

## การพยากรณ์ผลการสอบ TOEIC สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต โดยใช้เทคนิคเหมืองข้อมูล

### An Application of the Data Mining Technique to the Prediction of the TOEIC Test Results of English Majors, Rangsit University

ศศิธร ตุ่มเพชรรัตน์<sup>1\*</sup> สมบูรณ์ เอนกฤทธิมงคล<sup>2</sup> และ สุมนา เกษมสวัสดิ์<sup>3</sup>

Sasitorn Tumpetcharat<sup>1\*</sup> Somboon Anekritmongkol<sup>2</sup> and Sumana Kasemsawadi<sup>3</sup>

<sup>1</sup>นักศึกษาระดับปริญญาตรี หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาระบบสารสนเทศวิสาหกิจ  
วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต

<sup>2</sup>อาจารย์ประจำ หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ  
วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต

<sup>3</sup>อาจารย์ประจำ หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาระบบสารสนเทศวิสาหกิจ  
วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต

<sup>1</sup>Graduate student in Bachelor of Science Program in Enterprise Information Systems,  
College of Information and Communication Technology, Rangsit University

<sup>2</sup>Lecturer in Master of Science in Information Technology,

College of Information and Communication Technology, Rangsit University

<sup>3</sup>Lecturer in Bachelor of Science Program in Enterprise Information Systems,

College of Information and Communication Technology, Rangsit University

\*Corresponding author, E mail: sasitorn.t56@rsu.ac.th

#### บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลพยากรณ์ผลการสอบ TOEIC ของนักศึกษาคณะศิลปศาสตร์ ที่สามารถสอบได้ 500 คะแนนขึ้นไป ข้อมูลนักศึกษาคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต จำนวน 205 คน โดยรวบรวมผลการเรียนในรายวิชาหมวดวิชาชีพมาใช้ในการสร้างโมเดลพยากรณ์ ในการวิจัยครั้งนี้ผู้วิจัยได้สร้างโมเดลการพยากรณ์ทั้งหมด 3 โมเดล เพื่อหาวิธีการที่มีความถูกต้องมากที่สุดด้วยเทคนิคเหมืองข้อมูล ได้แก่ Decision Tree, Naïve-Bayes และ k-nearest neighbors จากการวัดประสิทธิภาพและเปรียบเทียบ พบว่าโมเดลที่สร้างด้วยวิธี k-nearest neighbor มีประสิทธิภาพให้ค่าความถูกต้องแม่นยำในการพยากรณ์สูงสุด โดยมีค่าความถูกต้องเฉลี่ยอยู่ที่ร้อยละ 98.86

คำสำคัญ: TOEIC , เหมืองข้อมูล, การพยากรณ์, ต้นไม้ตัดสินใจ, นาอ์เบย์, เคเนียร์สเนเบอร์

## Abstract

The research objective was to create a prediction model of the TOEIC test results for students at the Faculty of Liberal Arts whose test scores were 500 or higher. Data were collected from 205 students from the Faculty of Liberal Arts, Rangsit University. In order to create this prediction model, the data mining techniques (Decision Tree, Naïve-Bayes and k-nearest neighbors) were applied. After comparing the effectiveness of each model, the finding revealed that the most effective and precise model for prediction was k-nearest neighbor. The validation average was at 98.86 percent.

**Keywords:** TOEIC, Data mining, Decision Tree, Naïve-Bayes, k-nearest neighbors.

## 1. บทนำ

คณะศิลปศาสตร์ มหาวิทยาลัยรังสิต ได้เปิดสอนหลักสูตรสาขาวิชาภาษาอังกฤษครั้งแรกขึ้น ในปีการศึกษา 2542 นับเป็นเวลา 17 ปีแล้วที่ภาควิชาภาษาอังกฤษ ได้ผลิตบัณฑิตออกสู่สังคม โดยปัจจุบันสาขาวิชาภาษาอังกฤษ ได้ใช้ชื่อปริญญา ศิลปศาสตรบัณฑิต (ภาษาอังกฤษ), Bachelor of Arts (English) โดยมีชื่อย่อ ศศ.บ. (ภาษาอังกฤษ), B.A. (English) โดยมีวัตถุประสงค์เพื่อผลิตบัณฑิตที่มีความรู้ความสามารถในทางภาษาอังกฤษ มีความรู้ทางสาขาวิชาชีพ และมีความสามารถในการบูรณาการความรู้ทั้งสองด้านนี้ เพื่อเป็นพื้นฐานในการประกอบวิชาชีพทั้งในภาครัฐและเอกชน และเพื่อผลิตบัณฑิตที่มีคุณสมบัติพร้อมในด้านภาษาอังกฤษ เพื่อนำไปใช้ในการศึกษาระดับสูงขึ้นไป ในสถาบันการศึกษาในประเทศและต่างประเทศต่อไป โดยจะใช้เวลาในการศึกษาตลอดหลักสูตร 4 ปี มีจำนวนหน่วยกิตรวมตลอดหลักสูตร 136 หน่วยกิต แยกเป็นหมวดวิชาทั่วไป 30 หน่วยกิต หมวดวิชาเฉพาะ 100 หน่วยกิต และหมวดวิชาเลือกเสรี 6 หน่วยกิต ซึ่งบัณฑิตที่จบการศึกษา สามารถประกอบอาชีพได้หลายอาชีพ เช่น พนักงานสายการบิน พนักงานโรงแรม ล่าม ผู้ประกาศข่าว เลขานุการ ครู อาจารย์ นักวิชาการ และอื่นๆ

การสอบ TOEIC เป็นอีกหนึ่งในการทดสอบความรู้ความสามารถทางด้านภาษาอังกฤษ ของบัณฑิตสาขาวิชาภาษาอังกฤษได้เป็นอย่างดี ซึ่งคะแนนสอบ TOEIC จะมีคะแนนเต็มที่ 990 คะแนน โดยข้อสอบจะแบ่งออกเป็น 2 Part ได้แก่ PART 1 : LISTENING มีทั้งหมด 100 ข้อ คะแนนรวม 495 คะแนน เวลา 45 นาที ประกอบไปด้วย Photographs, Question-Response, Conversations และ Talks และ PART 2 : READING มีทั้งหมด 100 ข้อ คะแนนรวม 495 คะแนน เวลา 1 ชม. 15 นาที ประกอบไปด้วย Incomplete Sentences, Text Completion และ Reading Comprehension ใช้เวลาในการสอบ 1 ชั่วโมง 15 นาที รวมทั้งหมด 2 ชั่วโมง โดยบัณฑิตนั้นจะต้องมีผลคะแนนในการสอบ TOEIC ตามที่บริษัท นายจ้างได้กำหนดคะแนนไว้ ทั้งนี้คะแนนที่รับพิจารณาเพื่อคัดเลือกให้เป็นพนักงานขึ้นอยู่กับบริษัทนายจ้างเป็นผู้กำหนด เช่นบริษัทที่ให้บริการเกี่ยวกับการบินบางแห่งกำหนดคะแนน TOEIC อยู่ที่ 500 คะแนนขึ้นไป หรือบางแห่งอาจจะกำหนดอยู่ที่ 550 คะแนน เพื่อเป็นการยืนยันว่าบัณฑิตที่จะรับเข้าเป็นพนักงานนั้น เป็นผู้ที่มีความรู้ความสามารถทางด้านภาษาอังกฤษ และสามารถปฏิบัติหน้าที่ที่จะมอบหมายให้ เป็นไปได้ด้วยดี บริษัทโดยส่วนมากมักจะกำหนดคะแนนที่นำมาใช้ในการสมัครงานนั้นต้องมีอายุไม่เกิน 2 ปี นับตั้งแต่วันที่ได้รับผลสอบ ซึ่งในการสอบแต่ละครั้งจะมีทั้งบัณฑิตที่สอบที่ได้คะแนนสอบ TOEIC ได้ตามเกณฑ์ที่กำหนดและบัณฑิตบางคนก็ไม่สามารถสอบได้คะแนนตามเกณฑ์ที่กำหนด

เนื่องจากการสอบ TOEIC ของนักศึกษาคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต มีนักศึกษาบางส่วนที่สอบไม่ผ่านเกณฑ์ที่กำหนด เนื่องจากสาเหตุต่างๆ เช่น การเตรียมพร้อมในการสอบน้อยเกินไป องค์ประกอบความรู้ ความสามารถมีไม่มากพอ

จากปัญหาดังกล่าว ผู้วิจัยจึงเกิดแนวคิดที่จะนำผลสัมฤทธิ์ทางการเรียน ในหมวดวิชาชีพของนักศึกษา คณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต ที่จะสอบวัดความรู้ทางด้านภาษาอังกฤษ TOEIC เพื่อนำผลคะแนนไปสมัครงานหรือศึกษาต่อ มาวิเคราะห์พยากรณ์ ความเป็นไปได้ที่นักศึกษาจะสอบ TOEIC ได้คะแนนที่ 500 คะแนนขึ้นไป และเพื่อนำผลที่ได้จากการวิเคราะห์เบื้องต้นนี้ เป็นแนวทางในการพัฒนาหลักสูตรการสอน และเตรียมความพร้อมในการสอบ TOEIC ให้ดียิ่งขึ้น

ในการวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยได้เลือกใช้เทคนิค Data Mining (รศ. สายชล สัน สมบูรณ์ทอง, 2557) ซึ่งเป็นเทคนิคในการวิเคราะห์ข้อมูลอย่างหนึ่ง ซึ่งมาจากคำว่า “เหมืองข้อมูล” การทำเหมืองข้อมูลคือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบ (Patterns) ความสัมพันธ์และกฎ (Rules) ที่ซ่อนอยู่ในชุดข้อมูลนั้น

ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภททั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคมการทำเหมืองข้อมูลเปรียบเสมือนเป็นวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูลจากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย มาสู่การจัดเก็บในฐานข้อมูลที่สามารถดึงค่าสารสนเทศของข้อมูลมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูลนั้นๆ (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2557)

## 2. วัตถุประสงค์

เพื่อพยากรณ์ผลการสอบ TOEIC ของนักศึกษาคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต

## 3. อุปกรณ์และวิธีการ / วิธีดำเนินการวิจัย

ได้นำข้อมูลของนักศึกษาคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต จำนวน 205 คน โดยใช้ซอฟต์แวร์ RapidMiner Studio 7.3 (เอกสิทธิ์ พัทธวงศ์ศักดิ์, 2557) เพื่อประมวลผลเหมืองข้อมูลในวิธีการต่างๆ เพื่อหาวิธีการที่ถูกต้องมากที่สุดในการทำนายผลการสอบ TOEIC

โดยมีขั้นตอนดังนี้ เริ่มจากการจัดเตรียมข้อมูล ปรับโครงสร้างข้อมูล เพื่อให้พร้อมต่อการนำไปวิเคราะห์วิเคราะห์ข้อมูล พร้อมสรุปผล

งานวิจัยนี้ได้นำเทคนิคการทำเหมืองข้อมูลมาใช้เพื่อใช้ในการวิเคราะห์การพยากรณ์ผลการสอบ TOEIC ซึ่งคณะศิลปศาสตร์ 205 คน ได้มีข้อกำหนด สำหรับประเมินผู้สำเร็จการศึกษาในหลักสูตรศิลปศาสตรบัณฑิต สาขาวิชาภาษาอังกฤษ โดยข้อมูลที่ใช้ในการวิเคราะห์คือผลการเรียน โดยเลือกเฉพาะหมวดวิชาชีพประกอบไปด้วยรายวิชา (ENG212) APPLIED ENGLISH STRUCTURE, (ENG213) ENGLISH PHONETICS, (ENG314) CULTURE AND ENGLISH LANGUAGE, (ENG221) READING I, (ENG231) WRITING I, (ENG241) LISTENING – SPEAKING I, (ENG322) CRITICAL READING, (ENG323) ACADEMIC READING, (ENG332) WRITING II, (ENG333) ACADEMIC WRITING, (ENG342) LISTENING – SPEAKING II, (ENG343) ACADEMIC LISTENING – SPEAKING SKILLS

ขั้นตอนแรกใช้ Decision Tree เพื่อตรวจสอบว่ามีวิชาใดบางที่เป็นปัจจัยที่มีผลต่อการพยากรณ์ผลการสอบ หลังจากการสร้างโมเดลพบว่า มีบางวิชาที่ไม่ได้เป็นปัจจัยในการมาสร้างโมเดลใหม่ การสร้างโมเดลซึ่งปรับบางรายวิชาออก เลือกเฉพาะรายวิชาที่มีผลต่อการสอบ TOEIC ดังนี้ ENG323 , ENG212 , ENG231 , ENG322 , ENG221 และได้นำข้อมูลมาวิเคราะห์จำแนกคลาส (Classification) ด้วยเทคนิคการทำเหมืองข้อมูล โดยใช้ซอฟต์แวร์ RapidMiner Studio 7.3 เป็นเครื่องในการวิเคราะห์ข้อมูลและสร้างโมเดลการพยากรณ์ผลการสอบ ของบัณฑิตคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต โดยใช้ข้อมูลของผู้ที่เคยผ่านการสอบระหว่างปี พ.ศ. 2556 – 2559 มีจำนวนชุดข้อมูล 205 คน มีทั้งหมด 14 แอททริบิวต์ ดังตารางที่ 1

ซึ่งได้ข้อมูลผลการสอบ TOEIC ของคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษ และผลการเรียนจากระบบทะเบียนผลการเรียนของมหาวิทยาลัยรังสิต

ตารางที่ 1 การเก็บข้อมูลจำนวน 14 แอททริบิวต์ (ตัวอักษร) ที่ใช้ในการสร้างโมเดล (14 รายวิชา)

ลำดับ	รายวิชาที่ใช้ในการทำนาย			ผลการสอบ
	1	...	14	
	ENG212	...	ENG452	
1	A	...	B	Y
...	...	...	...	...
205	B+	...	C	N

ตารางที่ 2 การเก็บข้อมูลจำนวน 14 แอททริบิวต์ (ตัวเลข) ที่ใช้ในการสร้างโมเดล (14 รายวิชา)

ลำดับ	รายวิชาที่ใช้ในการทำนาย			ผลการสอบ
	1	...	14	
	ENG212	...	ENG452	
1	4	...	3.5	Y
...	...	...	...	...
205	3.5	...	2	N

จากตารางที่ 2 มีกระบวนการปรับเปลี่ยนข้อมูลเป็นตัวเลขแทนตัวอักษร เพื่อให้สามารถรันได้ทุกอัลกอริทึม โดยที่ A = 4 , B+ = 3.5 , B = 3 , C+ = 2.5 , C = 2 , D+ = 1.5 และ D = 1

ในการวิจัยครั้งนี้ผู้วิจัยได้สร้างโมเดลในการพยากรณ์ด้วยเทคนิคเหมืองข้อมูล 3 โมเดล ด้วยวิธี Decision Tree (กฤตยา ทองผาสุข, 2554), Naïve Bayes (Jiawei Han, Micheline Kamber, and Jian Pei. Data Mining, 2011) และ K-NN (K-Nearest Neighbors) (T. Hristea and Florentina, 2013) เพื่อวัดประสิทธิภาพของแต่ละโมเดล ด้วยวิธีการ 10-fold cross-validation ของโปรแกรม RapidMiner เพื่อให้กระบวนการเรียนรู้จะเพิ่มประสิทธิภาพของรูปแบบเพื่อให้พอดี (fit) กับข้อมูล Training เท่าที่จะเป็นไปได้ หากใช้วิธีการอื่นที่ใช้ข้อมูลที่อิสระเป็นข้อมูล Training อาจเกิดปัญหาที่เรียกว่า 'over-fitting' โปรแกรม RapidMiner จะแบ่งข้อมูลออกเป็นหลายส่วน (มักแสดงด้วยค่า k) โดยที่แต่ละส่วนจะมีข้อมูลเท่ากัน หลังจากนั้นข้อมูลหนึ่งส่วนจะใช้เป็นตัวทดสอบประสิทธิภาพโมเดล ทำวนไปเช่นนี้จนครบจำนวนที่

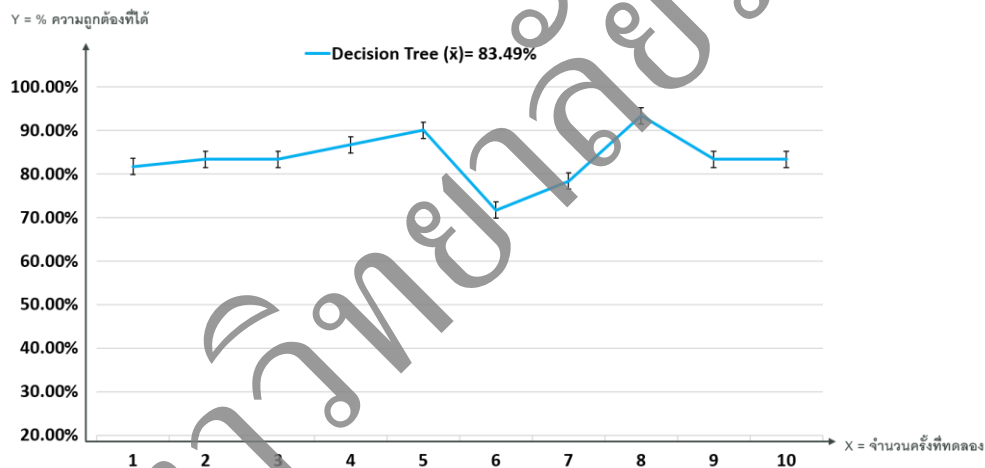
แบ่งไว้เพื่อเปรียบเทียบประสิทธิภาพของแต่ละโมเดลที่สร้างขึ้น ซึ่งทำการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง แล้วหาค่าเฉลี่ยและทำการทดสอบข้อมูลจริง โดยแบ่งข้อมูลออกเป็นข้อมูลเพื่อให้เรียนรู้โมเดล 70 และอีก 30 มาทดสอบโมเดล ทำการทดสอบ 10 ครั้ง โดยใช้โมเดลที่ต่างกัน 3 โมเดล เพื่อหาโมเดลที่มีความถูกต้องมากที่สุด

งานวิจัยนี้ผู้วิจัยได้นำอัลกอริทึม (Algorithm) ได้แก่ ต้นไม้ตัดสินใจ (decision tree) การเรียนรู้เบย์ (Naive Bayes) และ K-NN (K-Nearest Neighbors) เพื่อนำมาวิเคราะห์และเปรียบเทียบประสิทธิภาพทั้ง 3 อัลกอริทึม

จากงานวิจัยระบบสนับสนุนการตัดสินใจในการอนุมัติโครงการที่อยู่อาศัยโดยใช้เทคนิคต้นไม้ตัดสินใจ (ศุภชัย ประคองศิลป์ และ ณัฐวี อุดกฤษฎ์, 2552) ทำเหมืองข้อมูลด้วยวิธีต้นไม้ตัดสินใจเพียงอย่างเดียว แต่งานวิจัยนี้ทำงานเปรียบเทียบวิธีการ การทำเหมืองข้อมูลด้วยกัน 3 วิธีเพื่อค้นหาวิธีที่จะสามารถทำนายได้ถูกต้องมากที่สุด

#### 4. ผลการวิจัย

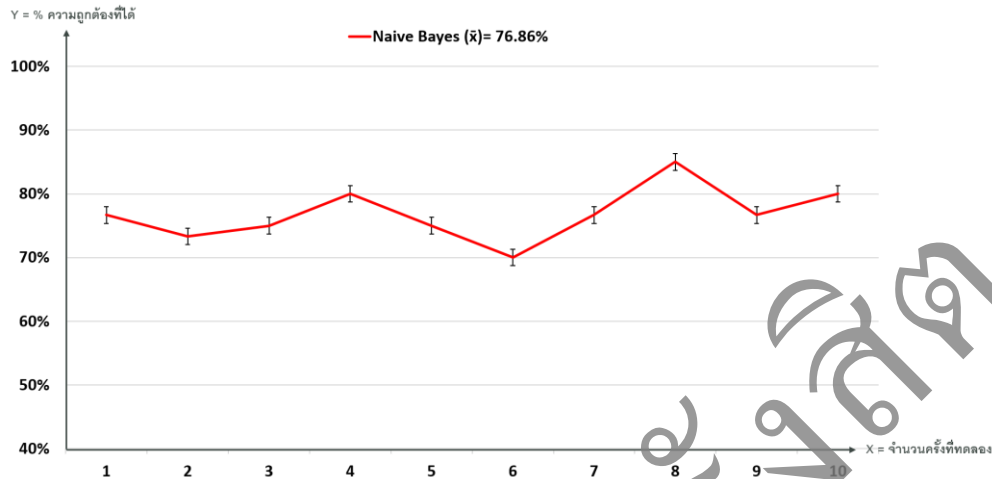
##### 4.1 ผลการทดสอบโมเดลโดยวิธี Decision Tree (ซัดซัย แก้วตา และ อัจฉรา มทาวีร์วัฒน์, 2553)



รูปที่ 2 ผลการทดสอบโมเดลด้วยวิธี Decision Tree

จำนวนข้อมูลทั้งหมด 205 ชุดข้อมูล ทำการทดสอบด้วยวิธีการ 10 Fold Cross Validation ซึ่งทำการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง ซึ่งผลการทดลองที่ได้แสดงในแต่ละจุดของกราฟเกิดจากการสุ่มเลือกของชุดข้อมูลที่เป็นตัว Training และ Testing เปอร์เซนต์ความถูกต้องไม่เท่ากัน เพราะชุดข้อมูล Training และ Testing ที่แตกต่างกัน จึงได้ผลการทดลองในแต่ละครั้งแตกต่างกันในแต่ละจุดของกราฟ โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 83.49% ดังรูปที่ 2

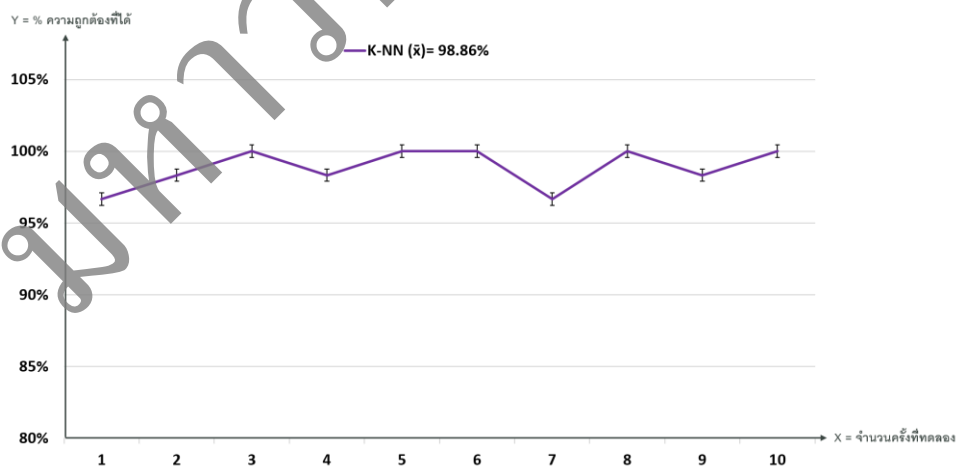
#### 4.2 ผลการทดสอบโมเดลโดยวิธี Naïve Bayes



รูปที่ 3 ผลการทดสอบโมเดลโดยวิธี Naïve Bayes

จำนวนข้อมูลทั้งหมด 205 ชุดข้อมูล ทำการทดสอบด้วยวิธีการ 10 Fold Cross Validation ซึ่งทำการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง ซึ่งผลการทดลองที่ได้แสดงในแต่ละจุดของกราฟเกิดจากการสุ่มเลือกของชุดข้อมูลที่เป็นตัว Training และ Testing เปอร์เซนต์ความถูกต้องไม่เท่ากัน เพราะชุดข้อมูล Training และ Testing ที่แตกต่างกัน จึงได้ผลการทดลองในแต่ละครั้งแตกต่างกันในแต่ละจุดของกราฟ โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 76.86% ดังรูปที่ 3

#### 4.3 ผลการทดสอบโมเดลโดยวิธี K-NN (K-Nearest Neighbors)

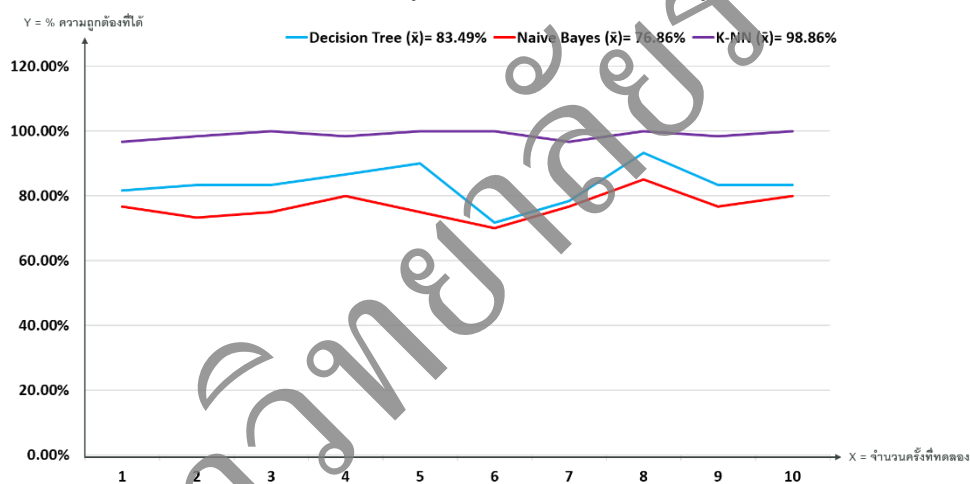


รูปที่ 4 ผลการทดสอบโมเดลโดยวิธี K-NN (K-Nearest Neighbors)

จำนวนข้อมูลทั้งหมด 205 ชุดข้อมูล ทำการทดสอบด้วยวิธีการ 10 Fold Cross Validation ซึ่งทำการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง ซึ่งผลการทดลองที่ได้แสดงในแต่ละจุดของ

กราฟเกิดจากการสุ่มเลือกของชุดข้อมูลที่เป็นตัว Training และ Testing เปรอ์เห็นถึงความถูกต้องไม่เท่ากัน เพราะชุดข้อมูล Training และ Testing ที่แตกต่างกัน จึงได้ผลการทดลองในแต่ละครั้งแตกต่างกันในแต่ละจุดของกราฟ โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 98.86% ดังรูปที่ 4

หลังจากการทดลองทดสอบโมเดลด้วยกัน 3 เทคนิควิธีและเปรียบเทียบประสิทธิภาพพบว่า โมเดลที่สร้างด้วยเทคนิควิธี K-NN (K-Nearest Neighbors) ให้ประสิทธิภาพสูงที่สุดในการทำนาย โดยได้ค่าเฉลี่ยจากการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 98.86% ลงลงมาคือโมเดลที่สร้างด้วยเทคนิควิธี Decision Tree ให้ประสิทธิภาพอยู่ในระดับปานกลางในการทำนาย โดยได้ค่าเฉลี่ยจากการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 83.49% และสุดท้ายคือโมเดลที่สร้างด้วยเทคนิควิธี Naive Bayes โดยให้ประสิทธิภาพอยู่ในระดับน้อยสุดในการทำนาย โดยได้ค่าเฉลี่ยจากการทดลองจำนวน 10 ครั้งด้วยกัน โดยใช้ข้อมูลชุดเดียวกันทั้งหมดในจำนวน 10 ครั้ง โดยให้ค่าความถูกต้องในการพยากรณ์เฉลี่ยอยู่ที่ 76.86%



รูปที่ 5 ผลการเปรียบเทียบ 3 โมเดล

หลังจากที่ได้ทดลองทดสอบ โมเดลทั้ง 3 โมเดลพบว่าโมเดลที่สร้างด้วยเทคนิควิธี K-NN (K-Nearest Neighbors) ให้ประสิทธิภาพสูงที่สุดซึ่งในแต่ละทุกๆ ช่วงของจุดของกราฟให้ค่าความถูกต้องในการพยากรณ์ไม่ต่ำกว่า 80% ดังรูปที่ 5

## 5. อภิปรายผล

เนื่องจากไม่มีเทคนิคใดที่ดีที่สุดในทุกชุดข้อมูลได้ จึงทำการวิจัยเพื่อหาเทคนิคเหมืองข้อมูลที่เหมาะสมกับการทำนายผลการสอบของ นักศึกษาคณะศิลปศาสตร สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต จากการวิจัยการทำเหมืองข้อมูลด้วยเทคนิควิธีทั้ง 3 โมเดล ในการพยากรณ์ผลการสอบ TOEIC ของนักศึกษาคณะศิลปศาสตร สาขาวิชาภาษาอังกฤษ มหาวิทยาลัยรังสิต สามารถทำนายได้ถูกต้องถึง 98.86% และได้เปรียบเทียบหลายเทคนิคเพื่อหาเทคนิคที่ดีที่สุด พบว่าเทคนิควิธีที่ดีที่สุดคือ K-NN (K-Nearest Neighbors) ค่าเฉลี่ยในการพยากรณ์อยู่ที่ 98.86% ลงลงมาคือเทคนิควิธี Decision Tree ค่าเฉลี่ยความถูกต้องในการพยากรณ์อยู่ที่ 83.49% และเทคนิควิธีที่ได้ค่าต่ำสุดคือ เทคนิควิธี

Naïve Bayes ค่าเฉลี่ยความถูกต้องในการพยากรณ์อยู่ที่ 76.86% งานวิจัยนี้สามารถแนะนำนักศึกษาคณะศิลปศาสตร์ สาขาวิชาภาษาอังกฤษว่าจะสอบผ่าน TOEIC หรือไม่ เพื่อเตรียมความพร้อมก่อนสอบได้

เนื่องจากวิธีการทำนายด้วยวิธีการ K-NN (K-Nearest Neighbors) มีความถูกต้องที่สูง แต่อาจมีวิธีการอื่นๆ ที่สามารถทำนายได้ถูกต้องมากกว่านี้และสามารถนำวิธี K-NN (K-Nearest Neighbors) ไปประยุกต์ใช้เพื่อทำนายในข้อมูลที่แตกต่างจากนี้ได้

## 6. บทสรุป

เนื่องจาก นักศึกษาคณะศิลปศาสตร์มหาวิทยาลัยรังสิต จำเป็นต้องสอบ TOEIC จากงานวิจัยที่สามารถนำวิธีการเหมืองข้อมูลโดยวิธี k-nearest neighbors มาทำนายผลการสอบ TOEIC ของนักศึกษาคณะศิลปศาสตร์ มหาวิทยาลัยรังสิตที่มีความถูกต้องสูงถึง 98.86 % โดยเน้นที่ผลการเรียนของวิชา ENG323, ENG212, ENG231, ENG322, ENG221 สามารถช่วยเตรียมความพร้อมให้นักศึกษา หรือสอนเสริม ที่ผลลัพธ์พยากรณ์ออกมาแล้วว่าคะแนนไม่ถึง 500 คะแนน ทำให้นักศึกษาสามารถสอบผ่าน TOEIC มากขึ้น

## 7. กิตติกรรมประกาศ

ขอขอบคุณคณะศิลปศาสตร์ มหาวิทยาลัยรังสิต สำนักงานทะเบียน มหาวิทยาลัยรังสิต ที่อนุเคราะห์ข้อมูลสำหรับงานวิจัยครั้งนี้ ผศ.ดร.สมบูรณ์ เอนกฤทธิ์มงคล รวมทั้งคณาจารย์และเจ้าหน้าที่วิทยาลัยเทคโนโลยีสารสนเทศและการสื่อสาร มหาวิทยาลัยรังสิต ที่ให้คำปรึกษา คำชี้แจงงานวิจัยสำเร็จลุล่วงไปได้ด้วยดี

## 8. เอกสารอ้างอิง

- กฤตยา ทองผาสุข. (2554). การเปรียบเทียบเทคนิคต้นไม้ตัดสินใจ กฎ นาอ็ฟเบย์ และเคเนียร์เซนเบอร์ เพื่อการจำแนกข้อมูล. ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ.
- ชัชชัย แก้วตา และ อัจฉรา มหาวีรวัฒน์. (2553) การวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ. ภาควิชาคณิตศาสตร์สถิติ และคอมพิวเตอร์คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี.
- รศ. สายชล สีนสมบูรณ์ทอง. (2557). การทำเหมืองข้อมูล Data Mining. บริษัท จามจุรี โปรดักส์ จำกัด.
- ศุภชัย ประคองศิลป์ และ ณัฐวี อุดกฤษฎ์. (2552). ระบบสนับสนุนการตัดสินใจในการอนุมัติโครงการที่อยู่อาศัยโดยใช้เทคนิคต้นไม้ตัดสินใจ. กรุงเทพมหานคร: The 5th National Conference on Computing and Information Technology.
- เอกสิทธิ์ พัทธวงศ์ศักดิ์. (2557). การวิเคราะห์ข้อมูลด้วยเทคนิคคต่าไมน์นิง เบื้องต้น. บริษัท เอเชีย ดิจิตอลการพิมพ์ จำกัด.
- Jiawei Han, Micheline Kamber, and Jian Pei. Data Mining. (2011). Concepts and Techniques, 3rd edition, Morgan.
- T. Hristea and Florentina. (2013). The Naïve Bayes Model for Unsupervised Word Sense Disambiguation Aspects Concerning Feature Selection. U.S.A.