Project Title: J P Morgan classification for legal documents

Problem Statement: "Automate the classification of various legal documents".

Here is a **step-by-step breakdown** using CRISP-DM Methodology:

Business Understanding:

1. **Objective** - Automation of legal contract review.
2. **Situation** - Lawyers and loan officers currently spend 360,000 hours per year on contract review. Errors in this process can lead to loan-servicing mistakes.
3. **Goals -**

- Decrease time spent on reviews
- Reduce errors
- Implement more complex filings, such as credit-default swaps and custody agreements.

4. **Project Plan** – Build a machine learning model to review entire documents with accuracy.

Data Understanding:

- Data consists of bank reports, loan contracts and other legal documents.
- Data can be types of documents, clause locations, textual formats, scanned images.
- Data includes clause types, Document length, clause frequency.
- Incomplete clauses, inconsistent language or formatting.

Data Preparation:

- Clean and format documents into clauses
- Label clauses according to categories
- Convert text to features
- Split data into training, validation, and test sets.

Data Modelling:

- Choose a model which is COIN in this case.
- Apply machine learning algorithms to identify patterns in these agreements
- Train it using labelled examples of clauses or agreements
- Fine tune it to get better accuracy
- Test it to make sure it performs reliably.

Evaluation:

- See how accurately the model understands and classifies clauses
- Use simple scores to measure its performance
- Look at where it makes mistakes and why
- Ask legal experts to review and confirm the results.

Deployment:

- Integrate the model into JP Morgan's existing systems
- Make it easy for legal teams to use the tool
- Monitor how it performs in real-world use
- Keep improving it based on feedback and new data.

Using this, JP Morgan can turn a time-consuming legal task into a fast, accurate, and smart process with COIN.

Here is the recorded video explanation link:
https://drive.google.com/file/d/195Cvmhzy0FOycG8kF7sdQFvxCGdhFa03/view?usp=drive_link