**THE CRISP DM**

**INTRODUCTION**

The CRoss Industry Standard Process for Data Mining (CRISP-DM) is a process model that provides an organized approach for planning a data mining project. Data mining is the process of uncovering patterns and valuable information from large datasets. The CRISP-DM methodology has six phases that naturally describe the data science life cycle.

I chose to employ this methodology first, because it has been well proven to be a guardrail in assisting a Data Scientist plan, organize, and implement a data science (or machine learning) project and second, because it is an idealized sequence of events that is flexible hence allows backtracking and repeating certain actions.

Therefore, the CRISP-DM methodology will help me achieve the goal of this project in determining the states that should be prioritized to win elections

The six phases of CRISP-DM

1. Business understanding

2. Data understanding

3. Data preparation

4. Modeling

5. Evaluation

6. Deployment

**Phase 1. Business Understanding**

For any project to be successful it is very important to have a deep understanding of the client's needs. In this case I will have to take note and understand the objectives of the project which will help me realize what the client really wants. I will then access the situation by determining the available resources, requirements, and also accessing the risks and contingencies and conduct a cost benefit analysis. I will then determine data mining goals and finally formulate a project plan. Planning is vital since it enables one to select technologies and tools to be used during the project period.

**Phase 2. Data Understanding**

This phase adds a foundation to Phase 1. It drives the focus to identify, collect, and analyze the data sets that can help you accomplish the project goals.

Therefore, this phase drives me towards acquiring the necessary data and loading it into my analysis tool(SQL). I will also be able to describe the data that I have collected in terms of data format, number of records or field identities. I will be in a position to explore data and verify data quality by examining how clean/dirty the data

**Phase 3. Data Preparation**

This phase is equally important since it helps in preparing data for modelling. I will be able to determine which datasets will be used, clean data, construct data, integrated from multiple sources and format data.

**Phase 4. Modeling**

Here I will be able to assess various models based on several different modelling techniques

I will be able to select modelling techniques by determining which algorithms to try e.g. regression. I will also be able to generate test designs, build models and assess those models based on the domain knowledge, the predefined and the test design. With CRISP-DM lifecycle, I can iterate modelling and assessment continuously until I find a ‘good enough’ model.

**Phase 5. Evaluation**

Whereas the Modeling face focuses on technical model assessment, the Evaluation Phase will assist me to broadly meet the business and what to do next. At this point I will be able to Evaluate the results, review the process and Determine the net step i.e. whether to proceed to deployment, iterate further or initiate new project

**Phase 6. Deployment**

Deployment phase can be as simple as generating a report or as complex as repeatable data mining depending on the requirements of the project. Hence this phase will enable me to present the requirements to the client.

I will be able to plan the deployment, plan monitoring and maintenance, produce a final report and finally review the project to see whether it went well or it needs to be improved.

**Conclusion**

Indeed, the flow of CRISP-DM is precise and more flexible as you can iterate quickly. For instance, in this project of determining which Grand Electors will win the presidential elections the CRISP-DM will ensure that the projects run with a flow and be successful at determining whether there are state that should be prioritized and why.