

▷ CRISP-DM :- This framework was originally developed by data miners in order to generalize the common approaches to defining and analyzing problem. CRISP-DM is "Problem solving frame work".

(CRISP- DM :- Cross industry Standard process of Data mining.

This frame work has 6 steps :-

- i) Business Issue Understanding
- ii) Data Understanding
- iii) Data Preparation.
- iv) Analysis / Modeling
- v) Validation
- vi) Presentation / visualization.

▷ Methodology Map :- The methodology map is a guide to determine the appropriate analytical techniques to solve a particular business question or problem.

The map outlines 2 main scenarios for business problems:

- i) Data analysis.

ii) Predictive Analysis.

- Data Analysis :- It is defined as process of cleaning, transforming and modeling data to discover useful information for business decision-making. The purpose of data analysis is to extract useful information from data and taking the decision based upon the data analysis.
- Predictive Analysis :- It describes the use of statistics and modeling to determine future performance based on current and historical data. Predictive analysis looks at pattern in data to determine if these pattern are likely to emerge again, which allows businesses and investors to adjust where they use their resources to take advantages of possible future events.

Business Problem.

Predict Outcome Data Analysis.

<u>Data Rich</u>	<u>Data Poor.</u>	<u>Geographical</u>
Numerical Continuous Time Based Linear regres. Decision Tree Forest Model	Classification Binary Logistic Regr. Decision Tree Deision Tree	A/B testing Non Binary Forest model Boosted model.
		Segmentation, Aggregation, Descriptive.

Business Problem:-

I need to know How many tickets per week on average we can expect from this client so we can ensure we have enough help desk resources in place.

Business Issue Understanding:-

The initial phase focuses on understanding the project objectives and requirements from a business perspective and then converting this knowledge into data mining problem definitions, and a preliminary plan designed to achieve the objectives. A decision model, especially one built using the Decision Model and Notation Standard can be used.

- Q. What decisions needs to be made?
- Q. What information is needed to inform those decisions?
- D. What type of analysis can provide the information needed to inform those decisions?

Example . Problem :-

→ How much electricity capacity does utility company needed to supply for any given hour ~~tomorrow~~ tomorrow?

- Firstly, we need "what decisions need to made".
- q. How do we supply the power we need?
- Second, Information to need to known.
- q. Electricity per hour of the day Tomorrow.
or
- q. How much electricity will be demanded in each hour of the day tomorrow. (This need to be predicted).
- Now, we know that we need a predictive analysis to help us obtain the data to inform the decision we need to make.

Data Understanding :-

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify

data quality problems, to discover first insights into the data, or to detect interesting subseries to form hypothesis for hidden information.

- Q. what data is needed?
- Q. What data is available?
- Q. What are the important characteristics of the data?
- Q. What factors drive electricity demand tomorrow?
for example:- Some factors that are likely to be important are the day of week, as well as the month and the year. It's also useful to know the hour of the day and the temperature at that hour and may be even the preceding hour.
- Q. How can we Predict hourly temperatures?
Now, we'll need to make predictions about the temperature at any given hour. Now we have some idea about our data needs, i.e. factors that will help us predict hourly temperatures.

D Data Preparation :-

The data preparation phase covers all activities

to construct the final dataset (data that will be fed into the modeling tool) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

- Common Steps used in Data Preparation :-

- i) Gathering :- When gathering data - we may need to collect data from multiple sources within your organization.
- ii) Cleansing :- The data set you are working with may have issues that you want to resolve prior to your analysis. This can be in the form of incorrect or missing data.
- iii) Formatting :- You may need to format the data by changing the way a date field appears, renaming a field, or even rotating the data, similar to using the vlookup function in Excel.

iv) Blending :- We may want to blend or combine the data with other datasets to enrich it with additional variables, similar to using the vlookup function in excel.

v) Sampling :- Lastly, we may want to sample the dataset and work with a more manageable number of records.

► Analysis and Modeling :-

In this Phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirement on the form of data. Therefore, stepping back on the data preparation phase is often needed.

- Important Steps :-

★ Q. Determine what methodology to use to solve the problem.

★ Q. Determine the important factors or variables

that will help solve the problem.

* Build a model to solve the problem.

* Run the model and move to the validation phase.

Example:- In last example we determined that we needed to predict the temperature for the next day using some of the factor we identified. And then we are going to predict the electricity usage using the predicted temperature as well as other factors.

So, we start by building out a predictive model that help us predict the temperature for the next day. We will create our initial model and learn during the process. 1st → Create first model and we'll validate it. Then determine if we need to update the model with addition information. Then we will repeat the process that we're comfortable with, then perform analysis).

► Validation:-

At this stage in the project you have built a model (or models) that appears to have high quality,

from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and ~~and~~ review the steps executed to construct the model to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

- Important steps :-

- * Q. Observe the key results on the model.
- * Q. Ensure the results make sense within the context of the business problem.
- * Q. Determine whether to proceed to the next step or return to a previous phase.
- * Q. Repeat as many times as necessary.

▷ Presentation and Visualization :-

Creation of the model is generally not the end of the project. Even if the purpose of the model

it to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that is useful to the customer. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data scoring (e.g. segment allocation) or data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. Even if the analyst deploys the model it is important for the customer to understand up front the actions which will need to be carried out in order to actually make use of the created models.

- Key Consideration.

- i) Determine the Best Method of presenting insights based on the analysis.
- ii) Determine the best method of presenting insights based on audience.
- iii) Make sure the amount of information shared is not overwhelming.
- iv) Use the results to tell a story to the audience.

- v) for more complex analysis, you may want to walk the audience through the analytical problem solving process.
- vi) Always reference the data sources used.
- vii) Make sure your analysis supports the decision that needs to be made.