

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

ปัจจุบันอินเทอร์เน็ตและระบบสารสนเทศเข้ามามีบทบาทสำคัญอย่างมากในชีวิตประจำวันของมนุษย์มาก ความเจริญก้าวหน้าของเทคโนโลยีในด้านต่าง ๆ ได้ถูกพัฒนาไปอย่างรวดเร็วและ เทคโนโลยีสารสนเทศได้ถูกนำมาประยุกต์ใช้ภายในองค์กร อาทิเช่น ด้านการติดต่อสื่อสาร ด้านอำนวยความสะดวกด้านการศึกษา ด้านการช่วยการตัดสินใจ ด้านการพยากรณ์ ด้านนำเสนอข้อมูลหรือด้านความรู้ ประเทศไทยนั้นได้มีพัฒนาความเจริญก้าวหน้าทางด้านเทคโนโลยีและข้อมูลมาอย่างต่อเนื่องจะเห็นได้ว่าเทคโนโลยีและข้อมูลนั้นมีบทบาทสำคัญในการดำเนินงานให้ความรู้หรือช่วยเหลือประชาชนมากขึ้นซึ่งเกิดจากการพัฒนาเทคโนโลยีและข้อมูลที่รวบรวมสร้างเป็นสารสนเทศให้เกิดประโยชน์ ยิ่งในปัจจุบันจะเห็นได้ว่า เทคโนโลยีมีความสำคัญมากในหลายด้าน แต่ที่เห็นได้ชัดคือ ด้านการศึกษา เนื่องด้วยเกิดโรคระบาด COVID-19 ทำให้เทคโนโลยีมีบทบาทสำคัญมากต่อการศึกษา แต่จะมีเรื่องของค่าใช้จ่ายมากยิ่งขึ้นซึ่งนักศึกษาส่วนหนึ่ง จะต้องเจอผลกระทบหนักเรื่องของค่าใช้จ่ายต่าง ๆ และอาจส่งผลกระทบต่อการศึกษา การได้รับทุนการศึกษาจะเป็นส่วนหนึ่งที่สามารถส่งเสริมให้นักศึกษามีแรงผลักดันในการศึกษาต่อไป

ด้วยระบบการศึกษาของประเทศไทยในปัจจุบัน ถือว่ามีความเหลื่อมล้ำทางโอกาสอยู่ค่อนข้างมาก เนื่องจากไม่มีมาตรการการช่วยเหลือด้านตัดสินใจของการศึกษา ทั้งปัจจัยความเหลื่อมล้ำฐานะทางเศรษฐกิจ และความเหลื่อมล้ำฐานะทางสังคมล้วนส่งผลต่อโอกาสทางการศึกษาทั้งสิ้น ซึ่งระบบการศึกษาไทยยังคงดำเนินด้วยเงินตราเป็นหลัก และการเข้าถึงโอกาสทางการศึกษาจะเป็นตัวกำหนดฐานะทางสังคม และการประกอบอาชีพในอนาคต แนวทางการแก้ปัญหาที่ เครื่องมือสำคัญที่ช่วยให้การเข้าถึงโอกาสทางการศึกษากระจายไปสู่คนในชั้นรากฐานของสังคมมากขึ้น ทำให้ในปัจจุบันมีหลายภาคส่วนที่ส่งเสริมการกระจายทุนการศึกษา ทั้งในระดับมัธยมศึกษา อุดมศึกษา และบัณฑิตศึกษา แต่ยังไม่เห็นเทคโนโลยีในการจัดการกระบวนการในด้านของทุนการศึกษา เพื่อให้เป็นระบบ ระเบียบ การจัดสรรทุนการศึกษายังไม่เพียงพอต่อความต้องการของนักศึกษาที่มีความประสงค์จะรับทุนการศึกษา ยังมีนักศึกษาจำนวนมากที่ต้องการและมีคุณสมบัติที่ตรงกับวัตถุประสงค์การให้ทุนการศึกษา แต่ยังไม่ได้รับการดูแลช่วยเหลือ

เพราะไม่มีมาตรการช่วยเหลือนักศึกษาอย่างเป็นระบบ จึงจำเป็นต้องมีแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ เข้ามาช่วยในการตัดสินใจ

จากปัญหาข้างต้นดังนั้นคณะผู้จัดทำจึงได้นำชุดข้อมูลตัวอย่างที่ได้มา เข้าสู่กระบวนการจำแนกตามรายการชุดข้อมูลโดยการใช้ขีดความสามารถในการประมวลผลของโปรแกรมคอมพิวเตอร์เพื่อเพิ่มความน่าเชื่อถือในการวิเคราะห์และประมวลผลข้อมูล โดยใช้เทคนิคต้นไม้ตัดสินใจในการสร้างแบบจำลองขึ้นมา เพื่อเป็นมาตรการการช่วยเหลือการตัดสินใจของด้านการศึกษาในการมอบทุนการศึกษา ให้กับนักเรียน นักศึกษาอย่างเป็นระบบ

1.2 วัตถุประสงค์

1.2.1 เพื่อพัฒนาแบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

1.2.2 เพื่อเสนอแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา

1.3 ประโยชน์ที่จะได้รับ

1.3.1 ได้แบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษา ด้วยเทคนิคต้นไม้ตัดสินใจ

1.3.2 ได้ข้อมูลแนวทางในการตัดสินใจที่จะมอบทุนการศึกษาให้กับนักศึกษา

1.4 ขอบเขต

1.4.1 ขอบเขตของข้อมูล

1.4.1.1 ข้อมูลที่นำมาวิเคราะห์ เป็นข้อมูลที่ได้จากเว็บไซต์ Kaggle.com จำนวน 1,044 ชุด

1.4.1.2 แบ่งข้อมูลที่นำมาวิเคราะห์ออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลสำหรับการเรียนรู้ (Training Data) 70% เป็นจำนวน 731 ชุด และกลุ่มข้อมูลสำหรับการทดสอบ (Testing Data) 30% จำนวน 313 ชุด

1.4.2 ขอบเขตการวิเคราะห์ข้อมูล

1.4.2.1 รวบรวมข้อมูล (Data Integration) เพื่อตรวจสอบรายละเอียดปริมาณและความน่าเชื่อถือของข้อมูลนักศึกษาเบื้องต้น เพื่อใช้สำหรับเป็นมาตรการช่วยเหลือในด้านของการศึกษา ที่ได้ชุดข้อมูลจากเว็บไซต์ Kaggle.com

1.4.2.2 คัดเลือกข้อมูล (Data Selection) โดยการเลือกคอลัมน์ที่มีความสัมพันธ์กันที่สามารถนำมาสร้างแบบจำลองต้นไม้ตัดสินใจได้ และนำมาใช้ในการนำมาวิเคราะห์

1.4.2.3 การทำความสะอาดข้อมูล (Data Cleaning) ตรวจสอบ เปลี่ยนแปลงข้อมูลให้ได้ข้อมูลที่สมบูรณ์และเหมาะสม เช่น ซึ่ขอลสถานศึกษา เป็นต้น ก่อนนำไปนำไปสร้างกฎในการตัดสินใจ

1.4.2.4 ทำเหมืองข้อมูล (Data Mining Engine) โดยใช้เทคนิคต้นไม้ตัดสินใจในการจำแนกประเภทในการ โดยการนำข้อมูลเข้าสู่โปรแกรม Weka สร้างกฎในการตัดสินใจ

1.4.2.5 ประเมินรูปแบบ (Pattern Evaluation) ที่ได้จากการทำเหมืองข้อมูลจากเทคนิคต้นไม้ตัดสินใจในขั้นตอนก่อนหน้า ซึ่งวัดค่าความยำโดยใช้ Confusion Matrix ในโปรแกรม Weka

1.4.3 ขอบเขตการปรับปรุงและการนำเสนอข้อมูล

1.4.3.1 นำเสนอชุดข้อมูลบนเว็บไซต์

1.4.3.2 นำเสนอรูปแบบโมเดล Decision Tree บนเว็บไซต์

1.4.3.3 นำเสนอข้อมูลสารสนเทศจากข้อมูลดิบ โดยใช้เครื่องมือการจัดการข้อมูล Data Visualization เพื่อนำเสนอเป็นแผนภาพแทนข้อมูล

1.4.3.4 ประยุกต์ใช้แบบจำลองมาตรการการช่วยเหลือด้านการศึกษา สำหรับนักศึกษาด้วยเทคนิคต้นไม้ตัดสินใจ บนเว็บไซต์

1.5 เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลและพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Acer Nitro 5 An515-43-R1N1 AMD Ryzen 5 3550H Ram 8 GB

1.5.1.2 Acer Nitro 5 An515-43-R1N1 AMD Ryzen 5 3550H Ram 16 GB

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public หรือ ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ

1.5.2.2 ชุดคำสั่งภาษา HTML4, HTML5 (Hyper Text Markup Language)

1.5.2.3 ชุดคำสั่ง CSS3 (Cascading Style Sheets)

1.5.2.4 โปรแกรม Atom ใช้ในการเขียนคำสั่งและเขียนโปรแกรม

1.5.2.5 โปรแกรม Adobe XD ใช้ในการออกแบบหน้าเว็บไซต์

1.5.2.6 ระบบปฏิบัติการ Window 10 ใช้ในการเปิดซอฟต์แวร์ต่าง ๆ

1.5.2.7 Microsoft Office 2016 ใช้ในการทำเอกสารต่าง ๆ

1.5.2.8 โปรแกรม Weka 3.8.4 ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคเหมืองข้อมูล

1.5.2.9 โปรแกรม Rapid Miner Studio 9.5.1 ใช้ในการทำด้วยเทคนิคเหมืองข้อมูล

1.5.2.10 โปรแกรม XAMPP 5.5.38 ใช้ในการจำลอง Web Server

1.5.2.11 โปรแกรม FileZilla 3.40.0 ใช้ในการติดต่อ Server เพื่อ Upload File

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

1.6.1 สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

1.6.2 แหล่งรวบรวมข้อมูล

เว็บไซต์ Kaggle.com

1.7 ระยะเวลาในการดำเนินการ

ตาราง 1.1 ระยะเวลาในการดำเนินงาน

แผนการดำเนินงาน	2563						
	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.
1. ศึกษาและกำหนดความต้องการ	←→						
2. ตรวจสอบความถูกต้องข้อมูล		←→					
3. วิเคราะห์ข้อมูลจำแนกตามกลุ่ม		←→					
4. ออกแบบระบบ			←→				
5. เขียนและทดสอบระบบ			←→				
6. ตรวจสอบระบบโดยรวม			←→				
7. ประเมินการใช้งานระบบ			←→				
8. จัดทำเอกสารประกอบ โครงการ	←→						

1.8 นิยามศัพท์เฉพาะ

การวิเคราะห์ข้อมูล (Analysis) หมายถึง เป็นการจัดระเบียบแยกแยะส่วนต่าง ๆ ของหลักฐานหรือข้อมูลที่ได้ออกเป็นหมวดหมู่ เพื่อหาคำตอบตามความมุ่งหมาย และตามสมมติฐานที่ได้กำหนดไว้ การวิเคราะห์ข้อมูลนี้เป็นขั้นตอนการทำงานที่ต่อเนื่องมาจากการวัด การนับ และจัดเรียงลำดับข้อมูล ส่วนใหญ่เป็นเรื่องเกี่ยวกับการนำเอาวิธีการทางสถิติมาวิเคราะห์หาค่าตัวแปรหรือหาลักษณะของตัวแปร

เว็บไซต์ (Website) หมายถึง สื่อนำเสนอข้อมูลบนเครื่องคอมพิวเตอร์ หรือคือการรวบรวมหน้าเว็บเพจหลายหน้า ซึ่งเชื่อมโยงกันผ่านทางไฮเปอร์ลิงก์ ซึ่งต้องเปิดด้วยโปรแกรมเฉพาะทางที่เรียกว่า Web Browser โดยถูกจัดเก็บไว้ในเวิร์ดไวด์เว็บ และเว็บไซต์นั้นถูกสร้างขึ้นด้วยภาษาทางคอมพิวเตอร์ที่เรียกว่า HTML (Hyper Text Markup Language)

ทุนการศึกษา (Scholarships) หมายถึง ทุนการศึกษามีมาช้านานแล้ว และมีหลายประเภท เช่น ทุนรางวัลผลการเรียนดี ให้แก่ผู้ซึ่งผ่านการสอบคัดเลือก แข่งขันความสามารถทางวิชาการ ทุนช่วยเหลือผู้ขาดแคลน ให้แก่เด็กและเยาวชนที่ยากจน หรืออยู่ในแดนทุรกันดาร ทุนส่งเสริมการศึกษาเฉพาะทาง เป็นทุนพิเศษ เพื่อส่งเสริมให้เรียนวิชาเฉพาะวิชาใดวิชาหนึ่ง ให้รู้สึกซึ่งสามารถนำวิชาไปถ่ายทอดแก่ผู้อื่น หรือปฏิบัติงานที่ต้องใช้ความเชี่ยวชาญเป็นพิเศษในเรื่องนั้น ๆ เป็นต้น

การศึกษา (Education) หมายถึง การศึกษาในความหมายทั่วไปอย่างกว้างที่สุด เป็นวิธีการส่งผ่านจุดมุ่งหมายและธรรมเนียมประเพณีให้ดำรงอยู่จากรุ่นหนึ่งสู่อีกรุ่นหนึ่ง โดยทั่วไป การศึกษาเกิดขึ้นผ่านประสบการณ์ใด ๆ ซึ่งมีผลกระทบเชิงพัฒนาต่อวิถีที่คนคนหนึ่งจะคิด รู้สึกหรือกระทำ แต่ในความหมายเทคนิคอย่างแคบ การศึกษาเป็นกระบวนการอย่างเป็นทางการซึ่งส่งผ่านความรู้ ทักษะ จารีตประเพณี และค่านิยมที่สั่งสมมาจาก รุ่นหนึ่ง ไปยังอีกรุ่นหนึ่ง นั่นคือ การสอนในสถานศึกษา สำหรับปัจจุบันนี้มีการแบ่งระดับชั้นทางการศึกษาออกเป็นชั้น ๆ เช่น การศึกษาปฐมวัย ประถมศึกษา มัธยมศึกษา ทั้งนี้รวมไปถึงระดับอาชีวศึกษา อุดมศึกษา และการฝึกงาน

แบบต้นไม้ตัดสินใจ (Decision tree) เป็นหนึ่งในวิธีการเรียนรู้ซึ่งใช้ในสถิติ, การเรียนรู้ของเครื่อง และการทำเหมืองข้อมูล โดยพิจารณาการสังเกตการแบ่งแยกข้อมูลโดยพิจารณาข้อมูลในการเรียนรู้ของเครื่อง (machine learning) ต้นไม้ตัดสินใจ เป็นโมเดลทางคณิตศาสตร์ที่ใช้ทำนายประเภทของวัตถุโดยพิจารณาจากลักษณะของวัตถุ บัพภายใน (inner node) ของต้นไม้จะแสดงตัวแปร ส่วนกิ่งจะแสดงค่าที่เป็นไปได้ของตัวแปร ส่วนบัพใบ (leaf node) จะแสดงประเภทของวัตถุ

ต้นไม้ตัดสินใจที่บัพใบแสดงถึงข้อมูลที่เป็นข้อมูลไม่ต่อเนื่อง (discrete values) จะเรียกว่าต้นไม้ตัดสินใจแบบจำแนก (classification trees) และต้นไม้ตัดสินใจที่บัพใบเป็นข้อมูลต่อเนื่อง (continuous values) จะเรียกว่าต้นไม้ตัดสินใจแบบถดถอย (regression trees)

ต้นไม้การตัดสินใจในการบริหารธุรกิจ เป็นแผนผังต้นไม้ช่วยในการตัดสินใจ โดยแสดงถึงมูลค่าของทรัพยากรที่จะใช้ ความเสี่ยงในการลงทุนและผลลัพธ์ที่มีโอกาสเกิดขึ้น ต้นไม้ตัดสินใจสร้างขึ้นเพื่อช่วยการตัดสินใจเพื่อใช้ในการสร้างแผนงาน นิยมใช้มากในการบริหารความเสี่ยง (risk management) ต้นไม้ตัดสินใจเป็นส่วนหนึ่งของทฤษฎีการตัดสินใจ (decision theory) และทฤษฎีกราฟ ต้นไม้ตัดสินใจเป็นวิธีการพื้นฐานอย่างหนึ่งสำหรับการทำเหมืองข้อมูล

1.9 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้สังเกตเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลนักศึกษาเบื้องต้น เพื่อใช้สำหรับเป็นมาตรการช่วยเหลือในด้านของการศึกษา ด้วยกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และเผยแพร่ข้อมูลสารสนเทศบน Web Browser ที่เป็นที่ยอมรับในยุคอินเทอร์เน็ตคือ การเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาเว็บไซต์สำหรับการเปิดเผยข้อมูลที่ได้มาจาก Kaggle.com เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูลหรือดำเนินงานต่าง ๆ ให้มีความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่ายและมีประสิทธิภาพ