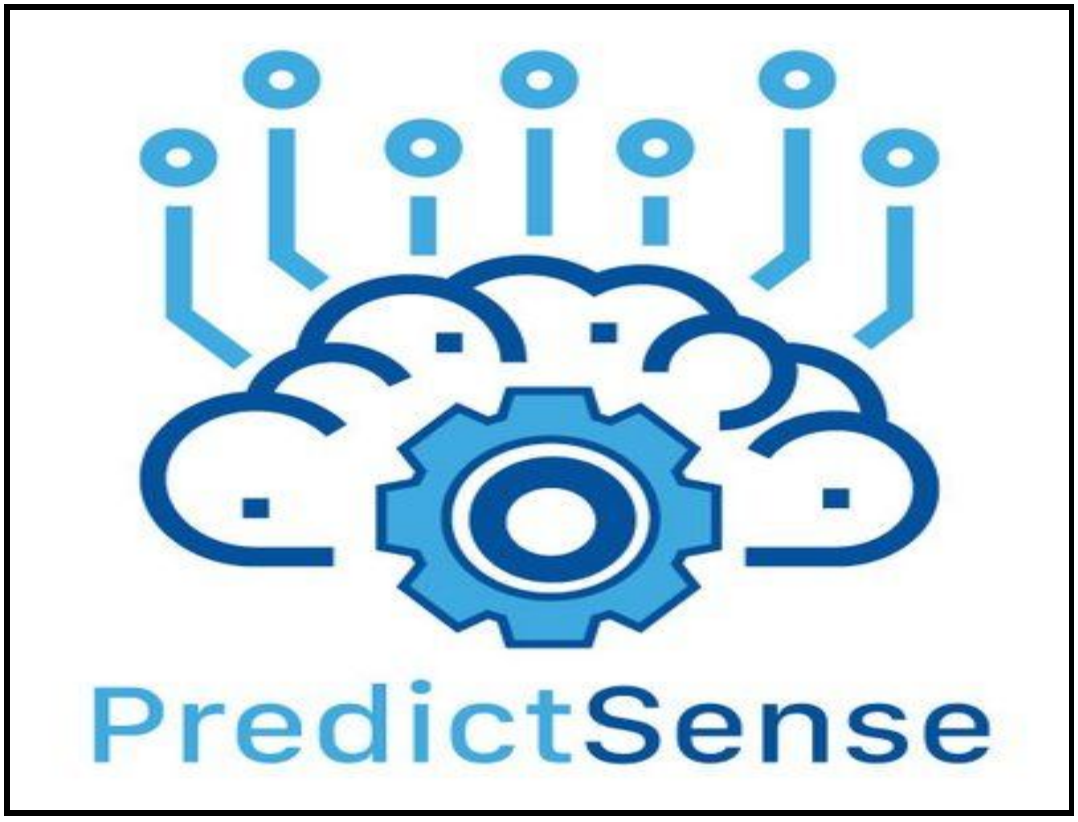
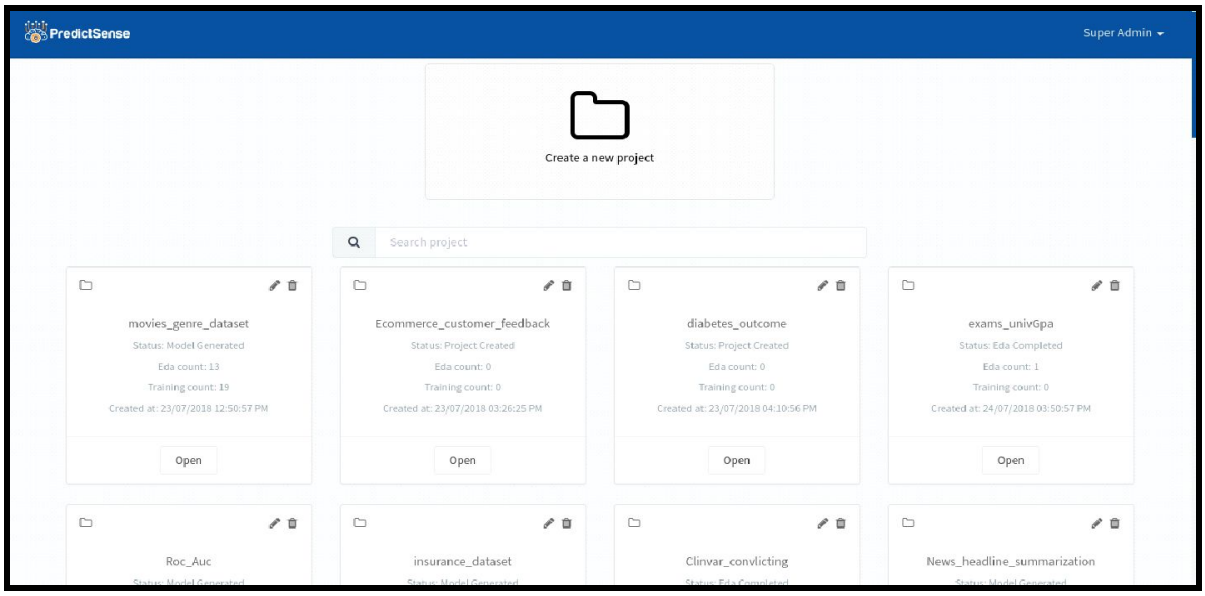
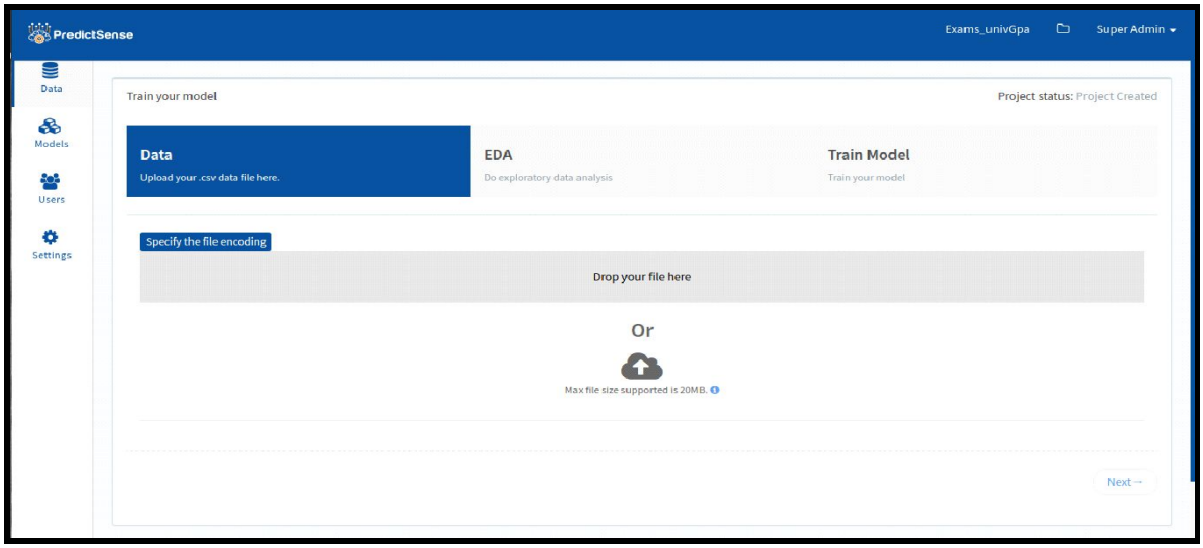

PredictSense

# PROJECT

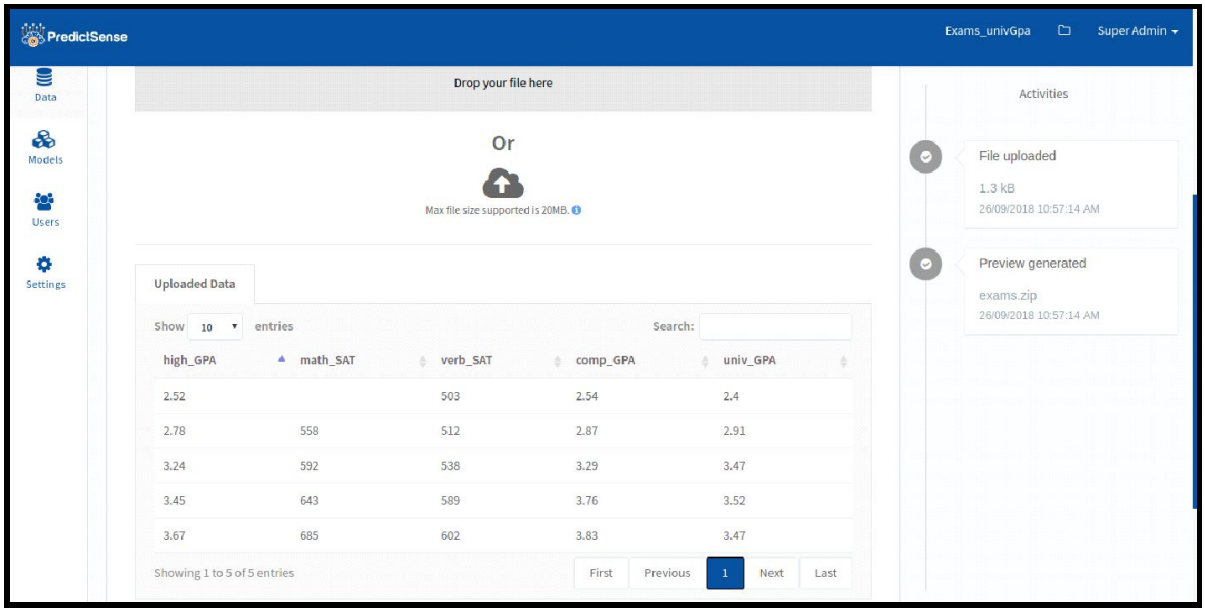



- Users can create a project by clicking on **“Create a new project”** and providing a project name according to the dataset and then click on **“Create project”**.

# DATA

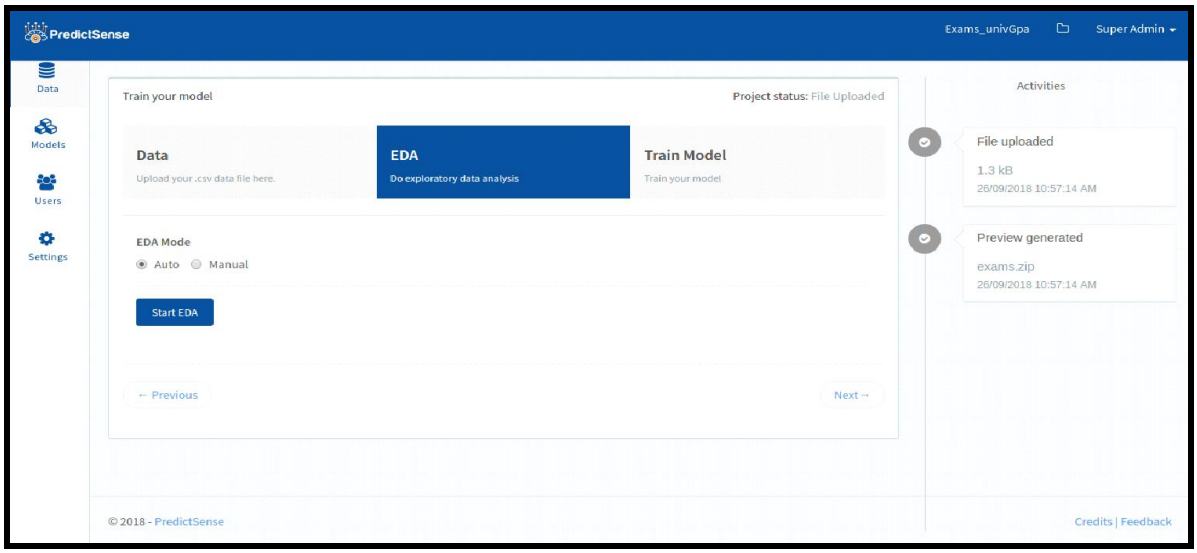



- After creating the project, Data page will be displayed where the Users will upload his dataset.

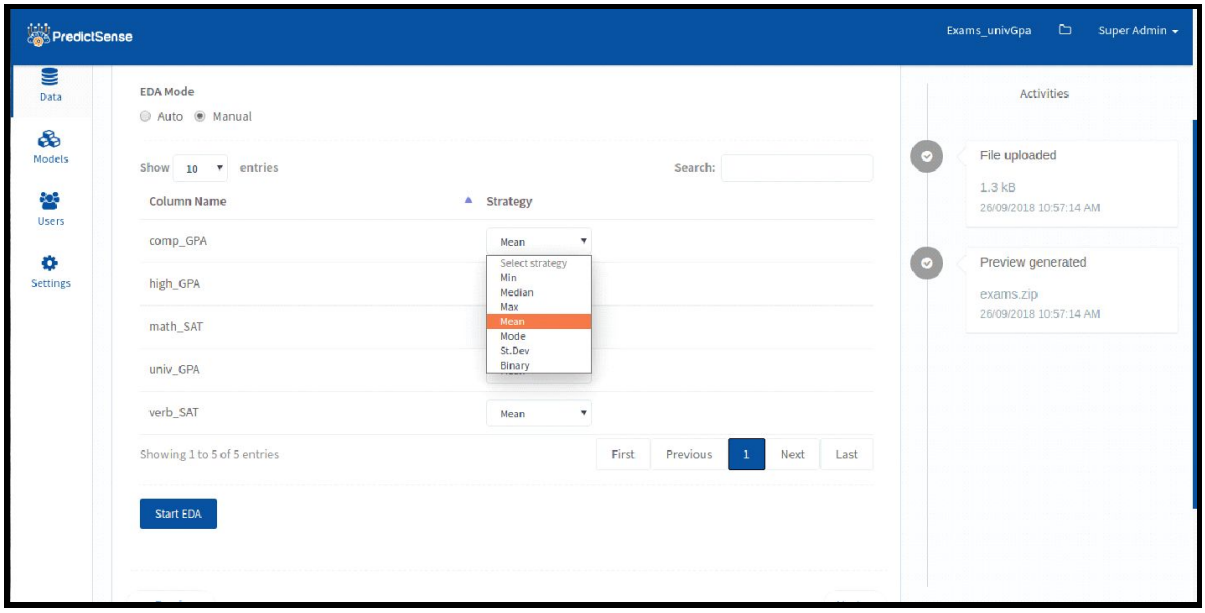



- ❖ After uploading the dataset, Users can preview dataset and file details.
- ❖ It will support **CSV, XLS, XLSX, PICKLE,Compressed(ZIP,RAR)** files in datasets.
- ❖ Then Users will click on **"Next"** to move to EDA.

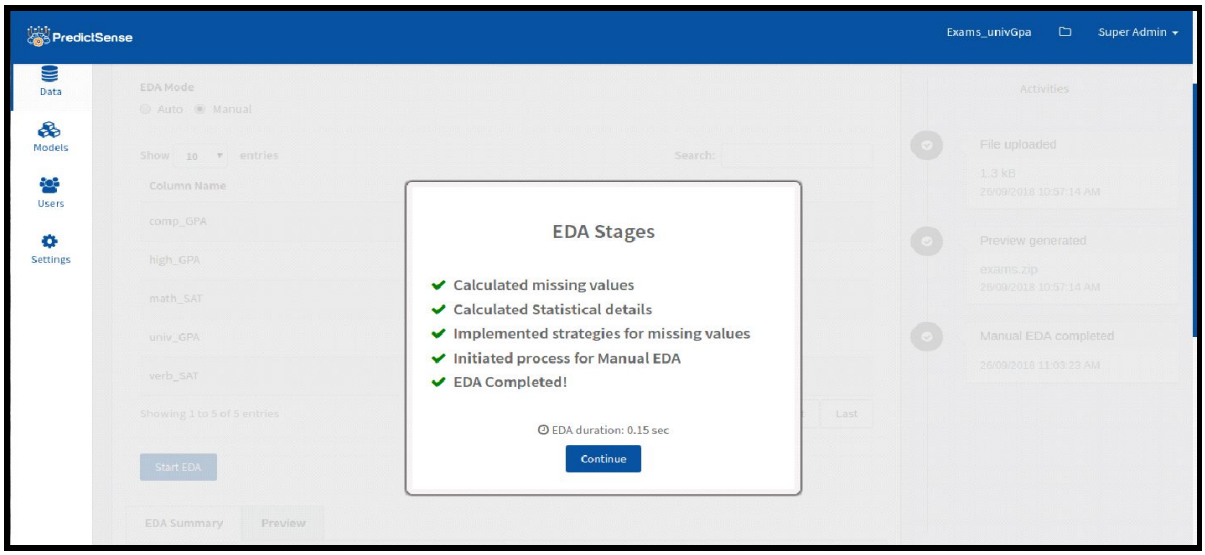
## EDA




- ❖ In Case of Manual EDA, Users can impute the missing values according to the preferred strategy and for Auto EDA, PredictSense will select the best strategy.

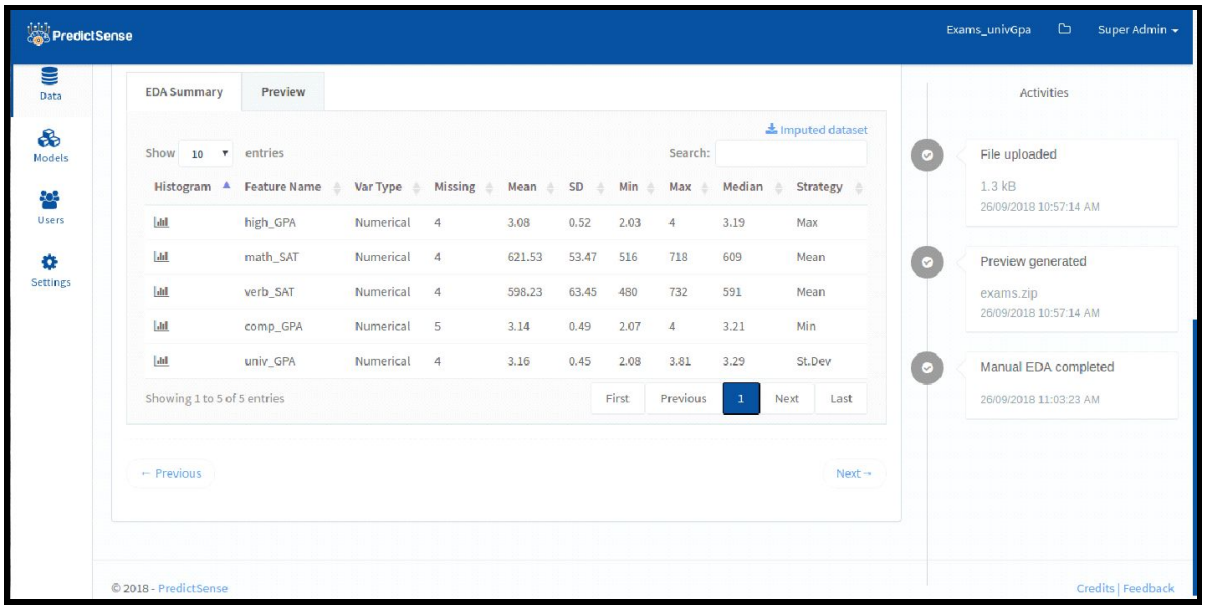



- The Statistical Strategies used are Min, Median, Max, Mean, Mode, Standard Deviation, and Binary.
- After selecting all the required strategies, click on **"Start EDA"** to proceed.

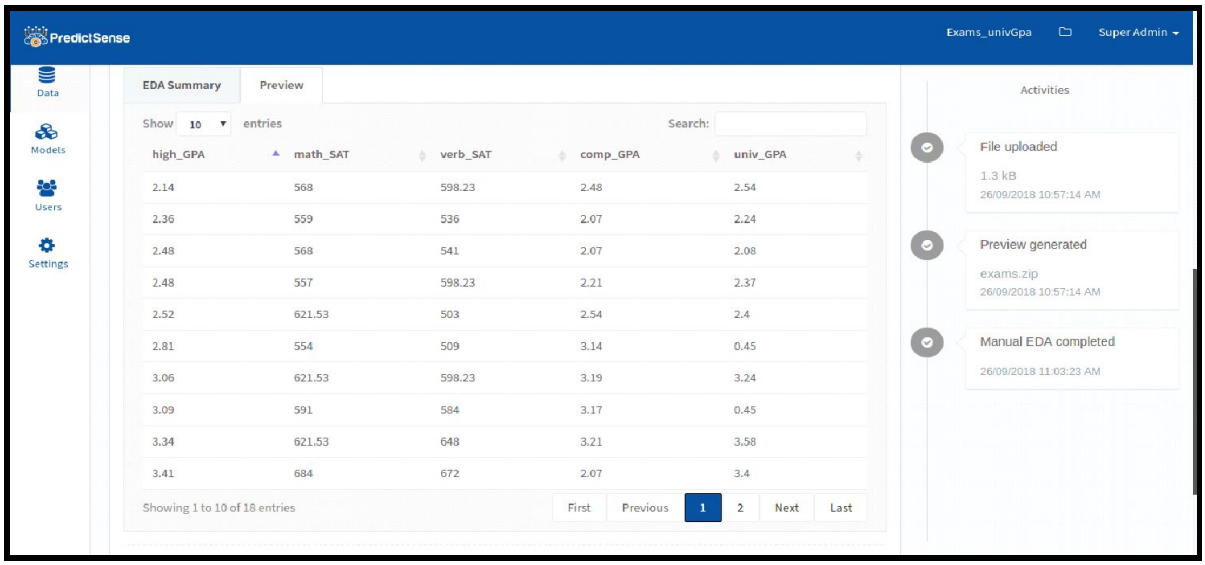



- Click on **"Continue"** to view the EDA Summary.

# EDA SUMMARY

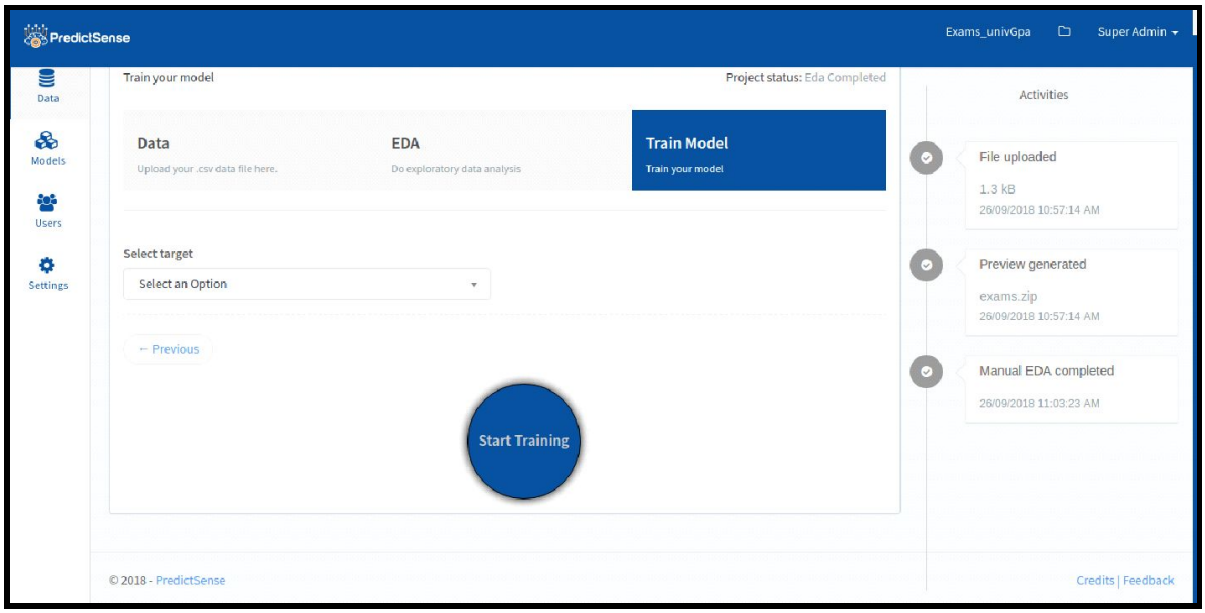



- EDA Summary contains the statistical details about the data, giving the user an insight into the data.
- The Histogram on the left provides the visualization of data across every the feature, revealing the pattern and its distribution over the range.
- Move to **"Preview EDA"** to view the columns imputed with their preferred strategies value.

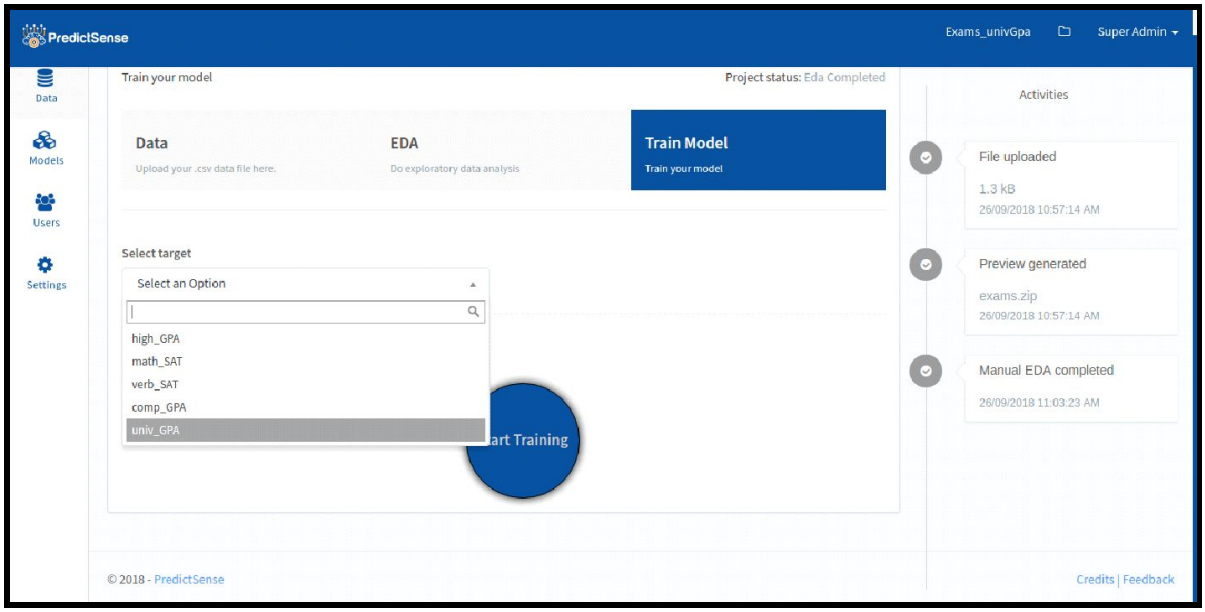



- Click on **"Next"** to move to TRAIN MODEL.

# TRAIN MODEL



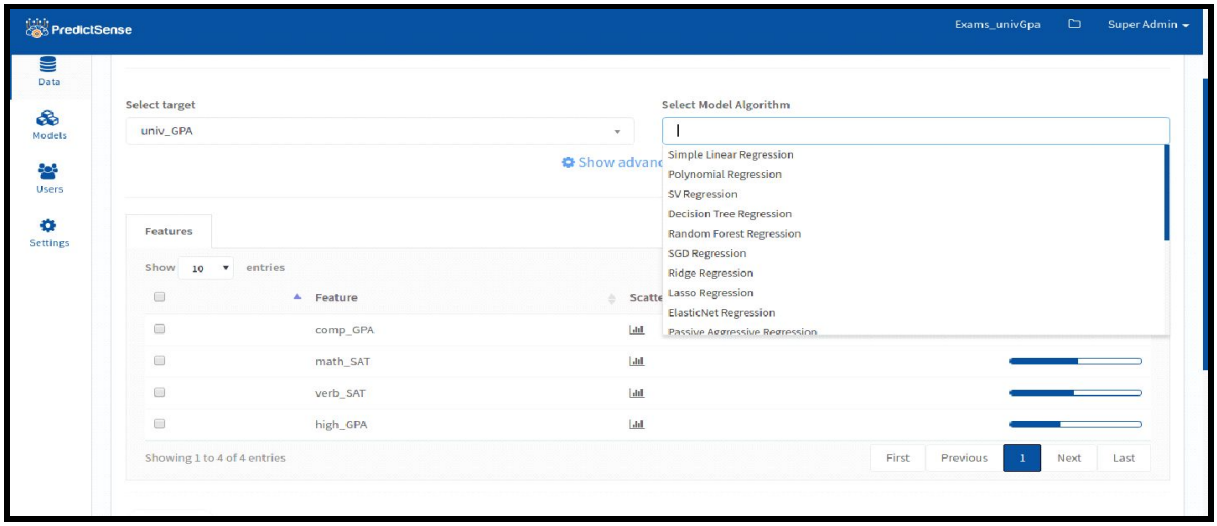

- Users can select a target feature for which prediction is to be done.
- This can be selected from **"Select target"** drop-down.




- After selecting a target PredictSense can automatically detect the type of problem, Classification/Regression. Will authenticate its pertinent algorithms and will display a list of algorithms in algorithm drop-down.

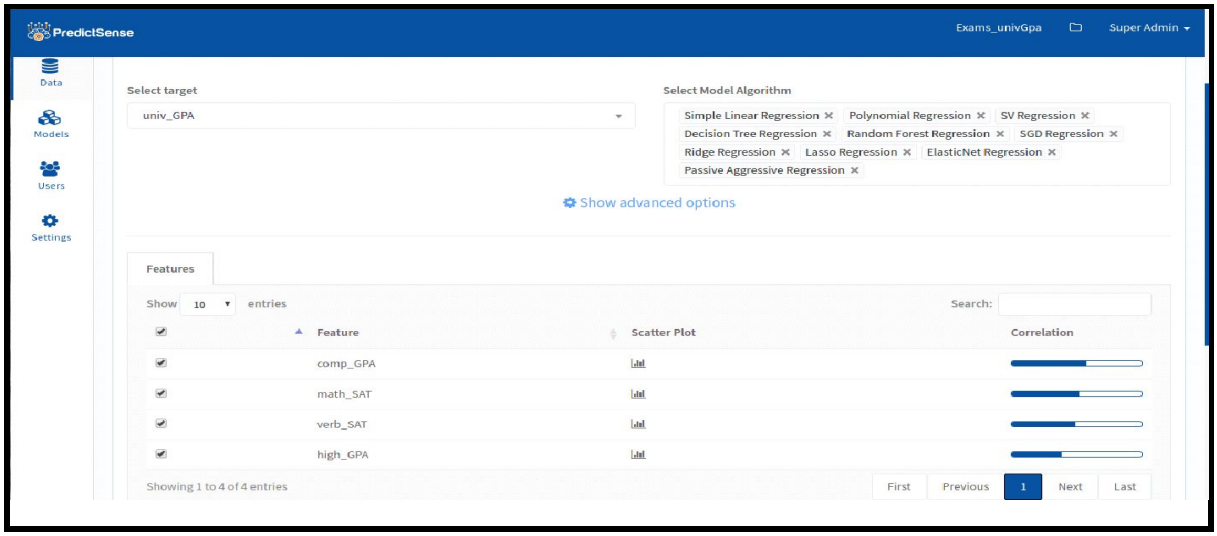
- Accordingly, from the extensive set of Algorithms, the algorithms specific to the problem will be populated.

## MODEL SELECTION



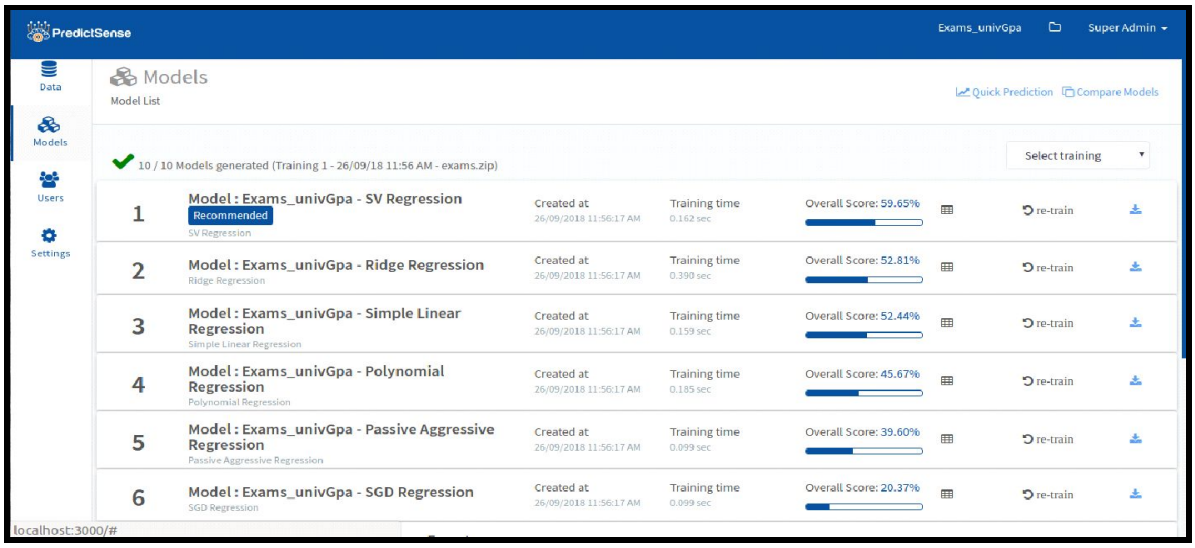

- After multi-selecting algorithms, go for feature selection where it will display all the independent features and these features are sorted on the basis of the correlation with the target variable making it easier for the Users to pick the independent features.

## FEATURE SELECTION



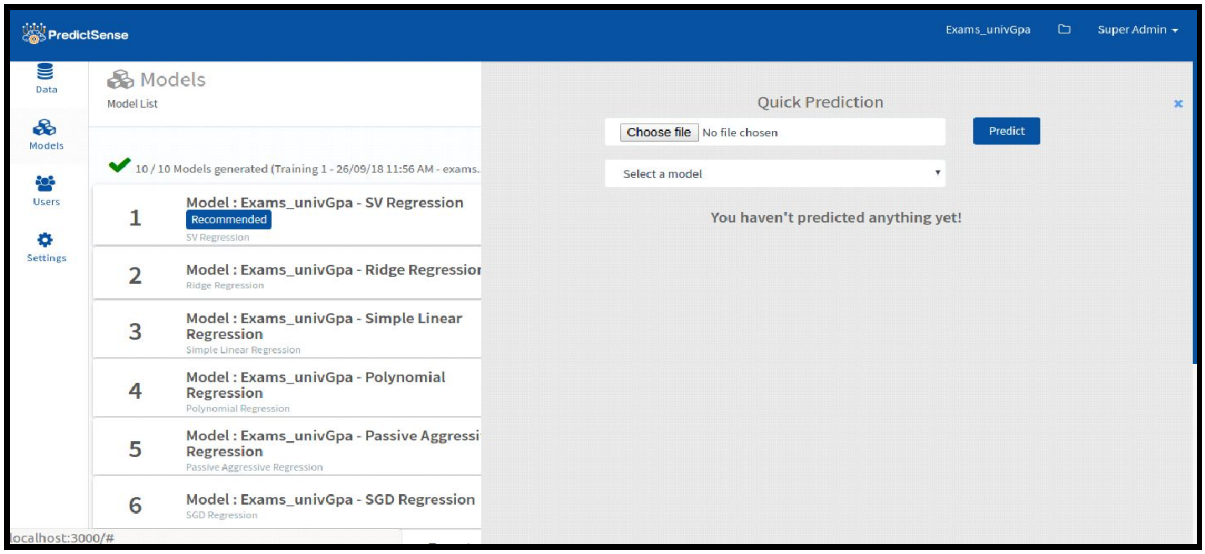

- After selecting the correlated features, click on **“Start Training”** to frame models for the selected algorithms.

# MODELS



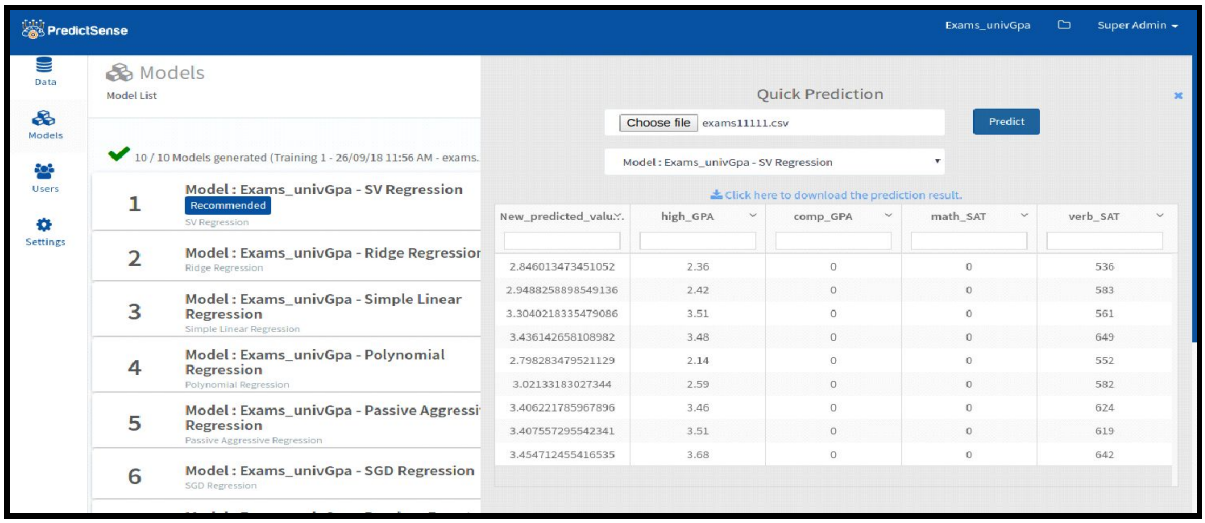

- Once the data is trained and models are generated, the models will be ranked on the model Leaderboard page.
- The models are ranked based on their performance metrics.
- These Models are deployment ready and can be downloaded within the Predict Sense Toolkit.

# QUICK PREDICTION




- Here, the user needs to upload his dataset on which the predictions are required and choose a model for prediction and click on **"Predict"** to get his predicted result.

## PREDICTED RESULTS

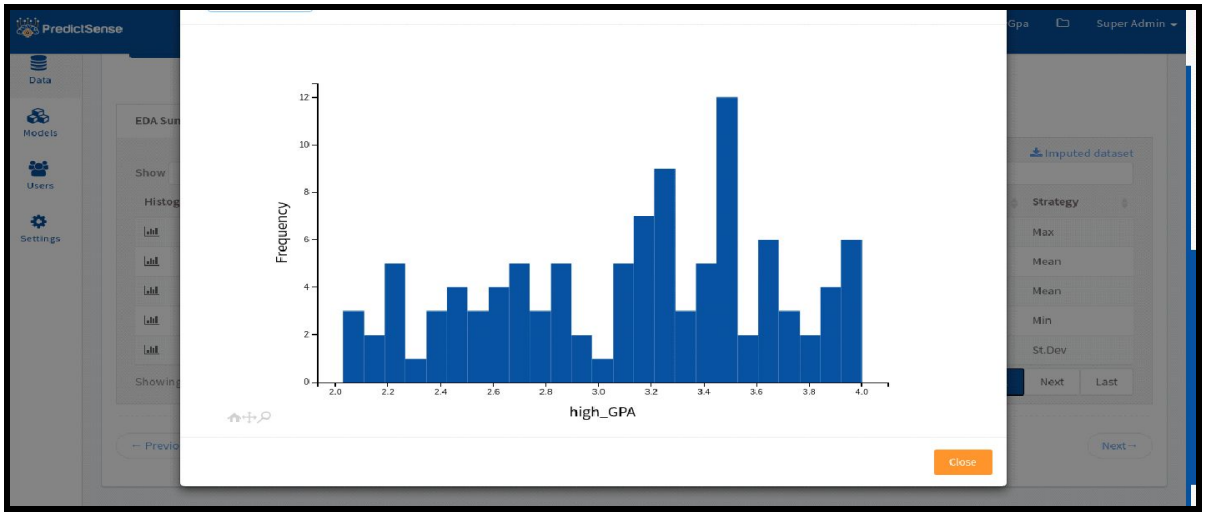



- Users can download the predicted results for their convenience.

## DATA VISUALISATION

- In Data visualization, Users can evaluate their data in the form of graphs and charts.

## HISTOGRAM

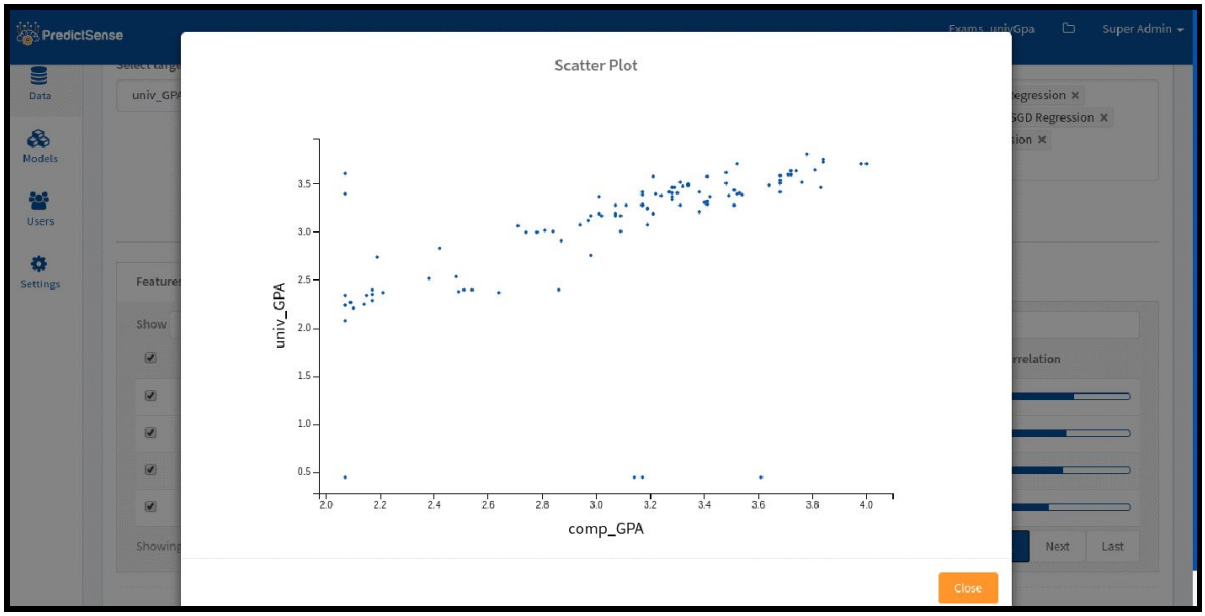



- The histogram will help to visualize how their data is frequently distributed which makes it easy to identify the outliers in the early stage. In the case of categorical Data, it will show how balanced is his dataset.

## WORD CLOUD

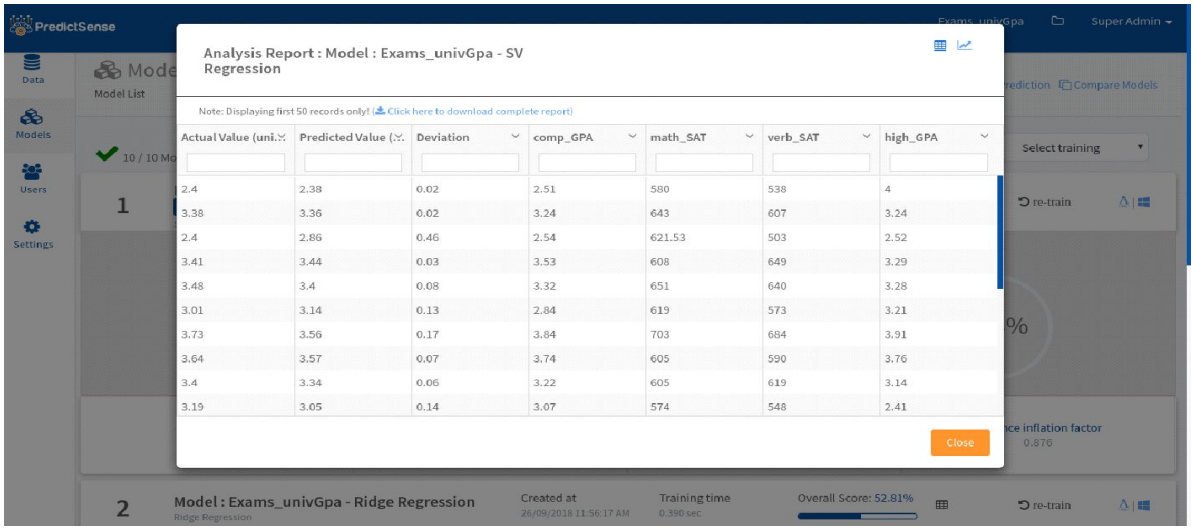
- Word cloud will display the most frequently used words in the text column, which will be assisting the Users to identify a correlation with their target.

## SCATTER PLOT



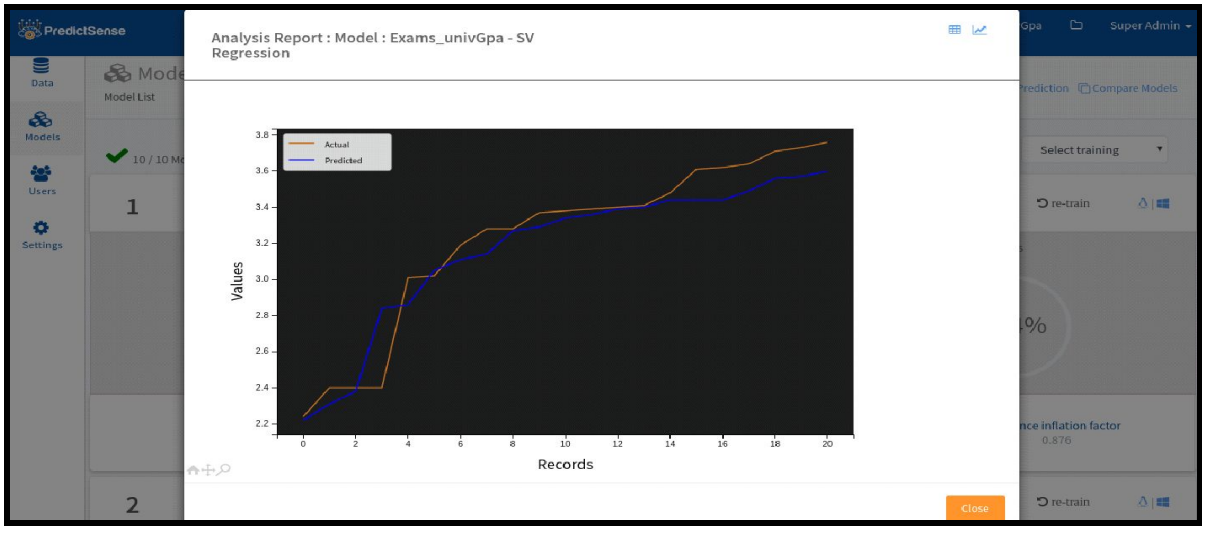

- Scatter plots help to identify how one variable is affected with another variable and linearly correlated with each other. It will also help to identify the potential root cause of problems at an earlier stage.

## ANALYSIS REPORT



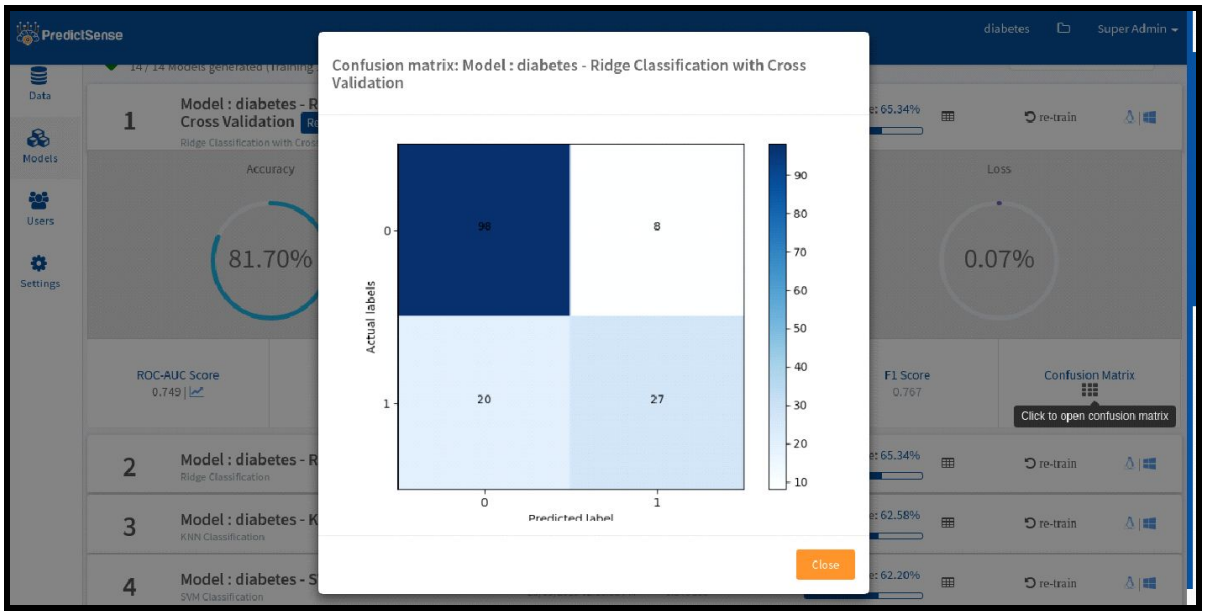

- Analysis report will assist to evaluate your actual and predicted results for the dataset of every model, It can be downloaded for your convenience.

## ANALYSIS REPORT VIZUALISATION




- Analysis Report Visualization gives an idea of plotting the results on the graph which will be displayed between the actual and predicted value of the model.

# CONFUSION MATRIX




- The Confusion Matrix is plotted between the actual and predicted value for categorical data, It will show the True and False positive count of the target data.