**Name = Manoj Kumar**

**1st Sep Batch Data Scientist Placement Guarantee Cource**

**Question: Distinct the process based on all the steps in CRISP-DM.**

**This is process of CRISP-DM for J P Morgan business.**

**Here is link Zip file :** [**https://drive.google.com/file/d/1F\_tOwaii9msGePNTDpfW\_0TsnrMf61Gv/view?usp=drive\_link**](https://drive.google.com/file/d/1F_tOwaii9msGePNTDpfW_0TsnrMf61Gv/view?usp=drive_link)

**Here is the directy watch video on Google Drive:**

[**https://drive.google.com/file/d/1QpwfbcjbMbAhKzvToA8dDgFjib-Vk164/view?usp=drive\_link**](https://drive.google.com/file/d/1QpwfbcjbMbAhKzvToA8dDgFjib-Vk164/view?usp=drive_link)

**BUSINESS UNDERSTANDING**

Objective and requirements of the software COIN by J P Morgan is automatically classify the legal documents ,which was taking 360,000 persons-hours manually before now it can happen within a second.

**(1)DEFINING BUSINESS OBJECTIVE**

The objective of J P Morgan business is the process of the reviewing legal documents of the commercial loan agreements should be fast and efficient to save time.

**(2)ACCESSING THE CURRENT SITUATION**

Current situation is to reviewing the document manually it is time consuming and there can be some errors also. J P Morgan intends to deploy COIN for more complex filings like credit-default swaps and custody agreement in the medium and long term.

**(3)DETERMINING DATA MINING GOALS**

With the help of the some algorithm should categorize and identify the clauses, and COIN software can detect the patterns which is difficult to review manually.

**PRODUCING OF PROJECT PLAN**

The project plan is the process of reviewing legal document should be fast and error free and J P Morgan intends to deploy COIN for more complex filings like credit-default swaps and custody agreements.

**DATA UNDERSTANDING:**

Understand the data structure and identify the data patterns, ensure the quality of the data which we got from legal documents, agreements and the data of bank's numerous contracts.

**(1)GATHERING INTIAL DATA**

Gathered data from commercial loan agreements, legal documents which used to reviewed manually before by lawyers and loan officers.

**(2)DESCRIBING DATA**

Described the data to understand the basic structure, properties and data type to identify the patterns and trends of the data.

**(3)EXPLORING THE DATA**

Explored data to understand complex patterns and relationships, which clauses frequently appearing etc.

**(4)VERIFYING DATA QUALITY**

Checks data quality like in which form data is text, numeric, complex and what is the type of data structured or unstructured.

**DATA PREPARATION**

To get prepared the data and get analyzed the data for modeling. like missing value treatment, outlier treatment , handling duplicate etc.

**(1)SELECTING DATA**

Selected the data from legal documents and commercial loan agreements, i think they selected the data which should cover the wide range of the clauses of the documents.

**(2)CLEANING DATA**

In the data cleaning process they have done missing value treatment , and they dealt with outliers, removed duplicated values , added new column if needed etc.

**(3)INTERGRATING DATA**

J P Morgan business team ensured that the data should be combined in single data set so that it will become for modeling.

**(4)FORMATTING DATA**

If there was any data in the form of text they have converted all the data into numeric form so that the data can be in same structure it will be easy for machine learning algorithms

**MODELING:**

In modeling phase they had to make a model that can accurately classify the legal documents and automate the process of reviewing documents.

**(1)SELECTING THE MODELING TECHNIQUES**

Here they have to decide which technique is good for reviewing document automatically , classification like Decision tree , Random Forest i think they have decided to use because it can best result.

**(2)DESIGNING TESTS**

I think they had splitted the data in training and testing datasets. Then trained the model done some metrics like accuracy, recall, F1-score etc.

**(3)BUILDING THE MODEL**

On the bases of selected techniques above they built the model ,trained the model on machine learning algorithms using prepared dataset, provided the training examples etc.

**(4)ACCESSING THE MODEL**

They had accessed the model that how is the model performing , compared the result with predefined metrics.

**EVALUATION:**

In this phase we have to evaluate the model so that we can achieve the goal of reviewing process.

**(1)EVALUATING THE RESULT**

I think the team of J P Morgan evaluated the output and performance of the metrics like accuracy, recall and F-1 score etc.to check efficiency of the model to classify legal documents.

**(2)REVIEWING THE PROCESS**

I think they reviewed the process like data preparation, model building , testing so that they can ensure that the model working perfectly if there is any need of improvement they can improve it.

**(3)DETERMINING THE NEXT STEP**

I think after that they have confirmed the model performance and have decided to deploy the model.

**DEPLOYMENT:**

Then they deployed the model in production area so that they can monitor the performance and review the project finalize the outcome.

**(1)PLANNING DEPLOYMENT**

They have made a detailed plan to deploy the model.

**(2)MONITORING AND MAINTANCE**

Then they have monitored the performance and tracked the performance if needed to improve something they improved.

**(3)REVIEWING THE PROJECT**

After the deployment they reviewed the project to access that all of the goals has achieved or not evaluated the outcome model performance again.

**(4)FINALIZING THE PROJECT**

After the reviewing project J P Morgan finalized the project, he validated the project objectives and ensured that all the goals has achieved successfully.