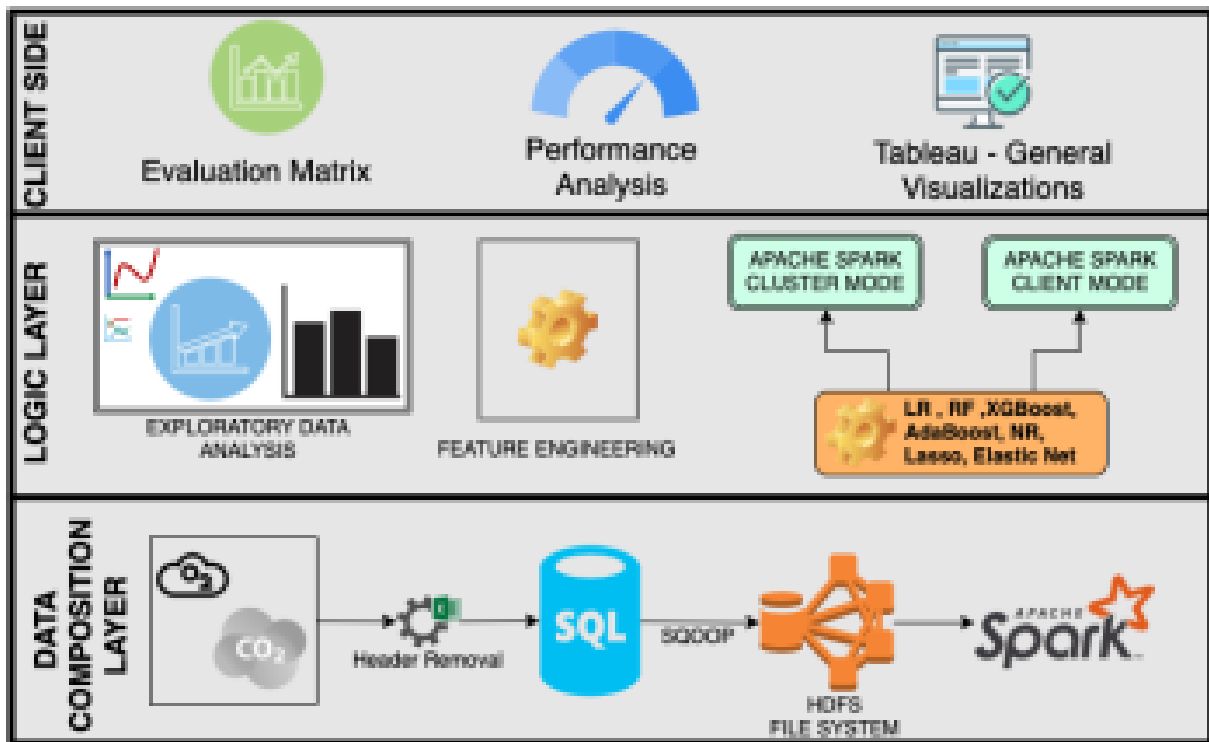# PROJECT: AIR QUALITY ANALYSIS IN TN.

ARCHITECTURE DIAGRAM

 The process flow architecture was designed for the analysis air quality in Tamil nadu to increase the robustness and accuracy of the research. And also architecture the polluntants level in the environment. It was composed of various elements and technologys. The architecture has a three tier structure.

The first layer in the architecture is data composition layer where the data is imported Data extraction is the process of obtaining raw data from a source and replicating that data somewhere else. The raw data can come from various sources, such as a database, Excel spreadsheet.

In Data cleaning phase, the data are detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

In load stage, the extracted data and the transformed data are loaded into the target database. To make the data load efficient, it is necessary to index the database and disable the constraints before loading the data.

The second layer in the diagram is logic layer which acts as the core of the entire project. The first step in the exploration data analysis which refers to the initial step in data analysis in which data analysts use data visualization and statistical techniques to describe dataset characterizations, such as size, quantity, and accuracy, in order to better understand the nature of the data

Data exploration techniques include both manual analysis and automated data exploration software solutions that visually explore and identify relationships between different data variables, the structure of the dataset, the presence of outliers, and the distribution of data values in order to reveal patterns and points of interest, enabling data analysts to gain greater insight into the raw data.

Next, Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available data to build more accurate and efficient machine learning models.

The success of machine learning models heavily depends on the quality of the features used to train them. Feature engineering involves a set of techniques that enable us to create new features by combining or transforming the existing ones. These techniques help to highlight the most important patterns and relationships in the data, which in turn helps the machine learning model to learn from the data more.After the completion following steps make link between the layers.

 Finally the third layer called as the client side is where the results obtained from the implementation part is presented and discussed using various tools. The main goal of the research is to predict the AQI with respect to the concentration of PM2.5 in the atmosphere of Tamil naduwhich is only possible with a strong architecture design.

In this architecture KDD methodology used is avtraditional design is modified according to the steps followed in this research . KDD is one of the methodology used in the area of machine learning to develop different models. Data mining is a part of the KDD process. Following a clearly modified methodology approach helps in building a robust model by proper implementation of various stages at the proper time, The model can only be developed with prior understanding of the domain and client goals. The modified methodology consist of five stages. The methodology is structured in such a way that each stage is given equal importance, starting from data collection which is the first stage of the research, second is the data pre-processing, third data transformation which includes feature engineering, fourth the implementation stage and finally the result