

## Project Design Phase-I

Date	23 September 2022
Team Id	PNT2022TMID40543
Project Name	Proposed Solution
Maximum Marks	2

### Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	<b>Goal of Prediction Analysis:</b> First there should be a clear prediction definition statement. It should be clearly say what business trend or patten and outcome they are trying to predict, for what period of time, and for how long the prediction should be valid. This will help form the statistical hypothesis a kind of conversion of the English like problem statement into a statistical problem. Unclear prediction definition problem statement often leads to prediction wrong problems.

2.	Idea/Solution description	Brainstorm within the management team to come up with statement that has clarity in the prerequisites, and assumption and deviation that are willing to be accepted as a part of the prediction analysis. Communicate the same to the data analytics team precisely, which can dissect the statement and convert into a statistical and data analytics problem.
3.	Novelty/Uniqueness	Predictive analytics is the use of data to predict future trends and events. It uses historical data to forecast potential scenarios that can help drive strategic decisions.
4.	Social Impact/Customer satisfaction	Predictive analytics are used to determine customer responses or purchases, as well as promote cross-sell opportunities
5.	Business Model (Revenue Model)	Predictive modeling is a commonly used statistical technique to predict future behavior. Predictive modeling solutions are a form of data-mining technology that works by analyzing historical and current data and generating a model to help predict future outcomes.
6.	Scalability of the Solution	i) Classification Model ii) Clustering Model iii) Forecast Model iv) Outliers Model v) Timeseries Model

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	<b>Quality of Data:</b> The quality of data we have should be accurate and reliable. Obviously, the outcome will solely depend on the data we put into the prediction. If the data is skewed, then the prediction which is dependent on it, will be skewed as well.
2.	Idea/Solution description	The first duty of a data analytics engineer is to confirm the randomization of the data given. If the given data is non-random then the correlation and relationships between the data attributes will be skewed leading to wrong outcomes. The next point is to see the standard deviation of the data after normalization. Decision about the outliers (as to whether to keep them or discard them) should be taken so that the data that is fed into the model is consistent. Also, in case of unstructured data, proper cleansing needs to be done. Appropriate Data Mining and Data Lake Processes should be followed.
3.	Novelty/Uniqueness	A novel in this case means unusual, data that are new and do not occur regularly or are simply different from the others.

4.	Social Impact/Customer satisfaction	Social Impact data is becoming an important tool for social purpose organizations when making decisions and measuring impact, but many struggles to do so. There seems to be a universal debate in the social sector about how data can be used for maximizing social impact.
5.	Business Model (Revenue Model)	Data quality is defined as: the degree to which data meets a company's expectations of accuracy, validity, completeness, and consistency. By tracking data quality, a business can pinpoint potential issues harming quality, and ensure that shared data is fit to be used for a given purpose.
6.	Scalability of the Solution	Scalability is the measure of a system's ability to increase or decrease in performance and cost in response to changes in application and system processing demands.