**DUX (Data utility with Excellence)**

**Topic**

Generic analysis of a dataset and development of a basic ML model based on user query in high-level language.

**Description**

Developing insights for a given dataset play a crucial role in addressing the real-world data problems. The process of gaining insights is usually time-consuming as it involves writing code to plot various graphs for different data points, understanding the statistics like outliers, range, distribution type, interquartile range and many more. Our aim is to streamline this process to the highest possible extent. What if the developer/user can understand all of this by typing-in the query in a high-level language (English) and gain all those insights? Let’s not stop there, what if we could predict label values based on the features, apply ML, just by processing the given high-level language statement? And this applies for almost every kind of dataset.

**Area of specialization**

* Machine Learning (Main-stream)
* Data Science
* Database
* Front-end

**Nature of project**

*What novelty do you see in the project?*

A conventional data scientist follows seven steps to develop a solution in-terms of Machine learning: Preprocess data, understand data, gain insights, develop a basic model, fine-tune parameters and features, scale the model and integrate. With this project, first ***FOUR*** steps in the conventional process would be optimized to the highest extent. A data scientist can preprocess, understand data, gain insights and even develop a basic model with performance evaluation by just typing-in the queries in English language which streamlines the process of developing a basic model.

*Feasibility of project*

This project is highly feasible with the available technology. This project integrates various technologies to come-out with a unique solution. The only

constraint would be an internet connection.

*Research gap*

Students have already working on a prototype which needs exploration or research of some services to scale this prototype to a higher level for generic analysis. Estimated time for research would be 3 months. A first iteration can be generated by the next 3 months.

*Outcomes*

1. Streamline the process of data analysis.
2. Maintaining a global perspective for solution developed.
3. Data analysis without a single line of code / without any programming knowledge.

4. Train a holistic model without any local-installed libraries

5. Most efficient Predictions based on the model trained, operating system independent.