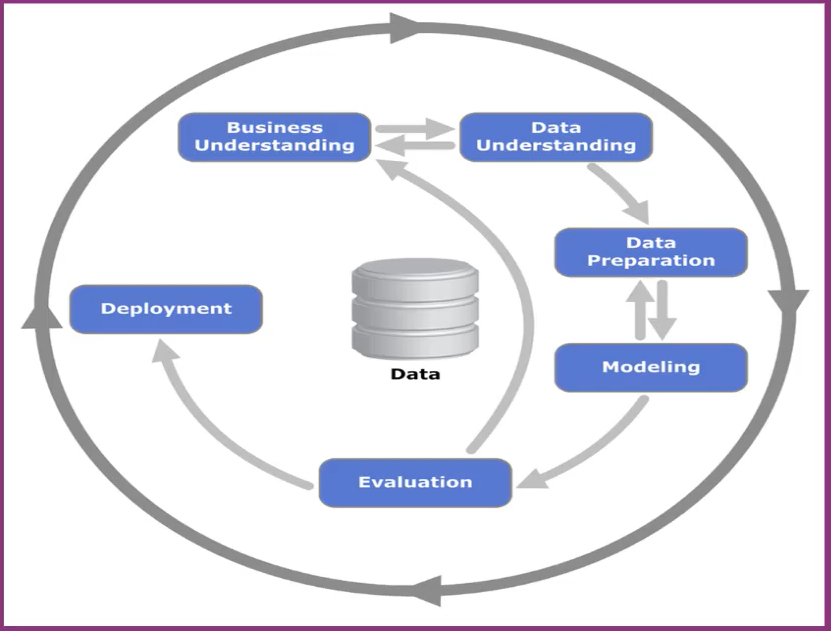
JP Morgan classification for legal documents by Crisp-dm

Name: - Ajay Email: - [2ajaythakre@gmail.com](mailto:2ajaythakre@gmail.com)

Video Explanation link: - <https://www.loom.com/share/8d356ca9782548f2874edfef692a1791?sid=eb832fdb-1753-4a84-adcf-592b5f1d8466>

## CRISP-DM Modal Chart



### Business Understanding: -

#### The Problem: -

The problem for JP Morgan was taking too much time for interpreting commercial loan agreements that consumes approx. 360k/hours of work/year by lawyers and loan officers

#### Business Objective: -

The objective is to reduce the costs, saves time, reduce human error and automate the process via ML model

#### Data Mining Goal: -

Build a model that categorize legal documents and automate the process.

This Model should take less time than taken by lawyers and officers

This modal should minimize the human errors.

### Data Understanding: -

According to crisp-dm modal now we have to collect data and describe data to make sure it meets our business objective

The data will be collected from bank past contracts and loan agreements

We can extract text for documents or images, pdf via OCR models and collect all the data in the digital form to perform data cleaning tasks in next phase

The data having repeated clauses, clause wording, headlines and location will be used for categorized.

we have to check data quality of scanned image/pdf they may have blurry images so we have to enhance them with image enhance technique

Data Preparation: -

The next step in Crisp-dm is to prepare the data so they are fit for analysis

Organize the data on the basis of repeated clauses and patterns.

Based on clause wording and location in the agreements into one of 150+ attributes

Data preparation step contains to check the quality of collected data does it have any error or missing values. we can replace them with a suitable value by predicting on the base of other other clauses wording sentences in there.

## Data Modeling: -

The next step in crisp-dm model is to modelling the data by ML models usually to find pattern inside and repititive clauses and words

it includes perform the analysis on the collected and clean data where we connect them together for meaningful insights categorize them on the basis of repeated words or demographical data like location.

## Evaluation: -

On the Evaluation phase we will implement the model in real world.

In this phase we will be checking how the model performs in business oriented environment and check does it correctly categorizing the data and check model output meets the real-world expectation.

Does it meet the speed and accuracy as required expectations

## Deployment: -

In deployment phase we will deploy the model in real-world   
it can now process the documents and categorize them well  
This will save hundreds of hours and improve overall efficiency.

## Conclusion: -

The JP Morgan project shows how analysis is very helpful in business problem and how it solve the problem via crisp-dm method and automate the repetitive tasks

By crisp-dm model we understand the problem, collect the data, prepare and analyse it

This model saves hundreds of hours and cost friendly and automate the manual repetitive tasks