บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของปัญหา

เห็ดเป็นสิ่งมีชีวิตจำพวกเห็ดรา มีทั้งที่ใช้เป็นอาหารและประเภทที่มีพิษ สายพันธุ์ของเห็ดมีมากกว่า 30,000 สายพันธุ์ กระจัดกระจายอยู่ทั่วโลกในจำนวนสายพันธุ์ดังกล่าวมีสายพันธุ์ที่มนุษย์สามารถนำมาบริโภคเป็นอาหารได้และสายพันธุ์ที่เป็นเห็ดที่มีพิษหรือเห็ดเมา ซึ่งถ้าบริโภคเข้าไปอาจเป็นอันตรายถึงแก่ชีวิตได้ เห็ดที่นำมาบริโภคเป็นอาหาร ในอดีตนั้นมีเพียงไม่กี่ชนิด ซึ่งในอดีตคนส่วนใหญ่มักจะมองและคิดถึงเห็ดในลักษณะของสิ่งที่ไม่น่าปลอดภัย ลึกลับและน่ากลัว ต่อมาคนเราได้มีการศึกษาและได้รับความรู้มากขึ้นจึงสามารถที่จะเลือกชนิดของเห็ดที่ปลอดภัยเพื่อนำมารับประทาน เดิมทีเป็นการเก็บเห็ดจากธรรมชาติมาบริโภค ทั้งนี้คนที่เก็บเห็ดจะต้องมีความรู้เกี่ยวกับเห็ดด้วยว่าเป็นเห็ดที่รับประทานได้หรือไม่ เพราะบางชนิดก็มีพิษหากรับประทานเข้าไปอาจเป็นอันตรายถึงชีวิตได้ ซึ่งในอดีตจนถึงปัจจุบันมีคนจำนวนมากเสียชีวิตจากการรับประทานเห็ดที่มีพิษ เนื่องจากการขาดความรู้ ความเชี่ยวชาญในเรื่องในรูปลักษณะของเห็ดทำให้ไม่สามารถแยกแยะได้ว่าเห็ดใดมีพิษ เนื่องจากเห็ดที่สามารถนำมารับประทานได้บางชนิดมีรูปร่างลักษณะคล้ายคลึง กับเห็ดมีพิษ ดังนั้นความรู้ในเรื่องลักษณะภายนอกของเห็ดจึงมีความสำคัญอย่างมากต่อการจำแนกสายพันธุ์เห็ด

การวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ คือการนำข้อมูลที่ได้ถูกรวบรวมไว้และถูกเผยแพร่ผ่านเว็บไซต์ kaggle.com ซึ่งเป็นเว็บไซต์ที่เก็บรวบรวมชุดข้อมูลต่าง ๆ เป็นแหล่งรวม Datasets หรือ ชุดข้อมูล สำหรับฝึกสอน Machine Learning ที่ใหญ่ที่สุดในโลกแห่งหนึ่ง มีข้อมูลทุกประเภทไม่ว่าจะเป็น Datasets ในหมวดหมู่ Finance, Business, Physics, Biology, Sports, News ซึ่งเป็นข้อมูลที่เปิดเผยได้ เพื่อให้ผู้ใช้บริการสามารถนำชุดข้อมูลไปศึกษาหรือวิเคราะห์ให้เกิดประโยชน์ต่อไปได้ การวิเคราะห์ข้อมูลปริมาณ

มหาศาลแสดงให้เห็นประโยชน์ต่าง ๆ ก็คือ ความรวดเร็วและประสิทธิภาพ ดำเนินการวิเคราะห์และค้นหาข้อมูลที่สามารถนำมาใช้ประกอบการตัดสินใจในอนาคตได้ ซึ่งการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ มาทำการวิเคราะห์นั้น เนื่องจากต้องการทราบถึงข้อมูลในเชิงลึกของคุณสมบัติใดที่บ่งบอกถึงเห็ดพิษมากที่สุด เช่น รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น เป็นต้น เห็ดถือเป็นกลุ่มสิ่งมีชีวิตที่สำคัญในระบบนิเวศ มีความสัมพันธ์ที่หลากหลาย กับสิ่งมีชีวิตชนิดอื่น ๆ ในระบบนิเวศ นอกจากนี้เห็ดยังเป็นสิ่งมีชีวิตกลุ่มหลักที่ใช้ในการชี้วัดความอุดมสมบูรณ์ของสภาพป่า และเป็นแหล่งอาหารสำหรับความหลากหลายของเห็ด ซึ่งสมควรมีการศึกษาอย่างเป็นระบบและมีการจัดทำการจำแนกประเภทของเห็ด เพื่อเก็บรวบรวมพันธุ์เห็ดต่าง ๆ ไว้ใช้ประโยชน์ในการจัดจำแนกเห็ดรับประทานได้หรือเห็ดพิษ ใช้เป็นข้อมูลที่สามารถนำไปศึกษาต่อยอดได้

ดังนั้น ทางผู้วิเคราะห์จึงมองเห็นปัญหานี้ เพื่อนำไปสู่กระบวนการวิเคราะห์ผลการรายงานผล ในการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนืออย่างถูกต้อง และแม่นยำ โดยนำข้อมูลมาวิเคราะห์หาคุณลักษณะใดที่บ่งบอกถึงเห็ดมีพิษมากที่สุด โดยมีปัจจัยที่เกี่ยวข้องเช่น รูปทรงของหมวกเห็ด พื้นผิวของหมวกเห็ด สีของหมวกเห็ด รอยช้ำ กลิ่น เป็นต้น ซึ่งได้ใช้เทคนิคการทำเหมืองข้อมูลแบบ Classification เพื่อให้เข้าใจถึงความสัมพันธ์ที่เกี่ยวข้องในการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ ให้ทราบถึงคุณลักษณะของเห็ดพิษได้อย่างชัดเจน สามารถนำข้อมูลที่วิเคราะห์ได้นั้นนำไปใช้ ในการดำเนินการแก้ไข ช่วยลดการรับประทานเห็ดพิษอย่างมีประสิทธิภาพ และทางผู้วิเคราะห์ข้อมูลยังได้ทำการนำเสนอข้อมูลแบบ visualization ด้วยการแสดงผลข้อมูลในรูปแบบของภาพด้วยโปรแกรม Tableau Public และเผยแพร่ข้อมูลสารสนเทศนี้บน Web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ ด้วยการสรุปจำแนกข้อมูลตามประเภทต่าง ๆ ได้สะดวกรวดเร็วยิ่งขึ้น รวมถึงมีความถูกต้อง สะดวก ลดความซ้ำซ้อนเพิ่มประสิทธิภาพให้กับการศึกษาค้นคว้า และทันต่อเวลา

1.2 วัตถุประสงค์

1.2.1 เพื่อวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

1.2.2 เพื่อเผยแพร่ข้อมูลบนเว็บไซต์

1.3 ประโยชน์ที่จะได้รับ

1.3.1 ได้ข้อมูลสารสนเทศของการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ

1.3.2 ได้เผยแพร่ข้อมูลบนเว็บไซต์

1.4 ขอบเขต

1.4.1 ขอบเขตผู้วิเคราะห์ข้อมูล

1.4.1.1 การรวบรวมข้อมูลที่ได้จากเว็บไซต์ kaggle.com

1.4.1.2 การทำเหมืองข้อมูลแบบ Classification เพื่อใช้ทำนายแนวโน้มการเกิดขึ้นของปัจจัย ที่บ่งบอกถึงเห็ดพิษด้วยการสร้างโมเดล Decision Tree

1.4.1.3 การประมวลผลโดยใช้โปรแกรม RapidMiner Studio และ โปรแกรม Weka

1.4.1.4 การนำเสนอข้อมูลแบบ Visualization ด้วยการแสดงผลข้อมูลในรูปแบบของกราฟ ด้วยโปรแกรม Tableau Public

1.4.1.5 การนำเสนอเผยแพร่ข้อมูลสารสนเทศนี้บน Web Browser โดยการใช้ภาษา HTML

1.4.1.6 ผู้วิเคราะห์ข้อมูลสามารถเพิ่มข้อมูลสารสนเทศเกี่ยวกับเห็ดบน Web Browser ได้

1.4.2 ขอบเขตผู้ใช้งานทั่วไปบนเว็บบราวเซอร์

1.4.2.1 สามารถดูข้อมูลสารสนเทศของการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ที่เผยแพร่ผ่านทางเว็บไซต์

1.4.2.2 สามารถดูข้อมูลสารสนเทศในลักษณะตารางสรุปข้อมูลบ่งชี้ถึงเห็ดพิษและเห็ดที่รับประทานได้

1.4.2.3 สามารถดูข้อมูลสารสนเทศในลักษณะรูปแบบของแผนภูมิจากโปรแกรม Tableau Public ได้

1.4.2.4 สามารถกรองข้อมูลสารสนเทศผ่านระบบของ Tableau Public บนWeb Browserได้

1.5 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

1.5.1 Hardware

1.5.1.1 Notebook: LENOVO IdeapadS410 core i5 Ram 4 GB

1.5.1.2 Notebook: Acer core i5 Ram 8 GB

1.5.2 Software

1.5.2.1 โปรแกรม Tableau Public ใช้สำหรับในการแสดงผลข้อมูลรูปแบบภาพ

1.5.2.2 ชุดคำสั่งภาษา HTML, CSS, JavaScript สำหรับพัฒนาหน้าจอแสดงผล

1.5.2.3 Adobe Dreamweaver ใช้ในการเขียนคำสั่งและออกเว็บไซต์

1.5.2.4 โปรแกรม Adobe XD ใช้สำหรับการออกแบบหน้าเว็บไซต์

1.5.2.5 ระบบปฏิบัติการ : Window 10 ใช้ในการเปิดซอฟต์แวร์ต่าง ๆ

1.5.2.6 โปรแกรม Microsoft SQL Server 2014 ใช้สำหรับจัดเตรียมข้อมูล

1.5.2.7 โปรแกรม Microsoft excel ใช้สำหรับการวิเคราะห์ข้อมูล

1.5.2.8 โปรแกรม Microsoft office 2016 ใช้สำหรับจัดการข้อมูลต่าง ๆ

1.6 สถานที่ใช้ในการดำเนินการศึกษาและรวบรวมข้อมูล

- สถานที่ใช้ในการดำเนินการศึกษา

มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนาภาคพายัพเชียงใหม่ 128 ถนนห้วยแก้ว ตำบลช้างเผือก อำเภอเมือง จังหวัดเชียงใหม่ 50300

- แหล่งรวบรวมข้อมูล

เว็บไซต์ Kaggle.com

1.6 ระยะเวลาในการดำเนินการ

ตารางที่ 1. 1 ระยะเวลาในการดำเนินการ

แผนการดำเนินการ	พ.ศ.2563						
	เม.ย	พ.ค	มิ.ย	ก.ค	ส.ค	ก.ย	ต.ค
1. ศึกษาและกำหนดความต้องการ	→						
2. ศึกษาข้อมูลและค้นหาข้อมูล	→	→					
3. การเตรียมข้อมูลและตรวจสอบความถูกต้อง			→				
4. วิเคราะห์ออกแบบฐานข้อมูล			→	→			
5. ออกแบบระบบ					→		
6. เขียนและทดสอบระบบ						→	
7. ตรวจสอบระบบโดยรวม							→
8. ประเมินผลและสรุปผลการวิเคราะห์ข้อมูล							→
9. จัดทำเอกสารประกอบโครงงาน		→	→	→	→	→	→

1.7 นิยามศัพท์เฉพาะ

1.7.1 อเมริกาเหนือ หมายความว่า ภูมิภาคอเมริกาเหนือมีเพียง 2 ประเทศ คือ ประเทศแคนาดา สหรัฐอเมริกา ซึ่งส่วนใหญ่ใช้ภาษาอังกฤษเป็นภาษาราชการ ส่วนอเมริกากลางใช้ภาษาสเปนเป็นหลัก ประกอบด้วยประเทศทั้งใหญ่และเล็ก รวมถึงประเทศที่เป็นหมู่เกาะจำนวน 22 ประเทศ โดยกรีนแลนด์เป็นประเทศอยู่เหนือที่สุด และประเทศปานามาอยู่ใต้สุด มีพรมแดนติดกับทวีปอเมริกาใต้

1.7.2 Visualization หมายความว่า การทำให้เห็นได้ หรือเรียกว่า จินตทัศน์ เป็นการกล่าวถึงการสร้างภาพ แผนผัง หรือ ภาพเคลื่อนไหว ใช้ในการสื่อสารแทนข้อความ โดยวิธีการนี้สามารถใช้ได้ทั้งในทางรูปธรรม และนามธรรม โดยมีทั้งการจำลองเหตุการณ์ที่เกิดขึ้นจริงในอดีต เหตุการณ์ที่ไม่สามารถมองเห็นได้ หรือการสร้างภาพในอนาคตเพื่อใช้ในการสื่อสาร การทำให้เห็นได้เป็นหนึ่งในวิธีการทำเหมืองข้อมูล โดยการสร้างภาพคอมพิวเตอร์กราฟิก ที่สามารถนำเสนอข้อมูลมากมายอย่างครบถ้วนแทนการใช้ข้อความนำเสนอข้อมูลที่มากมาย เราอาจพบข้อมูลที่ซ้อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยการทำให้เห็นได้

1.7.3 หมวกเห็ด(cap) หมายถึง เป็นส่วนประกอบปลายสุดของดอกที่เจริญเติบโตขึ้นไปในอากาศ เมื่อดอกบาน เต็มที่จะกางออก มีลักษณะรูปทรงเหมือนร่มกาง ขอบซุ้มลงหรือแบนราบ

หรือกลางหมวกเว้าลง เป็นแอ่ง มีรูปเหมือนกรวยปากกว้าง ผิวหมวก เห็ดด้านบนอาจจะเรียบ ขรุขระ มีเกล็ด (Scales) หรือมีขน แตกต่างกันไปแล้วแต่ชนิดของเห็ด เกล็ดหรือขนเป็นเนื้อเยื่อที่ยังคงเหลือติดจากกลุ่มเนื้อเยื่อห่อหุ้มดอกเห็ดอ่อน (outer veil) เนื้อของดอกเห็ดมีความหนาบางต่างกัน อาจเหนียว หรือฉีกขาดได้ง่าย สีของเนื้อเห็ดภายในและภายนอก อาจเป็นสีเดียวหรือแตกต่างกัน

1.7.4 ครีบ(gill) หมายถึง ด้านล่างของหมวกเห็ดมีครีบหรือเรียงเป็นรัศมีรอบก้านดอกห้อยแขวนลงมาจาก เนื้อของหมวกเห็ดที่อยู่ตอนบน เห็ดบางชนิดมีครีบหมวกด้านในยึดติด หรือไม่ยึดติดกับก้านดอก ด้านนอกเชื่อมติดกับขอบหมวกสองข้างของครีบหมวกเป็นที่เก็ดสปอร์ของดอกเห็ด ครีบหมวกนั้น อาจถูกย่อยให้ละลายเป็นของเหลวในเห็ดบางชนิด เช่น เห็ดหิงห้อย หรือเห็ดน้ำหมึก เห็ดแต่ละชนิดมีจำนวนครีบหมวกแตกต่างกัน และความหนาบางไม่เท่ากัน จำนวนครีบหมวกจึง ใช้เป็นลักษณะประกอบการจำแนกเห็ดด้วย สีของครับหมวกส่วนมากเป็นสีเดียวกับสปอร์ของเห็ด ซึ่งจัดเป็นลักษณะแตกต่างของเห็ดแต่ละชนิดโดยปกติมีสีขาว เหลือง ชมพู ม่วง น้ำตาลและดำ เห็ดบางสกุลไม่มีครีบแต่จะมีรู (pore หรือ tube)หรืออาจมีลักษณะคล้ายหนาม(spine) แทนครีบที่จะมีสปอร์อยู่ภายใน บางชนิดสปอร์จะถูกฝังอยู่ในเนื้อเยื่อก้อนวุ้น เช่น เห็ดหูหนู หรือมีสปอร์เกิด อยู่ในเปลือกหุ้มที่เป็นก้อนกลม เช่น เห็ดเผาะ

1.7.5 ก้านดอก(stalk หรือ stipe) หมายถึง ส่วนมากเป็นรูปทรงกระบอก บางชนิดมีโคน หรือปลายเรียวเล็ก มีขนาดและความยาวแตกต่างกัน ตอนบนยึดติดกับหมวกเห็ดหรือครับหมวกด้านใน ตอนล่าง ของเห็ดบางชนิดอาจมีเส้นใยหยาบรวมกันเป็นก้อนหรือเปลือกหุ้มโคน (volva) ซึ่งมีลักษณะ คล้ายถ้วยชาหงายรองรับอยู่ เช่น เปลือกหุ้มโคนในเห็ดฟาง ฯลฯ บนก้านดอกตอนบนของเห็ด บางชนิดมีวงแหวน (ring) หรือเยื่อบาง (ม่าน หรือ annulus) หุ้มอยู่โดยรอบ ก้านดอกเห็ดมีผิวเรียบ ขรุขระ หรือมีขน หรือมีเกล็ด เมื่อถูกสัมผัสด้วยมือหรืออากาศอาจเปลี่ยนสีได้

1.7.6 วงแหวน(ring หรือ annulus) หมายถึง เนื้อเยื่อบาง ๆ ยึดติดก้านดอกใต้หมวกเห็ดลงมาเล็กน้อย เป็นส่วนของเนื้อเยื่อห่อหุ้มครับเมื่อดอกเห็ดยังอ่อนที่เรียกว่า inner veilวงแหวนนี้อาจเลื่อนขึ้นลงได้ไม่ยึดติดกับก้านดอกในเห็ดบางชนิด

1.7.7 เปลือกหุ้ม (volva) หมายถึง เนื้อเยื่อชั้นนอกสุดที่ห่อหุ้มดอกเห็ดทั้งดอกไว้ในระยะที่เป็นดอกอ่อน หรือ outer veil นั่นเอง ซึ่งมีในเห็ดบางชนิด เช่น เห็ดฟาง และในเห็ดพิษหลายชนิดในสกุล Amanita เมื่อดอกเห็ดขยายใหญ่ขึ้นเปลือกหุ้มตอนบนจะแตกออก เพื่อให้หมวกเห็ดและ

ก้านดอก ยืดตัวชูสูงขึ้นมาในอากาศ ทิ้งให้เปลือกหุ้มอยู่ที่โคนก้าน มองดูเหมือนก้านดอกเห็ดอยู่ในถ้วย

1.8 บทสรุป

จากบทนำที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น ผู้วิเคราะห์ข้อมูลได้เล็งเห็นความสำคัญของการนำเทคโนโลยีมาประยุกต์ใช้ในการวิเคราะห์ข้อมูลการจำแนกเห็ดพิษหรือเห็ดที่รับประทานได้ในอเมริกาเหนือ ด้วยกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Classification ในรูปแบบของแผนภูมิต้นไม้ หรือที่เรียกว่า Decision tree และเผยแพร่ข้อมูลสารสนเทศบน web browser ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML ในการพัฒนาหน้าเว็บไซต์สำหรับการเปิดเผยข้อมูลที่ได้มาจาก kaggle.com เนื่องจากเทคโนโลยีปัจจุบันเข้ามาช่วยในการจัดการข้อมูล หรือดำเนินงานต่าง ๆ ให้มีความสะดวกสบายเป็นอย่างมาก อีกทั้งผู้ใช้งานยังเข้าถึงข้อมูลได้ง่าย และมีประสิทธิภาพ