Case Study:- J P Morgan classification for legal documents

**Question**: Distinct the process based on all the steps in CRISP-DM.(100 Marks)

**Learning Outcomes**: Convert a business problem into an analytical problem and the ability to break the process down using CRISP-DM.

Phase 1: Business Understanding

**Defining business objective**:- Business objective is to reduce the documentation review time

**Accessing the current situation**:- by manual it is time consuming and difficult to find the required documents.

**Determining data mining goals**:- the goal is to develop a software that automatically find the required documents in no time.

**Producing a project plan**:-defining the goals and the required team to deal with it and creating defined time slots for the particular work.

Phase 2: Data Understanding

**Gathering initial data**:-collect the legal documents from the concerned government department.

**Describing data**:- describe the different schemes ,tables and columns for the convenience.

**Exploring data**:- finding the common patterns.

**Verifying data quality**:-checking the data is complete and accurate. And deleting the missing data and improper data from the tables.

Phase 3: Data Preparation

**Selecting data**:-Selecting the sample data to find the insides.

**Cleaning data**:-remove the noise data , inconsistent data etc. for better understanding of the data.

**Integrating data**:- integrate the different fields of the data to reduce the redundancy.

**Formatting data**:- formatting the into a single form which suitable for the testing the different models.

Phase 4: Modeling

**Selecting modeling techniques**:-choosing the proper model for the data

**Designing tests**:- creating the test cases to test the model and cross verification it and also to find the error and bugs in the respective model.

**Building the model**:-After tests the model than we prepare the most defined model with the original dataset

**Assessing the model**:- see the accuracy, precision and other factors. And confirm the specific model.

Phase 5: Evaluation

**Evaluating results**:-check the model is meeting the required goals or not.

**Reviewing the process**:- while reviewing the process check for the area of improvement and changes in the models to optimize it.

**Determining the next steps**:- Doing the required improvements

Phase 6: Deployment

**Planning deployment**:-coordinating with the development team and plan the deployment.

**Monitoring and maintenance**:- Monitoring the performance in real time. Ans see the area of improvement.

**Reviewing the project**:- gathering the feedback from the usersand ask for the area of improvement and recommendations on the project.

**Finalizing the project**:- Making the final report on the project.