



รายงาน

เรื่อง การพยากรณ์การสอบผ่านวิชา IC3 โดยวิธีการเรียนรู้ของเครื่อง ด้วยโปรแกรม Orange Canvas

จัดทำโดย

| | |
|-------------------------|----------------|
| นางสาวมุกมณี ลาวัลย์ | 116510907007-8 |
| นายสาววิยะดา นุ่นกระจาย | 116510907041-7 |
| นายธนภูมิ โชคสัมฤทธิ์ผล | 116510907030-0 |

เสนอ

อาจารย์พิเชฐ คุณากรวงศ์

อาจารย์ปองพล นิลพฤษ

รายงานนี้เป็นส่วนหนึ่งของการศึกษาวิชา Machine Learning และ AI for Bigdata
สาขาการวิเคราะห์และจัดการข้อมูลขนาดใหญ่ คณะวิทยาศาสตร์และเทคโนโลยี

ภาคเรียนที่ 2 ปีการศึกษา 2566

มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

คำนำ

รายงานเรื่อง การพยากรณ์การสอบผ่านวิชา IC3 โดยวิธีการเรียนรู้ของเครื่องด้วยโปรแกรม Orange Canvas เป็นส่วนหนึ่งของรายวิชา Machine Learning และวิชา Artificial Intelligence โดยมีจุดประสงค์เพื่อศึกษาความรู้เกี่ยวกับการใช้โปรแกรม Orange Canvas เพื่อการพยากรณ์โดยใช้โมเดลภายในโปรแกรมเข้ามาช่วย ซึ่งรายงานเล่มนี้มีเนื้อหาเกี่ยวกับการทำงานของโปรแกรมการใช้ model เพื่อวิเคราะห์ข้อมูลและพยากรณ์ผลลัพธ์และรายงานผล

โดยรายงานเล่มนี้พวกเราได้วางแผนการดำเนินงานการศึกษาค้นคว้าเป็นเนื้อหาเพื่อให้ง่ายต่อการศึกษาต่อหรือนำไปประยุกต์ใช้ในงานต่างๆต่อไปได้ ทางกลุ่มของเราต้องขอขอบคุณ อาจารย์ผู้ให้ความรู้ และแนวทางในการศึกษา วิจัยที่เกี่ยวข้องกับชิ้นงาน และสมาชิกในกลุ่มที่ช่วยในการปฏิบัติงานตลอดมา

จึงหวังว่ารายงานเล่มนี้จะเป็นประโยชน์แก่ผู้ที่ได้อ่านหรือศึกษาต่อหากผิดพลาดประการใด

ขออภัยมา ณ ที่นี้ด้วย

คณะผู้จัดทำ

สารบัญ

| | หน้า |
|----------------------------------------|------|
| คำนำ | ก |
| สารบัญ | ข-ค |
| สารบัญรูป | จ |
| สารบัญตาราง | ฉ |
| บทที่ | |
| บทที่ 1 บทนำ | |
| ที่มาและความสำคัญ | 1 |
| วัตถุประสงค์ | 1 |
| ขอบเขตการดำเนินงาน | 1 |
| ผลที่คาดว่าจะได้รับ | 1 |
| บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง | |
| การพยากรณ์(Forecasting) | 2-7 |
| Orange Data Mining | 7 |
| การวัดผลโมเดล | 8 |
| IC3 Certificate | 8 |
| บทที่ 3 วิธีการดำเนินงาน | |
| การจัดเตรียมข้อมูล | 9 |
| การทำ Data Mining | 9-11 |

สารบัญ(ต่อ)

| บทที่ | หน้า |
|-------------------------------------------|-------|
| บทที่ 4 ผลการดำเนินงาน | |
| การ Test Data | 12 |
| การพยากรณ์ในมิติอื่นๆ | 13-15 |
| บทที่ 5 สรุปอภิปรายผลและข้อเสนอแนะ | |
| สรุปผลการพยากรณ์ | 16 |
| บรรณานุกรม | 17 |

สารบัญรูป

| ภาพที่ | | หน้า |
|--------|-------------------------------------|------|
| 1 | kNN | 4 |
| 2 | RNNs | 5 |
| 3 | Naive Bayes | 5 |
| 4 | Trees | 6 |
| 5 | Train Data | 10 |
| 6 | Test Data | 12 |
| 7 | การส่งงานมีผลต่อการสอบผ่าน | 13 |
| 8 | โมเดลผลมีผลต่อการสอบผ่าน | 13 |
| 9 | เวลาที่ใช้ในการสอบมีผลต่อการสอบผ่าน | 14 |
| 10 | คณะที่สังกัดมีผลต่อการสอบผ่าน | 14 |
| 11 | เกรดเฉลี่ยมีผลต่อการสอบผ่าน | 15 |

สารบัญตาราง

| ตารางที่ | | หน้า |
|----------|-------------------------------|------|
| 1 | วิธีการดำเนินงาน | 9 |
| 2 | สังกัดคณะ | 9 |
| 3 | ตัวแปรอิสระ | 10 |
| 4 | Test and Score Train | 11 |
| 5 | Confusion Matrix Of kNN Train | 11 |
| 6 | Confusion Matrix Of kNN Test | 12 |
| 7 | Test and Score Test | 12 |

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญ

เป็นข้อมูลผลการสอบวิชา IC3 เพื่อให้นักศึกษาได้ใช้กระบวนการเรียนรู้ของเครื่อง (Machine Learning) ในการพยากรณ์เพื่อทำนายโอกาสในการสอบผ่านรายวิชา IC3

1.2 วัตถุประสงค์

1.2.1 เพื่อพยากรณ์หาตัวแปรที่มีอิทธิพลต่อการสอบผ่านวิชา IC3

1.2.2 เพื่อพยากรณ์หรือทำนายการสอบผ่านวิชา IC3 ในอนาคต

1.2.3 เพื่อรายงานการพยากรณ์การสอบผ่านวิชา IC3

1.3 ขอบเขตการดำเนินงาน

1.3.1 ศึกษาข้อมูลผลการสอบของนักศึกษาในวิชา IC3

1.3.2 เลือกข้อมูลที่ต้องการใช้พยากรณ์ และจัดรูปแบบของข้อมูล

1.3.3 สร้างโมเดลการพยากรณ์การสอบผ่านจากผลการสอบในรายวิชา IC3

1.4 ผลที่คาดว่าจะได้รับ

1.4.1 โมเดลที่สามารถทำนายการสอบผ่านวิชา IC3 ได้

1.4.2 โมเดลที่สามารถวิเคราะห์ปัจจัยการสอบผ่านได้

1.4.3 โมเดลที่มีค่าความเชื่อมั่นมากกว่าเท่ากับ 0.95%

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 การพยากรณ์(Forecasting)

การพยากรณ์ คือ การประมาณ หรือ การคาดคะเนว่าอะไรจะเกิดขึ้นในอนาคต ด้วยการยึดข้อมูล และข้อเท็จจริงจากอดีตเป็นหลักในการพยากรณ์ เหตุการณ์ที่เกิดขึ้นในอดีตกับเหตุการณ์อย่างเดียวกันที่เกิดขึ้นอีกในปัจจุบัน จะส่งผลให้ได้ผลลัพธ์ที่คล้ายกับที่เกิดขึ้นมาแล้ว หรือกล่าวได้ว่า เป็นการใช้ข้อมูลจากข้อเท็จจริงมาสนับสนุน และใช้ความรู้ทางสถิติ คณิตศาสตร์ ตัวเลขทางสถิติ ผลลัพธ์จากการวิจัย เข้ามาช่วยในการวิเคราะห์ข้อมูล ทั้งนี้ การพยากรณ์ (Forecasting) อาจตรงหรือคลาดเคลื่อนจากสิ่งที่เกิดขึ้นจริง การพยากรณ์ที่ดีจึงควรจะต้องใช้ข้อมูลที่ใกล้เคียงกับความเป็นจริงให้มากที่สุด โดยทั่วไปจึงมักพบเห็นการใช้กระบวนการ Forecasting ในการพยากรณ์ภูมิอากาศ และในด้านธุรกิจ เช่น การพยากรณ์ยอดขาย การคาดการณ์แนวโน้มของตลาดผู้บริโภค เป็นต้น

การพยากรณ์ Forecasting สามารถแบ่งเป็นช่วงระยะเวลาในการพยากรณ์ ซึ่งครอบคลุมระยะเวลาของการพยากรณ์ขึ้นอยู่กับวัตถุประสงค์ของแต่ละองค์กร โดยทั่วไปสามารถแบ่งออกเป็น 3 ระยะ

1. การพยากรณ์ระยะสั้น วันหรือเดือน มุ่งเน้นไปที่เหตุการณ์หรือสถานการณ์ที่มีผลต่อธุรกิจในอนาคตสั้น ๆ เช่น การพยากรณ์ยอดขายเดือนถัดไป หรือพยากรณ์สภาพอากาศในอนาคตไม่กี่วันข้างหน้า
2. การพยากรณ์ระยะปานกลาง ระยะ 1-2 ปีขึ้นกับฤดูกาล เน้นไปที่ยุคหรือช่วงเวลาที่ยาวขึ้น เช่น ระยะเวลาปีหรือหลายปี การพยากรณ์ในระยะนี้อาจเน้นทางทัศนคติของตลาด เศรษฐกิจ หรือการพยากรณ์แนวโน้มของธุรกิจตลาดในระยะยาว
3. การพยากรณ์ระยะยาว 10-20 ปีขึ้นกับนโยบาย การพยากรณ์ในระยะยาวมักมีการดำเนินการที่เน้นไปที่การแบ่งปันความรู้ทางกายภาพหรือทางการวิจัยที่มีผลเป็นเวลานาน การพยากรณ์ระยะนี้อาจเป็นเช่นการพยากรณ์แนวโน้มของเทคโนโลยี การเปลี่ยนแปลงในกลุ่มประชากร หรือการพยากรณ์ผลกระทบของการเปลี่ยนแปลงสภาพโลกในระยะยาว

2.1.1 วิธีการพยากรณ์

วิธีการพยากรณ์ยังจะแบ่งออกเป็น 2 ประเภทใหญ่ๆ คือ วิธีการพยากรณ์เชิงปริมาณ และวิธีการพยากรณ์เชิงคุณภาพ มีรายละเอียดดังนี้

2.1.1.1 การพยากรณ์เชิงคุณภาพ มีจุดเด่นโดยการใช้ดุลพินิจของผู้มีประสบการณ์ใน การให้นำหนักในการพยากรณ์ โดยมีวิธีย่อย 5 วิธี คือ

1. วิธีเดลฟาย (Delphi Method) เป็นวิธีการพยากรณ์โดยใช้กลุ่มของผู้เชี่ยวชาญ มาวิเคราะห์และพยากรณ์ร่วมกัน โดยจะมีหน่วยงานที่ทำหน้าที่หลักดำเนินการโดยสร้าง แบบสอบถามอย่างต่อเนื่อง ถามไปเรื่อย ๆ กับผู้เชี่ยวชาญ เพื่อให้เกิดผลดีและลดการเกิดอิทธิพลจาก ผู้เชี่ยวชาญ คนหนึ่งจะส่งผลต่อความคิดกับผู้เชี่ยวชาญคนอื่น

2. วิธีวิจัยตลาด (Market Research) เป็นวิธีที่รวมเทคนิคการพยากรณ์เชิง ปริมาณมาใช้ ประโยชน์ โดยข้อมูลจะได้จากแบบสอบถาม การสำรวจทางโทรศัพท์ การสัมภาษณ์ และ การอภิปรายกลุ่ม แล้วนำข้อมูลที่ได้มาทดสอบสมมติฐานด้านการตลาด โดยวิธี Panel Consensus เป็นวิธีการที่ใช้ผู้เชี่ยวชาญมาอภิปรายกลุ่มเกี่ยวกับปัญหาต่างๆ เพื่อให้ได้มาซึ่งข้อสรุปที่เป็นค่าในการพยากรณ์

3. วิธี Grass-Roots Forecasting เป็นการพยากรณ์ที่เจาะจงขอบเขตเฉพาะ ส่วน โดยการสอบถามบุคคลที่ใกล้ชิดปัญหา เพื่อพยากรณ์ในขอบเขตที่รับผิดชอบ แล้วนำค่าแต่ละคน มาพยากรณ์รวมกัน

4. วิธีการพยากรณ์โดยยึดอดีตเป็นหลัก (Historical Analogy) เป็นการ ใช้ ข้อมูลของ เหตุการณ์หนึ่งในอดีต มาใช้พยากรณ์เหตุการณ์คล้ายกันที่เกิดขึ้น

2.1.1.2 การพยากรณ์เชิงปริมาณ เน้นการใช้ข้อมูลตัวเลขและเทคนิคทางคณิตศาสตร์เพื่อทำนาย ผลลัพธ์ในอนาคต โดยใช้แบบจำลองทางคณิตศาสตร์หรือวิธีการทางสถิติ

1. รูปแบบเชิงปัจจัยสาเหตุ (Associative models) เป็นวิธีการทางสถิติที่ใช้ในการหาความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระกับตัวแปรตาม โดยในทางธุรกิจการค้าอาจใช้เพื่อทำนาย ยอดขายของสินค้าตามตัวแปรต่าง ๆ เช่น การโฆษณาส่งผลต่อยอดขายหรือไม่ แคมเปญการตลาดนี้เหมาะสมในช่วงเวลานี้หรือไม่

2. อนุกรมเวลา (Time series) เป็นการใช้ข้อมูลที่ได้มาจากชุดข้อมูลตามเวลา เพื่อทำนาย แนวโน้มในอนาคต ตัวอย่างเช่น การพยากรณ์ยอดขายของสินค้าในแต่ละเดือนหรือการพยากรณ์ ราคาหุ้นในตลาดหุ้นโดยใช้ข้อมูลจากเวลาในปัจจุบันและอดีต

2.1.2 การเรียนรู้ของเครื่อง (Machine learning)

เป็นการสร้าง ตัวแบบเพื่อเป็นส่วนในการประมวลผลในการคำนวณเพื่อตั้งเงื่อนไขในการตัดสินใจ ต่างๆ โดยผู้พัฒนาจะนำชุดข้อมูลที่เป็นทั้งส่วนของข้อมูลดิบ (Raw Data) และผลลัพธ์(Output) เข้าไปให้ คอมพิวเตอร์เรียนรู้ เพื่อสร้างตัวแบบซึ่งจะเป็นเหมือนชุดความคิดหลักของโปรแกรม การใส่ชุดข้อมูลใน

เรียนรู้ของเครื่อง จะถือเป็นการสอน (Train) เพื่อให้โปรแกรมเรียนรู้และแยกแยะข้อมูลและสร้าง ตัวแบบ ขึ้นต้นก่อน จากนั้นนำข้อมูลอีกชุดมาทดสอบว่าตัวแบบที่โปรแกรมได้ทำมาได้ค่าคลาดเคลื่อนอยู่ในค่าที่ ยอมรับได้หรือไม่ จากนั้นอาจจะปรับเปลี่ยนบางส่วนของโปรแกรมเพื่อให้ได้ความคลาดเคลื่อนต่ำลง

2.1.2.1 ประเภทของการเรียนรู้ของเครื่อง

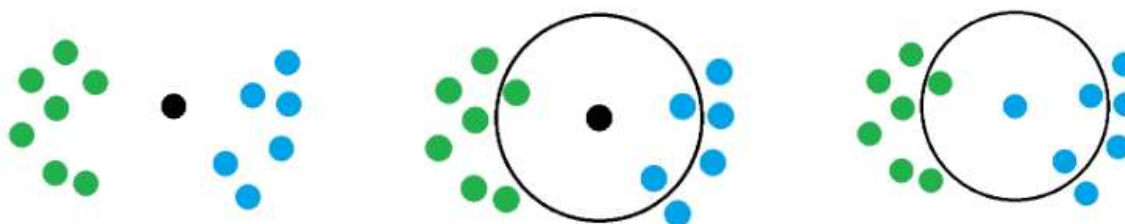
1. การเรียนรู้แบบมีการสอน (Supervised Learning) เป็นการเรียนรู้แบบต้องมีการสอน หรือให้ ข้อมูลเข้าไปสอน (Training Data) เมื่อผ่านการเรียนรู้แล้วจึงจะสามารถ Regression หรือ Classification) ออกมาได้

2. การเรียนรู้แบบไม่มีการสอน (Unsupervised Learning) เป็นการเรียนรู้ที่ไม่จำเป็นต้องใช้ผลลัพธ์ (Output) ในการทำนาย กล่าวคือป้อนเฉพาะข้อมูลที่จะทำนาย จากนั้นระบบจะทำการประมวลผลให้เอง ซึ่ง จะแบ่งออกเป็น การจัดกลุ่ม (Clustering) และ การหาความสัมพันธ์ (Association)

3. การเรียนรู้ด้วยการป้อนกลับผลลัพธ์ (Reinforcement Learning) เป็นการเรียนรู้ที่ให้ระบบ เรียนรู้แล้วป้อนกลับข้อมูลเพื่อให้ระบบสามารถปรับปรุงตัวเอง เช่น ระบบการทรงตัวของหุ่นยนต์

2.1.2.2 k-Nearest Neighbor (kNN)

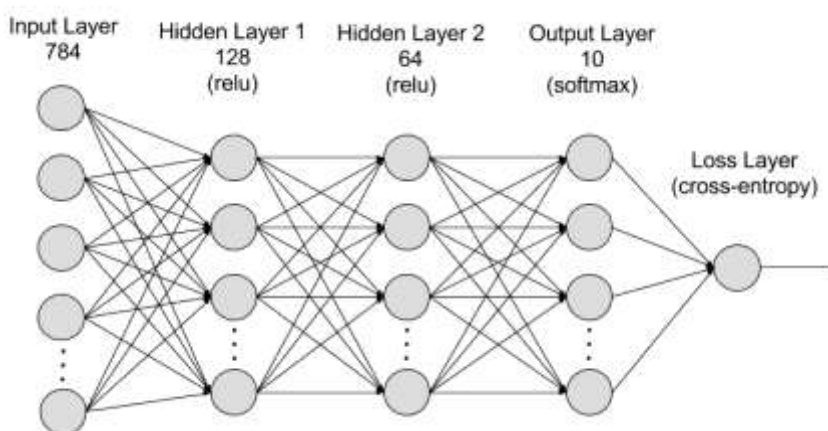
เป็นตัวแบบแบ่งกลุ่มของข้อมูล โดยนำข้อมูลที่ใกล้เคียงกันมาจัดกลุ่มด้วยกัน โดยมีหลักการนำ ข้อมูลอื่นๆมาเปรียบเทียบกับตัวข้อมูลที่สนใจ ว่ามีความใกล้เคียงกันมากแค่ไหน หากข้อมูลที่สนใจอยู่ใกล้ กับข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ที่สุด ใช้สำหรับแก้ปัญหาที่รู้ จำนวนกลุ่มที่แน่นอนอยู่แล้ว แต่มีข้อมูลบางตัวที่ไม่สามารถบอกได้ว่าข้อมูลนั้นอยู่กลุ่มไหน สามารถใช้ ระบบนี้เข้ามาช่วยเลือกกลุ่ม ซึ่งจะคล้าย ๆ กับการโหวตเสียงข้างมาก



รูป 1kNN

2.1.2.3 Neural Networks (RNNs)

นิวรัลเน็ตเวิร์กแบบวนกลับ (RNN) เป็นโมเดลดีปเลิร์นนิงที่ได้รับการฝึกฝนเพื่อประมวลผลและแปลงข้อมูลอินพุตเป็นเอาต์พุตข้อมูลตามลำดับเฉพาะ ข้อมูลตามลำดับคือข้อมูล เช่น คำ ประโยค หรือข้อมูลแบบอนุกรมเวลา ซึ่งส่วนประกอบตามลำดับจะสัมพันธ์กันตามความหมายที่ซับซ้อนและกฎไวยากรณ์ RNN เป็นระบบซอฟต์แวร์ที่ประกอบด้วยส่วนประกอบที่เชื่อมต่อกันมากมายที่เลียนแบบวิธีที่มนุษย์แปลงข้อมูลตามลำดับ เช่น การแปลข้อความจากภาษาหนึ่งเป็นอีกภาษาหนึ่ง RNN ส่วนใหญ่ถูกแทนที่ด้วยปัญญาประดิษฐ์ (AI) และโมเดลภาษาขนาดใหญ่ (LLM) ซึ่งมีประสิทธิภาพในการประมวลผลข้อมูลมากขึ้นตามลำดับ



รูป 2RNNs

2.1.2.4 Naive Bayes

Naive Bayes เป็นหนึ่งใน Classification Model ใช้ในการแบ่งกลุ่มหรือหาเหตุการณ์ที่จะเกิดขึ้นโดยการอิงทฤษฎีความน่าจะเป็นของ Bayes หรือ Bayesian ซึ่ง Target ของโมเดลจะมีความคล้ายคลึงกับ Logistic Regressionว่าจะเกิดเหตุการณ์นั้นหรือไม่โดยจะเพิ่มโอกาสในการเกิดเหตุการณ์เข้าไปด้วย โดยมักใช้ในการวิเคราะห์ข้อมูลที่มีความต่อเนื่องของเหตุการณ์ (Dependent Event) เช่น โอกาสในการเกิดโรคในกลุ่มประชากรที่เราสนใจ ซึ่งจำเป็นจะต้องอาศัยการคำนวณผ่านสูตรดังนี้

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

รูป 3Naive Bayes

และกำหนดให้ $P(A|B)$ คือความน่าจะเป็นในการเกิดเหตุการณ์ A โดยมี B เป็น Condition

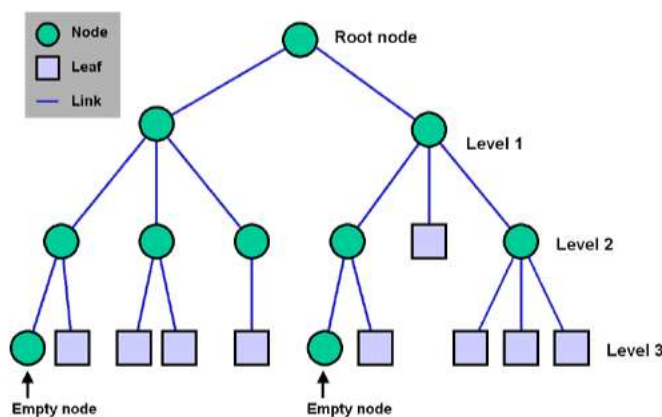
$P(B|A)$ คือความน่าจะเป็นในการเกิดเหตุการณ์ B โดยมี A เป็น Condition

$P(A)$ คือโอกาสในการเกิดเหตุการณ์ A จากเหตุการณ์ทั้งหมด

$P(B)$ คือโอกาสในการเกิดเหตุการณ์ B จากเหตุการณ์ทั้งหมด

2.1.2.5 Decision Trees

Decision Tree หรือต้นไม้ตัดสินใจเป็น Classification Model ที่ใช้กระบวนการ Rule-Based คือเป็นการสร้างกฎ If-Else ขึ้นมา อธิบายให้เข้าใจง่ายคือเป็นวิธีการแยกข้อมูลถ้าข้อมูลเข้ากฎที่ตั้งขึ้นก็จะแยกไปอยู่กลุ่มหนึ่ง ถ้าไม่เข้ากฎก็จะแยกไปอยู่อีกกลุ่มหนึ่ง โดย Decision Tree Model ซึ่งการทำ Decision Tree จะมีส่วนประกอบหลักได้ Root Node คือชุดข้อมูลและกฎเกณฑ์ที่ตั้งต้นเมื่อทำการวิเคราะห์ไปเรื่อย ๆ จนถึงกลุ่มสุดท้ายในการแบ่ง จะเรียกกลุ่มเหล่านั้นว่า Leaf Node



รูป 4 Trees

2.1.3 Artificial Intelligence (AI)

2.1.3.1 ปัญญาประดิษฐ์ (AI) หมายถึง ความสามารถทางปัญญาของเครื่องจักร ที่ถูกออกแบบมาให้เลียนแบบความสามารถทางปัญญาของมนุษย์

2.1.3.2 คุณสมบัติหลักของ (AI) การเรียนรู้เรียนรู้จากข้อมูลและประสบการณ์การคิดวิเคราะห์ข้อมูลและหาความสัมพันธ์การแก้ปัญหา หาทางออกที่ดีที่สุดสำหรับปัญหาการตัดสินใจตัดสินใจอย่างมีประสิทธิภาพปรับตัวเข้ากับสถานการณ์ใหม่

ตัวอย่างการประยุกต์ใช้ AIระบบแนะนำสินค้า แนะนำสินค้าที่ตรงกับความสนใจของลูกค้าการจดจำใบหน้า: ระบุบุคคลจากภาพถ่ายหรือวิดีโอการแปลภาษา แปลภาษาโดยอัตโนมัติการขับขี้อัตโนมัติ ความคุ้มครองโดยไม่ต้องมีคนขับการวินิจฉัยโรค วิเคราะห์ภาพถ่ายทางการแพทย์เพื่อวินิจฉัยโรค

2.1.3.3 ประเภทของ Machine Learning สามารถเรียนรู้จากข้อมูลโดยไม่ต้องเขียนโปรแกรมDeep Learning ประเภทของ Machine Learning ที่ใช้โมเดลประสาทเทียมNatural Language Processing(NLP) เข้าใจและประมวลผลภาษาธรรมชาติComputer Vision รับรู้และวิเคราะห์ภาพความแตกต่างระหว่าง

2.2 Orange Data Mining

Orange Data Mining เป็นซอฟต์แวร์สำหรับการวิเคราะห์ข้อมูล การเรียนรู้ของเครื่อง และการนำเสนอภาพข้อมูล โมเดลของ Orange Data Miningไม่ได้สร้างโมเดลของตัวเองโดยตรง แต่เป็นเครื่องมือที่ช่วยสร้างและวิเคราะห์โมเดล Machine Learning

Orange Data Mining มี Widget สำหรับรองรับการทำงานกับโมเดล Machine Learning หลากหลายประเภท

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)การจำแนกประเภท
2. Classificationแยกข้อมูลออกเป็นกลุ่ม เช่น จำแนกประเภทของสแปมอีเมลการถดถอย
3. Regressionทำนายค่าต่อเนื่อง เช่น คาดการณ์ราคาบ้านการเรียนรู้แบบไม่มีผู้สอน
4. Unsupervised Learningการจัดกลุ่มข้อมูล
5. Clustering จัดกลุ่มข้อมูลที่มีลักษณะคล้ายคลึงกันการลดมิติ
6. Dimensionality Reduction ลดจำนวนมิติของข้อมูลเพื่อการวิเคราะห์ที่ง่ายขึ้น

2.3 การวัดผลโมเดล

1. Accuracy (AUC) คือ ผลจากการ Predict ถูกต้อง จากข้อมูลทั้งหมดที่นำมา Predict คำนวณได้จากสูตร $Accuracy = (TP+TN)/(TP+TN+FP+FN)$ โดยค่า accuracy จะมีค่าอยู่ระหว่าง 0-1 ยิ่งเข้าใกล้ 1 แปลว่าโมเดลเราทำนายผลได้ดีมาก

2. Sensitivity หรือ Recall คือค่าที่โมเดลทายเป็นคลาสที่กำลังพิจารณาถูก/ค่าเหตุการณ์จริงเป็นคลาสที่กำลังพิจารณาทั้งถูกและผิด หาค่าได้จากสูตร $Sensitivity/Recall = TP/(TP+FN)$

3. Classification accuracy(CA) คือสัดส่วนของจำนวนความถูกต้องในการ Predict ต่อจำนวนที่นำมา Predict ทั้งหมด

4. F1-Score เป็นค่าที่ได้จากการเอาค่า precision และ recall มาคำนวณรวมกัน (F1 สร้างขึ้นมาเพื่อเป็น single metric ที่วัดความสามารถของโมเดล ไม่ต้องเลือกระหว่าง precision, recall เพราะเฉลี่ยให้แล้ว) หาค่าได้จากสูตร $F1 = 2 * [(precision * recall) / (precision + recall)]$

5. Precision(Prec) คือค่าความเที่ยง เป็นค่าที่จะบ่งบอกว่า การทำซ้ำ มีความสเถียรมากน้อยแค่ไหน

6. Matthews Correlation Coefficient คือสัมประสิทธิ์สหสัมพันธ์แมทธีวส์หรือค่าสัมประสิทธิ์ใช้ในการเรียนรู้เพื่อวัดคุณภาพของแบบจำลองในการจำแนกประเภทไบนารี ค่าสัมประสิทธิ์ +1 แสดงถึงการทำนายที่สมบูรณ์แบบ 0 ไม่ดีไปกว่าการทำนายสุ่มและ -1 บ่งชี้ว่าแบบจำลองไม่ดีพอที่จะจำแนกได้ทั้งหมด

2.4 IC3 Certificate

The Internet and Computing Core (IC3) Certificate คือ ประกาศนียบัตรที่รับรองความรู้และทักษะคอมพิวเตอร์ขั้นพื้นฐาน และเป็นประกาศนียบัตรที่ถูกกำหนดเป็นมาตรฐานเป็น Neutral Vendor Standard โดย Global Digital Literacy Council และรองรับโดยกระทรวงไอซีที กระทรวงวัฒนธรรม และสมาคมคอมพิวเตอร์แห่งประเทศไทย ซึ่งประกอบด้วย 3 Modules

1. Computing Fundamentals 2. Key Applications 3. Living Online

IC3 Certificate สามารถใช้สอนได้กับคนที่ไม่มีความรู้ทางด้านการใช้งานคอมพิวเตอร์เลย หรือสามารถเลือก module ใช้ใน แต่ละส่วนโดยเน้นองค์ความรู้ตามหลักสูตรมาใช้ในการเรียนการสอนได้ตามสมควรต่อกลุ่มเป้าหมาย

บทที่ 3

วิธีการดำเนินงาน

3.1 วิธีการดำเนินงาน

| ลำดับ | รายการ | พ.ย. | ธ.ค. | ม.ค. | ก.พ. | มี.ค. |
|-------|--------------------------|------|------|------|------|-------|
| 1 | ศึกษาและเก็บรวบรวมข้อมูล | ←→ | ←→ | | | |
| 2 | วิเคราะห์ข้อมูล | | ←→ | ←→ | | |
| 3 | ศึกษาโมเดลที่จะใช้ | ←→ | | | ←→ | |
| 4 | ทดสอบและแก้ไข | | | | ←→ | ←→ |
| 5 | สรุปผล | | | | | ←→ |

ตาราง 1 วิธีการดำเนินงาน

3.2 การจัดเตรียมข้อมูล

ข้อมูลที่ใช้ในการพยากรณ์เป็นข้อมูลของนักศึกษาในมหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี ที่สอบมาตรฐาน IC3 จำนวน 15,201 คน ประกอบไปด้วยคณะดังนี้

| ลำดับ | คณะ | จำนวน(คน) |
|-------|-------------------------|-----------|
| 1 | เทคโนโลยีการเกษตร | 940 |
| 2 | เทคโนโลยีคหกรรมศาสตร์ | 753 |
| 3 | เทคโนโลยีสื่อสารมวลชน | 1379 |
| 4 | การแพทย์บูรณาการ | 154 |
| 5 | ครุศาสตร์อุตสาหกรรม | 1498 |
| 6 | บริหารธุรกิจ | 4633 |
| 7 | พยาบาลศาสตร์ | 185 |
| 8 | วิทยาลัยการแพทย์แผนไทย | 253 |
| 9 | วิทยาศาสตร์และเทคโนโลยี | 730 |
| 10 | วิศวกรรมศาสตร์ | 2957 |
| 11 | ศิลปกรรมศาสตร์ | 641 |
| 12 | ศิลปศาสตร์ | 1024 |
| 13 | สถาปัตยกรรมศาสตร์ | 54 |

ตาราง 2 สังกัดคณะ

3.2.1 การหาตัวแปรอิสระ ผู้จัดทำเลือกที่จะใช้ตัวแปรอิสระที่อาจจะส่งผลต่อการสอบผ่าน ซึ่งมีรายละเอียดดังต่อไปนี้

| ลำดับ | รายการ | ประเภท | คำอธิบาย | ขอบเขต |
|-------|----------------------------------------|---------|----------------------|-------------------------------------------------------------------------------------------|
| 1 | IC3_MODULE_NAME | Text | ชื่อโมดูล | IC3 GS5 - Computing Fundamentals IC3 GS5 - Key Applications IC3 GS5 - Living Online |
| 2 | IC3_SCORE | Integer | คะแนนสอบ | 0 ถึง 1,000 คะแนน |
| 3 | IC3_RESULT | Text | ผลการสอบ | Fail/ Pass |
| 4 | STD_CURRENT_GPA | Float | เกรดเฉลี่ยปัจจุบัน | 0.0 ถึง 4.0 |
| 5 | STD_FACULTYNAME_THAI | Text | คณะ | 14 คณะ |
| 6 | ONLINE_ASSIGNMENT_SUBMISSION_FREQUENCY | Text | การส่งงานของนักศึกษา | Hight,Medium,Low,Lowest |
| 7 | IC3_EXAM_TIMEUSED | Integer | เวลาที่ใช้ในการสอบ | 0 ถึง 50 นาที |

ตาราง 3 ตัวแปรอิสระ

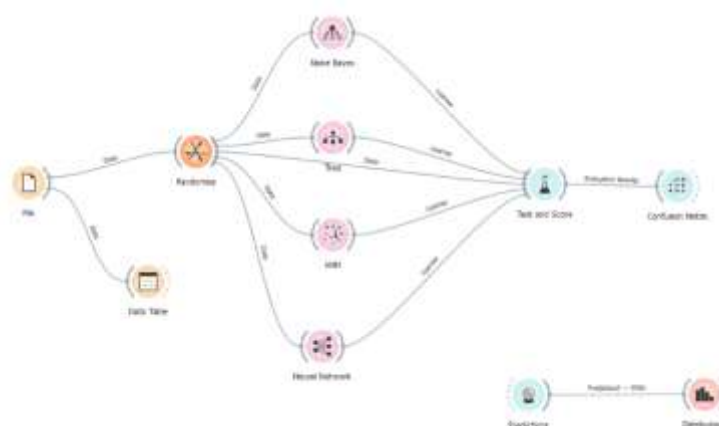
3.2.2 การแบ่งข้อมูล ข้อมูลทั้งหมดมีจำนวน 45,603 ข้อมูล ข้อมูลจะถูกแบ่งเป็น 3 ชุด คือ Training Set ,Testing Set และ Randomize

1. Training Set มีจำนวนข้อมูลทั้งสิ้น 27,362 ข้อมูล คิดเป็น 60%
2. Testing Set มีจำนวนข้อมูลทั้งหมด 13,681 ข้อมูล คิดเป็น 30%
3. Randomize มีจำนวนข้อมูลทั้งหมด 4,560 ข้อมูล คิดเป็น 10%

3.3 การทำ Data Mining

3.3.1 การ Train Data

ในการทำ Data Mining เลือกใช้โมเดลในการเทรน 4 แบบ ได้แก่ Naïve Bayes , Tree , kNN และ Neural Network เพื่อทดสอบประสิทธิภาพของโมเดล เป็นจำนวน 2 ครั้ง



รูป 5 Train Data

ได้ผล Test and Score ดังนี้

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|----------------|-------|-------|-------|-------|--------|-------|
| Tree | 0.942 | 0.957 | 0.957 | 0.957 | 0.957 | 0.904 |
| kNN | 0.960 | 0.955 | 0.955 | 0.955 | 0.955 | 0.901 |
| Naïve Bayes | 0.947 | 0.881 | 0.880 | 0.880 | 0.881 | 0.734 |
| Neural Network | 0.954 | 0.906 | 0.904 | 0.907 | 0.906 | 0.790 |

ตาราง 4 Test and Score Train

จากตารางจะเห็นได้ว่า โมเดล kNN มีค่าประสิทธิภาพของโมเดลสูงที่สุดเป็น 0.960 มีค่าความถูกต้องโมเดลในการจำแนกประเภทเป็น 0.955 รองลงมาเป็น Neural Network Naïve Bayes Tree มีค่าเป็น 0.954 0.947 และ 0.942 ตามลำดับ จึงเลือกใช้โมเดล kNN ในการพยากรณ์

Confusion Matrix Of kNN

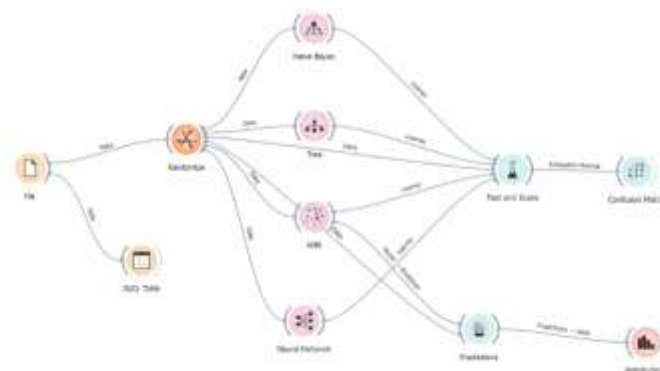
| | | Predicted | | |
|--------|------|-----------|------|-------|
| | | Fail | Pass | Sum |
| Actual | Fail | 17404 | 566 | 17920 |
| | Pass | 652 | 8740 | 9392 |
| | Sum | 18956 | 9306 | 27362 |

ตาราง 5 Confusion Matrix Of kNN Train

บทที่ 4

ผลการดำเนินงาน

4.1 การ Test Data



รูป 6 Test Data

4.1.1 ผลการพยากรณ์ด้วย kNN

| | | Predicted | | |
|--------|------|-----------|------|-------|
| | | Fail | Pass | Sum |
| Actual | Fail | 8999 | 5 | 9004 |
| | Pass | 5 | 4672 | 4677 |
| | Sum | 9004 | 4677 | 13681 |

ตาราง 6 Confusion Matrix Of kNN Test

1. True Positives (TP) = 4672
2. True Negatives (TN) = 8999
3. False Positives (FP) = 5
4. False Negatives (FN) = 5

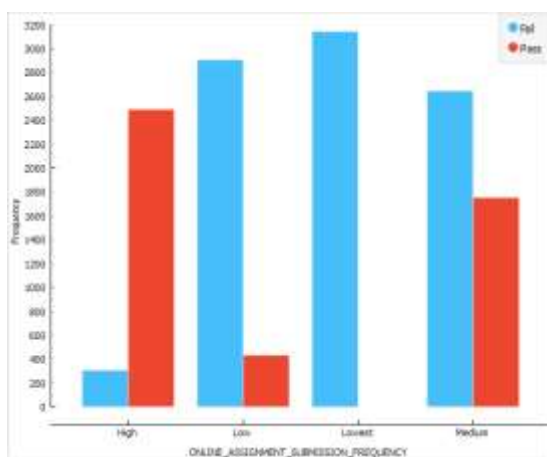
ได้ผล Test and Score ดังนี้

| Model | AUC | CA | F1 | Prec | Recall | MCC |
|-------|-------|-------|-------|-------|--------|-------|
| kNN | 1.000 | 0.999 | 0.999 | 0.999 | 0.999 | 0.998 |

ตาราง 7 Test and Score Test

4.2 การพยากรณ์ในมิติอื่น ๆ

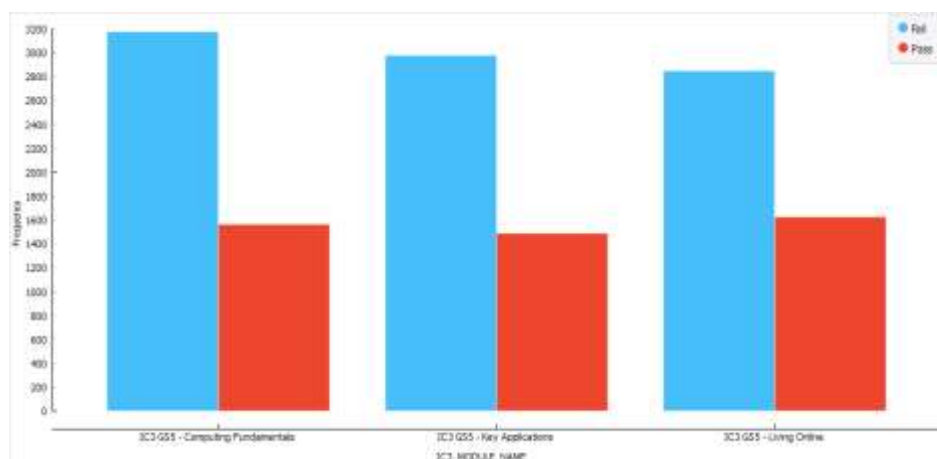
4.2.1 การพยากรณ์เมื่อการส่งงานมีผลต่อการสอบผ่าน IC3



รูป 7 การส่งงานมีผลต่อการสอบผ่าน

ผลการพยากรณ์คือ นักศึกษาที่มีความถี่การส่งงานสูง(Hight) จะมีโอกาสสอบผ่าน IC3 2,488 คน สอบตก 310 คน นักศึกษาที่มีความถี่การส่งงานปานกลาง(Medium) จะมีโอกาสสอบผ่าน IC3 1,753 คน สอบตก 2,646 คน นักศึกษาที่มีความถี่การส่งงานน้อย(Low) จะมีโอกาสสอบผ่าน IC3 435 คน สอบตก 2,905 คน นักศึกษาที่มีความถี่การส่งงานน้อยมาก(Lowest) จะมีโอกาสสอบผ่าน IC3 1 คน สอบตก 3,143 คน

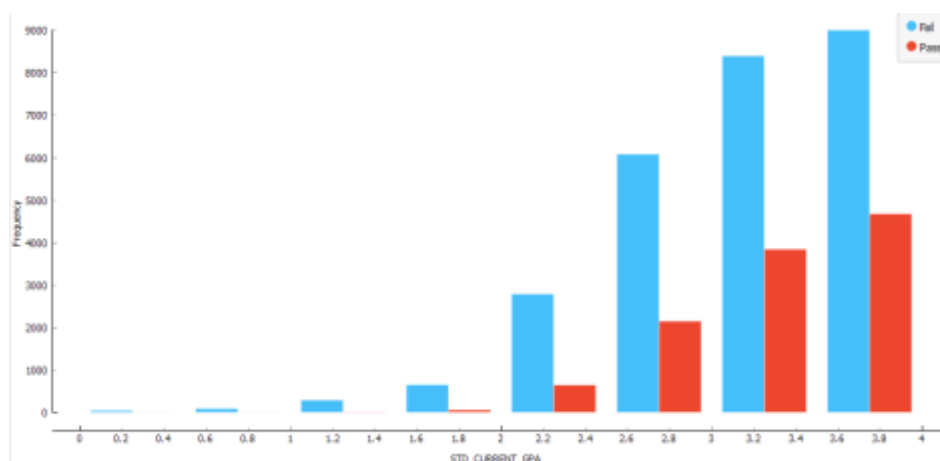
4.2.2 การพยากรณ์เมื่อโมดูลผลมีผลต่อการสอบผ่าน IC3



รูป 8 โมดูลผลมีผลต่อการสอบผ่าน

ผลการพยากรณ์คือ นักศึกษาที่สังกัดคณะบริหารธุรกิจ สอบผ่าน 1,123 คน สอบตก 3,007 คน คณะวิศวกรรมศาสตร์ สอบผ่าน 1,444 คน สอบตก 1,235 คน คณะบริหารธุรกิจ สอบผ่าน 1,123 คน สอบตก 3,007 คน คณะครุศาสตร์อุตสาหกรรม สอบผ่าน 480 คน สอบตก 888 คน คณะเทคโนโลยีสื่อสารมวลชน สอบผ่าน 304 คน สอบตก 946 คน คณะศิลปศาสตร์ สอบผ่าน 286 คน สอบตก 637 คน คณะเทคโนโลยีการเกษตร สอบผ่าน 131 คน สอบตก 723 คน คณะเทคโนโลยีคหกรรมศาสตร์ สอบผ่าน 135 คน สอบตก 560 คน คณะวิทยาศาสตร์และเทคโนโลยี สอบผ่าน 272 คน สอบตก 361 คน คณะศิลปกรรมศาสตร์ สอบผ่าน 81 คน สอบตก 508 คน วิทยาลัยการแพทย์แผนไทย สอบผ่าน 126 คน สอบตก 95 คน คณะพยาบาลศาสตร์ สอบผ่าน 141 คน สอบตก 15 คน คณะการแพทย์บูรณาการ สอบผ่าน 114 คน สอบตก 15 คน และคณะสถาปัตยกรรมศาสตร์ สอบผ่าน 40 คน สอบตก 14 คน

4.2.5 การพยากรณ์เมื่อเกรดเฉลี่ยมีผลต่อการสอบผ่าน IC3



รูป 11 เกรดเฉลี่ยมีผลต่อการสอบผ่าน

ผลการพยากรณ์คือ นักศึกษาที่มีเกรดเฉลี่ยน้อยกว่า 0.5 สอบผ่าน 3 คน สอบตก 45 คน น้อยกว่า 1.0 สอบผ่าน 4 คน สอบตก 90 คน น้อยกว่า 1.5 สอบผ่าน 9 คน สอบตก 292 คน น้อยกว่า 2.0 สอบผ่าน 60 คน สอบตก 653 คน น้อยกว่า 2.5 สอบผ่าน 647 คน สอบตก 2,790 คน น้อยกว่า 3.0 สอบผ่าน 2,149 คน สอบตก 6,081 คน น้อยกว่า 3.5 สอบผ่าน 3,843 คน สอบตก 8,394 คน และน้อยกว่าเท่ากับ 4.0 สอบผ่าน 4,677 คน สอบตก 9,004 คน

บทที่ 5

สรุปอภิปรายผลและข้อเสนอแนะ

5.1 สรุปผลการพยากรณ์

การพยากรณ์การสอบผ่านวิชา IC3 จากผู้เข้าสอบ 13,681 คน มีผู้สอบผ่าน 4,672 คน และสอบไม่ผ่าน 8,999 คน ทำนายผิดว่าสอบผ่าน 5 คน และทำนายผิดว่าสอบไม่ผ่าน 5 คน

ปัจจัยที่ส่งผลต่อการสอบผ่านวิชา IC3 ของนักศึกษาได้แก่ การส่งงานเป็นประจำ(Hight) เวลาที่ใช้ในการทำข้อสอบมากกว่าเท่ากับ 45 นาที และศึกษาอยู่สังกัดคณะวิศวกรรมศาสตร์ มีโอกาสในการสอบผ่านวิชา IC3 ปัจจัยที่ไม่ส่งผลต่อการสอบผ่านวิชา IC3 ได้แก่ โมดุลในการสอบ และเกรดเฉลี่ย

บรรณานุกรม

KongRuksiam Studio. (19 มีนาคม 2567). *Machine Learning*(EP.5). เข้าถึงได้จาก <https://shorturl.asia/3fqVB>

OLARIK SURINTA. (19 มีนาคม 2567). *ORANGE*. เข้าถึงได้จาก http://olarik.it.msu.ac.th/?page_id=412

Porntiva Visitsora. (20 มีนาคม 2567). *Metrics* พื้นฐานสำหรับวัดประสิทธิภาพของโมเดล *Machine Learning*. เข้าถึงได้จาก <https://citly.me/ZXbqG>

Sirawich Smitsomboon. (19 มีนาคม 2567). *สรุปความเข้าใจ RNN, LSTM, GRU (24/10/2020)*. เข้าถึงได้จาก <https://shorturl.asia/ubU7V>

training.rmutt. (21 มีนาคม 2567). *ทักษะการใช้คอมพิวเตอร์และเทคโนโลยีสารสนเทศด้วยเกมมิฟิเคชัน*. เข้าถึงได้จาก <https://ic3.rmutt.ac.th/>

ภิรัชญญา. (19 มีนาคม 2567). *การพยากรณ์ (Forecasting)*. เข้าถึงได้จาก https://ssc-monitor.com/content.php?n_menu5=1278

สถาบันนวัตกรรมและกรรมาภิบาลข้อมูล. (19 มีนาคม 2567). *เข้าใจใน 5 นาที! Classification Model คืออะไร*. เข้าถึงได้จาก <https://digi.data.go.th/blog/what-is-classification-model/>