TRIBE-D

BUSINESS REQUIREMENT DOCUMENT (BRD)

FOR ORGANIZING UNSTRUCTURED DATA

PROBLEM STATEMENT: Each second a huge amount of data is created and collected. By large margin, most of the data that organizations collect is unstructured data that doesn't easily conform to an existing data model like structured data. This data is of no use. It's worthless without machine learning. With the help of machine learning and machine learning models unstructured data can be converted to structured one.

Say for e.g., Predicting student's performance in their academics using previous information available such as test marks, no. of absentees, parental education, time devoted by student in their studies.

OBJECTIVE: Nowadays prediction of student's performance has turned out to be an urgent desire in most of the educational institutes. Education sector in the world in adapting machine learning methods to spot struggling students thereby increasing success rate. So, the main aim of this project is to prove the possibility of training and modelling a small dataset size and the feasibility of creating a prediction model with credible accuracy rate. Modification in teaching and learning experiences personalize engagement providing an individual to adapt for a better education. The information which cannot be gleaned by human brain can easily be done with data analytics using machine learning. These assists grappled students and at the same time challenges the gifted ones.

OUTLINE: The conversion of unstructured data into structured data will be attained with the help of machine learning algorithms such as classification algorithms (Random Forest, Decision Tree, Naïve Bayes Classifier, etc). These algorithms will help us in building prediction model to predict the performance of student. Then the visualization of data is done to identify the hidden pattern and draw insights.

Steps involved:

1. Data collection
2. Data cleaning
3. Normalizing/ Standardizing
4. Data evaluation
5. Analysing → Drawing insights
6. Data modelling
7. Data visualization

TECHNICAL STACK:

Languages- Python, MongoDB
Technology- Artificial Intelligence, Machine Learning
Applications- Python Jupyter Notebook, Excel Sheet (Data collection & Data Storing)
GitHub for Project Management

SCOPE OF THE PROJECT: The ultimate goal of any educational institution is offering the best educational experience and knowledge to the students. Identifying the students who need extra support and taking the appropriate actions to enhance their performance plays an important role in achieving that goal.

SCHEDULE, TIMELINE AND DEADLINES:

S. NO.	TASK NAME	DURATION	START DATE	FINISH DATE
1	Data collection	3-4 days	28/6/21	01/7/21
2	Importing libraries	2 days	02/7/21	03/7/21
3	Loading dataset	1 day	04/7/21	04/7/21
4	Data cleaning	1 day	05/7/21	05/7/21
5	Normalizing	2 days	06/7/21	07/7/21
6	Training/ Test sets	2 days	08/7/21	09/7/21
7	Classification	3 days	10/7/21	12/7/21
8	Applying algos	3-4 days	13/7/21	16/7/21
9	Data modelling	5 days	17/7/21	21/7/21
10	Refining values	2 days	22/7/21	23/7/21
11	Accuracy check	2 days	24/7/21	25/7/21
12	Prediction	2-3 days	26/7/21	29/7/21