PredictSense

User instructions

WINJIT TECHNOLOGIES PVT. LTD.

Contents

1 PredictSense 7

1.1 Overview 7

1.2 Create a project 8

1.3 Predictive modeling 9

1.4 Forecasting 11

1.5 Clustering 12

1.6 Create project from existing sample projects 13

2 Home 14

3 Data 15

3.1 Data Connection 15

3.1.1 File upload 17

3.1.2 Pull from URL 18

3.1.3 Pull data from mysql, mssql, postgresql 19

3.1.4 Database connection configuration 19

3.1.5 MSSQL , MySQL , PostgreSQL – 20

3.1.6 SFTP connection configuration 21

3.1.7 Pull from SFTP server 22

3.1.8 Snowflake 23

3.1.9 Amazon s3 24

3.1.10 Google bigquery 25

3.1.11 Create folder 26

3.1.12 Sample Datasets 26

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.2 Data flow 27

3.2.1 Interceptor data flow 27

3.2.2 Direct flow 30

3.3 Data set 31

3.3.1 Data preview 32

3.3.2 Detailed report 32

3.4 3.3.3 Merge dataset 33

3.3.4 Row wise merge 33

3.4.1 Column wise merge 34

4 Text analysis 36

4.1 Cleaning/transformation - 38

4.2 POS tagging- 39

4.3 NER 41

4.4 Sentiment analysis- 42

4.5 Text summarization 43

4.6 Word Frequency 44

4.7 Word Embedding – 44

4.8 Rule Based Matching – 46

4.9 Save experiment 48

5 Upload model 50

6 EDA module 53

6.1 EDA details 55

6.2 Manual selection 56

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

6.3 Feature analysis 56

6.4 Imputation Feature 58

6.5 Share report 58

6.6 Target distribution 61

6.7 Split Dataset 62

7 Train model 64

7.1 Training steps 64

7.2 Training steps with advanced options 65

8 Models leaderboard 66

8.1 Model tuning 68

8.2 Delete training 71

8.3 Delete model 72

9 Quick prediction 73

9.1 Single row input prediction 73

9.2 File upload prediction 74

10 Clustering model 77

10.3 EDA operation 78

10.4 Train model 79

10.4.1 Training steps 79

10.4.2 Training steps with advanced options 81

10.5 Models leaderboard 83

10.6 Quick prediction 90

10.6.1 File prediction 90

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

10.6.2 Single row input prediction 90

10.7 Pipeline 91

11 Timeseries model 92

11.1 Project management page 92

11.2 Data connection 92

11.3 EDA operation for timeseries 93

11.4 Train model 94

11.4.1 Training steps 94

11.4.2 Training steps with advanced options 96

11.5 Models leaderboard 97

11.6 EDA operation for multiple timeseries:- 99

11.7 Train model for multiple timeseries 101

11.8 Models leaderboard 104

11.9 Pipeline 106

12 Model deployment using PredictSense 107

12.1 Configure model deployment setting 107

12.2 Deploy model 108

13 Inference Engine 110

13.1 Download Inference Engine skeleton for docker 110

13.2 For Docker Toolkit 110

13.3 Dashboard 112

13.4 Manage 113

13.5 Monitoring 114

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13.6 Predictions 115

13.7 Drift 116

13.8 Retrain model 120

13.9 Scheduler module 123

13.10 Alert 126

13.11 Connection 127

13.12 Download Docker Toolkit Lite 128

14 Drift Analysis 132

14.1 Drift settings 132

14.2 Generate drift report 134

14.3 Drift report module 135

14.4 Feature wise drift report 136

15 Notebook 137

15.1 Notebook functionalities 138

15.2 Reading data 139

15.3 Data analysis 140

15.4 Create shareable reports 141

15.5 Share functions 142

15.6 Load shared function 143

15.7 Save data 143

16 Report 145

17 Ticket, Guided tour, Feedback and Profile 150

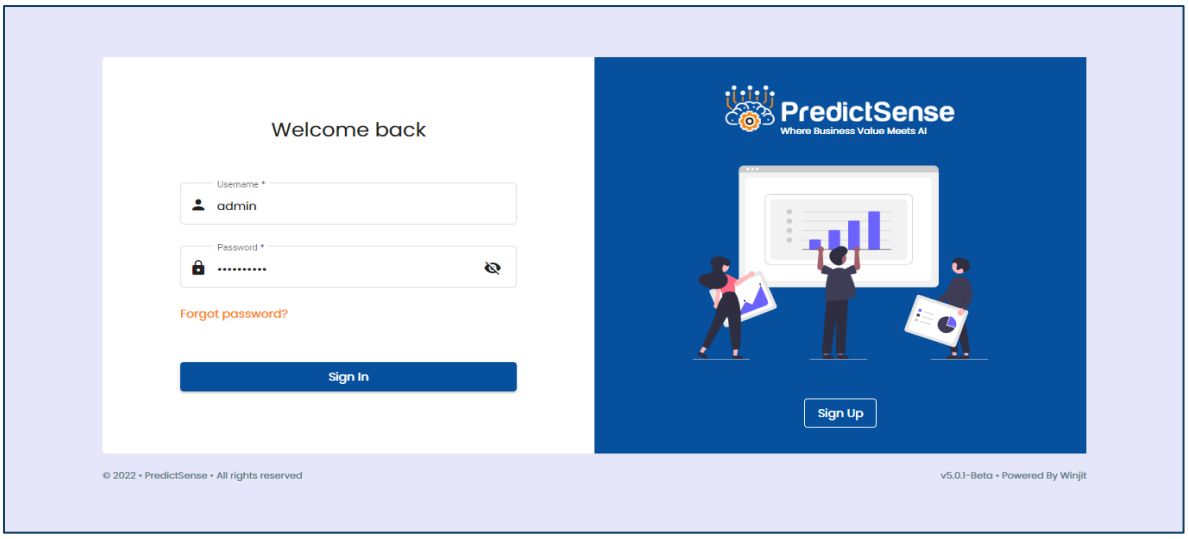
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

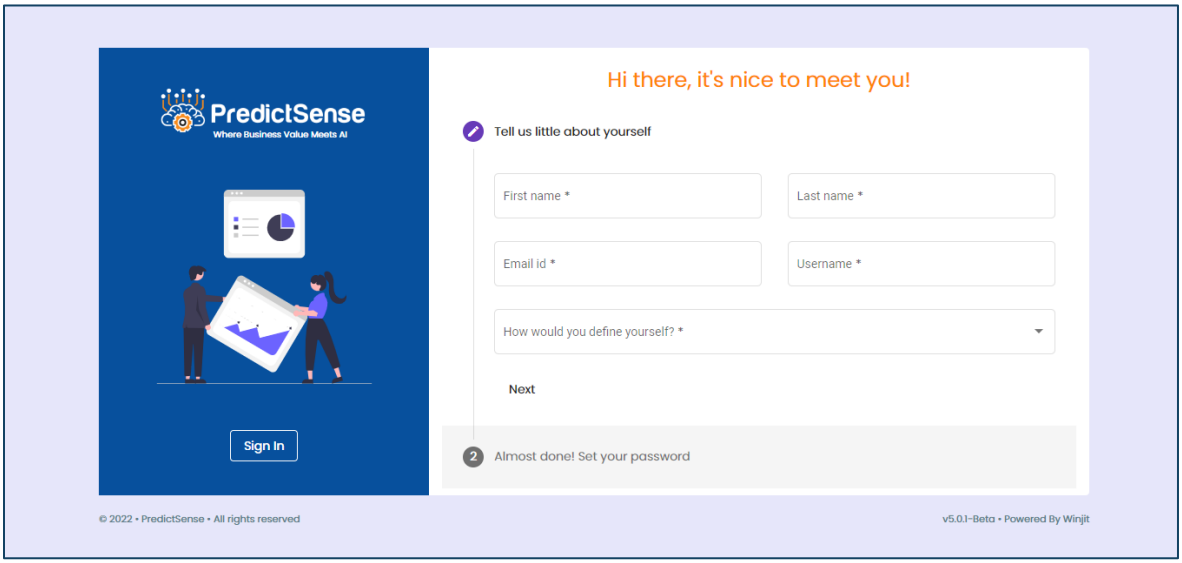
1 PredictSense

1.1 Overview

'PredictSense', is an automated machine learning platform built on an open API structure using efficient algorithms. It helps you to solve complex real -time business problems with its high-power algorithms in a very less amount of operational time. It builds predictive models which helps you to take precise and optimum business decisions

Let's get started with PredictSense:





1.2 Project Management Page

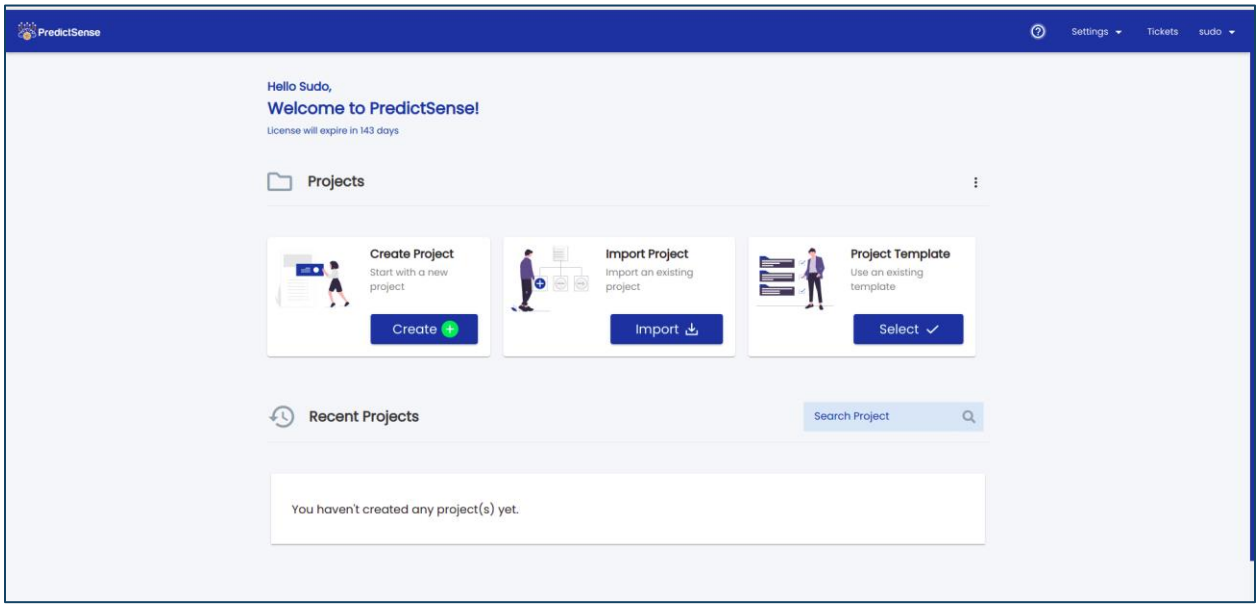
1. Create a new project

2. Import existing project

3. Create project from existing templates

4. Displayed the recent created projects

5. Create Ticket and view ticket



Project Management Page

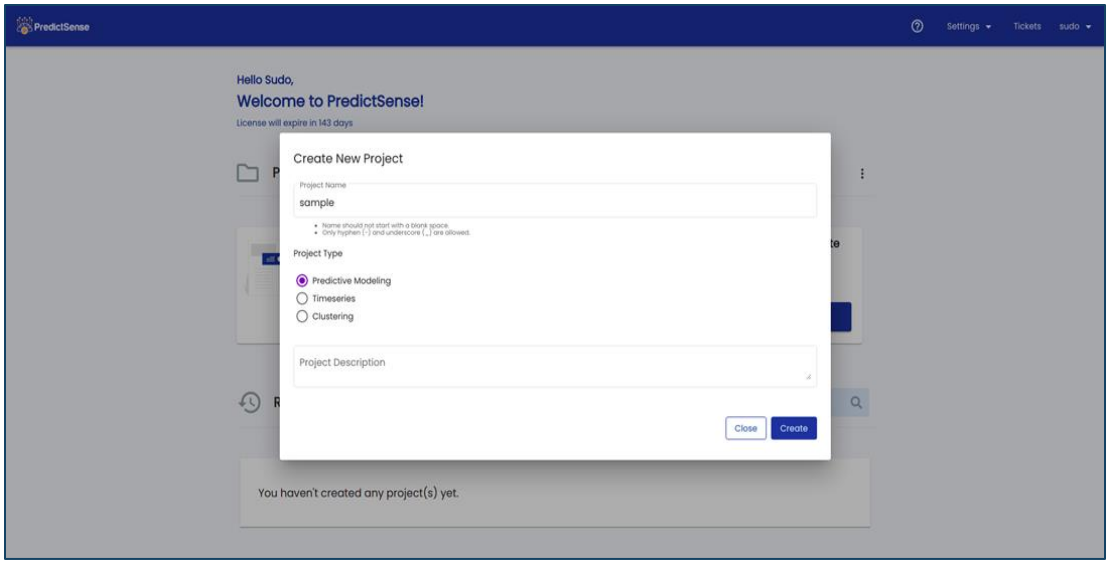
1.2 Create a project

User can create project with three different machine learning problems. Create project is segregated in to three parts Predictive modeling, Forecasting, Clustering

Steps to create project-

1. Entering **"Project name"**
2. Selecting **"Project type"** & **"Description"**
3. Clicking on **"Create project"** button.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Create Project

1.3 Predictive modeling

Predictive modeling is the general concept of building a model that is capable of making predictions. Typically, such a model includes a machine learning algorithm that learns certain properties from a training dataset and use them for predictions.

Predictive modeling can be divided further into two sub areas: Regression and classification. Regression models are based on the analysis of relationships between variables and trends in order to make predictions about continuous variables, e.g., the prediction of the maximum temperature for the upcoming days in weather forecasting.

In contrast to regression models, the task of classification is to assign discrete class labels to particular observations as outcomes of a prediction. To go back to the above example: A pattern classification task in weather forecasting could be the prediction of a sunny, rainy, or snowy day.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

Regression algorithms:

1. Linear regression
2. Polynomial regression
3. Decision tree regression
4. Random forest regression
5. SGD regression
6. Ridge regression
7. Lasso regression
8. ElasticNet regression
9. Passive aggressive regression
10. ARD regression
11. MLP regression
12. Kernel ridge regression
13. nuSV regression
14. LassoLars regression
15. Ridge CV regression
16. Bayesian ridge regression
17. TheilSen regression
18. XGBoost regression
19. Artificial neural network regression
20. LightGBM regression
21. SVM regression

Classification algorithms:

1. Logistic regression
2. Logistic regression with cross validation
3. KNN classification
4. Gaussian NB classification

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

5. Decision tree classification
6.Random forest classification
7. Extra tree classification
8.Ensemble extra trees classification
9.MLP classification
10. Nearest centroid classification
11. Ridge classification
12. Ridge classification with cross validation
13. Passive aggressive classification
14. Artificial neural network classification
15. XGBoost classification
16. LIghtGBM classification
17.SVM classification

1.4 Forecasting

Forecasting is the process of making predictions of the future based on past and present data and most commonly by analyzing trends. A common place example might be estimation of some variable of interest at some specified future date

Forecasting algorithms:

1. Sarimax
2. Prophet

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

1.5 Clustering

Clustering is a Machine Learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to classify each data point into a specific group. Data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis.

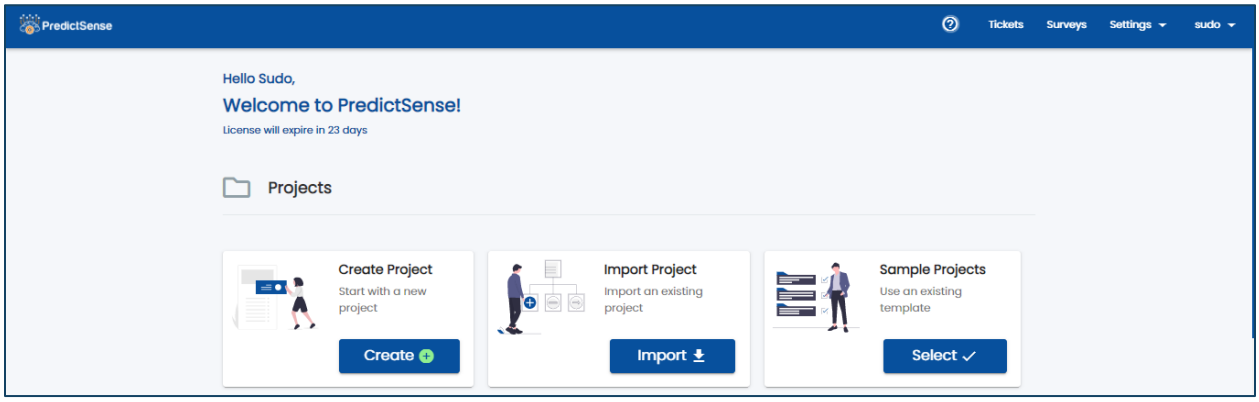
Clustering algorithms:

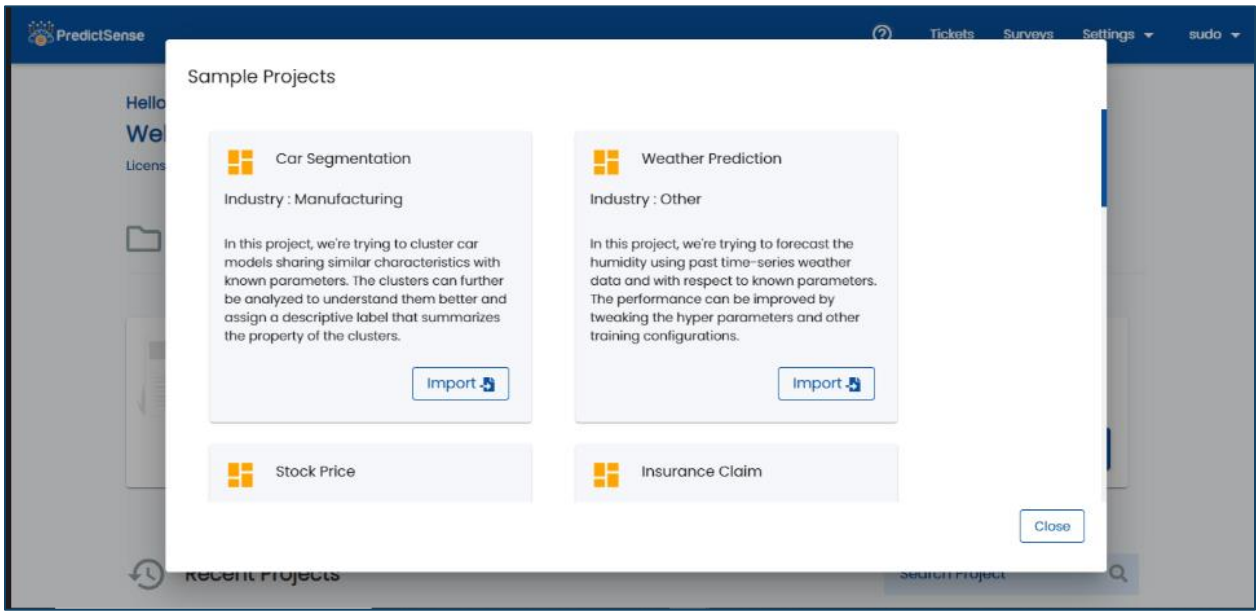
1. DBSCAN
2. KMeans
3. Agglomerative

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

1.6 Create project from existing sample projects

Application allow user to select sampled projects from sample project option. where user can find default sample project created and directly import same in project.



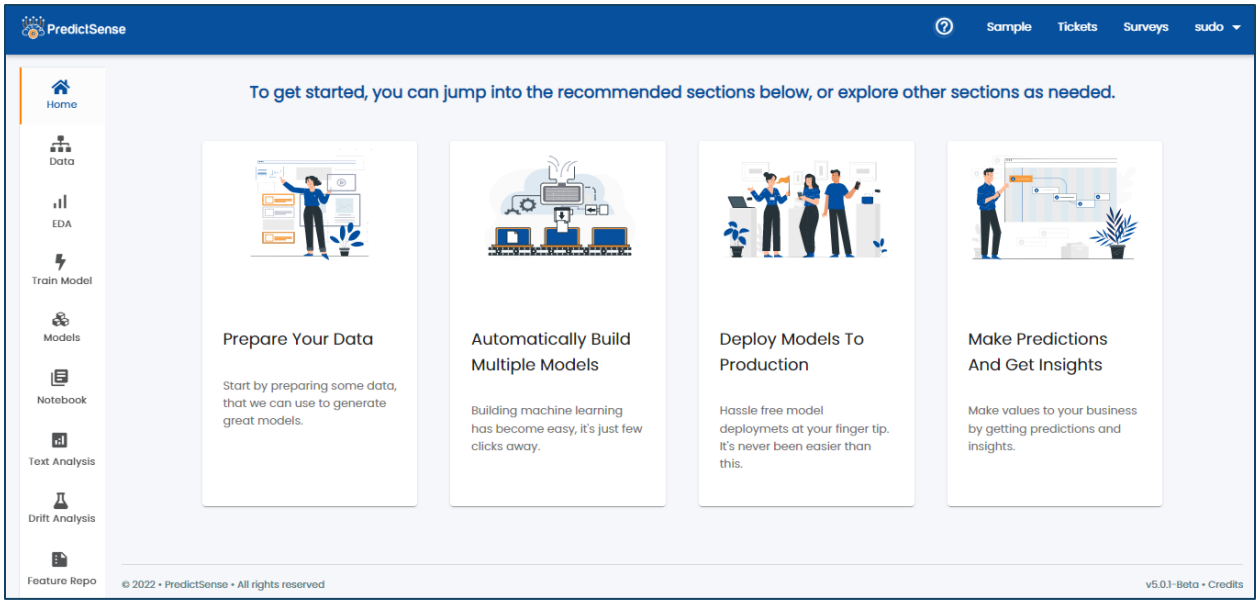


Sample Project

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

2 Home

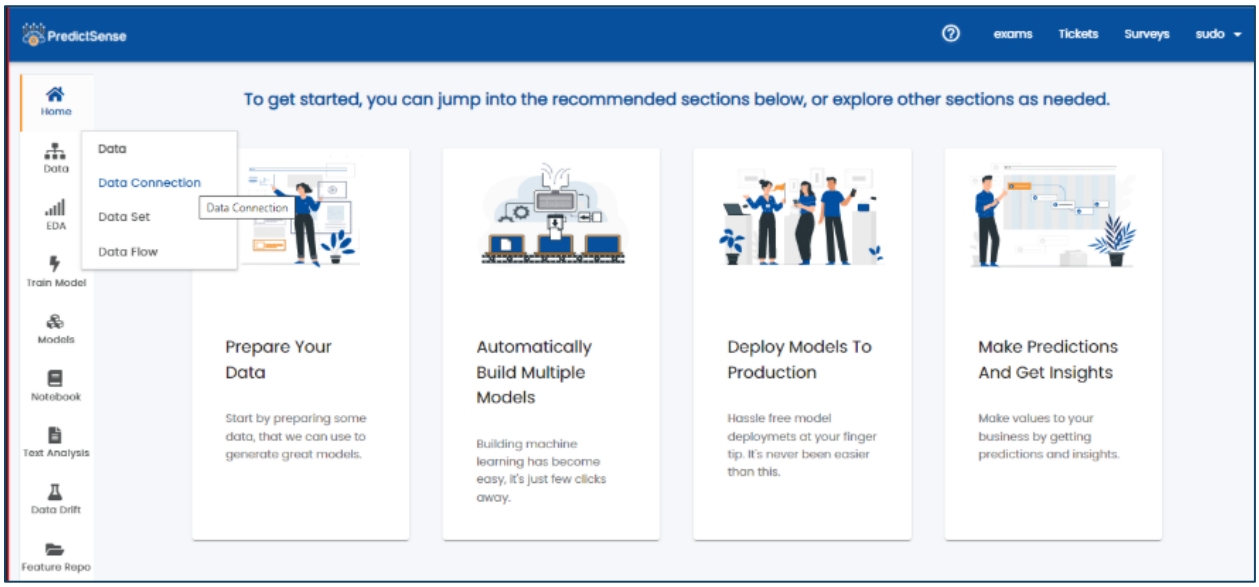
Home page will display the flow of PredictSense and different steps to perform the different operation like prepare your data, automatically build multiple models, deploy model to production, make prediction to get insights etc.



Home page

3 Data

It contains the following operations like data connection module, dataset module and data flow module etc.

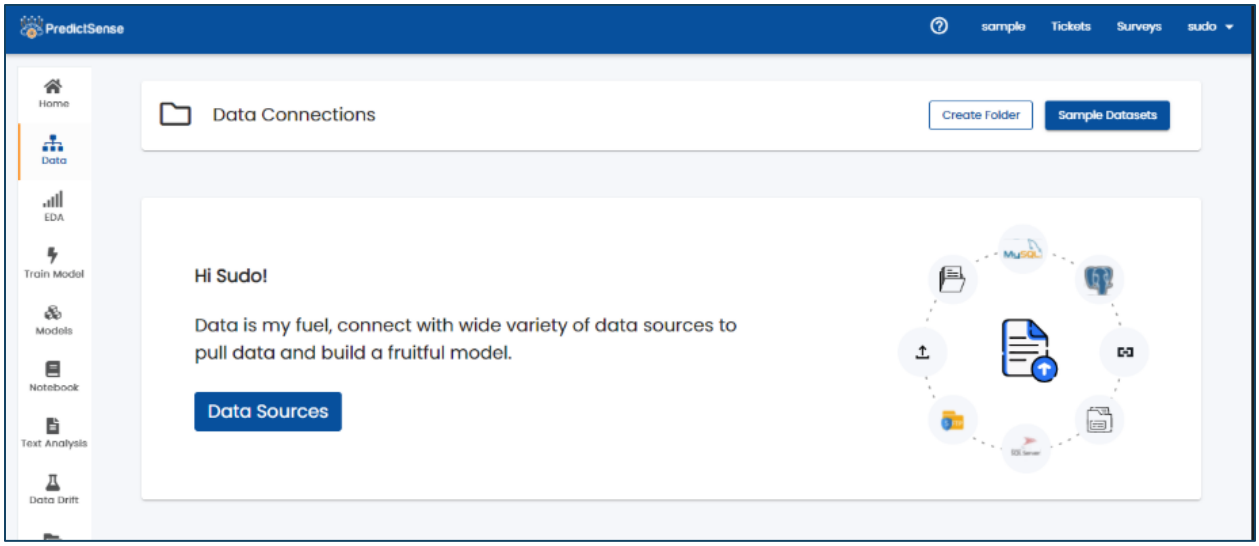


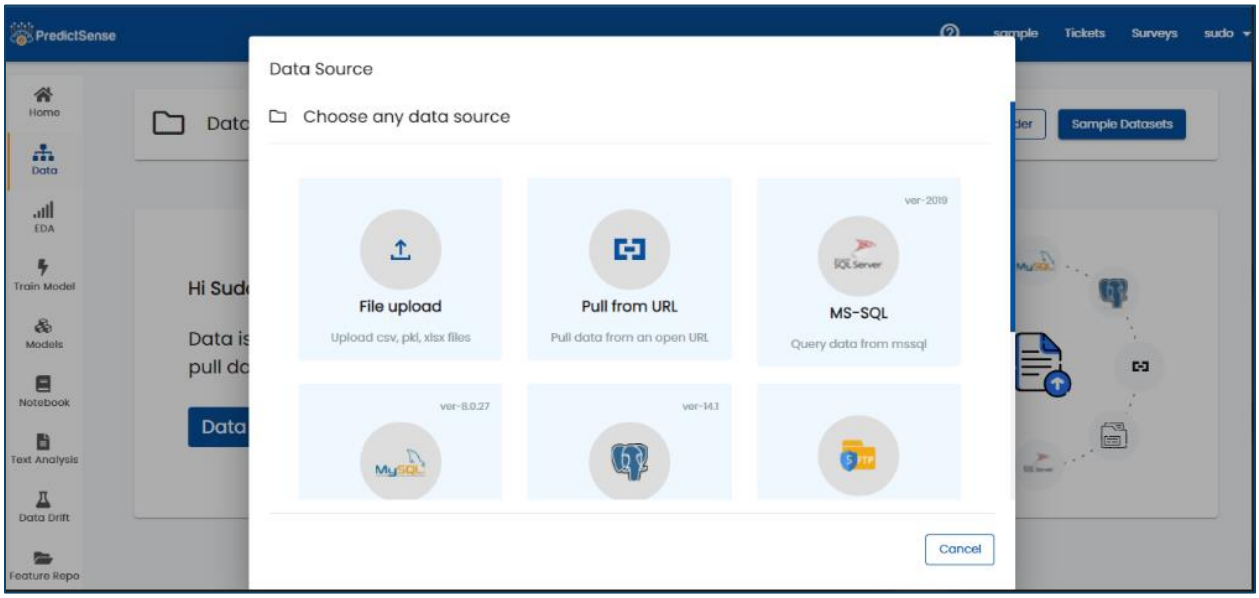
Data Page

3.1 Data Connection

Data connection utility help user to create connections for data sources like pull from URL, MSSQL, MySQL, PostgreSQL, snowflake, SFTP, Amazon s3 and Google BigQuery options. You can automate the data connection using scheduler activity. It will pull the data after every minute, hour, day and week based on scheduler selection in specific data connection. In data connection module you can see only raw files.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.





Data Connection

Users have various options for importing the data into PredictSense.

1.**File Upload:** Data from local device can be uploaded in the form of **.csv, .pkl, .xlsx** format

2.**Pull from URL:** Data can pulled directly from URL with **JSON or csv** format. Also, you can schedule the pull activity.

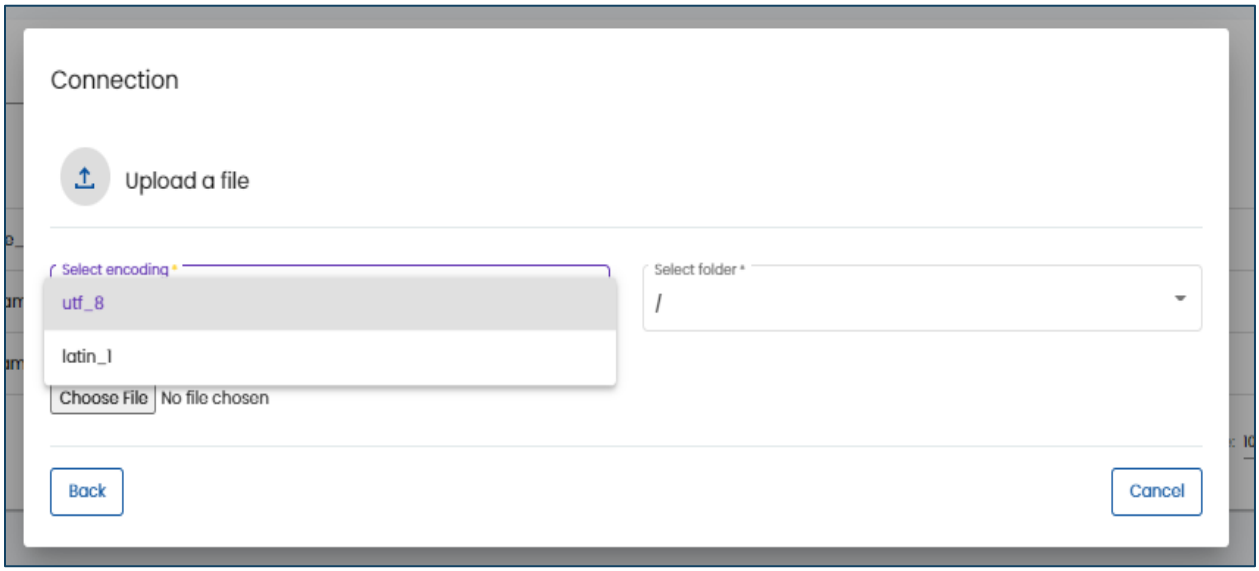
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.**MSSQL, MySQL, PostgreSQL, snowflake and Google bigquery** : Data can be query from existing **database** and used to make predictions. Also you can scheduler the pull from MSSQL, MySQL, PostgreSQL, snowflake and Google bigquery etc.

4. **SFTP :** Data can pulled from SFTP connection in different format and also user can setup scheduler to data for data pull from SFTP connection by adding scheduler.

3.1.1 File upload

Upload files (csv,xlsx,pkl,html) to PredictSense from your computer by providing relevant information like "encoding" and "folder name". Uploaded file will be added to **"Datasets"** tab and also in **"Data connection".** Application supports file encoding latin_1, unstructured file under file upload.



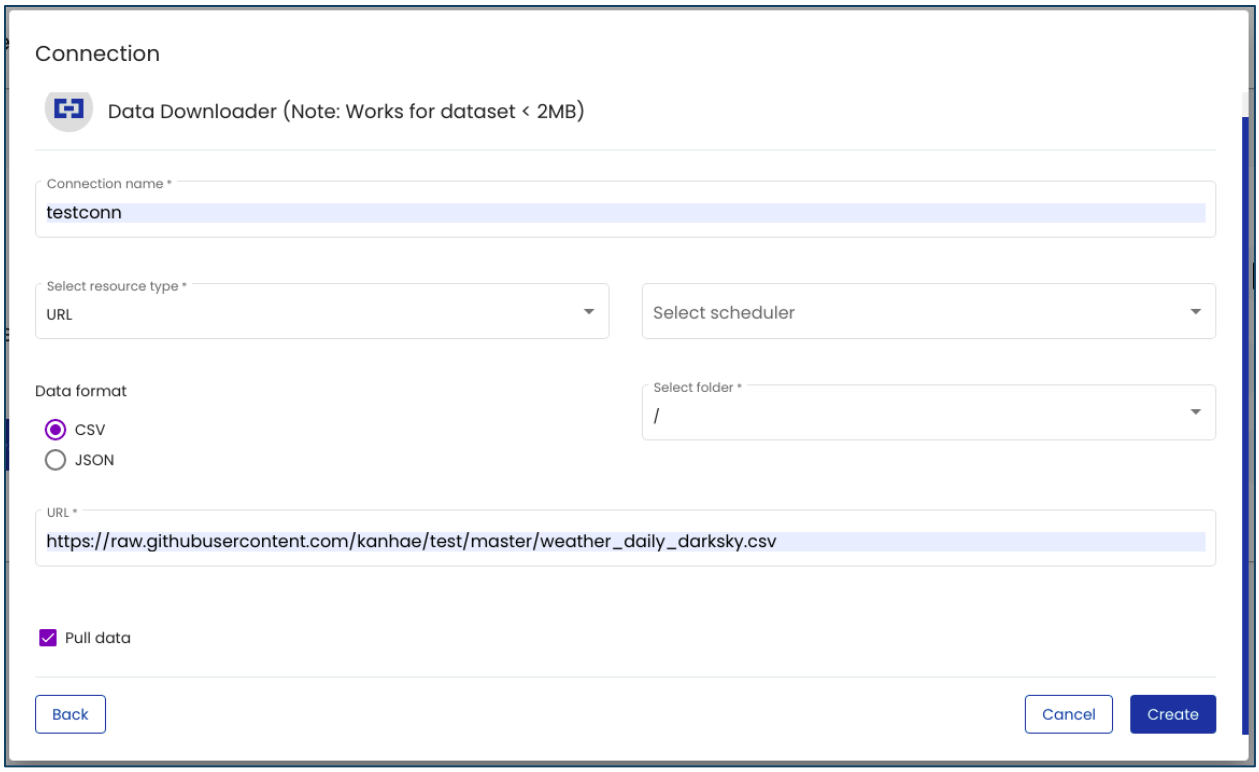
File Upload

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

3.1.2 Pull from URL

You can create a connection for pulling data (csv, json, html etc) from different URL's.

Data can be pulled from URL connection by entering **"filename", "Folder Name", "Select Scheduler", "Select type"** and **"URL"** and selecting type of data format "**CSV or JSON"** and after clicking on "**create data connection"** button.

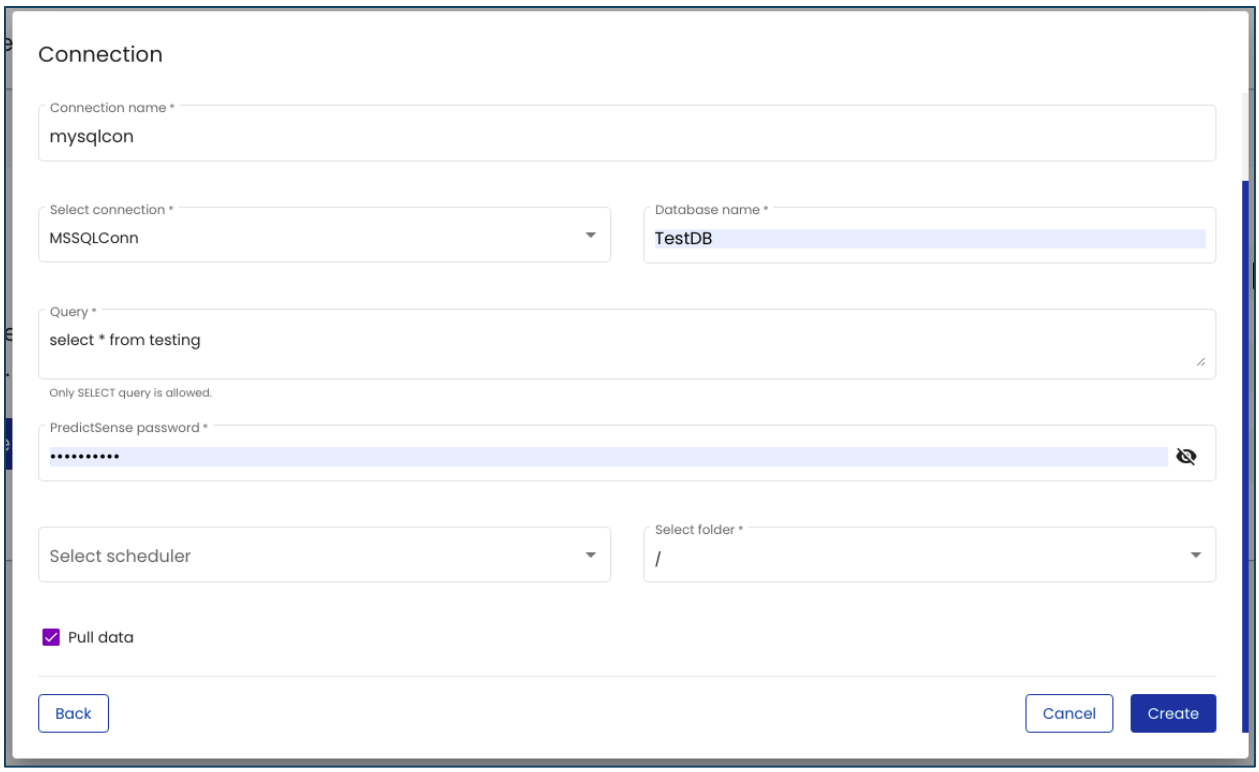


Pull from URL Source

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.3 Pull data from mysql, mssql, postgresql

Data set pulled UI will show like below for MSSQL, MySQL, PostgreSQL etc. fill all the required details, enter database query and pull the data from. User required query to configure database connections to pull the Dataset.



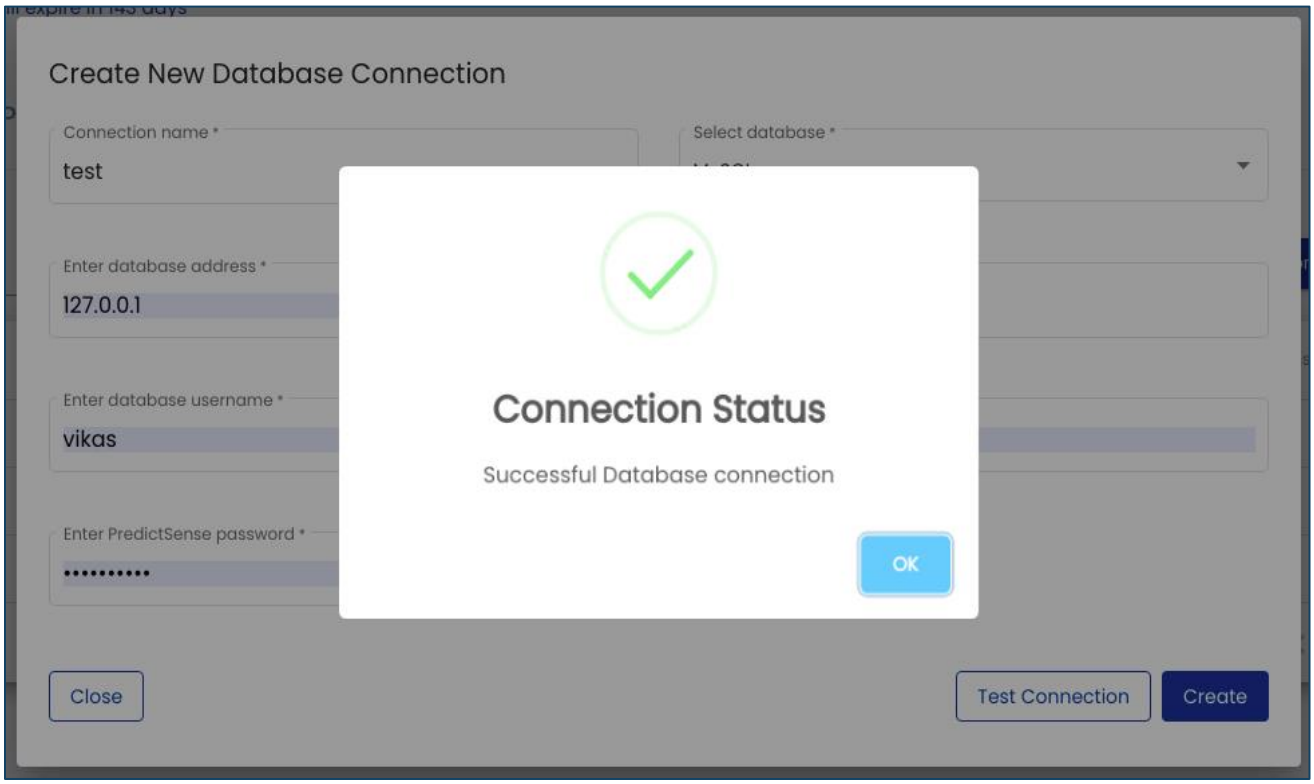
Pull from Database

3.1.4 Database connection configuration

Database connection is required to pull the datasets from any data connections like MSSQL, MySQL, PostgreSQL, Snowflake, Amazon s3 and Google bigquery etc. Go to the settings select database connection option in settings dropdown where you can configure database connection for MSSQL, MySQL and PostgreSQL, Snowflake, Amazon s3 and Google bigquery etc. by filling all the required details. User validate connection by clicking on test connection button. Use submit button for save the settings.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.5 MSSQL , MySQL , PostgreSQL –

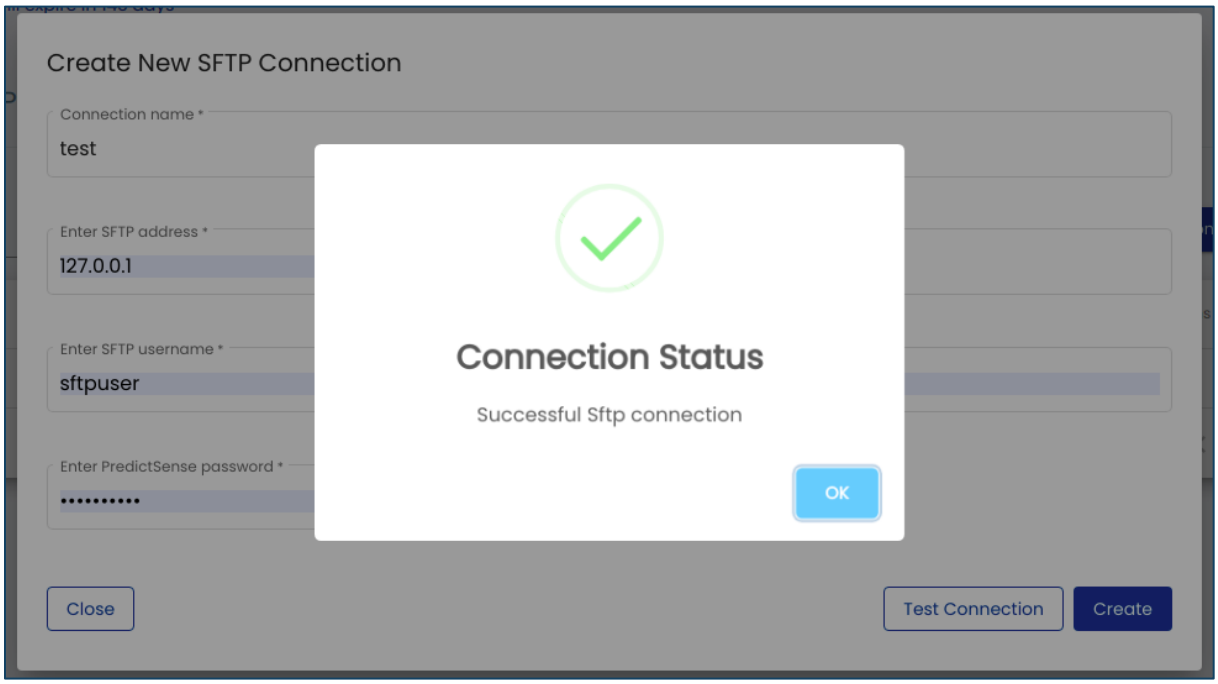


Database Connection Status

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

3.1.6 SFTP connection configuration

You can configure the SFTP connection by adding all the required details.

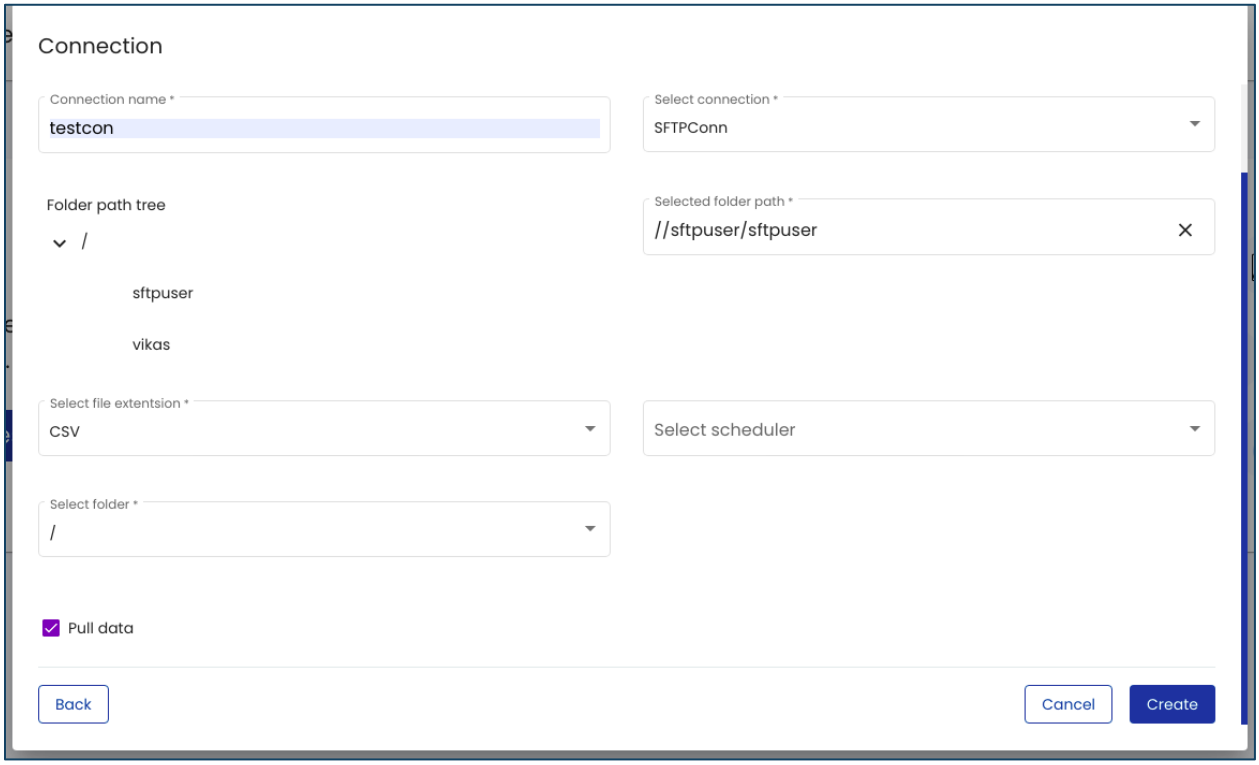


SFTP Connection Status

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.7 Pull from SFTP server

To pull the data from SFTP connection. First you need to configure the SFTP connection and pull the specified format of data like csv, pdf etc.

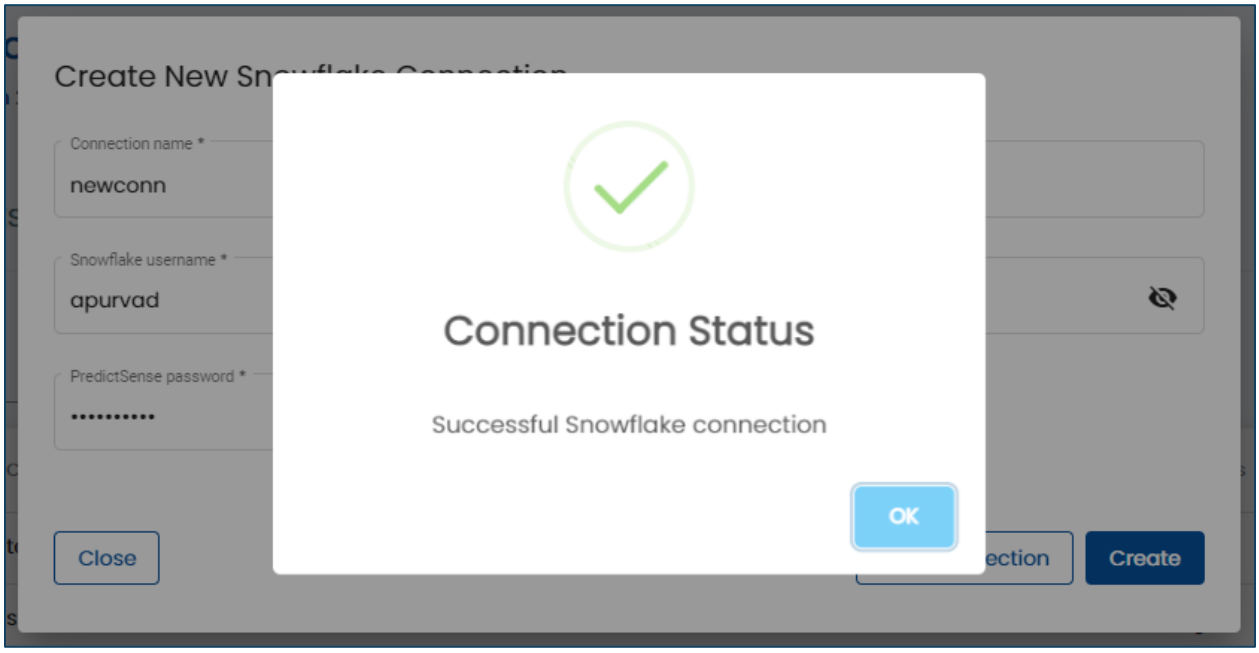


Pull Data from SFTP

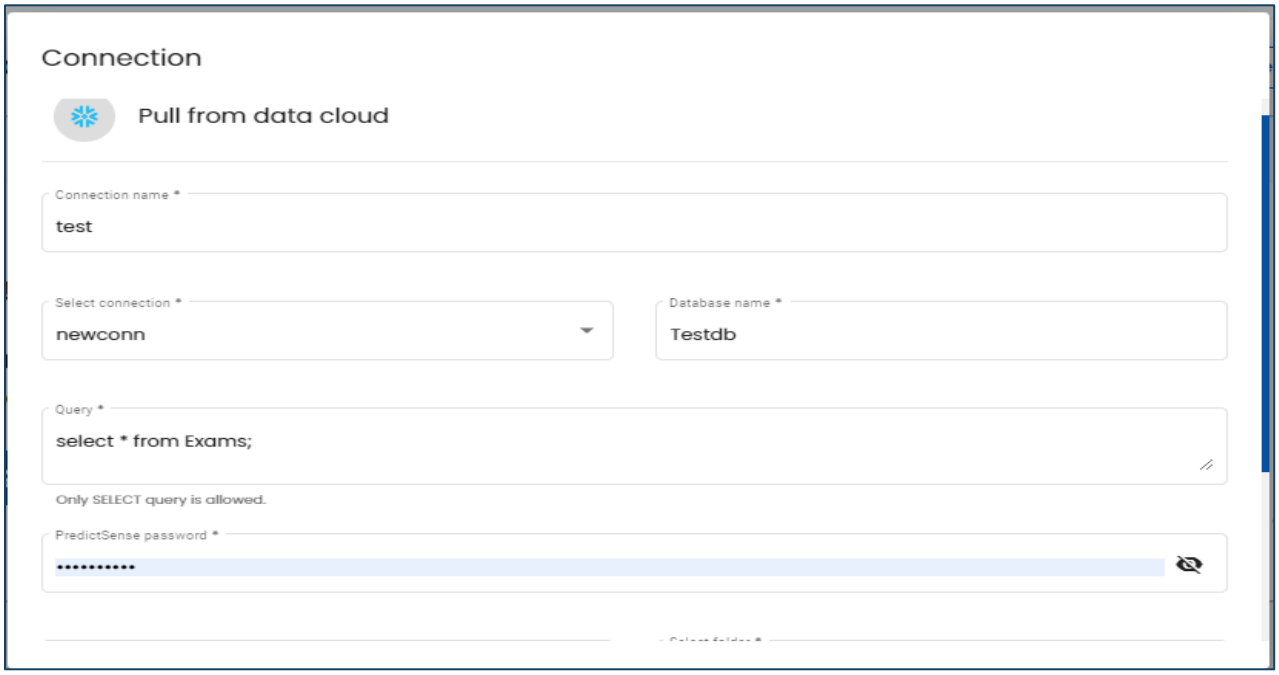
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

3.1.8 Snowflake

You can configure the snowflake connection by adding all the required details and pull the specified format of data like csv.



Snowflake Connection Status

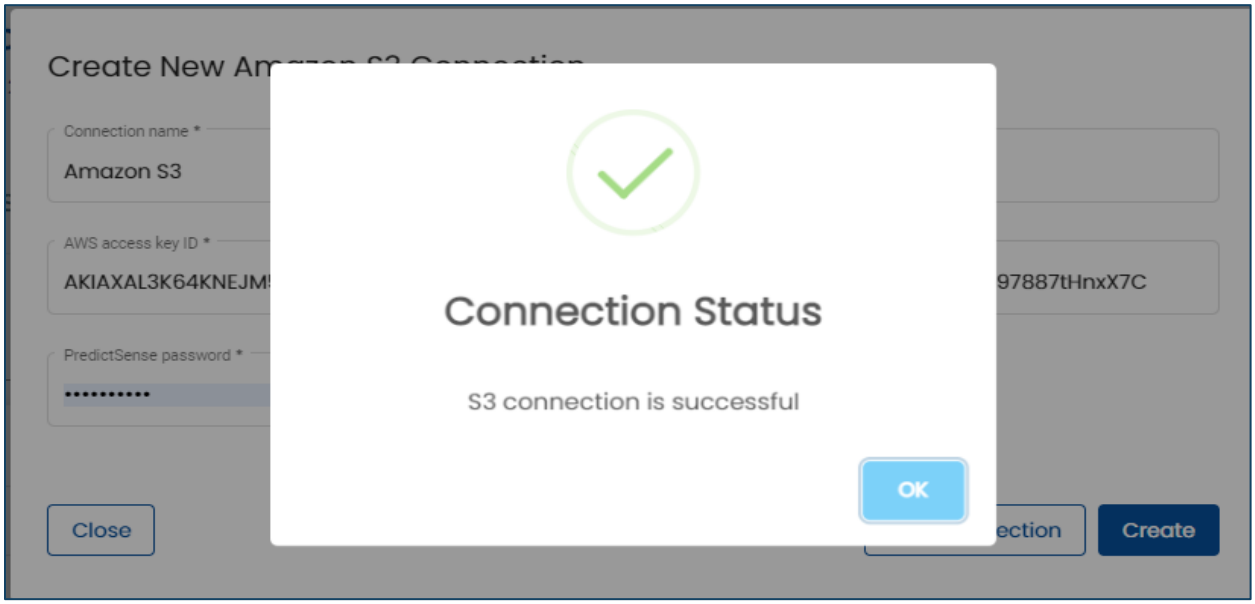


Pull Data from Snowflake

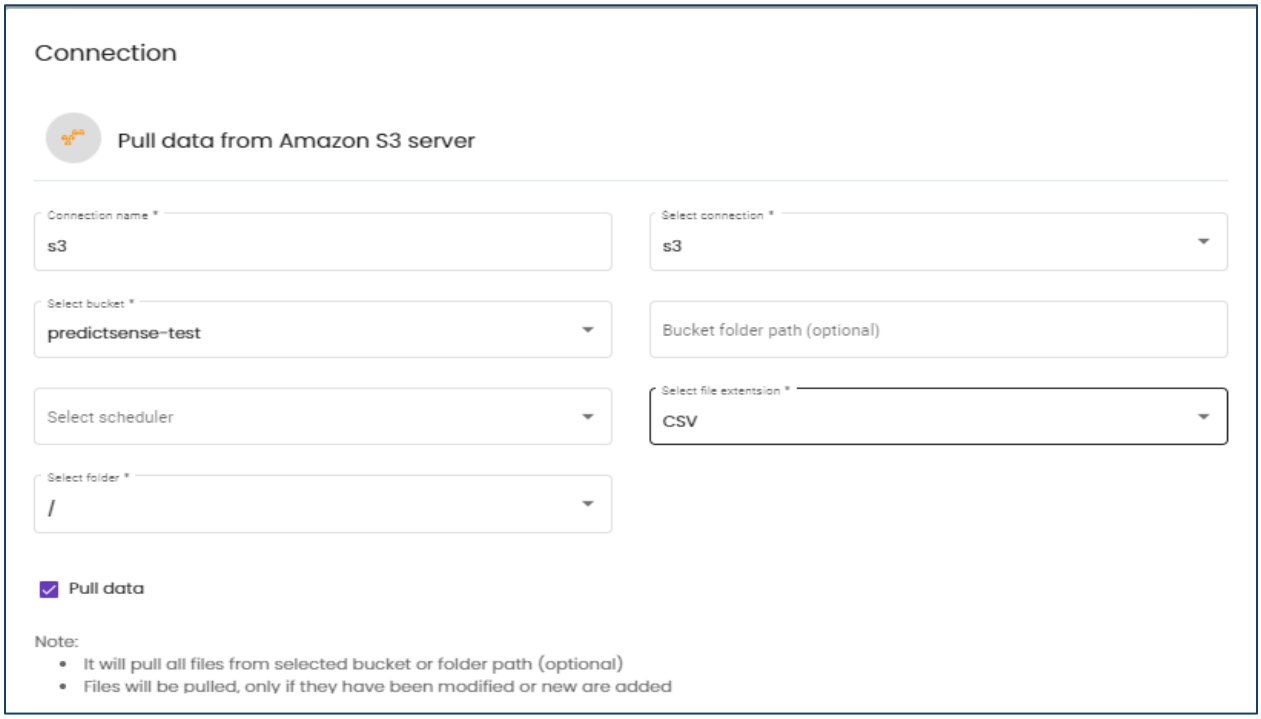
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.9 Amazon s3

You can configure the amazon s3 connection by adding all the required details and pull the specified format of data like csv, xlsx etc.



Amazon s3 Connection Status

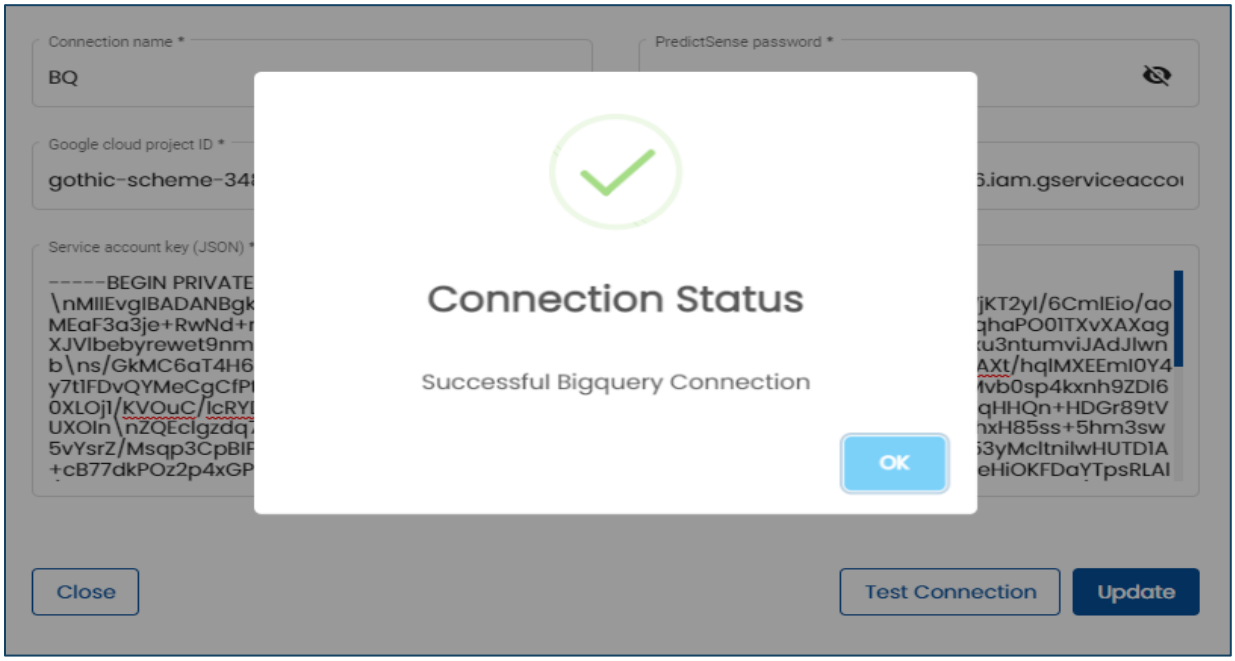


Pull Data from Amazon s3

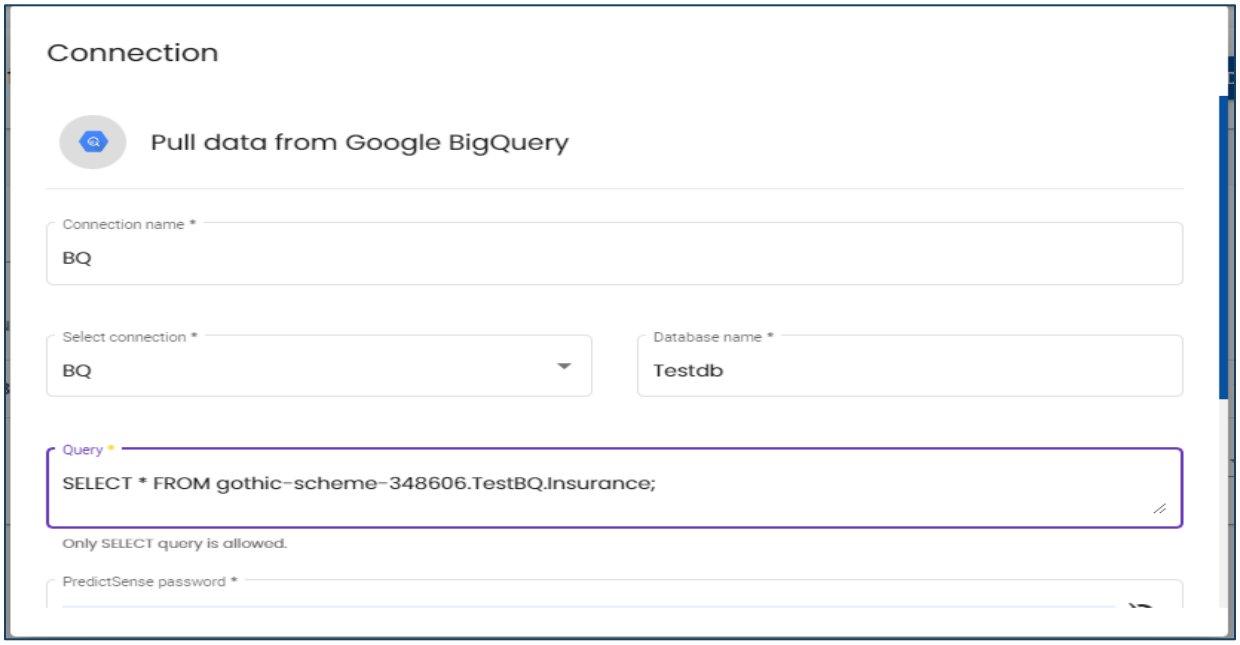
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.10 Google bigquery

You can configure the bigquery connection by adding all the required details and pull the specified format of data like csv.



Google bigquery connection status

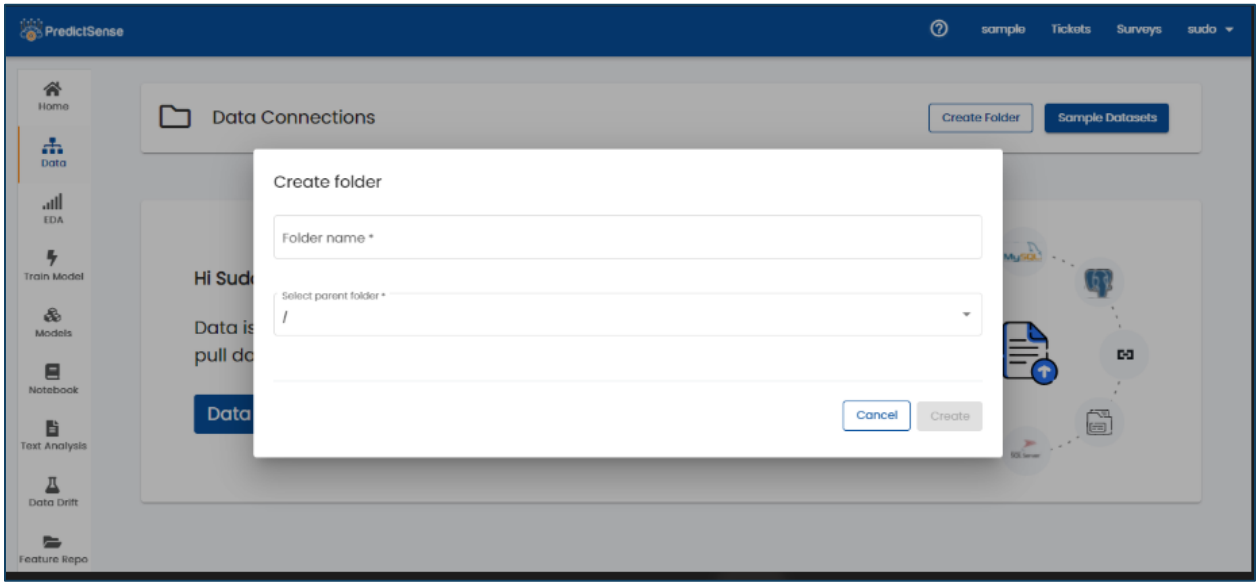


Pull Data from Google bigquery

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

3.1.11 Create folder

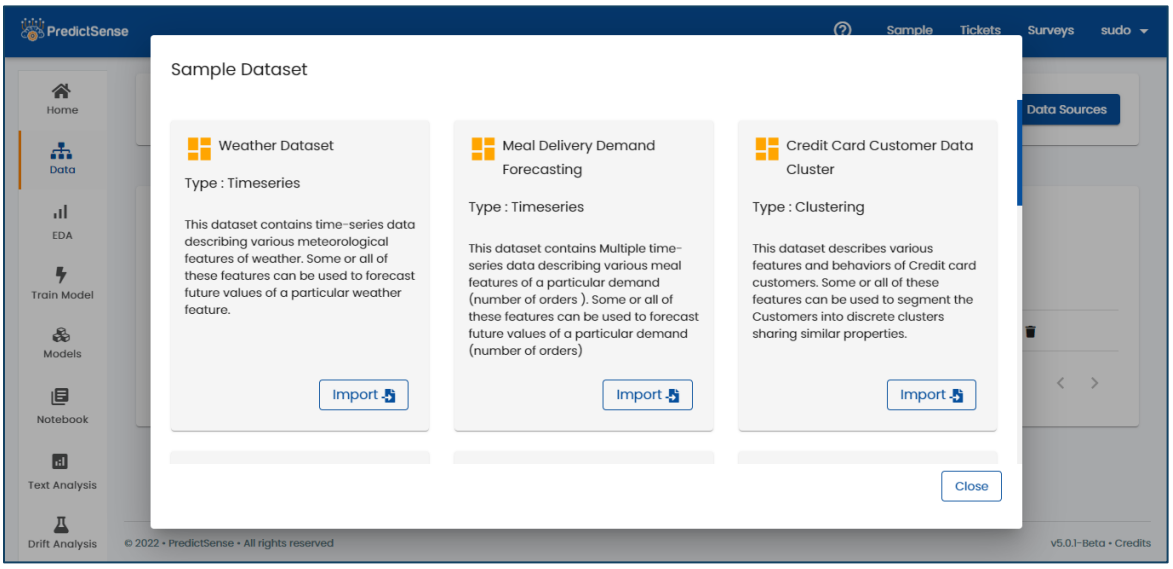
You can organize different data connections and files in multiple folders. You can create it while you are creating a connection or by clicking the Create Folder button in data connection module.



Create Folder

3.1.12 Sample Datasets

User can use sample dataset for training model.



Sample datasets

3.2 Data flow

Flow based tool for programming, used for fetching, transforming, and storing data in PredictSense. It can be used for creating the logical flow for any type of data. Data flow helps to process the unstructured file in to structured file.

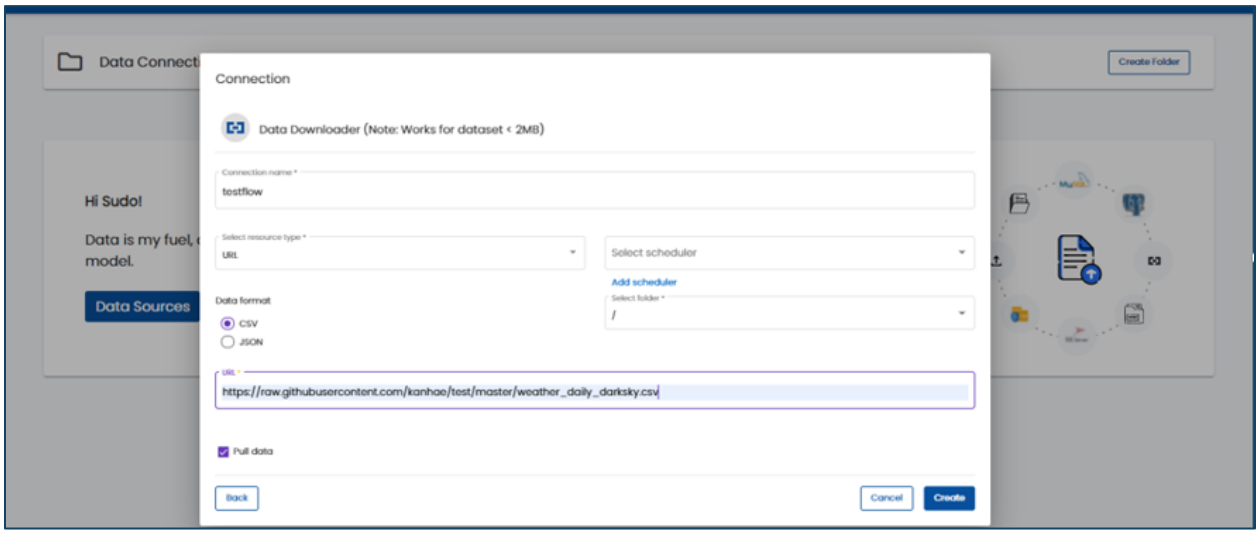
Types of data flow

1. Interceptor data flow
2. Direct data flow

3.2.1 Interceptor data flow

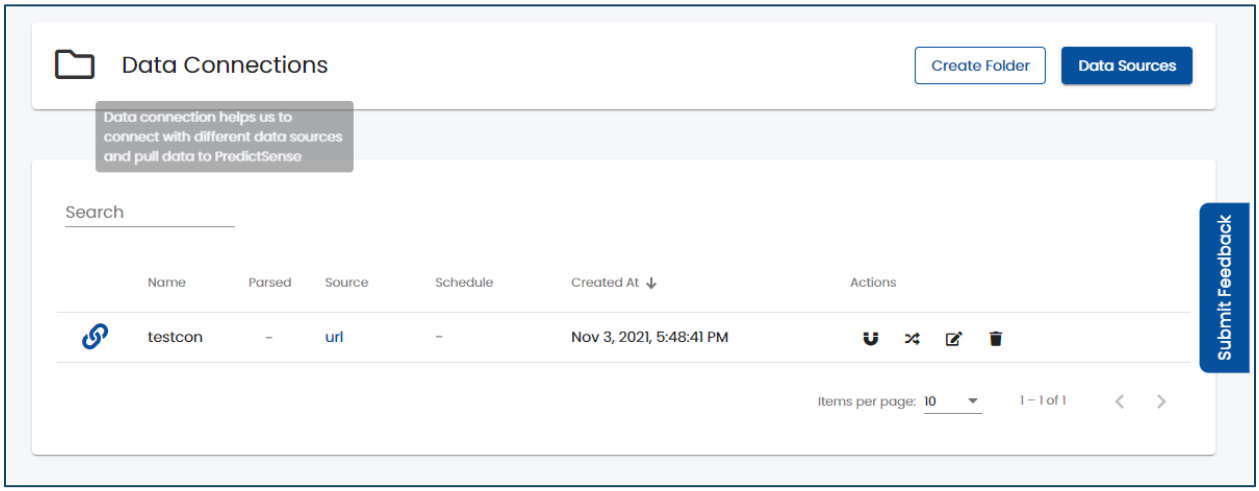
You can attach a data flow to a data connection, so whenever the data connection is executed it will be passed through the data will generate the output. This type of flow is called as interceptor flow. After creating a connection, you can click on the data flow option to create a interceptor data flow.

a) Go to data connection page and create connection for URL by filling all the required filed and assign the scheduler to connection. It will pull the dataset using the URL.



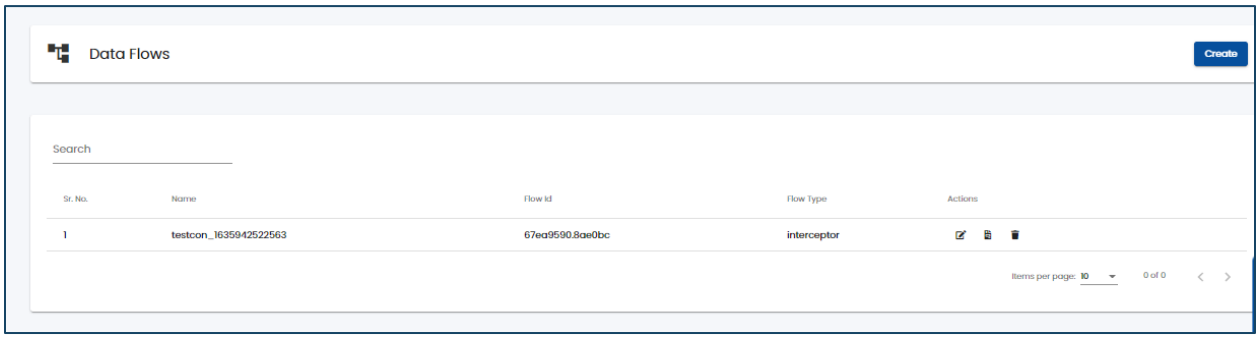
Pull data from URL

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

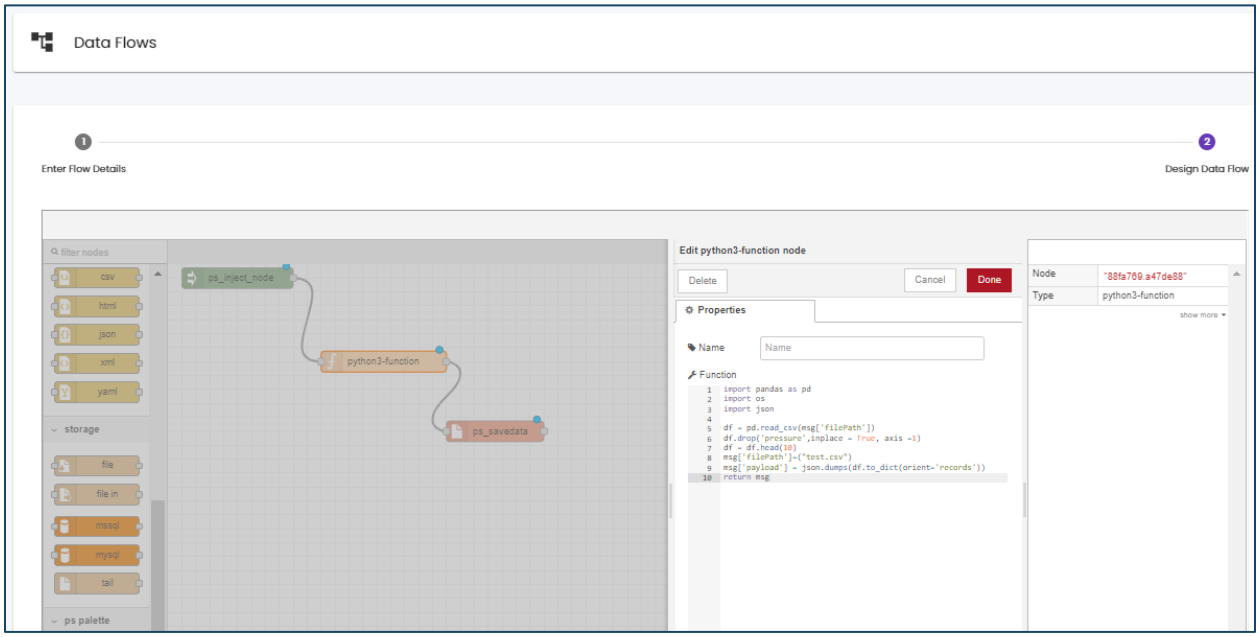


Creating interceptor data flow

b) By clicking on dataflow icon in data connection page you can create interceptor data flow. You can drag and drop the necessary node to create the flow. For any flow the inject node is mandatory to add. Please refer to the below screenshot example added one inject node, python3function, savedata node and passed the logic to drop column and save the dataset.

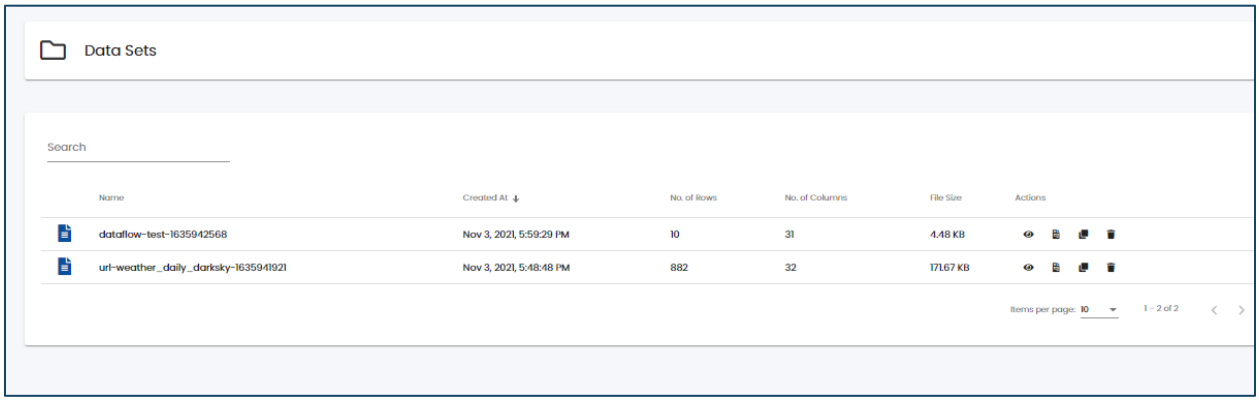


© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Data Flow

c) Execute the flow from data connection page. Click on source and it will open the popup window were you can start the flow execution by clicking on execute data flow button and you can see the saved data in dataset module.

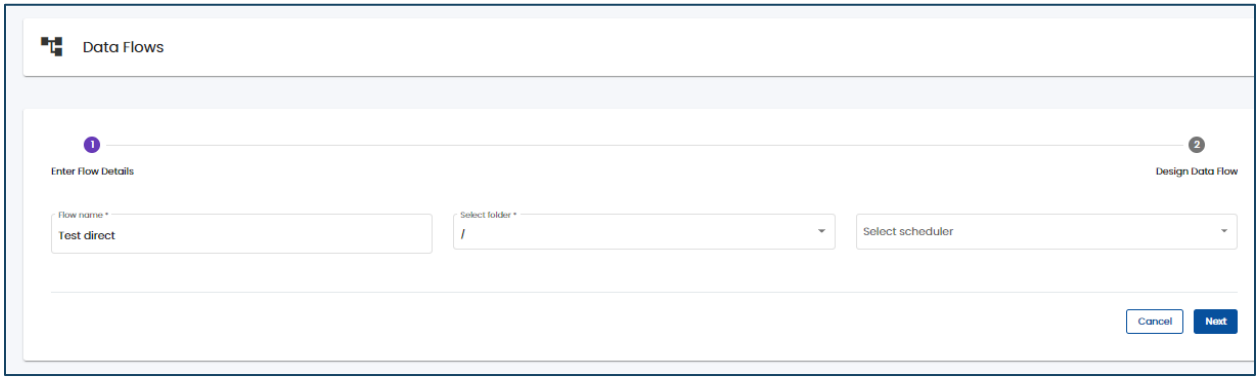


Dataset after flow execution

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

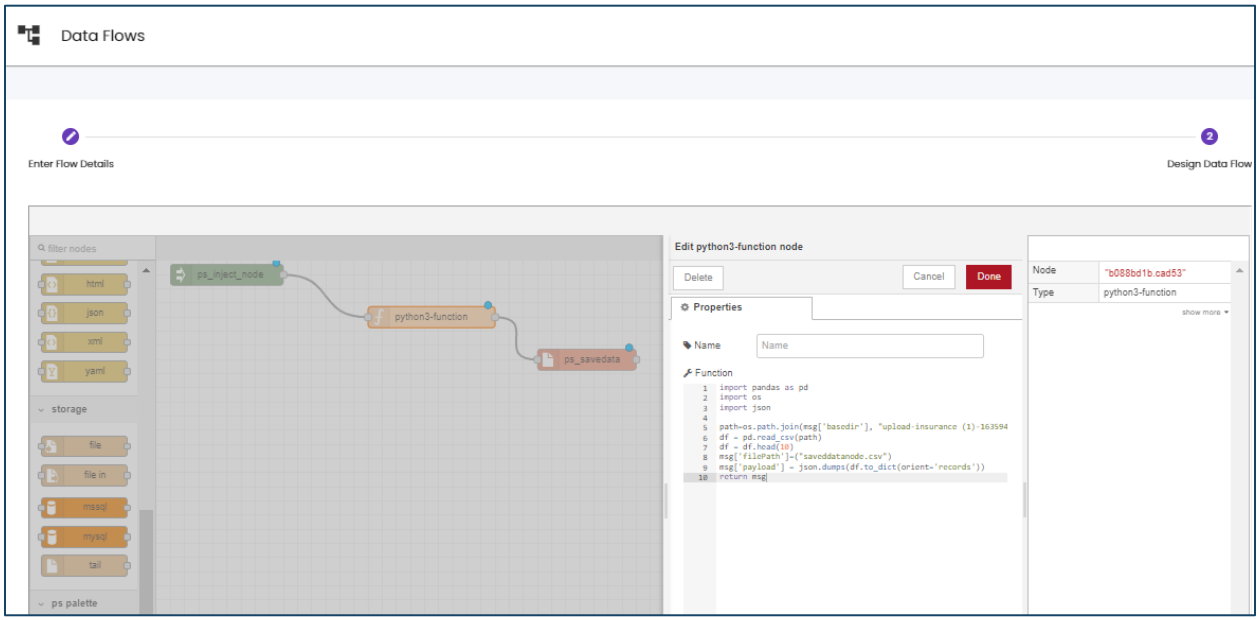
3.2.2 Direct flow

Whenever we need to transform some data which is not coming from data connections we can use direct data flow.



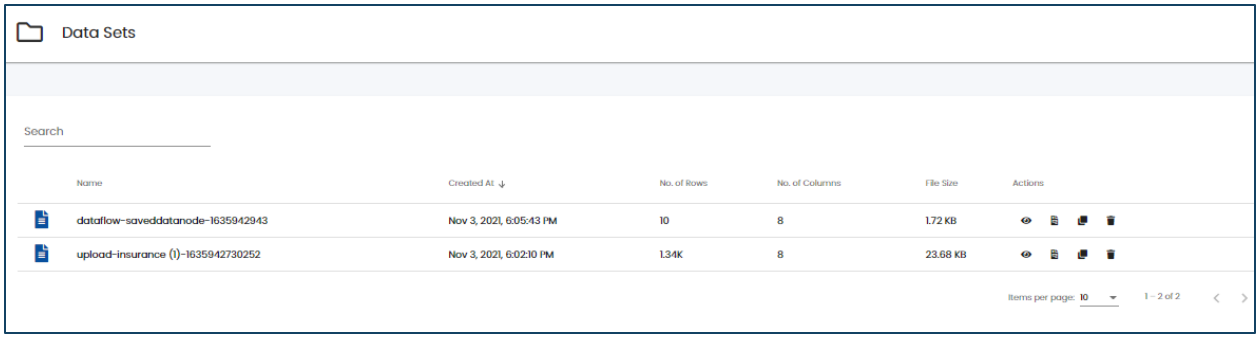
Direct Data Flow Creation

For example: you can create direct flow using **savedata** node and Append data node. savedata node will create a new file whenever the flow get executed and **appenddata** node will create on processed file in dataset page and one downloadable file in data connection module.



Adding Python code

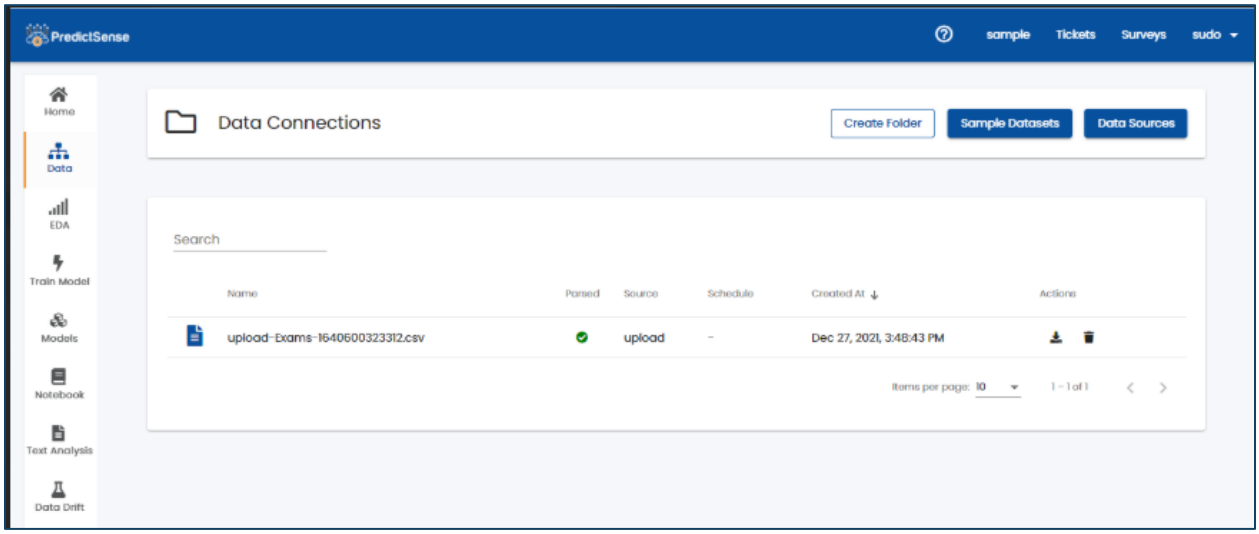
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Processed file

3.3 Data set

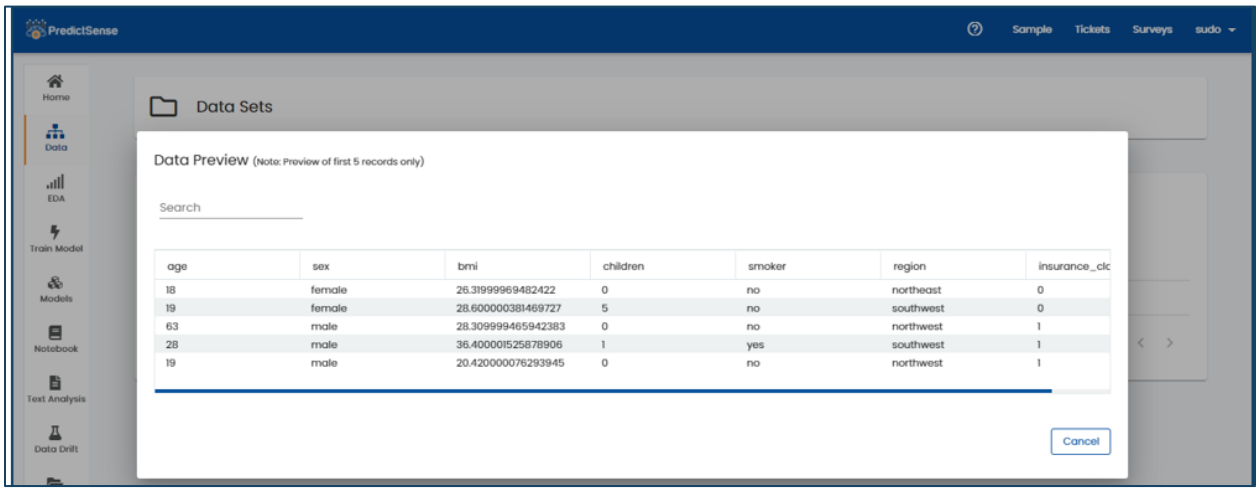
Data uploaded from different sources and connections are listed over here. All the files listed over here are structured and ready for further process. Preview of this datasets are available by **"click on data preview".** Also detailed report is available by **"click on detailed report"** and user can merge the two datasets from dataset module**.**



Dataset page

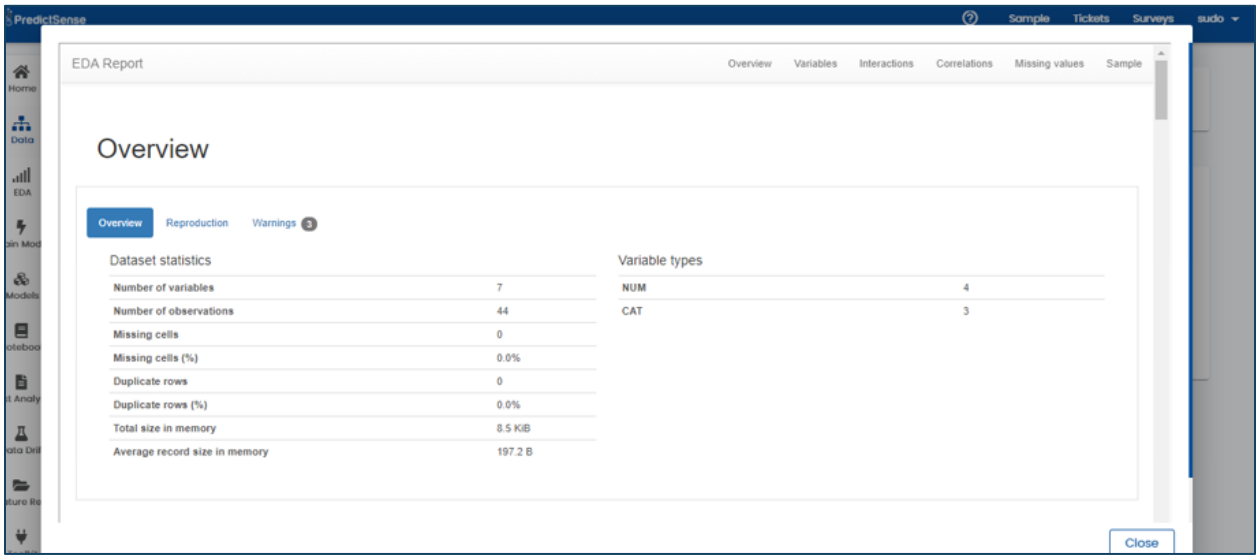
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

3.3.1 Data preview



Dataset Preview

3.3.2 Detailed report



Detail Report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

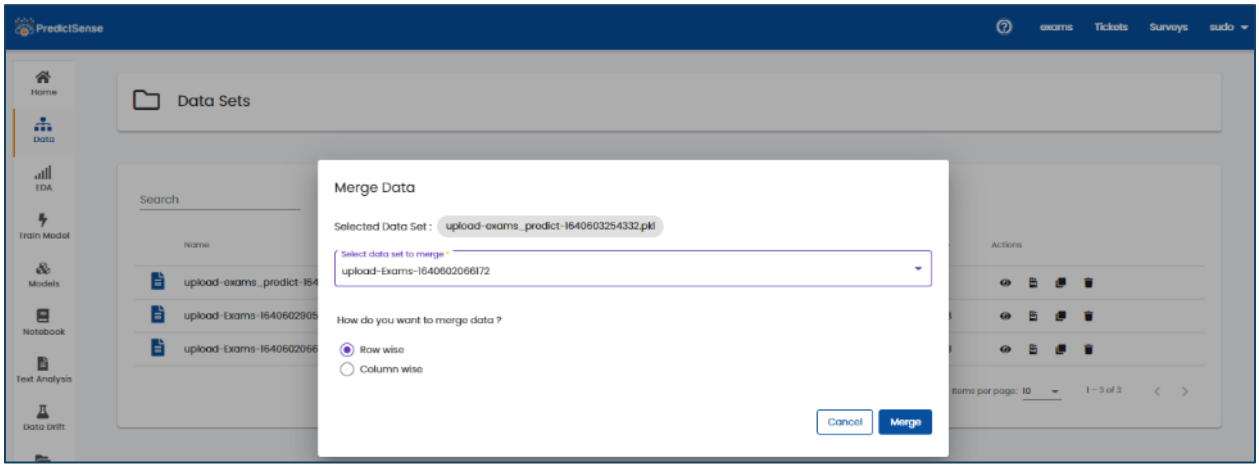
3.4 3.3.3 Merge dataset

Merge Dataset feature provide the option to merge two different datasets based on common column in both the datasets.

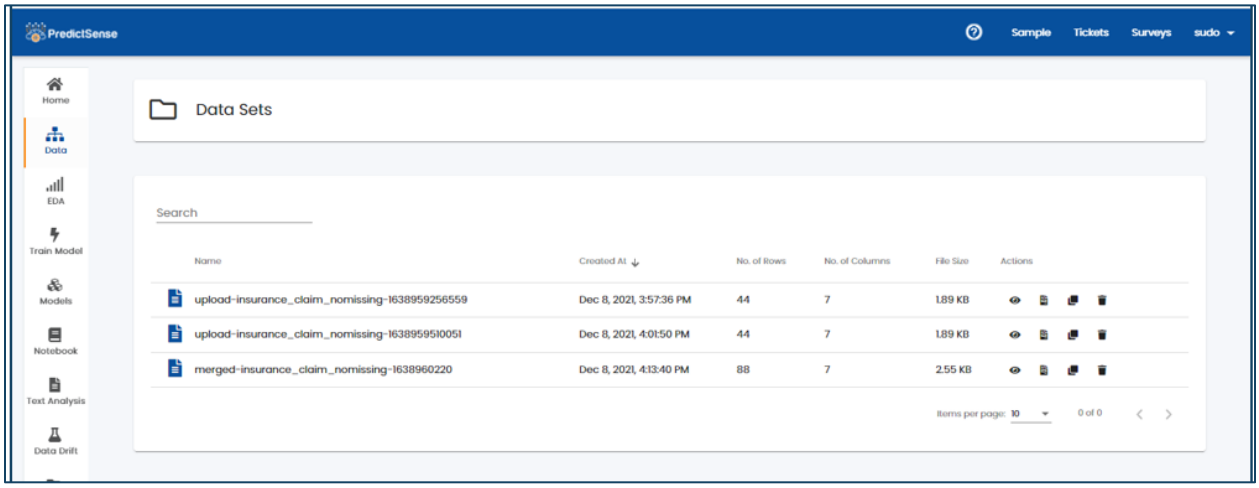
1. Row wise Merge
2. Column wise Merge

3.3.4 Row wise merge

Merging two datasets is the process of bringing two datasets together into one, and aligning the rows from each based on common attributes.



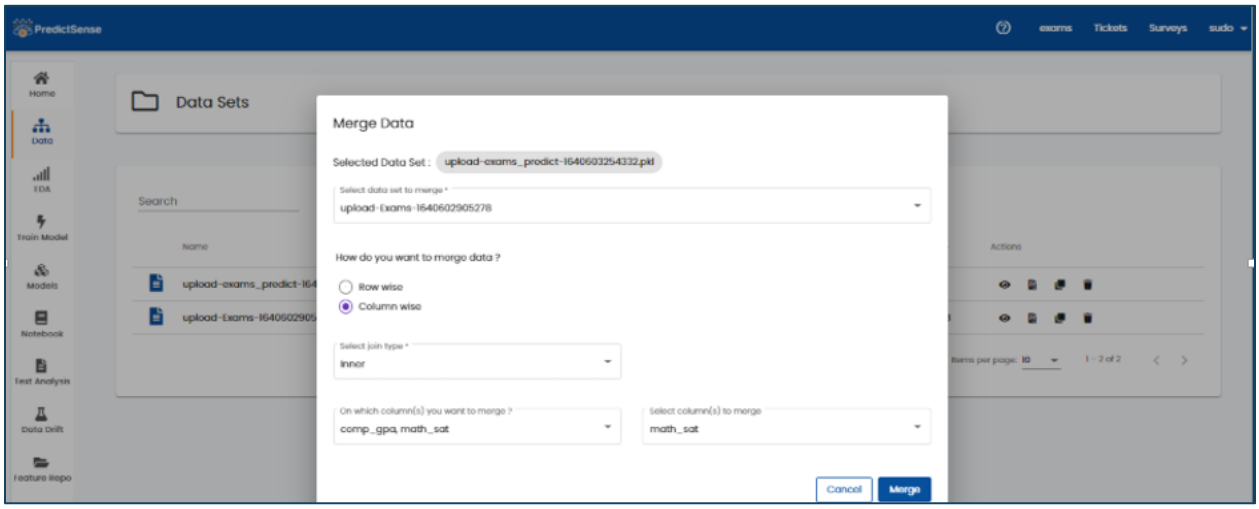
Row wise Merge



Data set page after row merge

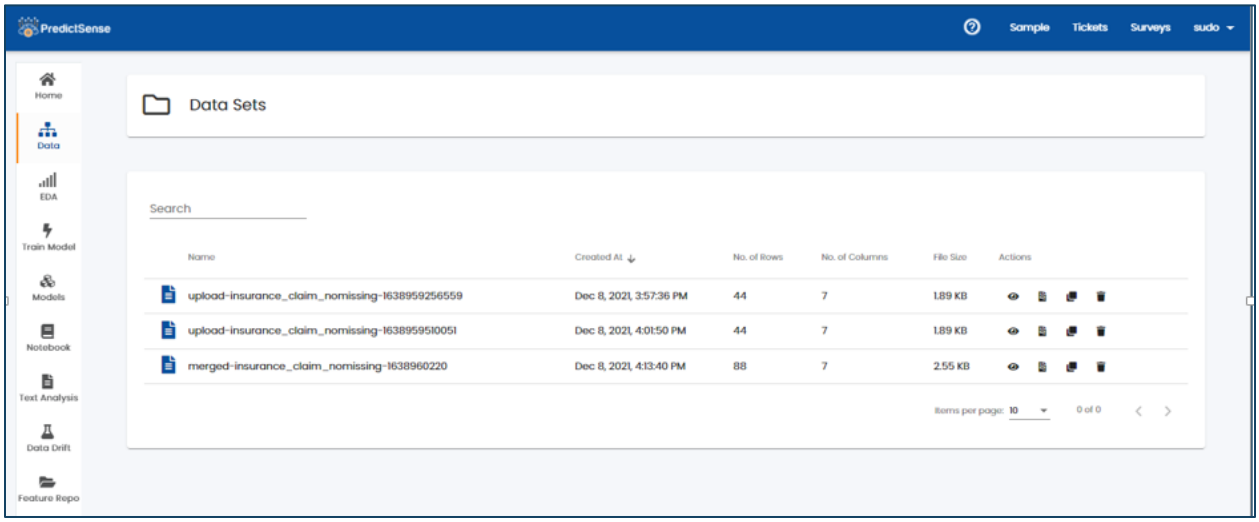
3.4.1 Column wise merge

It is the process of bringing two datasets together into one, and aligning the columns from each based on common attributes.



Column wise Merge

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Data set page after column merge

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

4 Text analysis

Text Analysis aims to extract machine-readable information from unstructured text in order to enable data-driven approaches towards managing content. To overcome the ambiguity of human language and achieve high accuracy for a specific domain, requires the development of customized text mining pipelines. Text Analysis is about parsing texts in order to extract machine-readable facts from them. The purpose of Text Analysis is to create structured data out of free text content.

Text analysis support following options like-

4.1. Cleaning/transformation

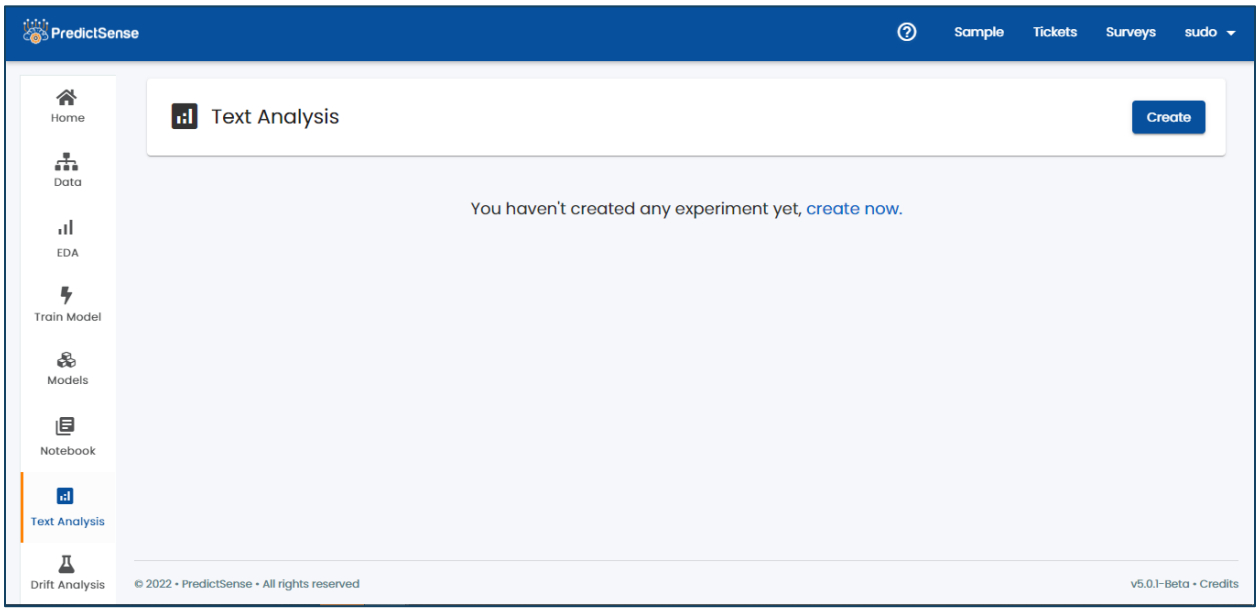
4.2. POS tagging

4.3. NER

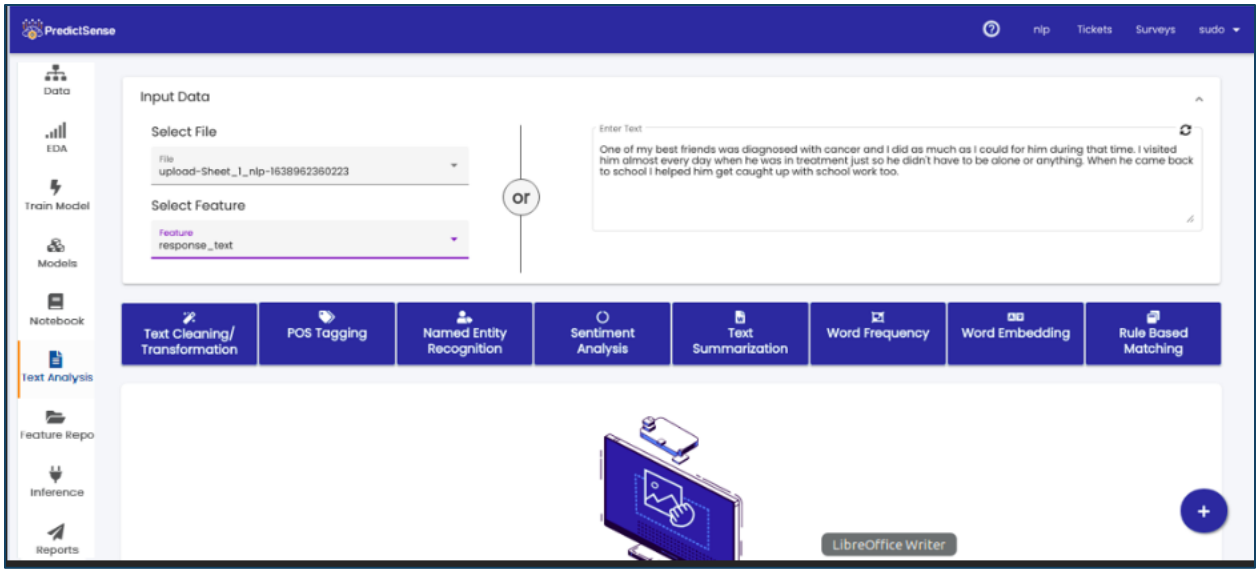
4.4 Sentiment analysis

4.5. Text summarization

4.6. Word Frequency



Text Analysis page

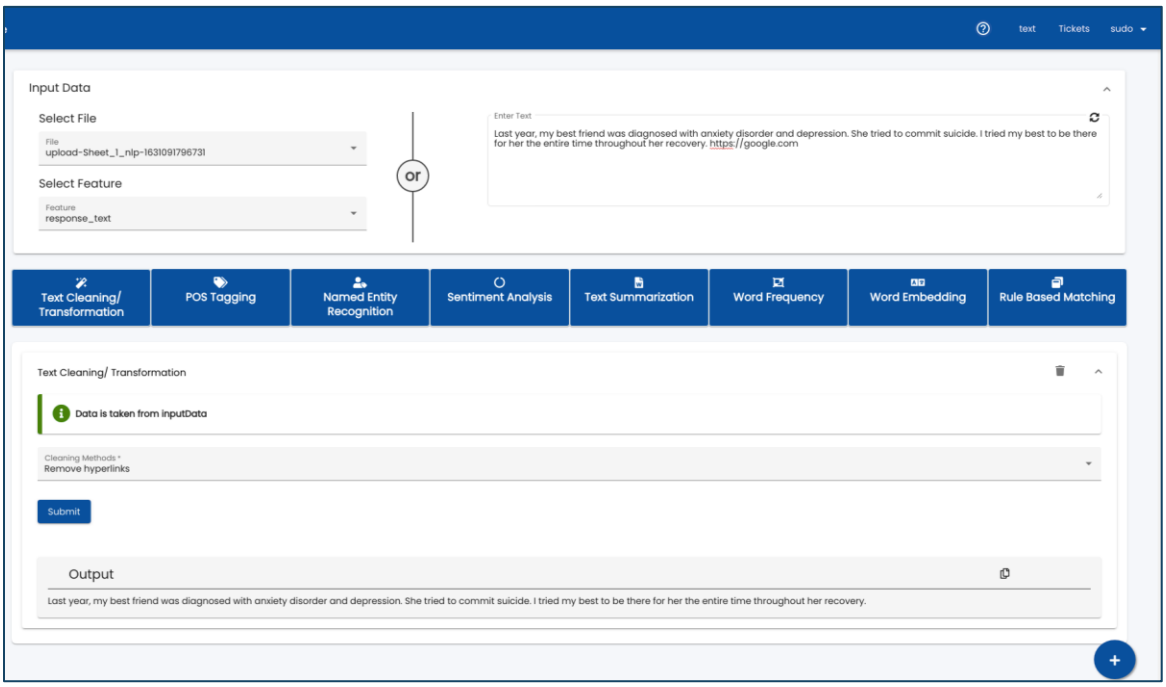


Input Data page

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

4.1 Cleaning/transformation -

Text contains different symbols and words which doesn't convey meaning to the model while training. So we will remove them before feeding to the model in an efficient way. This method is called text cleaning.

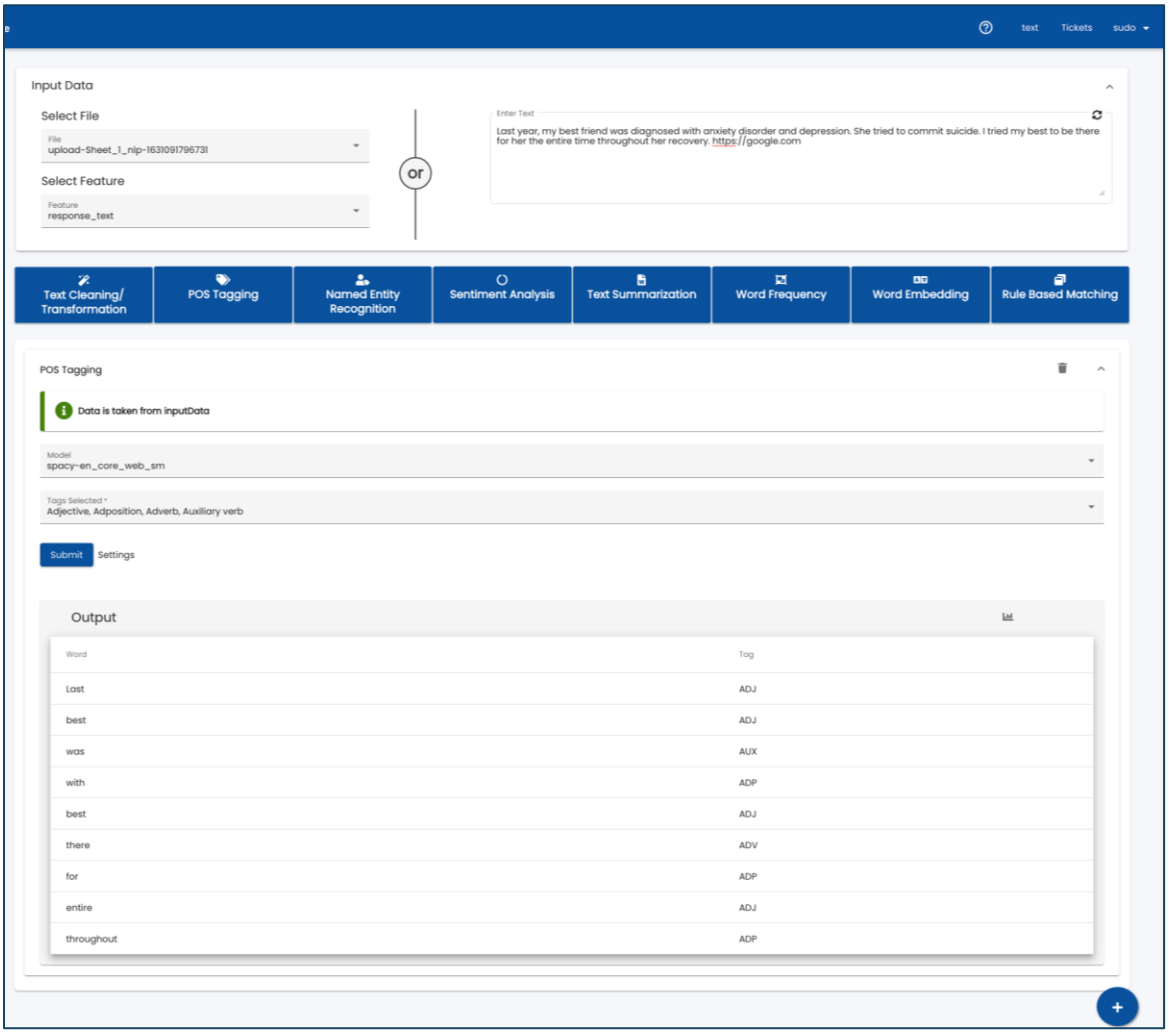


Text Cleaning Transformation

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

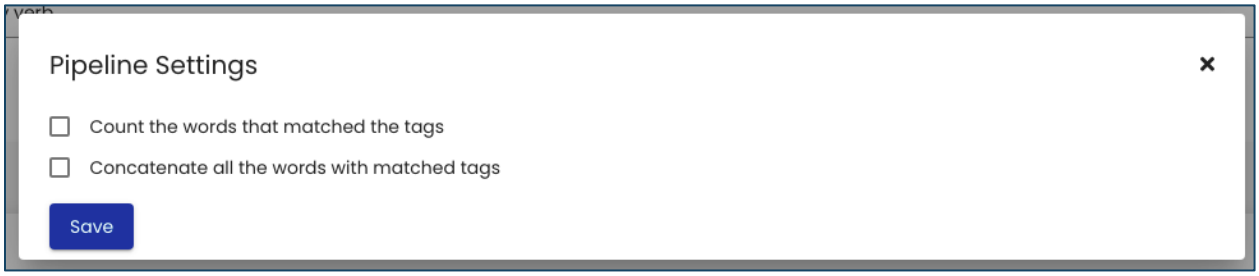
4.2 POS tagging-

Part-of-Speech tagging is a well-known task in Natural Language Processing. It refers to the process of classifying words into their parts of speech (also known as words classes or lexical categories).

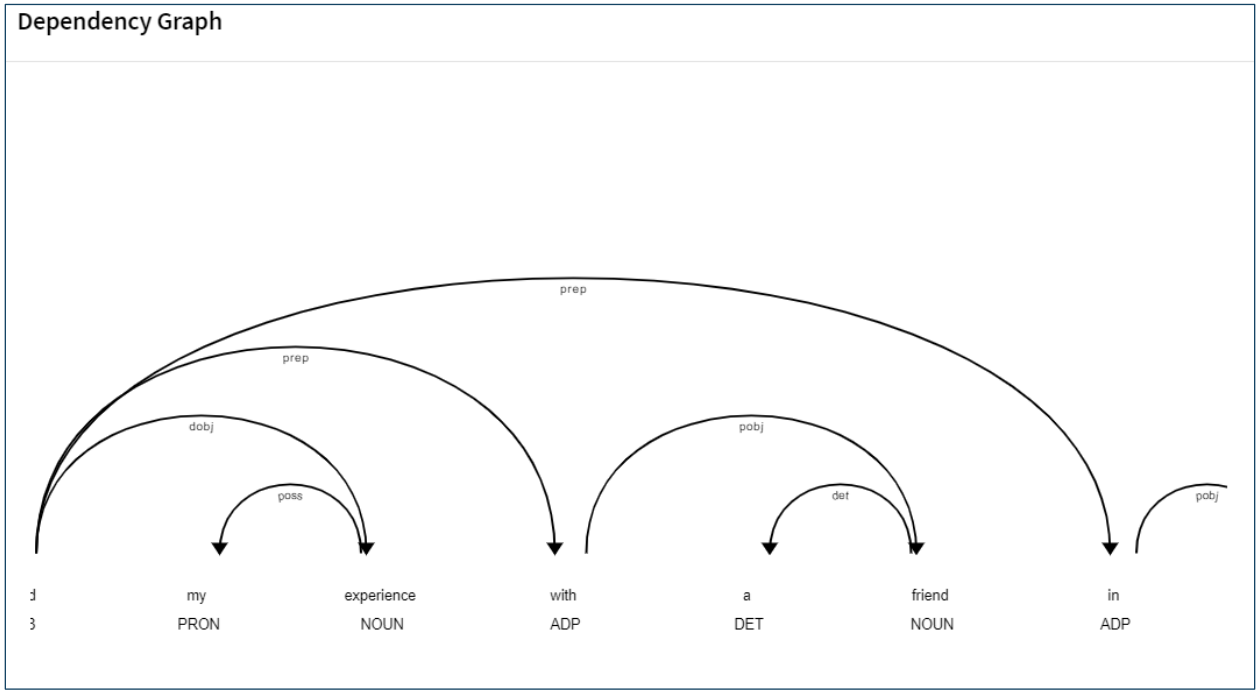


POS tagging

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Pipeline Setting Page

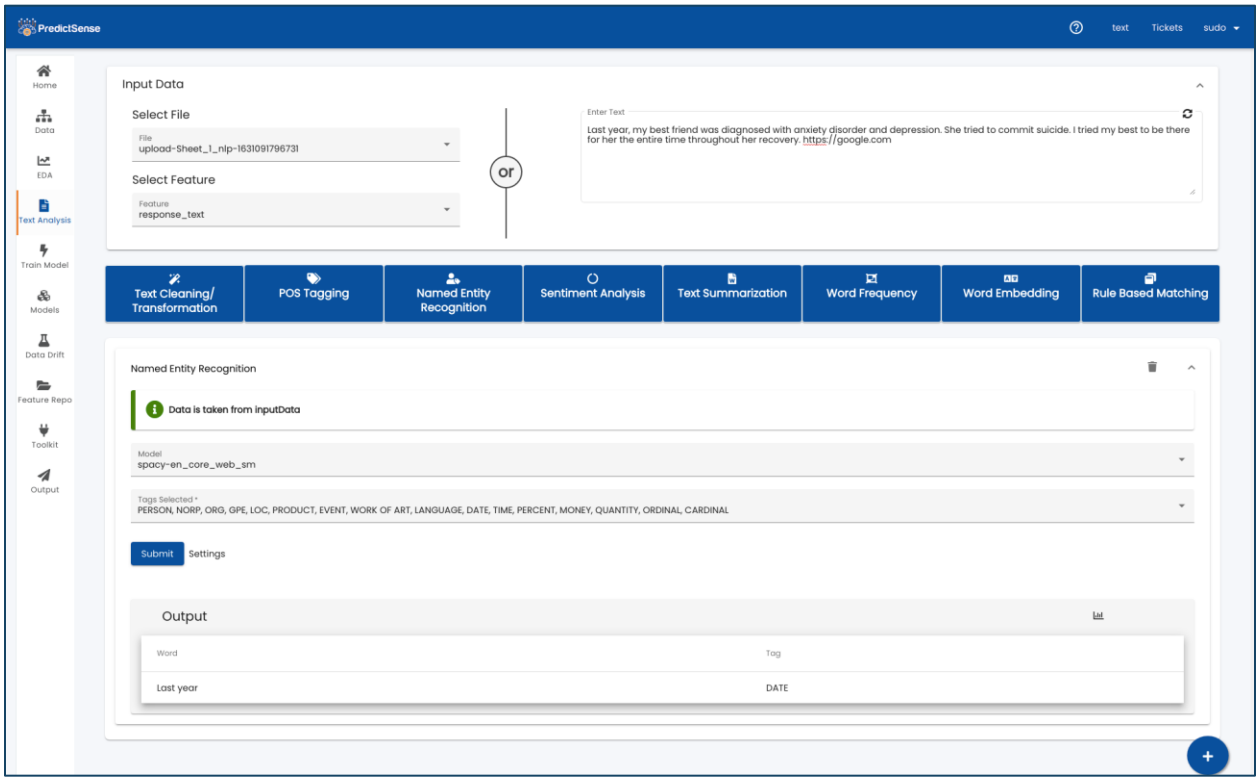


Dependency Graph

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

4.3 NER

NER is an information extraction technique to identify and classify named entities in text. These entities can be pre-defined and generic like location names, organizations, time etc.



NER page

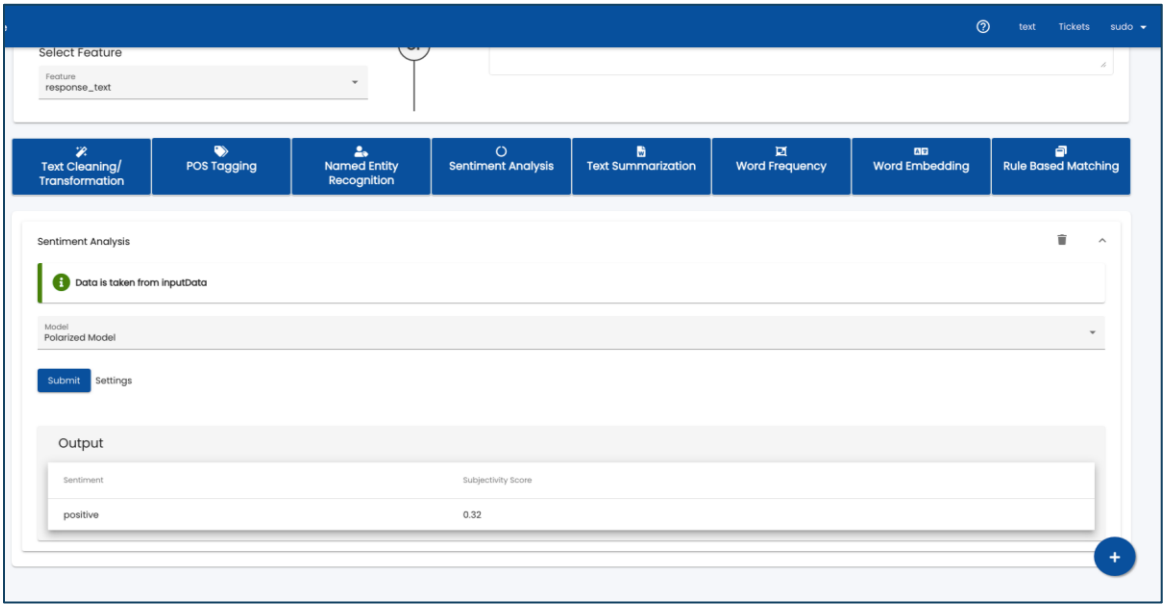
NER Graph

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

Pipeline setting for NER

4.4 Sentiment analysis-

Sentiment analysis is a machine learning tool that analyzes texts for polarity, from positive to negative. By training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.



Sentiment analysis

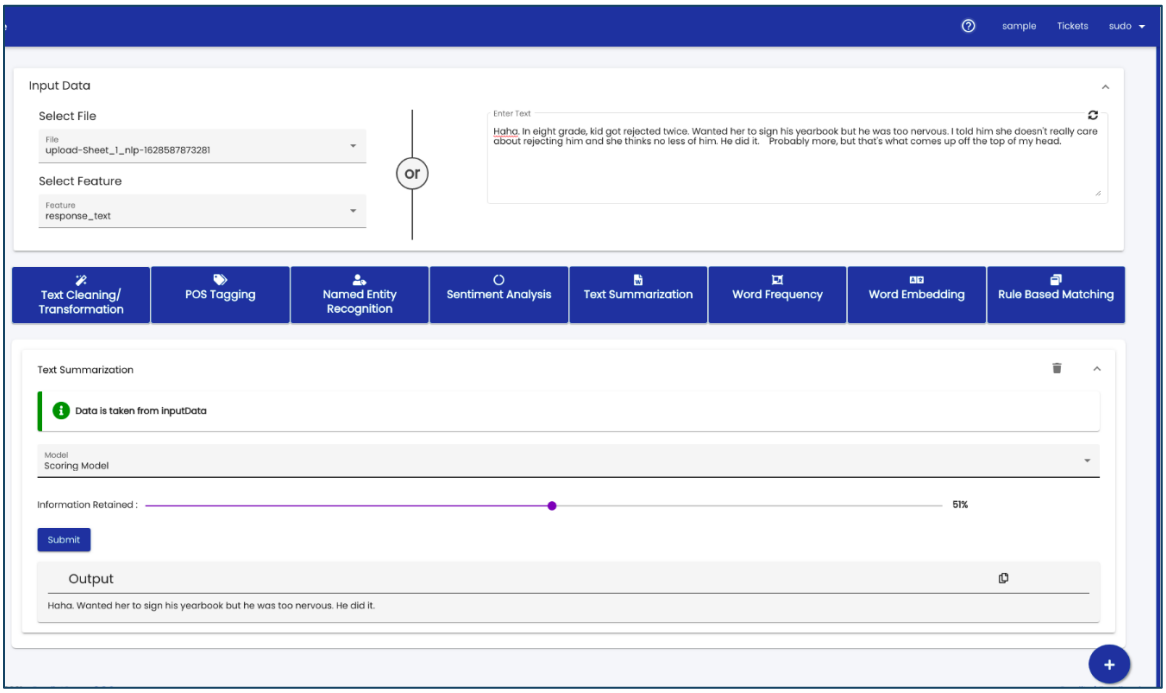
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

Pipeline setting for Sentiment analysis

4.5 Text summarization

Text summarization refers to the technique of shortening long pieces of text. The intention is to create a coherent and fluent summary having only the main points outlined in the document.

Automatic text summarization is a common problem in machine learning and natural language processing (NLP).

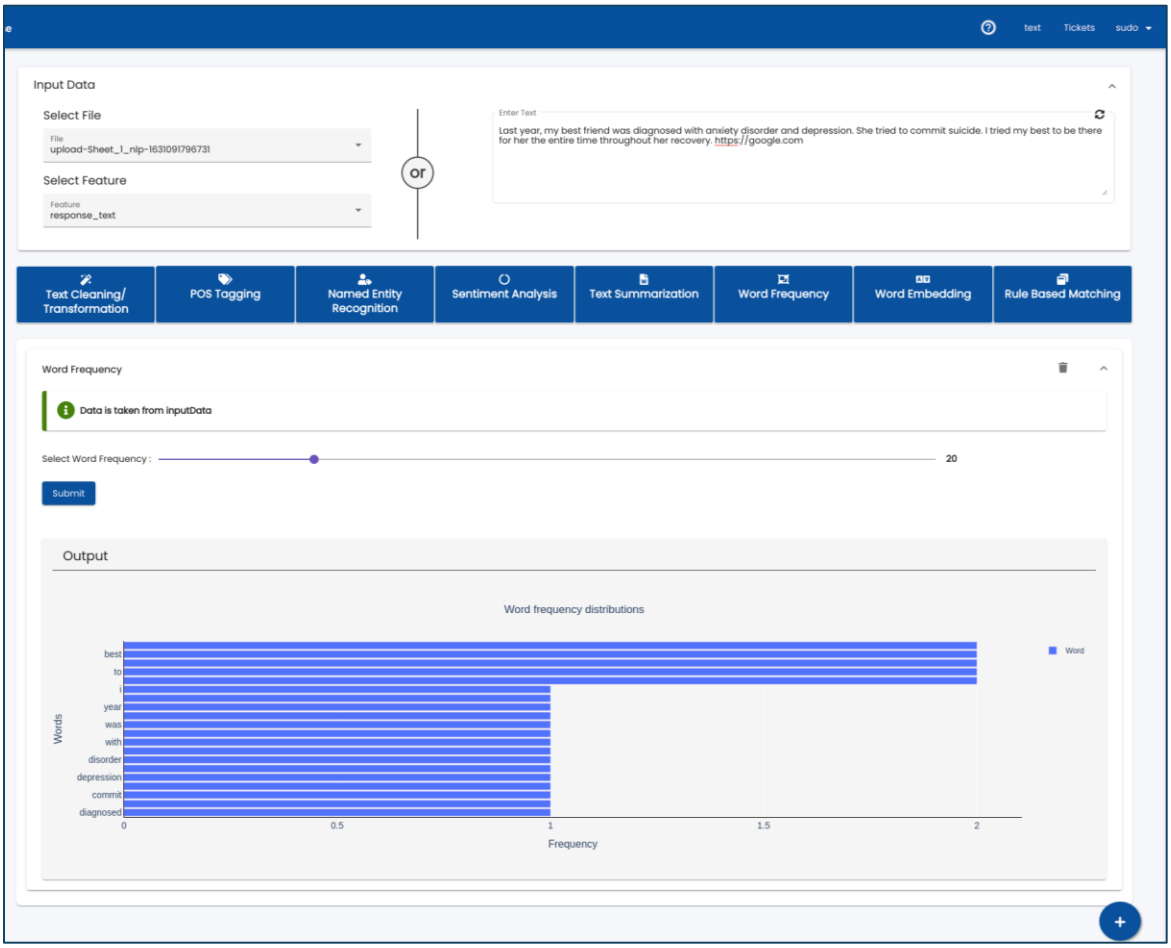


Text summarization

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

4.6 Word Frequency

It lets you know that how many times a particular word is used in an entire text. It counts every word of your text and displays the number of repetition.



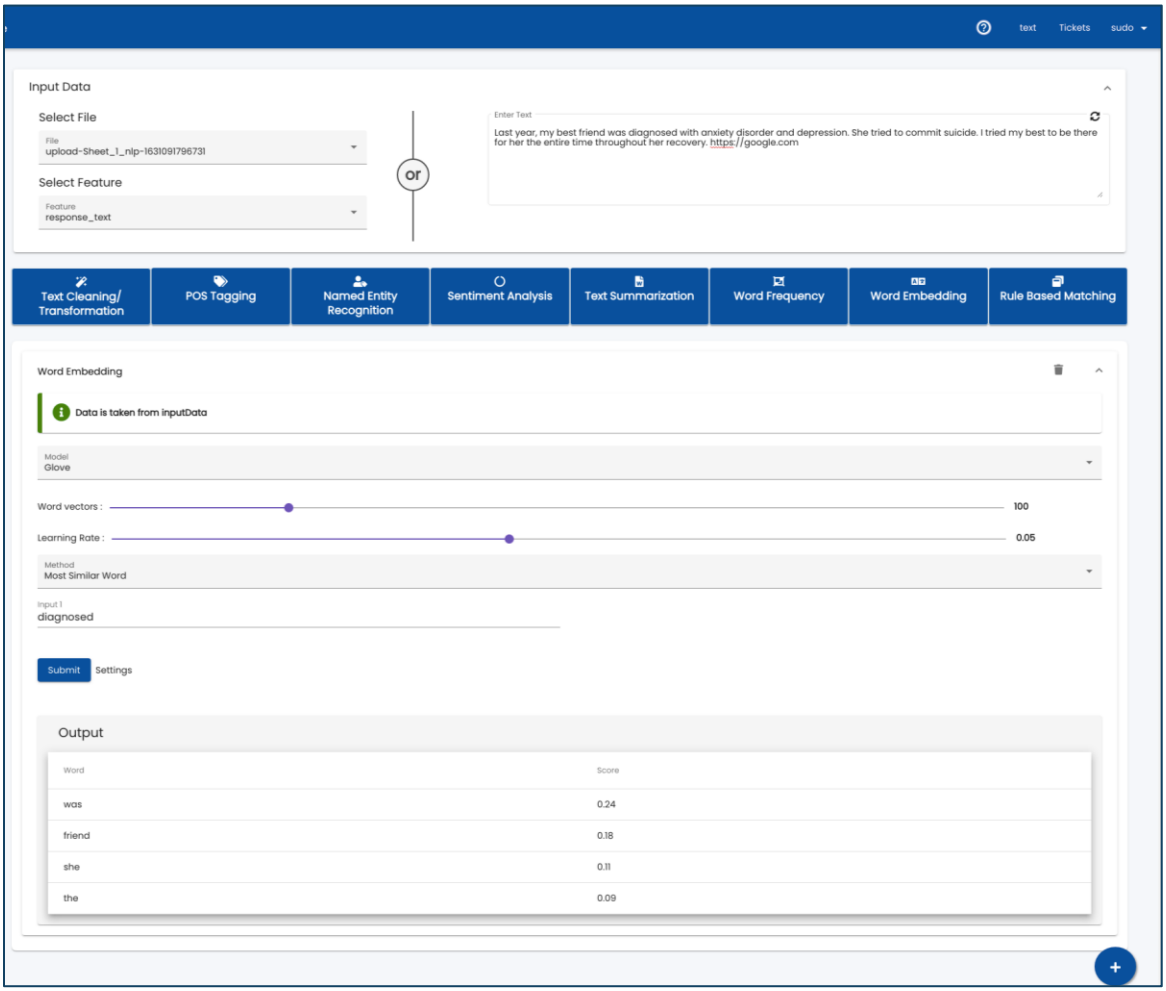
Word Frequency

4.7 Word Embedding –

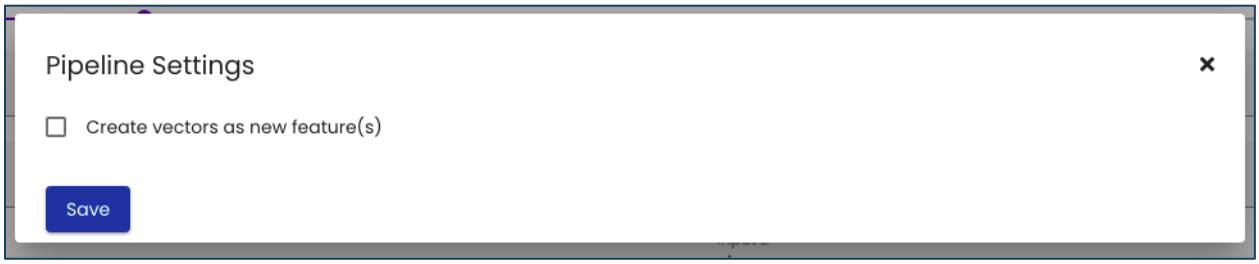
Word embedding are a type of word representation that allows words with similar meaning to have a similar representation. You can check the most similar word, next

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

output word, similarity score between words. Word embedding is a term used for the representation of words for text analysis, typically in the form of a real-valued vector that encodes the meaning of the word such that the words that are closer in the vector space are expected to be similar in meaning.



Word embedding



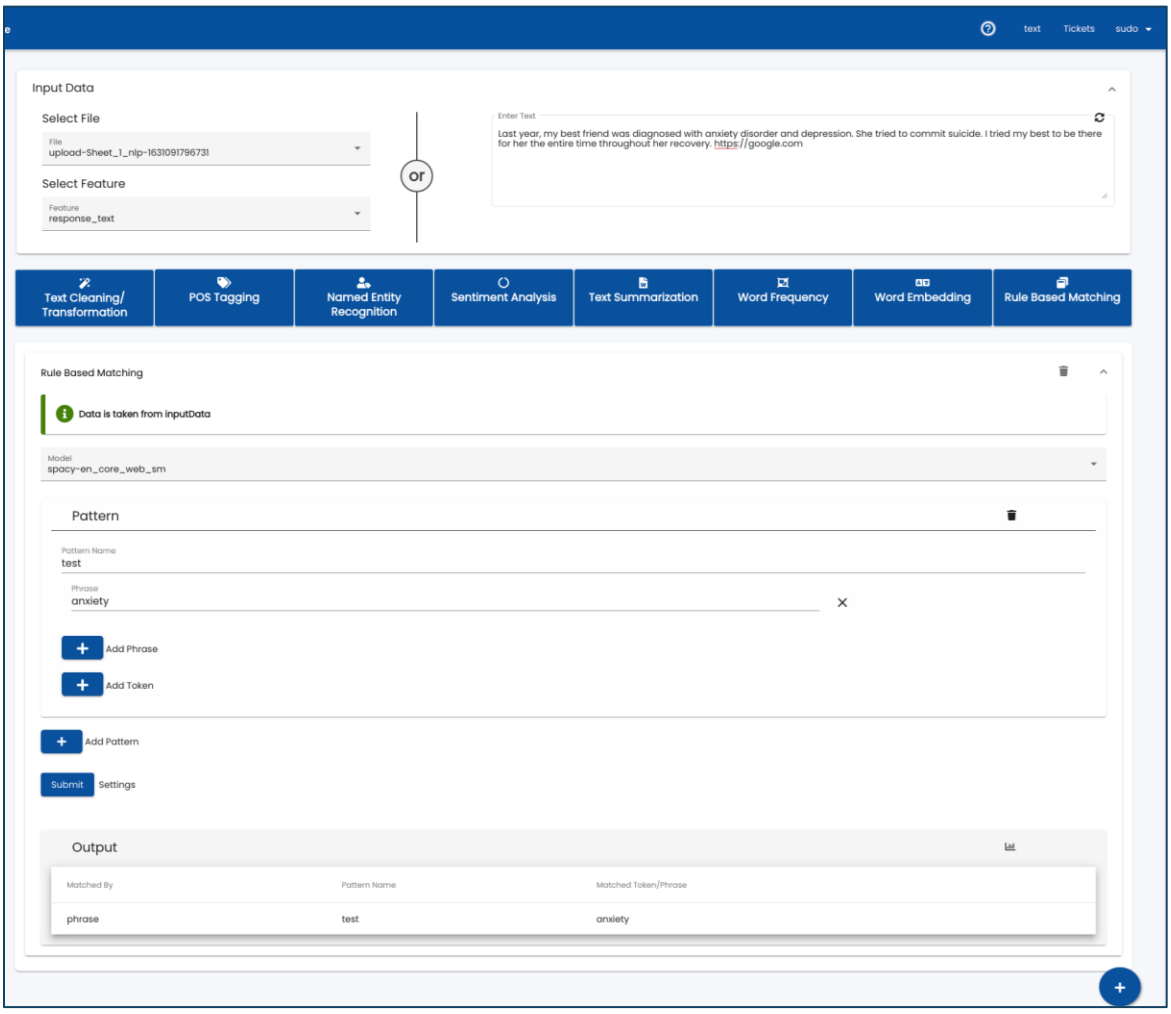
Pipeline setting for Word embedding

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

4.8 Rule Based Matching –

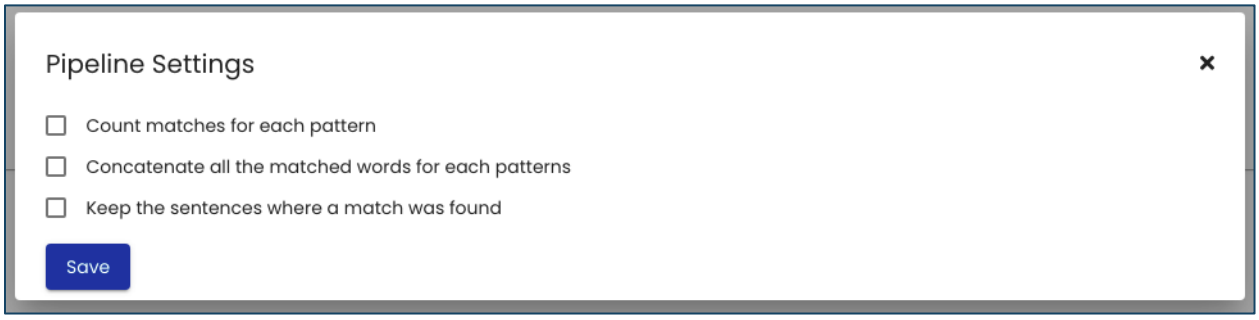
This helps us match token, phrases and entities of words and sentences according to some pre-set patterns along with the features such as parts-of-speech, entity types, dependency parsing, lemmatization and many more.

By selecting the checkboxes in pipeline settings you can create the new features and save the experiment to use the newly generated features to build the models.



Rule Based Matching

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



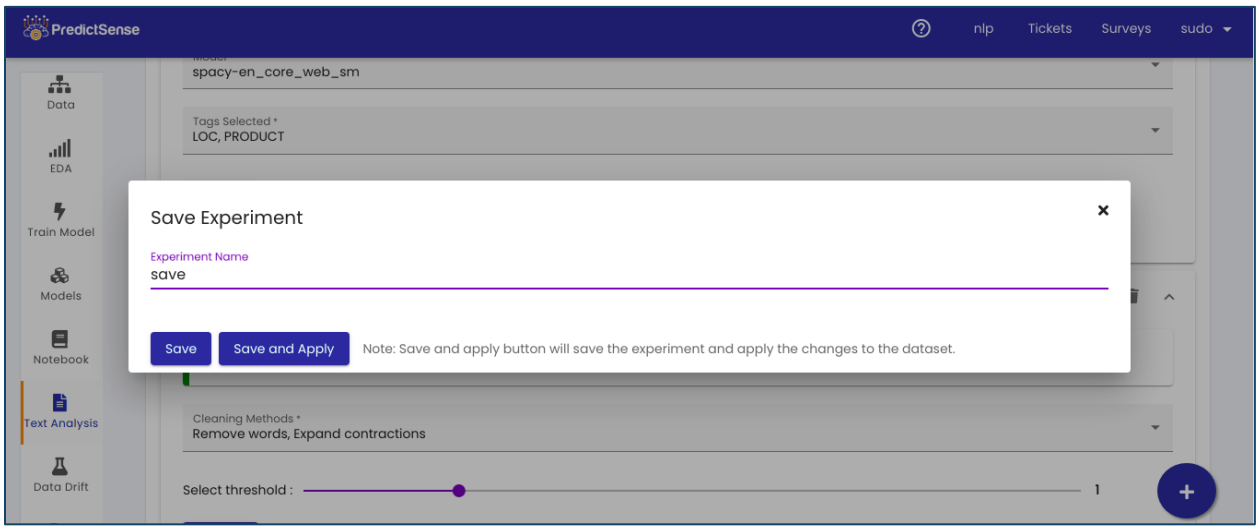
Pipeline setting for Rule Based Matching

Rule Based Matching Graph

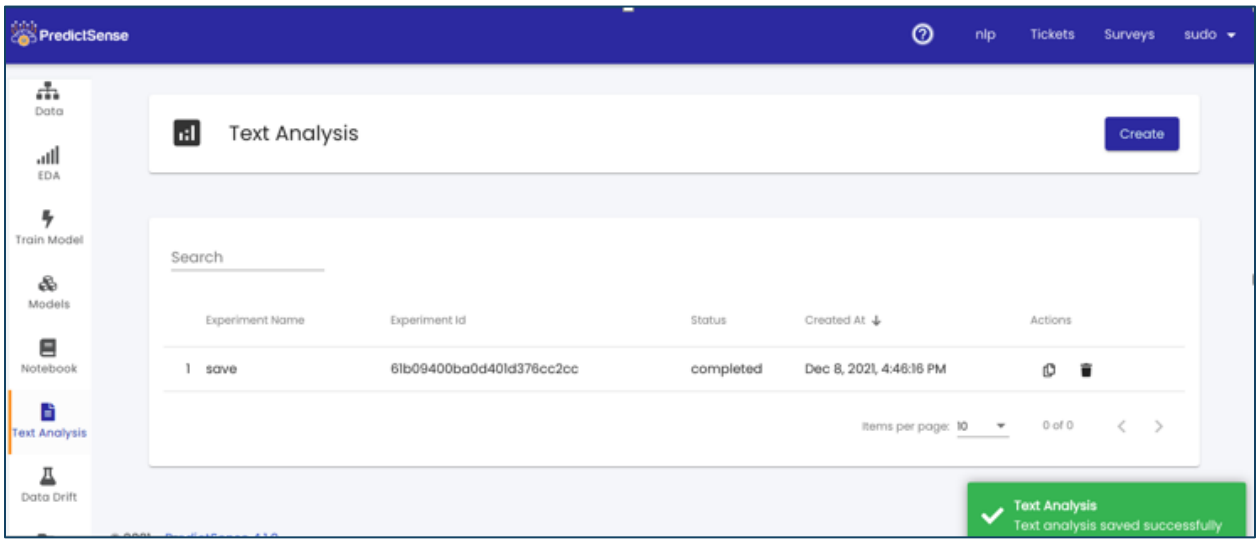
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

4.9 Save experiment

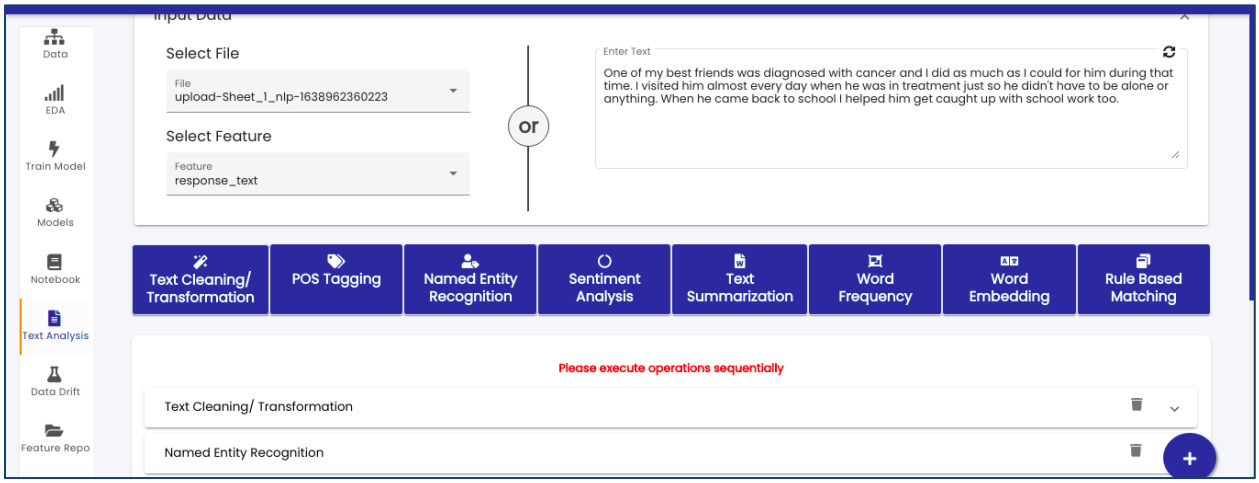
User can save all the operations done in text analysis module by clicking on save button. The text analysis module shows the list of saved operations with following details like experiment name, experiment id, created at and actions fields. Also user can edit the save experiment and need to execute operations.



Save experiment



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

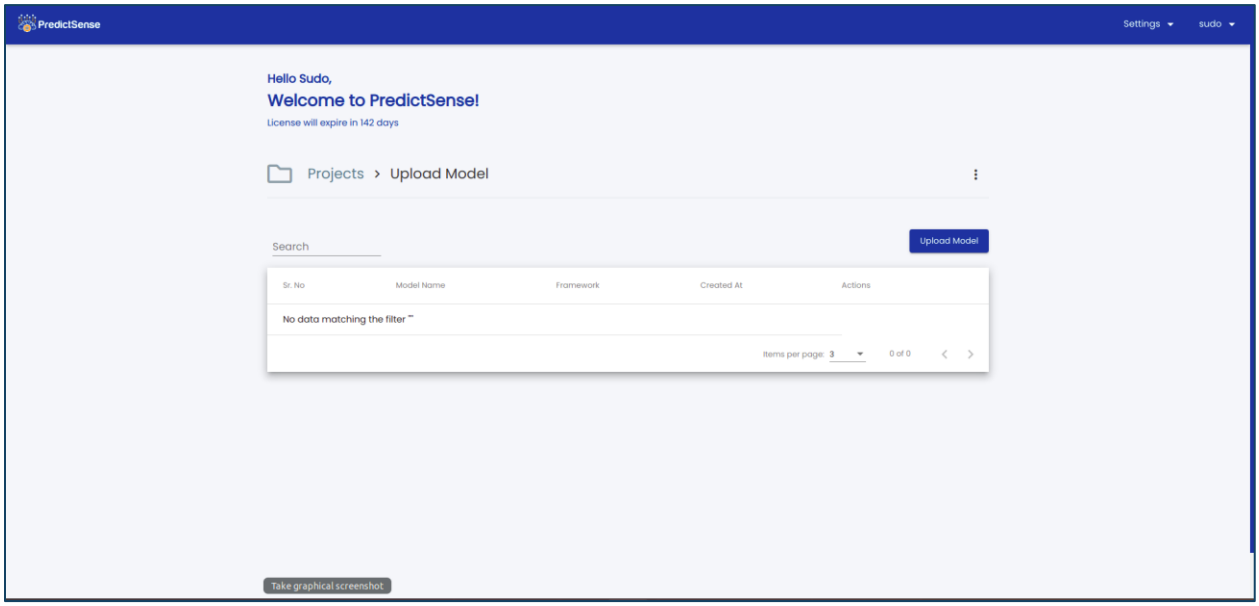


Edit experiment

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

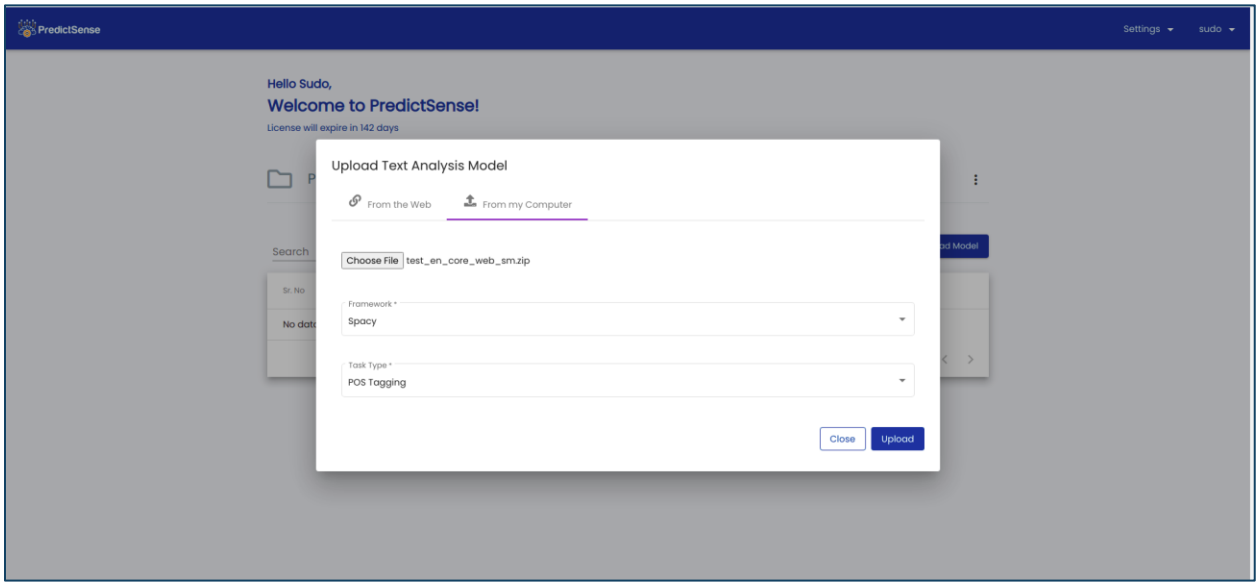
5 Upload model

In upload model you can upload the custom models and also you can use that uploaded model in text analysis for POS and NER to perform the operations on text data.

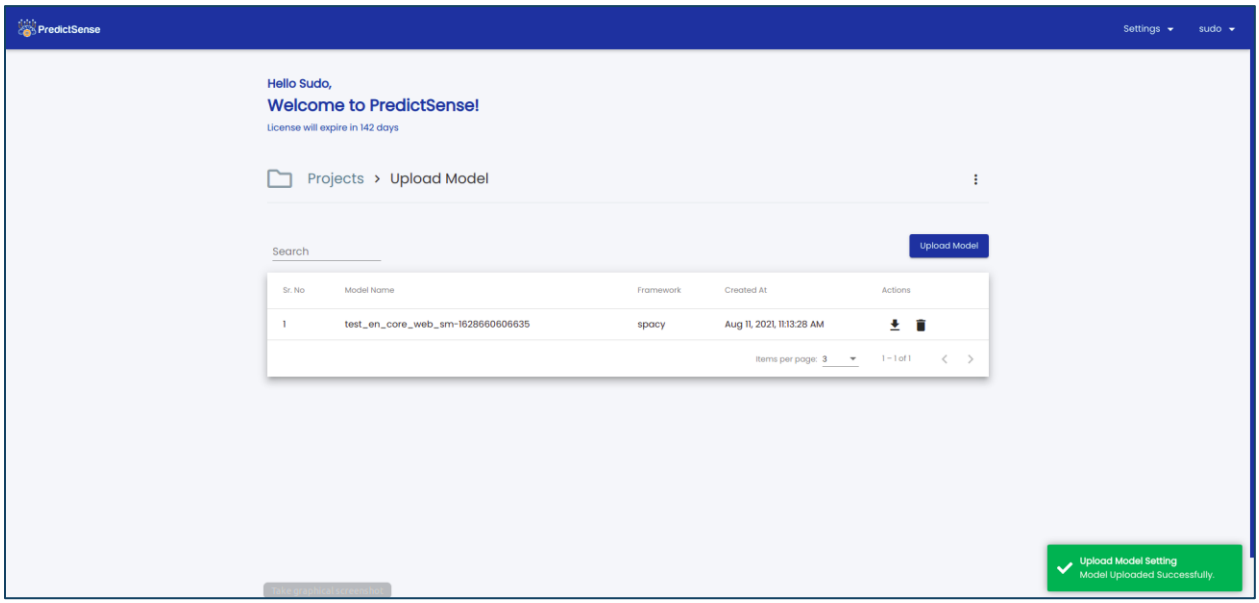


Upload model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

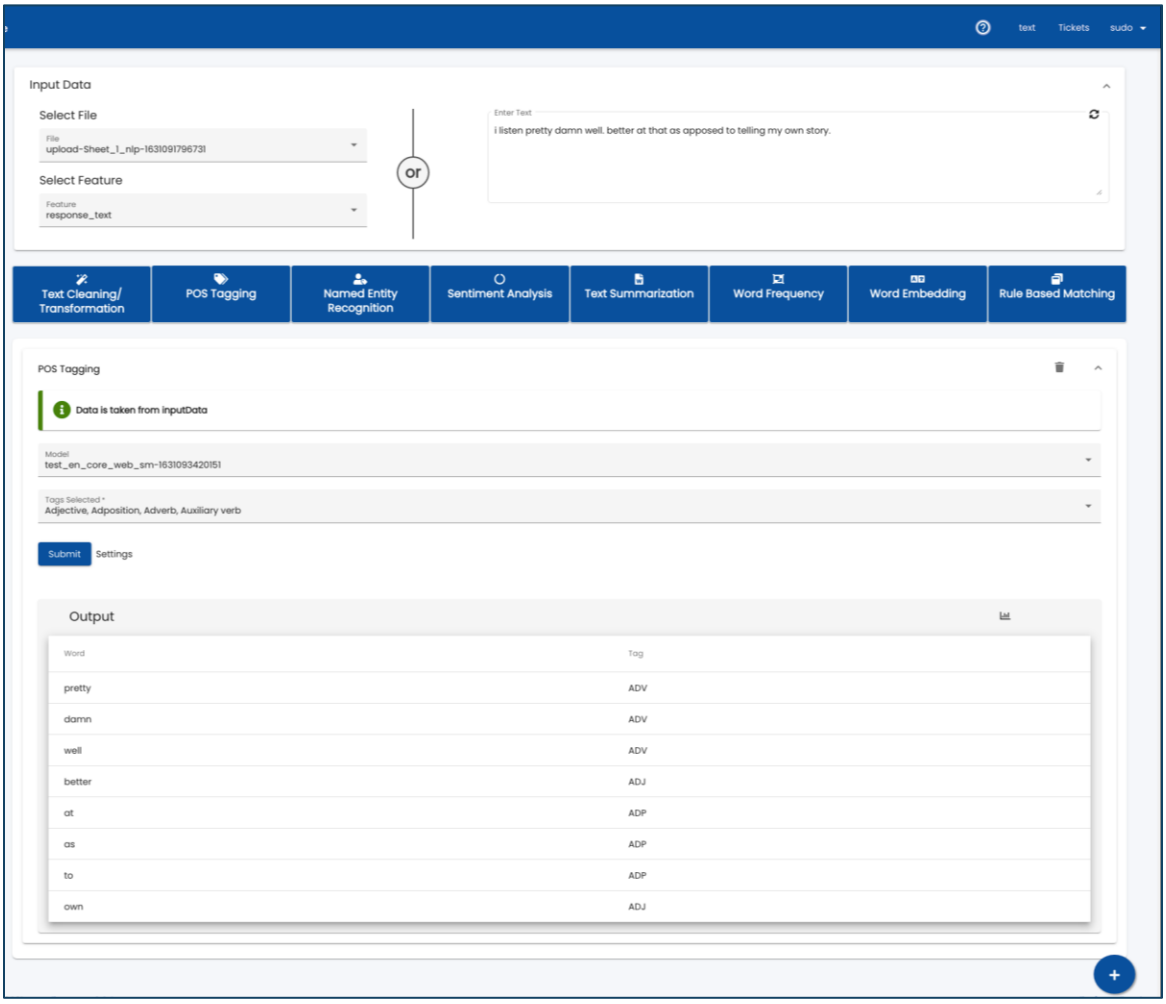


Upload Text analysis model



After upload page

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

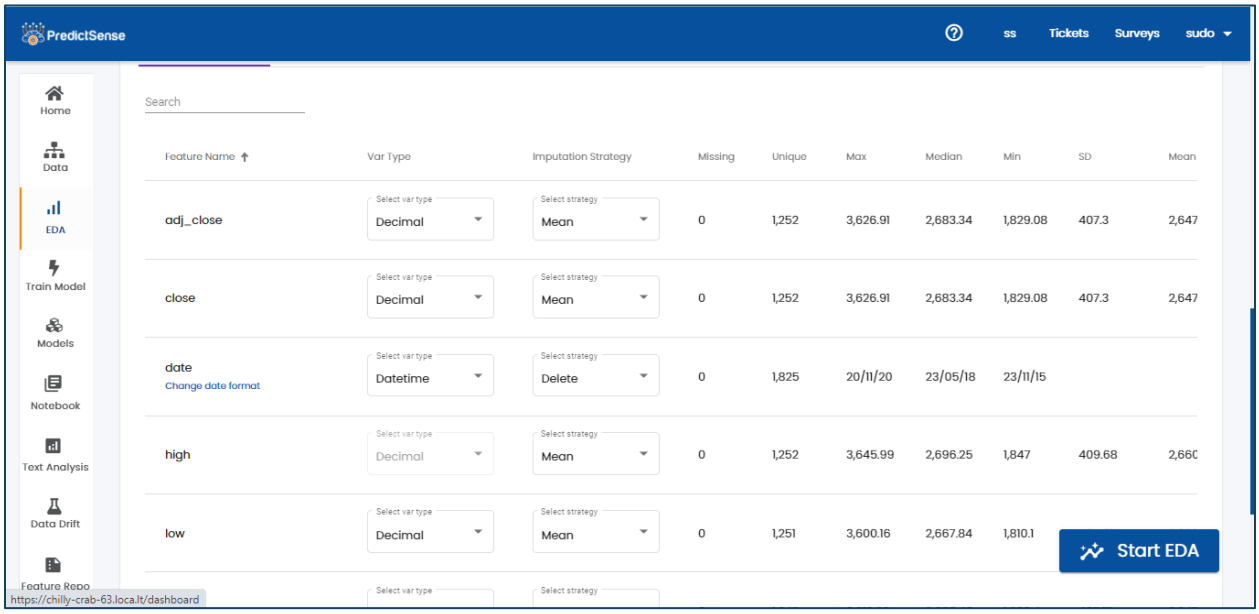


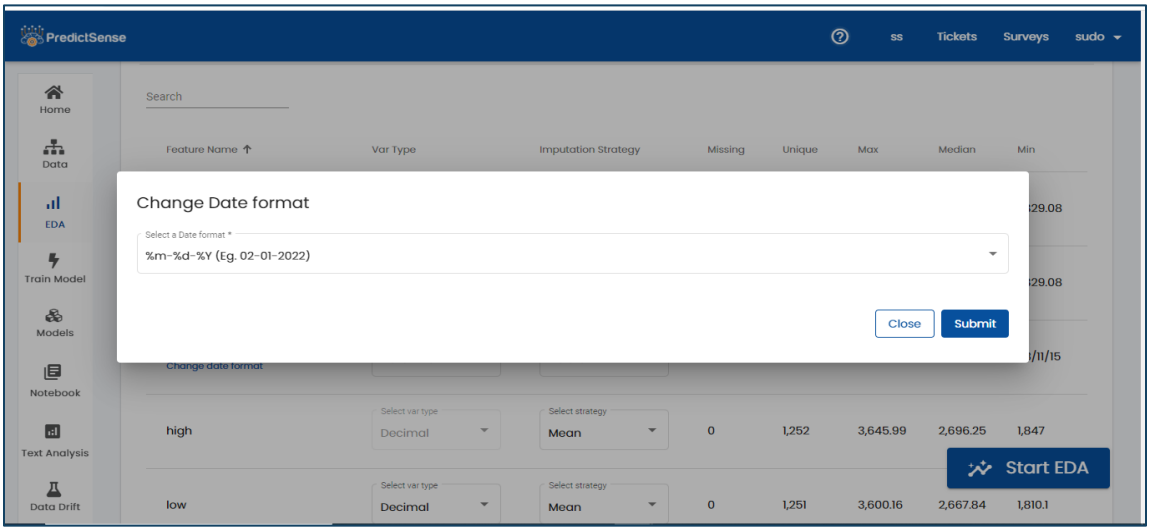
Text model upload view

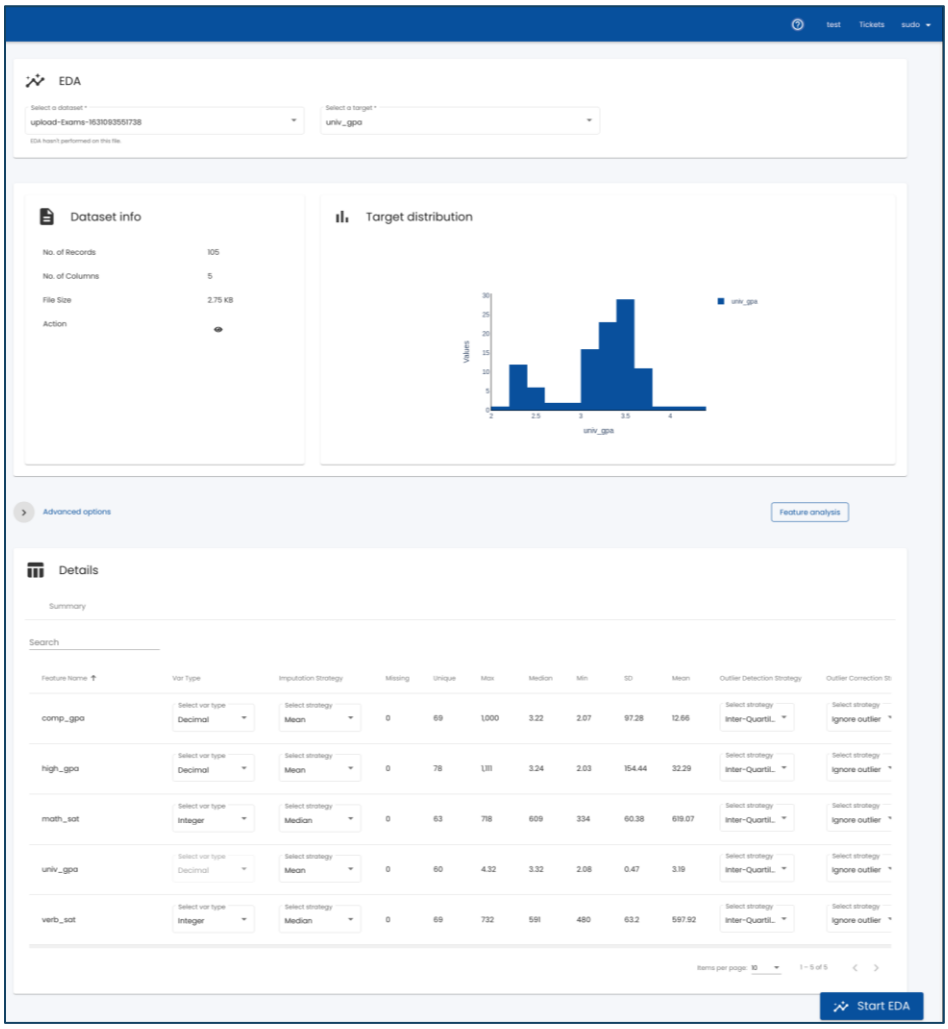
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

6 EDA module

EDA is the process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypotheses with the help of summary statistics and graphical representations. EDA can be done by selecting dataset from select dropdown “Dataset” and target feature from select target dropdown. And after that clicking on “start EDA”. In EDA there different stages for data cleaning and imputation of missing values, Handling outliers and calculating different statistics. Date feature will be manipulated as per date change format.

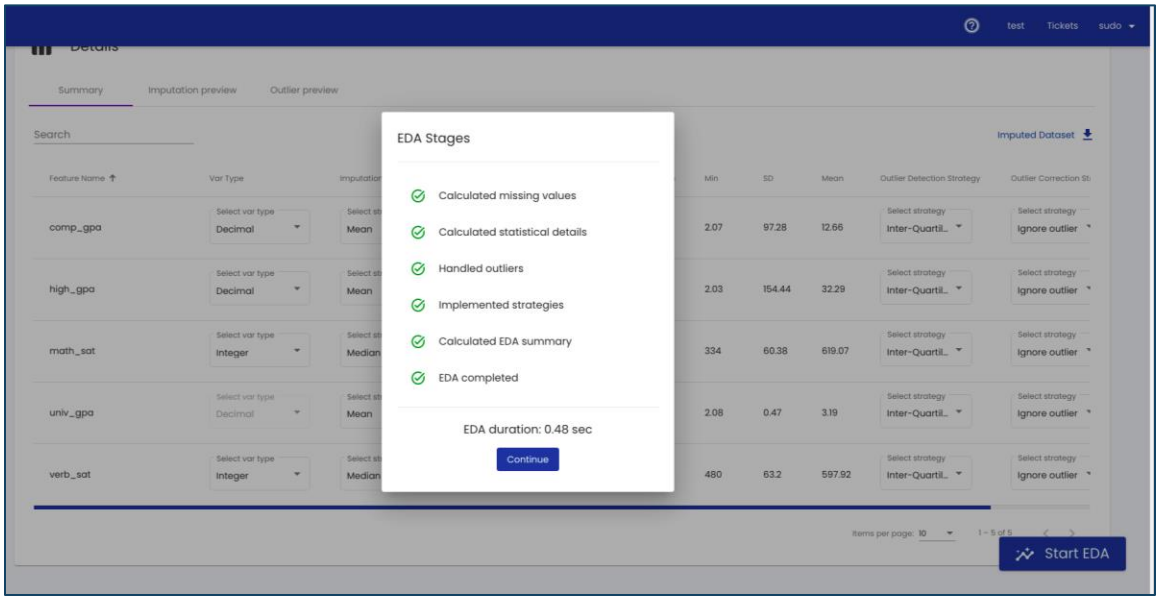






EDA Page

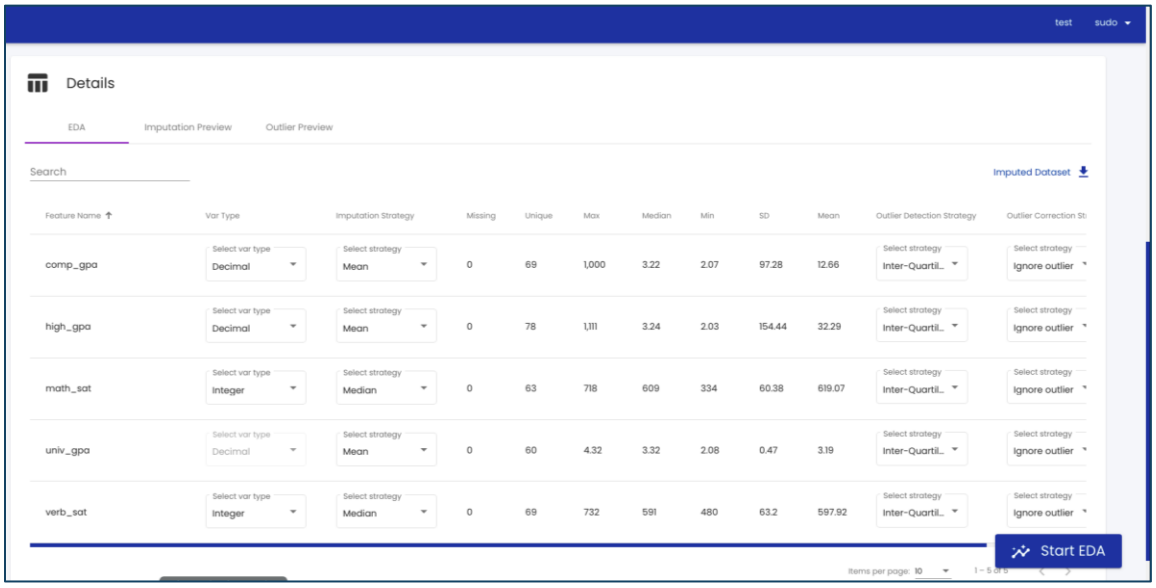
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



EDA Stages

After completing of EDA, EDA details are displayed with EDA summary, Missing value imputation and Outlier preview.

6.1 EDA details



EDA details

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

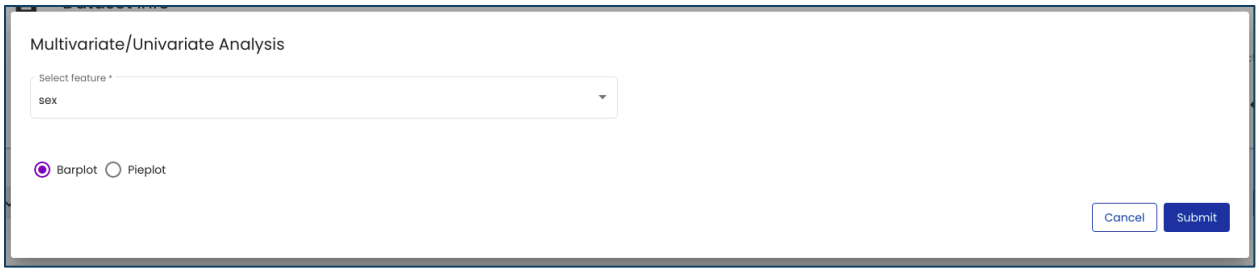
6.2 Manual selection

User can change the default methods for "**Imputation of missing value", "Outlier detection strategy", "Outlier correction strategy", "Var type"** as well as default from given options in EDA.

6.3 Feature analysis

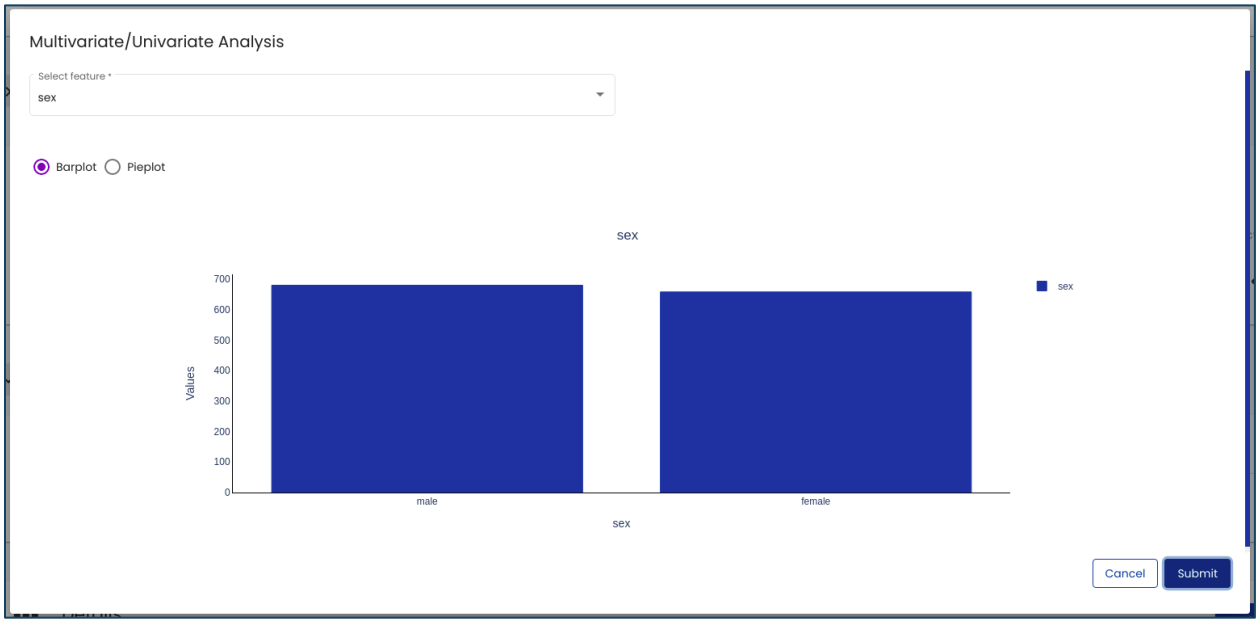
It shows graphical representation of features with statistical details. With more than one feature the comparison of different features in the form of "**bar plot", "box plot", "scatter plot", "Pie plot", "Histogram plot", "word cloud", "2D" & "3D" etc.**

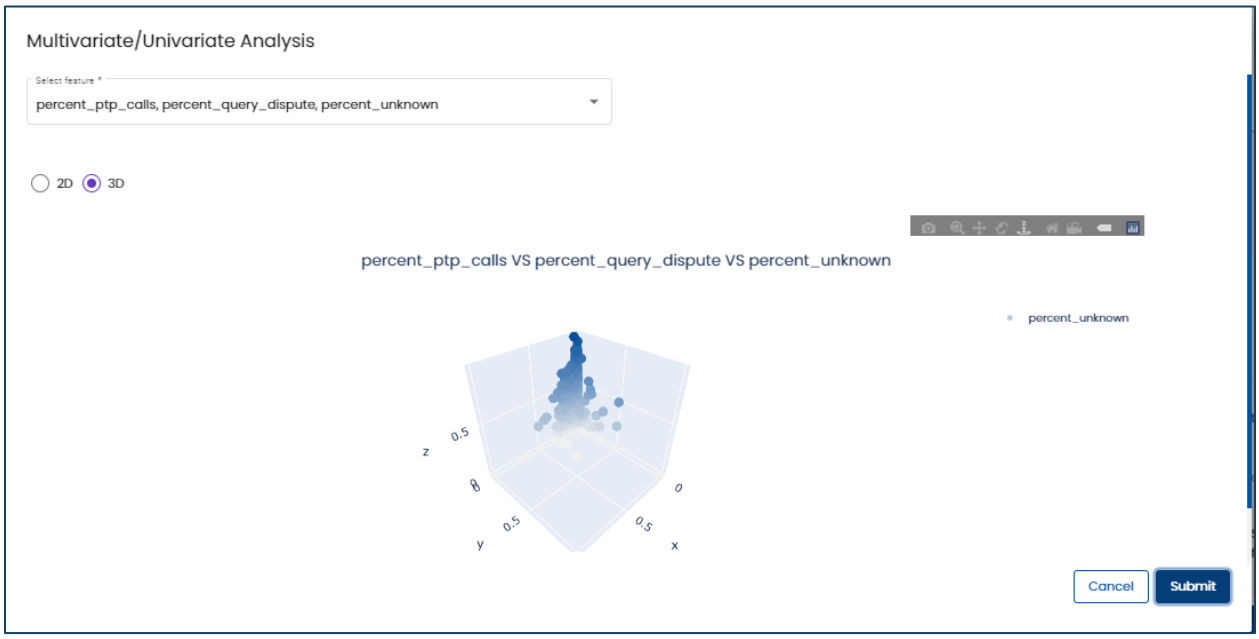
User can select feature from dropdown and after selecting **"bar plot", "line plot", "box plot" and "Histogram plot"** from radio button and submitting it will give detailed visualization and comparison of data. user can also visualize two-dimensional and three-dimensional plot for combination of three numeric feature. Systematic sampling is performed when dataset row count is exceed above 50k row for numeric feature.



Feature analysis

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



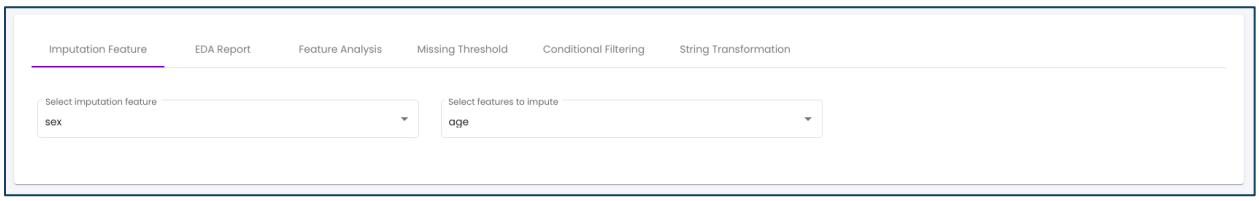


Feature analysis report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

6.4 Imputation Feature

If user knows target variable in advance then selecting that as imputation feature, all the operations of manual eda are handled according to the imputation feature

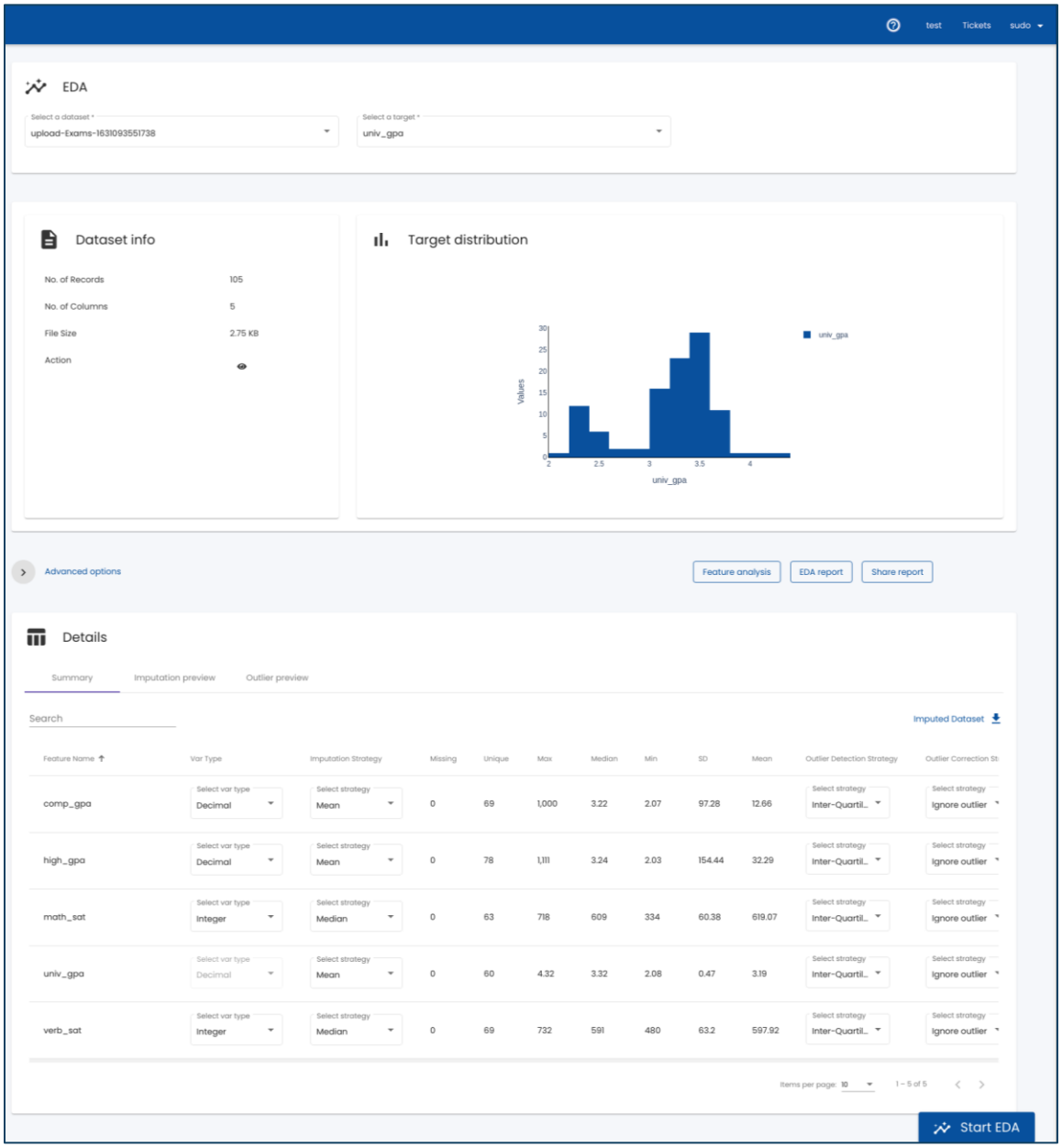


Imputation Feature

6.5 Share report

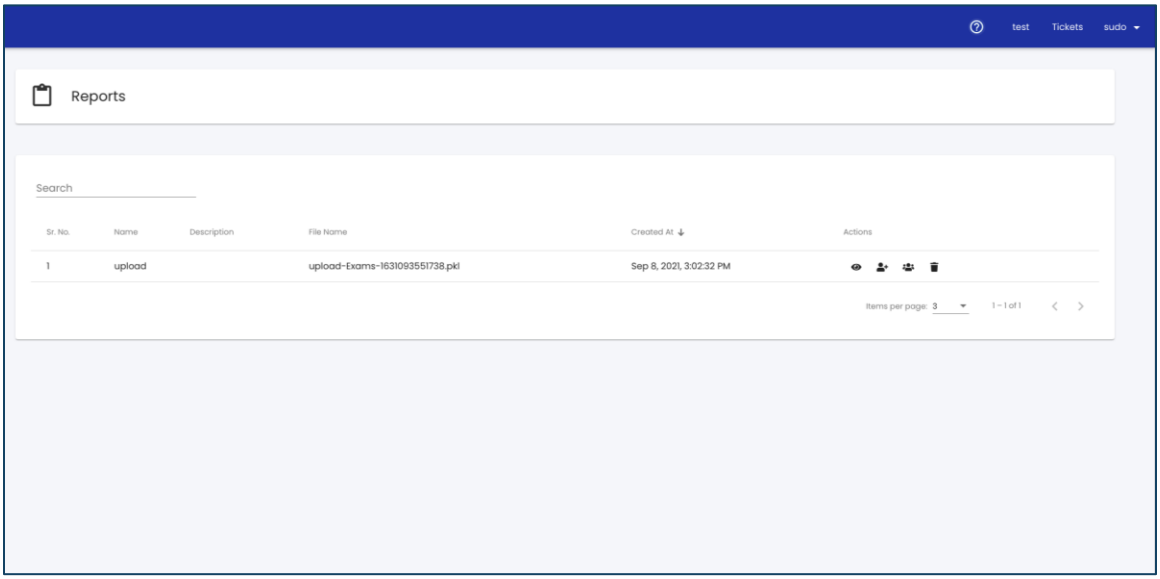
In EDA Module share report option is enable after performing eda operation. Once you click on share report. It will open the output module page were you can see the share report. You can also share the report with another user by passing email id and by clicking on share button.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

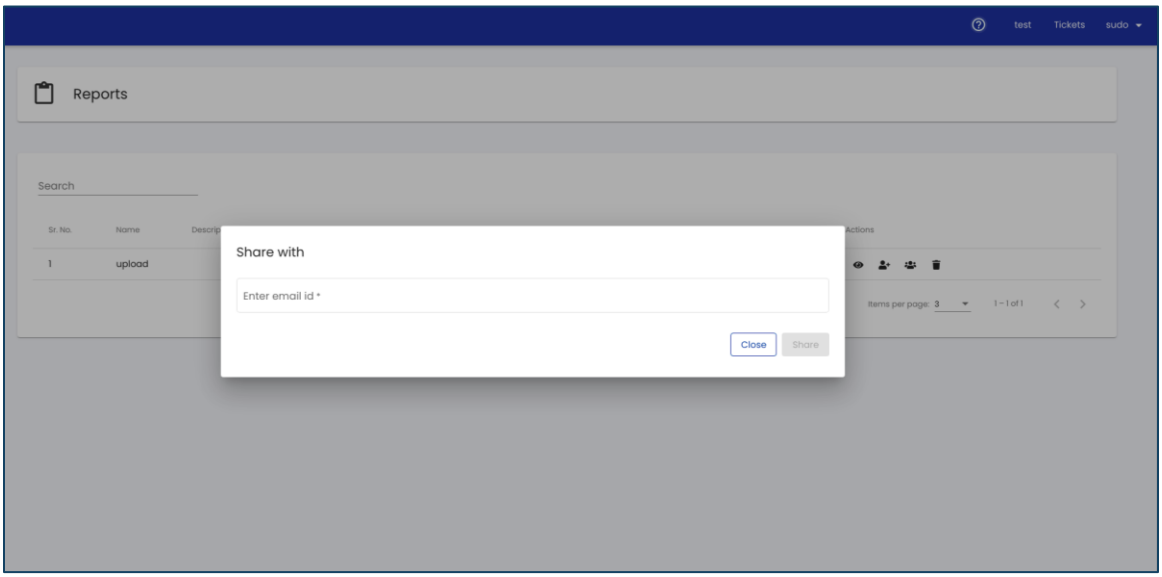


EDA Share Report button

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



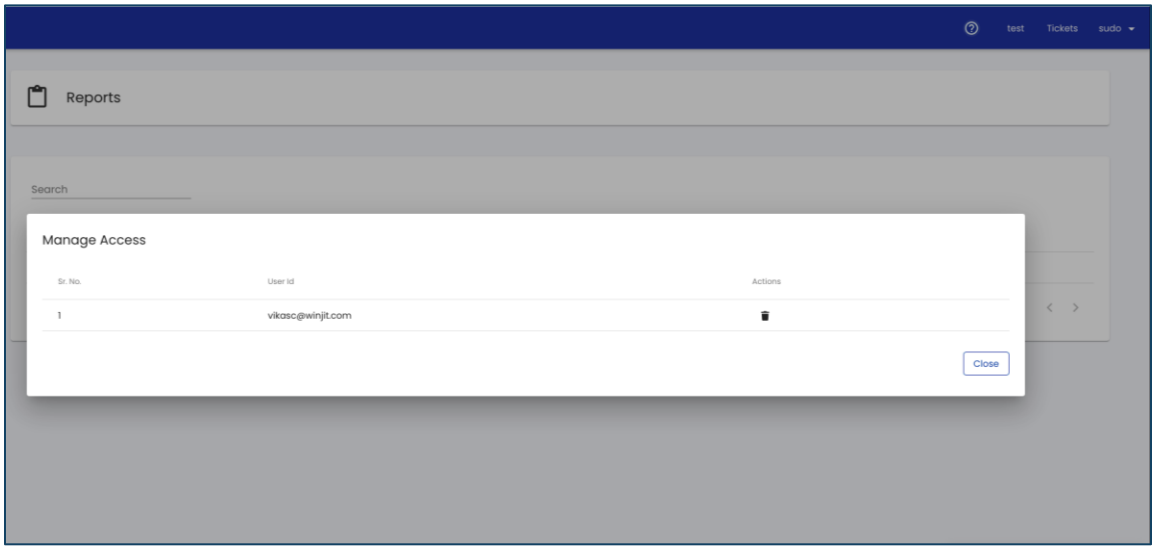
Share Report



Share with user

Using share with user feature you can share the report with different user like developer, report user, admin user role.

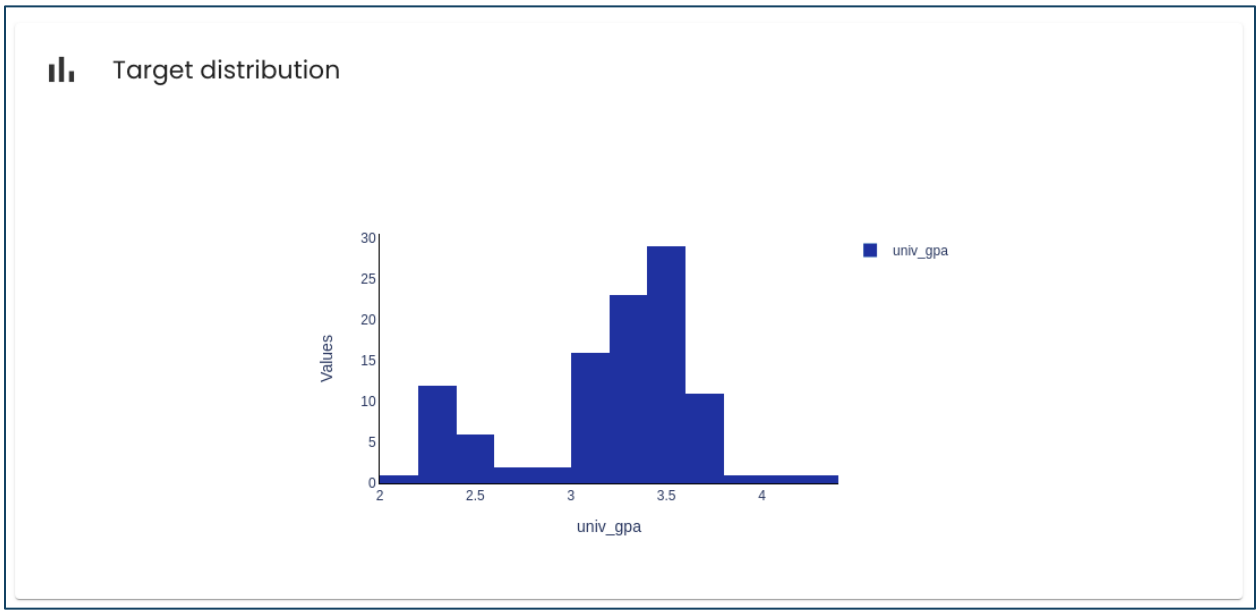
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Manage access to share with user

6.6 Target distribution

Based on target feature selection in eda module you can see the different plots in target distribution.



Target Feature Graph

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

6.7 Split Dataset

The data should ideally be divided into 3 sets – namely, train, test, and holdout cross-validation or development (dev) set. Let's first understand in brief what these sets mean and what type of data they should have.

1. **Train Set:**

The sample of data used to fit the model that is the actual subset of the dataset that we use to train the model (estimating the weights and biases in the case of Neural Network). The model observes and learns from this data and optimize its parameters.

2. **Dev Set:**

The development set is used to validate the trained model. This is the most important setting as it will form the basis of our model evaluation. If the difference between error on the training set and error on the dev set is huge, it means the model as high variance and hence, a case of over-fitting.

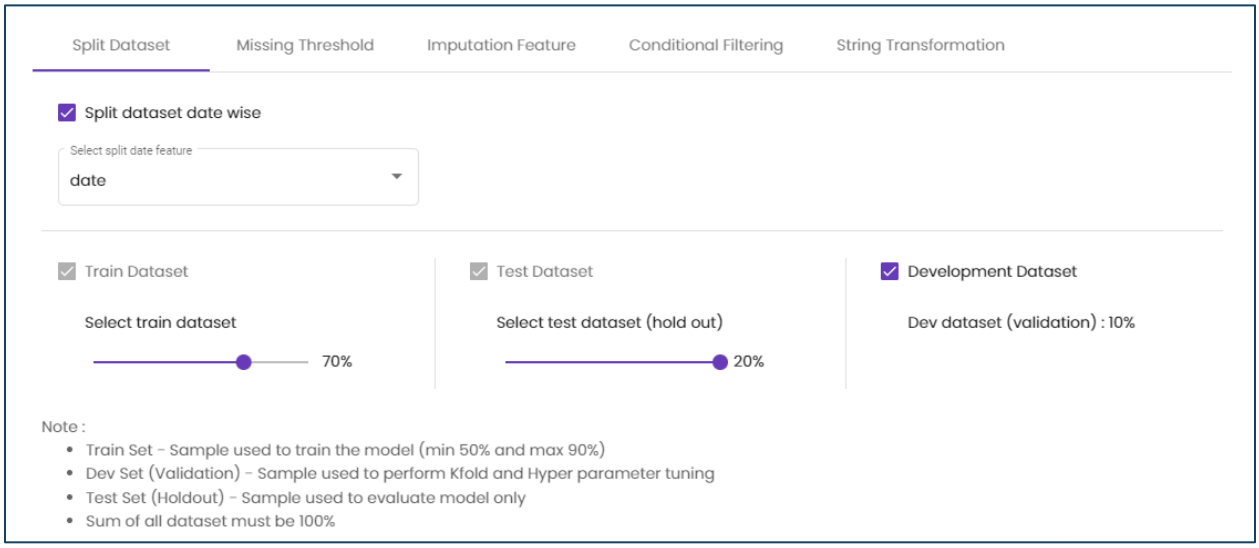
3. **Test Set:**

The test set contains the data on which we test the trained and validated model. It tells us how efficient our overall model is and how likely is it going to predict something which does not make sense. There are a plethora of evaluation metrics (like precision, recall, accuracy, etc.) which can be used to measure the performance of our model.

4. **Date wise:**

Only for predictive modelling the data containing date feature user can split the dataset. This split happens and data get divided in above the sets. It is an optional feature.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Split dataset

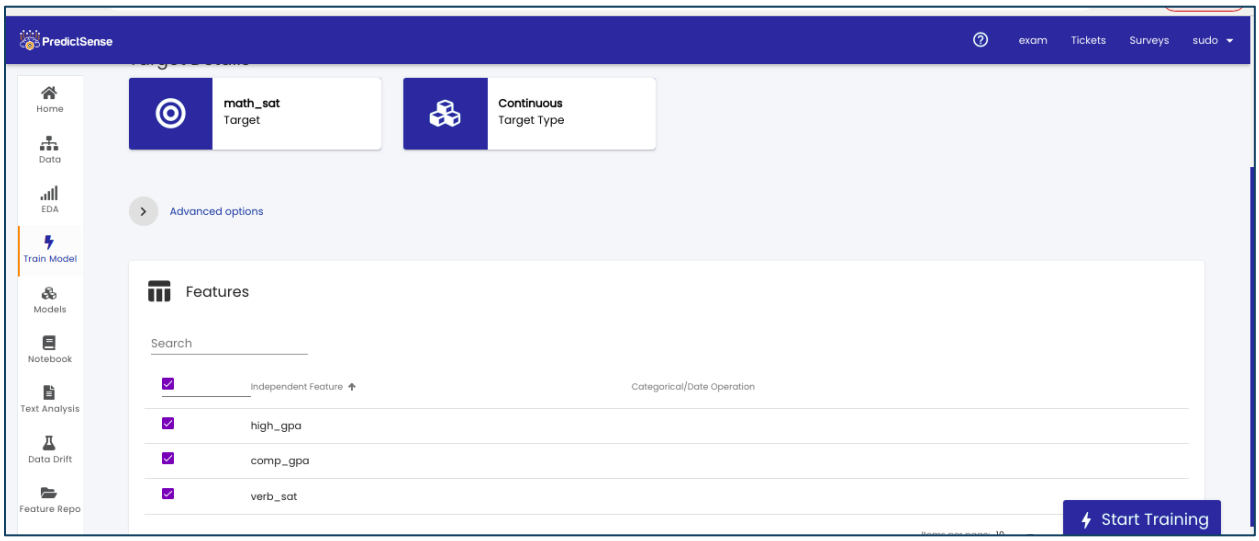
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

7 Train model

User can train model on target feature which decides the problem type to classification or regression. Based on the selected target in EDA module PredictSense will automatically suggest relevant algorithms in train model pages For Training a specific model, various advanced training options like **feature generation, feature transformation, feature reduction, sampling, Algorithm Parameters** can be selected to improve the training of model.

7.1 Training steps

1. Select dataset
2. Select different algorithms from list of algorithms
3. Select multiple independent features
4. Click on start training

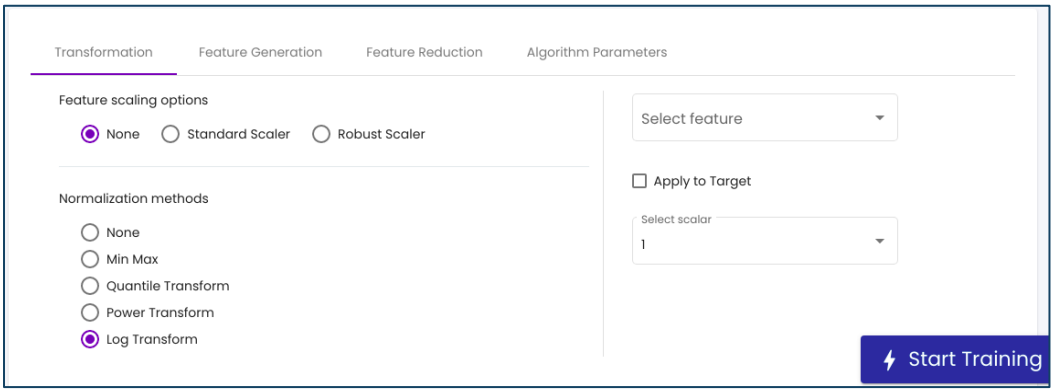


Train model page

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

7.2 Training steps with advanced options

1. Select dataset
2. Select different algorithms from list of algorithms
3. Select various options from advanced options by switching between tabs like **"Feature generation", "Transformation", "Feature reduction", "Algorithm Parameters"** etc different methods for **scaling, normalization, dimensionality reduction, sampling, Feature Selection.**
4. Select multiple independent features
5. Click on start training



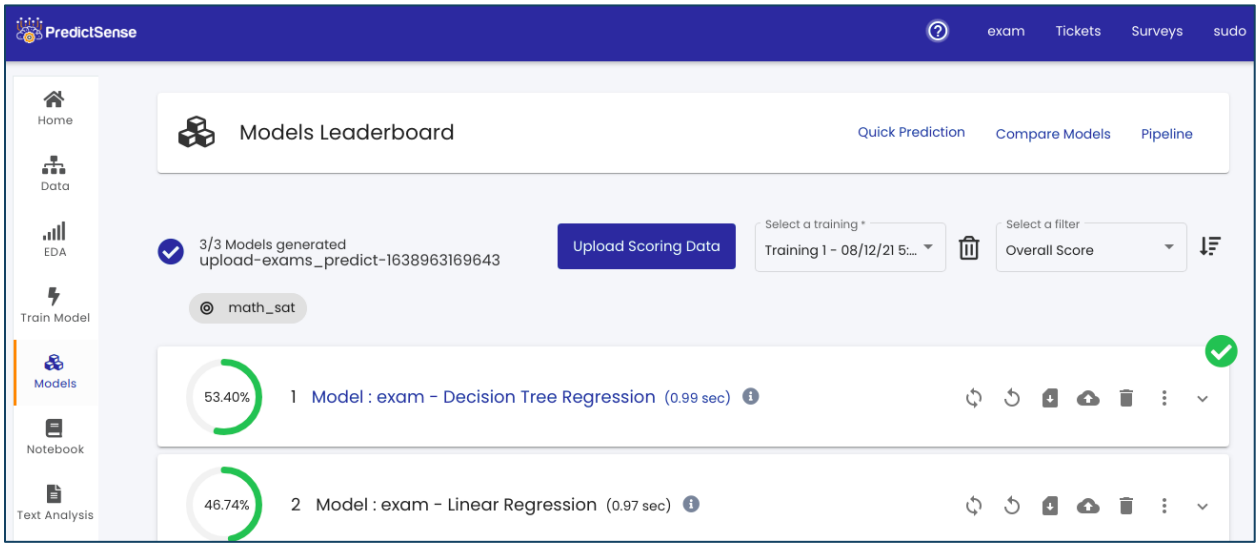
Advanced features

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

8 Models leaderboard

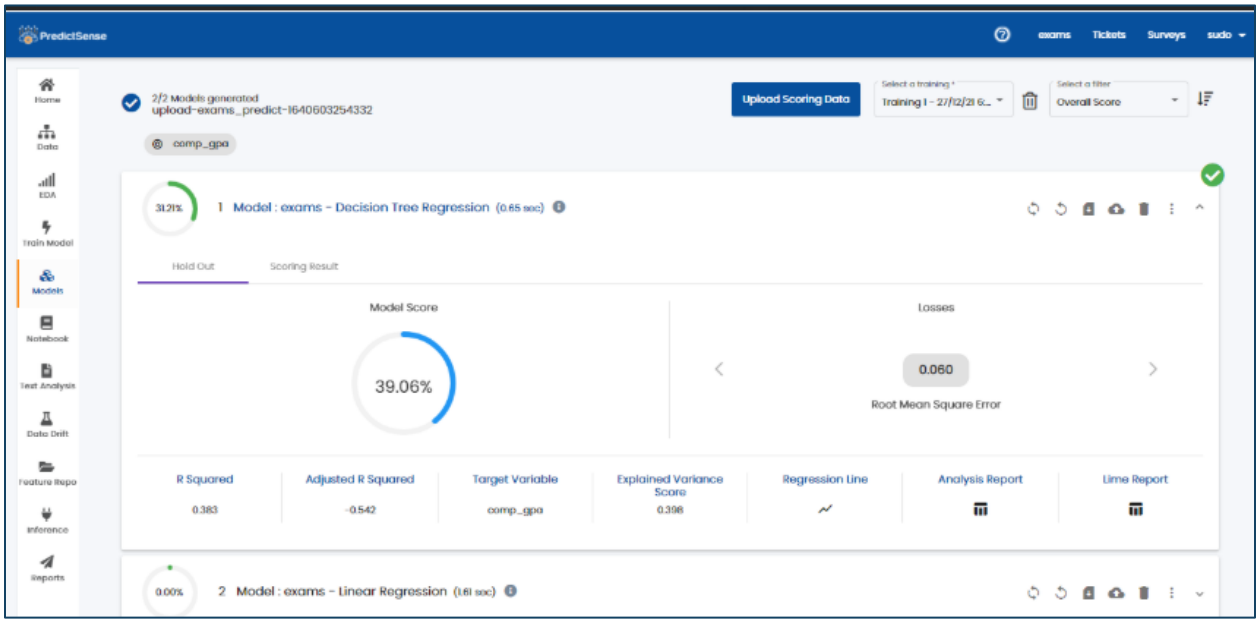
Model leader board shows the entire generated model based on training options selected by user. Leader model is the one with highest overall score.

All the models have different performance matrices like **regression curve, classification report, confusion matrix, lift-gain chart, Accuracy and other parameters of losses.**



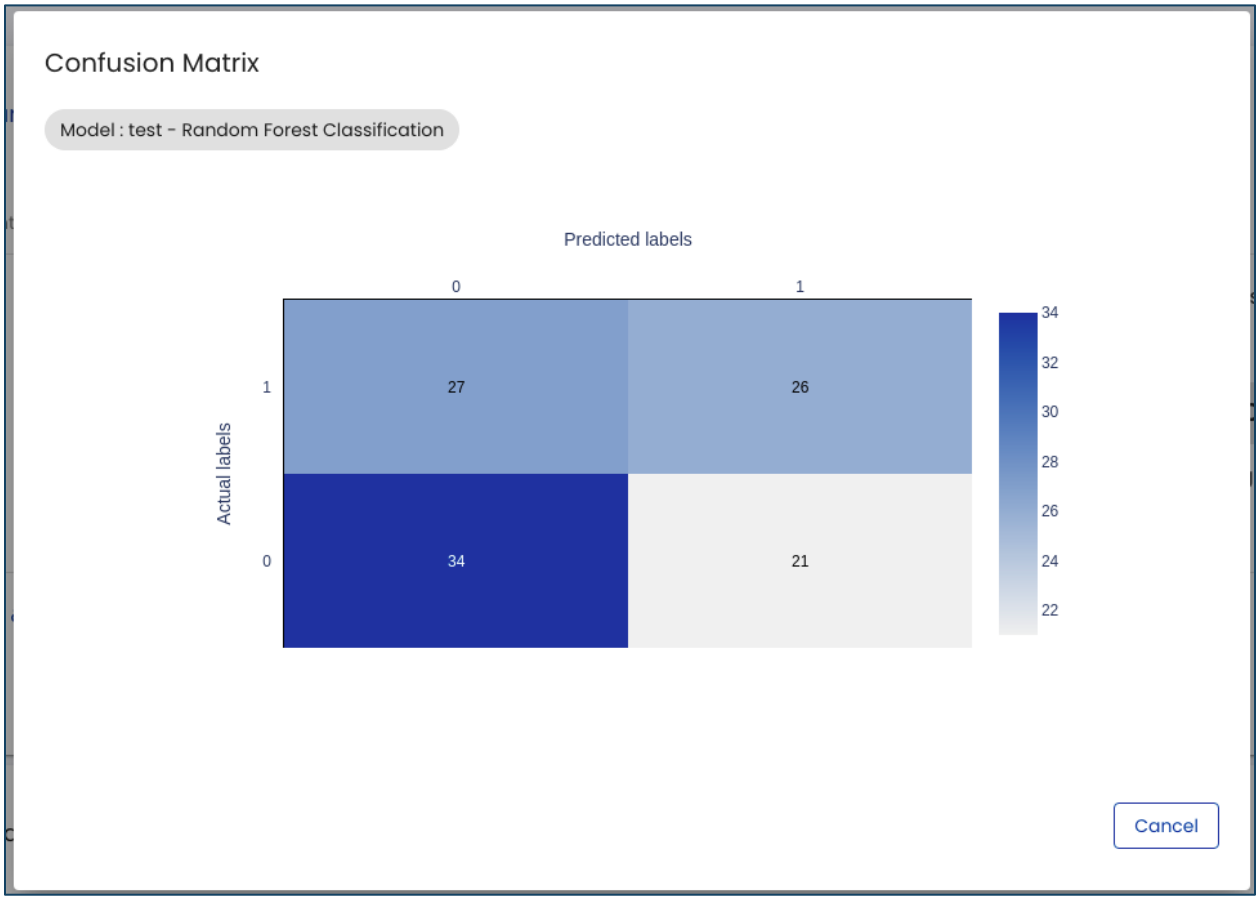
Model Leaderboard

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Detail Page

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



Confusion Matrix

8.1 Model tuning

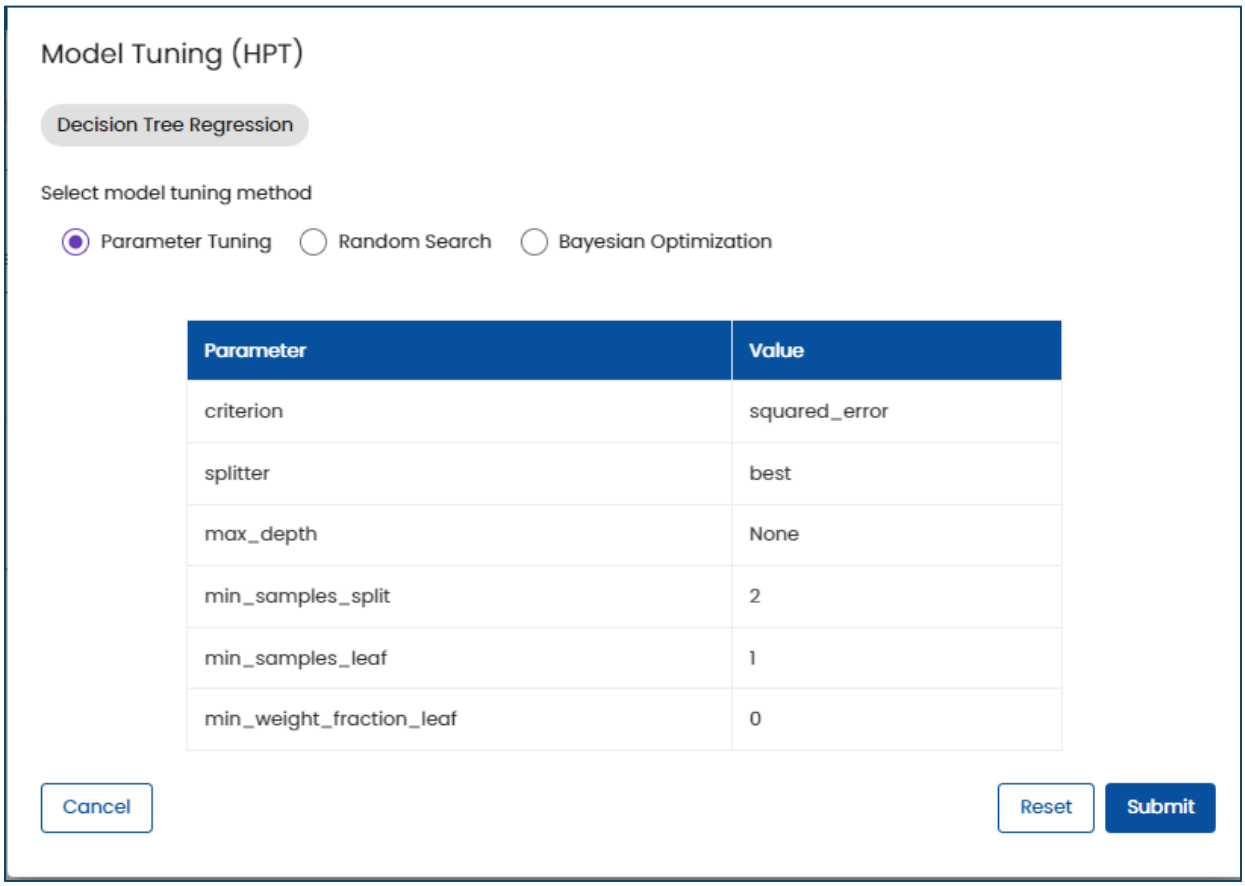
User can change the parameters used for model building by choosing the best parameters and training the model again on that parameters to get best predictions.

There are three methods for hyper parameter tuning.

1. Parameter tuning
2. Random search
3. Bayesian optimization

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

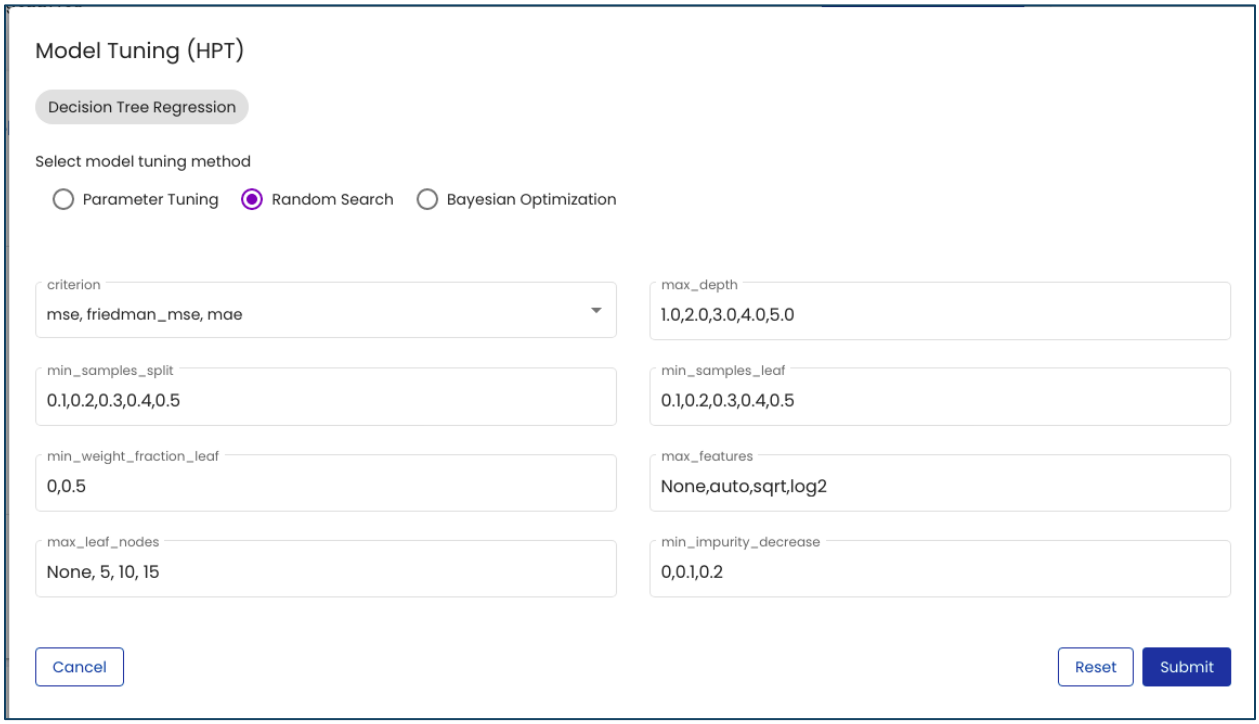
User can change the parameters with reference to parameter used in last turning and clicking on **"start HPT"** will train model with best parameters



Parameter	Value
criterion	squared_error
splitter	best
max_depth	None
min_samples_split	2
min_samples_leaf	1
min_weight_fraction_leaf	0

HPT

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

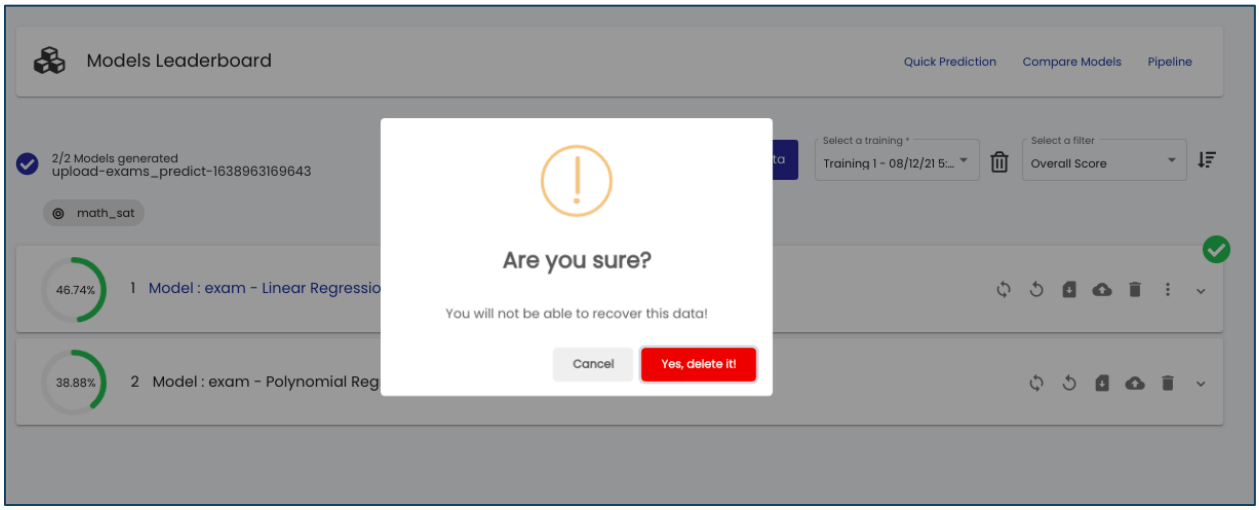


Model Tuning for train model

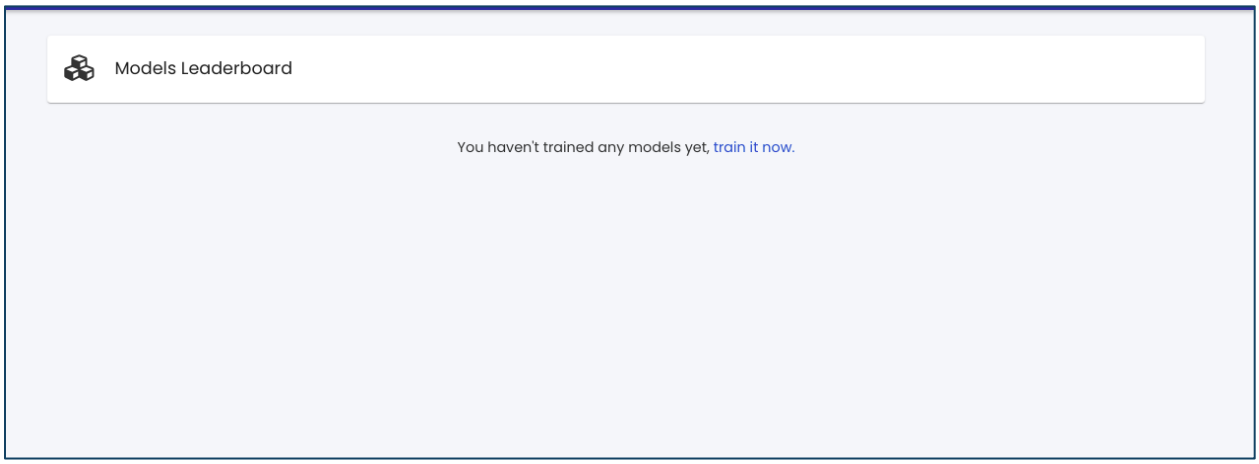
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

8.2 Delete training

Delete training feature is used to delete the any no. of training session. In PredictSense you can train multiple model and also you can delete the different training sessions.



Pop up to validate Delete Training

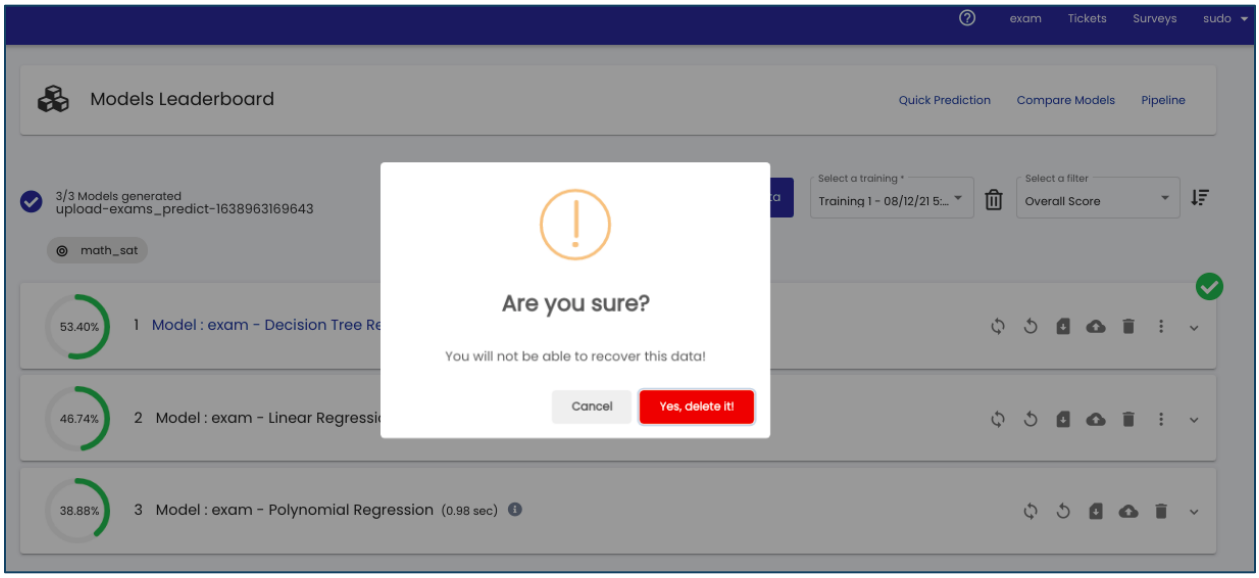


After Delete

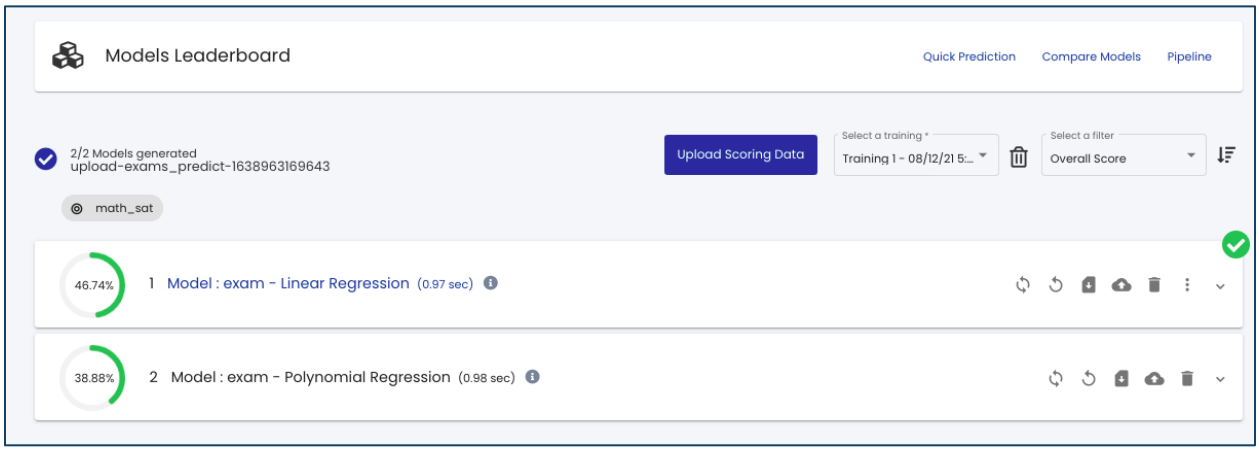
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

8.3 Delete model

Now PredictSense provide the delete model functionality. You can delete any model from any training session by clicking on delete button in models window.



Pop up to validate Delete model



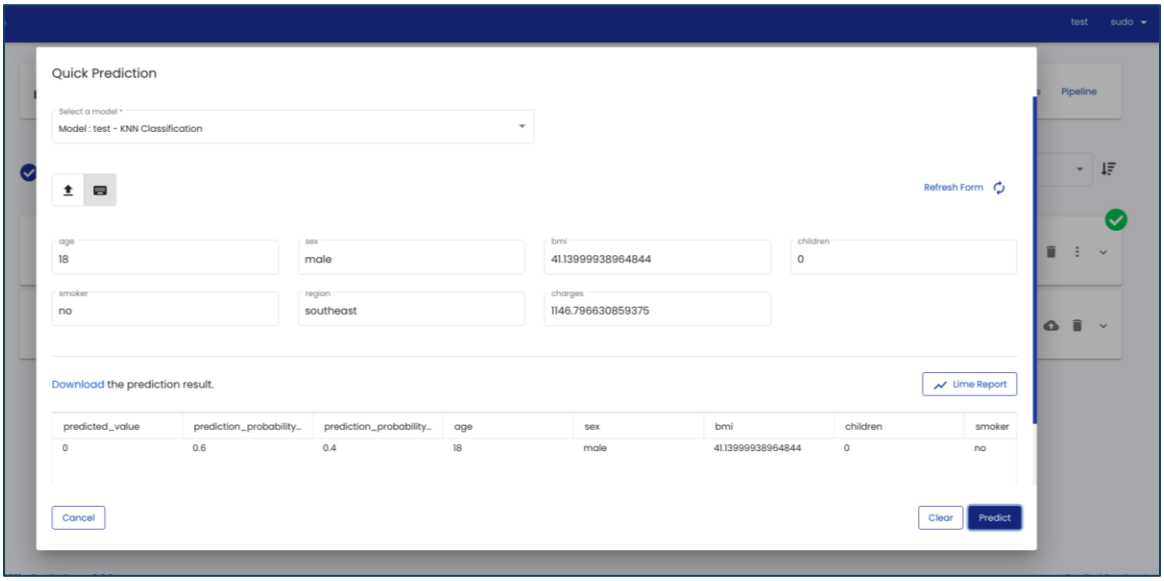
After Delete model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

9 Quick prediction

Using quick prediction option, we can quickly evaluate the prediction for any models which we have generated by providing input data in the form of a file or form input.

9.1 Single row input prediction

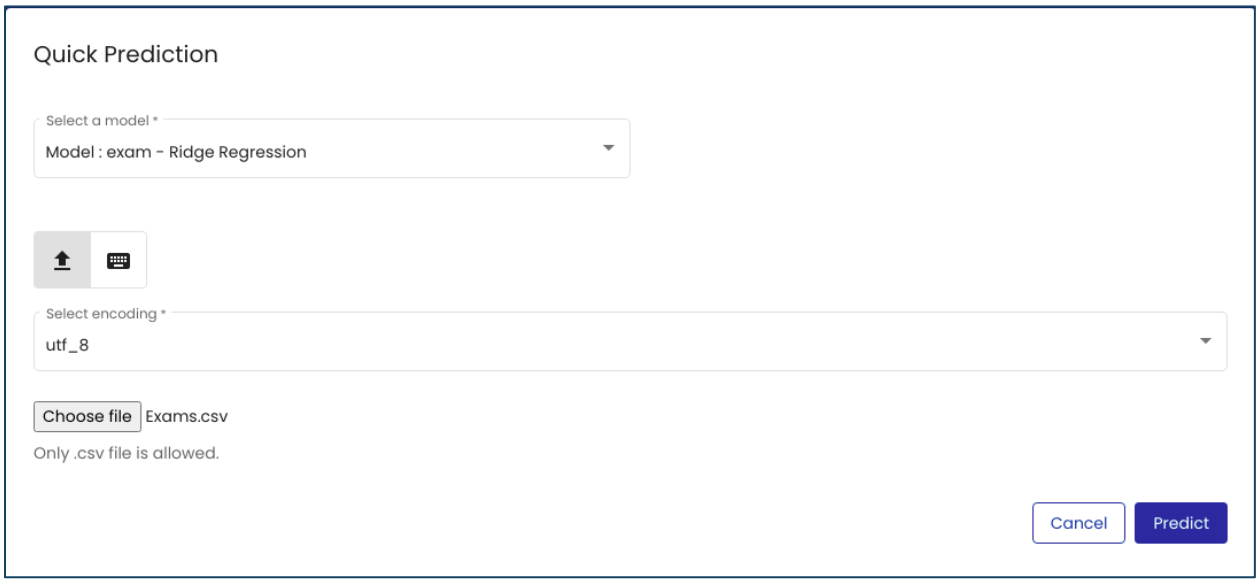


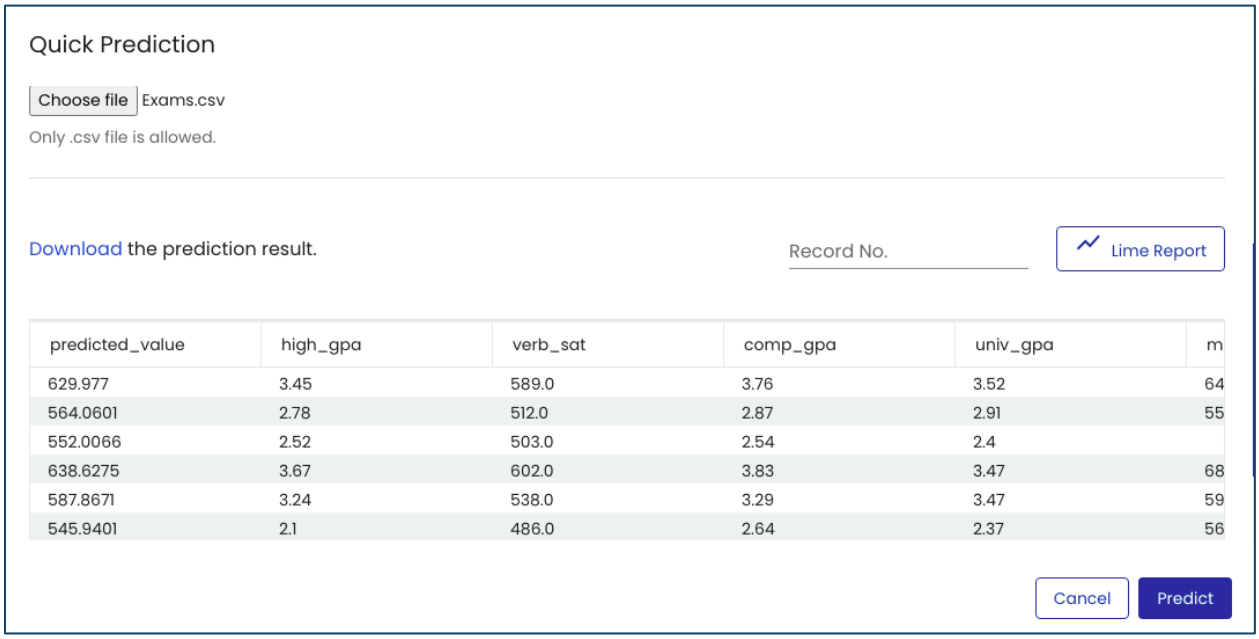
Quick prediction page

Lime Report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

9.2 File upload prediction

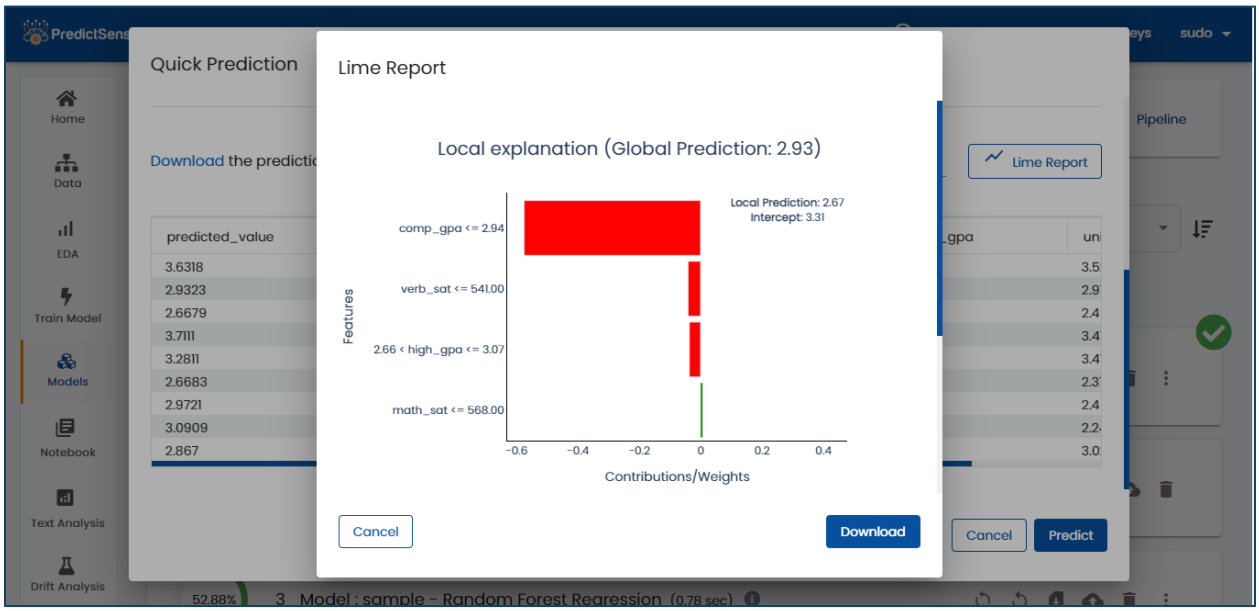




predicted_value	high_gpa	verb_sat	comp_gpa	univ_gpa	m
629.977	3.45	589.0	3.76	3.52	64
564.0601	2.78	512.0	2.87	2.91	55
552.0066	2.52	503.0	2.54	2.4	
638.6275	3.67	602.0	3.83	3.47	68
587.8671	3.24	538.0	3.29	3.47	59
545.9401	2.1	486.0	2.64	2.37	56

File upload in Quick Prediction

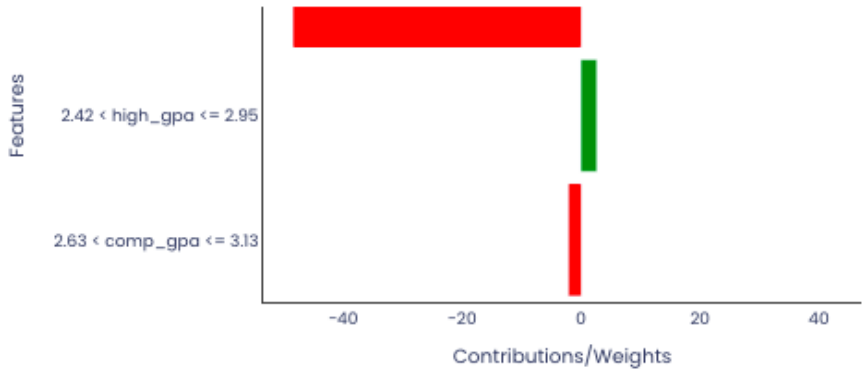
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



Lime report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

Lime Report



Data on which lime report generated.

comp_gpa	high_gpa	index	verb_sat
2.869999885559082	2.7799999713897705	1	512

FAQ's

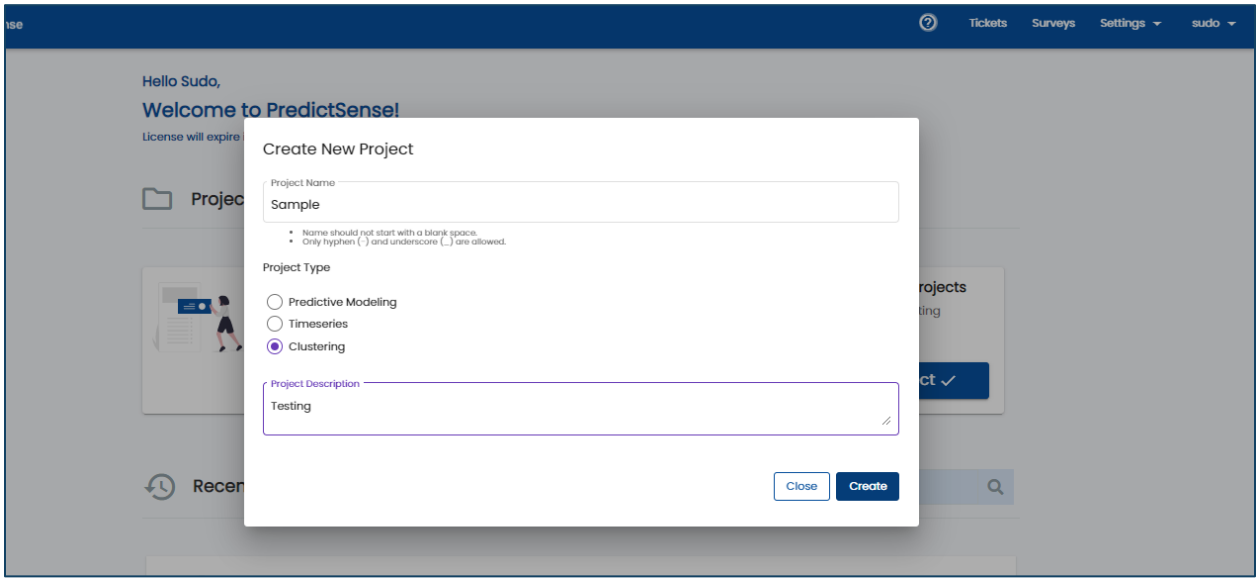
Cancel

Download

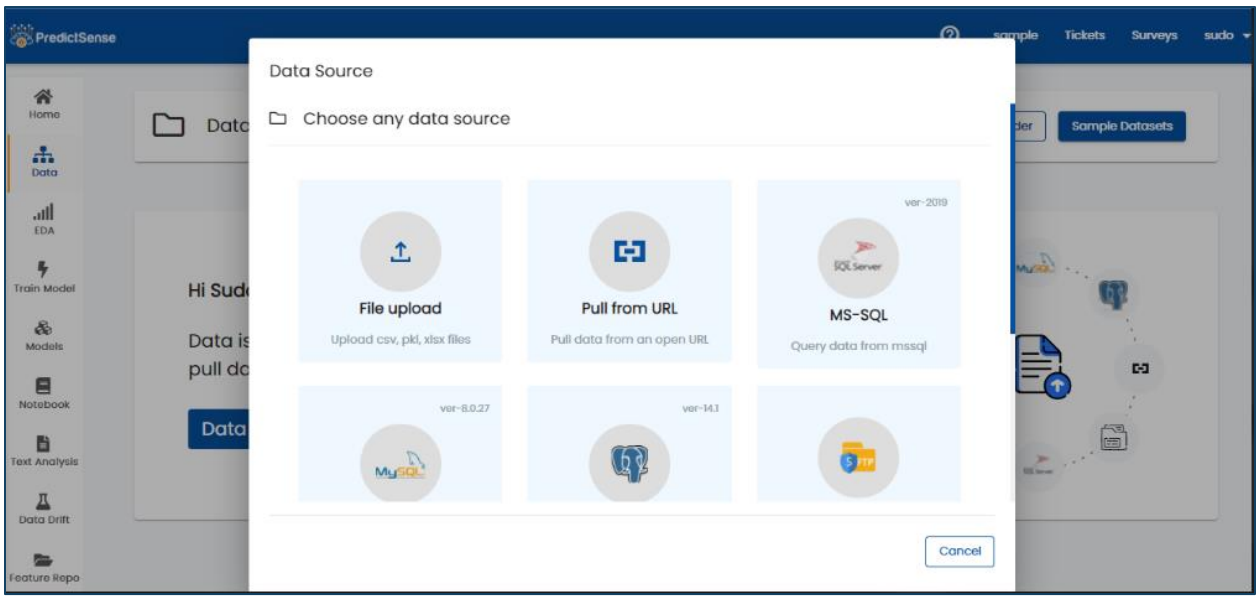
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

10 Clustering model

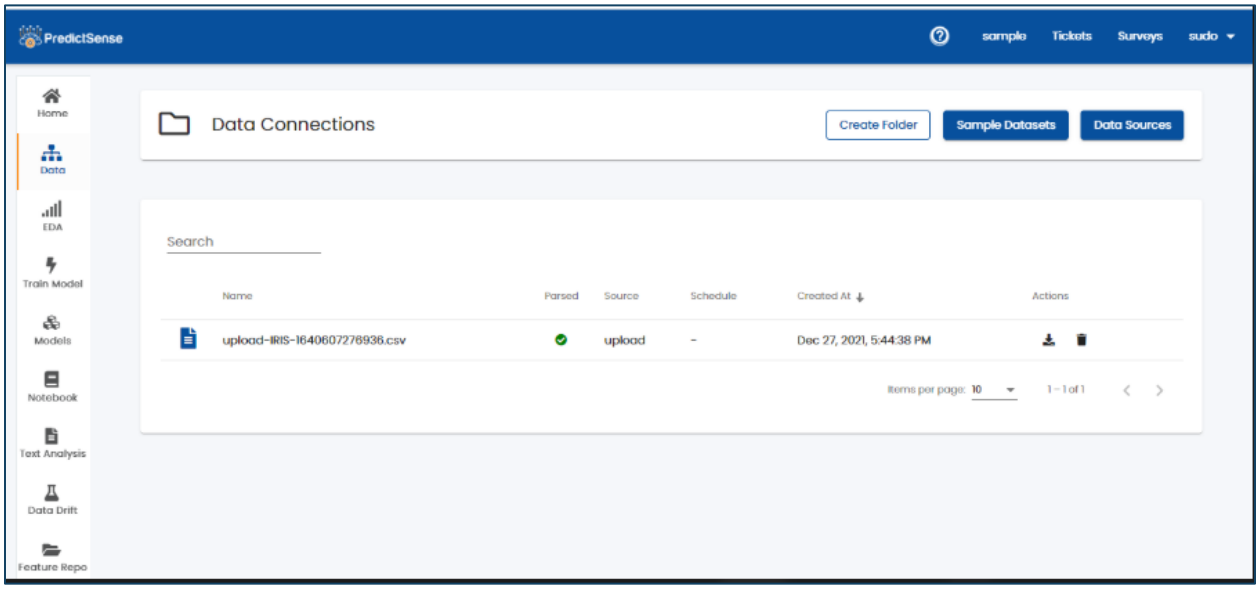
10.1 Project management page



10.2 Data Connection

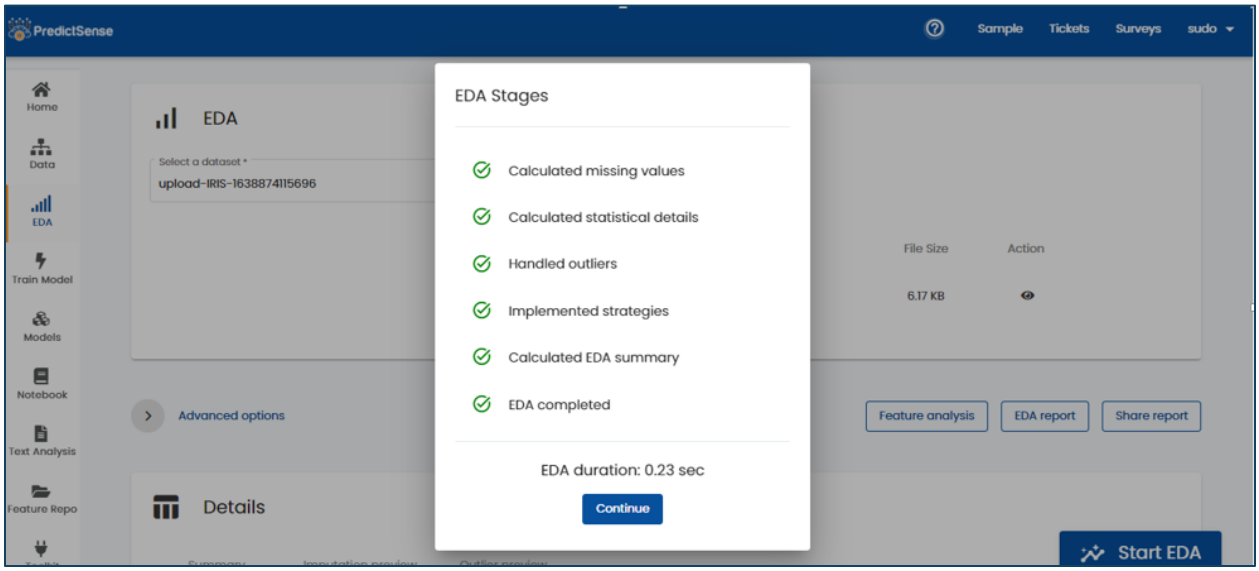


© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Data Set Page

10.3 EDA operation



EDA stage

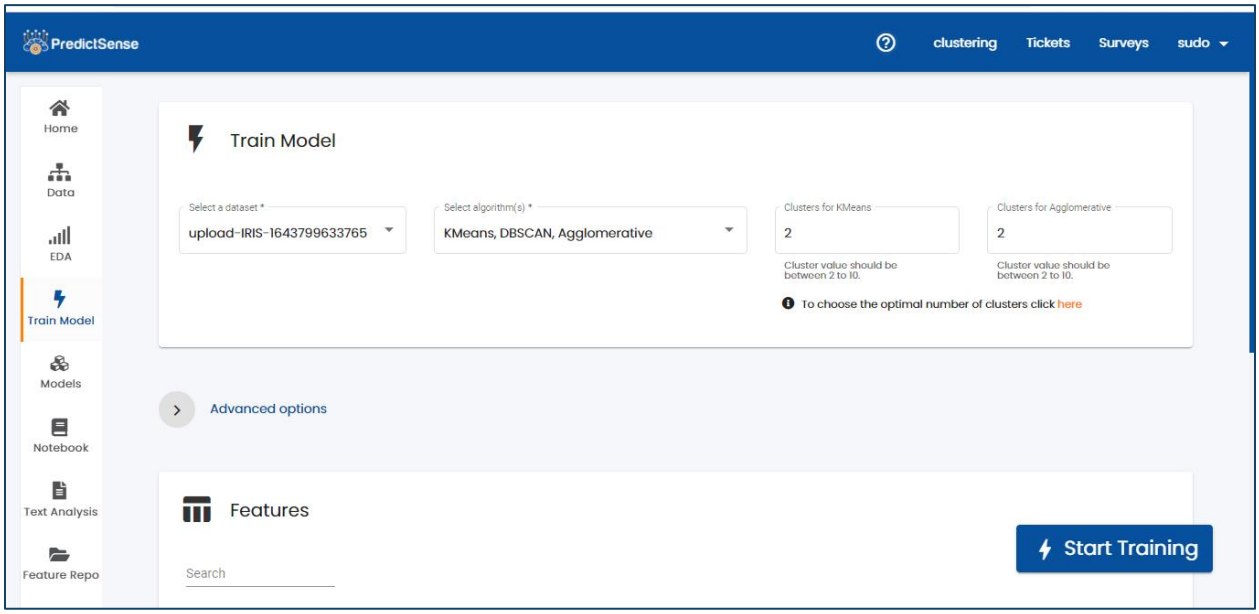
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

10.4 Train model

User can train model on clustering dataset for all **"KMeans", "DBSCAN" and "Agglomerative"**. For Training a specific model, various advanced training options like **feature transformation, feature reduction, Algorithm Parameters** can be selected to improve the training of model.

10.4.1 Training steps

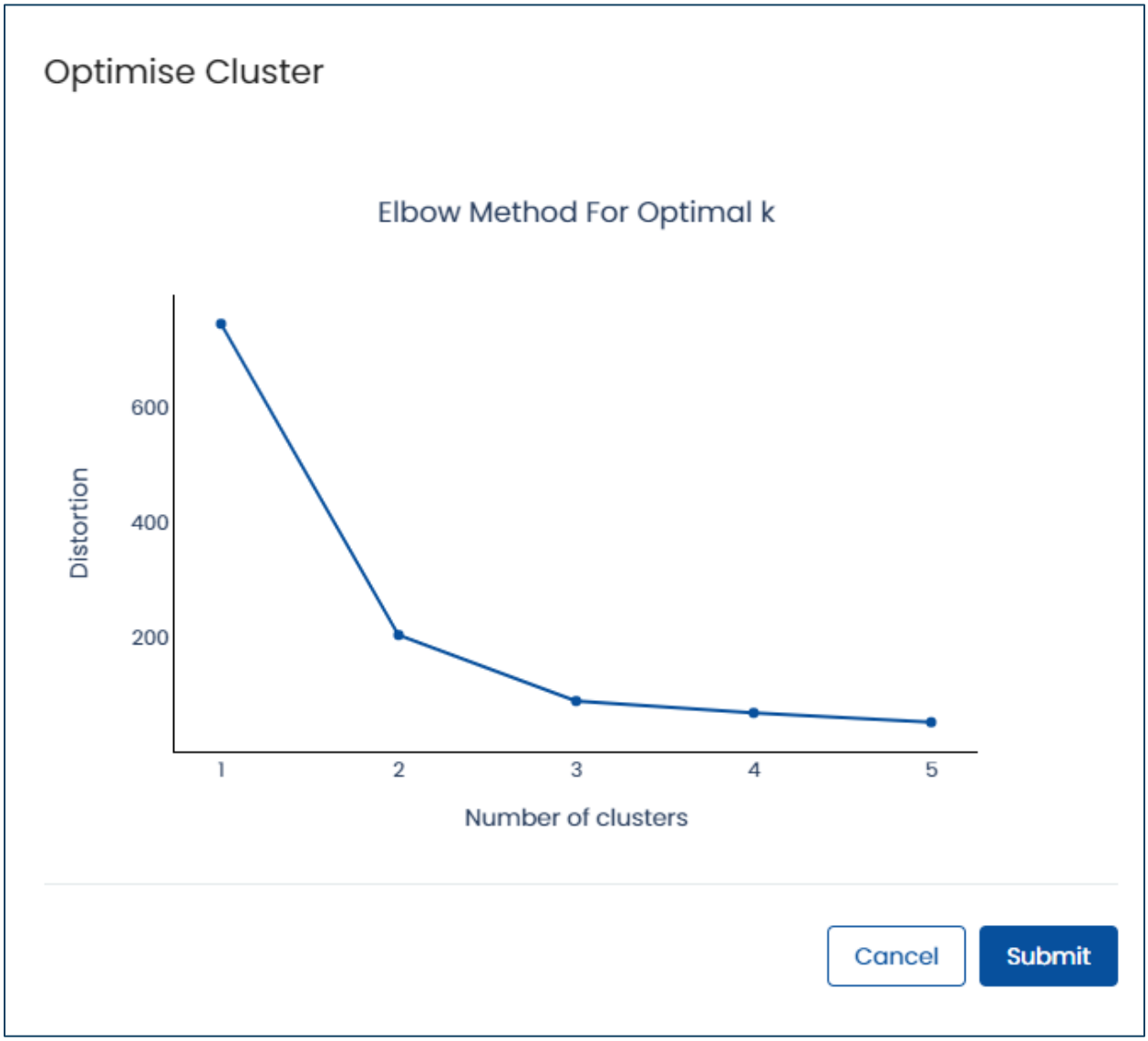
1. Select dataset
2. Select algorithms from list of algorithms
3. Add no. of clusters for KMeans and Agglomerative
4. Select multiple independent features
5. Click on start training



Train model page

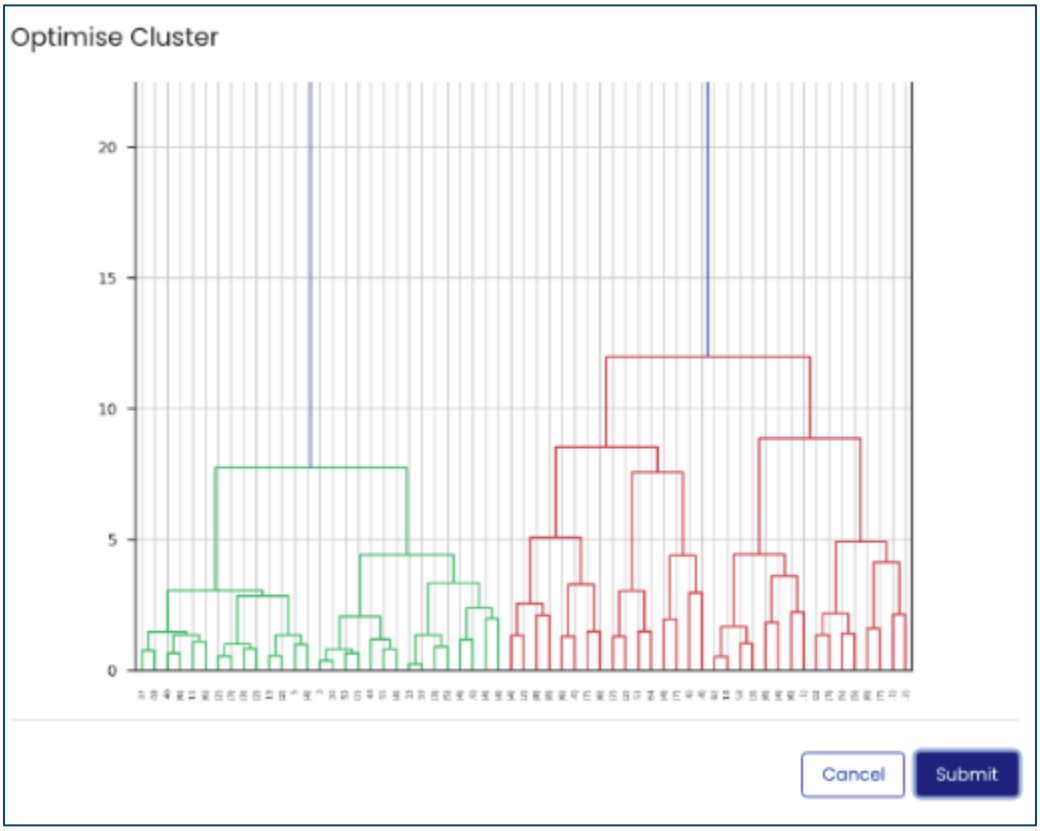
When user click on the optimal number of cluster, able to generate elbow graph for kmeans and dendrogram for agglomerative.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Elbow graph

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

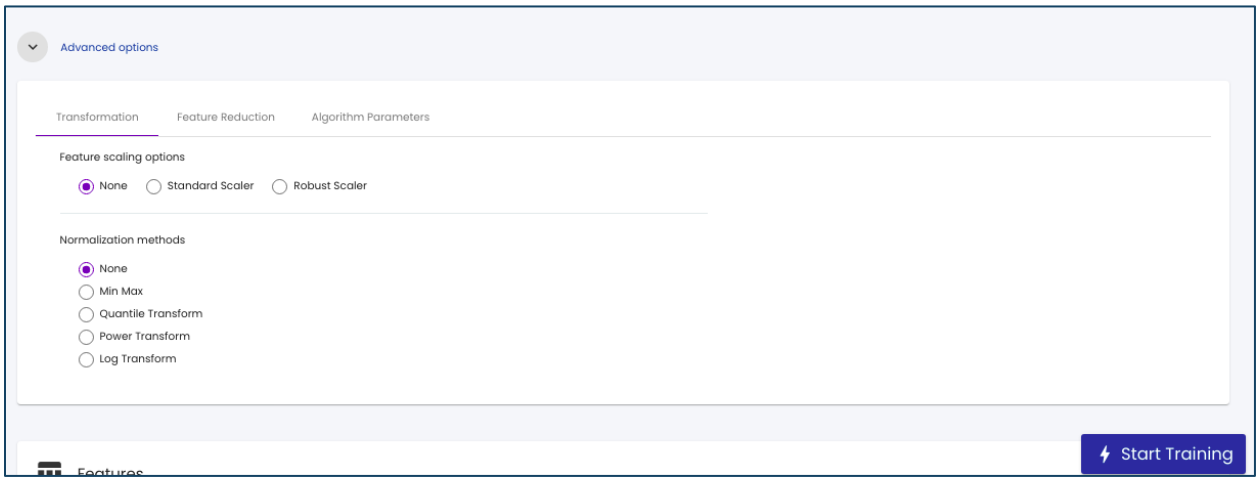


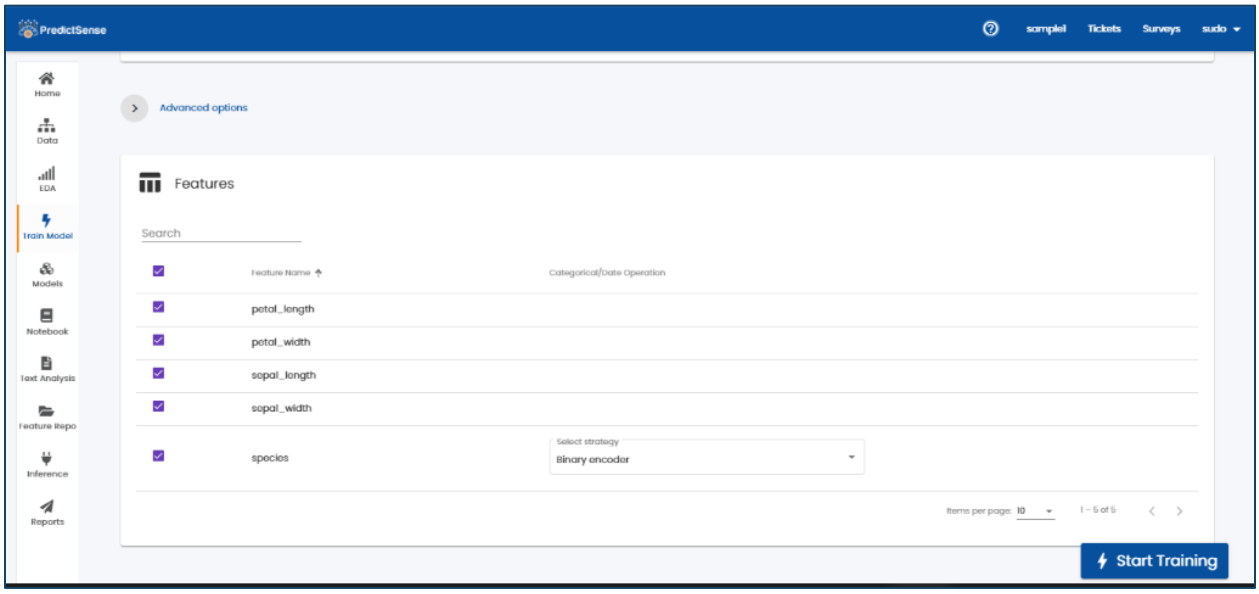
Dendrogram

10.4.2 Training steps with advanced options

1. Select dataset
2. Select algorithms from list of algorithms
3. Add no. of clusters for KMeans and agglomerative
4. Select various options from advanced options by switching between tabs like **"transformation", "feature reduction", "Algorithm Parameters" etc** different methods for **scaling, normalization, dimensionality reduction.**
5. Select multiple independent features
6. Click on start training

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

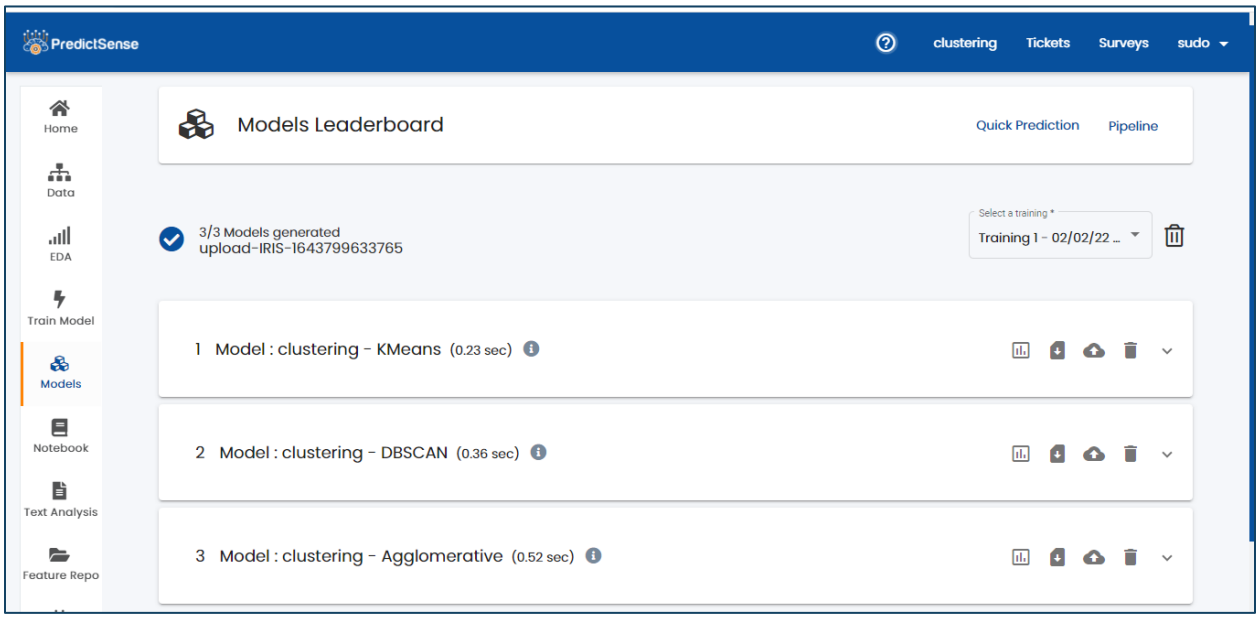




Feature selection and Start training

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

10.5 Models leaderboard

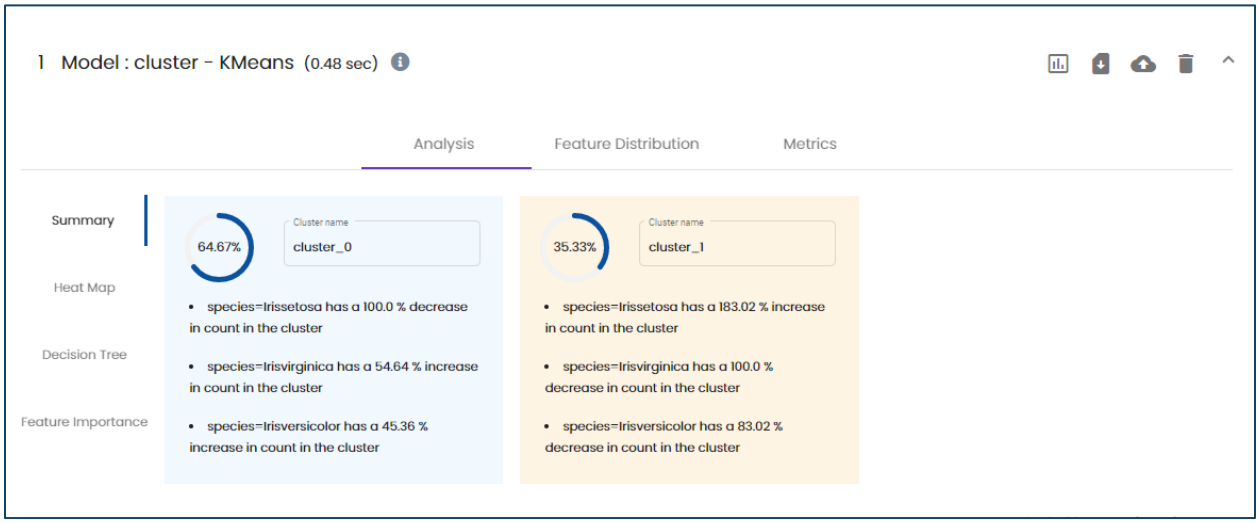


Model leaderboard Page

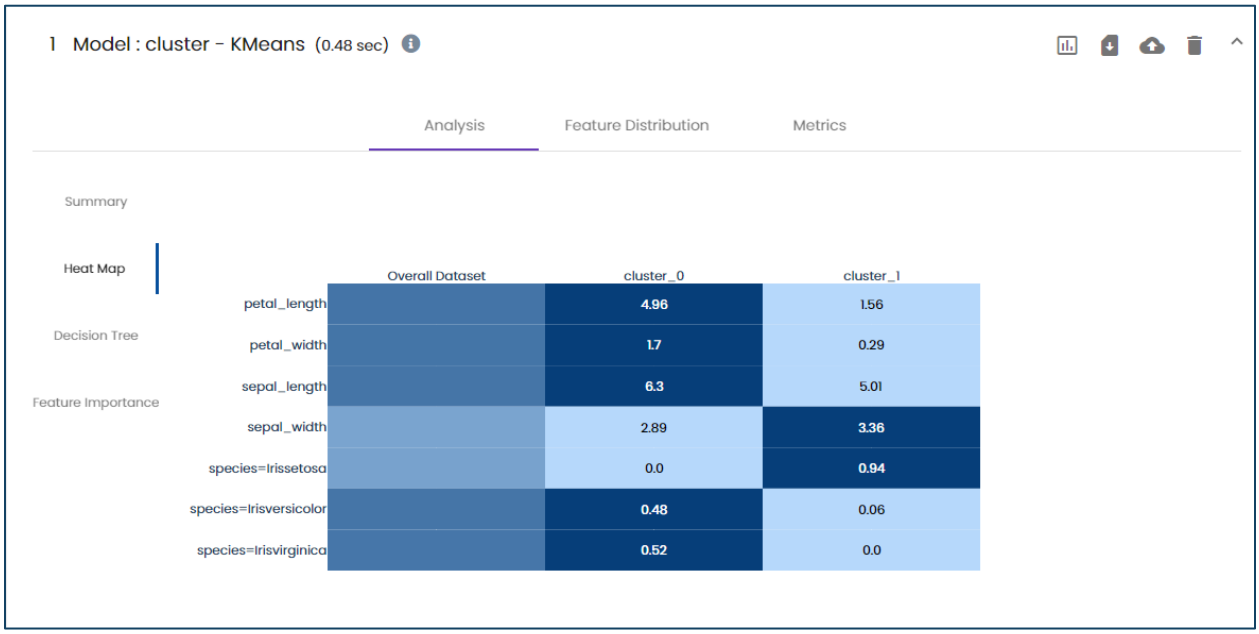
10.5.1 Analysis

The Cluster analysis helps to better understand the clustered output dataset on how the samples are segmented and describes statistical interpretations of the features for each cluster to then allow the user to give a defined label to each cluster. There are four tabs Summary, heat map, decision tree, feature importance. In summary all cluster are been displayed and user can also edit cluster name which will be reflect in heatmap as well. Heatmaps visualize a data matrix by drawing a rectangular grid corresponding to rows and columns in the matrix, and coloring the cells by their values.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Cluster analysis

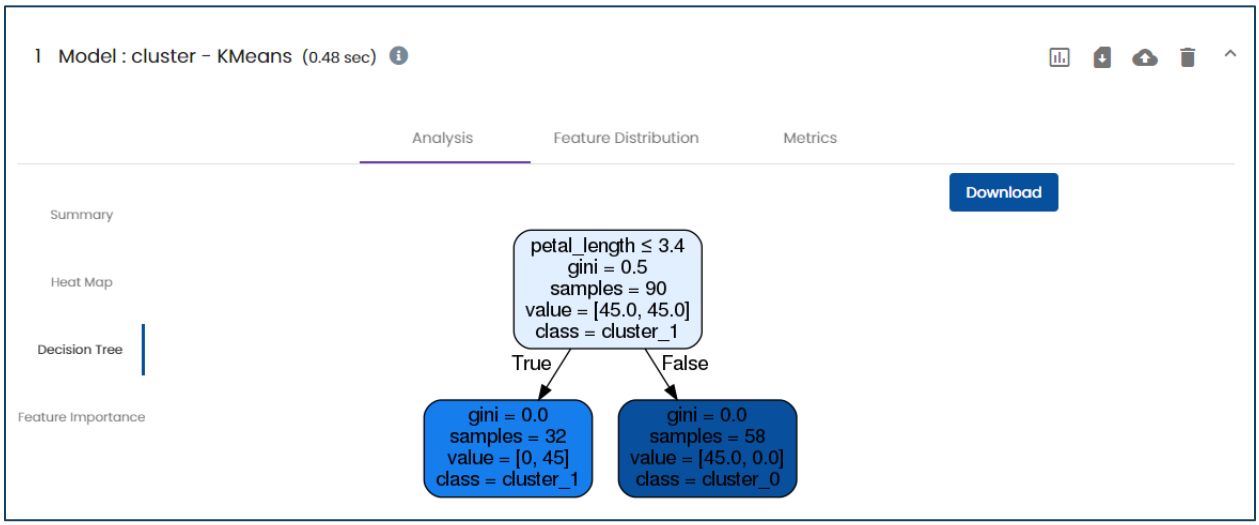


Heat map

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

Decision Tree:

The decision tree graph is a binary tree to help visualize how the data is getting segmented to its respective cluster showing the top features on which it is getting clustered by training a decision tree model using the clustered output dataset.

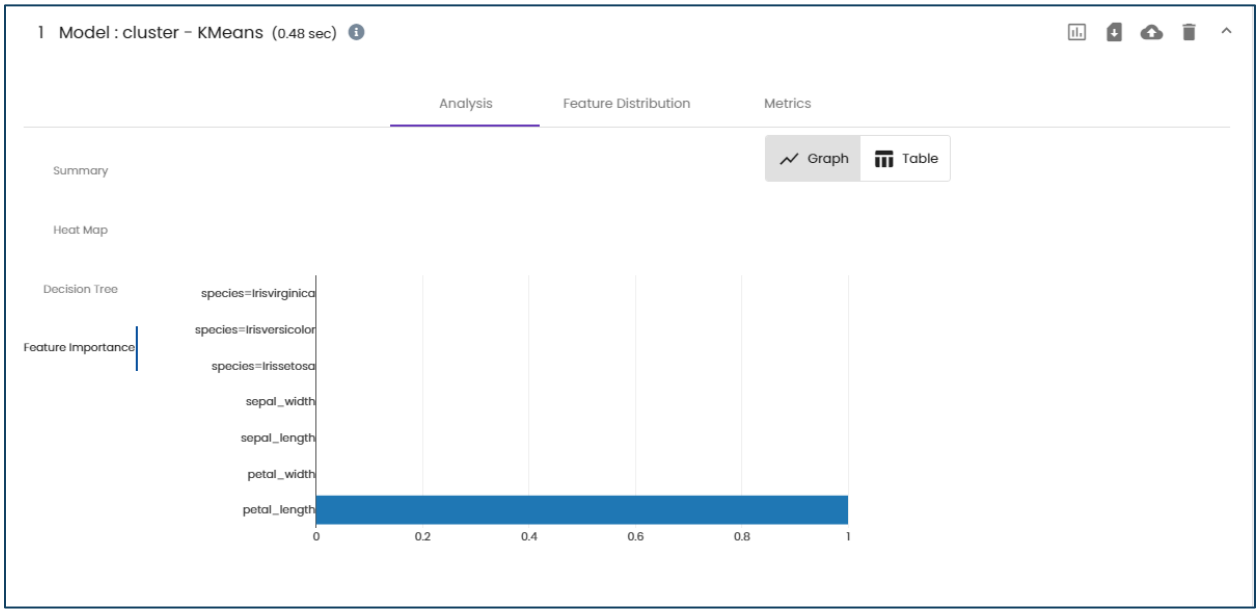


Decision tree plot

Feature Importance:

The feature importance plot displays the score for each feature in relatively contributing for the prediction of the cluster ranging from 0-1 and adding up to 1 for all features which is generated from the trained decision tree model on the clustered output dataset.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

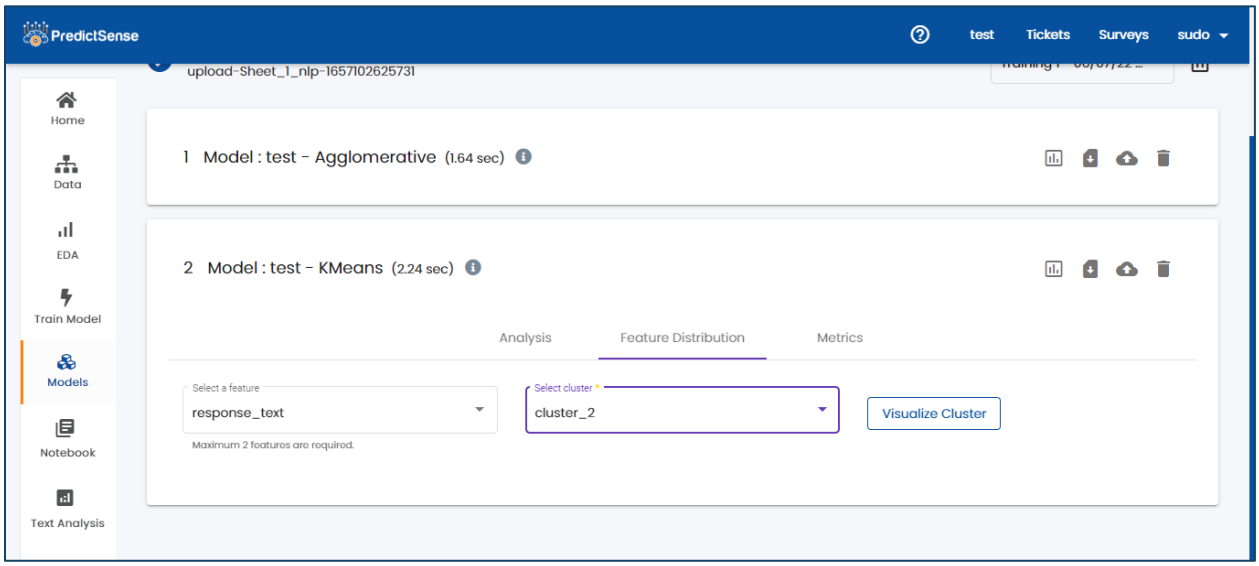


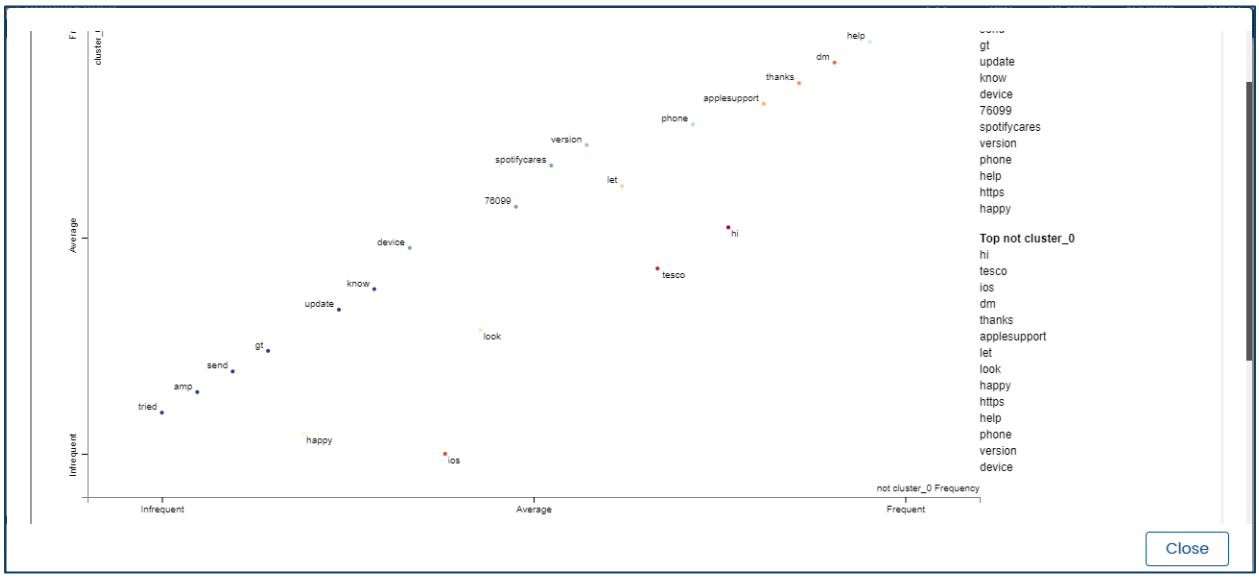
Feature Importance

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

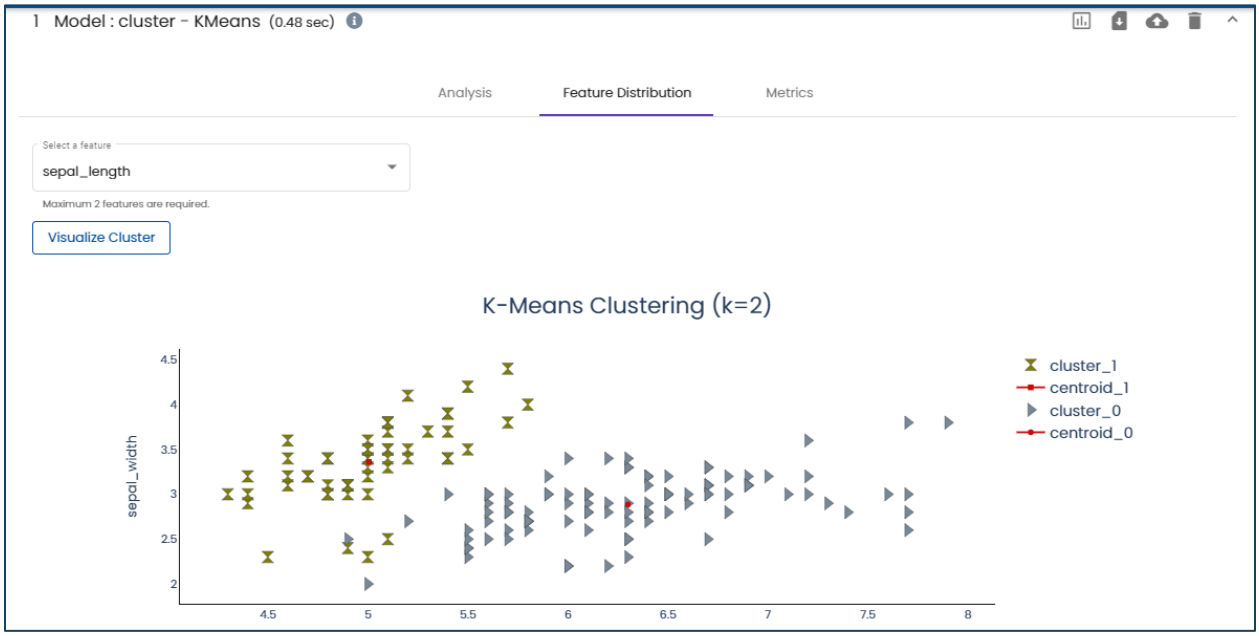
10.5.2 Feature Distribution

User will able to generate scatter plot for text features to visualize words or phrases that are characteristic to two categories. It helps to inspect words or phrases based on their frequencies in each category.





Scatter text plot



Feature Distribution

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

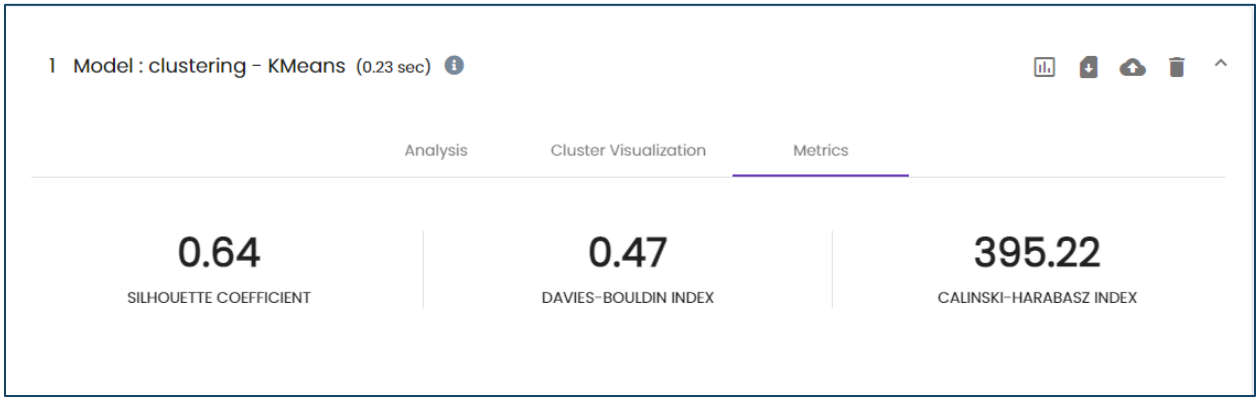
10.5.3 Metrics:-

This are the three metrics added in clustering module:-

Silhouette coefficient

Davies-bouldin index

Calinski-Harabasz index

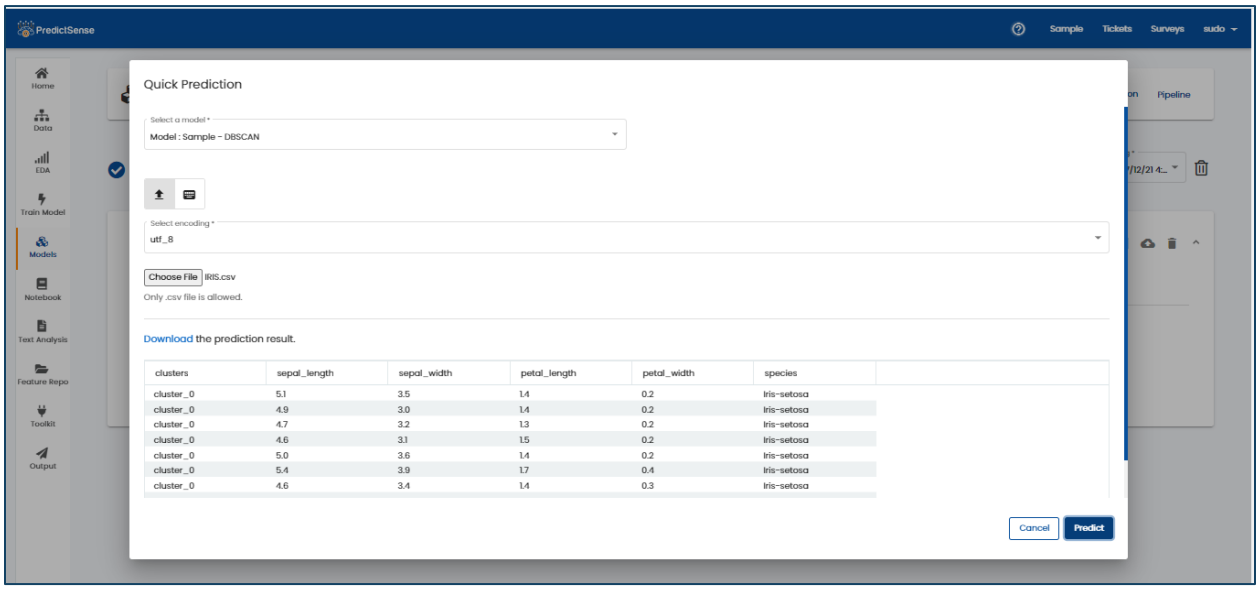


Metrics

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

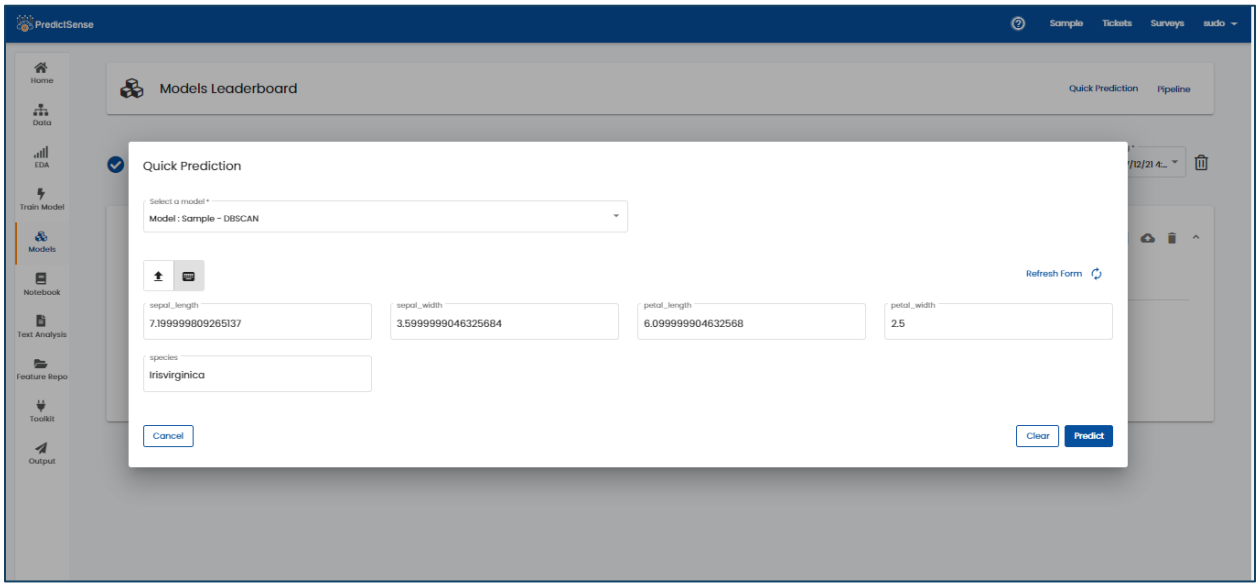
10.6 Quick prediction

10.6.1 File prediction



File Prediction

10.6.2 Single row input prediction

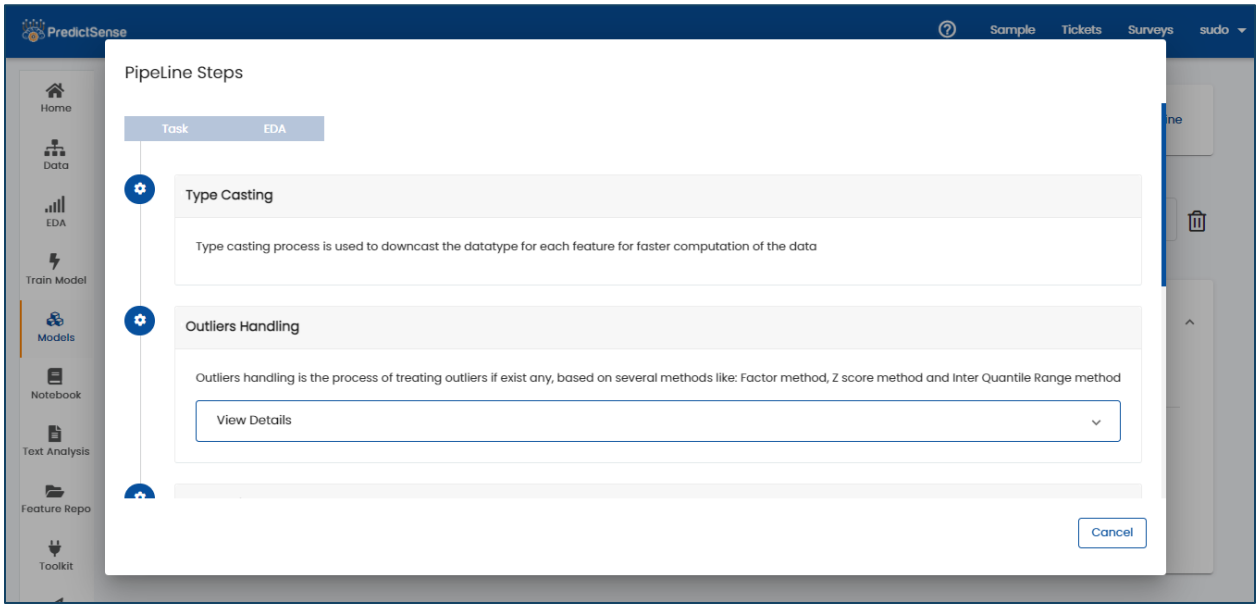


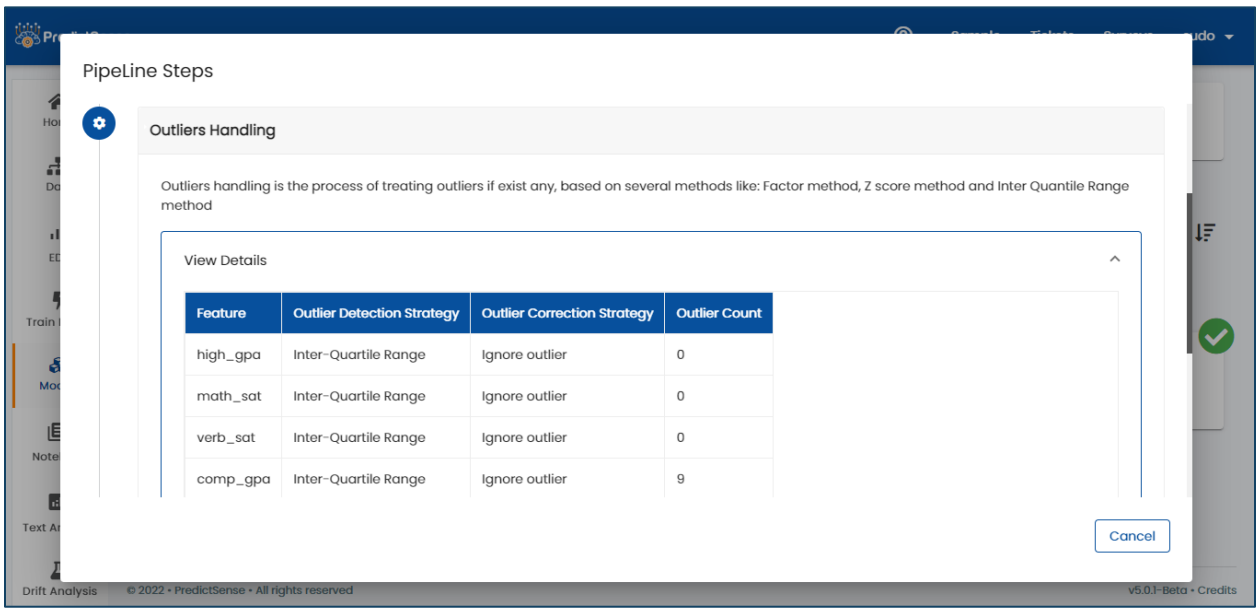
Single row input prediction

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

10.7 Pipeline

A machine learning pipeline is used to help automate machine learning workflows. They operate by enabling a sequence of data to be transformed and correlated together in a model that can be tested and evaluated to achieve an outcome.

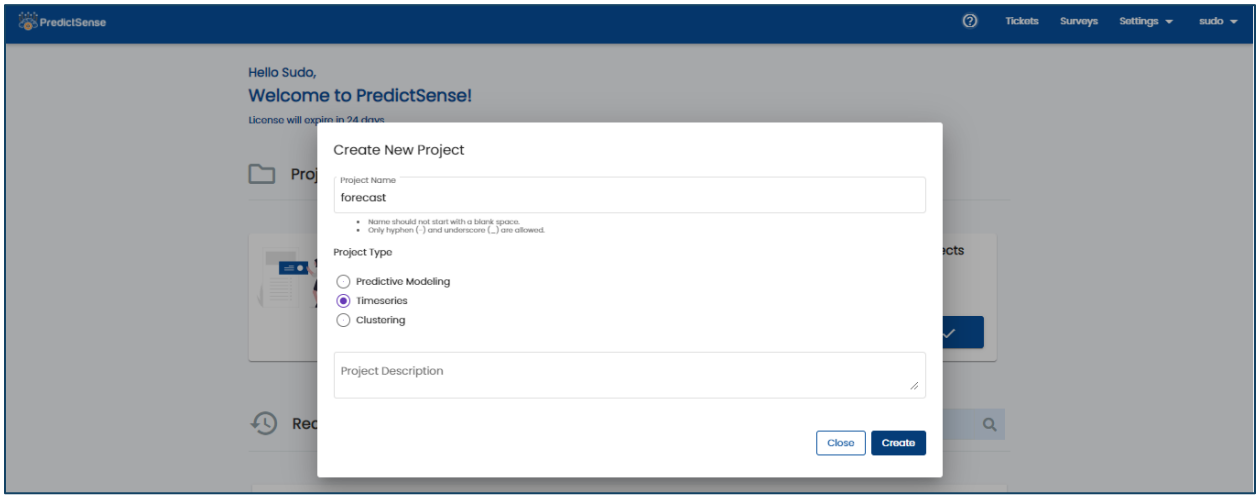




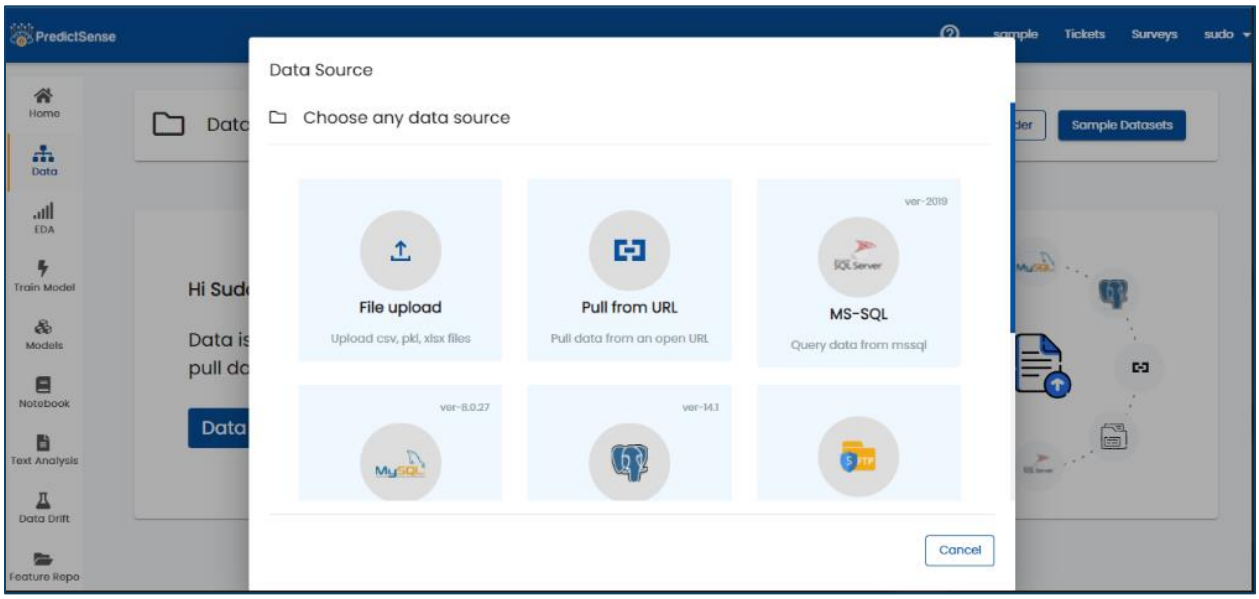
Pipeline Steps

11 Timeseries model

11.1 Project management page

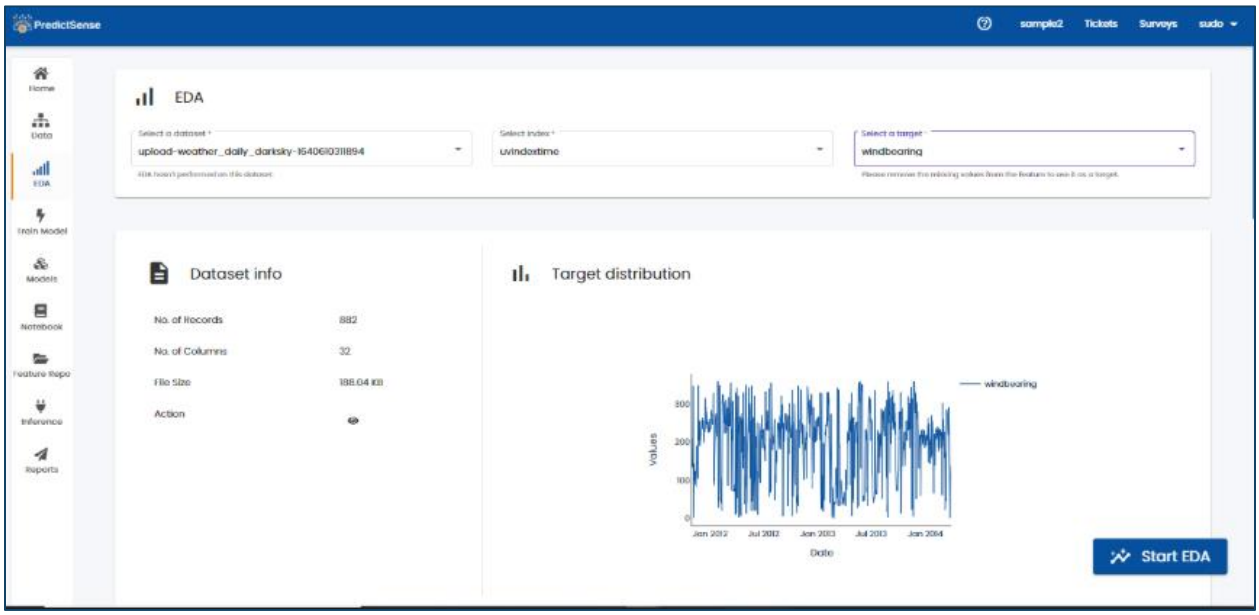


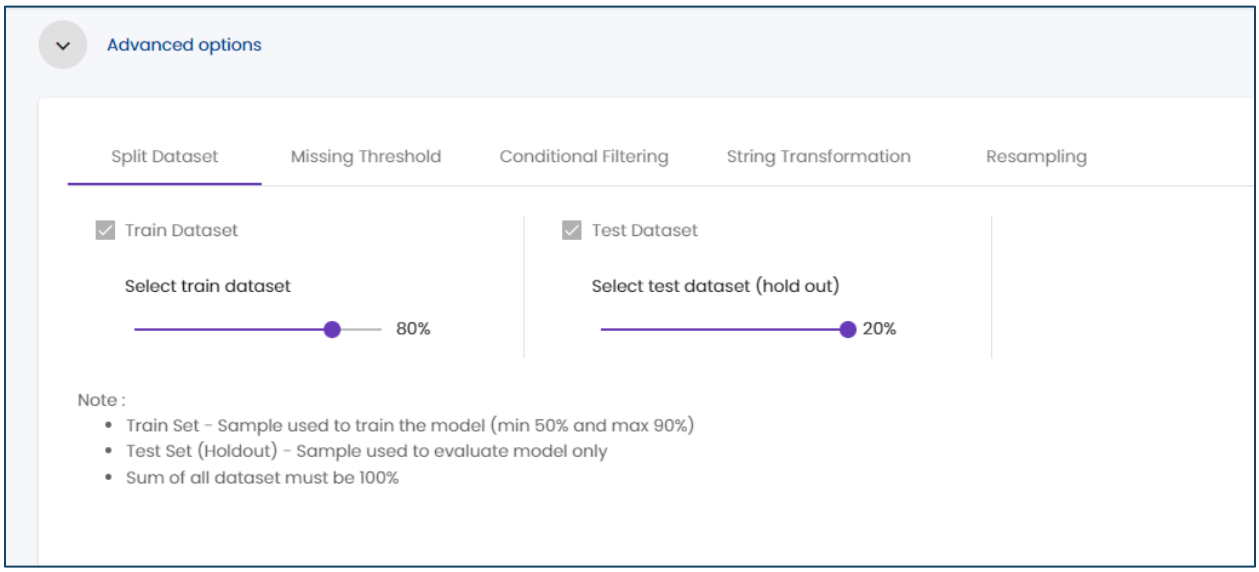
11.2 Data connection



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

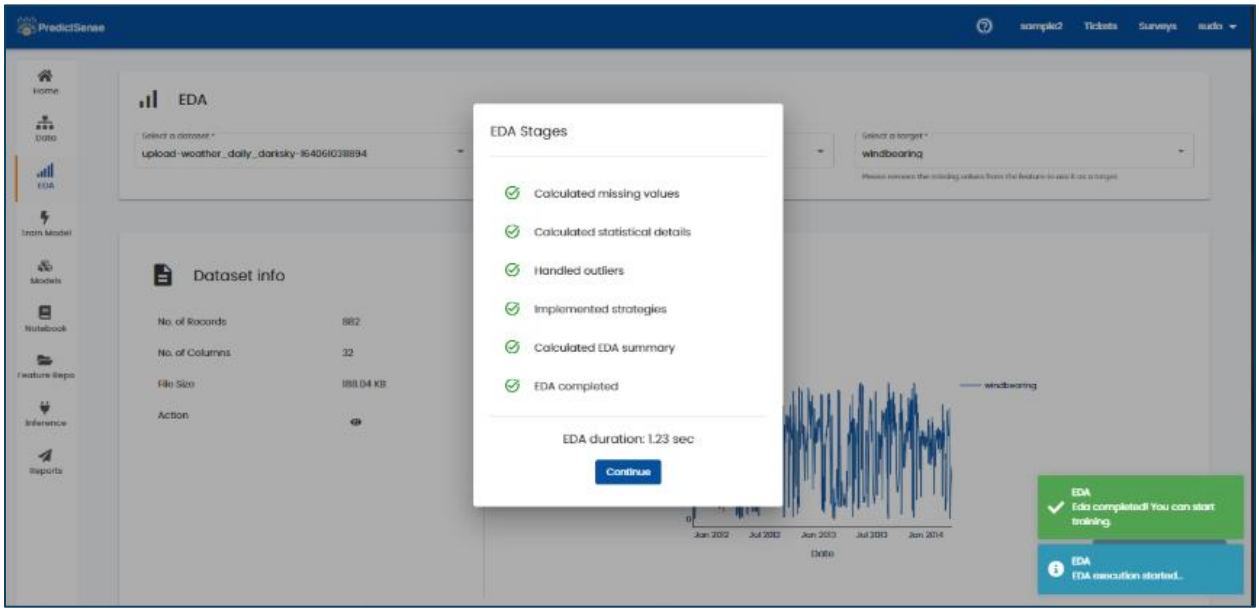
11.3 EDA operation for timeseries





Advanced options

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



EDA Stage

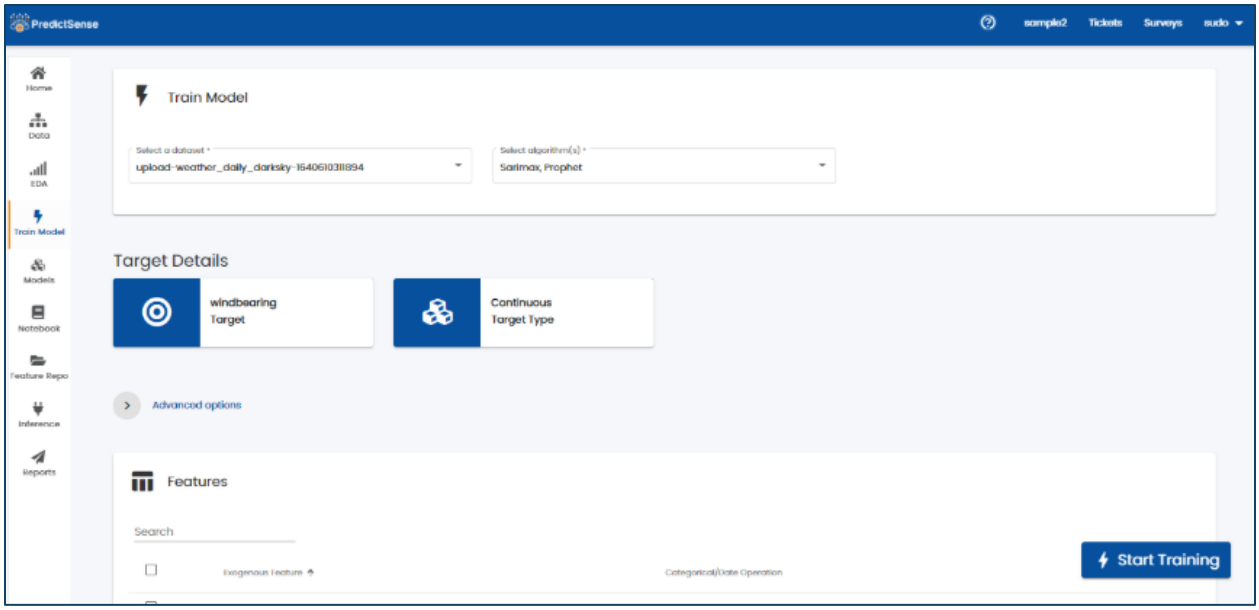
11.4 Train model

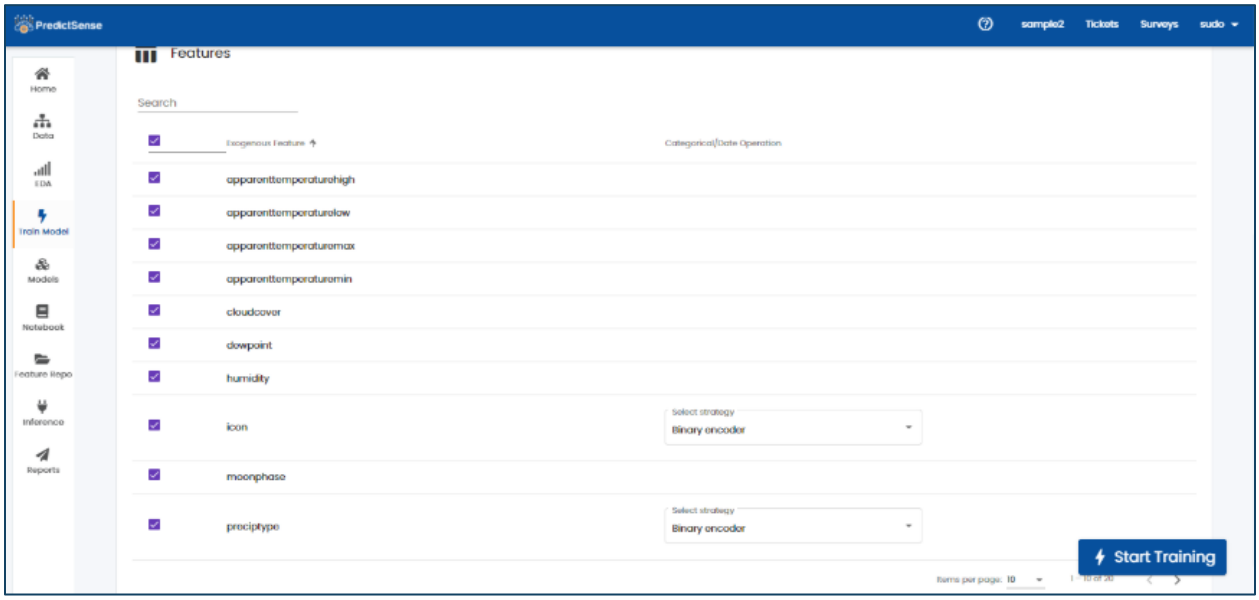
User can train model for both **"Sarimax"** and **"Prophet"**. For training a specific model, various advanced training options like **Components, Stationarity Test, feature transformation, Auto Correlation, Algorithm Parameters** can be selected to improve the training of model. For multiple timeseries user can train model for **"Random forest"** and "**Linear Regression Ts".**

11.4.1 Training steps

1. Select dataset
2. Select algorithms from list of algorithms
3. Select multiple independent features
4. Click on start training

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



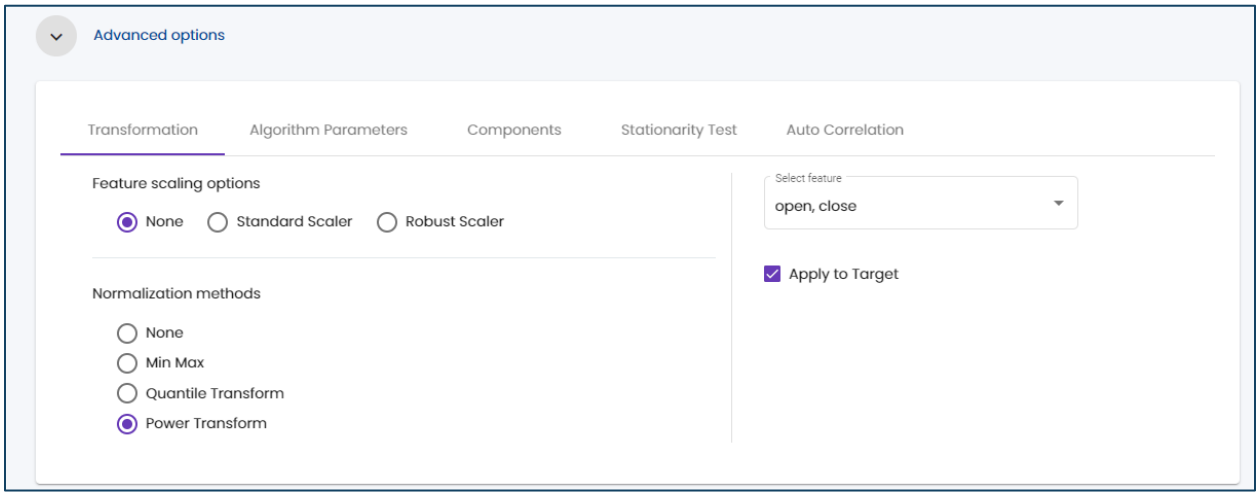


Train model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

11.4.2 Training steps with advanced options

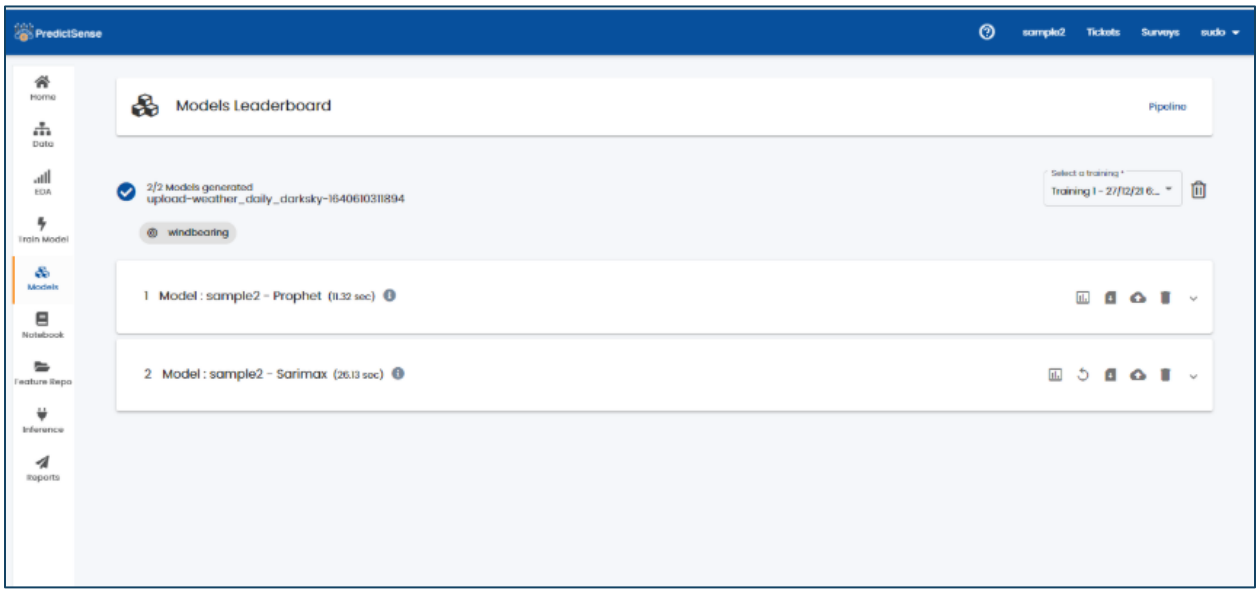
1. Select dataset
2. Select target feature
3. Select algorithms from list of algorithms
4. Select various options from advanced options by switching between tabs like **"Components", "Stationarity Test", "Transformation", "Auto Correlation"** and **"Algorithm Parameters" etc.**
5. Select multiple independent features
6. Click on start training

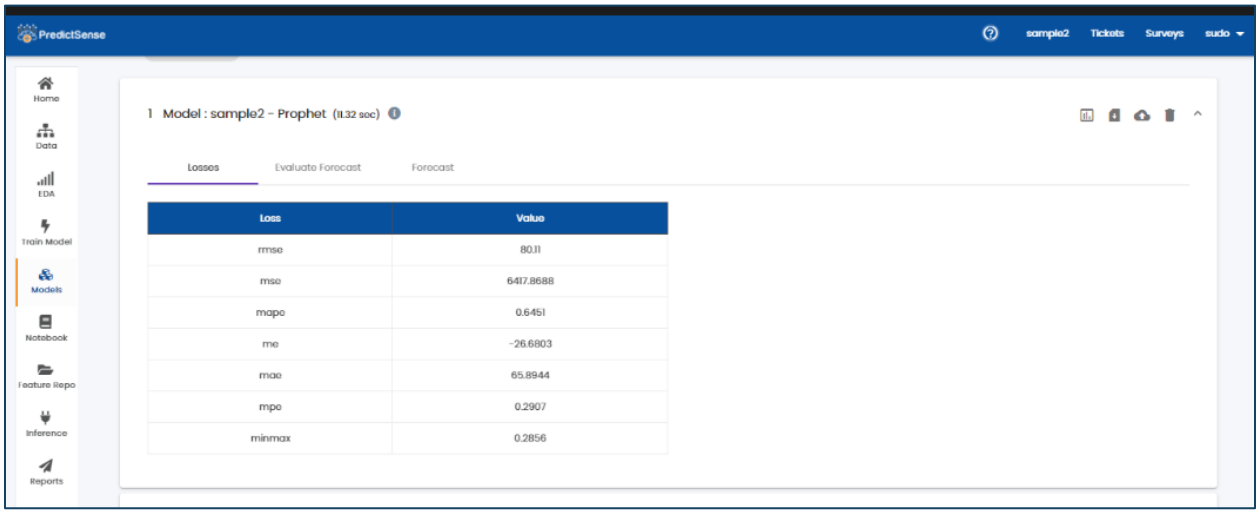


Advanced options

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

11.5 Models leaderboard

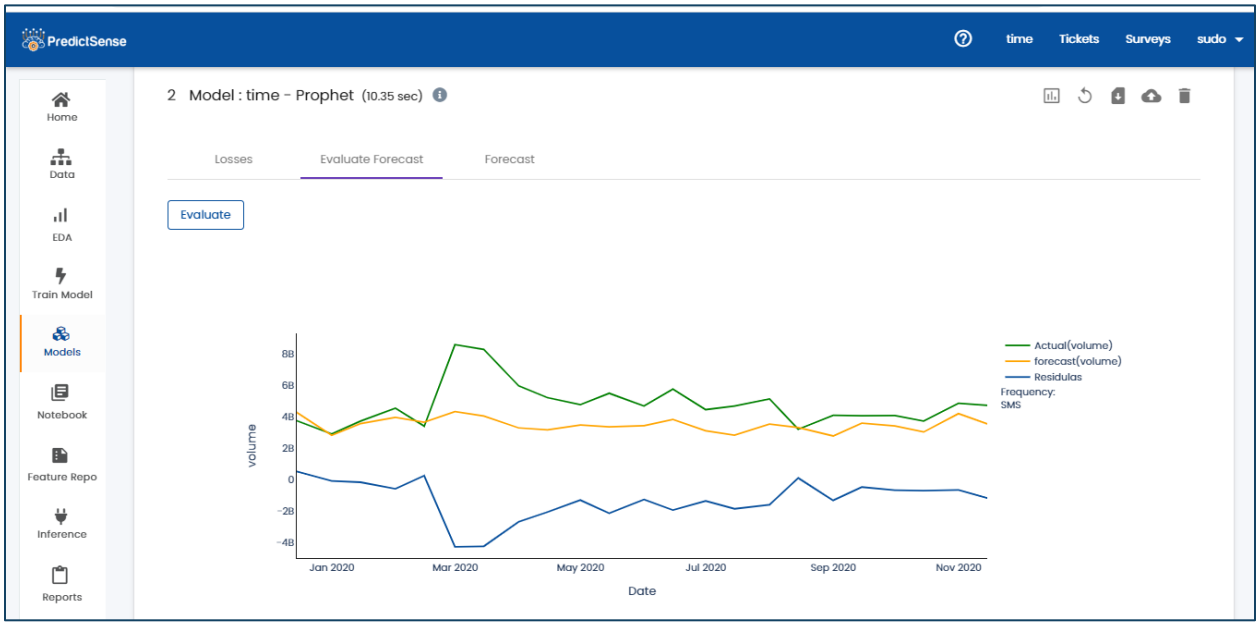




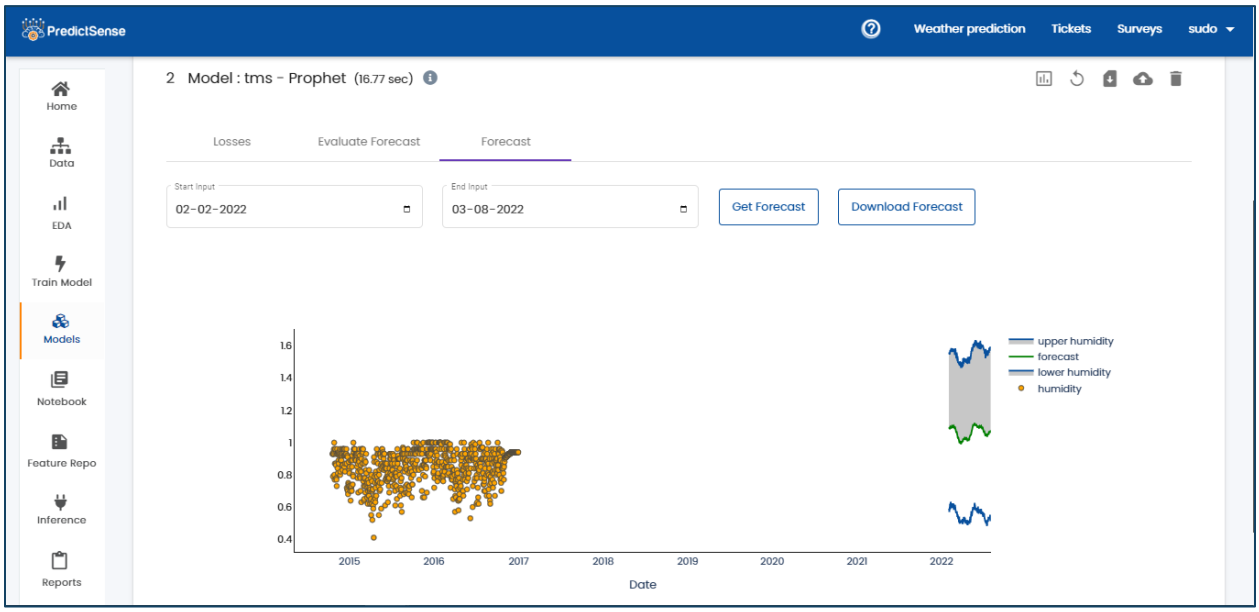
Loss	Value
rmse	80.11
mse	6417.8688
mape	0.6451
me	-26.6803
mae	65.8944
mpe	0.2907
minmax	0.2856

Details model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

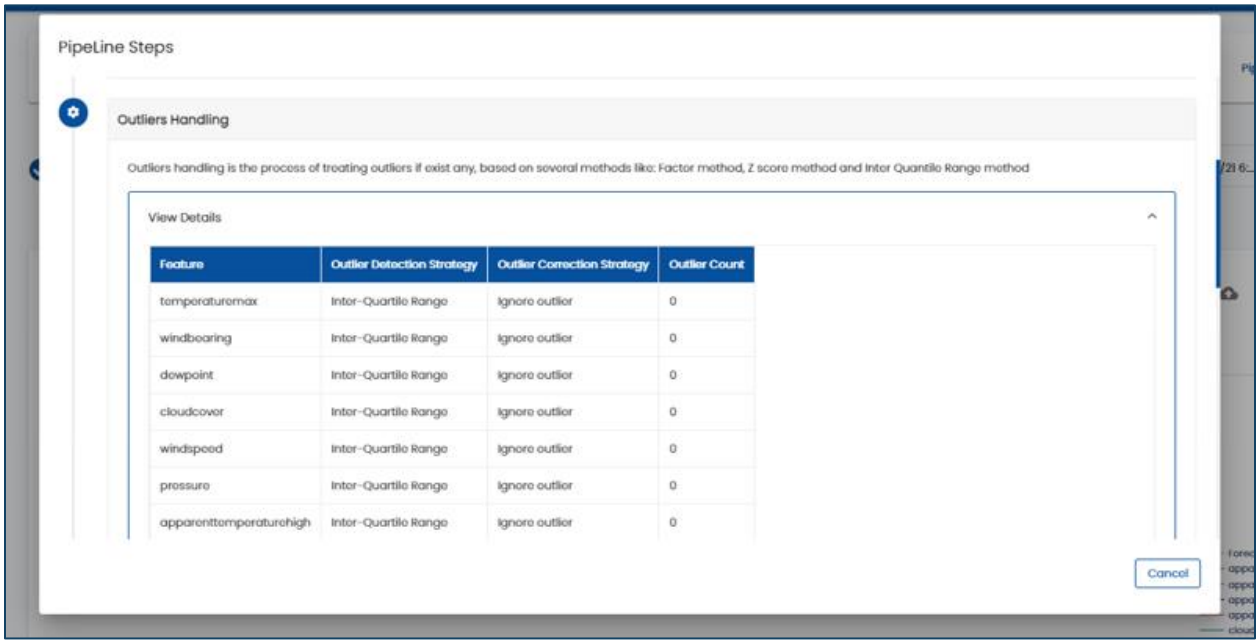


Evaluate



Forecast

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



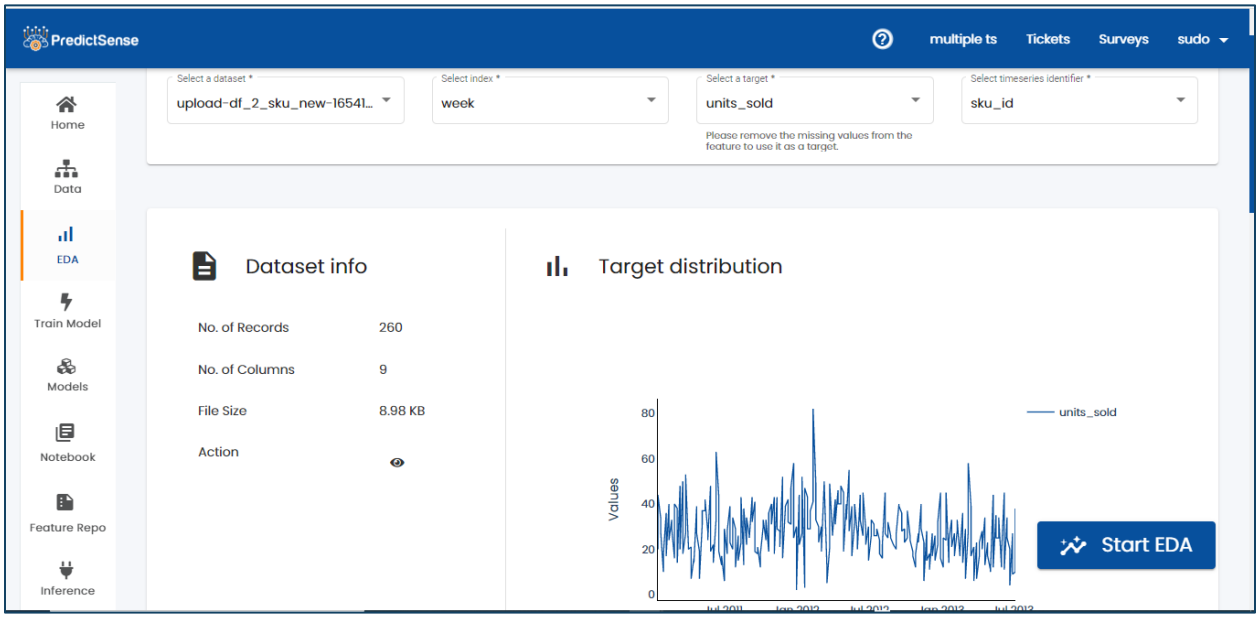
Feature	Outlier Detection Strategy	Outlier Correction Strategy	Outlier Count
temperaturemax	Inter-Quartile Range	Ignore outlier	0
windbearing	Inter-Quartile Range	Ignore outlier	0
dewpoint	Inter-Quartile Range	Ignore outlier	0
cloudcover	Inter-Quartile Range	Ignore outlier	0
windspeed	Inter-Quartile Range	Ignore outlier	0
pressure	Inter-Quartile Range	Ignore outlier	0
apparenttemperaturehigh	Inter-Quartile Range	Ignore outlier	0

Pipeline

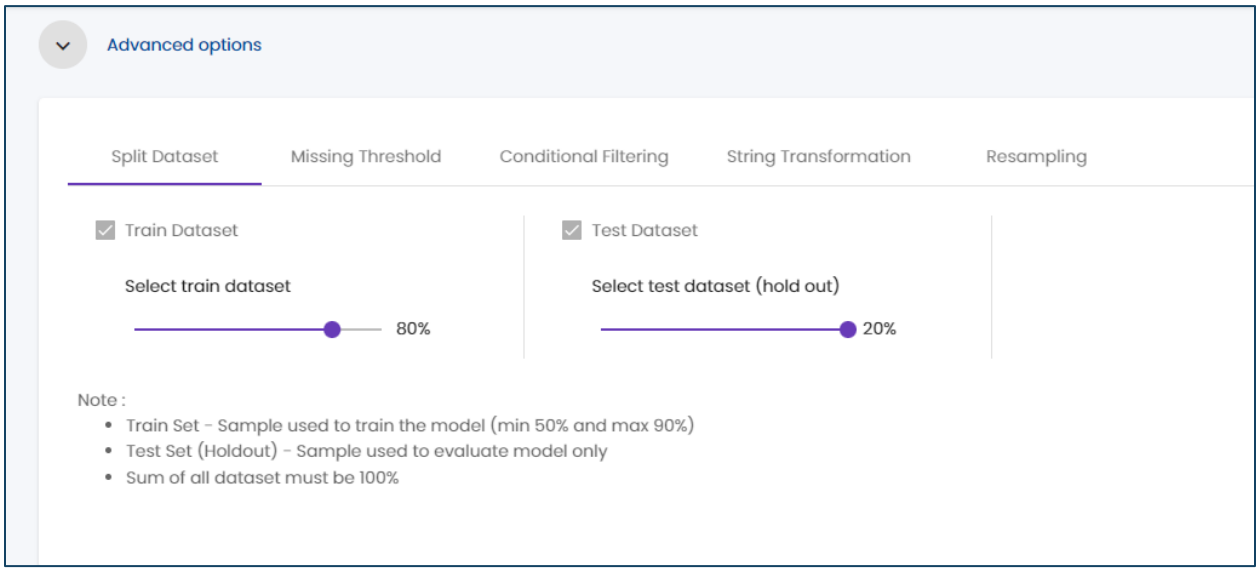
11.6 EDA operation for multiple timeseries:-

User will also able to train Multiple Timeseries model. Multiple Timeseries is basically a timeseries having multiple individual timeseries. Traditionally, if we wanted to forecast for this data, we had to train multiple models for individual timeseries data. But using this functionality, we can do it using single model. This can be used where user have multiple data in one and wants to forecast for all.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

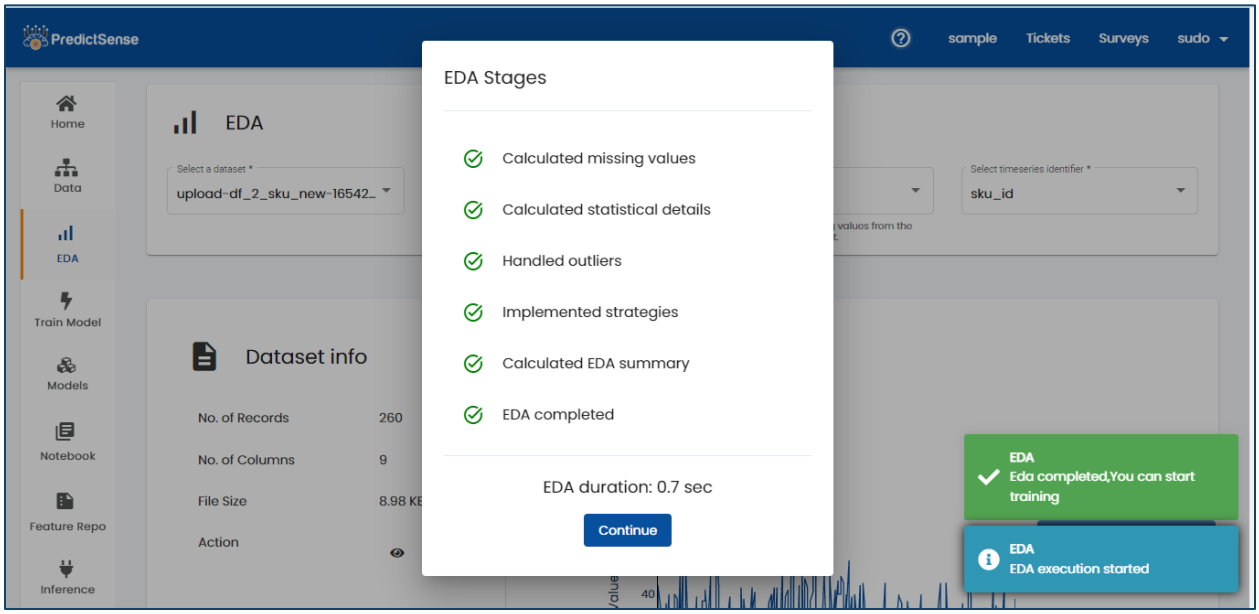


EDA



Advanced options

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



EDA Stage

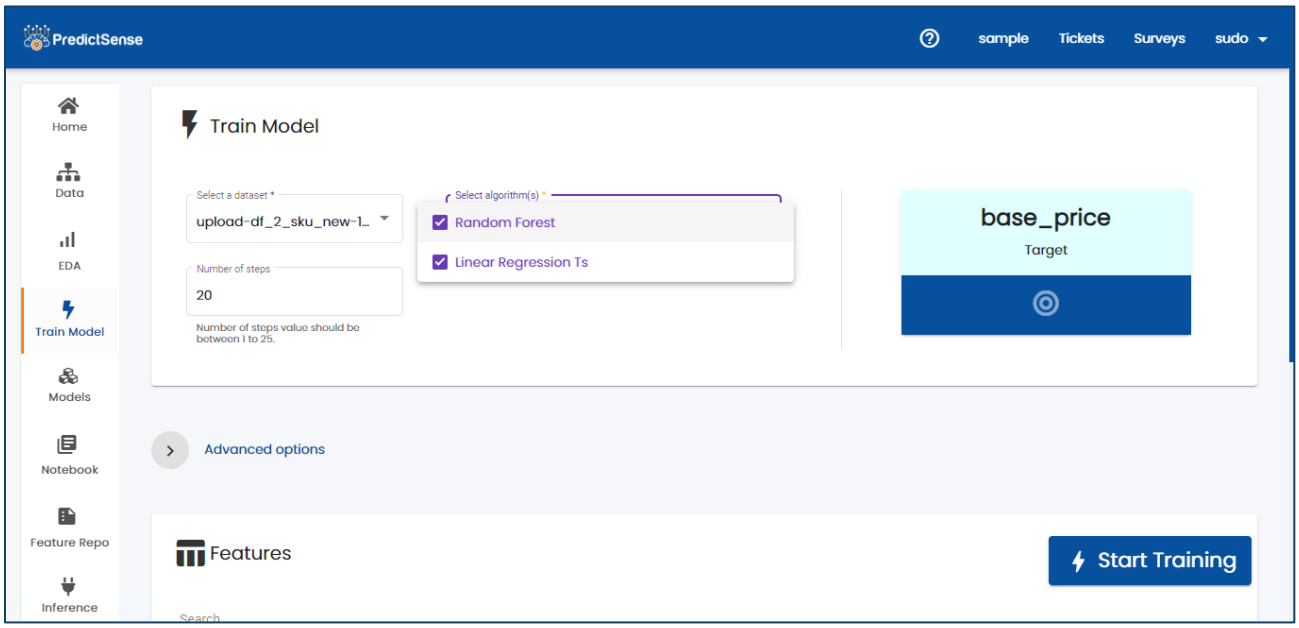
11.7 Train model for multiple timeseries

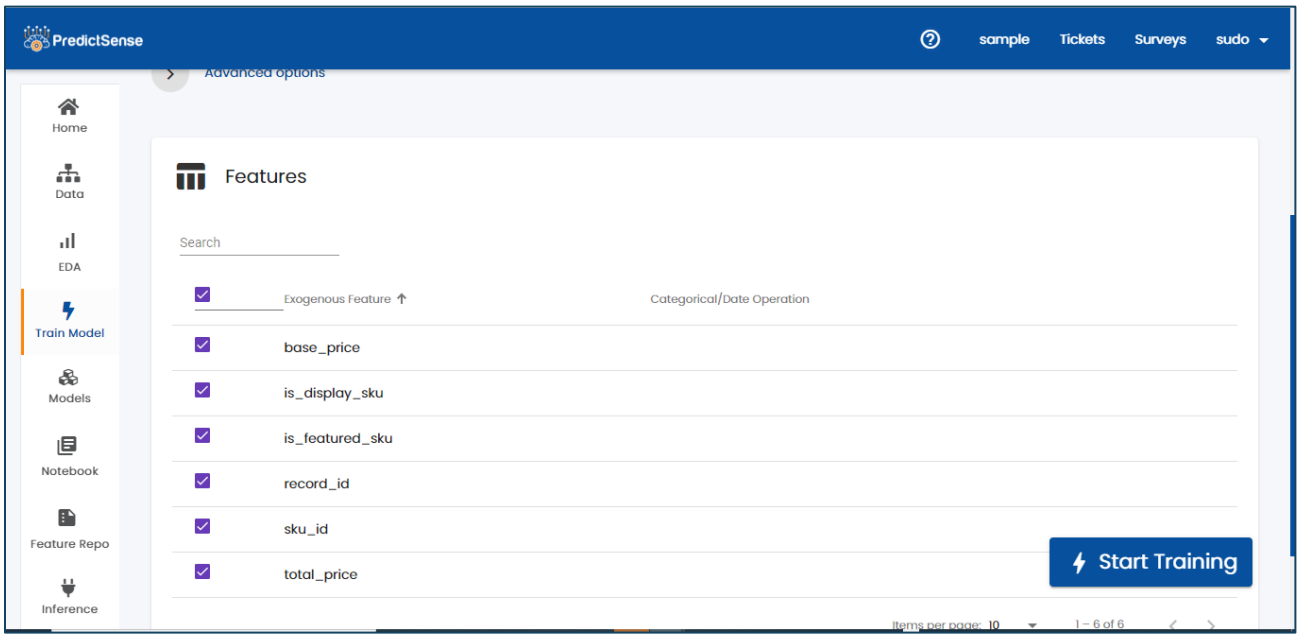
User will able to train model for **"Random forest" and "Linear Regression Ts".** For training a model, various advanced training options like **Components, Stationarity Test, feature transformation, Auto Correlation, Algorithm Parameters** can be selected to improve the training of model.

11.7.1 Training steps

1. Select dataset
2. Select algorithm from list
3. Enter number of steps
4. Select multiple independent features
5. Click on start training

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



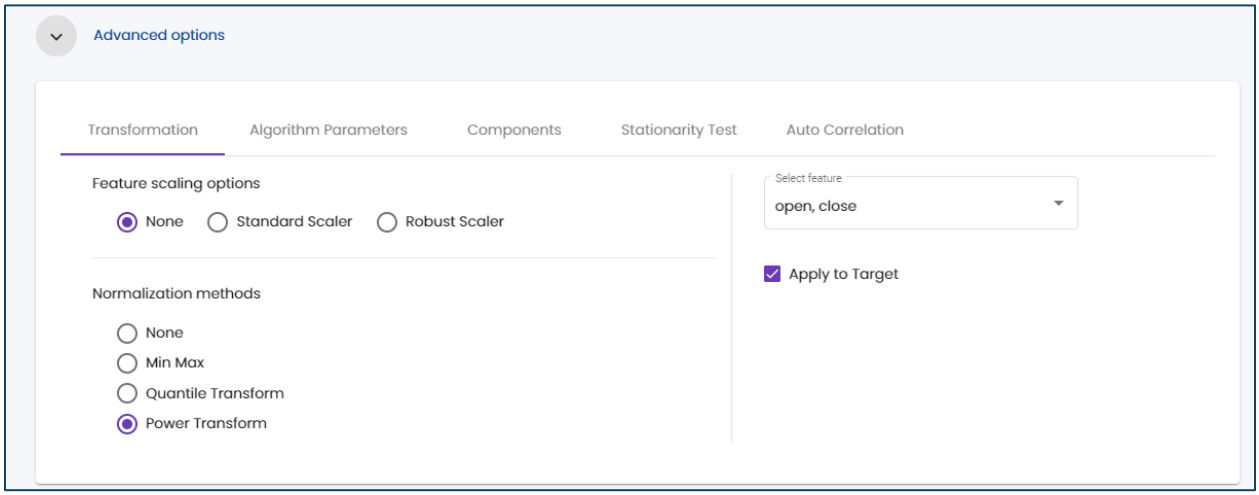


Train model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

11.7.2 Training steps with advanced options

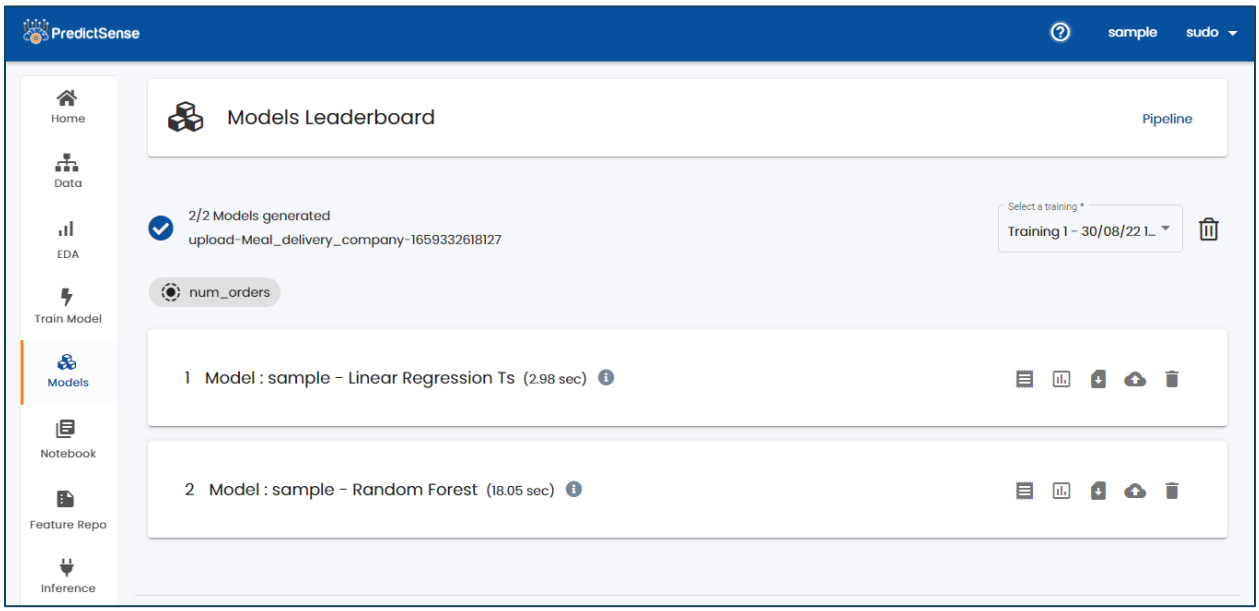
1. Select dataset
2. Select algorithm from list
3. Enter number of steps
4. Select various options from advanced options by switching between tabs like **"Components", "Stationarity Test", "Transformation", "Auto Correlation",** and **"Algorithm Parameters" etc.**
5. Select multiple independent features
6. Click on start training

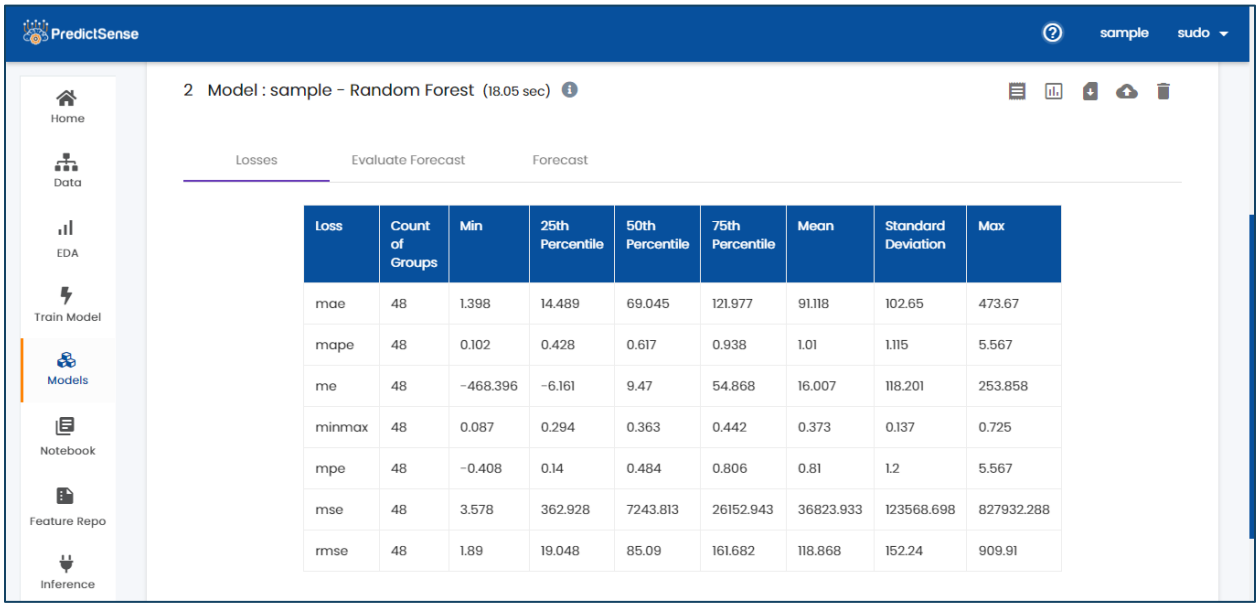


Advanced options

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

11.8 Models leaderboard

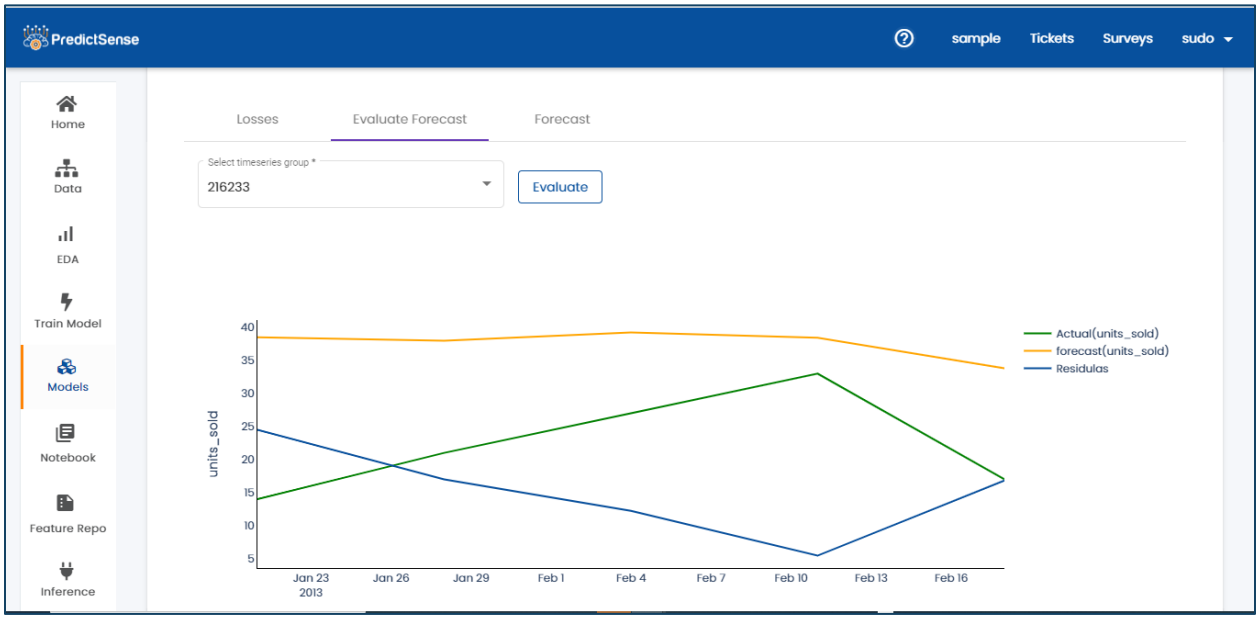




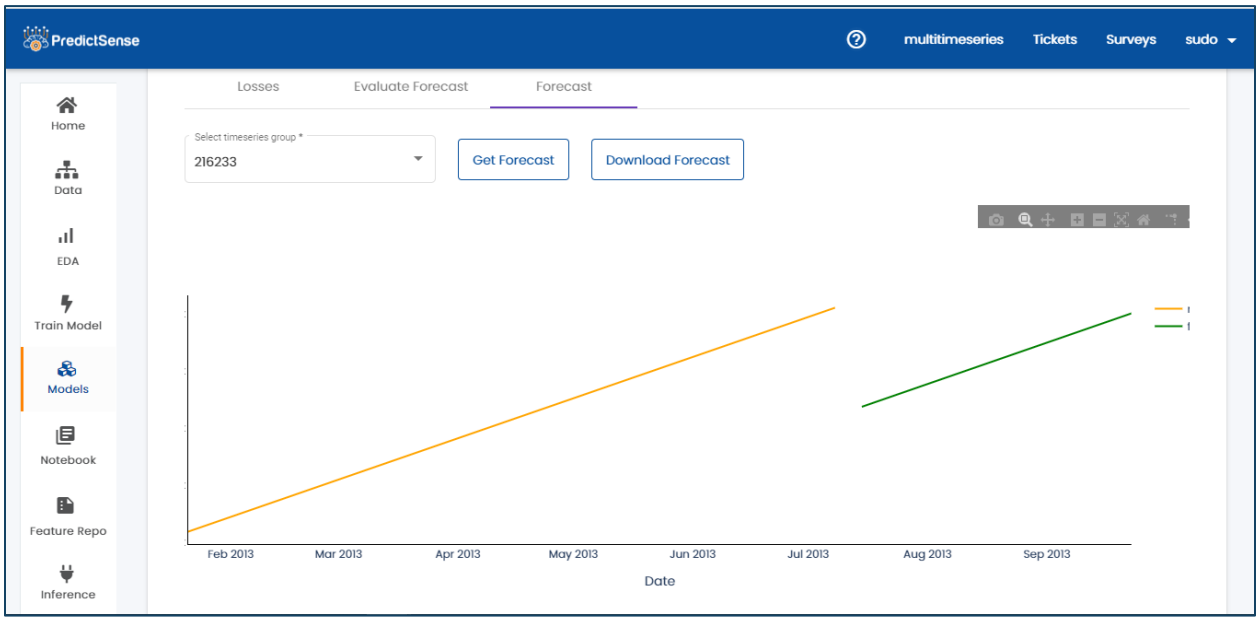
Loss	Count of Groups	Min	25th Percentile	50th Percentile	75th Percentile	Mean	Standard Deviation	Max
mae	48	1.398	14.489	69.045	121.977	91.118	102.65	473.67
mape	48	0.102	0.428	0.617	0.938	1.01	1.115	5.567
me	48	-468.396	-6.161	9.47	54.868	16.007	118.201	253.858
minmax	48	0.087	0.294	0.363	0.442	0.373	0.137	0.725
mpe	48	-0.408	0.14	0.484	0.806	0.81	1.2	5.567
mse	48	3.578	362.928	7243.813	26152.943	36823.933	123568.698	827932.288
rmse	48	1.89	19.048	85.09	161.682	118.868	152.24	909.91

Details model

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Evaluate forecast



Forecast

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

11.9 Pipeline

PipeLine Steps

Outliers Handling

Outliers handling is the process of treating outliers if exist any, based on several methods like: Factor method, Z score method and Inter Quantile Range method

View Details

Feature	Outlier Detection Strategy	Outlier Correction Strategy	Outlier Count
record_id	Inter-Quartile Range	Ignore outlier	0
store_id	Inter-Quartile Range	Ignore outlier	0
sku_id	Inter-Quartile Range	Ignore outlier	0
total_price	Inter-Quartile Range	Ignore outlier	0

Cancel

Pipeline

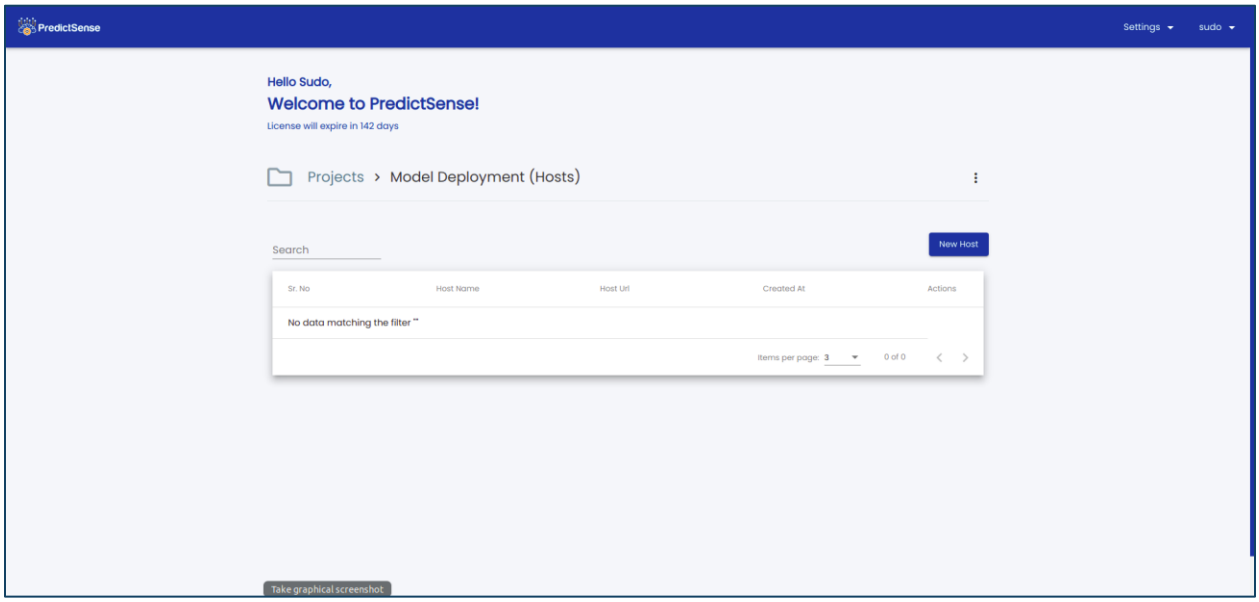
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

12 Model deployment using PredictSense

Model deployment functionality in PredictSense helps the user to deploy the model in one click based on the deployment settings.

12.1 Configure model deployment setting

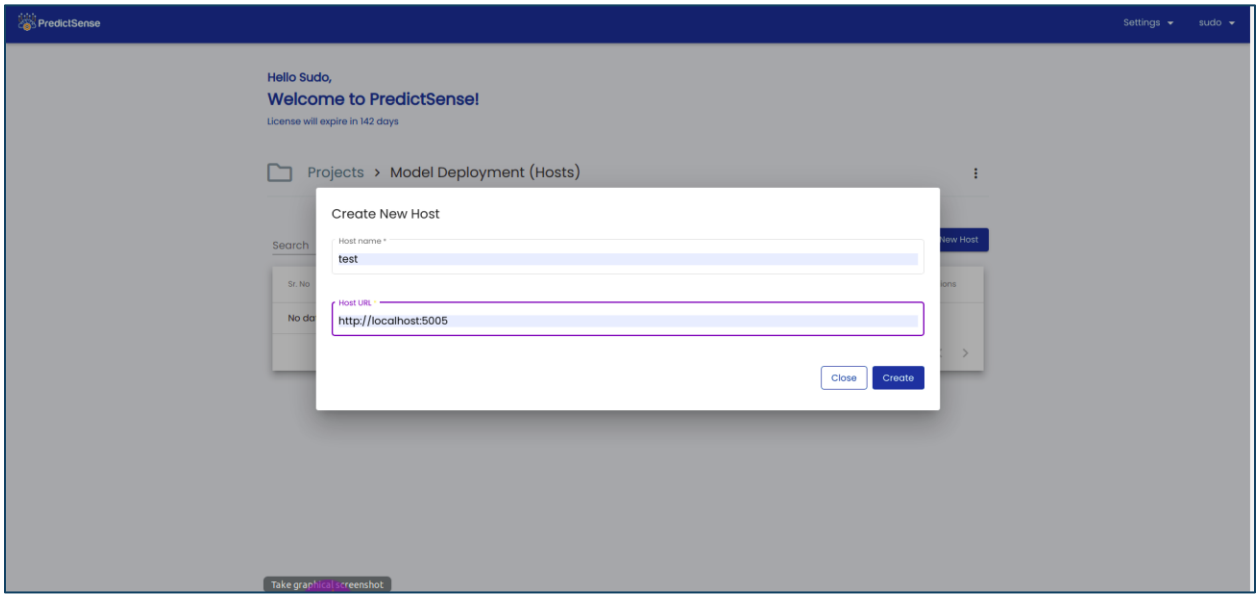
Login to PredictSense and click on model deployment under settings dropdown. Enter the Host name and Host URL like http://localhost:5005 or you can use the cloud URL wherever the Inference Engine is running and click on save button to save the settings.



Model deployment

a) You can add deployment settings for different servers and can deploy model on different server as well by selecting the correct host name in models deployment popup screen.

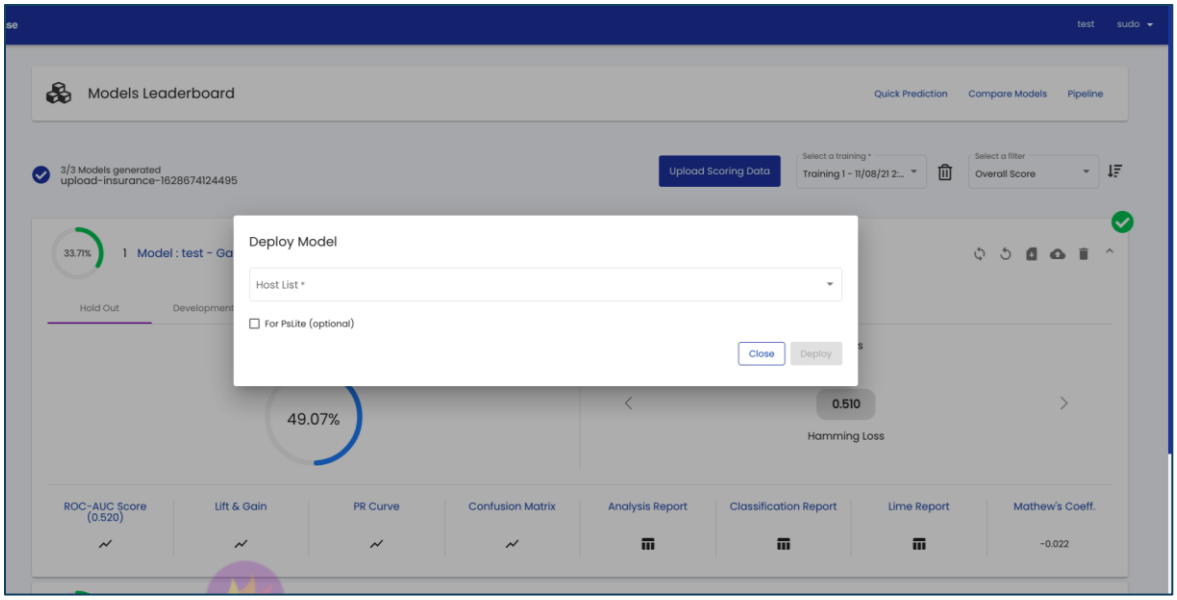
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



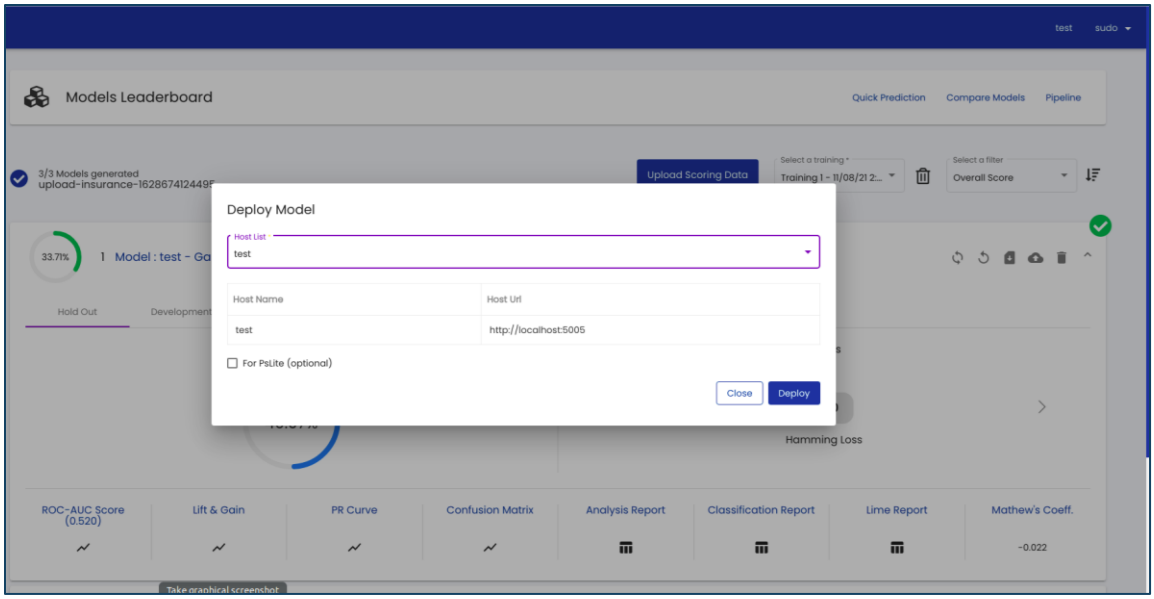
Create Host

12.2 Deploy model

a) Go to the models dashboard page and click on model deployment icon. It will open the deployment popup window where you can select the host name on which server you want deploy the model and click on deploy button

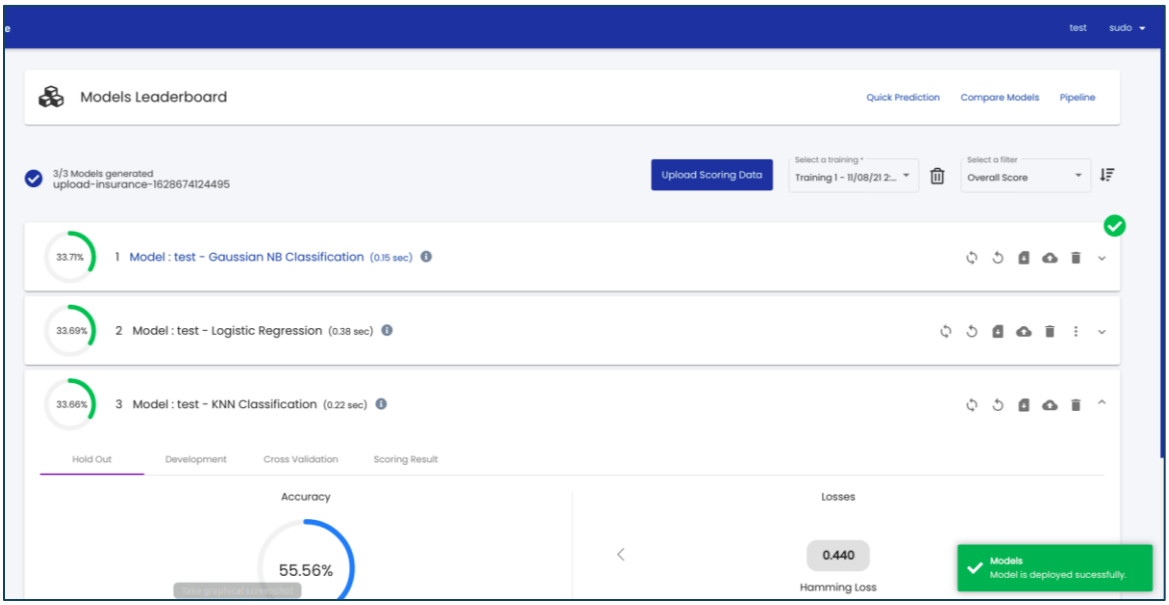


© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Deploy Model

b) Select host name from dropdown list and click on deploy button to deploy the model on server.



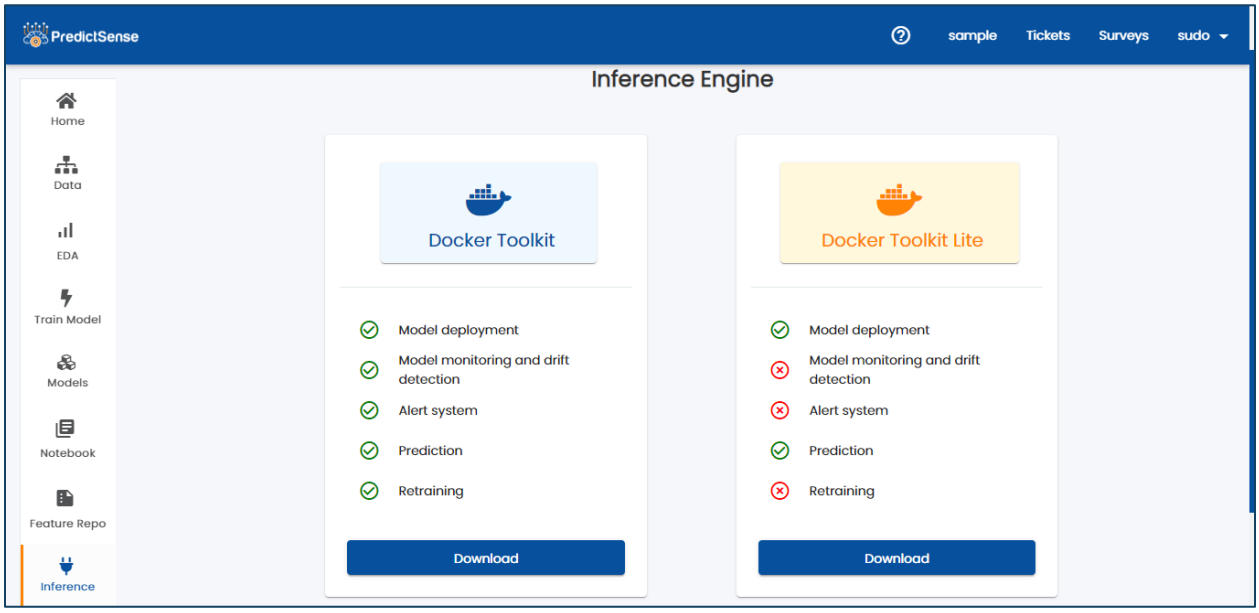
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13 Inference Engine

Inference utility helps to deploy and run the model anywhere on windows platform. You can run the Inference Engine and check the Prediction for any model before that you need to deploy the models from PredictSense in to Inference Engine. You can also retrain and schedule the retraining activity for SQL option.

13.1 Download Inference Engine skeleton for docker

Login to PredictSense, create any type of Project and download Inference Engine skeleton for docker.

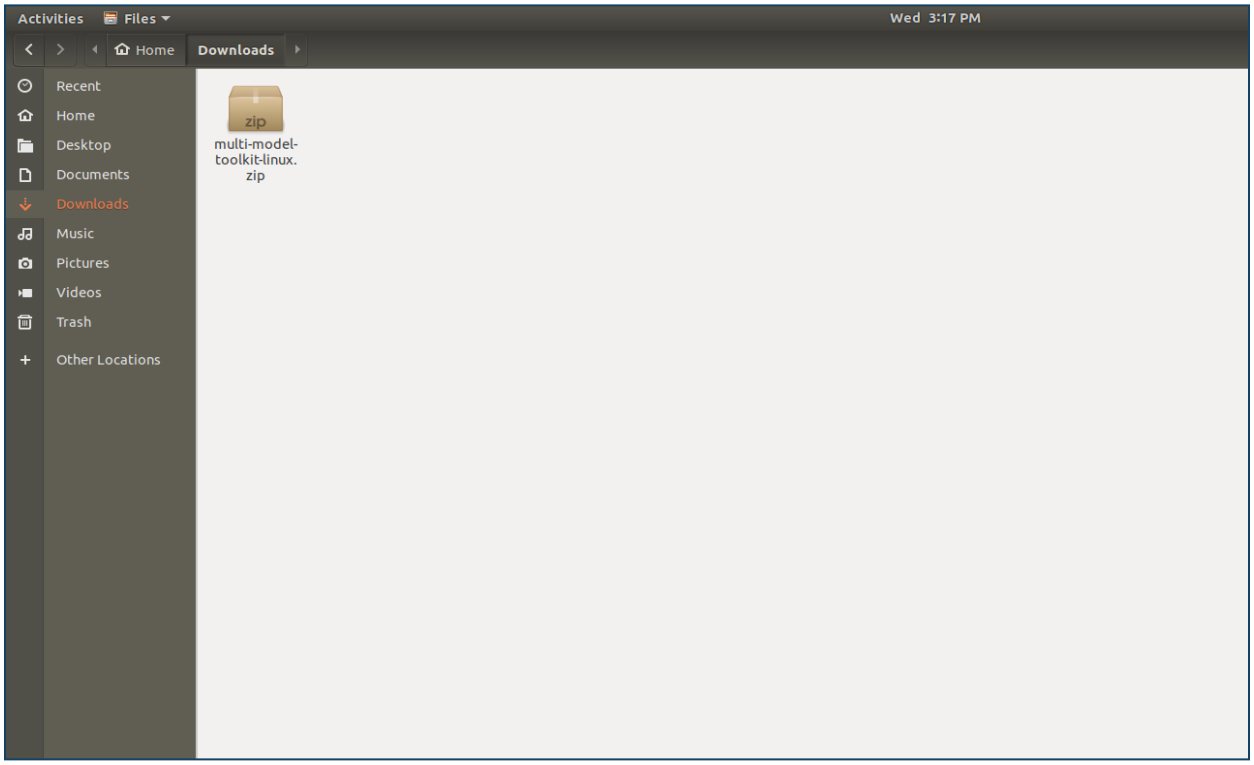


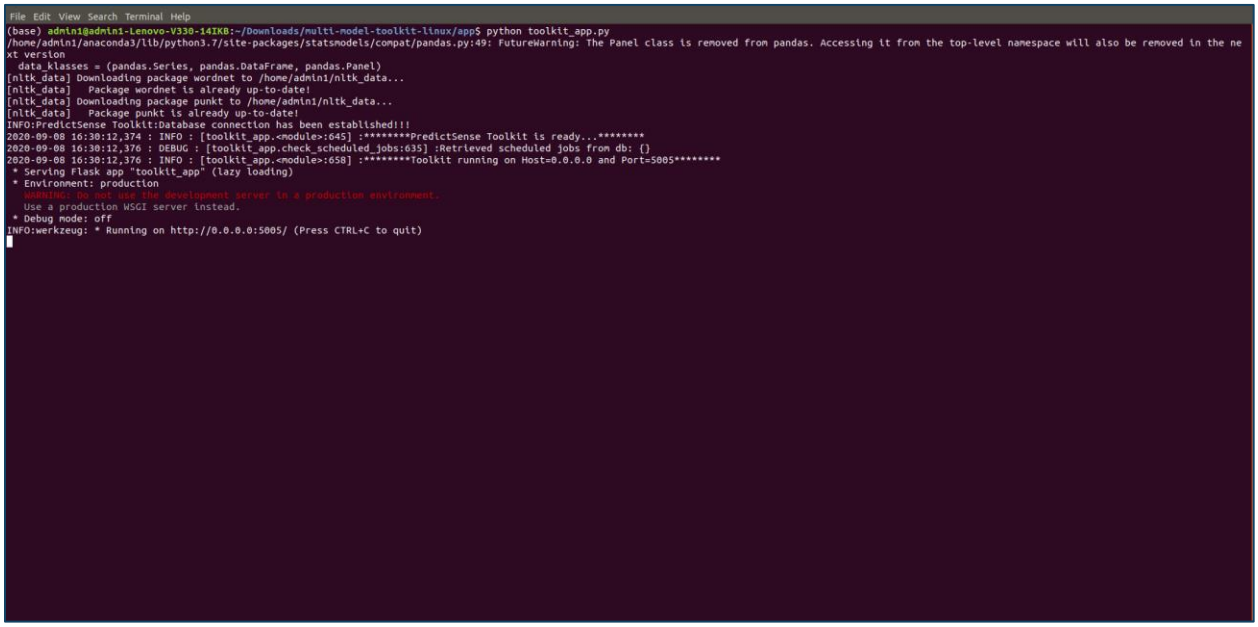
Inference Engine page

13.2 For Docker Toolkit

Run docker Inference Engine using **docker-compose up –build** command. Once the above operation is completed successfully, Inference Engine will run on **http://localhost:5005**.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

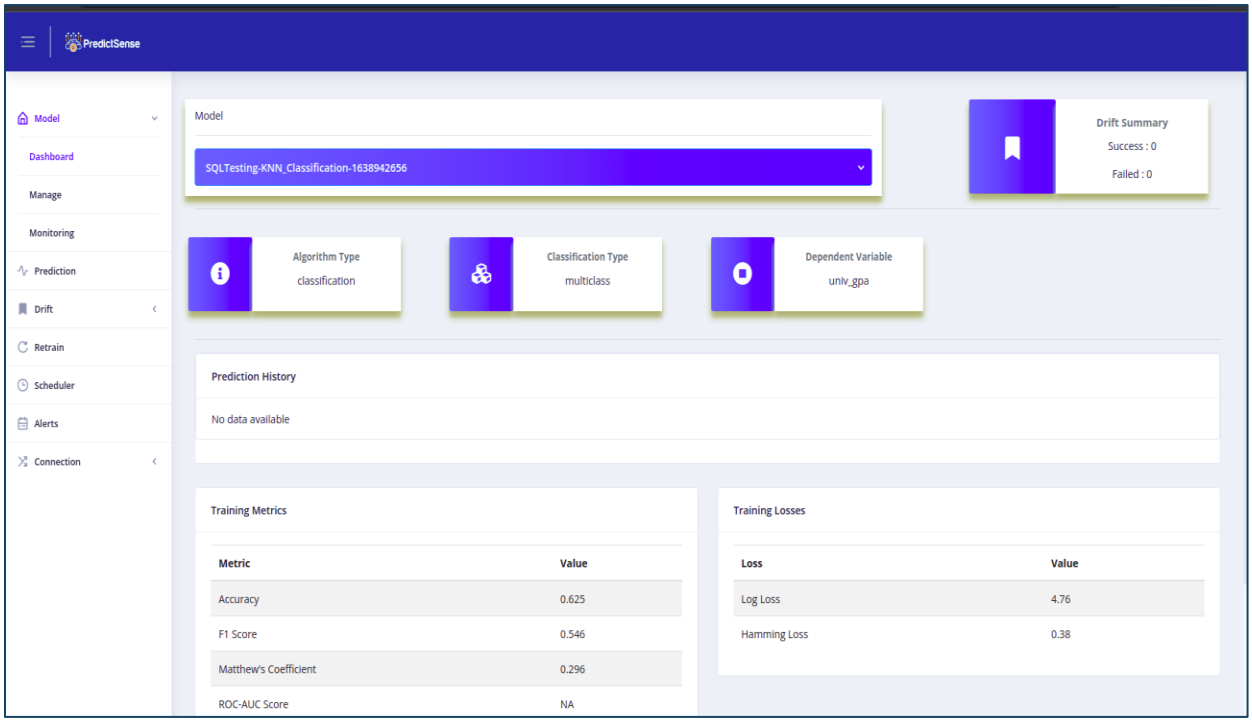




© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

13.3 Dashboard

Inference Engine dashboard shows the model related info like drift summary, algorithm type, no. of independent variable, dependent variable, training metrics, training losses, prediction history, classification report, confusion matrix and feature coefficient for the selected model. Also, in Inference Engine we can monitor the model performance for week, day, month etc.

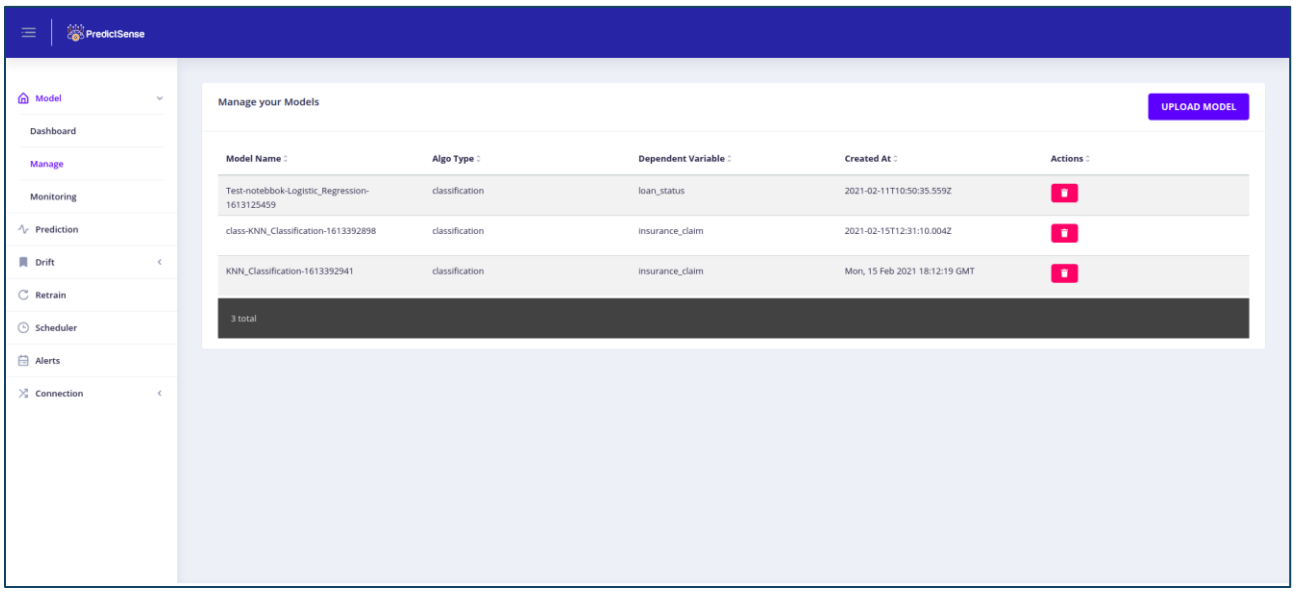


Dashboard

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

13.4 Manage

In manage module you can see the all the deployed model list along with the following details model name, algorithm type, dependent variable, created at, delete action, upload model. You can upload model zip file by clicking on upload model button and perform the prediction, retraining, monitoring model etc.

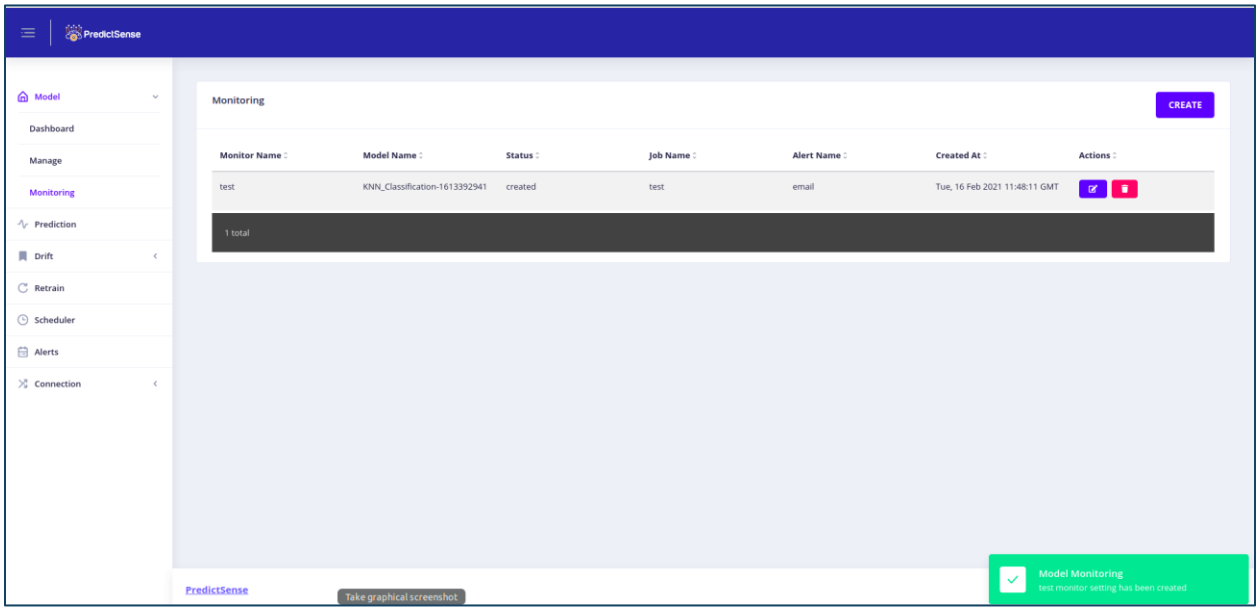


Manage screen

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13.5 Monitoring

In this module you can monitor the performance and compute the drift report of any model for hour, day, week, month. To compute drift report first need to add the drift config. In dashboard page you can see the drift summary status in success and fail. Once you click on drift summary it will redirect to drift report page were you can view the reports etc.



Model monitoring

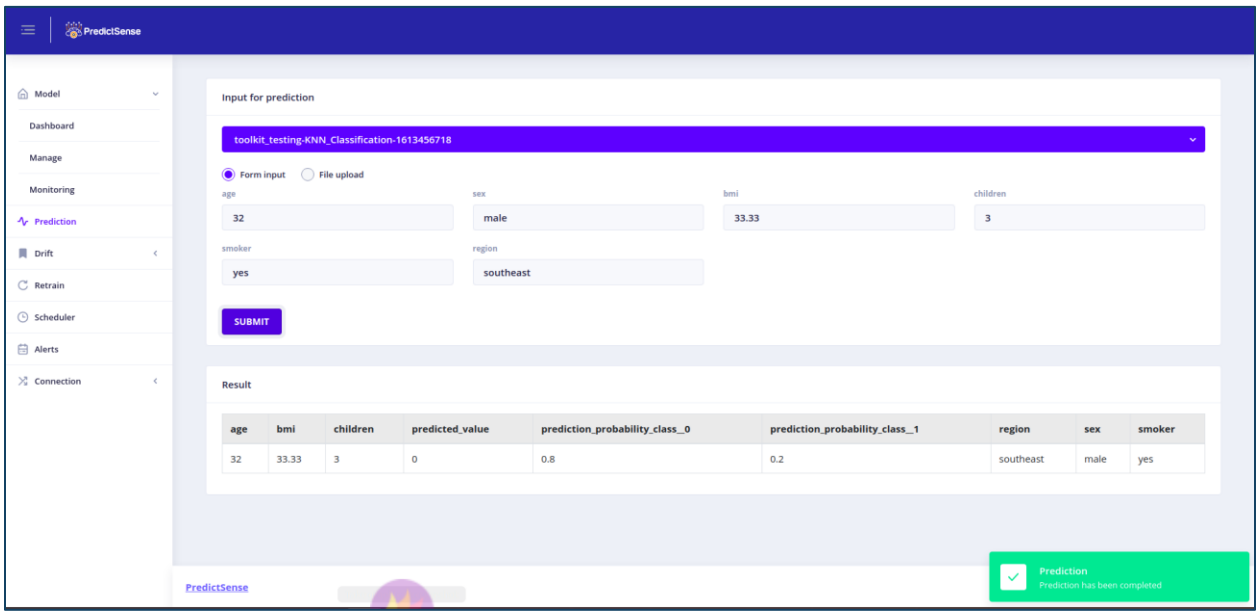
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13.6 Predictions

Refresh the page and go to the prediction page, check in dropdown deployed model is showing or not.

a) **Form input prediction**

you can pass the input parameters and click on submit button.

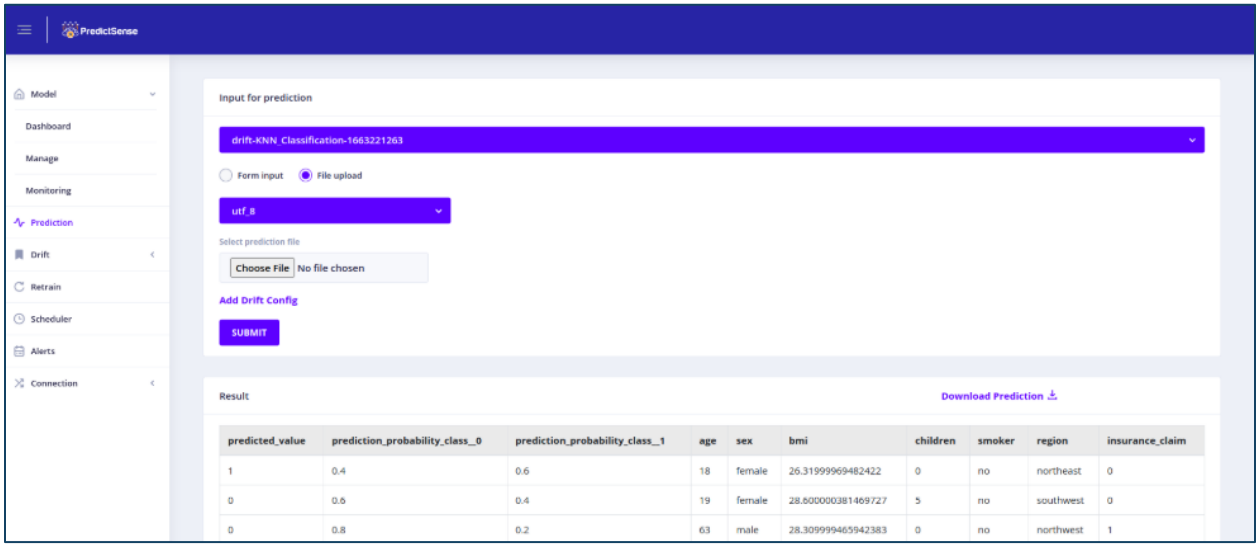


Input Prediction

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

b) File upload prediction:

You can upload the .csv file and select the file encoding like utf-8, latin_1 and click on submit button to check the predicted values.

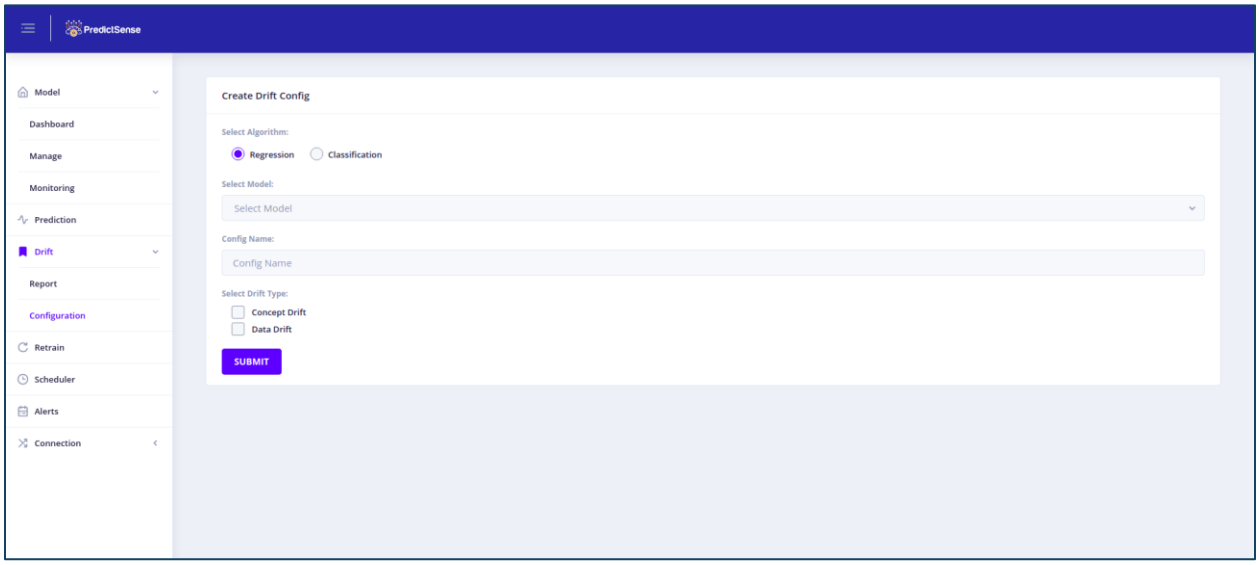


File Upload Prediction

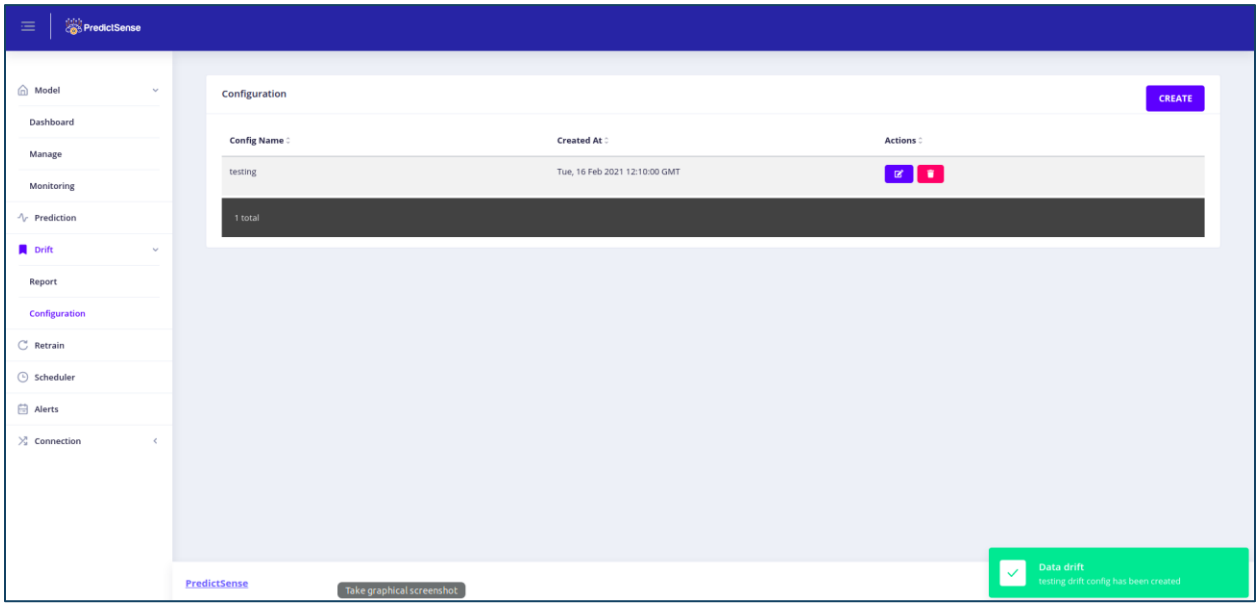
13.7 Drift

This module segregated in two-part report and configuration. In configuration module you can create the configuration for any regression and classification module and use it during the prediction to compute the drift report. You can see all the reports in report module.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

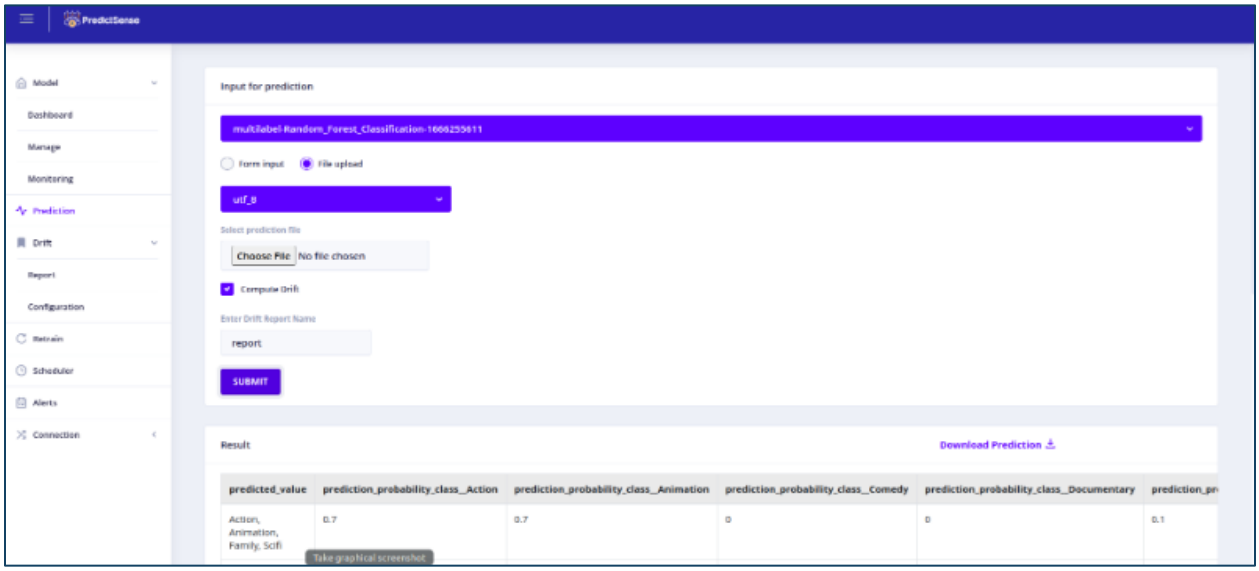


Create Drift Config

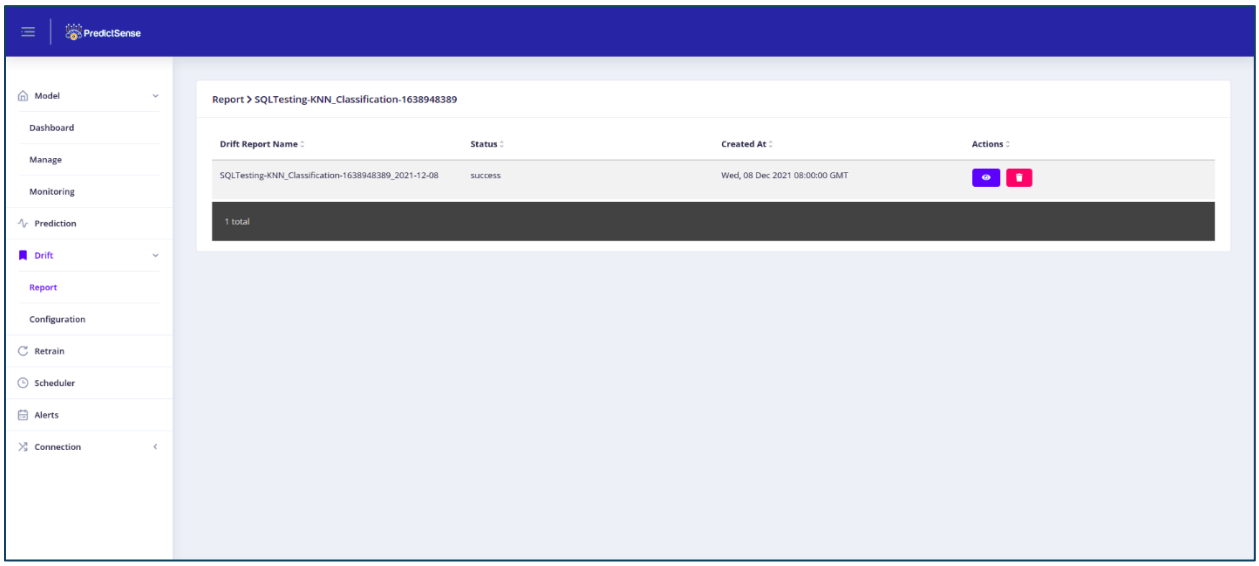


Configuration

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

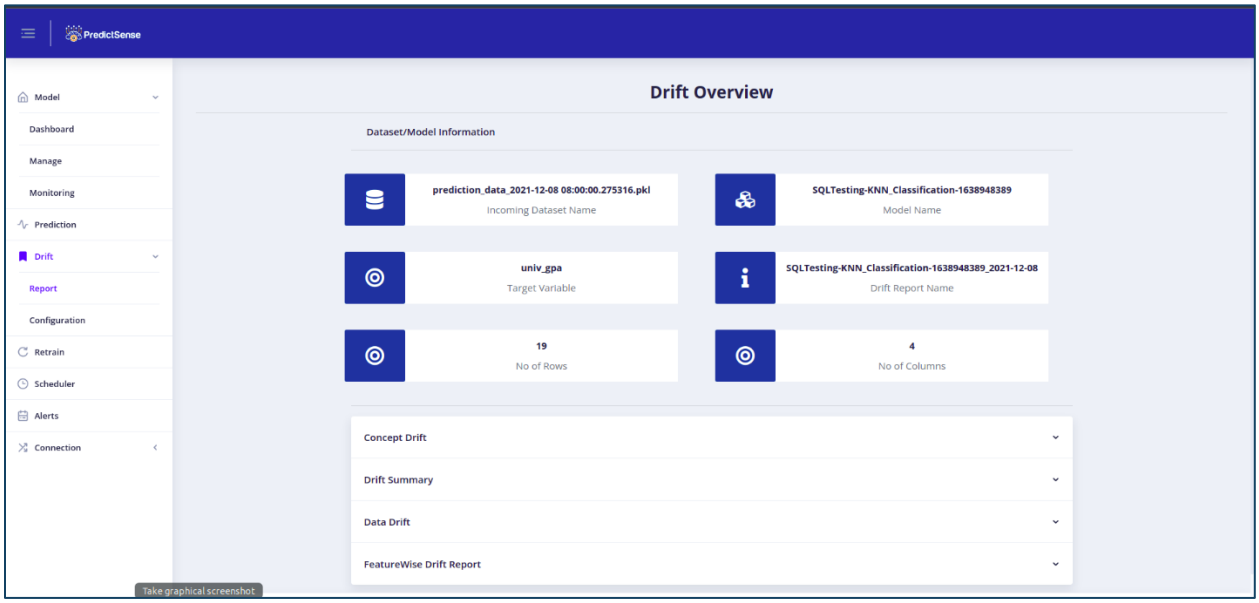


Compute Drift



Drift Report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

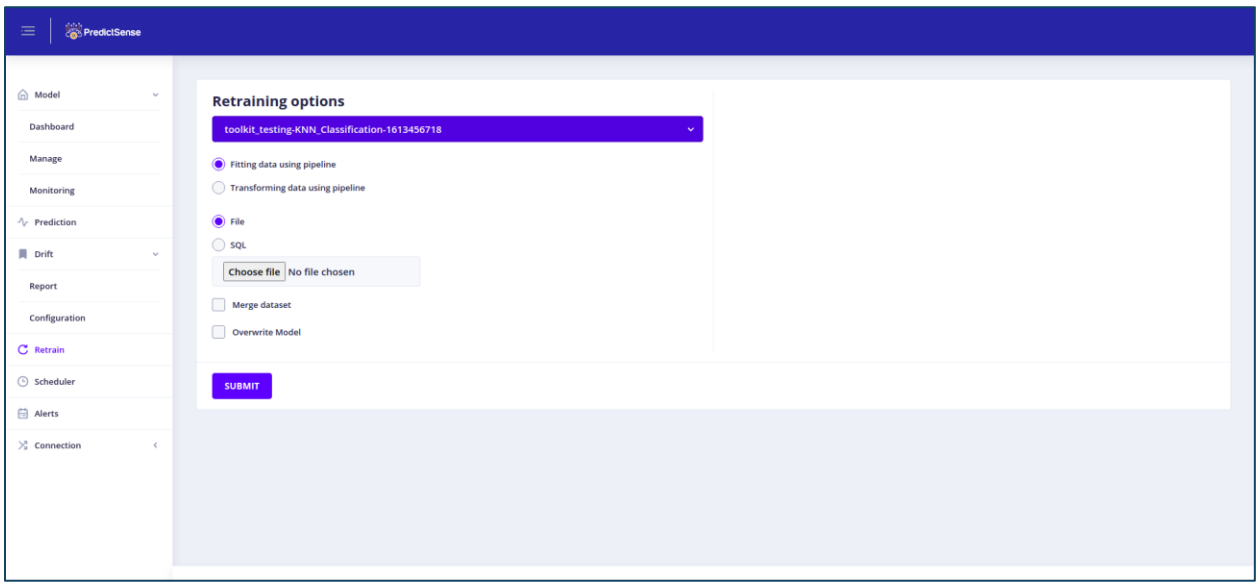


Drift Overview

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

13.8 Retrain model

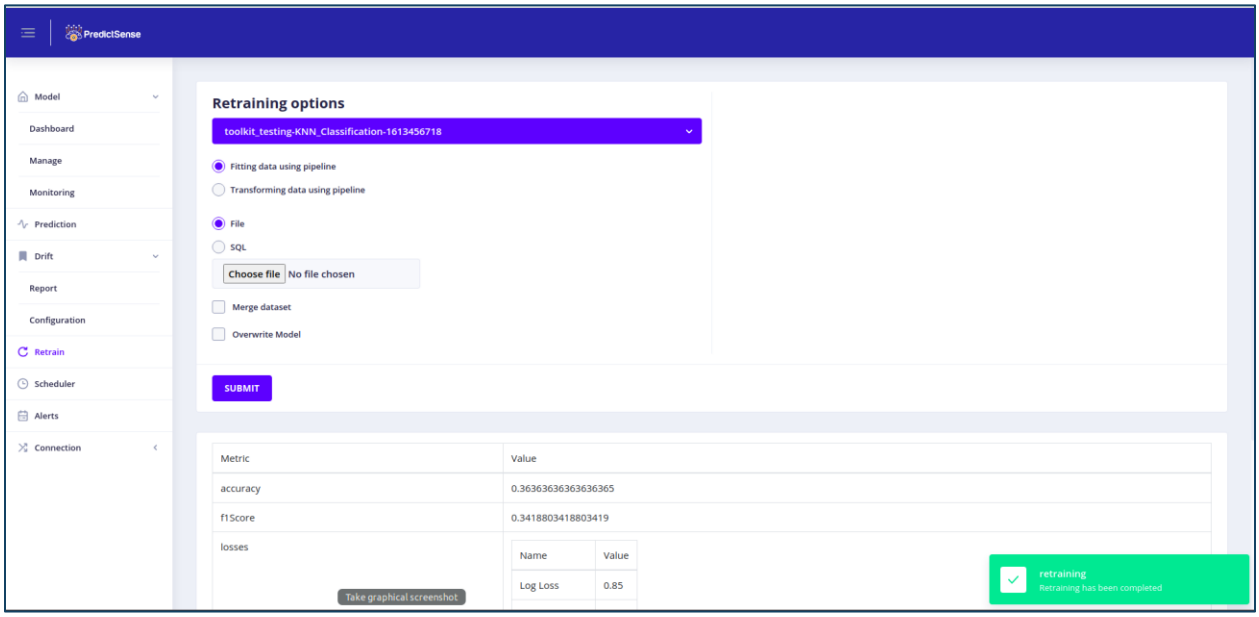
Retrain model provide the functionality to retrain the existing model on different datasets. It also provides the merge dataset and overwrite the model functionality.



a) **Fitting data using pipeline**

By using this option you can retrain the existing model with new datasets and create the new model. It will display with same name but different ID.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

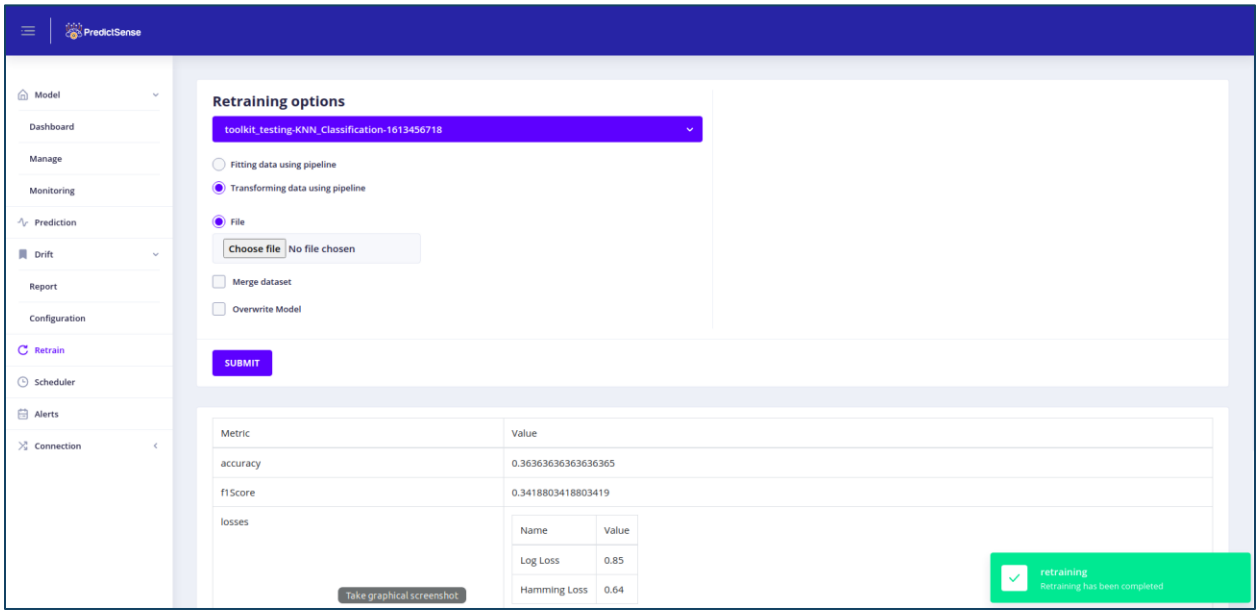


Fitting data using pipeline

b) **Transforming data using pipeline**

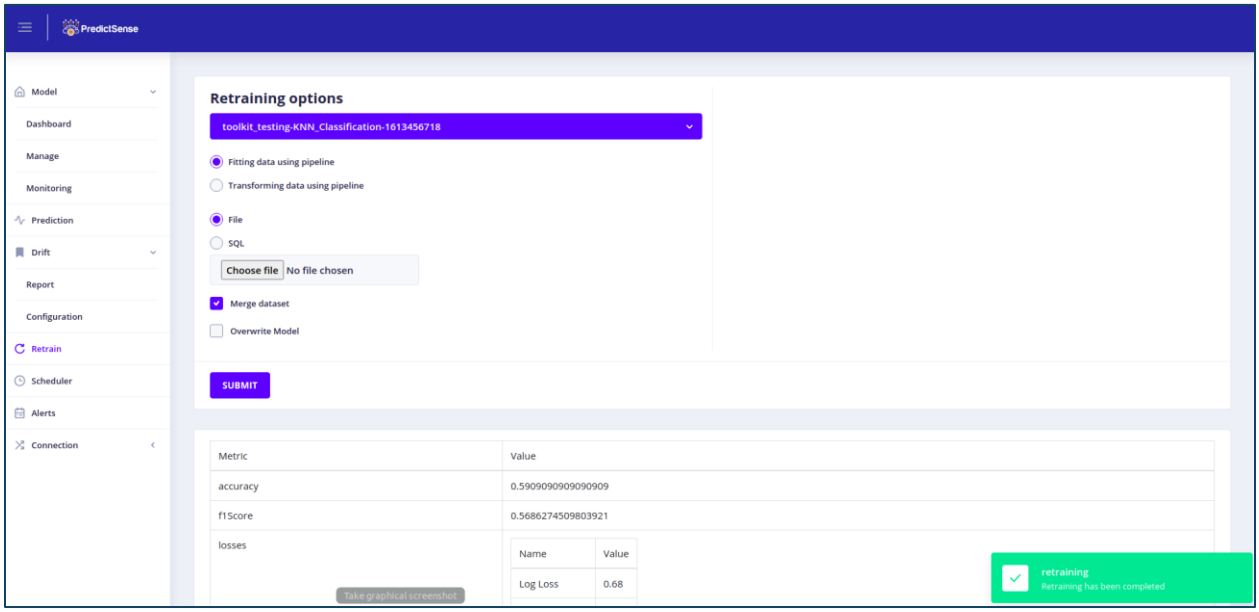
By using this option you can retrain the existing model with new datasets and create the new model. It will display with same name but different ID and you can see the newly generated model in dropdown list.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



c) **Merge dataset**

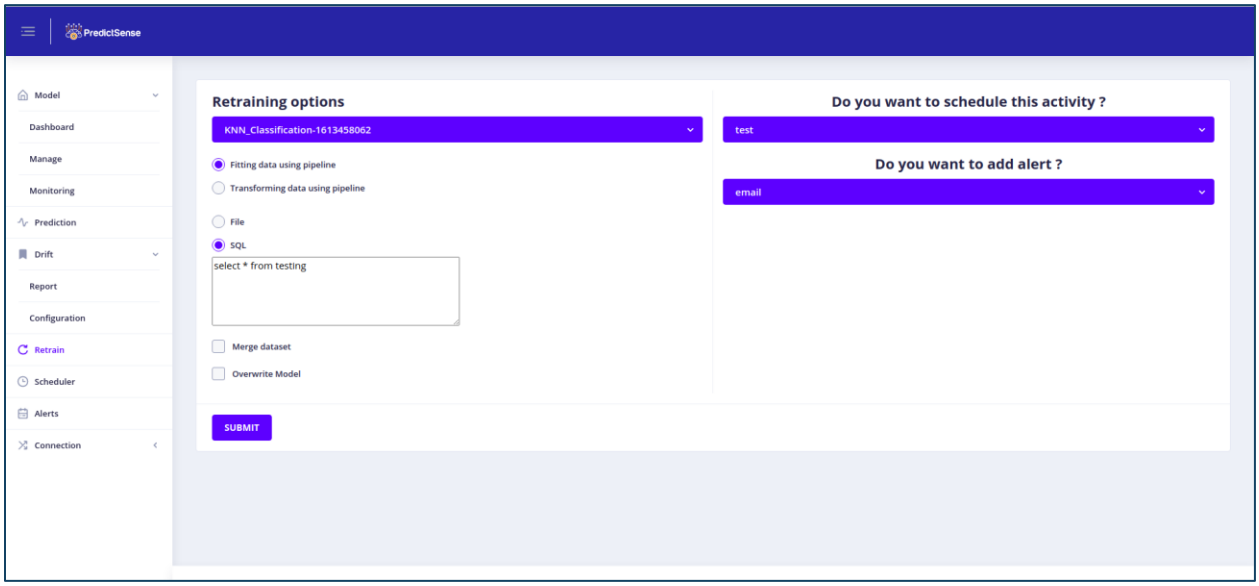
By using the merge dataset option during retraining you can merge two datasets and it will create the new model.



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

d) **SQL**

Use the sql option if your original model has been trained on a dataset pulled from any of the sql sources. You can even modify the query and also you can schedule the retraining activity for sql option. First you need to create scheduler from scheduler module and use it while retraining. You can add the alert setting so whenever any new model got retrained you will get the notification alert message to respective email id.

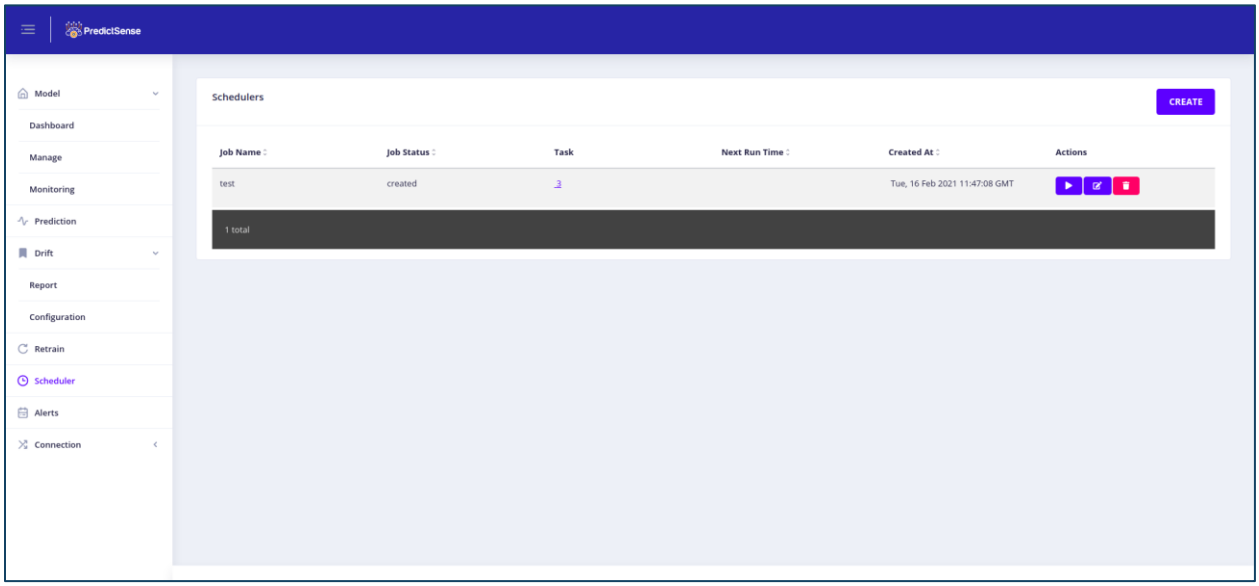


13.9 Scheduler module

Now you can create scheduler in Inference Engine app and schedule the retraining activity for SQL , model monitoring options. Also you can add multiple task in single scheduler and create multiple scheduler for multiple tasks.

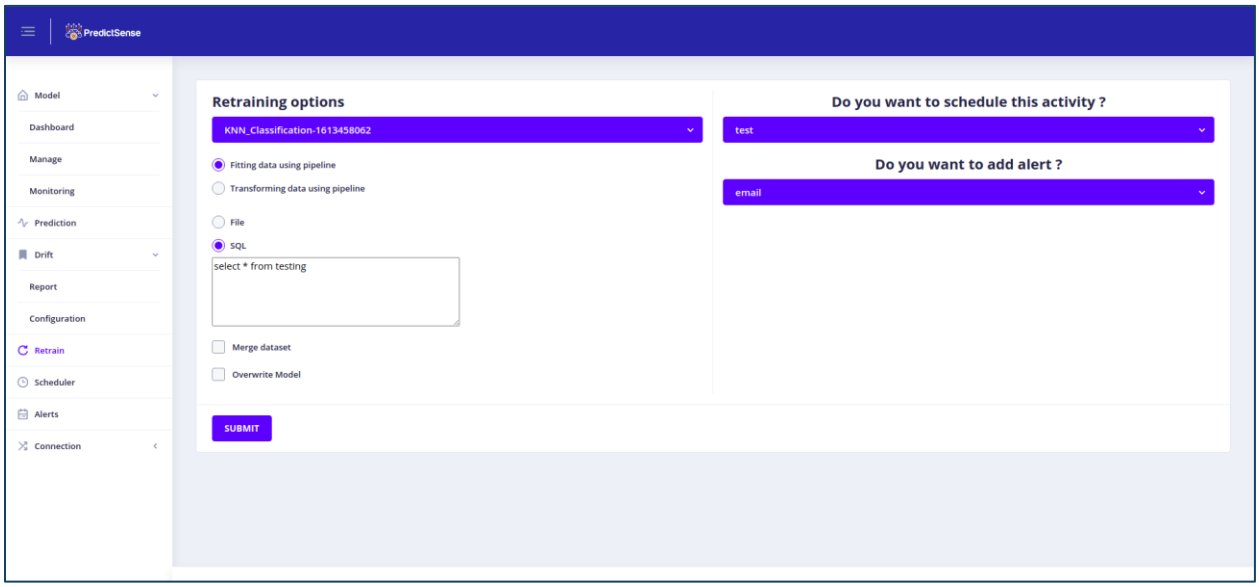
a) Go to scheduler page and create scheduler for hours, day, week, months etc . You can also delete the scheduler and pause the running scheduler activity.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Scheduler

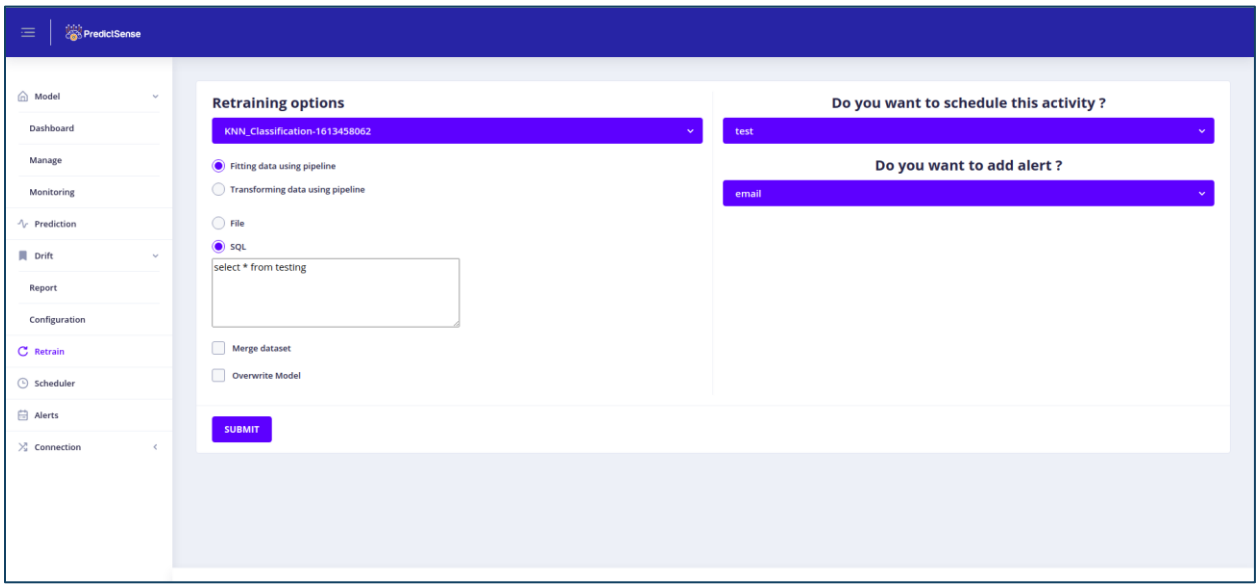
b) Go to the retraining page and select SQL trained model and select SQL option in right side of the screen it will display the scheduler activity were you can any scheduler activity from dropdown and click on submit button.



Retrain

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

c) you can see the next retraining and wakeup time in scheduler module.

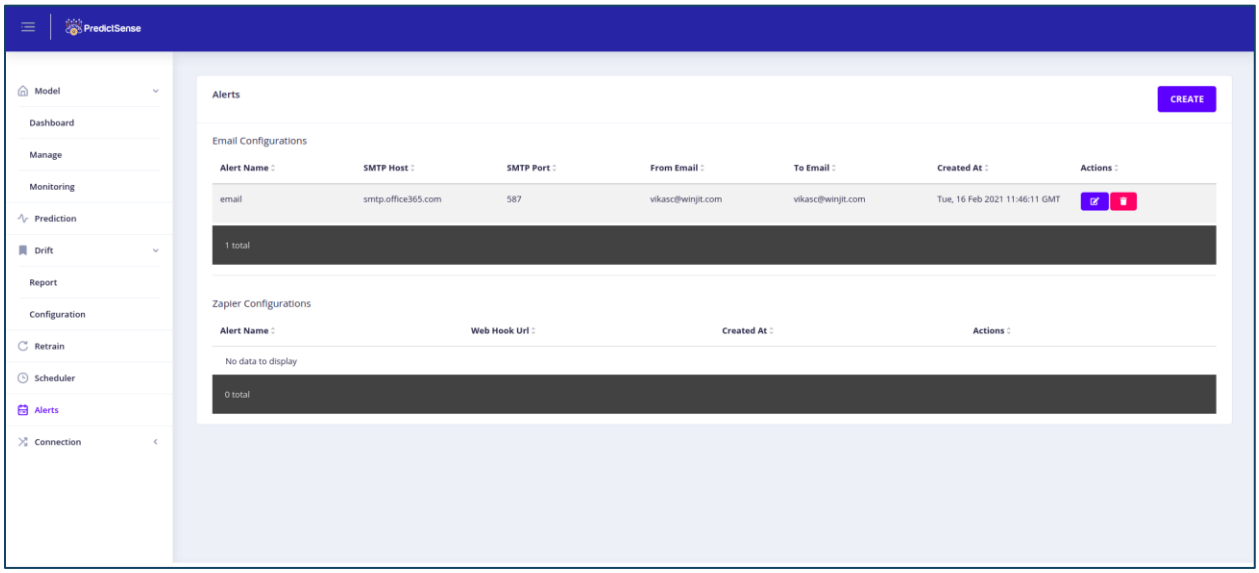


Retraining using scheduler

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13.10 Alert

In this module we can configure the alert for gmail, office365 and zapier. We can use the alert for retraining, model monitoring task to get notification during each run of scheduler.

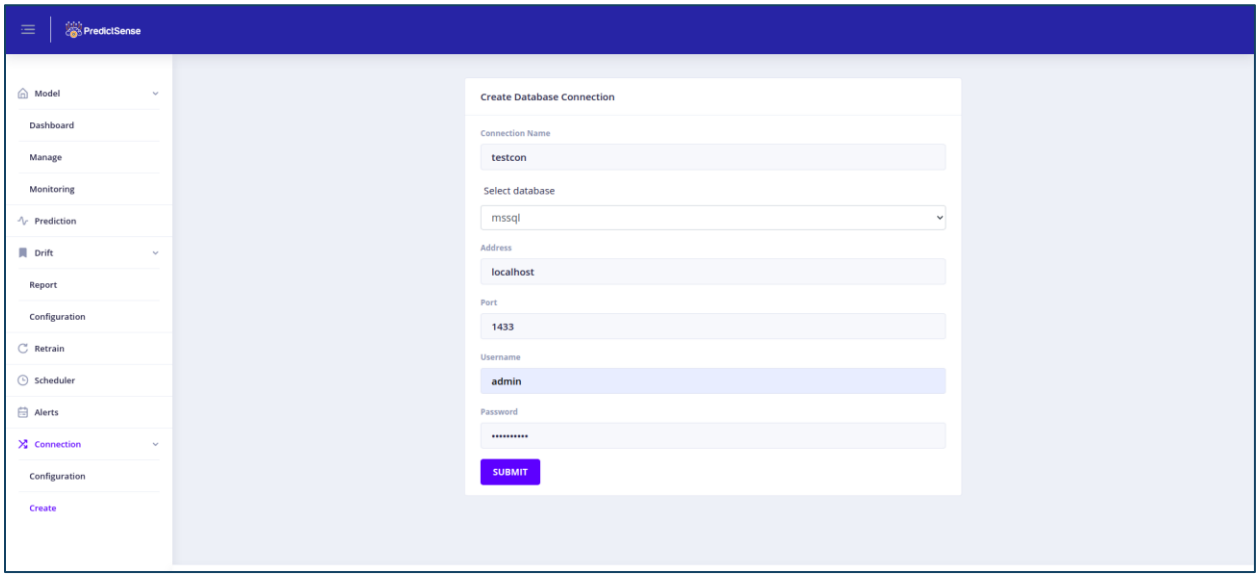


Alert

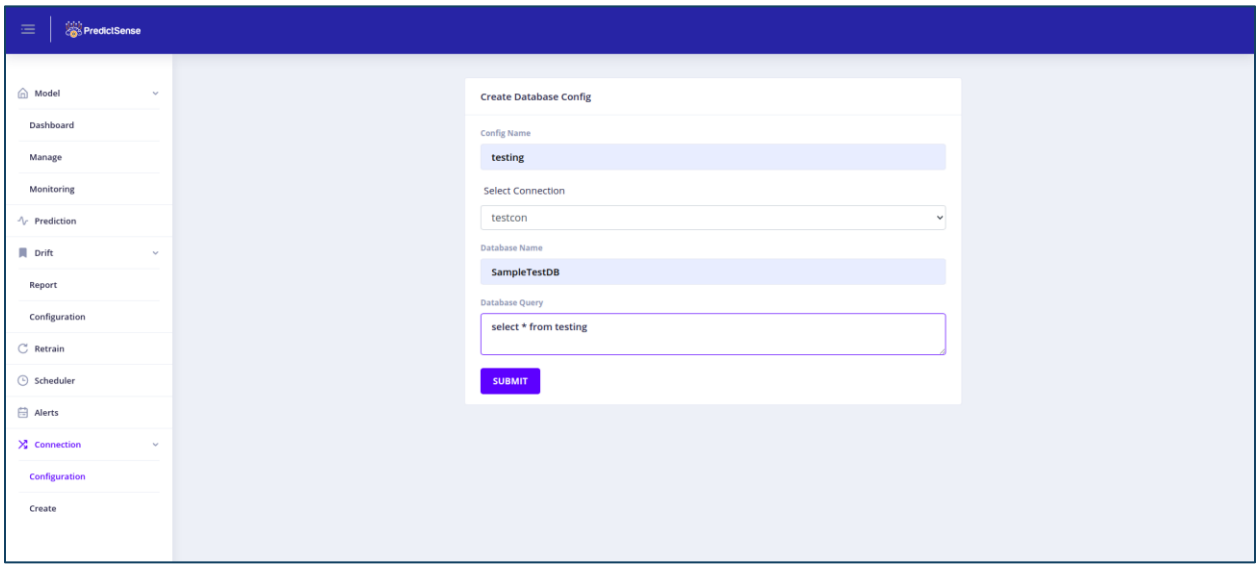
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

13.11 Connection

It is segregated in to two parts create and configuration. In create module you can create database connection for MSSQL, MySQL and PostgreSQL. In configuration module you can configure the query and use it during retraining.



Create Connection

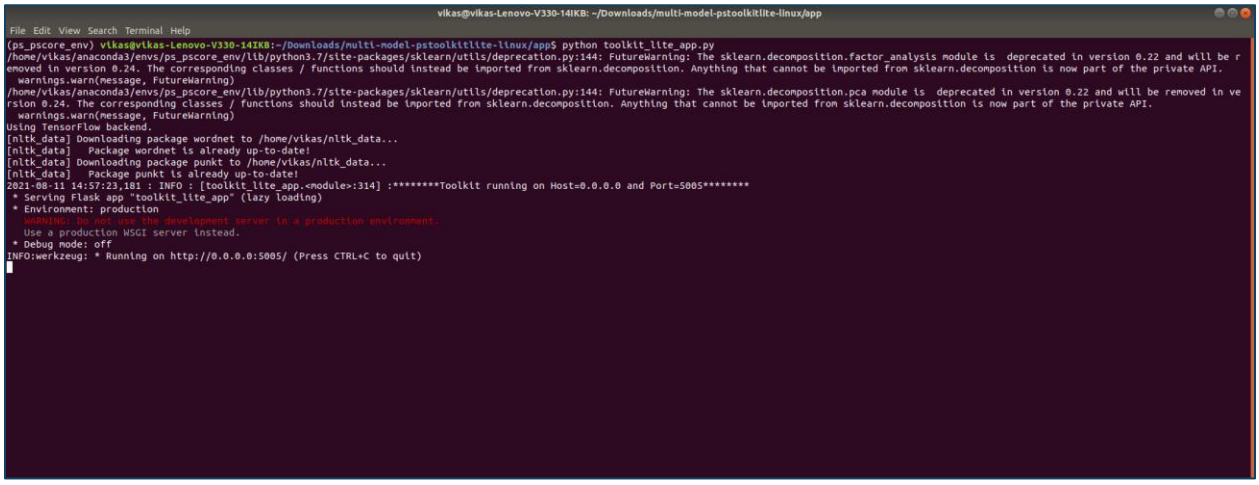


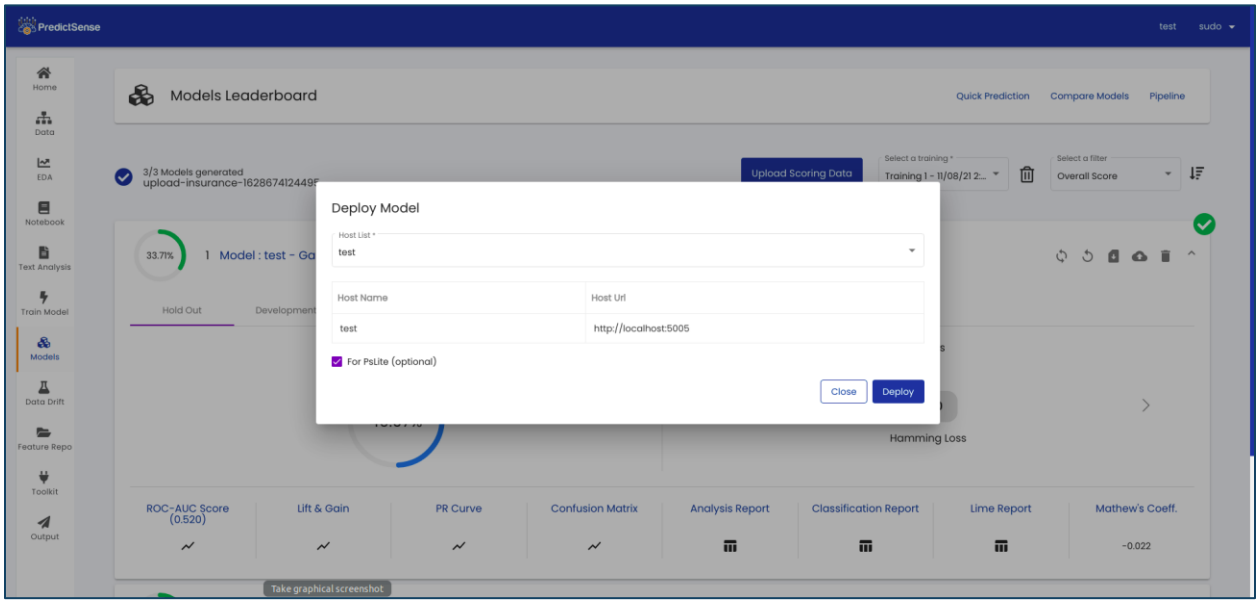
Configure database

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

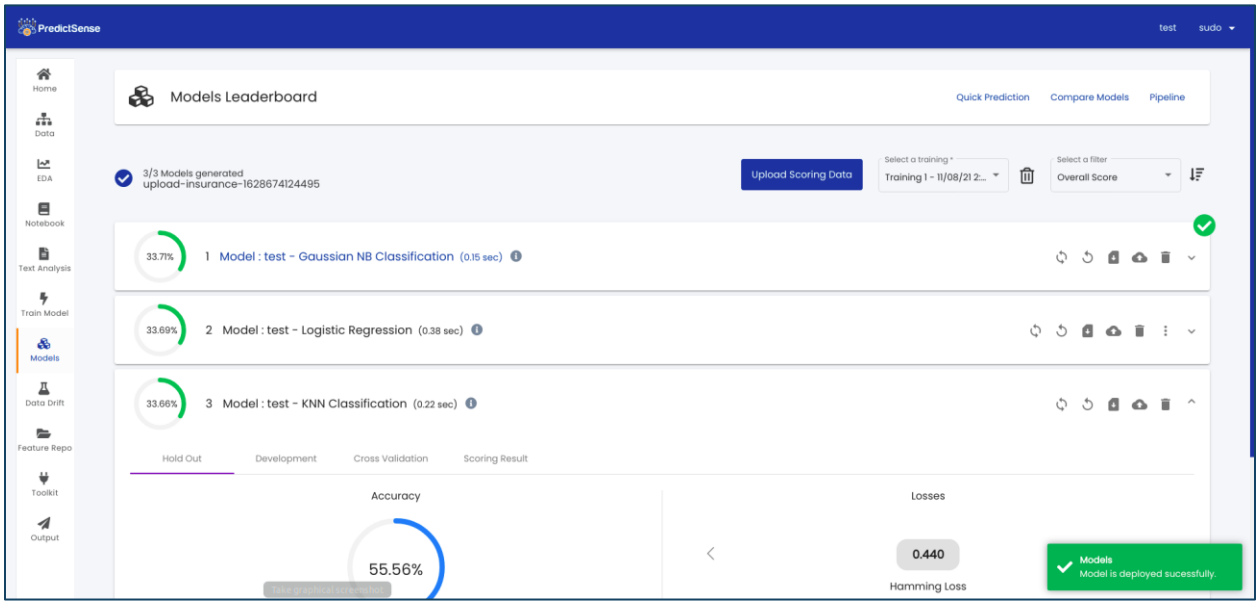
13.12 Download Docker Toolkit Lite

Login to PredictSense, create any type of Project and download Inference Engine skeleton for docker.

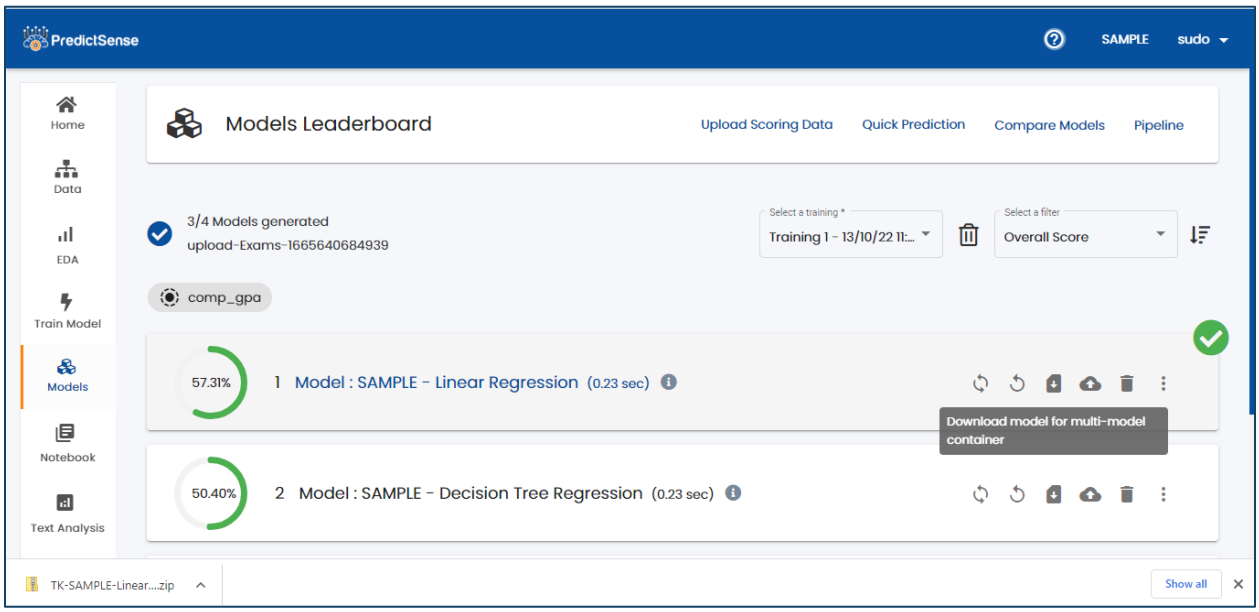




© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

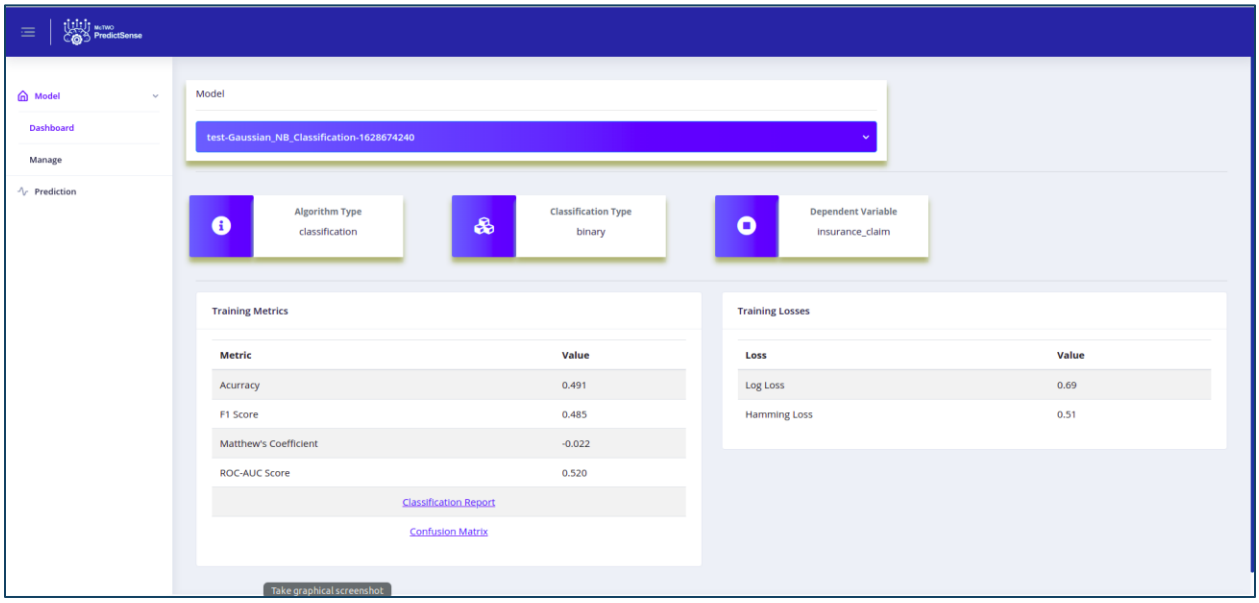


Deploy model

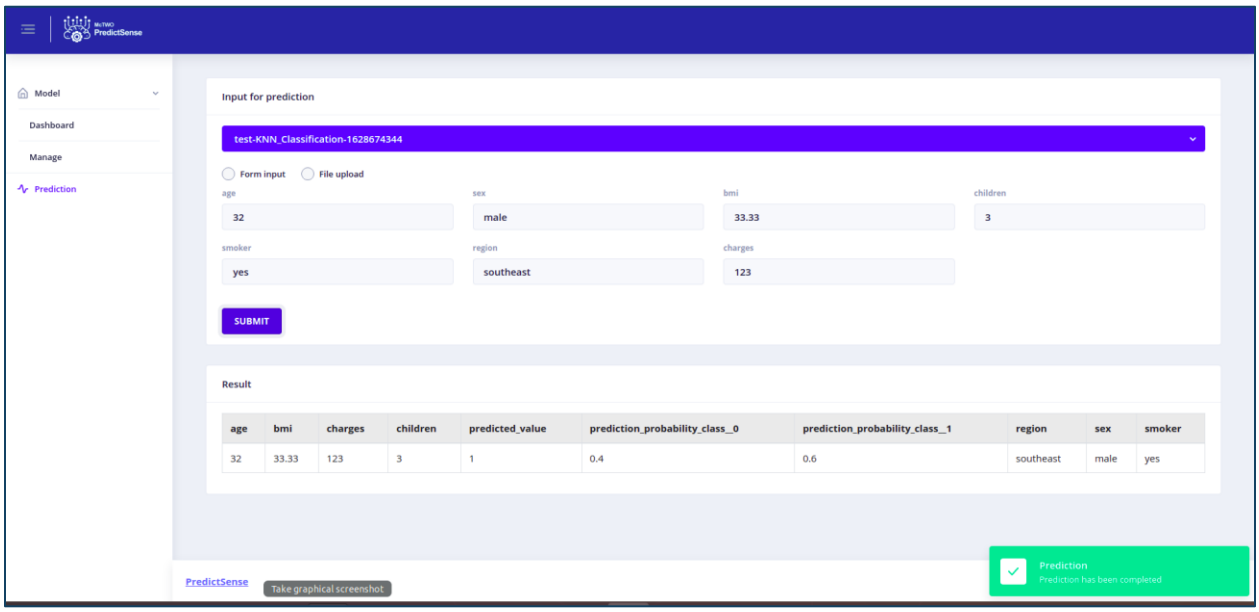


Upload model

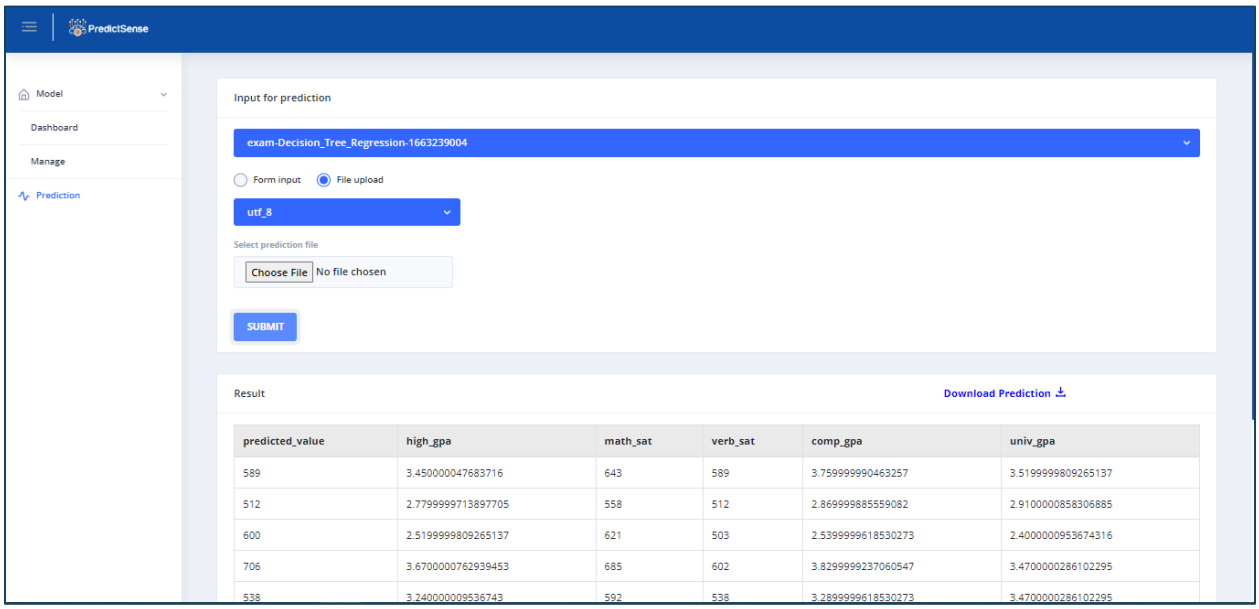
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



Dashboard



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



Input prediction

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

14 Drift Analysis

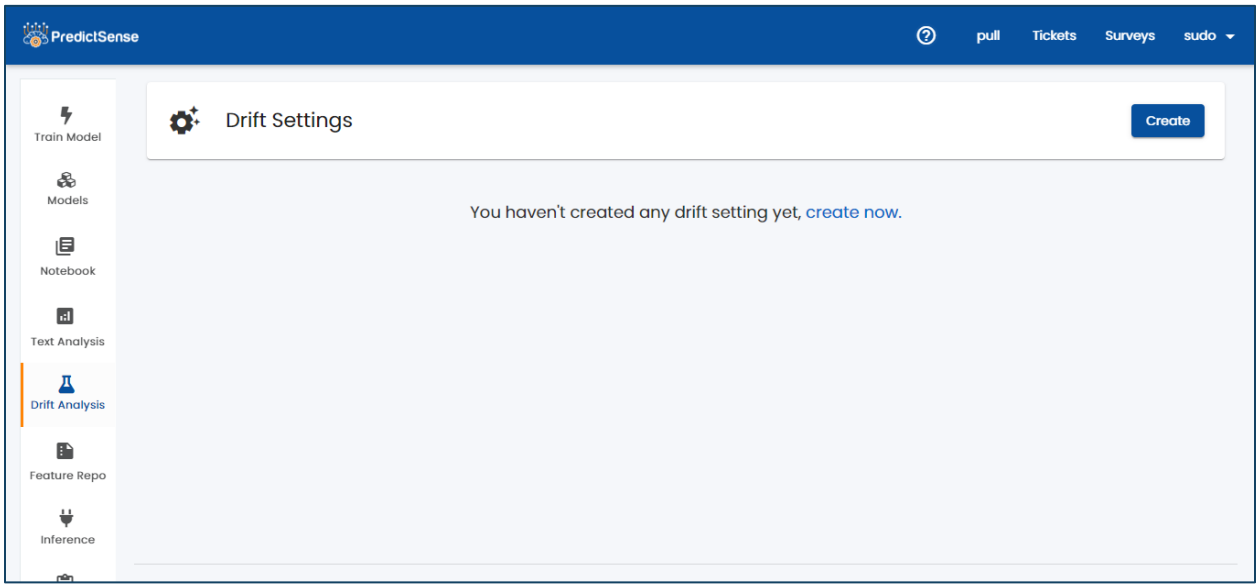
Drift analysis in machine learning refers to the change in the relationships between input and output data in the underlying problem over time.

For example, one concept in weather data may be the season that is not explicitly specified in temperature data, but may influence temperature data.

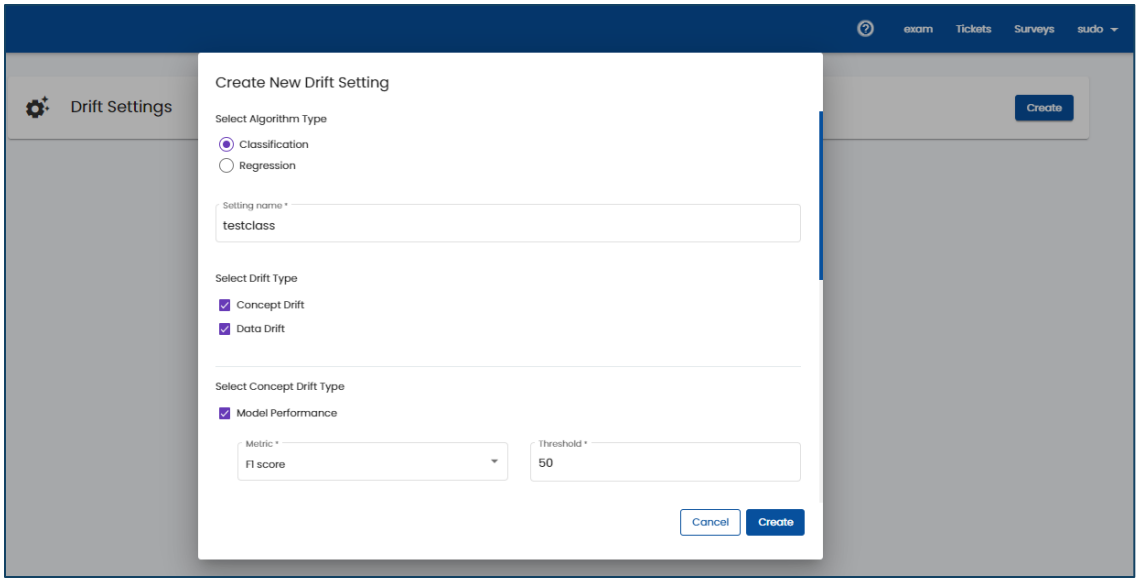
Uploaded two datasets in which one dataset is having less no. of feature and another one is having more no. of feature. You can add more no. of labels in any one column and check the drift is present or not.

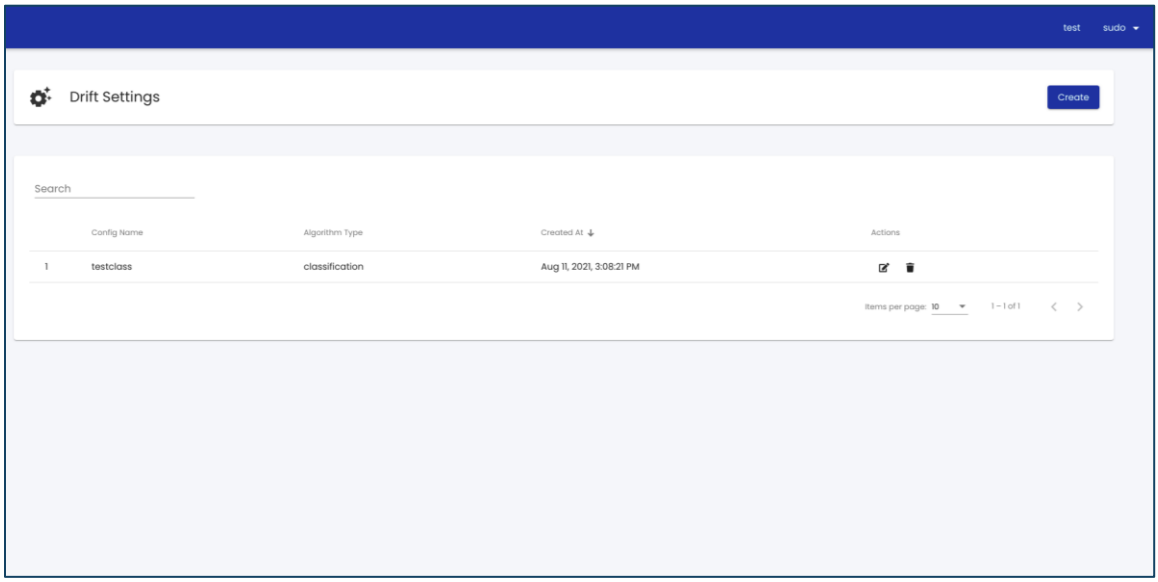
14.1 Drift settings

In this module you can add the drift configuration for regression and classification model. In models page whenever you compute the drift report you need to select the drift configuration and pass the drift report name.



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|





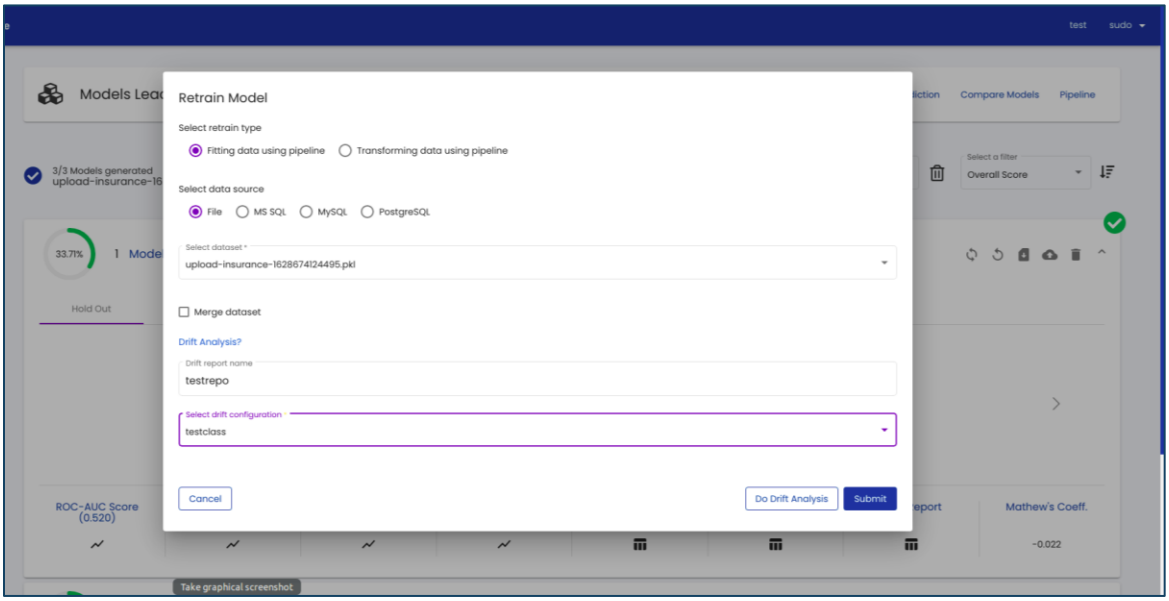
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

14.2 Generate drift report

Go to the models dashboard page and select any model. Click on retrain option it will open the retrain popup along with drift analysis option

Steps :-

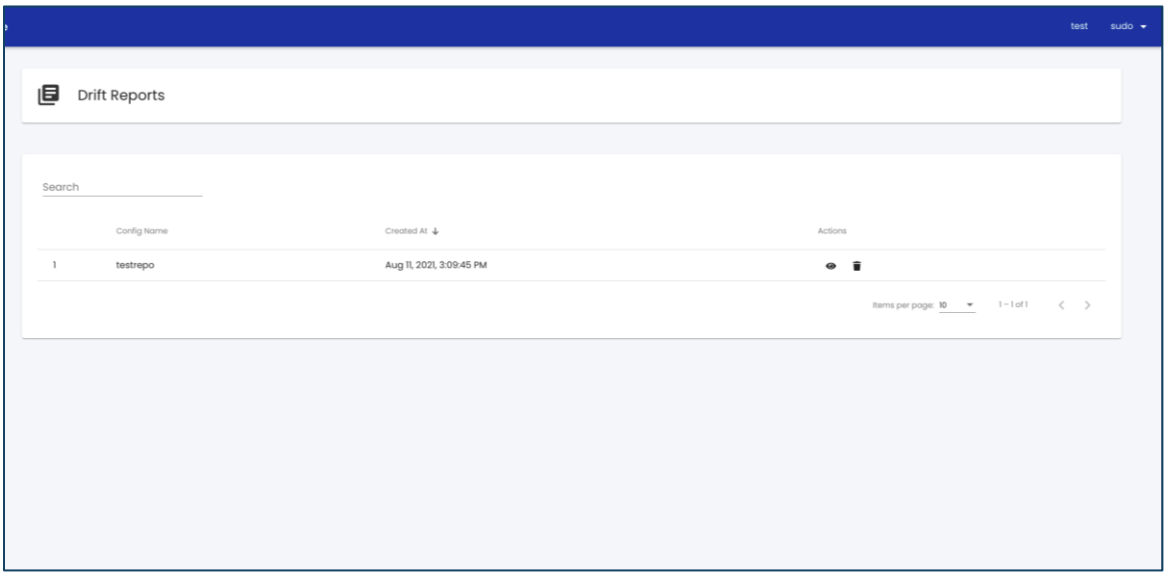
1. Click on retrain button
2. Click on Drift analysis option in retrain popup window
3. Enter the Report name
4. Select the drift configuration
5. click on Do Drift Analysis button



Retrain model

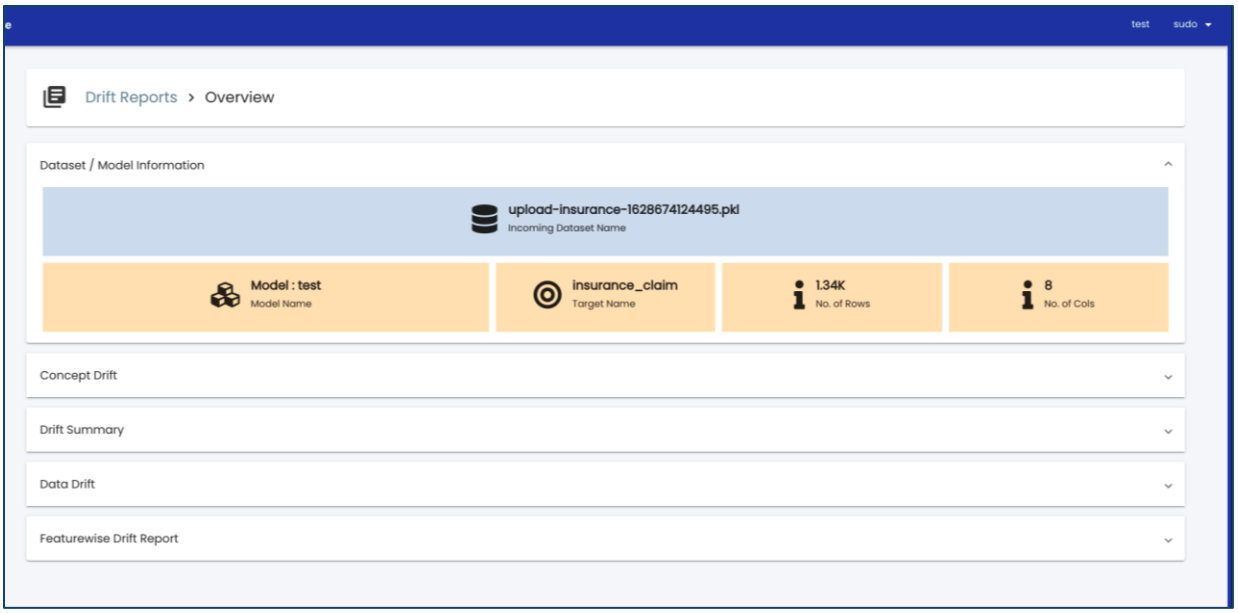
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

14.3 Drift report module



Drift Reports

You can see all the Generated reports in drift report module. Drift report shows the different performance metric like retraining dataset info, accuracy, % drift contribution, overall % rows have drifted. Also you can generate the feature wise drift report as well.



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

14.4 Feature wise drift report

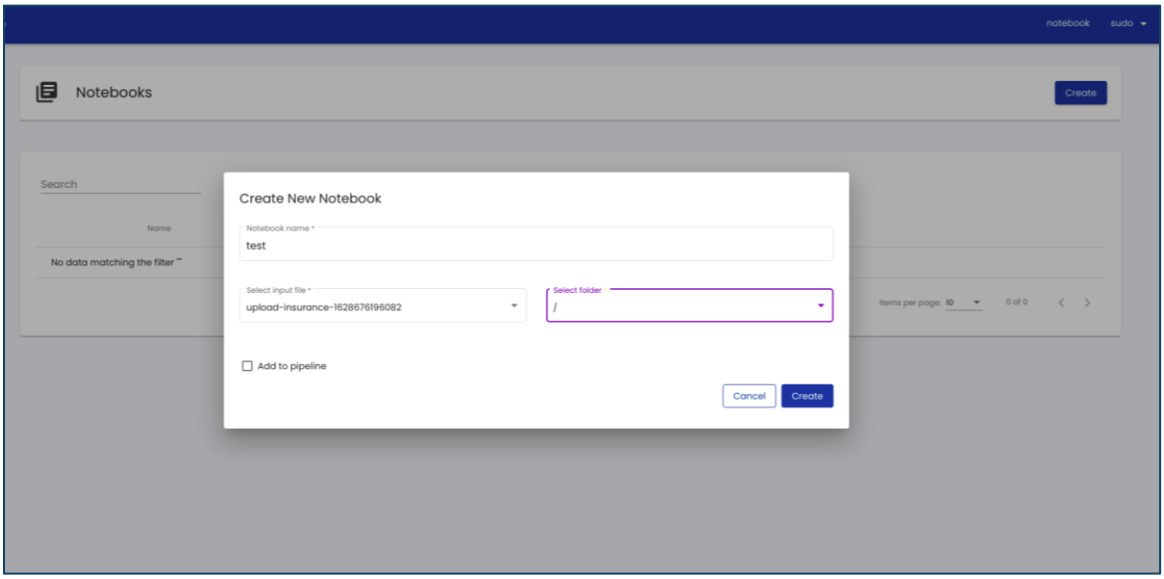
You can analyze drift for any feature in dataset by using the feature wise drift report functionality in PredictSense. Go to the Drift report model and select any feature from input feature dropdown. It will show the drift plot along with the baseline data stats and incoming data stats

Feature wise Drift Report

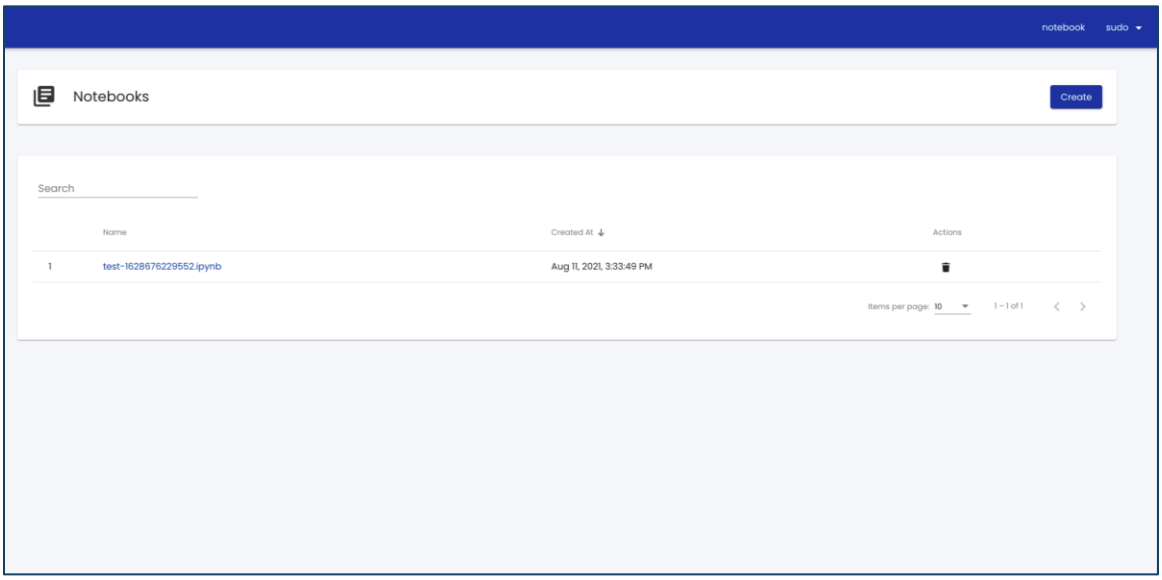
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

15 Notebook

The notebook section allows user to write custom code for analysis or data transformation.



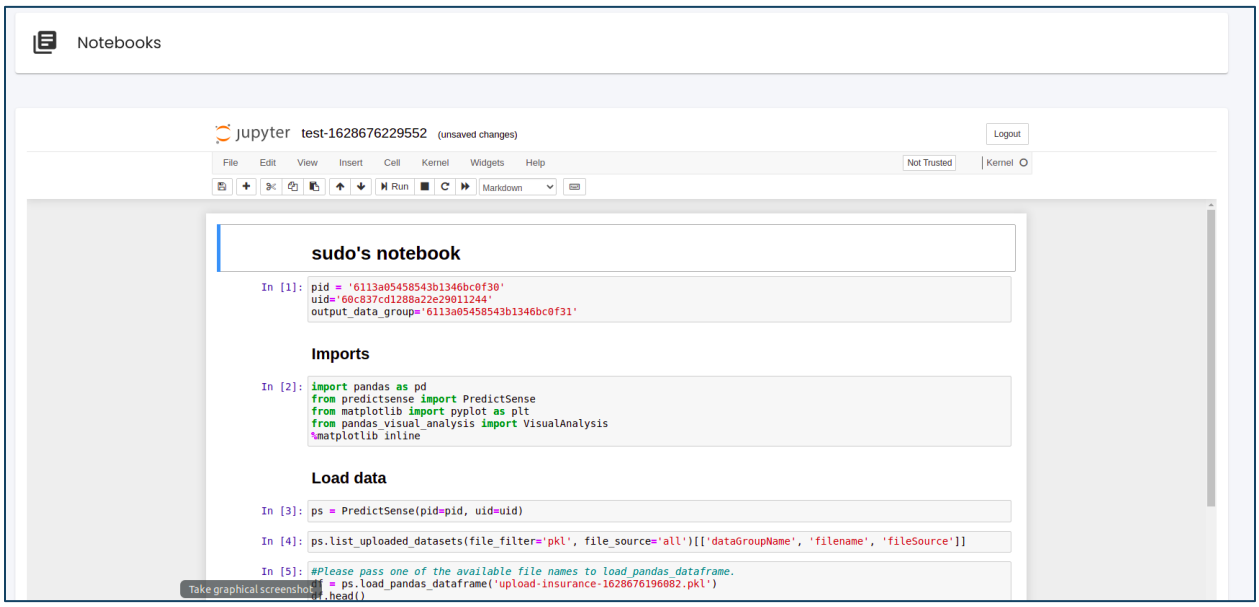
Notebook



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

15.1 Notebook functionalities

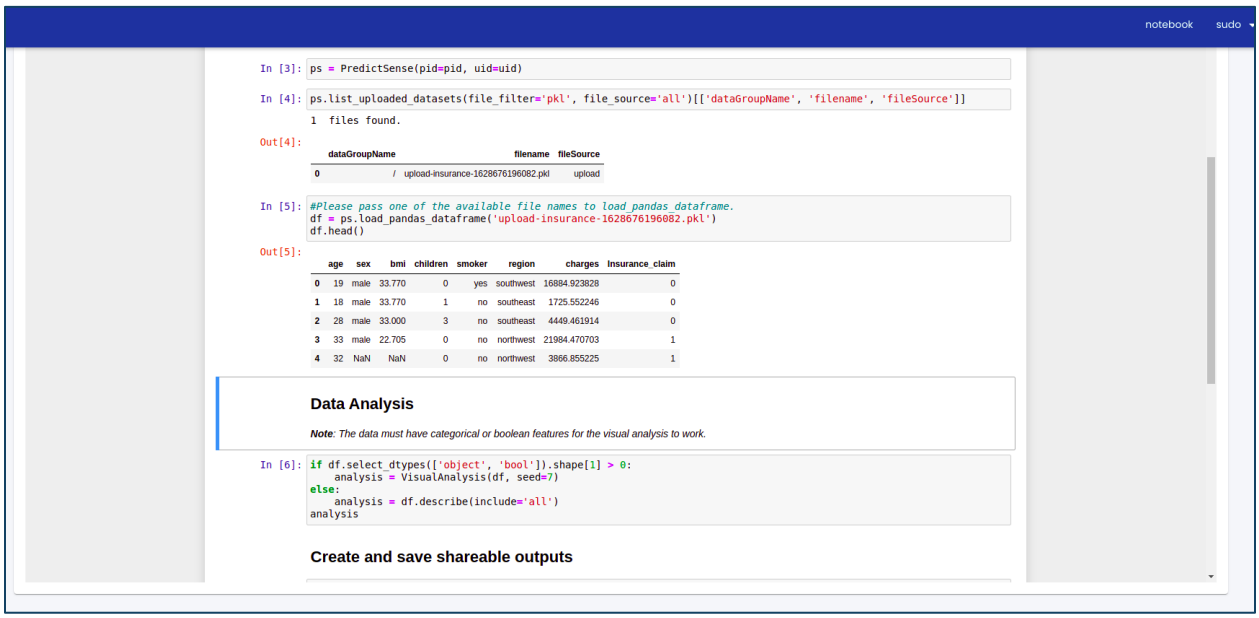
Notebook is equipped with a python-based utility to interact with the PredictSense system. And on creating a new notebook, some cells are populated with starter code for the user to start interacting with PredictSense.



Python code

15.2 Reading data

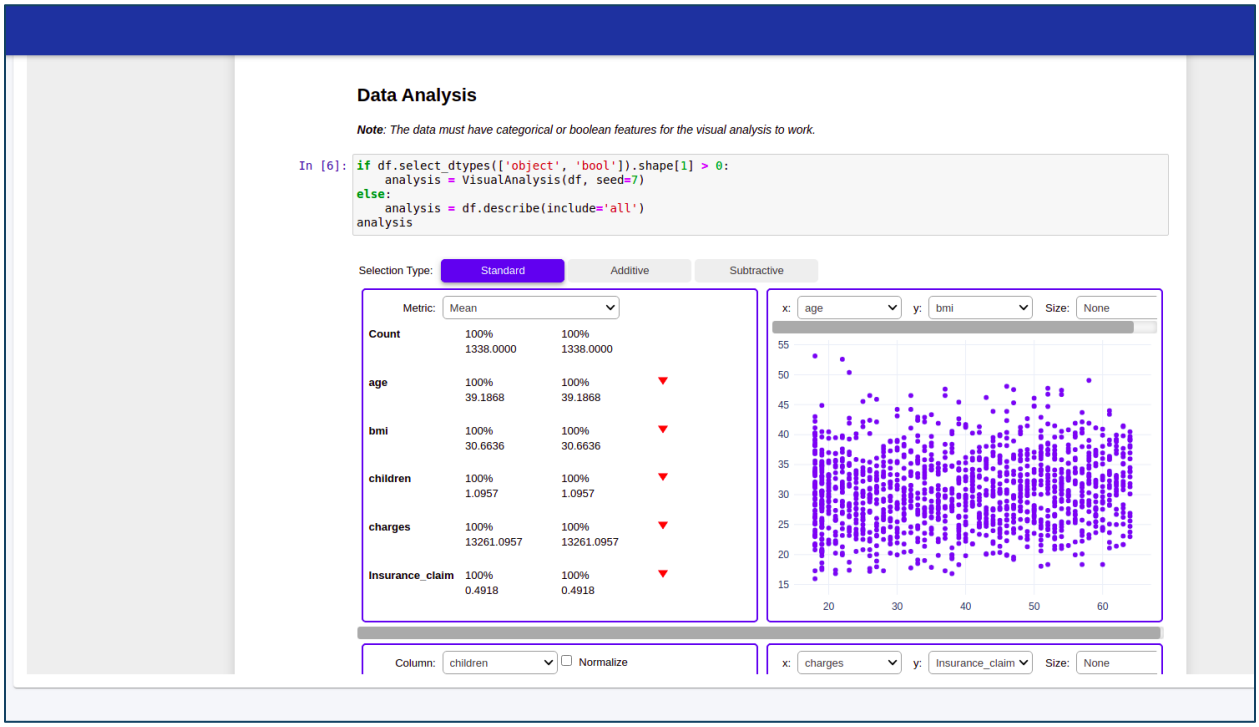
The user can use the **"list_uploaded_datasets"** API to list the available files. The **"load_pandas_dataframe"** helps to load the selected file.



Read data

15.3 Data analysis

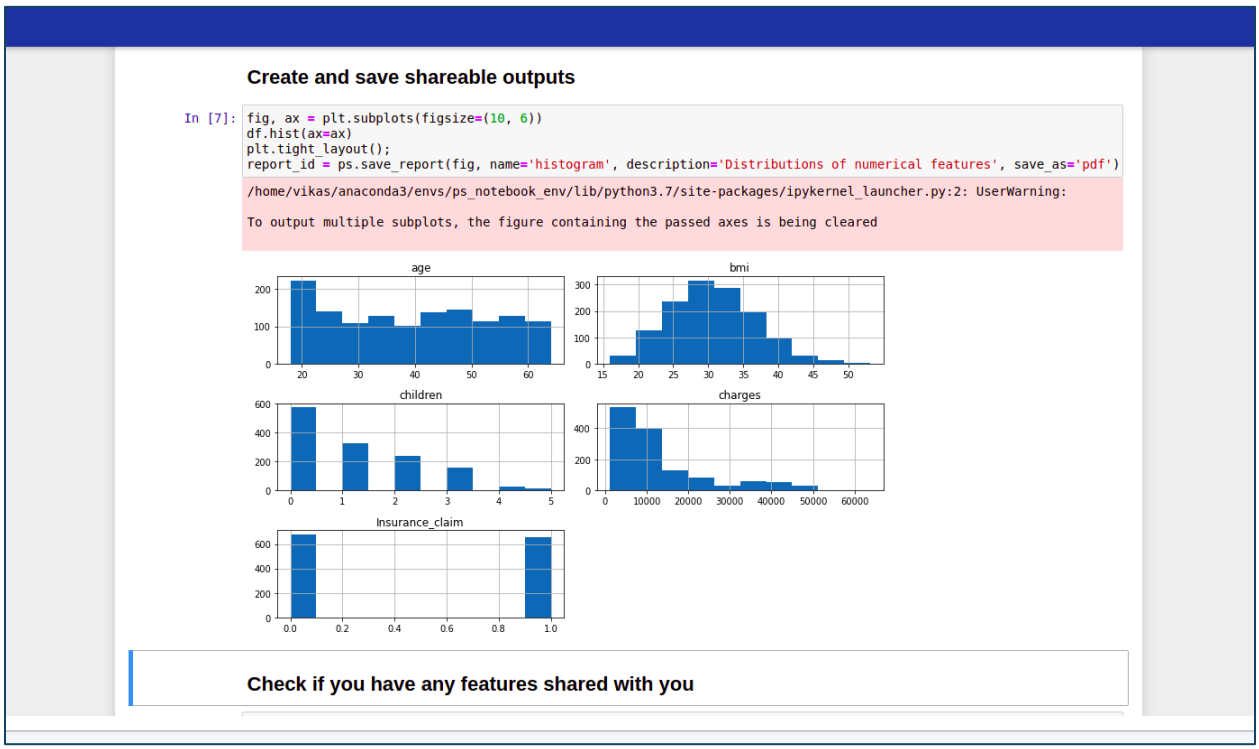
With the help of **"VisualAnalysis"** API, the user can also generate a set of interactive data analysis widgets to inspect and analyse the relationships in the data.



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

15.4 Create shareable reports

"save_report" API allows user to save plots or dataframes to the output module. These reports then can be shared with different users.



Create Share Report

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

15.5 Share functions

PredictSense's "**share_function"** utility allows user to collaborate with other users via **"feature sharing"** functionality. A function can be shared with other users by providing the function and the intended email address.

Share function with user

In [12]:
```
def compute_square(number):
    """
    Compute square of a give number.

    Parameters
    --------------
    number: int/float
        The number to compute the square of.

    Returns
    ---------------
    int/float
        Squared number

    Examples
    ----------------
    >>> squared_number = compute_square(number=5)
    >>> print(squared_number)
    25
    """
    return number ** 2
```

In [15]:
```
compute_square(5.0)
```

Out[15]: 25.0

In [16]:
```
ps.share_function(
    function_object=compute_square, description="Compute square of a given number",
    email_addy="admin@winjit.com")
```

Out[16]: 'Function shared successfully.'

Share with user

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

15.6 Load shared function

A shared feature can be utilized by loading it in memory with **"get_shared_function"** API.

Check if you have any features shared with you

In [17]:
```
shared_feature_list = ps.list_shared_features()
if len(shared_feature_list) > 0:
    print ('Top 5 shared features')
    print (shared_feature_list.head())
```

```
Top 5 shared features
                      createdAt            name      type sharedFeaturesList  \
0  Thu, 08 Oct 2020 10:44:51 GMT  compute_square  function                []

  sharedIndex            email
0          []  ver3@winjit.com
```

Load one of the shared features

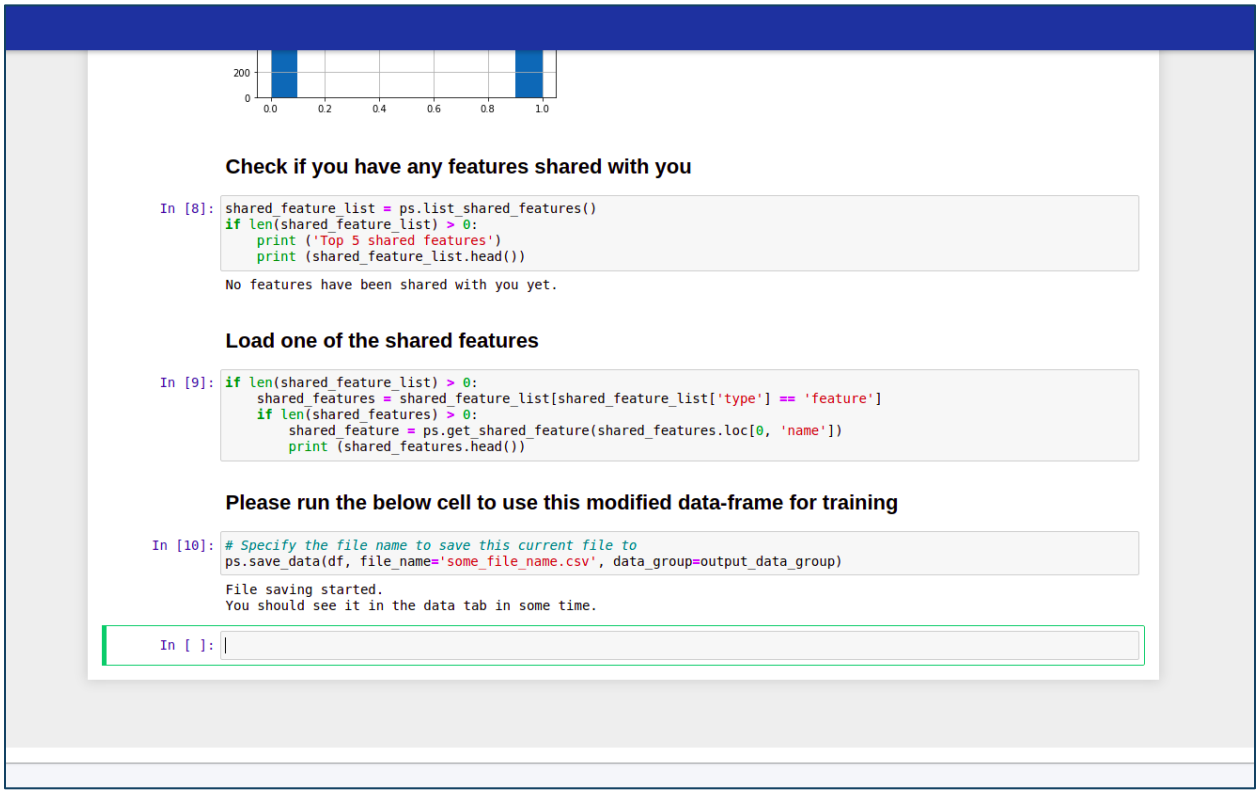
In [19]:
```
func = ps.get_shared_function('compute_square')
```

```
Function loaded successfully.

def compute_square(number):
    """
    Compute square of a give number.

    Parameters
    --------------
    number: int/float
        The number to compute the square of.

    Returns
    ---------------
    int/float
        Squared number

    Examples
    ----------------
    >>> squared_number = compute_square(number=5)
```

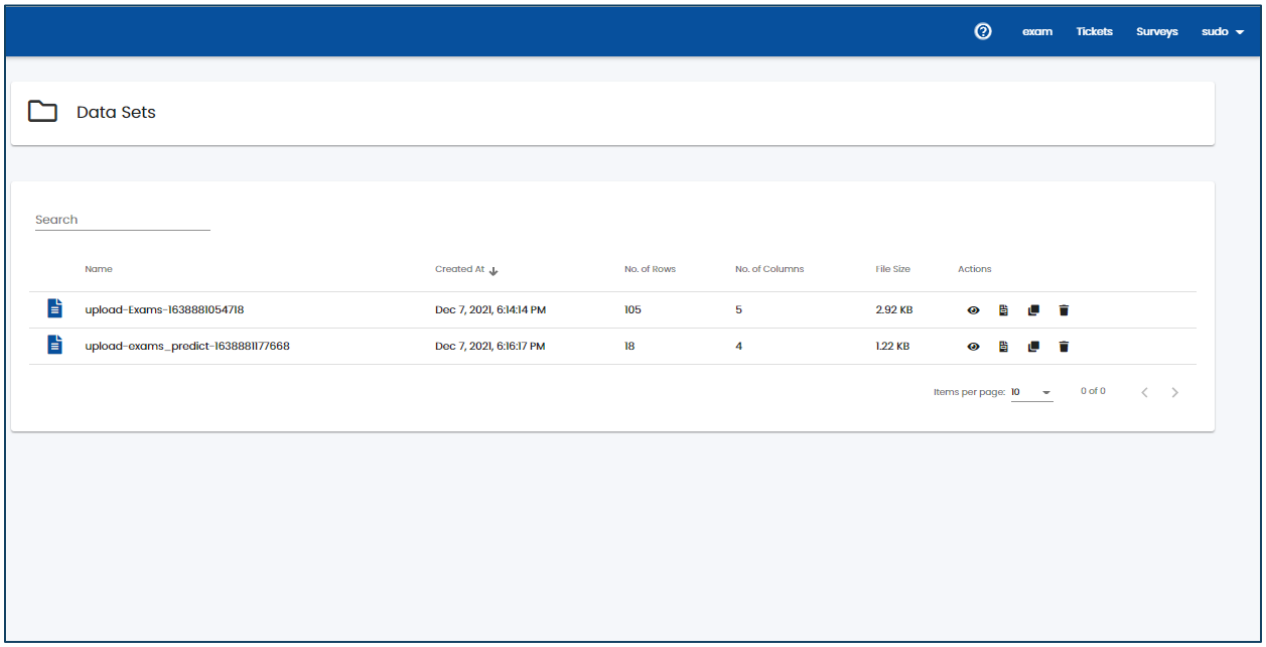
15.7 Save data

Once the custom data transformation is done, the user can use this engineered data to use for further processing in PredictSense by calling the **"save_data"** API. This will register the newly created data to the data pages and hence allow user to perform EDA and training.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



Save data



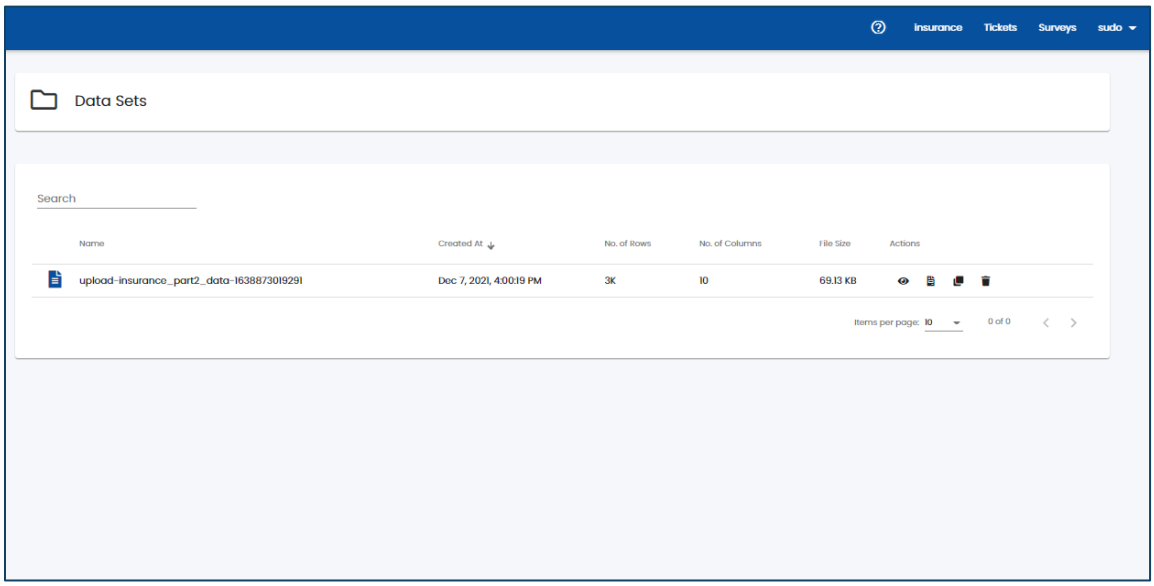
Data sets

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

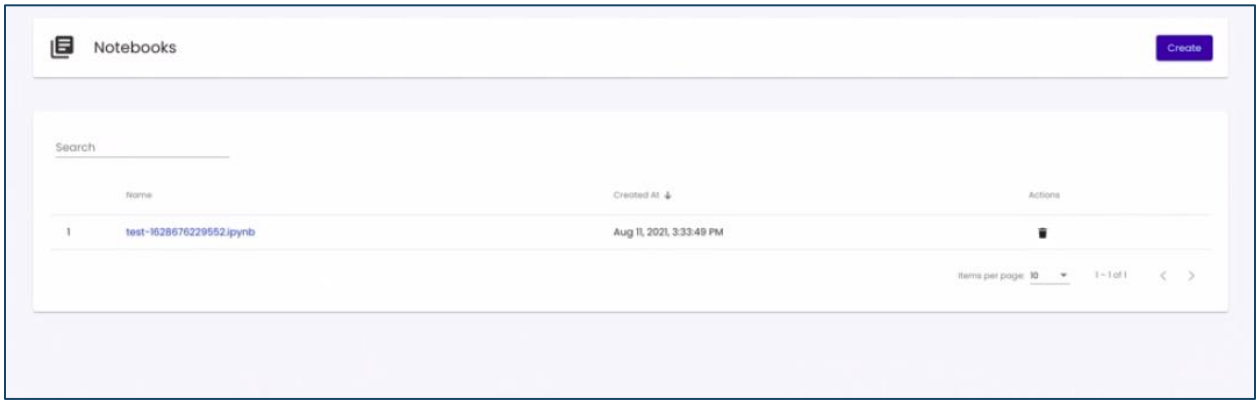
16 Report

In output module user can see all the shared and saved report.

a) You can create and save shareable report in output model. **"save_report"** API allows user to save plots or dataframes to the output module. These reports then can be shared with different users. First you need to create notebook and execute the **"create and save shareable output"** cell.

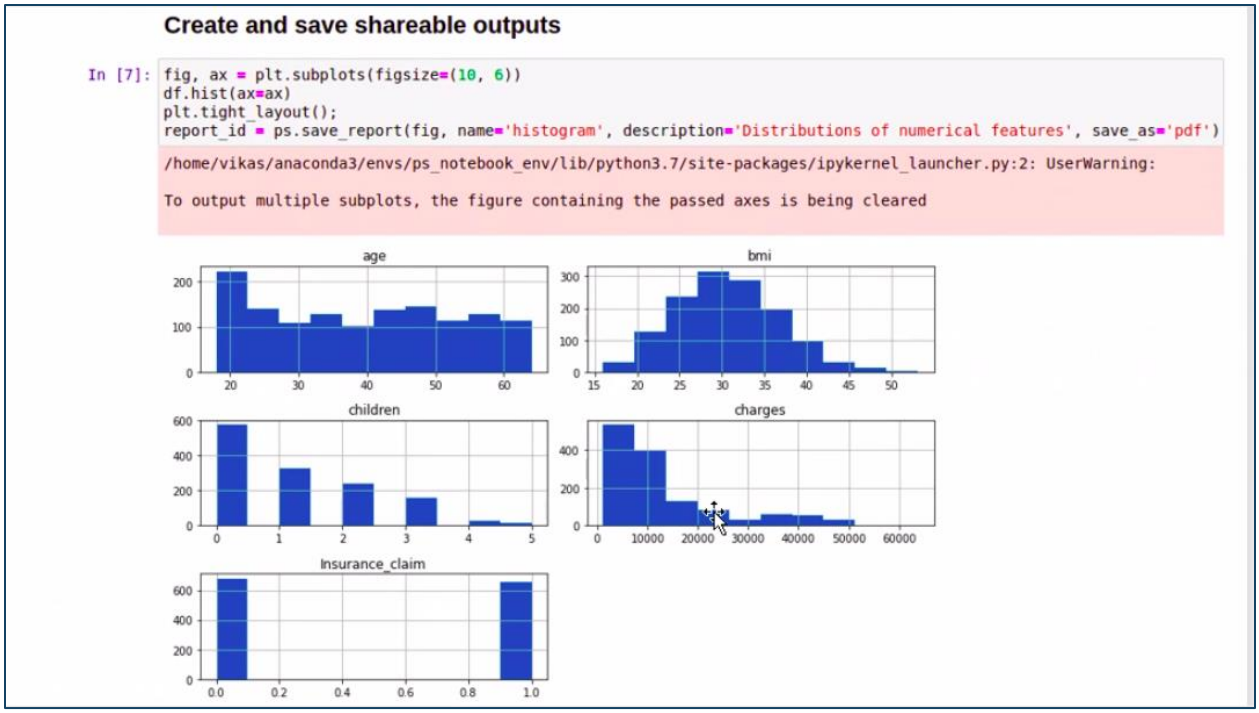


Upload data set

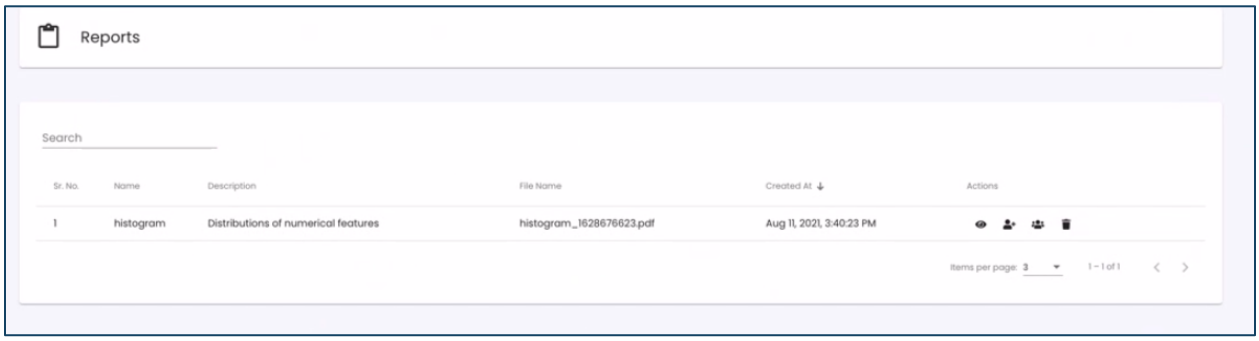


Create notebook

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



Go to the report model it will show the created and saved file.



Reports

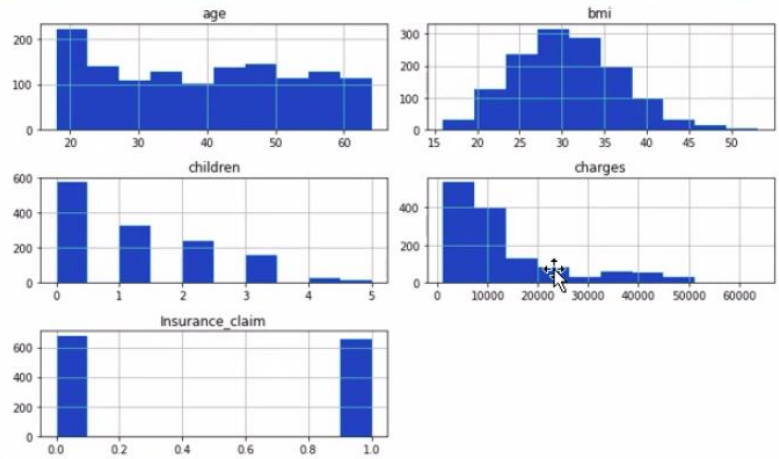
b) You can view the report by clicking on preview button please refer the below screenshot

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

Create and save shareable outputs

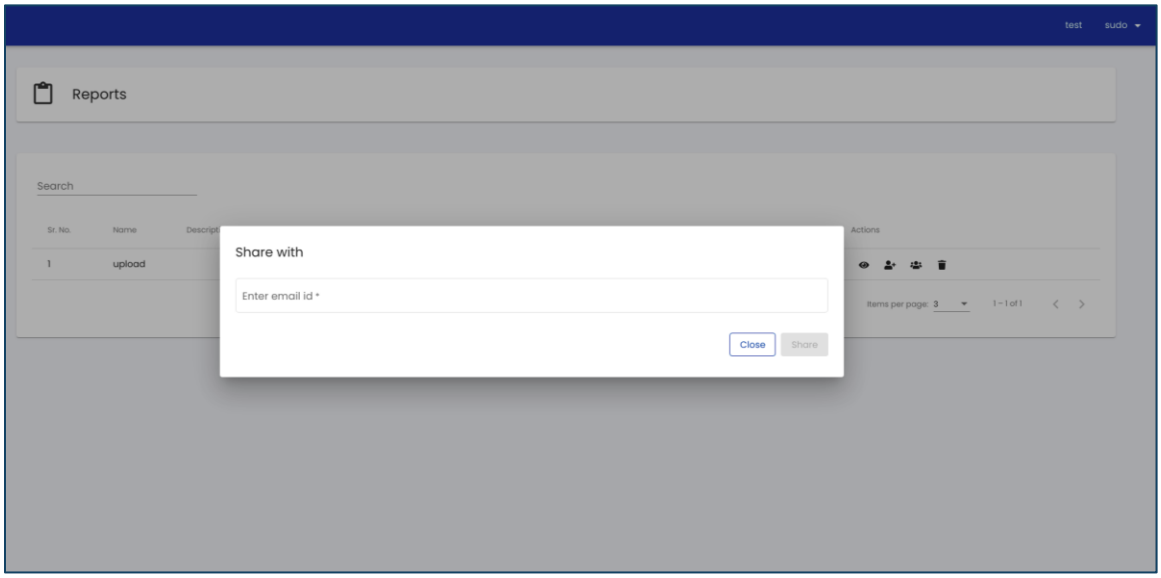
In [7]:
```
fig, ax = plt.subplots(figsize=(10, 6))
df.hist(ax=ax)
plt.tight_layout();
report_id = ps.save_report(fig, name='histogram', description='Distributions of numerical features', save_as='pdf')
```

```
/home/vikas/anaconda3/envs/ps_notebook_env/lib/python3.7/site-packages/ipykernel_launcher.py:2: UserWarning:

To output multiple subplots, the figure containing the passed axes is being cleared
```

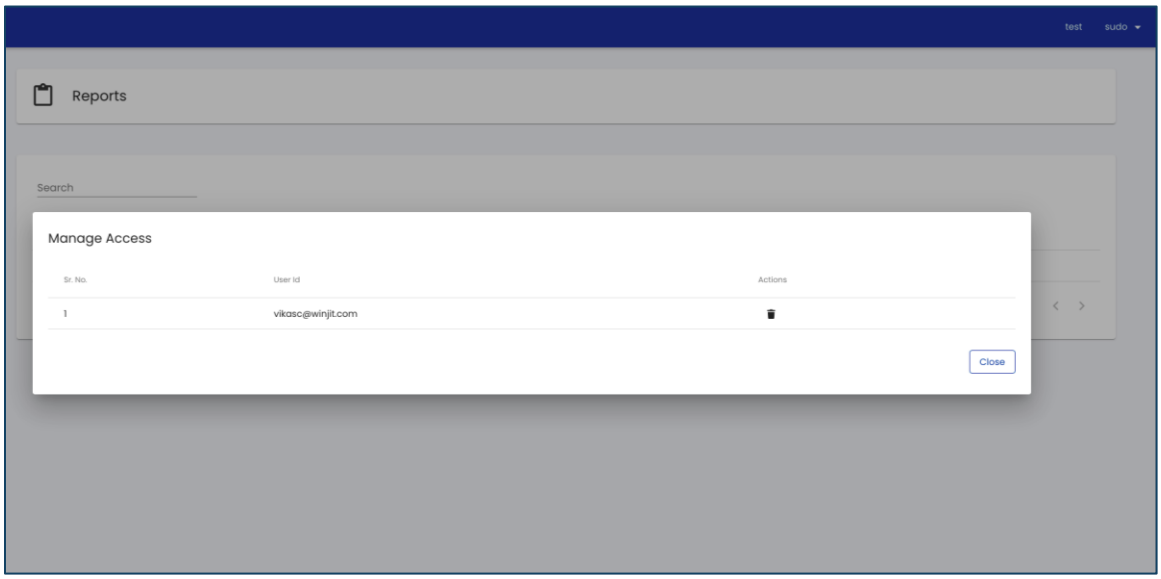


c) You can share the report with other user by clicking on manage access button it will open the popup screen were you can pass the email Id of user and click on share button.

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.



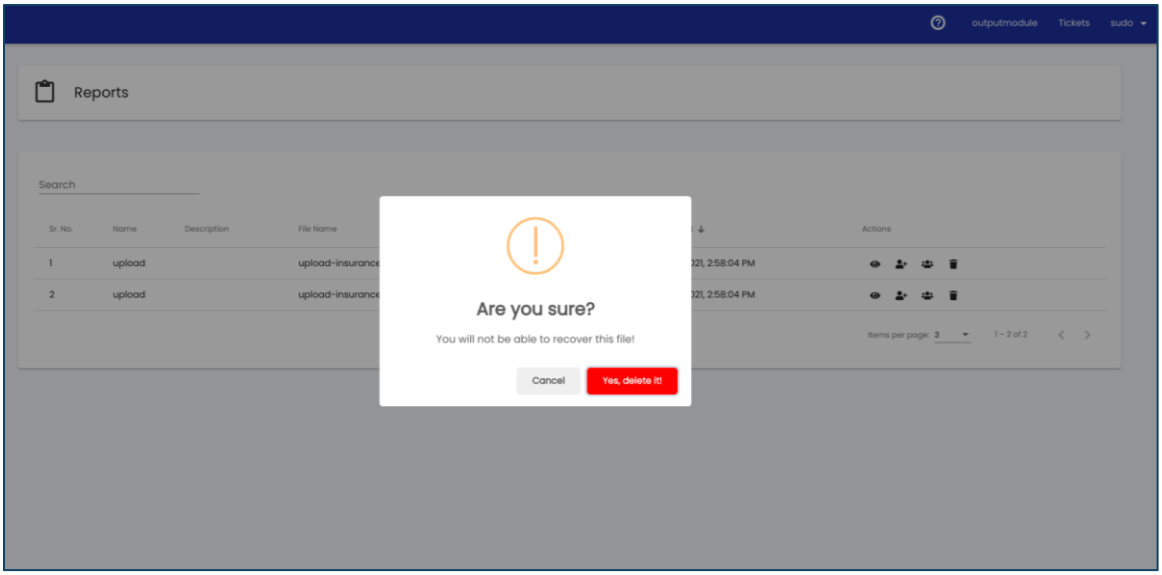
Share report



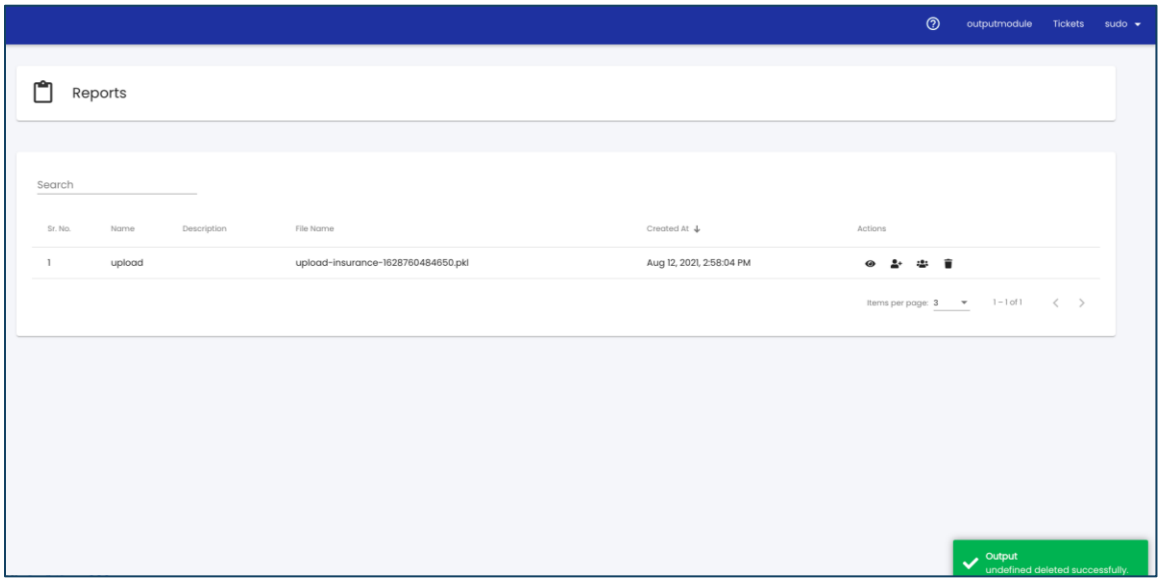
Manage Access

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

e) You can delete the reports from this module list by clicking on delete button.



Delete the Report

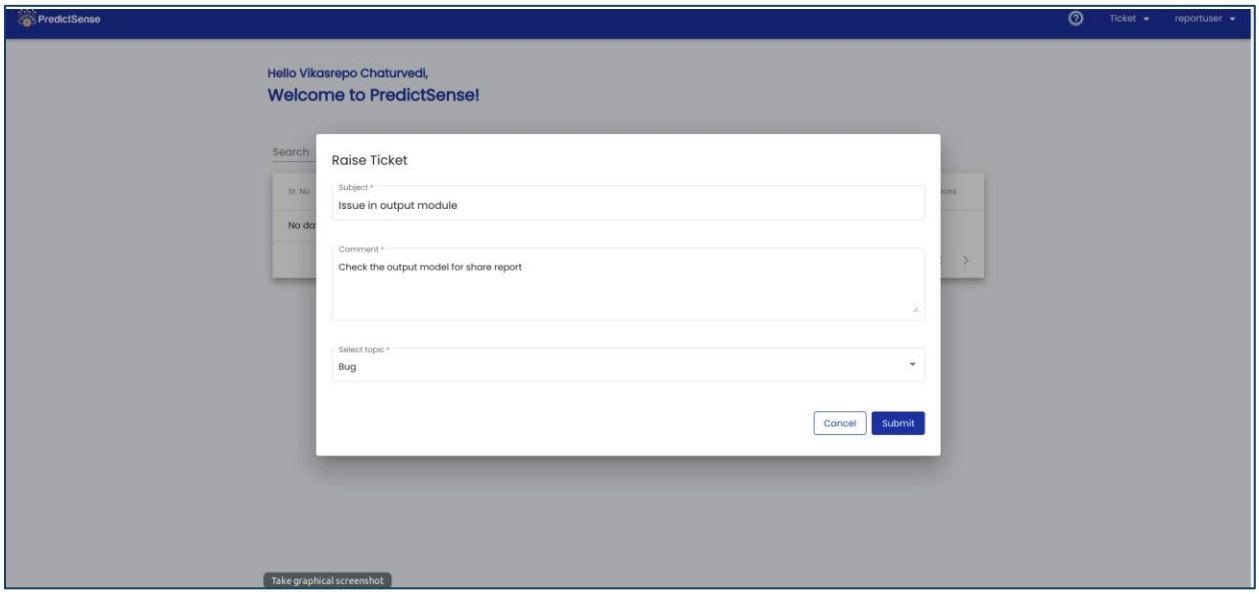


After delete

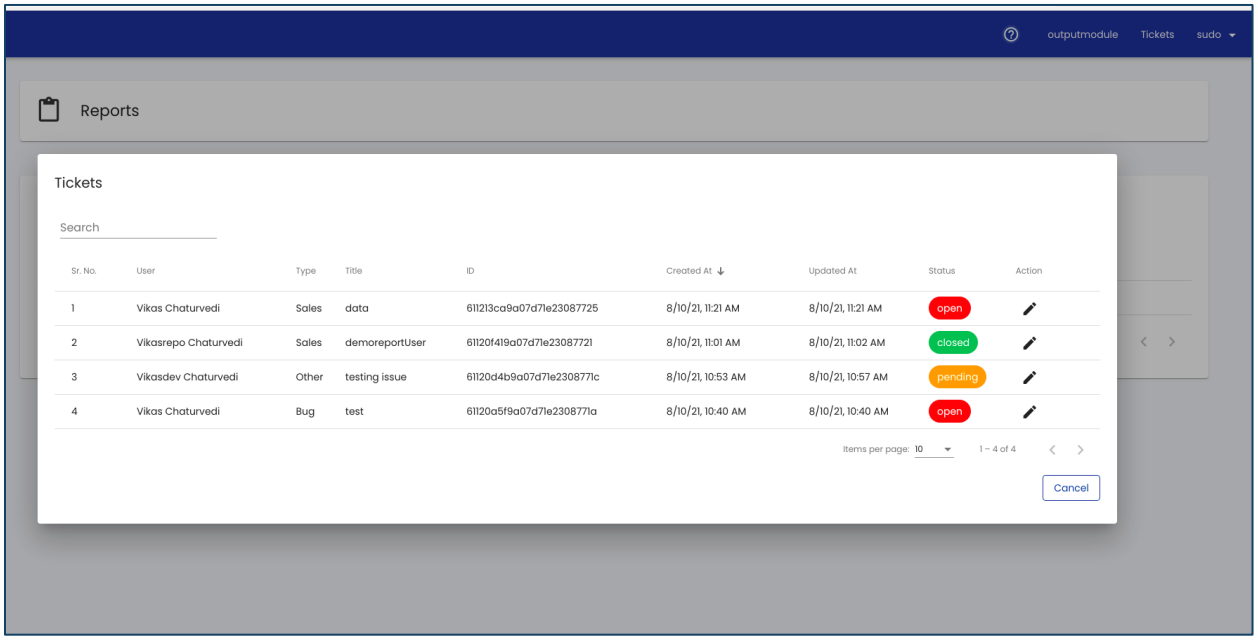
© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.

17 Ticket, Guided tour, Feedback and Profile

In application allow user to add ticket for bug, new enhancement and feedback etc. Once it's added user receive mail o his registration email and once resolved same intimated by mail. In view option user able to see current status for tickets.



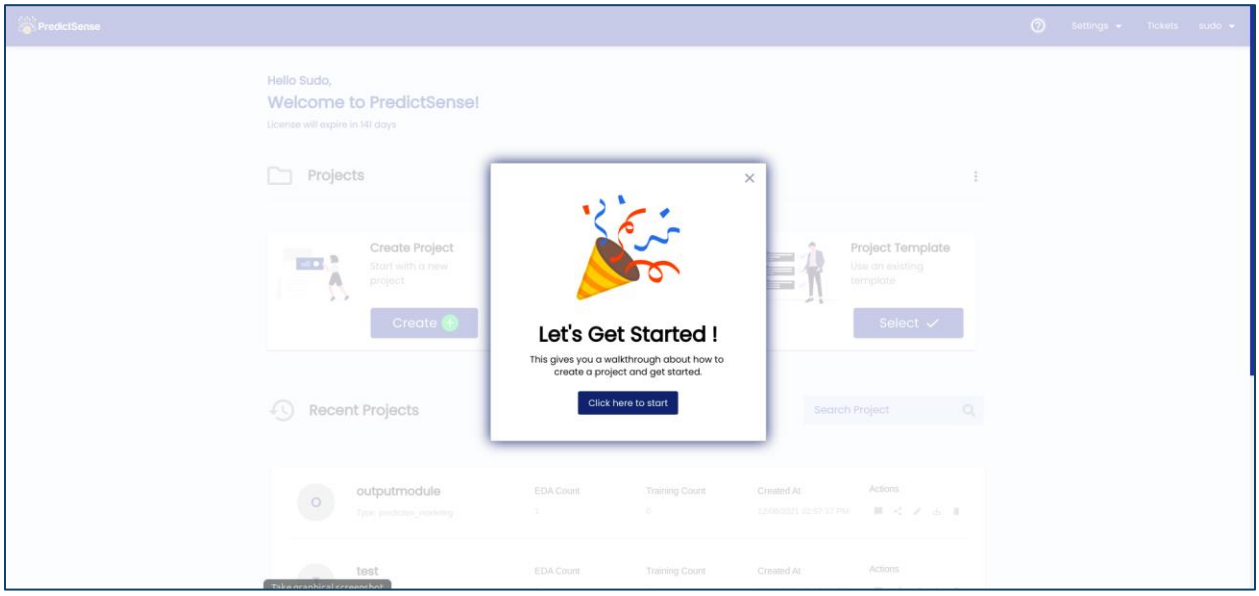
Raise Ticket



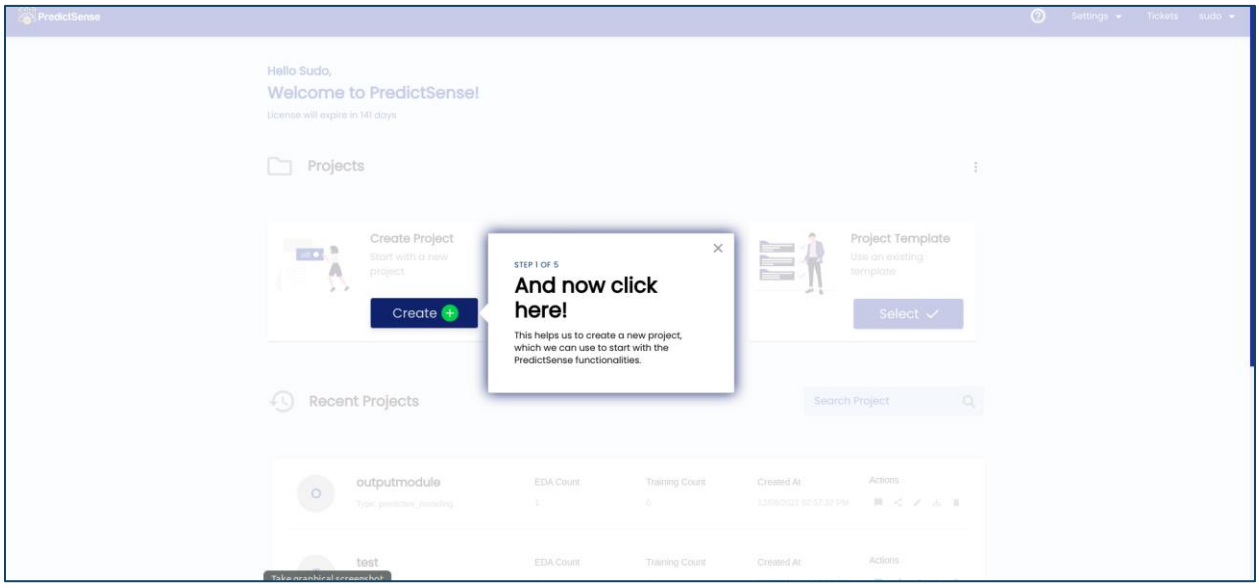
View Ticket

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

Guided tour helps user to understand application flow and help if user get stuck at any point using ? Help option.



Guided Tour

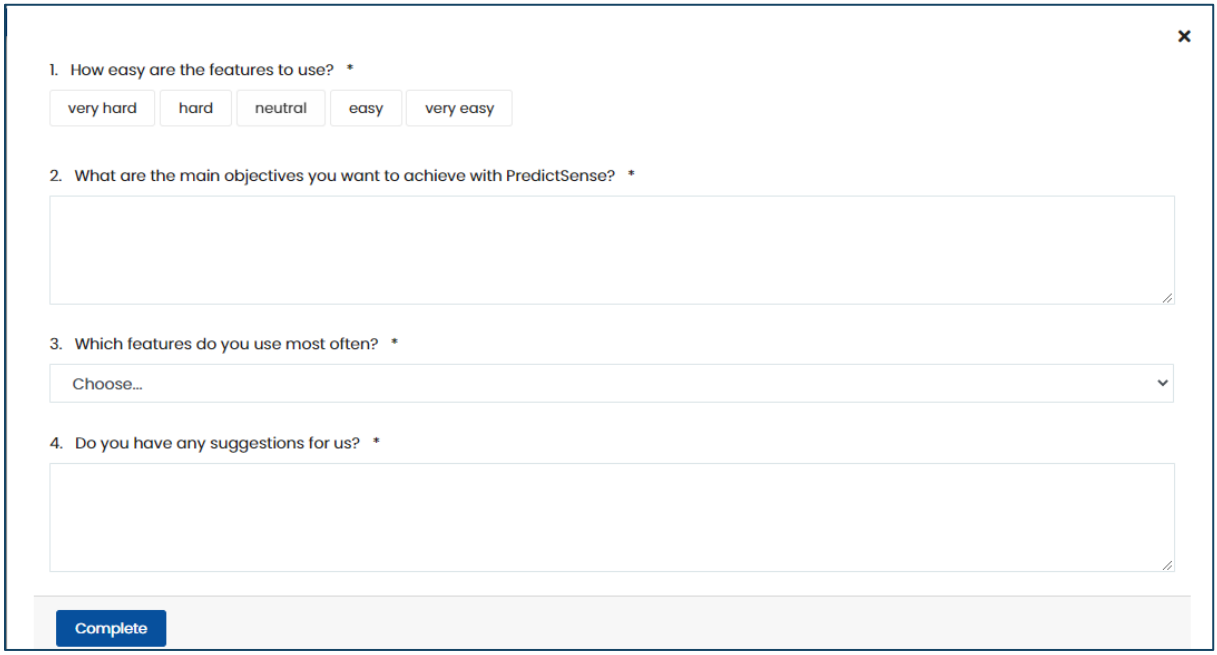


Guided Tour

© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|

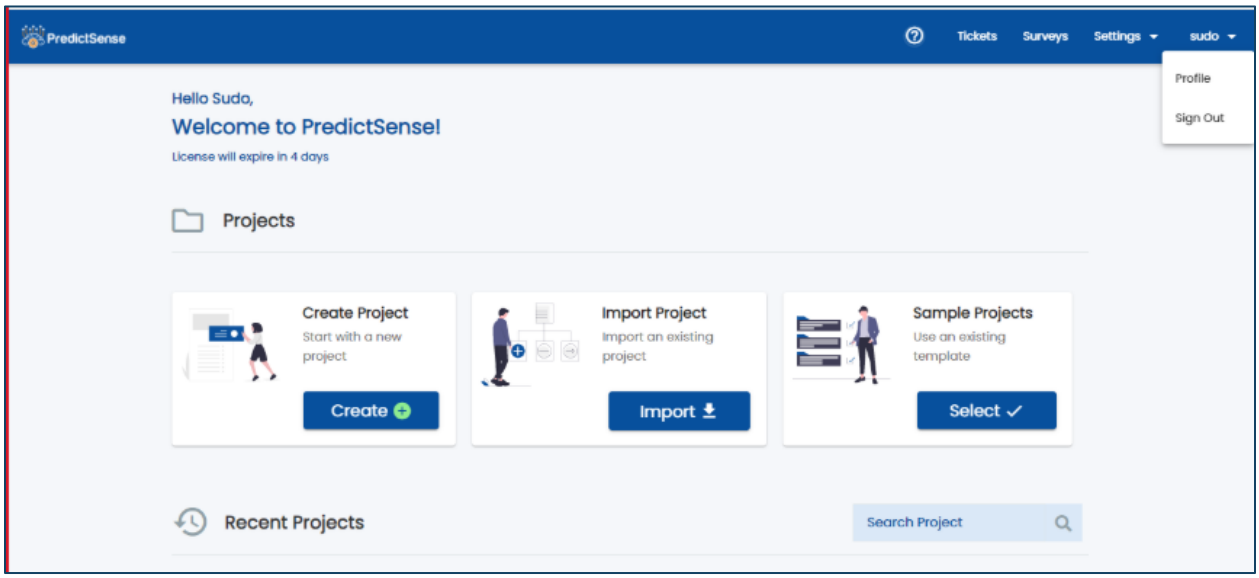
Feedback –

Application allow user to submit feedback and his overall experience about PredictSense by answering few questionnaires'.

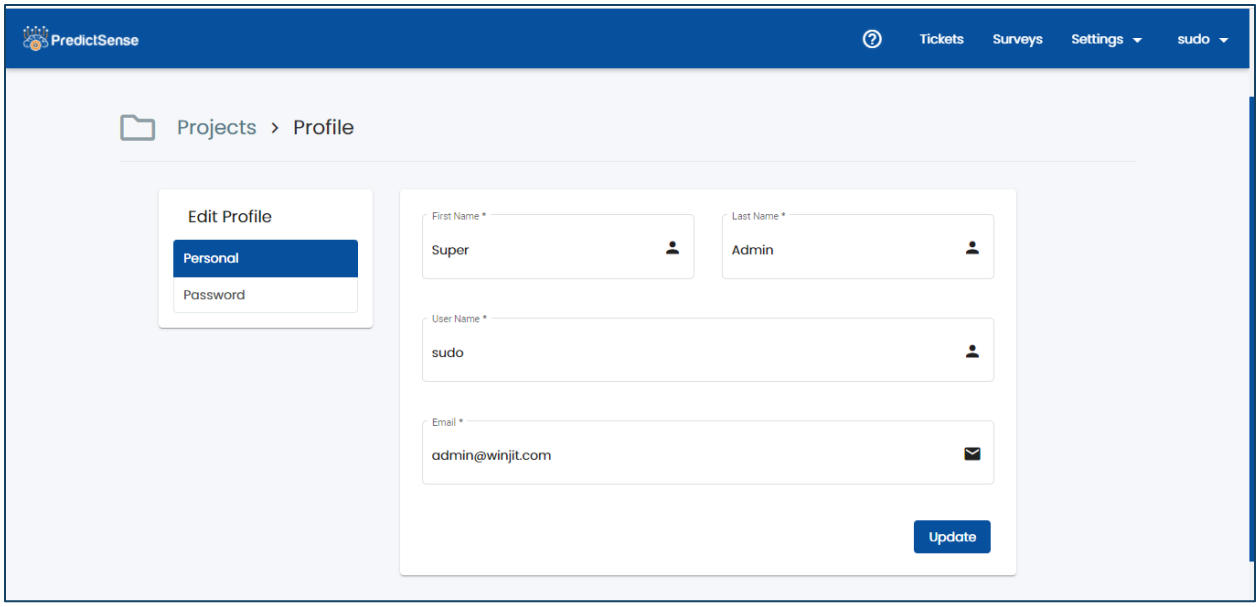


Profile-

Application allow user to update profile by changing user infomation and password.



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|



© 2022, All rights reserved at Winjit Technologies Pvt. Ltd.|