**บทที่ 2**

**แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง**

ในการจัดทำโครงการเรื่อง การวิเคราะห์ความคิดเห็นของนักเรียน นักศึกษาวิทยาลัยอาชีวศึกษานครปฐมบนสื่อสังคมออนไลน์โดยใช้การประมวลผลภาษาธรรมชาติ คณะผู้จัดทำได้ศึกษาค้นคว้า แนวคิด ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้อง ได้แก่ สื่อสังคมออนไลน์ (Social Media) ทวิตเตอร์ (Twitter) การเรียนรู้ของเครื่องจักร (Machine Learning) การประมวลภาษาธรรมชาติ (Natural Language Processing : NLP) การจำแนกประเภทข้อความ(Text Classification) การวิเคราะห์ความรู้สึก (Sentiment Analysis) และกระดานสรุปข้อมูล(Dashboard) โดยมีรายละเอียดดังต่อไปนี้

**2.1 สื่อสังคมออนไลน์ (Social Media)**

สื่อสังคมออนไลน์ หมายถึง สื่อดิจิทัลที่เป็นเครื่องมือเพื่อใช้ในการสื่อสารระหว่างกันในเครือข่ายทางสังคม(Social Network) ผ่านทางเว็บไซต์ หรือโปรแกรมประยุกต์บนสื่อต่างๆ ที่มีการเชื่อมต่อกับอินเทอร์เน็ต โดยเน้นให้ผู้ใช้ทั้งที่เป็นผู้ส่งสาร และผู้รับสารมีส่วนร่วมอย่างสร้างสรรค์ในรูปแบบของข้อความ วีดีโอ รูปภาพ หรือเสียง

เครือข่ายทางสังคมเป็นช่องทางที่ใช้สำหรับเชื่อมต่อระหว่างบุคคล หรือกลุ่มบุคคล เพื่อให้เกิดเป็นกลุ่มสังคม เพื่อแลกเปลี่ยนและแบ่งปันระหว่างข้อมูลทั้งด้านธุรกิจ การเมือง การเรียน ในปัจจุบันมีบริเครือข่ายทางสังคมอย่างมากมาย เช่น เฟซบุ๊ค(Facebook) ไลน์(Line) อินสตาแกรม(Instagram) ทวิตเตอร์(Twitter) ดิสคอร์ด(Discord) ซูม(Zoom) และ ยูทูบ(YouTube) เป็นต้น ซึ่งแต่ละแพลตฟอร์มมีรูปแบบการแลกเปลี่ยนและแบ่งปันระหว่างข้อมูลที่แตกต่างกันออกไป เช่น Facebook จะเป็นการส่งข้อความหากันระหว่างบุคคล หรือกลุ่มบุคคล และแชร์เรื่องราว รวมถึงการคอมเมนต์ YouTube จะมีรูปแบบการแบ่งปันข้อมูล ในรูปแบบของวีดีโอ Twitter จะมีรูปแบบการแบ่งปันข้อมูลในรูปแบบของแฮชแท็ก เพื่อให้การสนทนาอยู่ในขอบเขตของหัวข้อการสนทนานั้นๆ เป็นต้น

**2.2 ทวิตเตอร์ (Twitter)**

ทวิตเตอร์ คือ บริการเครือข่ายสังคมออนไลน์จำพวกไมโครบล็อก (Micro Blog) โดยผู้ใช้สามารถส่งข้อความยาวไม่เกิน 280 ตัวอักษร ว่าตัวเองกำลังทำอะไร หรือเผยแพร่(Retweet) ข่าวสาร ที่น่าสนใจของผู้อื่น และข้อความที่ส่งถึงกันมีศัพท์เรียกว่า ทวีต(Tweet) ซึ่งเปรียบเหมือนเสียงนกร้องอยู่ตลอดเวลา

นอกเหนือจากการให้บริการทวีตข้อความบนทวิตเตอร์แล้วยังมีบริการ Twitter API เพื่อใช้ในการเข้าถึงข้อมูลสาธารณะของทวิตเตอร์ สามารถใช้เพื่อดึงข้อมูลโดย Twitter API ให้บริการทรัพยากรข้อมูลต่างๆ ดังตารางที่ 2-1

**ตารางที่**\***2-1**\*\*ข้อมูลการเข้าถึงทรัพยากร

|  |  |
| --- | --- |
| ทรัพยากร | รายละเอียด |
| Tweets | ดึงข้อมูลบล็อกการสนทนาที่ผู้ใช้โพสและโต้ตอบกัน |
| Users | ดึงข้อมูลของ Users |
| Spaces | ดึงข้อมูลกลุ่มของ Users บนทวิตเตอร์ |
| Direct Messages | รับและส่งข้อความไปยัง Target ที่ต้องการได้ |
| Trends | ดึงข้อมูลแนวโน้มและเนื้อหาตามพื้นที่ต่างๆ ที่เป็นที่สนใจ |
| Places | ดึงข้อมูลเชิงสถานที่ |

**2.3 การเรียนรู้ของเครื่องจักร (Machine Learning)**

การเรียนรู้ของเครื่องจักร (Machine Learning :ML) หมายถึง ศาสตร์แขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence) ที่ใช้ในการคิดและตัดสินใจ ที่ใช้สร้างการเรียนรู้ให้กับเครื่องโดยอาศัยวิธีการทางคณิตศาสตร์และสถิติในการสร้างแบบจำลอง การสร้าง AI ยังต้องนำความรู้ด้านวิทยาศาสตร์ข้อมูล (Data Science) มาเป็นพื้นฐานการคำนวณวิวัฒนาการของ Machine Learning สามารถเรียงลำดับความเป็นมาที่น่าสนใจบางส่วนตามลำดับปี พ.ศ. ได้ ดังตารางที่ 2-2

**ตารางที่**\***2-2**\*

|  |  |
| --- | --- |
| **ปี พ.ศ.** | **วิวัฒนาการ** |
| **2468** | Walter Pitts และ Warren McColloch ได้ร่วมกันสร้างทฤษฎี Neural Network โดยอาศัยหลักการทำงานจากสมองของมนุษย์มาเป็นการคิดคำนวณทางคณิตศาสตร์ และยังเป็นมาตรฐานมาจนถึงปัจจุบัน |
| **2493** | Alan Turing ได้สร้างแบบทดสอบที่ชื่อ Turing Test เพื่อตอบคำถาม Can machines think? โดยการทดสอบดังกล่าวก็เพื่อพิสูจน์ความฉลาดของคอมพิวเตอร์ ว่าสามารถคิดเองได้หรือไม่ |
| **2494** | Marvin Minsky และ Dean Enmonds ได้สร้าง Artificial Neural Network (โครงข่ายประสาทเทียม) ที่ประกอบด้วยการเชื่อมต่อกันทั้งหมด 40 นิวรอน |
| **2495** | Arthur Samuel คือผู้เริ่มต้นใช้คำว่า Machine Learning เป็นคนแรก โดยนำมาเป็นหลักการเพื่อสร้างโอกาสที่จะชนะในเกมคอมพิวเตอร์ |
| **2500** | Frank Rosenblatt ได้สร้าง Perceptron ซึ่งเป็นองค์ประกอบหนึ่งหน่วยของ Neural Network โดยจำลองมาจากเซลล์ประสาทมนุษย์ และใช้เรื่อยมาจนถึงปัจจุบัน |
| **2508** | Alexey (Oleksii) และ Valentin Lepa ได้พัฒนา Multi-layer Perception และถือว่าเป็นจุดเริ่มต้นที่พัฒนาไปสู่การเรียนรู้เชิงลึก (Deep Learning) |
| **2510** | Thomas Cover and Peter E. Hart จากมหาวิทยาลัย Stanford ได้สร้างโมเดลการเรียนรู้ที่เรียกว่า Nearest Neighbors ทั้งรูปแบบของ Classification และ Regression |
| **2512** | Marvin Minsky ได้แก้ไขปัญหาบางอย่างของ Perceptron ซึ่งสร้างโดย Rosenblatt ที่ไม่สามารถใช้กับฟังก์ชันที่ซับซ้อนได้ เช่น XOR ด้วยการแทนที่ Perceptron ลงใน Multiple Hidden Layer |
| **2513** | Seppo Linnainmaa ได้เผยแพร่หลักการของ Backpropagation ใน Neural Network โดยวิธี Differential ซึ่งก็ได้รับการพัฒนาเรื่อยมาจนถึงปัจจุบัน |
| **2522** | Kunihiko Fukushima ได้พัฒนาการจดจำรูปแบบ (Pattern Recognition) จากวิธีการของ Multi-layer Network และเป็นพื้นฐานที่ใช้งานกันในปัจจุบัน เช่น การจดจำใบหน้าหรือลายมือ และวัตถุต่าง ๆ เป็นต้น |

**ตารางที่\*2-2 (ต่อ)\***

|  |  |
| --- | --- |
| **ปี พ.ศ.** | **วิวัฒนาการ** |
| **2533** | Robert Schiapire และ Yoav Freund ได้สร้างโมเดลของ Machine Learning ที่เรียกว่า Boosting โดยการนำหลาย ๆ โมเดลมาใช้งานร่วมกัน |
| **2536** | Cortes และ Vapnil ได้สร้างโมเดลที่ชื่อ Support Vector Machines (SVM) โดยเริ่มสร้างตั้งแต่ปี 2536 และได้เผยแพร่ปี 2538 ซึ่งถือว่าเป็นอีกโมเดลหนึ่งที่ได้รับความนิยมของ Machine Learning |
| **2538** | Tin Kam Ho ได้สร้างโมเดล Random Forest ได้รวมการตัดสินใจแบบต้นไม้ หลาย ๆ อันเข้าด้วยกัน คล้ายกับต้นไม้หลายต้นที่รวมกันเป็นป่า |
| **2539** | Yann LeCun ได้เสนอวิธีการคำนวณแบบ Stochastic Gradient Descent ซึ่งต่อมาได้ถูกนำไปใช้ในขั้นตอนของ Backpropagation ของ Neural Network รวมถึง Deep Learning |

2.3.1 หลักการทำงานของ Machine Learning

การทำงานของ ML แตกต่างจากการเขียนโปรแกรมแบบดั้งเดิม(Traditional Programming) เป็นกระบวนการแบบแมนนวลที่ผู้เขียนโปรแกรมจะต้องกำหนดกฎเกณฑ์ว่าจะให้ทำงานอย่างไร ให้โปรแกรมทำงานแบบไหน และไปในทิศทางไหนตามอัลกิริธึมที่กำหนดไว้ ขณะที่การเรียนรู้ของเครื่องจักรจะกำหนดกฎจากข้อมูลโดยอัตโนมัติ กล่าวคือ ป้อนผลลัพธ์ที่ต้องการเข้าไป ตัวอย่างเช่น ใช้หลักการ Machine Learning ในการพัฒนาโมเดลให้สามารถแยกภาพสุนัขกับแมวโดยชุดข้อมูลรูปภาพ เรียกชุดข้อมูลที่จะใช้ในการฝึกโมเดลว่า Data Set จากนั้นป้อนข้อมูลและบอกกับ เครื่องจักรว่ารูปภาพนี้คือสุนัข รูปภาพนี้คือแมว จากนั้นเครื่องจักรจะวิเคราะห์ (Feature Extraction) และเรียนรู้เองว่าสุนัข และแมวต่างมีองค์ประกอบอะไรบ้างและจะสร้างกฎในการแบ่งกลุ่ม (Classification) ระหว่างสุนัข และแมวโดยอัตโนมัติ ภาพที่2-1 แสดงความแตกต่างของการเขียนโปรแกรมแบบดั้งเดิม และแนวทางของการเรียนรู้ของเครื่องจักร

Diagram

Description automatically generated

ภาพที่ 2-1 หลักการทำงานของ Machine Learning

2.3.2 หลักการทำนายผลของ Machine Learning

วัตถุประสงค์หลักการเรียนรู้ของเครื่องจักร การทำนายผลเพื่อใช้ในการตัดสินใจโดยอาศัยการเลียนแบบแนวคิดจากมนุษย์ แต่เนื่องจากคอมพิวเตอร์นั้นไม่มีสมองที่ใช้จดจำเรื่องราวหรือบันทึกประสบการณ์เหมือนกับมนุษย์ ดังนั้นจะต้องอาศัย ข้อมูลที่มีอยู่แล้ว เพื่อใช้สร้างแบบจำลองในการทำนายผลเกี่ยวกับสิ่งนั้น ดังตารางที่ 2-3

**ตารางที่**\***2-3**\*\*เก็บข้อมูลสภาพอากาศที่บ่งชี้ว่าฝนตกหรือไม่

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ลำดับ** | **อุณหภูมิ** | **ความชื้น** | **เมฆ** | **ฝน** |
| 1 | ต่ำ | สูง | มาก | ตก |
| 2 | ต่ำ | สูง | น้อย | ไม่ตก |
| 3 | สูง | สูง | น้อย | ไม่ตก |
| 4 | สูง | ต่ำ | มาก | ตก |
| 5 | ต่ำ | ต่ำ | น้อย | ไม่ตก |
| 6 | สูง | สูง | มาก | ??? |

ในแถวสุดท้ายของตารางที่ 2-3 คือลักษณะสภาพอากาศทีทำนายผล นั่นคือ ถ้าอุณหภูมิ : สูง ความชื้น : สูง และเมฆ : มาก ฝนจะตกหรือไม่ตก ซึ่งหากเราใช้หลักการคิดแบบ Machine Learning ในเบื้องต้น ผลลัพธ์จะได้ดังนี้

เมื่อดูที่คอลัมน์ อุณหภูมิ : สูง ดังในตารางถัดไปจะเห็นได้ว่า มีทั้งกรณีที่ฝนตกและฝน ไม่ตก ดังนั้นจึงตัดสินใจไม่ได้ว่า ถ้าอุณหภูมิสูงแล้วฝนจะตกหรือไม่ ซึ่งต้องพิจารณาตารางที่ 2-4 ดังต่อไปนี้

**ตารางที่**\***2-4**\*\*คอลัมน์อุณหภูมิ : สูง

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ลำดับ** | **อุณหภูมิ** | **ความชื้น** | **เมฆ** | **ฝน** |
| 1 | ต่ำ | สูง | มาก | ตก |
| 2 | ต่ำ | สูง | น้อย | ไม่ตก |
| 3 | สูง | สูง | น้อย | ไม่ตก |
| 4 | สูง | ต่ำ | มาก | ตก |
| 5 | ต่ำ | ต่ำ | น้อย | ไม่ตก |

เมื่อดูที่คอลัมน์ ความชื้น : สูง ดังในตารางถัดไปจะเห็นได้ว่า มีทั้งกรณีที่ฝนตกและฝนไม่ตก ดังนั้นจึงตัดสินใจไม่ได้ว่า ถ้าความชื้นสูงแล้วฝนจะตกหรือไม่ ซึ่งต้องพิจารณาตารางที่ 2-5 ดังต่อไปนี้

**ตารางที่**\***2-5**\*\*คอลัมน์ความชื้น : สูง

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ลำดับ** | **อุณหภูมิ** | **ความชื้น** | **เมฆ** | **ฝน** |
| 1 | ต่ำ | สูง | มาก | ตก |
| 2 | ต่ำ | สูง | น้อย | ไม่ตก |
| 3 | สูง | สูง | น้อย | ไม่ตก |
| 4 | สูง | ต่ำ | มาก | ตก |
| 5 | ต่ำ | ต่ำ | น้อย | ไม่ตก |

เมื่อดูที่คอลัมน์ เมฆ : มาก ในตารางถัดไปจะเห็นได้ว่า มีเฉพาะกรณีที่ฝนตกดังนั้นจึงสามารถตัดสินใจได้แล้วว่า ถ้าเมฆมากแล้วฝนจะตกซึ่งต้องพิจารณาตารางที่ 2-6 ดังต่อไปนี้

**ตารางที่**\***2-6**\*\*คอลัมน์เมฆ : มาก

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ลำดับ** | **อุณหภูมิ** | **ความชื้น** | **เมฆ** | **ฝน** |
| 1 | ต่ำ | สูง | มาก | ตก |
| 2 | ต่ำ | สูง | น้อย | ไม่ตก |
| 3 | สูง | สูง | น้อย | ไม่ตก |
| 4 | สูง | ต่ำ | มาก | ตก |
| 5 | ต่ำ | ต่ำ | น้อย | ไม่ตก |

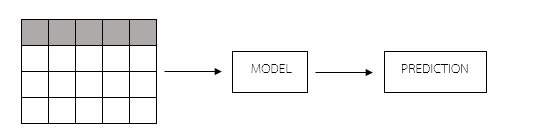
จากการเปรียบเทียบตามแนวคิดของ Machine Learning ดังที่กล่าวมา จึงสามารถสรุปได้ว่า ถ้าอุณหภูมิ : สูง และเมฆ : มาก จะทำนายได้ว่า : ฝนตก ดังตารางที่ 2-7

**ตารางที่**\***2-7**\*\*เปรียบเทียบตามแนวคิดของ Machine Learning

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ลำดับ** | **อุณหภูมิ** | **ความชื้น** | **เมฆ** | **ฝน** |
| 1 | ต่ำ | สูง | มาก | ตก |
| 2 | ต่ำ | สูง | น้อย | ไม่ตก |
| 3 | สูง | สูง | น้อย | ไม่ตก |
| 4 | สูง | ต่ำ | มาก | ตก |
| 5 | ต่ำ | ต่ำ | น้อย | ไม่ตก |
| ทำนาย | สูง | สูง | มาก | **ตก** |

2.3.3 โมเดลของ Machine Learning

กระบวนการพัฒนาโมเดล Machine Learning เริ่มจากการนำชุดข้อมูลตัวอย่าง (Sample Data หรือ Dataset) ที่มีอยู่ไปสร้างเป็นแบบจำลองโมเดล (Model) เพื่อการทำนายผล ดังภาพที่2-2



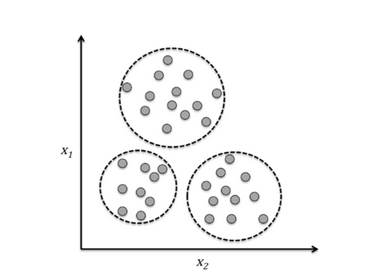
ภาพที่ 2-2 โมเดลของ Machine Learning

Machine Learning โมเดลมรการใช้รูปแบบการเรียนรู้ของเครื่องจักร 3 รูปแบบด้วยกัน ได้แก่ การเรียนรู้ของเครื่องจักรแบบมีผู้สอน(Supervised Learning) การเรียนรู้ของเครื่องจักรแบบไม่มีผู้สอน(Unsupervised Learning) และการเรียนรู้แบบเสริมแรง (Reinforcement Learning)

2.3.3.1 การเรียนรู้ของเครื่องจักรแบบมีผู้สอน(Supervised Learning)

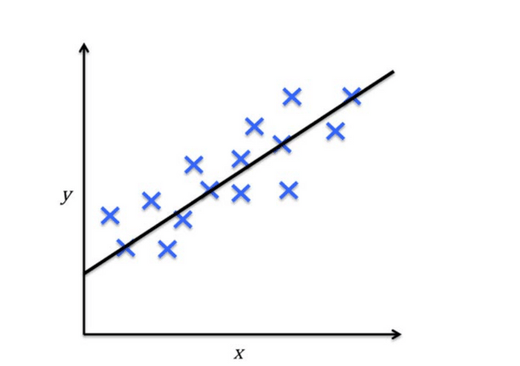
การทำให้คอมพิวเตอร์สามารถหาคำตอบของปัญหาได้ด้วยตนเอง “การมีผู้สอน” ในที่นี้หมายความว่า ในข้อมูลที่ใช้ฝึก (Training Data) เราจะมีมนุษย์มาคอยแยกประเภทหรือบอกผลลัพธ์ (Label) ที่ควรจะเป็นไปไว้ด้วย จากนั้นเราจะนำข้อมูลที่ใช้ฝึกไปผ่านอัลกอริทึมสำหรับสร้างโมเดลที่ไว้ทำนายผลลัพธ์ เมื่อเราได้โมเดลที่ไว้ทำนายแล้ว เราจะนำข้อมูลใหม่ที่เครื่องไม่เคยเห็น กล่าวคือไม่ใช่ข้อมูลชุดเดียวกันกับข้อมูลฝึกหัด เครื่องจะต้องทำนาย (Predict) ว่าคำตอบที่ได้ควรจะเป็นอะไร ประเภทของการเรียนรู้แบบมีผู้สอนจะแยกออกเป็นสองประเภท

ก) การแยกประเภท (Classification) การแยกประเภทเป็นหนึ่งในการเรียนรู้แบบมีผู้สอน และเป็นประเภทที่พบบ่อย เช่น ให้จดจำว่ามะม่วงที่มีสีเหลืองสุก และสีเขียวดิบ โดยมีภาพของมะม่วงจำนวน 100 รูปภาพ เราจะต้อง Label รูปภาพทั้งหมด โดยแบ่งมะม่วงได้ 2 ประเภท คือ สุก และดิบ ซึ่งประเภทที่แบ่งนั้นก็คือ Classification ดังภาพที่2-3



ภาพที่ 2-3 การแยกประเภทของข้อมูล (Classification)

ข) การถดถอยเชิงสถิติ(Statistical Regression) การถดถอยเชิงสถิติเป็นการหาความสัมพันธ์ระหว่างตัวแปร เช่น พื้นที่ของที่ดิน (x) กับราคาที่ดิน (y) โมเดลที่ใช้ในการหาความสัมพันธ์มีด้วยกันอยู่หลายแบบ แต่ที่นิยมกันคือการถดถอยเชิงเส้น (Linear Regression) ดังภาพที่ 2-4



ภาพที่ 2-4 การถดถอยเชิงสถิติ (Statistical Regression)

2.3.3.2 การเรียนรู้ของเครื่องจักรแบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอนจะไม่รู้คำตอบที่แน่ชัด แต่เราต้องการให้เครื่องตามหาโครงสร้างข้อมูลที่เราไม่รู้จัก (Unknown Structure) เช่น เราป้อนรูปเสือ และส้มโอ เมื่อเครื่องนำไปวิเคราะห์ก็จะสามารถเรียนรู้ถึงความแตกต่างระหว่างสองภาพนี้ แต่เราไม่ได้ส่ง Label ให้แก่เครื่องมันก็ไม่สามารถนำไปประมวลผล และจัดหมวดหมู(Classification) มันจึงใช้วิธีการแบ่งกลุ่มแทน(Clustering)

2.3.3.3 การเรียนรู้แบบเสริมแรง (Reinforcement Learning)

กลุ่มโมเดลประเภทนี้ คือการให้โมเดลกระทำการบางอย่าง(Action) และให้มันคอยสังเกตว่าการทำ Action แต่ละอย่างแล้วจะได้คะแนนสุดท้ายออกมาเป็นเท่าไร ถ้าทำ Action ตามลำดับต่างๆ ออกมาแล้วคะแนนดีก็จะได้รางวัล (Reward) เครื่องจะเริ่มจดจำลำดับการทำ Action แล้วได้คะแนนดี และพยายามจะทำ Action นั้นเรื่อยๆ การที่มันได้รางวัลเท่ากับเป็นการสนับสนุนให้โปรแกรมทำ Action นั้นซ้ำๆ ยิ่ง Action นั้นได้รางวัลมาก ก็จะทำแบบนั้นบ่อยๆ สิ่งที่สำคัญคือ เราต้องคิดโมเดลการให้คะแนนที่ดี

**2.4 การประมวลภาษาธรรมชาติ (Natural Language Processing : NLP)**

การประมวลภาษาธรรมชาติ หมายถึง วิทยาการแขนงหนึ่งในหมวดหมู่ของเทคโนโลยีปัญญาประดิษฐ์ซึ่งช่วยให้คอมพิวเตอร์สามารถเข้าใจ ตลอดจนตีความและใช้งานภาษาปกติที่มนุษย์ใช้สื่อสารได้ โดยเทคโนโลยีการประมวลภาษาธรรมชาติ มีรากฐานจากวิทยาการหลากหลายสาขาด้วยกัน โดยเฉพาะด้านวิทยาการคอมพิวเตอร์ (Computer Science) และ ภาษาศาสตร์เชิงคำนวณ (Computational Linguistics) เพื่อวัตถุประสงค์ในการพัฒนาให้เครื่องจักร หรือคอมพิวเตอร์สามารถเข้าใจภาษาของมนุษย์

2.4.1 ความสำคัญของการประมวลภาษาธรรมชาติ

2.4.1.1 ช่วยในการรับมือกับข้อมูลข้อความที่มีปริมาณมหาศาล การประมวลผลภาษาธรรมชาตินั้นช่วยให้อุปกรณ์คอมพิวเตอร์ต่าง ๆ สามารถสื่อสารกับมนุษย์ได้ด้วยการใช้งานภาษาของเครื่องเอง และดำเนินการทำงานต่าง ๆ ที่เกี่ยวข้องกับภาษาได้ ยกตัวอย่างเช่น การประมวลภาษาธรรมชาตินั้น ช่วยให้อุปกรณ์และคอมพิวเตอร์สามารถอ่านอักขระภาษาปกติ หรือทำความเข้าใจและตีความคำพูดของมนุษย์ ไปจนถึงการวัดอารมณ์ ความรู้สึกที่แฝงอยู่ในข้อความเหล่านั้นและกลั่นกรองใจความหรือนัยยะที่สำคัญออกมาเพื่อใช้งานระบบที่ทันสมัยในปัจจุบันสามารถวิเคราะห์ข้อมูลในปริมาณมหาศาลเกินกว่าขีดความสามารถของมนุษย์ โดยตัดข้อจำกัดเรื่องความเหน็ดเหนื่อยออกไป และสามารถทำงานด้วยความแม่นยำ คงเส้นคงวา และปราศจากอคติ การทำงานในปัจจุบัน มักต้องรับมือกับข้อมูลดิบจำนวนมหาศาล ซึ่งเกิดขึ้นอย่างต่อเนื่องในแต่ละวัน ไม่ว่าจะเป็นการทำงานในด้านประวัติคนไข้และทางการแพทย์ ไปจนถึงข้อมูลจากสื่อสังคมออนไลน์ ซึ่งการทำงานโดยอัตโนมัติจากปัญญาประดิษฐ์จะเป็นกุญแจสำคัญในการวิเคราะห์ข้อมูลเหล่านี้ได้ ไม่ว่าจะเป็นข้อมูลในรูปข้อความหรือคำพูด

2.4.1.2 ช่วยในการจัดระเบียบข้อมูลในลักษณะที่ไร้รูปแบบต่าง ๆ เนื่องจากภาษาที่มนุษย์ใช้กันนั้น มีความซับซ้อนและหลากหลายอย่างยิ่ง เพราะมนุษย์มีวิธีการแสดงออกมากมายนับไม่ถ้วน ทั้งในด้านการสื่อสารด้วยคำพูดหรือข้อความที่เกิดขึ้นด้วยการเขียน นอกจากการมีภาษานับร้อย ๆ พัน ๆ ภาษา ซึ่งต่างมีภาษาถิ่นแยกย่อยลงไปอีกนั้น ทุกภาษายังทวีความซับซ้อนยิ่งขึ้นไปอีกด้วยการมีชุดไวยากรณ์และโครงสร้างทางภาษาเฉพาะตัวของตนเอง รวมถึงคำ กลุ่มคำ และแม้แต่ศัพท์แสลงต่าง ๆ และเมื่อมนุษย์เราใช้ภาษาในการสื่อสารกันนั้น เรายังมักนิยมเขียนข้อความในรูปแบบย่อ ละเครื่องหมายวรรคตอนออกไป หรือแม้แต่การสะกดคำผิด ส่วนการสื่อสารด้วยวาจานั้นก็ยังมีประเด็นท้าทายของภาษาถิ่นและสำเนียงเฉพาะของแต่ละภูมิภาค แม้แต่ในภาษาเดียวกัน รวมถึงการพูดที่ไม่ชัดเจน อ้ำอึ้ง หรือใช้คำทับศัพท์แทรกแม้ว่าเทคนิคการทำงานทั้งแบบ Supervised Learning และ Unsupervised Learning โดยเฉพาะอย่างยิ่งกระบวนการทำงานแบบ Deep Learning จะได้ถูกนำมาใช้งานอย่างแพร่หลายในการสร้างแบบจำลองวิเคราะห์ภาษาของมนุษย์แล้วก็ตามก็ยังคงมีความจำเป็นในการสร้างความเข้าใจทางภาษาศาสตร์ที่ลึกและซับซ้อนยิ่งขึ้น รวมถึงความรู้ความเข้าใจเฉพาะด้าน ซึ่งแตกแขนงความชำนาญย่อยออกไปจากเทคนิค Machine Learning ตามปกติอีกด้วย ด้วยเหตุนี้ NLP จึงมีความสำคัญในการลดความสับสนทางการวิเคราะห์ภาษาลง และเพิ่มมิติให้แก่ข้อมูลในรูปของตัวเลข เพื่อการนำไปใช้งานต่าง ๆ ต่อไป เช่น ในการทำ Speech Recognition หรือการใช้งาน Text Analytics

2.4.2 กระบวนการทำงานของการประมวลภาษาธรรมชาติ

กระบวนการทำงานของการประมวลภาษาธรรมชาติประกอบด้วยหลากหลายวิธีการประมวลผลและแปลความหมายของภาษาปกติของมนุษย์ เช่น ระเบียบวิธีทางสถิติและการเรียนรู้ของเครื่องที่หลากหลาย ไปจนถึงกระบวนการทำงานตามขั้นตอน กฎเกณฑ์ และอัลกอริทึมที่ซับซ้อน ซึ่งเทคนิคในการรับมือข้อมูลเหล่านี้จำเป็นต่อการทำงาน เนื่องจากข้อมูลในรูปแบบข้อความคำและเสียงพูดนั้น มีความแตกต่างและหลากหลายค่อนข้างมาก รวมถึงสามารถใช้งานได้ในหลายรูปแบบเช่นกัน การทำงานขั้นพื้นฐานของการประมวลภาษาธรรมชาติ เช่นการทำ Tokenization Parsing Lemmatization/Stemming Part-of-speech Tagging Language Detection และ Identification of Semantic Relationships เป็นต้น หากจะกล่าวโดยสรุปแล้ว นั่นคือ การประมวลภาษาธรรมชาติ จะทำหน้าที่ย่อยข้อความหรือประโยคในภาษาที่มนุษย์ใช้สื่อสารลงเป็นส่วนย่อย ๆ หรือชิ้นเล็กๆ ที่ง่ายต่อการทำความเข้าใจของอุปกรณ์ จากนั้นจึงวิเคราะห์ความสัมพันธ์และประมวลผลระหว่างชิ้นส่วนต่างๆ ซึ่งกระบวนการทำงานย่อยๆ เหล่านี้ มักใช้งานในกระบวนการทำงานด้านการประมวลภาษาธรรมชาติ ที่ซับซ้อนยิ่งขึ้นไปอีกด้วย

2.4.2.1 Content Categorization การสรุปใจความหลักที่สำคัญของข้อความ หรือชุดการสื่อสารต่าง ๆ ซึ่งรวมถึงการสืบค้น การทำหมายเหตุ การแจ้งเตือนเนื้อหาที่สำคัญและการตรวจตราคำซ้ำ

2.4.2.2 Document Summarization การสร้างการสรุปสาระสำคัญจากข้อความปริมาณมากโดยอัตโนมัติ

2.4.2.3 Topic Discovery and Modeling เป็นการทำงานเพื่อจับใจความสำคัญและภาพรวมที่อยู่ในชุดข้อความขนาดใหญ่ และนำเทคนิคการวิเคราะห์ระดับสูง มาใช้กับข้อความเหล่านั้น

2.4.2.4 Contextual Extraction เป็นกระบวนการสกัดข้อมูลที่มีโครงสร้าง ออกมาจากข้อมูลรูปแบบข้อความทั่วไป

2.4.2.5 Sentiment Analysis การตรวจหาความรู้สึก อารมณ์ หรือมุมมองความคิดเห็นของผู้คนหรือสาธารณชน ซึ่งอยู่ในข้อมูลข้อความหรือคำพูด ซึ่งรวมไปถึงความรู้สึกโดยเฉลี่ยและการทำเหมืองข้อมูลเพื่อสำรวจความคิดเห็นจากคนจำนวนมาก

2.4.2.6 Machine Translation การแปลข้อความจากภาษาของมนุษย์จากภาษาหนึ่ง ไปยังอีกภาษาหนึ่ง โดยอัตโนมัติ

**2.5 การจำแนกประเภทข้อความ (Text Classification)**

การจำแนกประเภทข้อความ คือ การจัดแบ่งประเภทของเอกสาร เช่น อีเมล์ รีวิวสินค้า ข้อความในแชท ให้อยู่ในหมวดหมู่ที่กำหนด โดยหนึ่งเอกสารอาจจะอยู่มากกว่าหนึ่งหมวด ตัวอย่างหมวดหมู่เช่น อีเมล์ดี หรือ อีเมล์ขยะ, คะแนนรีวิวดี กลาง ต่ำ แย่ ตลอดจนการแยกแยะว่าภาษาในช่องสนทนานั้นคือภาษาอะไร การทำ Text Classification สามารถใช้หลากหลายวิธีการดังนี้

2.5.1 ระบบตามกฎ (Rule-Based Approach)

ระบบตามกฎ หมายถึง วิธีการคือจะกำหนดวิธีการที่อิงตามกฎจะจำแนกข้อความออกเป็นกลุ่มๆ โดยใช้ชุดภาษาศาสตร์ที่ประดิษฐ์ขึ้นเอง และจะพิจราณาความหมายของข้อความเพื่อระบุหมวดหมู่ที่เกี่ยวข้องตามเนื้อหา เช่น ต้องการแบ่งบทความข่าวออกเป็นสองกลุ่ม กีฬา และการเมือง โดยจะกำหนดรายการของคำที่แสดงถึงลักษณะเฉพาะของแต่ละกลุ่ม (เช่น คำที่เกี่ยวข้องกับกีฬา เช่น ฟุตบอล บาสเก็ตบอล เลอบรอน เจมส์ ลฯ และคำที่เกี่ยวข้องการกับเมืองเช่น โดนัลด์ทรัมป์ ปูติน เป็นต้น)

เมื่อต้องการจัดประเภทข้อความที่เข้ามาใหม่ จะต้องนับตำนวนคำที่เกี่ยวข้องกับหัวข้อที่ปรากฎในข้อความ เช่น “เกมแรกของเลอบรอน เจมส์กับทีมเลเกอร์สคือเมื่อไหร่” นับจำนวนคำที่เกี่ยวข้องกับกีฬา หากจำนวนคำที่เกี่ยวข้องกับกีฬามากกว่าจำนวนคำที่เกี่ยวข้องกับการเมือง ข้อความนั้นจะถูกจัดประเภทเป็นกีฬา

2.5.2 ระบบที่ใช้ในการเรียนรู้ของเครื่อง (Machine Learning Approach)

การจัดประเภทข้อความของแมชชีนเลินร์นิ่งจะเรียนรู้การจัดหมวดหมู่ตามข้อสังเกตในอดีต โดยใช้ตัวอย่างที่ติดป้ายกำกับไว้ล่วงหน้าเป็นข้อมูลการฝึกอบรม อัลกอริธึมการเรียนรู้ของเครื่องสามารถเรียนรู้ความเชื่อมโยงที่แตกต่างกันระหว่างส่วนต่างๆ ของข้อความ และคาดหวังผลลัพธ์เฉพาะ(แท็กข้อความ) สำหรับอินพุต(ข้อความ) “แท็ก” คือการจัดประเภทหรือหมวดหมู่ที่กำหนดไว้ล่วงหน้า เช่น ข้อความคือ “เมสซี่ทำแฮชทริค นำทีมพาชัย เข้าสู่รอบรองชนะเลิศ” ติดแท็กข้อความคือ “กีฬา” เป็นต้น

ขั้นตอนแรกในการฝึกอบรมตัวแยกประเภท NLP ของแมชชีนเลิร์นิงคือการแยกคุณลักษณะโดยการแปลงข้อความแต่ละรายการให้เป็นตัวเลขในรูปแบบของเวกเตอร์ หนึ่งในแนวทางที่ใช้บ่อยที่สุดคือการทำ Bag of Word (BOW) โดยที่เวกเตอร์แสดงถึงความถี่ของคำในพจนานุกรมของคำที่กำหนดไว้ล่วงหน้า จากนั้นอัลกอริธึมของแมชชีนเลิร์นนิงจะป้อนข้อมูลการฝึกที่ประกอบด้วยคู่ของชุดคุณลักษณะ ก็คือข้อความและแท็กของข้อความเพื่อสร้างแบบจำลองการจำแนกประเภทของข้อความ โดยจะอาศัยอัลกอรึธึมในการเรียนรู้ของเครื่อง เช่น Random Forest, Support Vector Machine, Naive Bayes และ Logistic Regression เป็นต้น

**2.6 การวิเคราะห์ความรู้สึก (Sentiment Analysis)**

การวิเคราะห์ความรู้สึกจากข้อความ หรือ Sentiment Analysis เป็นสาขาหนึ่งของงานด้าน NLP เราสามารถใช้หลักการ Sentiment Analysis มาช่วยในการวิเคราะห์ความรู้สึกของบุคคล เพื่อช่วยให้เข้าใจพฤติกรรมความรู้สึกที่มีต่อสิ่งนั้นๆ และสามารถนำผลที่ได้มาปรับให้เข้ากับธุรกิจ เพื่อตอบสนองความต้องการของคนเหล่านี้ได้ตรงตามจุด และมีประสิทธิภาพ ซึ่งกระบวนการ Sentiment Analysis มีดังนี้

2.6.1 Data Collection รวบรวมข้อมูลความคิดเห็นจากแหล่งต่างๆ เช่น คอมเมนต์ตามเฟสบุ๊ค รีวิวจากร้านค้าใน Lazada เป็นต้น

2.6.2 Data Pre-Processing เตรียมข้อมูลให้พร้อมก่อนนำไปวิเคราะห์ ได้แก่

2.6.2.1 การลบเครื่องหมายออกจากข้อความ ลบ Stop words (Tokenization)

2.6.2.2 นำคำที่ได้มาวิเคราะห์และใช้รากศัพท์ของคำ (Stemming) เช่น ate เปลี่ยนมาใช้รากศัพท์คือ eat

2.6.3 สร้างคลังคำศัพท์จากข้อความ(Bag of words) เช่น คลังคำศัพท์ของ Positive ได้แก่ สวย ดี ถูก คลังคำศัพท์ของ Negative ได้แก่ ราคาแพง พังง่าย ไม่ดีเลย แย่ เป็นต้น

2.6.4 Sentiment Analysis นำข้อมูลที่ผ่านการ Pre-process แล้วมาทำการวิเคราห์ เพื่อจำแนกประเภทของข้อความ (Text Classification) ว่าเป็น Positive (เชิงบวก) Negative (เชิงลบ) หรือ Neutral (เป็นกลาง) โดยวิธีการจำแนกแบบแนวทางตามกฎ คือ นับจำนวนคำที่ปรากฎในข้อความ หากมีคำที่เป็นเชิงบวกมากกว่า ระบบก็จะส่งกลับความรู้สึกเชิงบวก เช่น “สินค้าตัวนี้ดีมากๆเลย ราคาเหมาะสม” แต่หากตัวเลขที่นับเป็นบวก และลบเท่ากับ ระบบจะส่งกลับความรู้สึกเป็นกลาง

วิธีการอัตโนมติจะตรงกันข้ามกับระบบที่อ้างอิงตามกฎ ไม่ต้องพึ่งพากฎที่สร้างขึ้นด้วยตัวเอง แต่ใช้เทคนิคการเรียนรู้ของเครื่องจำแนกประเภทข้อความ(Text Classification) ป้อนข้อมูลการฝึกที่ประกอบด้วยคู่ของชุดคุณลักษณะ ก็คือข้อความและแท็กของข้อความเพื่อฝึกแบบจำลองการจำแนกข้อความ Sentiment Analysis

**2.7 กระดานสรุปข้อมูล (Dashboard)**

กระดานสรุปข้อมูล (Dashboard) คือการนำข้อมูลต่างๆที่อาจจะเป็นข้อมูลใน Report อยู่แล้วหรือข้อมูลใหม่ๆมาสรุปให้สามารถเห็นภาพได้ในหน้าเดียวเพื่อให้ดูง่ายและสามารถตีความได้ง่าย และเป็นข้อมูลที่อัพเดทสม่ำเสมอ การสร้าง Dashboard นั้นมีด้วยกันหลายรูปแบบ โดยแบ่งตามวัตถุประสงค์การใช้งาน ดังต่อไปนี้

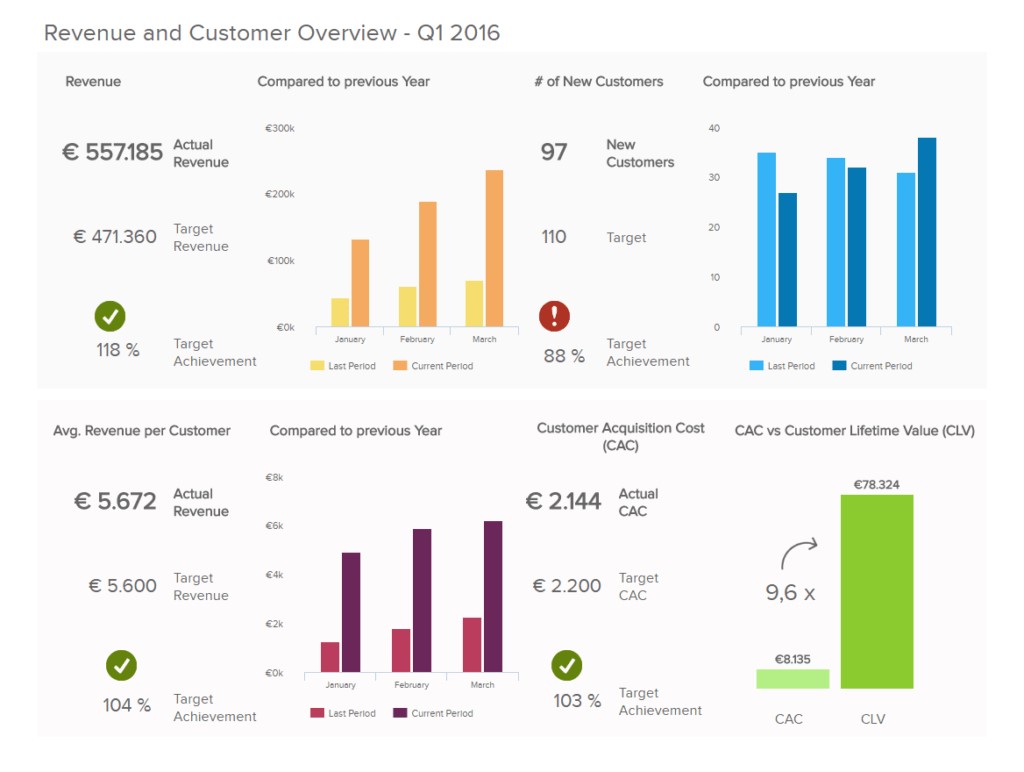
2.7.1 Data Dashboard เครื่องมือในการจัดการข้อมูลที่ติดตาม วิเคราะห์และแสดงดัชนีชี้วัดความสำเร็จของงาน (KPIs – key performance indicators) หรือ ดัชนีชี้วัดอื่นๆ รวมถึงจุดสำคัญของข้อมูลออกมาในรูปแบบ visual ทำให้สามารถเห็นความเป็นไปของธุรกิจ

2.7.2 ประโยชน์ของ Dashboard คือ สังเกตค่าตัวชี้วัดหลายตัวพร้อมกัน สามารถเห็นภาพรวมแบบ real-time ดึงข้อมูลมาทำเป็นรอบๆ ประหยัดทรัพยากรและเวลาในการทำรายงานแบบดั้งเดิม เป็นมาตรฐานที่เข้าใจตรงกัน ตรงประเด็น ไม่ซับซ้อน มีการตอบโต้ระหว่างคอมพิวเตอร์และผู้ใช้

2.7.3 รูปแบบของ Dashboard

2.7.3.1 Strategic Dashboard สำหรับ C-level

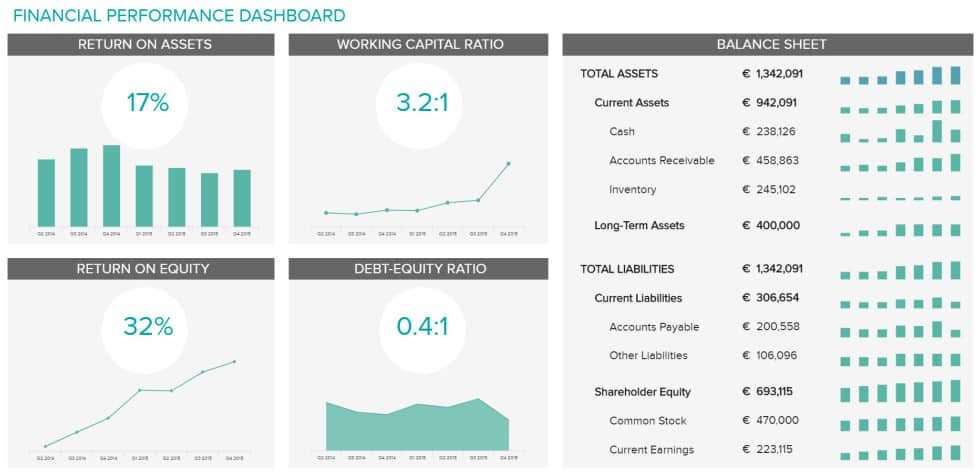
ใช้สำหรับในการวางแผนกลยุทธ์ ช่วยในการตัดสินใจของผู้จัดการทุกระดับ เน้นแสดงตัวชี้วัดที่สำคัญของบริษัท มุมมองเชิงเปรียบเทียบของบริษัทและบริษัทอื่น ตัวอย่างเช่น รายงานภาพรวมของการให้บริการลูกค้า โดยมีเป้าหมาย ตัวชี้วัดที่สำคัญ เพื่อเข้าใจธุรกิจและวางแผนจัดการปัญหา ดังภาพที่2-5



ภาพที่ 2-5 Strategic Dashboard สำหรับ C-level

2.7.3.2 Operational Dashboard สำหรับ Management & Team Lead level

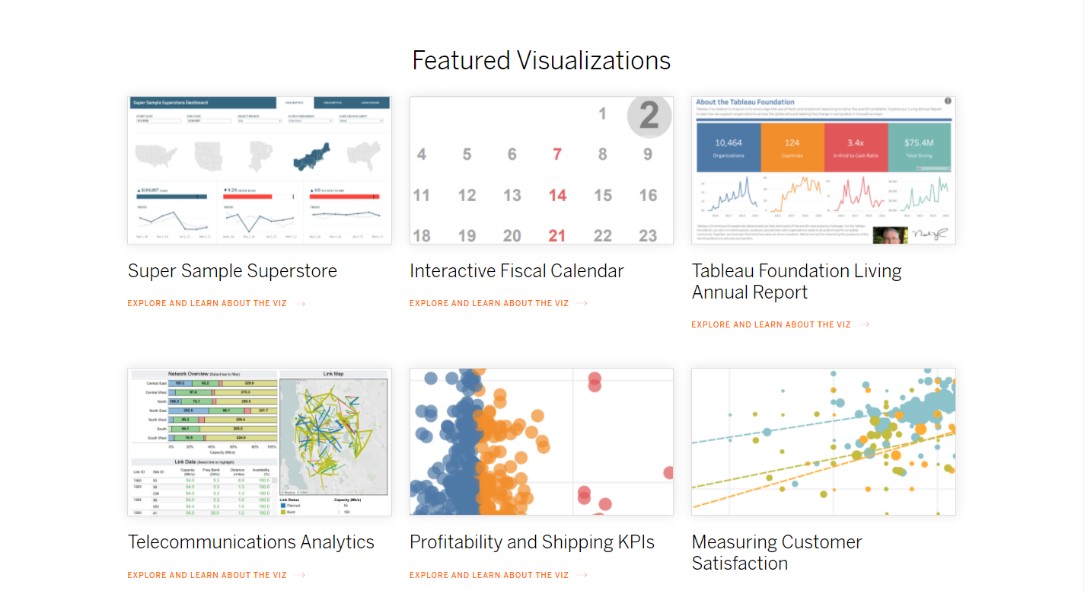
ใช้สำหรับตรวจสอบและควบคุมความคืบหน้าของงาน แสดงภาพรวมปัจจุบันของสายผลิตภัณฑ์ และแสดงผลเมื่อเกิดปัญหาแบบ Real-Time โดยรูปแบบนี้จะไม่เจาะลึกถึงรายละเอียดมากนัก ดังภาพที่2-6



ภาพที่ 2-6 Operational Dashboard สำหรับ Management & Team Lead level

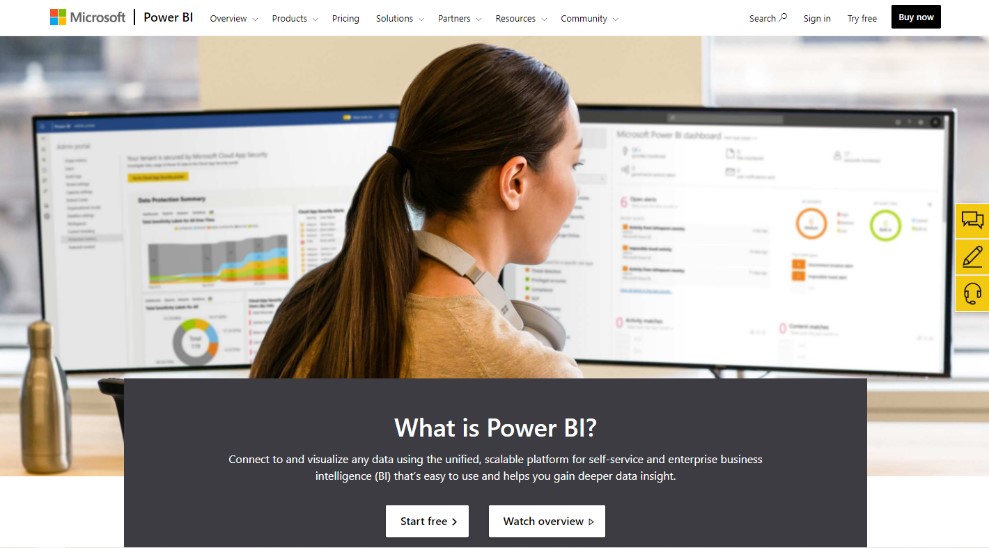
2.7.4 เครื่องมือที่ใช้ในการสร้าง Dashboard

2.7.4.1 Tableau ซอฟแวร์ที่ได้รับความนิยม และสร้าง Dashboard ใช้ในบริษัทที่มีงบค่อนข้างสูง ทั้งเวอร์ชั่นที่ไม่ต้องเสียค่าใช้จ่าย แต่ก็ใช้งานได้จำกัด และเวอร์ชั่นที่ใช้งานได้เต็มประสิทธิภาพ Community ที่ค่อนข้างใหญ่คอยช่วยเหลือและมีตัวอย่างการใช้งานที่หลากหลาย ดังภาพที่ 2-7

****

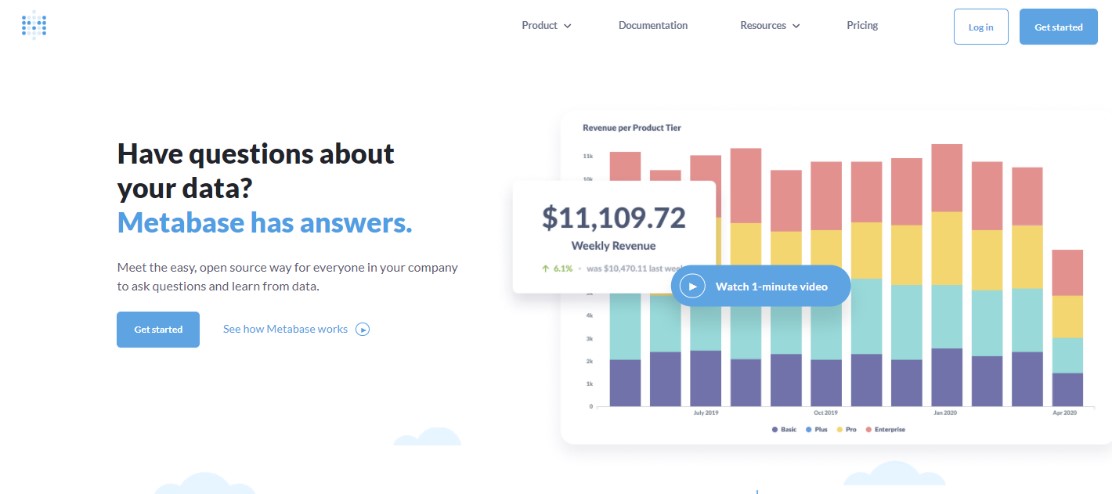
ภาพที่ 2-7 เครื่องมือที่ใช้ในการสร้าง Dashboard – Tableau

2.7.4.2 Microsoft PowerBI ซอฟแวร์ที่ไว้สร้าง Dashboard และช่วยในการวิเคราะห์ข้อมูล (data analysis) ทางธุรกิจ สำหรับคนที่ใช้ผลิตภัณฑ์ Microsoft Azure Office 365 และ Microsoft Excel



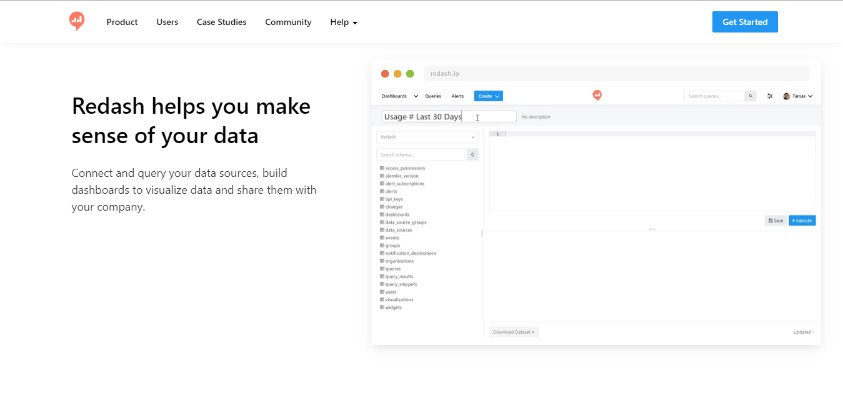
ภาพที่ 2-8 เครื่องมือที่ใช้ในการสร้าง Dashboard – Microsoft PowerBI

2.7.4.3 Metabase ซอฟแวร์ใช้งานง่ายในการสร้างกราฟ Dashboard และ Ad hoc Query ได้โดยลากวาง หรือใช้ SQL ก็ได้ สามารถเห็นรายละเอียดข้อมูลได้โดยตรง ดังภาพที่ 2-9



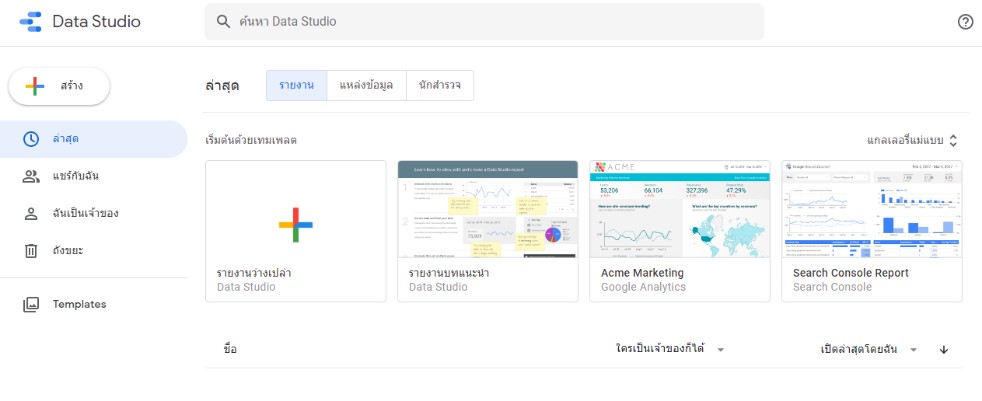
ภาพที่ 2-9 เครื่องมือที่ใช้ในการสร้าง Dashboard – Metabase

2.7.4.4 Redash เครื่องมือจากบริษัทลูกของ Databricks ที่ทำให้สามารถเชื่อมต่อกับ ข้อมูลและดึงข้อมูลจากหลายแหล่งพร้อมกัน เช่น SQL, NoSQL หรือ API มาสร้าง Dashboard เพื่อตอบคำถาม หา insight และแชร์กับคนอื่น ดังภาพที่ 2-10



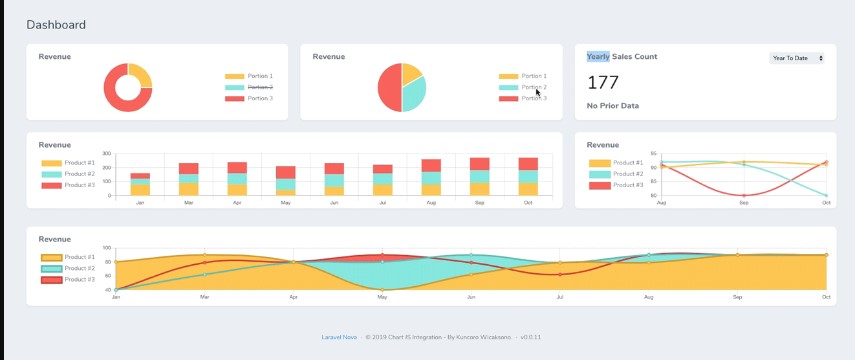
ภาพที่ 2-10 เครื่องมือที่ใช้ในการสร้าง Dashboard – Redash

2.7.4.5 Google Data Studio เครื่องมือจากทาง Google สำหรับใช้ดึงข้อมูลต่างๆ เพื่อมาสร้าง Dashboard ดูข้อมูลได้ทั้งข้อมูลเว็บไซต์ แพลตฟอร์ม หรือจากเครื่องมือต่างๆ ช่วยให้ดูข้อมูลได้ง่าย โดย Data Source ที่ Data Studio สามารถดึงมาตั้งแต่แพลตฟอร์มและเครื่องของ Google เอง เช่น Google Ads, YouTube, Google Analytics, Google Search Console เป็นต้น รวมทั้ง ข้อมูลจากแพลตฟอร์มหรือเครื่องมืออื่นๆ อย่าง Facebook, Twitter หรือระบบ CRM แอก็สามารถดึงได้ ผ่าน Data Connector ซึ่งเป็นพาร์ทเนอร์กับ GoogleData Studio สามารถเริ่มใช้งานได้ฟรีและง่ายมากๆ เพียงแค่มี Gmail หรือ Google Account ก็สามารถสมัครใช้งานได้เลย ดังภาพที่ 2-11



ภาพที่ 2-11 เครื่องมือที่ใช้ในการสร้าง Dashboard – Google Data Studio

2.7.4.6 Chart JS เครื่องมือเขียนกราฟสามารถใช้กับ Bootstrap 4 ได้อย่างดีเนื่องจากเขียนด้วย JavaScript และสามารถใช้ Option ของ Responsive ของ Chart JS ทำให้ปรับขนาดของกราฟได้ตามขนาดหน้าจอของอุปกรณ์ที่ใช้ โดยสามารถใช้กับ Bootstrap 4 ดังภาพที่ 2-12



ภาพที่ 2-12 เครื่องมือที่ใช้ในการสร้าง Dashboard – Chart JS

2.7.4.7 D3.js เป็น JavaScript Library ตัวหนึ่งสำหรับสร้าง Document ขึ้นมา โดย Document นั้นจะเปลี่ยนแปลงไปตาม Data ที่มอบให้ไปประมวลผลชื่อเต็ม Data-Driven Documents นั้นเองโดย d3.js จะช่วยจัดกลุ่ม Data เป็น element ย่อยๆ ด้านใน element อีกที ในรูปแบบของ DOM ( Document Object Model ) จากนั้นก็สามารถใช้ HTML , CSS และ SVG ในการแสดงข้อมูลนั้นๆออกมา ดังภาพที่ 2-13



ภาพที่ 2-13 เครื่องมือที่ใช้ในการสร้าง Dashboard – D3.js

**2.8 งานวิจัยที่เกี่ยวข้อง**

Vishwasa Navada [1] ได้ศึกษาการสกัดความรู้สึกโดยการนำข้อความจากสื่อโซเชียลมีเดียคือ ทวิตเตอร์(Twitter) มาทำการวิเคราะห์ด้วยเทคนิคอานีฟเบย์แบบเรีบลไทม์ โดยจำแนกผลการวิเคราะห์ออกเป็นส่วนคือ เป็นลบ(Nagitive) เป็นกลาง(Neutral) และเป็นบวก(Positive) ผู้วิจัยให้เหตุผลสำหรับการชิเทคนิคนี้ว่า เป็นเทคนิคที่ใช้งานง่ายและให้ประสิทธิภาพ ผลลัพธ์ของการวิจัยคือ สร้างเครื่องมือช่วยในการตัดสินใจ ในการทำการตลอด โดยแสดงข้อมูลในรูปแบบกราฟ ในงานวิจัยชิ้นนี้มีความสอดคล้องกับงานวิจัยของคณะผู้จัดทำ ทั้งแหล่งรวบรวมข้อมูล การสกัดความรู้สึก และการทำกราฟ

พนิดา ทรงรัมย์ [2] ได้ศึกษาการจำแนกความคิดเห็นทางการเมืองบนเครือข่ายสังคม โดยใช้วิธีการจำแนกแบบสัมพันธ์ ซึ่งเก็บรวบรวมความคิดเห็นเกี่ยวกับการดำเนินงานของ คสช. ระหว่างวันที่ 22 พฤษภาคม 2557 ถึง วันทที่ 17 กรกฎาคม 2557 โดยการใช้ Facebook Graph API ร่วมกับภาษา PHP เพื่อรวบรวมข้อความ มีข้อมูลที่รวบรวมได้ทั้งหมด 467 ข้อความ และนำไปจำแนก คือ ความคิดเห็นเชิงบวก และความคิดเห็นเชิงลบ โดยชิวิธีการจำแนกแบบความสัมพันธ์ของข้อมูลที่เกิดร่วมกันบ่อยด้วยขั้นตอนวิธี CBA (Classification Based on Associations) และใช้ 10-fold Cross Validation ในการแบ่งข้อมูลเรียนรู้ และข้อมูลทดสอบ ผลการจำแนกได้ความถูกต้อง 77.75%

สมัคร ชัยสงวน [3] ได้ศึกษาการวิเคราะห์ความรู้สึกแบบเรียลไทม์ โดยจำแนก ข้อมูลอานีฟเบย์สำหรับภาษาไทย ซึ่งเก็บรวบรวมความคิดเห็นบนเพจเฟซบุ๊ก ของมหาวิทยาลัยสุโขทัยธรรมาธิราช และนำมาวิเคราะห์ความรู้สึกของนักศึกษาที่มีต่อมหาวิทยาลัย เพื่อที่จะนำข้อมูลการวิเคราะห์มาพิจารณาถึงแนวทางในการปรับปรุงการจัดการเรียนการสอน ตลอดจนการพัฒนาระบบการบริการของมหาวิทยาลัย มีจุดประสงค์การพัฒนาตามความคาดหวังของผู้ที่จะเข้ามาศึกษาในระดับอุดมศึกษา ทำการวิเคราะห์ด้วยเทคนิคอานีฟเบย์ แบบเรียลไทม์ โดยจำแนกผลการวิเคราะห์ออกเป็นสามส่วน คือ เป็นบวก (Positive) เป็นลบ (Negative) และเป็นกลาง(Neutral) ผลการประเมินค่าความแม่นยำ (Precision) 80.03% ผลการประเมินค่าเรียกคืน (Recall) 90.52% ผลการประเมินหาค่าความถูกต้อง (Accuracy) 89.63% และผลการประเมินค่าการวัดประสิทธิภาพโดยรวม (F-measure) อยู่ที่ 84.81%

ชูชาติ หฤไชยะศักดิ์ [4] ได้วิจัยและพัฒนาเฟรมเวิร์ก(Framework) การวิเคราะห์ความรู้สึกสำหรับโซเชียลมีเดีย S-Sense ซึ่งเป็นเครื่องมือสำหรับรวบรวม ติดตาม และวิเคราะห์ประมวลผลข้อความบนเว็บไซต์เครือข่ายทางสังคมและเว็บบอร์ด เช่น เฟซบุ๊ค (Facebook) ทวิตเตอร์ (Twitter) พันทิป (Pantip) และ ยูทูบ (YouTube) เป็นต้น ข้อความส่วนใหญ่ที่ปรากฎบนอินเทอร์เน็ตนิยมใช้ภาษาที่ไม่เป็นทางการ มีโครงสร้างประโยคที่ไม่ถูกต้องตามไวยกรณ์ภาษาไทย ดังนั้นเทคโนโลยี S-Sense จึงถูกคิดค้นและพัฒนาขึ้นด้วยเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) และการทำเหมืองข้อความ (Text Minning)

**กรอบแนวคิดในการจัดทำโครงการ**

จากการศึกษางานวิจัยที่เกี่ยวข้องคณะผู้จัดทำมีแนวทางในการจัดทำโครงการเรื่อง การวิเคราะห์ความคิดเห็นของนักเรียน นักศึกษาวิทยาลัยอาชีวศึกษานครปฐมบนสื่อสังคมออนไลน์โดยใช้การประมวลผลภาษาธรรมชาติ ประกอบการพัฒนาโมเดลจำแนกข้อมูลตามความคิดเห็น และความรู้สึกของนักเรียนวิทยาลัยอาชีวศึกษานครปฐมที่ปรากฎบนสื่อสังคมออนไลน์ทวิตเตอร์ โดยใช้วิธีทางการเรียนรู้ของเครื่องจักร และการประมวลผลภาษาธรรมชาติ และแสดงผลการวิเคราะห์ในรูปแบบของกระดานข้อมูล (Dashboard) เพื่อเสนอต่อผู้ที่เกี่ยวข้องได้รับทราบถึงปัญหา ให้สามารถดำเนินการแก้ไขได้อย่างตรงจุด และมีประสิทธิภาพ