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

**Course 1: Introduction to Data Analytics**  
**Student Name: Gongala.Saraswathi\_MAY 1st  
  
Problem Statement/ Business objectives:** “Automate the classification of various legal documents”.   
  
**Learning Outcomes**: Convert a business problem into an analytical problem and the ability to break the process down using CRISP-DM.  
  
**CRISP-DM consists of six steps as mentioned below:**  
  
Step-1 – Business Understanding  
Step-2 – Data Understanding  
Step-3 – Data Preparation  
Step-4 – Modeling  
Step-5 – Evaluation  
Step-6 – Deployment

**1. Business Understanding**:

This is the first step of the CRISP-DM method which focuses on defining business objective and goals and requirements for this project.  
  
JP Morgan needs to reduce the huge amount of time lawyers and staff spend reviewing legal documents, especially commercial loan agreements. The current process takes over **360,000 hours per year**, which is a lot! Also can increase the chances of errors.  
  
**Business Goals:**

* Automate clause **classification in legal documents**.
* Reduce manual labor and time.
* Improve accuracy and **reduce loan-servicing mistakes**.
* Extend automation to complex filings (credit-default swaps, custody agreements).

**2. Data Understanding**:

This step focuses on collecting, describing, and exploring the data relevant to the business problem. **Data Understanding** phase could include:

### Data Collection: Gather all the legal documents needed for classification based on the project goals. Make sure to collect examples of each type.

### Data Description: Learn about the features of the data and check the metadata of each document to help with better analysis.

### Data Exploration: Organize the data to better understand it, find any missing values, and do a basic analysis to see how different types of documents are spread out.

### Data Quality: Check the data for any mistakes, repeated entries, or inconsistencies to make sure it’s clean and reliable. **3**. Data Preparation:

This phase is dedicated to cleaning and transforming raw data into a suitable format.

**Data Cleaning and Transformation:** Remove duplicate entries, fix or handle missing values, and clean the text by converting raw data into a standard format for easier processing.

**Data Integration and Selection**: Combine all the data into one usable set and choose the right format of the cleaned data for building the model.

### This part may take a lot of time, especially if the documents are messy or inconsistent.

### **4**. Modeling:

In this step with clean data in-hand, various modelling techniques are applied. This step involves **Selecting and Building the model, Designing Test and Assessing the model.**

**Model Selection**: Based on the business requirements we need to select appropriate machine learning algorithm for text classification that can be applied and the model name is **COIN (Contract Intelligence).**

In this state the model is dedicated to cleaning and transforming raw data into a suitable format

for modelling.

In this state the model is dedicated to cleaning and transforming raw data into a suitable format

for modelling.

In this state the model is dedicated to cleaning and transforming raw data into a suitable format

for modelling.

Data Collection: According to the goals and objectives identified collect all the legal documents

that is required for classification and state classification of legal documents and collect samples

of each type.

Data Collection: According to the goals and objectives identified collect all the legal documents

that is required for classification and state classification of legal documents and collect samples

of each type.

Model Selection: Select appropriate machine learning algorithm for text classification that can

be applied and we can apply Natural Language Processing (NLP) and Deep learning model.

Model Selection: Select appropriate machine learning algorithm for text classification that can

be applied and we can apply Natural Language Processing (NLP) and Deep learning model.

Model Selection: Select appropriate machine learning algorithm for text classification that can

be applied and we can apply Natural Language Processing (NLP) and Deep learning model.

Model Selection: Select appropriate machine learning algorithm for text classification that can

be applied and we can apply Natural Language Processing (NLP) and Deep learning model.

**Model Training**: After selecting the appropriate model, we need to train that model and we can apply Deep learning model. After training, the model needs to be built and tested to help improve its accuracy and make it work better.

**5. Evaluation**

Before proceeding to deployment the model’s performance is thoroughly evaluated. It ensures that it meets the business objective set in the first phase. This step includes **evaluating the results, reviewing the process, and determining the next steps.**  
  
**Evaluating Results:** Test the model with new data to see if it works well and meets business goals. Review all steps taken so far to plan the next actions.

**Reviewing Process:** Check each stage for any issues. If errors are found, analyse them and look for ways to improve the model. If needed go back to the previous steps and make the required changes.  
  
**6. Deployment:**

After completing these six stages of crisp-Dm this is the stage where the model is put into use in the real world, and a few important tasks are done to make sure everything runs smoothly.

**Planning for Deployment**: We make sure the model can handle a large number of documents without slowing down. We also figure out the best way to connect it with the system it will be used in.

**Monitoring**: We set up tools that keep an eye on how the model is performing. If something starts to go wrong or the model isn’t working as well, we get alerts so we can fix it quickly.

**Final Checks**: Before everything goes live, we review the project, clean up any loose ends, and make sure everything is ready to go.

**Conclusion:**This project illustrated the potential of data analytics to address actual business problems, such as the automation of legal document review. With the use of machine learning model **COIN**, it was apparent how previously laborious tasks could be simplified. Using the **CRISP-DM** framework enabled decomposition of a sophisticated problem into simple steps, allowing for a structured way to approach the project. The exercise illustrated how this approach can be used across different analytics problems, and as such, it is a useful tool for solving many types of business issues.