นข้อมูลทำนาย (Unseen Data) นำเข้าสู่แบบจำลองเพื่อให้ได้ผลลัพธ์เป็นผลการเรียนเมื่อสิ้นสุดการศึกษา ผลพยากรณ์ดังกล่าวจะทำให้ผู้เรียนสามารถใช้ประโยชน์ในการวางแผนการเรียนในภาคการศึกษาถัดไปได้อย่างมีประสิทธิภาพ

ส่วนที่สองคือ ส่วนของอาจารย์ประจำหลักสูตรหรืออาจารย์ที่ปรึกษา ซึ่งมีหน้าที่สำคัญในการดูแลผู้เรียนที่ศึกษาในหลักสูตรให้มีความสำเร็จในการเรียน รวมทั้งลดปริมาณผู้เรียนที่ล้มเหลวในการเรียนให้ได้มากที่สุดจากแบบจำลองที่ได้จากการวิจัย สามารถนำผลการเรียนของผู้เรียนในหลักสูตรทุกคนทุกชั้นปีมาพยากรณ์ผลการเรียนเมื่อสิ้นสุดภาคการศึกษา เพื่อระบุระดับความเสี่ยงต่อการล้มเหลวในการเรียนของผู้เรียนในหลักสูตร และหาแนวทางในการลดปัญหาดังกล่าวต่อไป

เอกสารอ้างอิง

- Frias-Martinez, E., Sanchez, A., and Velez, J. (2006). Support vector machines versus multi-layer perceptrons for efficient off-line signature recognition. *Engineering Applications of Artificial Intelligence*, 19(6), 693–704.
- Guyon, I., and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.*, 3, 1157–1182.
- Han, J. (2012). *Data mining: concepts and techniques* (3rd ed). Burlington, MA: Elsevier.
- Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.
- Jananii, G. (2012). Applying multi layer feed forward neural networks on large scale data. *Bonfring International Journal of Man Machine Interface*, 2(4), 1.
- Ladha, L., and Deepa, T. (2011). Feature selection methods and algorithms. *International Journal on Computer Science and Engineering*, 3(5), 1787–1797.
- Roiger, R. J. (2017). *Data mining: a tutorial-based primer*. CRC Press.
- Shevade, S. K., Keerthi, S. S., Bhattacharyya, C., and Murthy, K. R. K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transactions on Neural Networks*, 11(5), 1188–1193.
- Smola, A. J., and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- Vilailuck, S., Jaroenpuntaruk, V., and Wichadakul, D. (2015). Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center. *Veridian E-Journal Science and Technology Silpakorn University*, 2(2), 1–17. (in Thai)
- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Yadav, S. K., Bharadwaj, B., and Pal, S. (2012). Data Mining Applications: A comparative Study for Predicting Student's performance. *World of Computer Science and Information Technology Journal WCSIT*, 2012(2), 51–55.