

## บทที่ 1

### บทนำ

#### 1.1 ความเป็นมาของปัญหา

ปัญญาประดิษฐ์มีบทบาทสำคัญต่อการบริหารและการจัดการ โดยการจัดเก็บชุดข้อมูลถือเป็นปัจจัยสำคัญในการที่จะนำไปสู่การพยากรณ์ หรือการสร้างโมเดลที่ช่วยซัพพอร์ตงานในองค์กรได้ ปัจจุบันมีองค์กรหลายองค์กรที่พยายามเก็บรวบรวมข้อมูลในหลายรูปแบบ เช่น รูปภาพ เสียง สัญญาณ และข้อความ โดยข้อมูลรูปภาพถือว่าเป็นชุดข้อมูลที่สำคัญที่ช่วยมาแก้ปัญหาในงานด้านธุรกิจ เช่น การรู้จำใบหน้า เพื่อยืนยันตัวตน การทำนายวัตถุ เพื่อใช้ในการตรวจสอบจำแนกข้อผิดพลาดของผลผลิต หรือผลิตภัณฑ์ต่างๆ เป็นต้น แต่ปัจจุบันคนส่วนใหญ่ยังเก็บข้อมูลในรูปแบบไฟล์ ทำให้เกิดปัญหาเกิดความล่าช้า หรือใช้เวลานาน ในกรณีการศึกษานี้ ผู้จัดทำจึงได้จัดเก็บเป็นรูปภาพจำพวก Classification รวบรวมข้อมูลให้กับคอมพิวเตอร์ และส่งไปให้ Machine Learning เพื่อจดจำการเรียนรู้

จากปัญหาดังกล่าว ผู้จัดทำจึงมีแนวคิดที่จะพัฒนา ระบบจัดเก็บชุดข้อมูลรูปภาพ โดยผู้พัฒนาจะจัดทำเว็บไซต์จัดเก็บชุดข้อมูล ที่สามารถค้นหา วิเคราะห์ สืบค้น และเรียนรู้เพิ่มเติมเกี่ยวกับข้อมูลประเภทรูปภาพ เป็นต้น เพื่อช่วยให้ง่ายต่อการทำนายผลและคาดการณ์ง่ายต่อการวางแผน การประเมินผลในหลายๆด้าน และอำนวยความสะดวกในกระบวนการวิเคราะห์ และคัดกรองข้อมูล เช่น การวิเคราะห์โครงสร้างข้อมูล หรือการคัดกรองข้อมูลที่ไม่เหมาะสมหรือไม่เหมือนจริง

#### 1.2 วัตถุประสงค์

- 1.2.1 เพื่อพัฒนาเว็บไซต์สำหรับจัดเก็บชุดข้อมูลที่สามารถนำข้อมูลประเภทรูปภาพไปใช้งานในการทดลองเพื่อทำนายผลด้วยโมเดลการจำแนกประเภทข้อมูล Classification
- 1.2.2 เพื่อช่วยอำนวยความสะดวก และลดโอกาสในการเกิดข้อผิดพลาดของชุดข้อมูลรูปภาพ
- 1.2.3 เพื่อให้ผู้ใช้งานสามารถจัดเก็บข้อมูล และเผยแพร่ข้อมูลต่อสาธารณชนที่สามารถนำไปใช้ประโยชน์

### 1.3 ขอบเขตโครงการ

ระบบจัดเก็บชุดข้อมูล มีการแบ่งงานออกเป็น 3 ส่วนหลักๆ คือ ส่วนการจัดการข้อมูลพื้นฐาน ส่วนการจัดเก็บชุดข้อมูล และส่วนสุดท้ายการออกรายงาน โดยแบ่งกลุ่มผู้ใช้งานออกเป็น 2 กลุ่ม ได้แก่ ผู้ดูแลระบบ และผู้ใช้งานทั่วไป ซึ่งแต่ละส่วนมีรายละเอียด ดังนี้

#### 1.3.1 ส่วนการจัดการข้อมูลพื้นฐาน

##### ส่วนของผู้ดูแลระบบ

1.3.1.1 สามารถเพิ่ม แก้ไข ค้นหา ข้อมูลผู้ใช้งานระบบ

1.3.1.2 สามารถเพิ่ม แก้ไข ค้นหา ข้อมูลประเภทชุดข้อมูล

##### ส่วนของผู้ใช้งานทั่วไป

1.3.1.3 สามารถ เพิ่ม แก้ไข ข้อมูลของผู้ใช้งานระบบ

#### 1.3.2 ส่วนการจัดเก็บชุดข้อมูล

##### ส่วนของผู้ดูแลระบบ

1.3.2.1 สามารถ ค้นหา เพิ่ม แก้ไข ข้อมูลการจัดเก็บชุดข้อมูลรูปภาพทีละไฟล์

1.3.2.2 สามารถ ค้นหา เพิ่ม แก้ไข ข้อมูลการจัดเก็บชุดข้อมูลรูปภาพหลายไฟล์

1.3.2.3 สามารถกำหนดสิทธิ์การเผยแพร่ชุดข้อมูล

1.3.2.4 สามารถกำหนดขนาดTraining และ Testing Datasets เพื่อดาวนโหลดชุดข้อมูล พร้อมไฟล์เฉลยในรูปแบบ CSV

##### ส่วนของผู้ใช้งานทั่วไป

1.3.2.5 สามารถ ค้นหา เพิ่ม แก้ไข ข้อมูลการจัดเก็บชุดข้อมูลรูปภาพทีละไฟล์

1.3.2.6 สามารถ ค้นหา เพิ่ม แก้ไข ข้อมูลการจัดเก็บชุดข้อมูลรูปภาพหลายไฟล์

1.3.2.7 สามารถกำหนดขนาดTraining และ Testing Datasets เพื่อดาวนโหลดชุดข้อมูล พร้อมไฟล์เฉลยในรูปแบบ CSV

#### 1.3.3 ออกรายงาน

##### ส่วนของผู้ดูแลระบบ

1.3.3.1 รายงานสรุปผลดาวนโหลดชุดข้อมูล

1.3.3.2 รายงานข้อมูลผู้ใช้งาน

1.3.3.3 รายงานยอดผู้เข้าชมชุดข้อมูล

##### ส่วนของผู้ใช้งานทั่วไป

1.3.3.4 รายงานชุดข้อมูลที่ดาวนโหลด

#### 1.3.4 ขอบเขตด้านฮาร์ดแวร์

1.3.4.1 Laptop ASUS TUF Gaming A15	1 ชุด
1.3.4.2 หน่วยประมวลผลกลาง (CPU) AMD Ryzen 7	1 หน่วย
1.3.4.3 หน่วยความจำหลัก (SDRAM)	16 GB
1.3.4.4 หน่วยความจำสำรอง (Hard disk)	512 GB
1.3.4.5 จอภาพ 15.6 นิ้ว	1 หน่วย
1.3.4.6 เมาส์	1 ชุด

#### 1.3.5 ขอบเขตด้านซอฟต์แวร์

- 1.3.5.1 ระบบปฏิบัติการ Window 11
- 1.3.5.2 ฐานข้อมูล MySQL
- 1.3.5.3 เครื่องมือสำหรับเขียนโปรแกรม Visual Studio Code
- 1.3.5.4 ภาษา PHP
- 1.3.5.5 โปรแกรม xampp v3.3.0
- 1.3.5.6 โปรแกรมสำหรับการแสดงผล Microsoft Edge

### 1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1.4.1 ช่วยให้สามารถนำข้อมูลประเภทรูปภาพไปใช้งานเพื่อทำนายผลด้วยโมเดลการจำแนกประเภทข้อมูล Classification
- 1.4.2 ช่วยอำนวยความสะดวก และลดโอกาสในการเกิดข้อผิดพลาดของชุดข้อมูลรูปภาพ
- 1.4.3 ช่วยให้การจัดการเก็บชุดข้อมูล และสามารถดาวน์โหลดเผยแพร่ต่อสาธารณชน

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการจัดทำโครงงานครั้งนี้ ผู้เขียนมีจุดมุ่งหมายที่จะศึกษาถึงกระบวนการจัดเก็บข้อมูลของการทำ Machine learning โดยมุ่งโฟกัสของการรวบรวม Classification โดยทฤษฎีและแนวทางที่จะนำมาศึกษาในการทำโครงงานครั้งนี้ ประกอบด้วย

2.1 กระบวนการเก็บข้อมูล

2.2 ข้อมูลของ Classification

2.3 เครื่องมือและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

#### 2.1 กระบวนการเก็บข้อมูล

การเก็บรวบรวมข้อมูล (Data Collection) คือ การเก็บข้อมูลขึ้นมาใหม่ และการรวบรวมข้อมูลของตัวแปรหรือเหตุการณ์ที่เราสนใจอย่างเป็นระบบ เพื่อที่จะนำไปวิเคราะห์หาคำตอบที่ถูกต้องในรูปแบบที่เหมาะสมอาจเป็นได้ทั้งข้อมูลปฐมภูมิและทุติยภูมิ ทั้งในเชิงปริมาณและคุณภาพ ด้วยวิธีต่างๆ เช่น การทำแบบสอบถาม การสัมภาษณ์ การสังเกตการณ์

##### 2.1.1 ประเภทของข้อมูลจากการเก็บข้อมูล

2.1.1.1 Quantitative Data (ข้อมูลเชิงปริมาณ) คือ ข้อมูลที่ได้จากการเก็บข้อมูลออกมาในเชิงตัวเลข (Numerical data) เพื่อแสดงปริมาณของสิ่งที่นับหรือสิ่งที่วัดได้ สามารถแบ่งได้เป็น 2 แบบ คือ

- ข้อมูลปริมาณแบบต่อเนื่อง (Continuous Data) เป็นข้อมูลที่มีค่าต่อเนื่องกันในช่วงที่กำหนด เช่น อายุ น้ำหนัก ส่วนสูง
- ข้อมูลเชิงปริมาณแบบไม่ต่อเนื่อง (Discrete Data) เป็นข้อมูลจำนวนเต็มหรือจำนวนนับ เช่น จำนวนรถยนต์ในกรุงเทพฯ จำนวนนักศึกษา

2.1.1.2 Qualitative Data (ข้อมูลเชิงคุณภาพ) คือ ข้อมูลที่ไม่สามารถวัดค่าได้ด้วยตัวเลขว่ามากหรือน้อย แต่จะเป็นข้อมูลที่แสดงถึงสถานภาพ คุณลักษณะ ทักษะ ทักษะ หรือคุณสมบัติ มักจะอยู่ในรูปแบบของคำพูด การบรรยาย การอธิบาย ตัวหนังสือ รูปภาพ หรือสัญลักษณ์ต่างๆ เช่น สีต่างๆ สถานที่ที่ชอบไป เชื้อชาติ สถานภาพสมรส ศาสนา

##### 2.1.2 แหล่งที่มาของข้อมูล

2.1.2.1 Primary Data (ข้อมูลปฐมภูมิ) คือ ข้อมูลที่เก็บจากแหล่งกำเนิดข้อมูลโดยตรง ซึ่งจะใช้วิธีการเก็บรวบรวมข้อมูลปฐมภูมิด้วยการ ออกแบบสอบถาม, การสัมภาษณ์, การสำรวจ , การสังเกตการณ์

2.1.2.2 Secondary Data (ข้อมูลทุติยภูมิ) คือ ข้อมูลที่ผู้ต้องการใช้ไม่ได้ทำการเก็บรวบรวมเอง แต่ได้จากผู้อื่นที่ทำการรวบรวมข้อมูลไว้ก่อนแล้ว ซึ่งวิธีการเก็บรวบรวมข้อมูลทุติยภูมิจะได้จากการรวบรวมข้อมูลจากแหล่งต่างๆ เช่น รายงานที่ตีพิมพ์แล้ว การเก็บรวบรวมข้อมูลสถิติ กราฟต่างๆ เป็นต้น

### 2.1.3 วิธีการเก็บรวบรวมข้อมูล

2.1.3.1 Questionnaires (แบบสอบถาม)

2.1.3.2 Interview (การสัมภาษณ์)

2.1.3.3 Focus Group (การสนทนากลุ่ม)

2.1.3.4 Observation (การสังเกตการณ์)

2.1.3.5 Experiment (การวิจัยเชิงทดลอง)

2.1.3.6 Secondary Data Collection (การเก็บรวบรวมข้อมูลทุติยภูมิ)

### 2.1.4 ขั้นตอนการรวบรวมข้อมูล



## 2.2 ข้อมูลของ Classification

Classification เป็นจำแนกประเภทข้อมูล ของ Machine Learning แบบ Supervised Learning โดยมีตัวอย่างในชุดข้อมูลสอน (training set) จะมีคุณลักษณะหนึ่งซึ่งบอกค่าประเภทของตัวอย่างนั้น เราเรียกค่าคุณลักษณะนี้ว่าฉลากบอกประเภท (class label) ซึ่งเป็นค่าข้อมูลแบบ categorical Classification มี 2 ขั้นตอนหลัก คือ

### 2.2.1 การสร้างแบบจำลอง (Model construction)

- เซตของตัวอย่างที่ใช้ในการสร้างแบบจำลอง เรียกว่า ชุดข้อมูลสอน
- ตัวอย่างแต่ละตัวจะมีคุณลักษณะหนึ่ง ซึ่งบอกค่าประเภทที่กำหนดไว้ล่วงหน้า
- แบบจำลองที่สร้างขึ้น แสดงผลลัพธ์การเรียนรู้การจำแนกประเภทในรูปแบบ กฎการจำแนกประเภท (classification rules), ต้นไม้ตัดสินใจ (decision trees) หรือสูตรทางคณิตศาสตร์ ต่างๆ

### 2.2.2 การนำแบบจำลองที่ได้ไปใช้ (Model usage)

- เพื่อการจำแนกประเภทตัวอย่างในอนาคต โดยจะต้องมีการประมาณค่าความแม่นยำ (accuracy) ของแบบจำลองที่ได้เสียก่อนการนำไปใช้
- เปรียบเทียบค่า Label ประเภทที่ทราบล่วงหน้าของตัวอย่างในชุดทดสอบ (test set) กับค่าผลลัพธ์การจำแนกประเภทที่ได้จากแบบจำลอง
- อัตราความแม่นยำ (accuracy rate) คำนวณได้จากเปอร์เซ็นต์ผลการจำแนกประเภทได้อย่างถูกต้องของแบบจำลองที่สร้างขึ้น
- ชุดข้อมูลทดสอบเป็นอิสระไม่ขึ้นต่อชุดข้อมูลสอน มิเช่นนั้นจะเกิด overfitting

## 2.3 เครื่องมือและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

### 2.3.1 ภาษาที่ใช้ในการพัฒนา

ภาษา PHP (Hypertext Preprocessor หรือชื่อเดิม Personal Home Page) เป็นภาษาสคริปต์โอเพ่นซอร์สทั่วไปที่ใช้กันอย่างแพร่หลายที่ทำงานฝั่งเซิร์ฟเวอร์ ใช้สำหรับการทำเว็บไซต์ สามารถฝังไฟล์ HTML ได้อย่างง่ายดาย สามารถเขียนโค้ด HTML ลงในไฟล์ PHP ได้ โค้ด PHP และไฟล์มีนามสกุล “.php” ทำงานบนฝั่งเซิร์ฟเวอร์ ในขณะที่โค้ด HTML ถูกดำเนินการโดยตรงบนเบราว์เซอร์

### 2.3.2 MySQL

เป็นระบบจัดการฐานข้อมูล หรือ Database Management System (DBMS) แบบข้อมูลเชิงสัมพันธ์ หรือ Relational Database Management System (RDBMS) ซึ่งเป็นระบบฐานข้อมูลที่จัดเก็บรวบรวมข้อมูลในรูปแบบตาราง โดยมีการแบ่งข้อมูลออกเป็นแถว (Row) และในแต่ละแถวแบ่งออกเป็นคอลัมน์ (Column) เพื่อเชื่อมโยงระหว่างข้อมูลในตารางกับข้อมูลในคอลัมน์ที่กำหนด แทนการเก็บข้อมูลที่แยกออกจากกัน โดยไม่มีความเชื่อมโยงกัน ซึ่งประกอบด้วยข้อมูล (Attribute) ที่มีความสัมพันธ์เชื่อมโยงกัน (Relation) โดยใช้ RDBMS Tools สำหรับการควบคุมและจัดเก็บฐานข้อมูลที่จำเป็น ทำให้นำไปประยุกต์ใช้งานได้ง่าย ช่วยเพิ่มประสิทธิภาพในการทำงาน

ให้มี ความยืดหยุ่นและรวดเร็วได้มากยิ่งขึ้น รวมถึงเชื่อมโยงข้อมูล ที่จัดแบ่งกลุ่มข้อมูลแต่ละประเภท ได้ตามต้องการ จึงทำให้MySQL เป็นโปรแกรมระบบจัดฐานข้อมูลที่ได้รับความนิยมสูง

### 2.3.3 Visual Studio Code

เป็นโปรแกรม Code Editor ที่ใช้ในการแก้ไขและปรับแต่งโค้ด จากค่าย ไมโครซอฟท์ มีการพัฒนาออกมาในรูปแบบของโอเพนซอร์ส ซึ่ง Visual Studio Code นั้น เหมาะสำหรับนักพัฒนาโปรแกรมที่ต้องการใช้งานข้ามแพลตฟอร์ม รองรับการใช้งานทั้งบน Windows, macOS และ Linux สนับสนุนทั้งภาษา JavaScript, TypeScript และ Node.js สามารถเชื่อมต่อกับ Git ได้มีเครื่องมือส่วนขยายต่าง ๆ ให้เลือกใช้อย่างมากมาย ไม่ว่าจะเป็นการเปิด ใช้งานภาษาอื่น ๆ ทั้ง ภาษา C++, C#, Java, Python, PHP, Go เป็นต้น

### 2.3.4 Xampp

โปรแกรมสำหรับจำลองเครื่องคอมพิวเตอร์ส่วนบุคคลของเราให้ทำงานในลักษณะ ของ Web Server ซึ่งเครื่องคอมพิวเตอร์ของเราจะเป็นทั้งเครื่องแม่ และเครื่องลูกในเครื่องเดียวกัน โดยไม่ต้องเชื่อมต่อกับ Internet ก็สามารถทดสอบกับเว็บไซต์ที่เราสร้างขึ้นมาได้ทุกที่ทุกเวลา

### บทที่ 3

#### วิธีการดำเนินงาน

จากทฤษฎีที่เกี่ยวข้องนำมาซึ่งการออกแบบของระบบ โดยผู้เขียนได้ทำการวิเคราะห์และออกแบบระบบโดยใช้คอนเท็กซ์ไดอะแกรม (Context Diagram) และออกแบบฐานข้อมูลด้วยแผนภาพ Entity Relationship Model