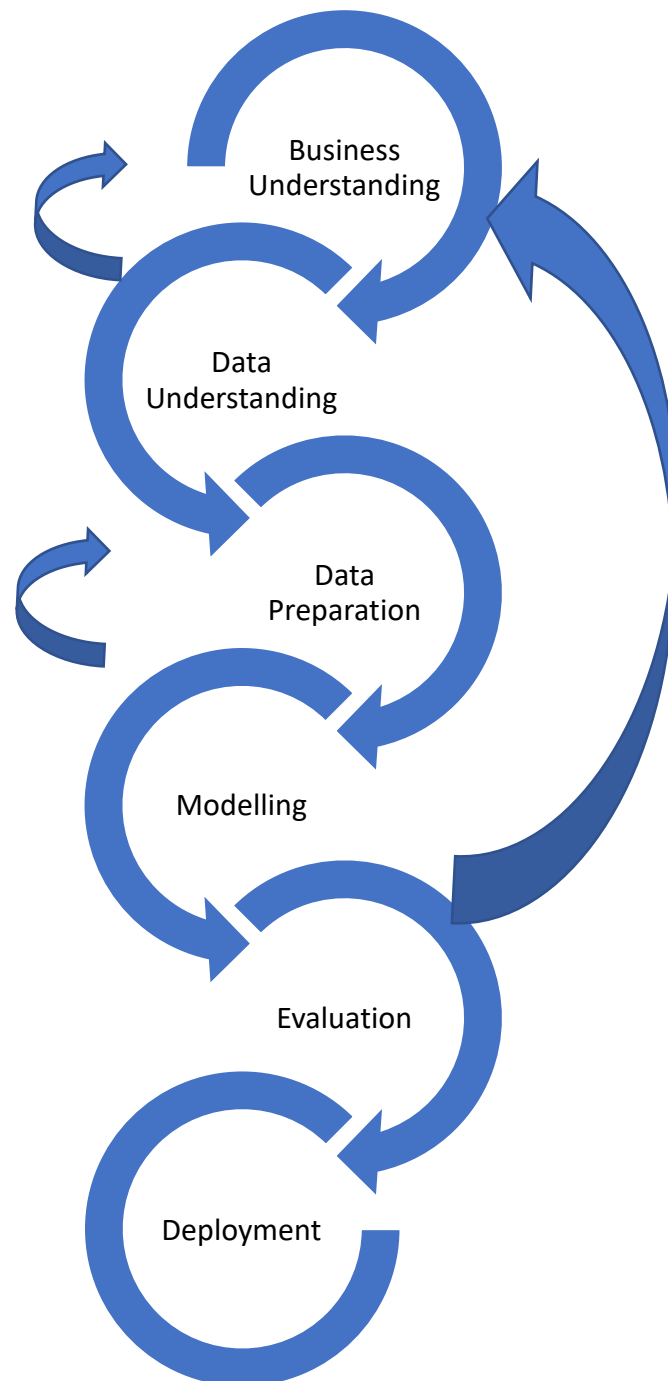


**Project Title:** J P Morgan classification for legal documents

To break down the problem of automating the classification of various legal documents using the CRISP-DM methodology using Analytical Project Framework Cycle.



## **1.Business Understanding:**

### **Defining Business Objective:**

Automate the classification of legal documents to reduce manual review time and increase accuracy.

### **Accessing Current Situation:**

Identify key stakeholders (Legal teams, loan officers, and compliance officers).

### **Determining Data Mining Goals:**

- Reduce document review time by 90%.
- Increase accuracy of document classification by 95%.
- Decrease loan-servicing mistakes by 80%.

### **Producing a Project Plan**

## **2.Data Understanding:**

### **Gathering initial data:**

Identify relevant data sources (Commercial loan agreements, credit-default swaps, custody agreements, and regulatory documents).

### **Describing the data:**

Collect sample datasets for training and testing. i.e., Texts , Images

### **Exploring data:**

Analyze document structures, clause patterns, and classification attributes.

### **Verifying data quality:**

Explore data quality issues i.e., Inconsistent formatting, missing information, and varying clause structures.

### **3.Data Preparation:**

#### **Selecting data:**

digitized commercial loan agreements

#### **Cleaning data:**

Remove noise, missing values extract relevant clauses and attributes.

#### **Integrating data:**

Split data into training, testing, and validation sets.

#### **Formatting data:**

tokenize, and normalize text.

### **4.Modelling:**

#### **Selecting Modelling technique:**

Image recognition, natural language processing (NLP), and machine learning techniques (e.g., supervised learning, deep learning).

#### **Designing tests:**

Performance metrics(accuracy,precision,recall,F1-score).

#### **Building Models:**

Evaluate models: Compare performance metrics (accuracy, precision, recall, F1-score) for different models.

### **5.Evaluation:**

#### **Evaluating results:**

Model performance on test data.

#### **Reviewing the Process:**

Compare results with manual review process.

#### **Determining the next steps:**

Identify areas for improvement.

## **6.Deployment:**

### **Planning deployment:**

Integrate COIN with J P Morgan's document review process.

### **Monitoring and Maintenance:**

Update the model with new data, retrain and re-evaluate.

### **Reviewing the project:**

Continuously watch the system's performance and retrain models according to their accuracy and precision efficiency.

### **Feedback:**

- Collect user feedback on the automated classification system.
- Identify new document types or regulatory changes requiring updates to the model.
- Continuously improve the system to maintain high accuracy and efficiency.

By following the CRISP-DM methodology, you can systematically break down the problem of automating legal document classification and develop a robust solution that meets J P Morgan's business objectives.

Here is the video explanation link:

[https://drive.google.com/file/d/15l43LC3GDqtfqN33RSgAownwf43ROwHP/view?usp=drive\\_link](https://drive.google.com/file/d/15l43LC3GDqtfqN33RSgAownwf43ROwHP/view?usp=drive_link)