**IMPLEMENTATION:**

**MODULES:**

* **User**
* **Admin**
* **Data Preprocess**
* **AutoML**

**MODULES DESCRIPTION:**

**User:**

The User can register the first. While registering he required a valid user email and mobile for further communications. Once the user register then admin can activate the customer. Once admin activated the customer then user can login into our system. User can do the data preprocess. At the time of data preprocess the h2o auto Ml server will start automatically and initiate the data from dataset of the adult’s data. This data publically available at shap server in the central repository. In the h2o models will load automatically in the project and will split the data our requirements. The files will be pickled and stored in file path locations. Later user can test the salary vs experience dataset. Here user can give the dynamically test split size. Based this size the salary dataset can split and will train to our model and fetch the predicted results. Use can compare the original and predicted results.

.

**Admin:**

Admin can login with his credentials. Once he login he can activate the users. The activated user only login in our applications. The admin can set the training and testing data for the project dynamically to the code. When h2o servers starts we can see the all models loading process. Admin can also view the user perfumed results. The test size and acquired scores also displayed in the admin page.

**Data Preprocess:**

Data pre-processing guarantees the delivery of quality data derived from the original dataset. It is an important step due to the unavailability of quality data as a large portion of information generated and stored is usually semi-structural or even non-structured in form. However, even though it is a crucial part of any machine learning pipeline, it is reported to be the least enjoyable part, with authors stating that 60-80% of data scientists finding it to be the most mundane and tedious job. In AutoML, certain data-preprocessing operations are hardcoded, which are then applied to a given dataset in certain combinations such that the overall clarity and usability of the data increases. We have largely classified these operations into the following categories based on our surveys of recent papers.

**AutoML:**

H2O is a fully open-source, distributed in-memory machine learning platform with linear scalability. H2O supports the most widely used statistical & machine learning algorithms, including gradient boosted machines, generalized linear models, deep learning, and many more. We suggest The incorporation of data-mining methods as they can deal with such unformed data. This can allow AutoML pipelines to create models capable of learning from Internet sources. In feature engineering, it should be noted that most methods used until now adhere to supervised learning. However, dataset specificity is high, and therefore, AutoML pipelines should be as generic as possible to accommodate the diverse datasets.