FINDABILITY PLATFORM

FPPredict+

FPPredict And FPForecast User Manual

May 2020

Contents

Table of Figures

Table of Tables

Version History

Version	Date	Nature of Amendment
1.0	7th March, 2018	First Version
2.0	24th September, 2018	Added Description of the Report. Added Forecast Job Details.
2.1	25th November, 2018	Added Support section. Updates as per Nov 2018 release.
2.2	28th December, 2018	Updates as per Dec 2018 release. Added Appendix for binary target prediction. Overall general changes in the content.
2.3	28th January, 2019	Added section for understanding of the downloaded report. Added section for dataset management. Added updates to categorical target variable job results for FPPredict. Added Updates for Important variables in FPForecast jobs.
2.3.1	15th February, 2019	Text in section 2.6.2.5 Metrics modified
2.3.2	21st February, 2019	Text in Nature of Amendment for Version 2.3 in version history modified. Page numbers in continuity. Section 2.9 modified
2.4	20th June, 2019	Added sections for Multimodel Forecast, Continuous Target Variable processing, Advanced Settings for FPPredict, zero models generated view, and Floating add job.
2.5	7th August,2019	Removed Multimodel forecast without actuals from dashboard
3.0	23rd October,2019	Added Report changes in FPForecast and FPPredict, Pagination in job summary, Modeling and Forecast Metrics for FPForecast, Modeling and Forecast metrics for FPPredict-CTV, Models view for Forecast, variables importance for CTV and forecast without groups, Default value of Job Description.
3.1	18th March, 2020	Added changes related to Login, Change password, Instance Registration, License validation and API key generation.

Definitions, Abbreviations and Acronyms

SL	Term	Description
1	FP	Findability Platform
2	FS	Findability Sciences
3	CSV	Comma Separated Values
4	Training File	A CSV file containing the historical dataset required to train the system and create the models.
5	Target Variable	The variable present in the Training file for which the system is being trained.
6	Prediction File	A CSV file containing the dataset where the values of target variables are to be predicted. Values of the target variable may or may not be known beforehand.
7	Id variable	The variable which is used to uniquely identify the records in the dataset.
8	Models file	System generated models file, which may be used to run prediction only jobs. This file is generated for Modeling only and Modeling and Prediction Jobs.
9	Decile	10^{th} part of the Results
10	Hits Percent	Percentage of hits for a given decile
11	Random Probability	Probability of finding the hits in the whole data set
12	Lift	Hits Percent / Random Probability
13	Outlier	An outlier is an observation that lies outside the overall pattern of a distribution
14	Binary target variable	Only two unique values present in the target variable.
15	Discrete target variable Case (DTV)	Case where the predicted value must be one of the values present in target variable in training file.
16	Continuous Target variable Case (CTV)	Case where the predicted value can be beyond the values present in target variable in training file.
17	License Details	Credentials required for the users to register the instance.
18	EULA	End User License Agreement

1. Overview

1.1 Intended Audience

This document is intended to cover the functionalities and features of the Findability Platform FPPredict+. The audience for this is the end users of the system.

1.2 Scope

The User Manual contains all essential information for the user to make full use of the information system. This manual includes a description of the system functions and capabilities, and procedures for system access and use.

2. System Functionalities

2.1 Login

Users may access the system only after successful authentication. The credentials for authentication will be provided by the FS representative.

For Redhat marketplace Users and ICP4D, Please use default credentials to Login

- Username: admin@findabiitysciences.com
- Password: fppr3dictpLu5

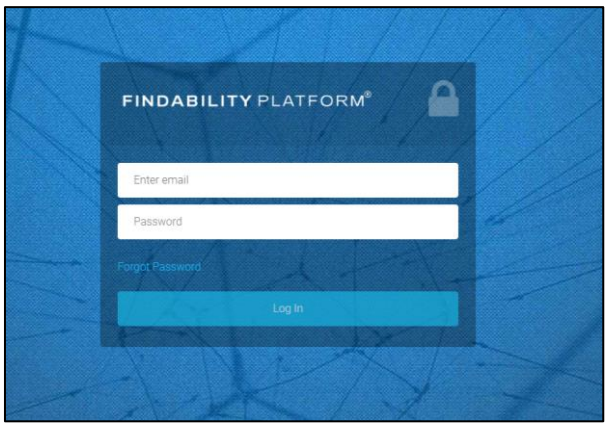


Figure 1 - Login Page

2.2 Forgot Password

In the event users don't remember their password, they can use forgot password functionality.

On Clicking "Forgot Password" link, users are provided with a form to enter their "Email Address", using which they login into the system.

This functionality is currently not available for ICP4D and Redhat Marketplace Users.

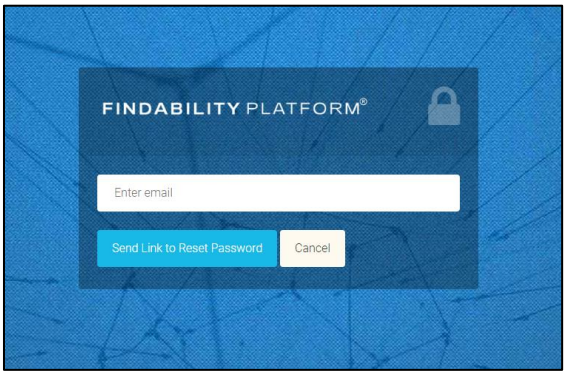


Figure 2 - Forgot Password

Users would then get an email with the link to reset password, along with a unique code, which can then be used to reset the password. The link and code remain valid for 4 hours.

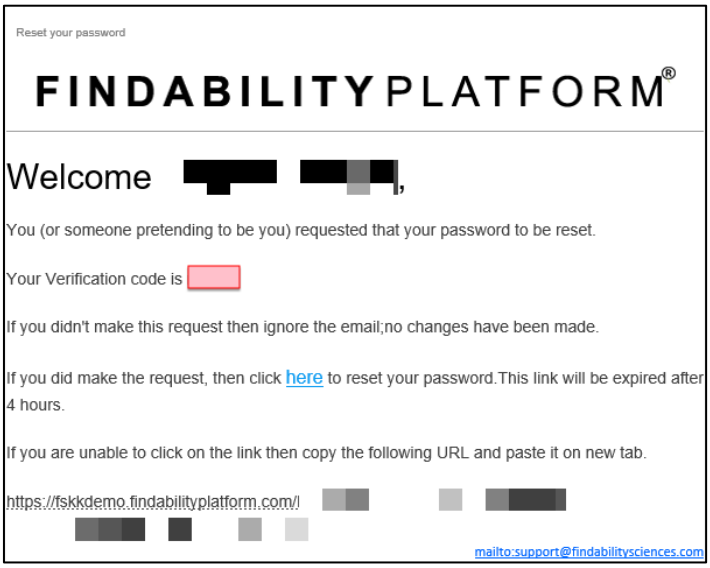


Figure 3 - Forgot Password Email

On clicking on the link provided in the mail, following page will open.

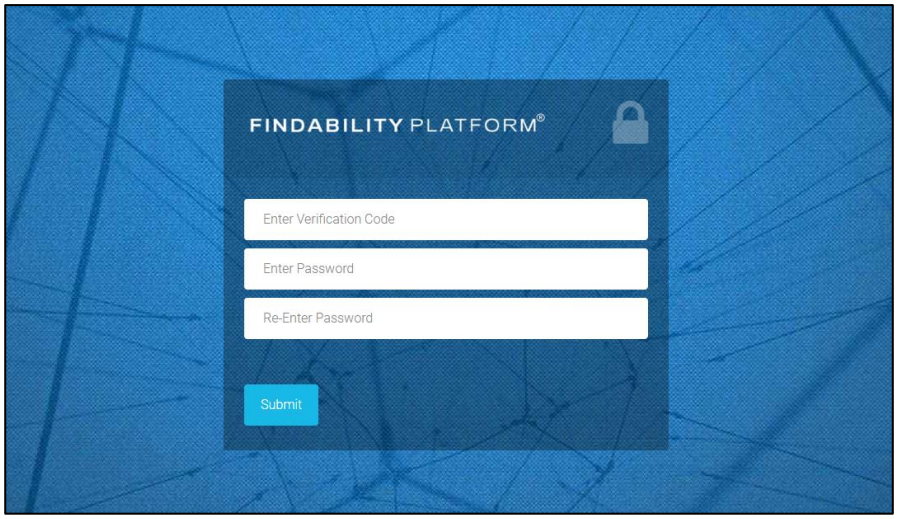


Figure 4 - Reset Password Form

User is supposed to enter the verification code provided in the mail along with new password, re-enter the new password and submit the form. On successful submission, the reset password process is complete, and users can use new credentials to log into the system.

2.3 EULA

On first Login and on Every License Agreement update, Users will be asked to agree to Terms and Conditions. Users will be able to submit jobs only after agreeing the Terms and conditions.

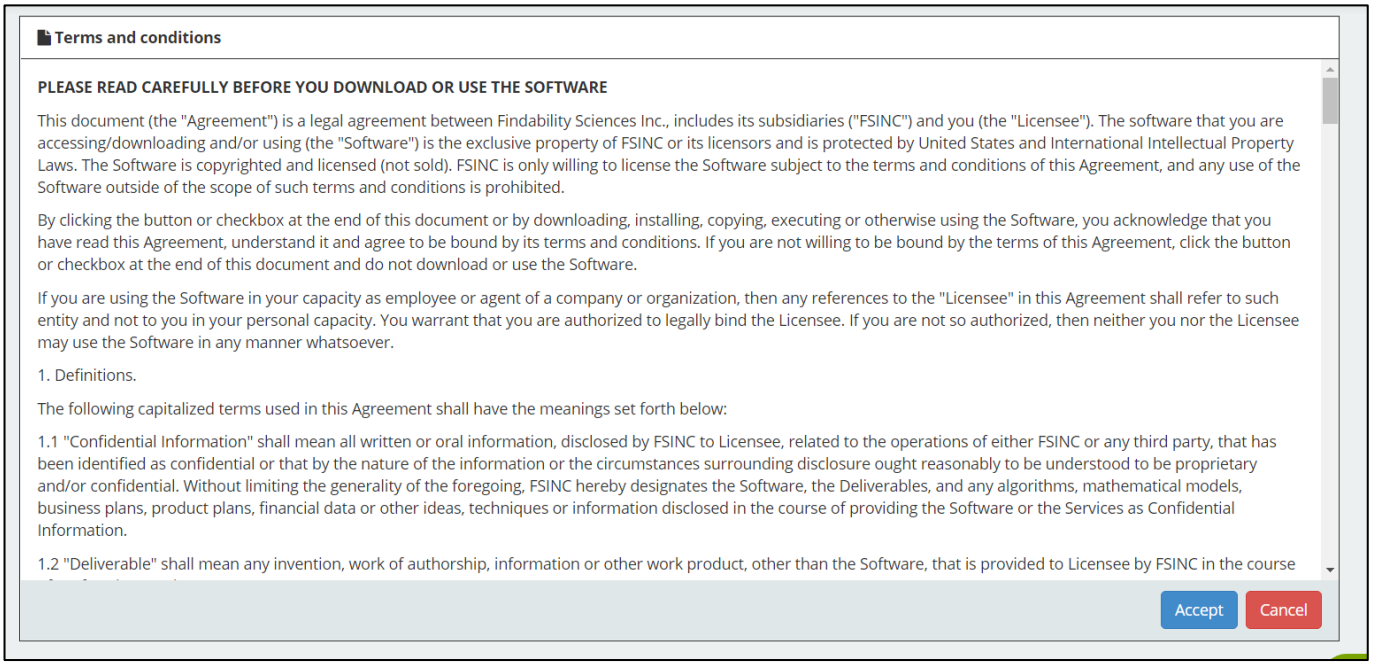


Figure 5 - Terms and Conditions

2.4 Language Settings

Users will be able to change the UI language as per preference. System currently provides two language options:

1. English
2. Japanese

Figure 6 - Language Options

2.5 Change Password

After successful login, users will be able to Change Password, by clicking on “Change Password” Option from the header.

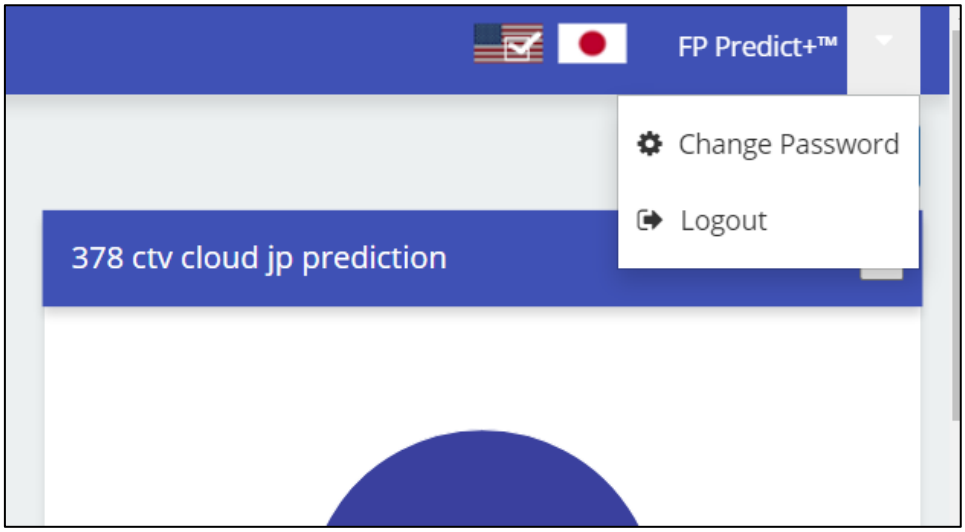


Figure 7 - Change Password Option

After clicking on change password option, system will produce following dialog box.

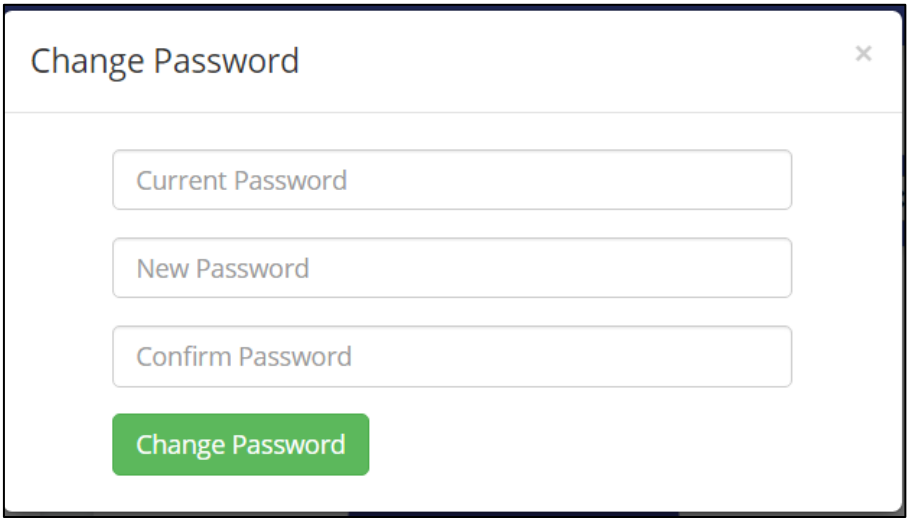


Figure 8 - Forgot Password Dialog

Here after providing current password, New password and Confirming new password and clicking change password, users will be to change their account's password.

2.6 Modules Navigation

Users will be able to navigate to different pages of the application using the navigation panel present in the side bar. The modules available in the system are as follows:

1. Dashboard
2. Analytics
3. Dataset Management
4. License Information

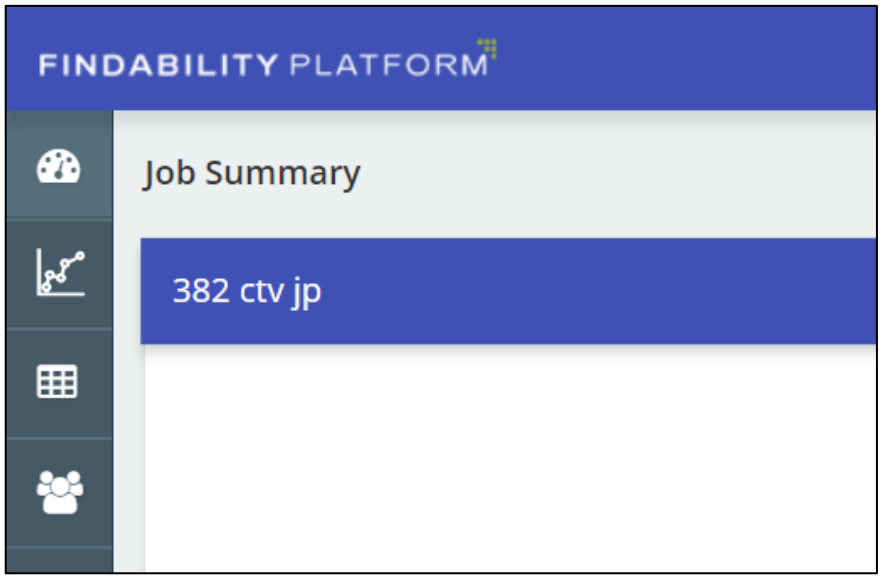


Figure 9 - Modules Navigation

2.7 Instance Registration

As soon as you Login in the system, system will prompt the Users to register the system by in a two step process:

1. Download System Information and share it with support.
2. Upload License file shared by support.

Figure 10 – Register Instance

In case of any issues please contact support@findabilitysciences.zendesk.com

On successful registration, system will redirect to License information page.

2.8 Dashboard

2.8.1 Job Summary

This page provides the summary of processed Jobs. Based on the job type, the results can be one of the following:

2.5.1.1 Modeling and Prediction, and Prediction Only Jobs

The successfully complete jobs show the Probability distribution of the prediction results in a pie chart in case system processes them as DTV job. The distribution has four categories:

a. **90% - 100%:** Count of records predicted with probability greater than 90% and Less than or equal to 100%.
b. **80% - 90%:** Count of records predicted with probability greater than 80% and Less than or equal to 90%.
c. **70% - 80%:** Count of records predicted with probability greater than 70% and Less than or equal to 80%.
d. **Others:** Count of records predicted with probability less than or equal to 70%.

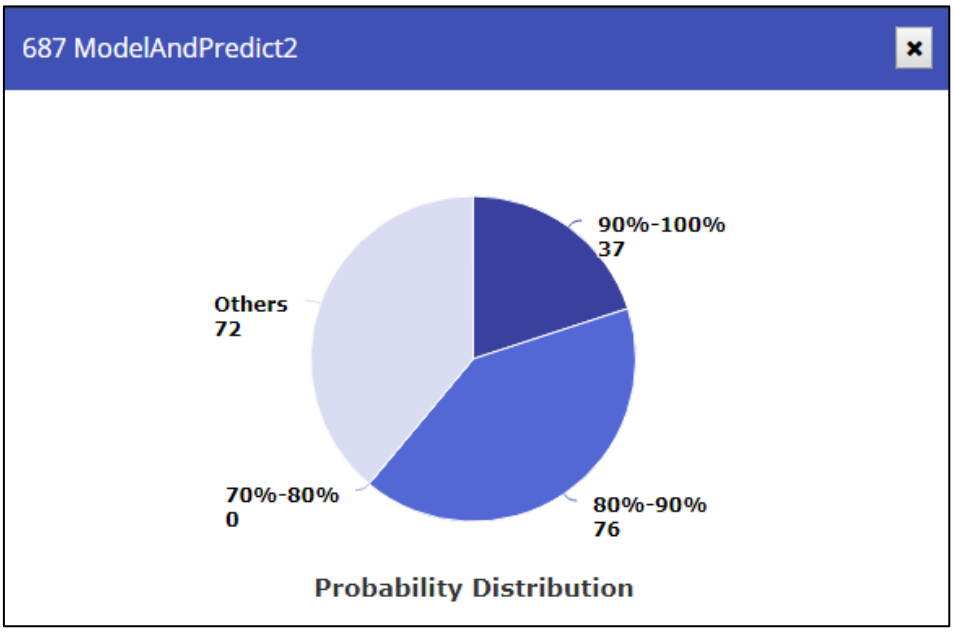


Figure 11 - Completed Prediction Jobs - Discrete Target Variable

On the other hand, the successfully completed jobs processed as CTV jobs show the distribution of the prediction in a pie chart, representing the number of records predicted using a model.

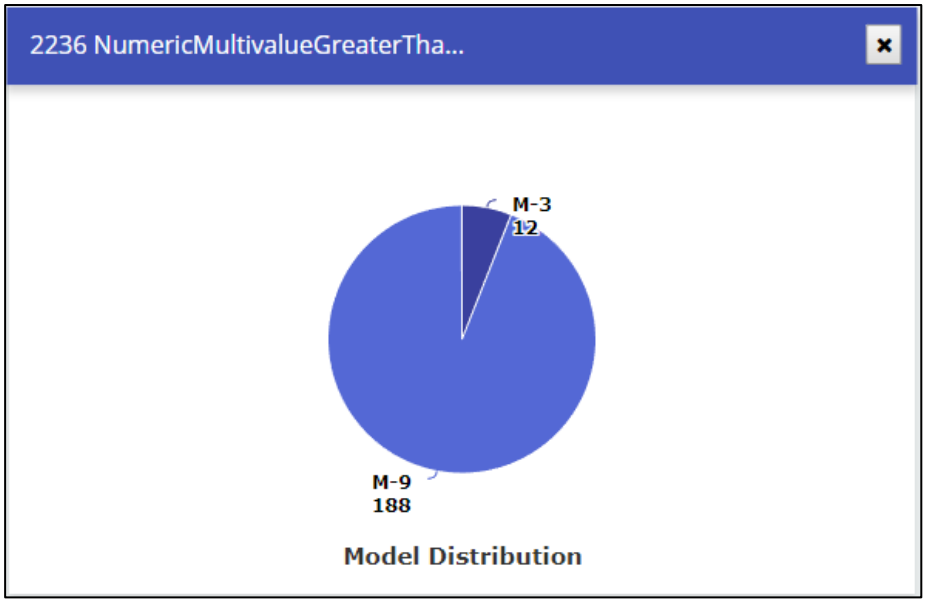


Figure 12 - Completed Prediction Jobs - Continuous Target Variable

On clicking the chart, Users will be redirected to Analytics Page for the job.

2.5.1.2 Modeling Only Job

In this case, System indicates that this is a Modeling only job and will provide button to download the model archive file.

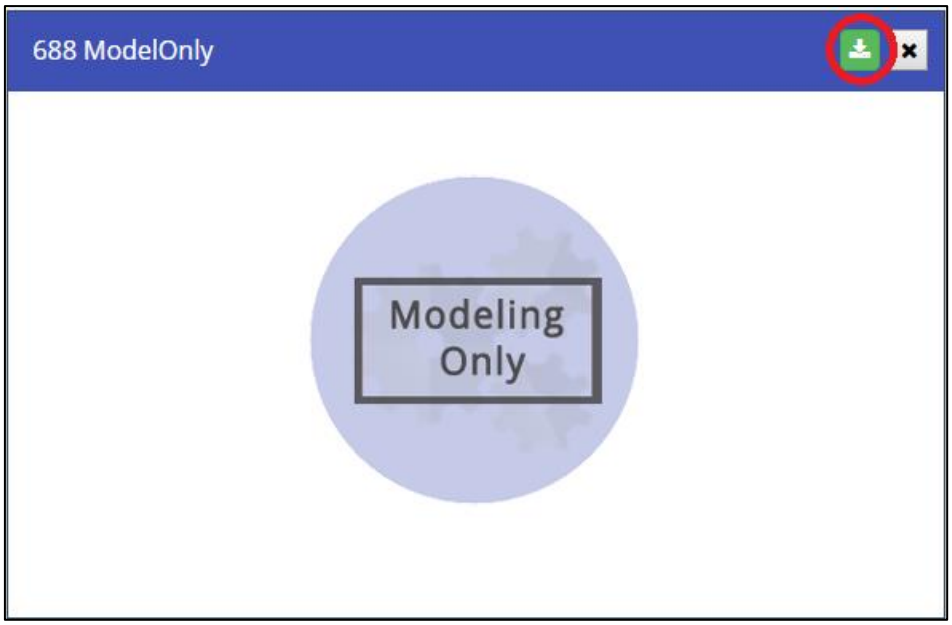


Figure 13 - Modeling Only Jobs

2.5.1.3 Not Successful Jobs

In case the jobs were not completed successfully, System will show the job as the image below.

Figure 14 - Failed Jobs

2.5.1.4 Zero models generated

In case there are no models generated, System will show the job as the image below.

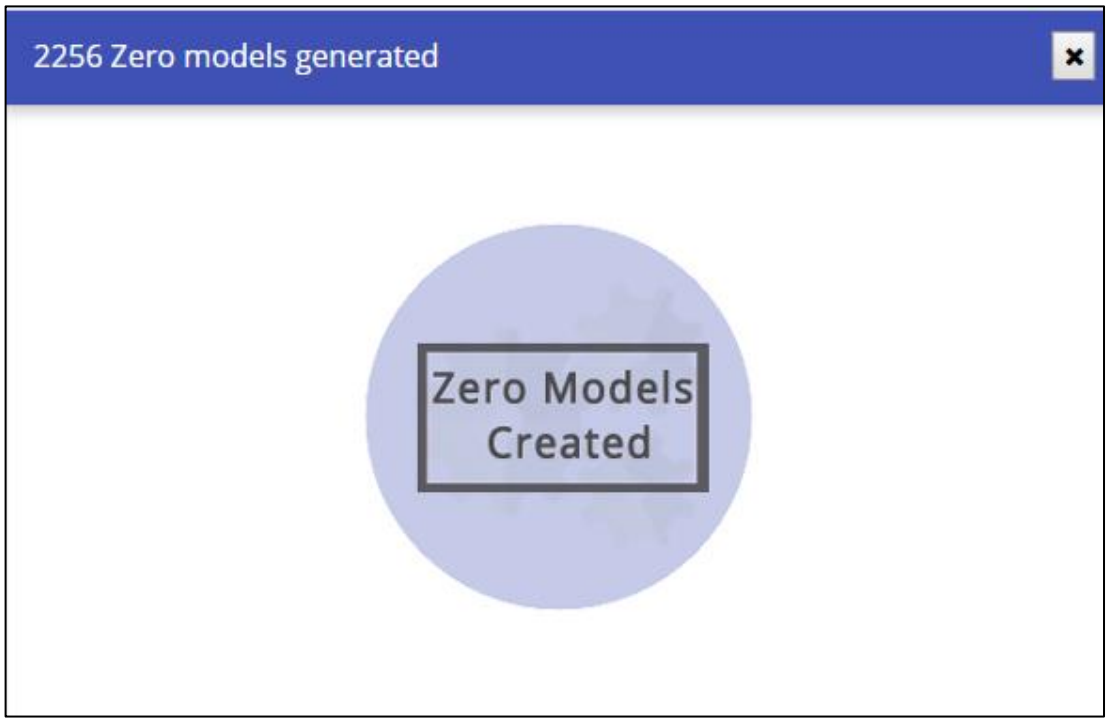


Figure 15 - Zero Models generated

2.5.1.5 Forecast Jobs

These are the jobs where multiple values of future events are predicted, where the data is a series of values of a quantity, obtained at successive times, with equal intervals between them.

When there is no group identifier selected the figure shows following trends:

a. **Historical values**: Shown in the light blue shade, the existing trend present in the data. In case of prediction only jobs, these values will not be shown.
b. **Forecast Values**: Shown in the darker blue shade, the forecast value for each period.
c. **Actual Values:** Shown in the darkest blue shade, the actual value for each period. These values are used for comparison with the forecast values. The values will not be present, if actual values associated with the period are not uploaded.

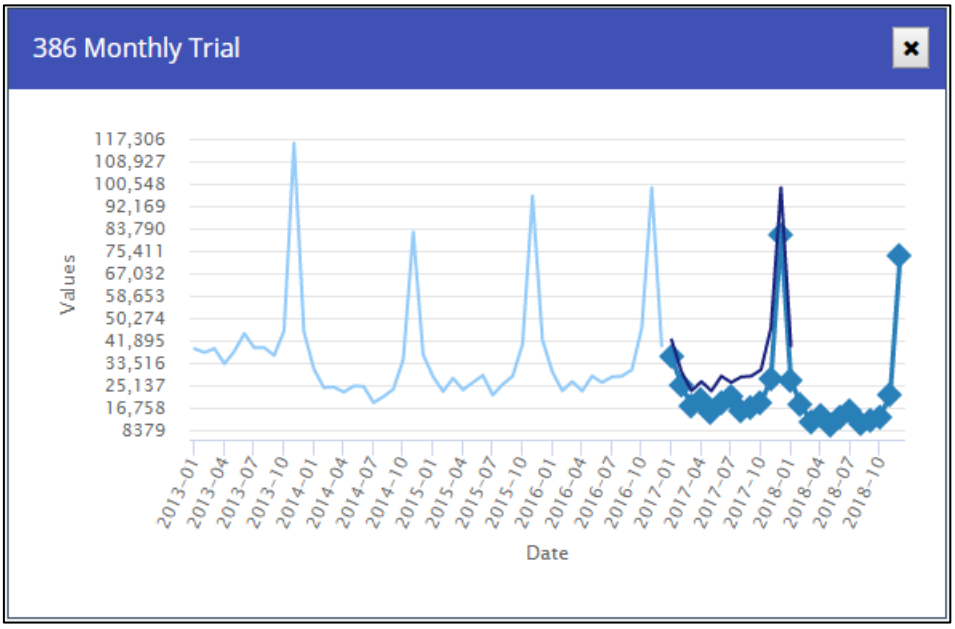


Figure 16 - Forecast Jobs without group identifier

However, if group identifier is selected (MultiModel Forecast), system will show the distribution of records predicted by a model in a tree map as follows:

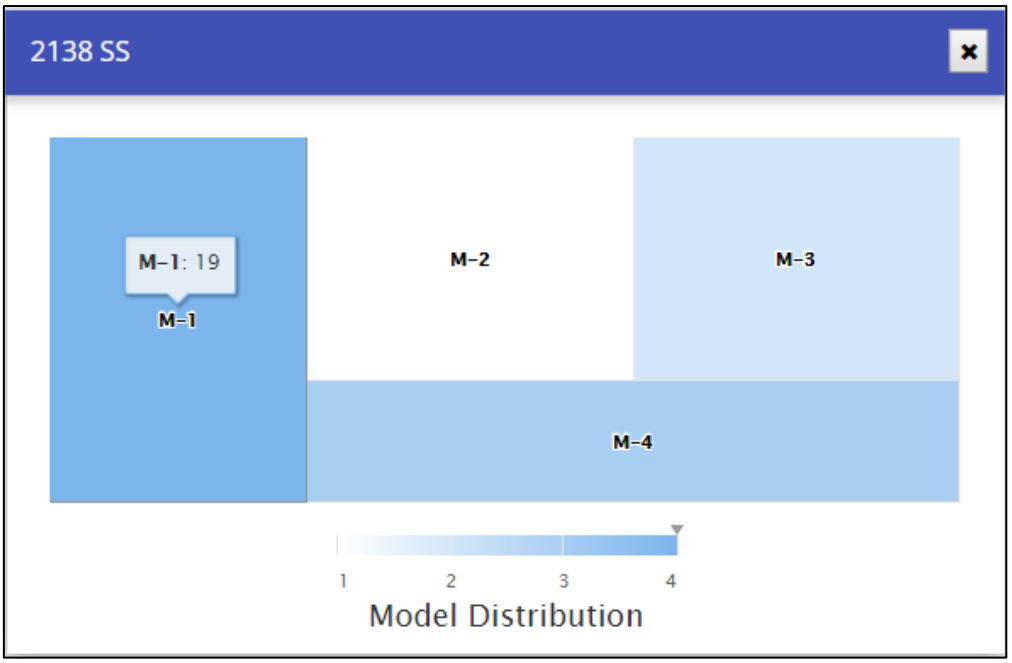


Figure 17 - MultiModel Forecast

On clicking the chart, Users will be redirected to Analytics Page for the job.

2.5.1.6 Remove Job

In case User wants to remove the job from the Job summary, System provides the functionality to remove the job by clicking on the remove the job button.

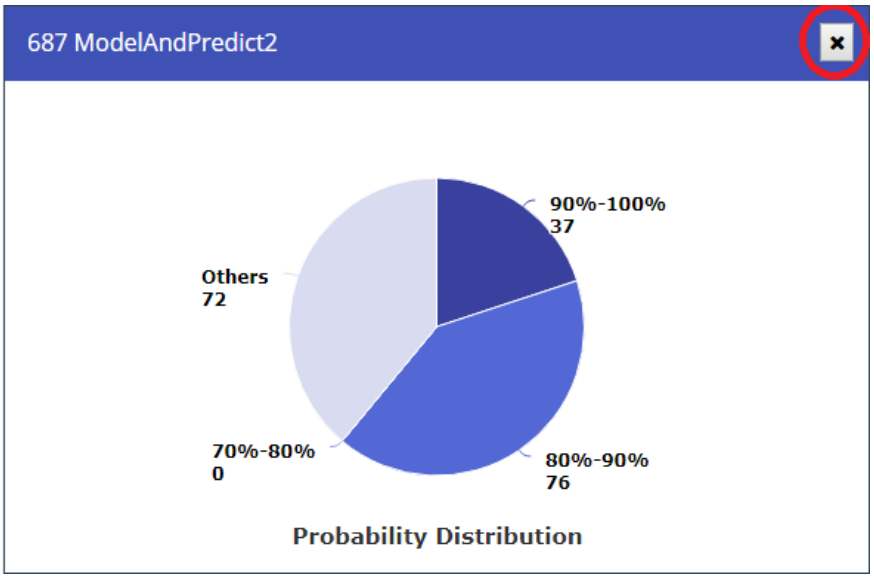


Figure 18 - Remove Job

2.8.2 Start Job

User can start a new job by clicking on the start button. Users will only be able to run jobs in case the instance is registered.

Figure 19 - Start Job Button

When users have scrolled down a little, the button will have a change in appearance and will be floating over the results.

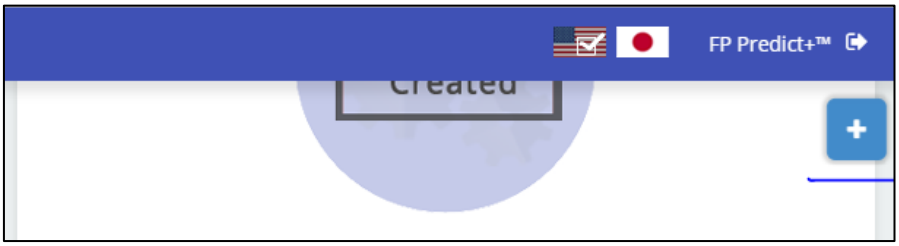


Figure 20 - Start Job Button - Floating

When user clicks on the start button, system asks ed with following 2 questions:

1. Does Your Data Contain date/time (Timestamp)?

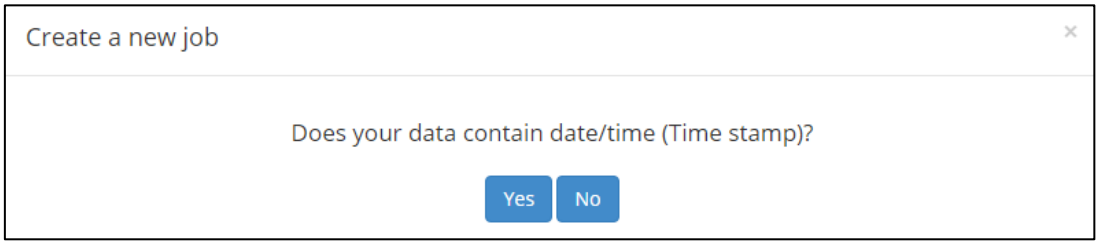


Figure 21 - Create Job Question 1

2. If the answer for first question is Yes, system asks the next question as “Do you need predicted values for multiple future periods?”

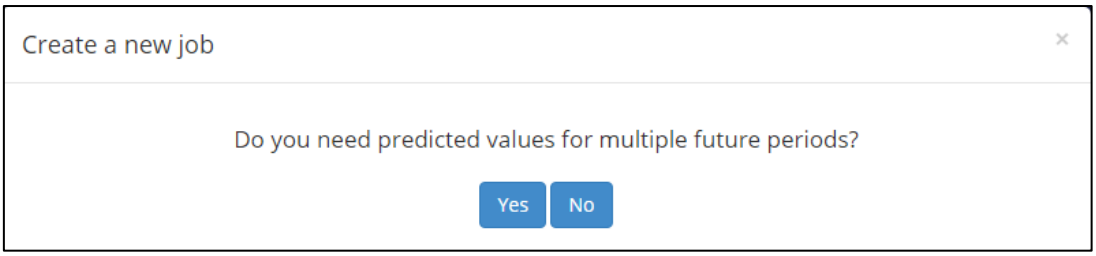


Figure 22 - Create Job Question 2

Here based on the user response, System will ask the following details for the Prediction or Forecast Job.

a. **Prediction**: When any of the answers is No.
b. **Forecast**: When Both answers are Yes.

1. **Prediction Job**: The details required for the job are as follows:

SL no.	Field name	Description	Comments
1.	Job Name	User Defined Name of the Job	Maximum 50 characters
2.	Job Description	User Defined Description regarding the Job. It may contain details about target, data sets, variables etc.	Will be set with job name by default.
3.	Job Type	Type of job. Can be one of the following: 1. **Model+Predict**: Modeling and Prediction 2. **Model**: Modeling Only 3. **Predict**: Prediction Only	
4.	Training File	CSV file for training the system and creating the models for prediction	Required for Modeling and Prediction, and Modeling Only jobs
5.	Target Variable	Header/ Column name in the data set for which the training/models are to be created	Required for Modeling and Prediction, and Modeling Only jobs
6.	Tracker Variable Checkbox	Checkbox allows users to select tracker variable for the modeling phase	
7.	Tracker Variable dropdown	Header/ Column name in the data set to be used as tracker variable. Refer Appendix #4 Applicable only for DTV jobs	Mandatory when Tracker Varaible checkbox is checked.
8.	Training Options	In case the target variable is Binary (contains only two unique values), should modelling be done to predict Less frequent value only, More Frequent value only or Both values. The details of Binary target prediction are present in Appendix #3	Required for Modeling and Prediction, and Modeling Only jobs
9.	Models File	Models File downloaded at the end of Modeling and Prediction or Modeling Only job, required for Prediction	Required for Prediction Only Job
10.	Prediction File	CSV file for prediction.	Required for Modeling and Prediction, and Prediction Only Jobs
11.	Id variable	The variable which is used to uniquely identify the records in the dataset.	Required for all Job types
12.	Advanced Settings	Settings providing the user ability to define the system behaviour for the Prediction Processing.	Advanced settings will not be available in case the job is Predict Only.
12.1	Automated	By choosing this option, user is allowing system to identify the type of use case and process it accordingly.	This will be the default setting for every job.
12.2	Manual Override	When user selects this option, system helps user decide the applicable use case asking a question “Predicted value should be part of Target variable values present in Training data”. System offers two options to choose as response to this question. Based on selected answer, system understands the category of the use case, which is as follows: 1. **“It Must be”**: Discrete Target Variable Case. 2. **“Not Necessarily”**: Continuous Target Variable Case.	If system finds the selected option to be inapplicable, based on the training data provided, system will ignore the manual override and process the data for applicable use case. System will provide an alert in such situation.

Table 1 - Prediction Job Parameters

The guidelines for the Training and prediction datasets are present in the Appendix #1

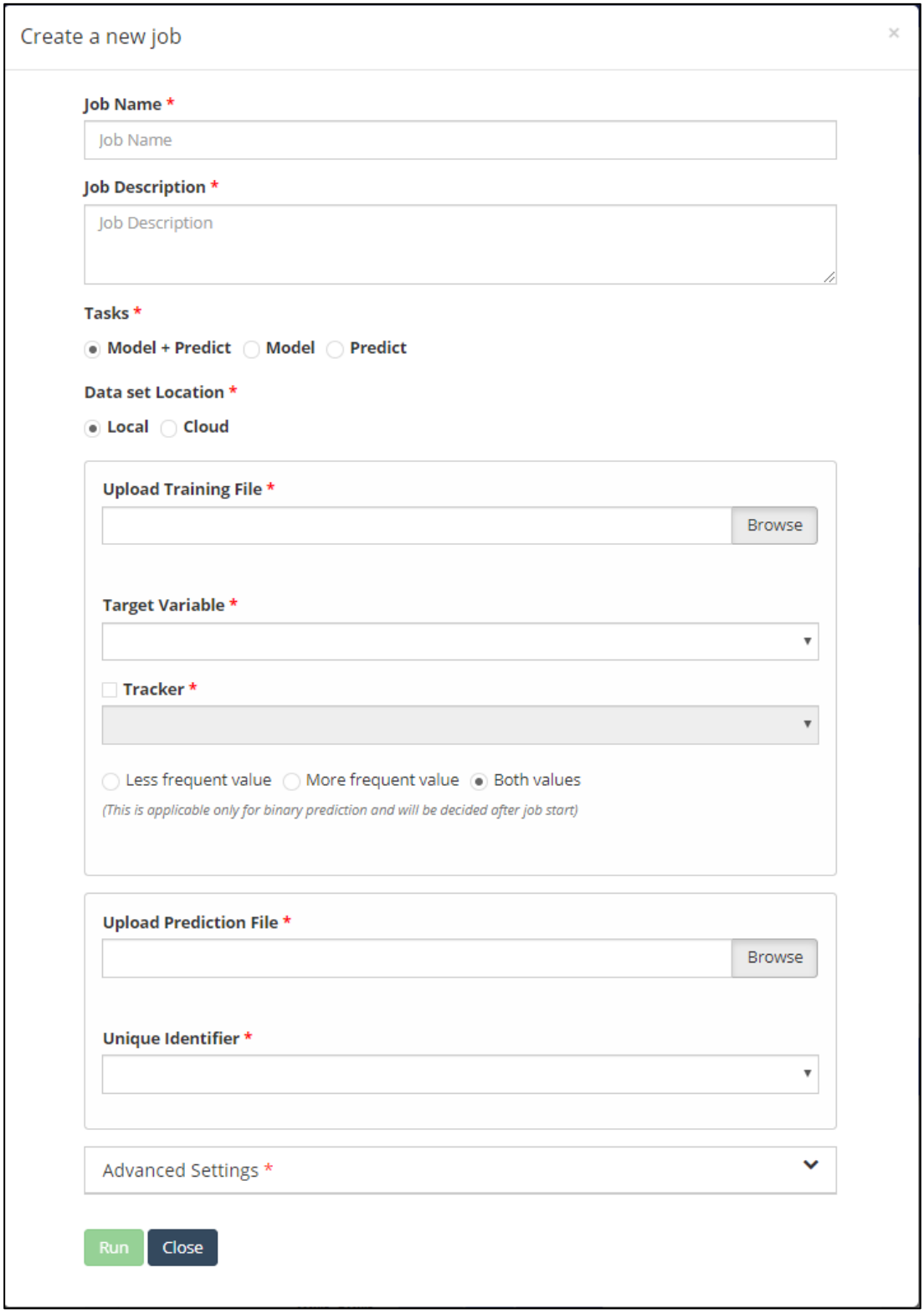


Figure 23 - Start Job Form - FPPredict

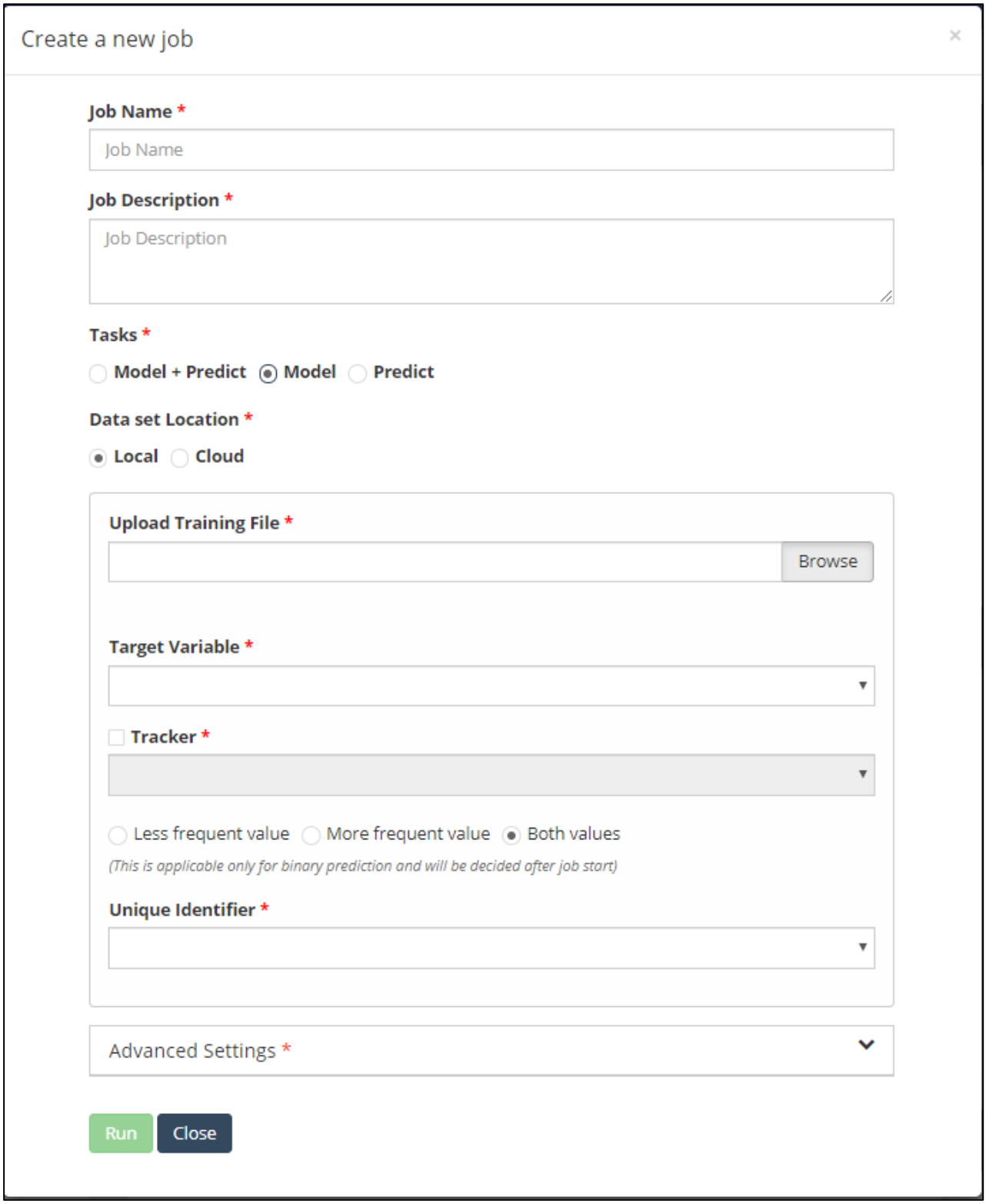


Figure 24 - Start Job Form - Modeling Only

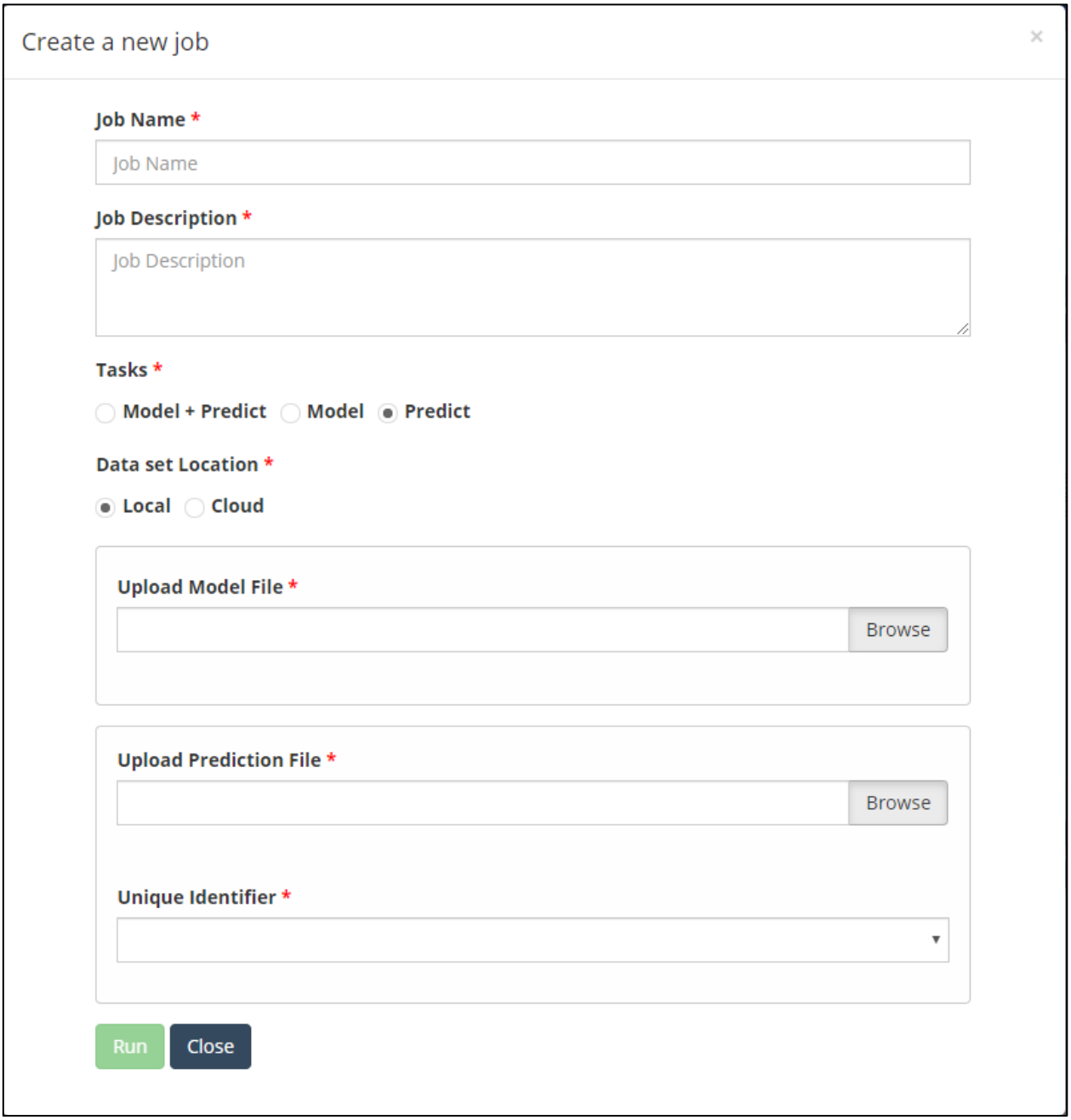


Figure 25 - Start Job Form - Prediction Only

On Starting a job, system shows the running status of the job.

Figure 26 - Job Status Dialog - Job Running

The dialog provides following two functionalities:

1. **Show Job Summary:** Closes the dialog and allows Users to see the results of previously completed jobs.
2. **Cancel job:** Cancels the current job.

Once the job finishes, system will show the user the status.

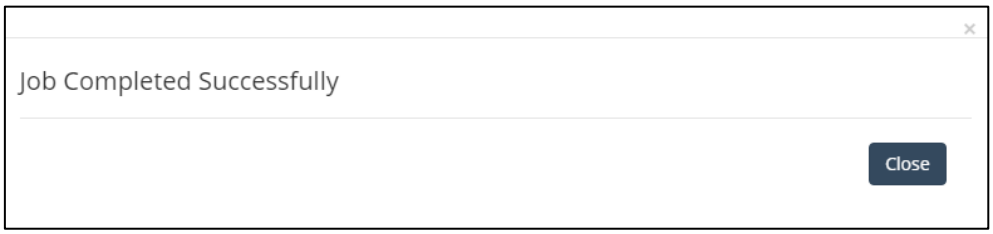


Figure 27 - Job Status Dialog - Job Finished

2. Forecast Job: The details required for the job are as follows:

SL no.	Field name	Description	Comments
1.	Job Name	User Defined Name of the Job	Maximum 50 characters
2.	Job Description	User Defined Description regarding Job. It may contain details about target, the data sets, variables etc.	Will be set with job name by default.
3.	Data Interval	Time interval present between the two consecutive observations given in the data set.	
4.	Tasks (Job Type)	Type of forecast job. Can be one of the following: 1. **Model+Forecast**: Modeling and Forecast 2. **Model**: Modeling Only 3. **Forecast**: Forecast Only	
5.	Training File	CSV file for training the system and creating the models for prediction.	Required for Modeling and Forecast, and Modeling Only jobs
6.	Target Variable	Header/ Column name in the data set for which the training/models are to be created	Required for Modeling and Forecast, and Modeling Only jobs
7.	Models File	Models File downloaded at the end of Modeling and Prediction or Modeling Only job required for Prediction	Required for Forecast Only Job
8.	Forecast File	CSV file for Forecast.	Required for Modeling and Forecast, and Forecast Only Jobs
9.	Timestamp variable	Header/ Column name in the data set which represents the time period associated with the record.	Required for Modeling and Forecast, and Forecast Only Jobs
10	Timestamp format	Time period format.	The dropdown value changes based upon the data interval
11.1	Group Identifier	A checkbox which allows users to select the group identifier column from the dataset	
11.2	Group Identifier	Header/ Column name in the data set which contains the Group Ids, used to identify subgroups present in the dataset	Optional parameter, only required if above parameter is checked.
12.	Remove Outliers	Whether System should remove outlier data from training process automatically	

Table 2 - Forecast Job Parameters

The guidelines of the Training and Forecast datasets are present in the Appendix #2

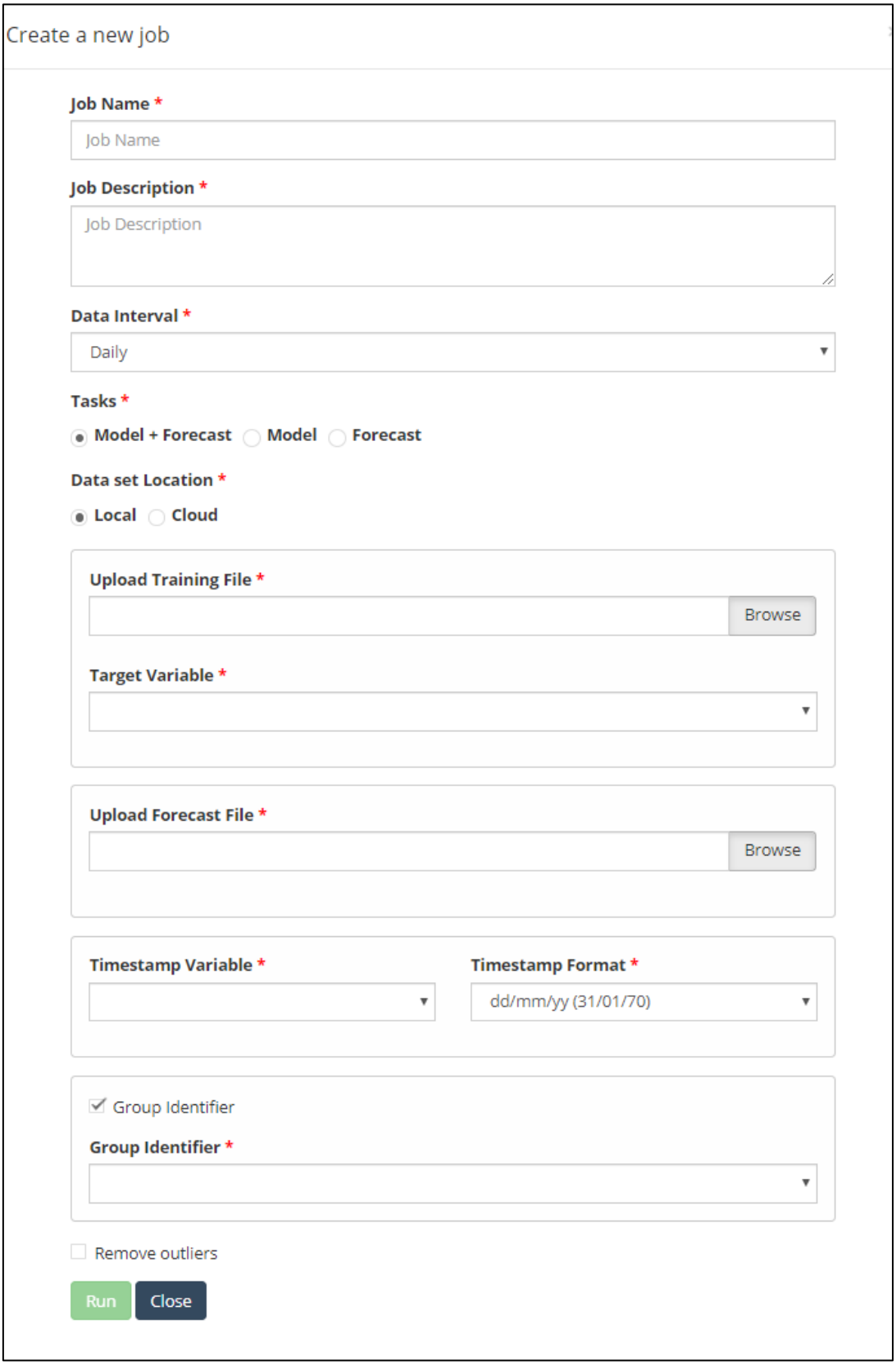


Figure 28 - Start Job Form - Modeling and Forecast

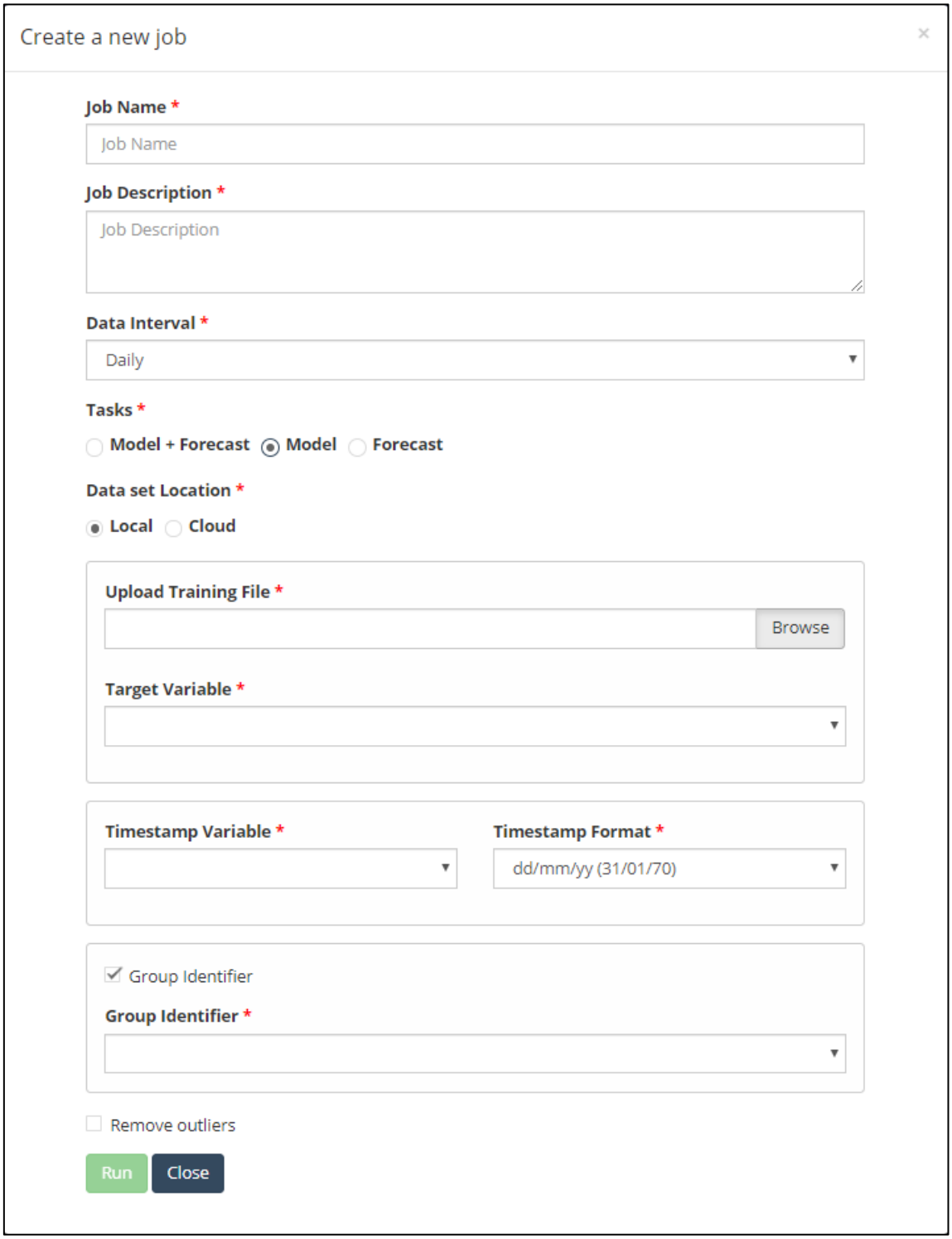


Figure 29 - Start Job Form - Modeling

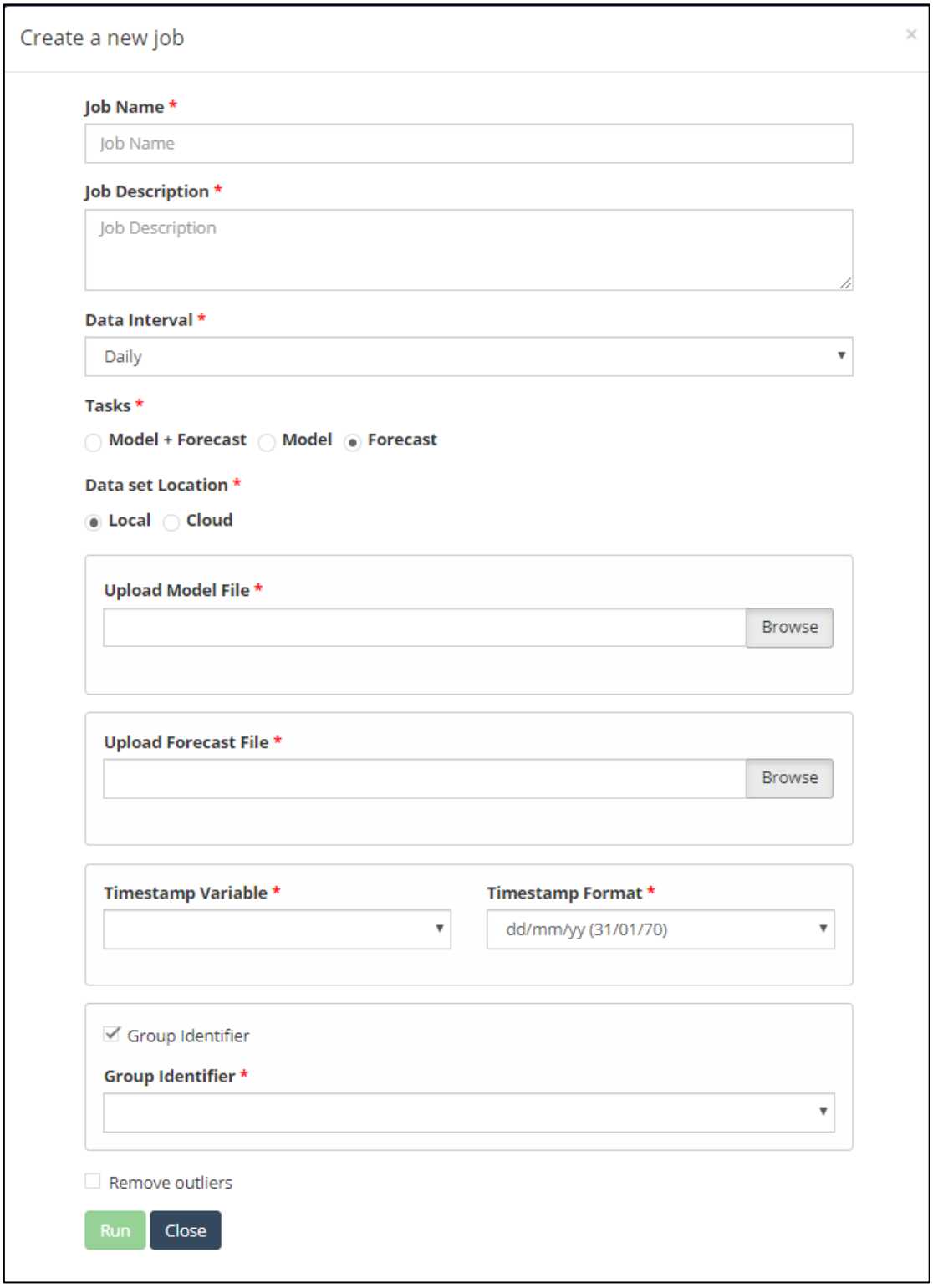


Figure 30 - Start Job Form - Forecast

At any given moment, System will allow only one running job.

2.8.3 Job Pagination

At any given moment, system will only show the latest 15 jobs. For view the older jobs, use the pagination bar at the bottom of the page.

Figure 31 - Job Summary - Pagination

2.9 Analytics

This page provides the analytics of the results for the successful prediction or forecast jobs. The list will only show the Jobs with following criteria:

1. **Prediction Jobs**: Completed Modeling and Prediction, or Prediction only jobs will be shown. Modeling only jobs and Failed jobs will not be part of the list.
2. **Forecast Jobs**: Completed Modeling and Forecast jobs or Forecast Only jobs will be shown. Modeling only jobs and Failed jobs will not be part of the list.

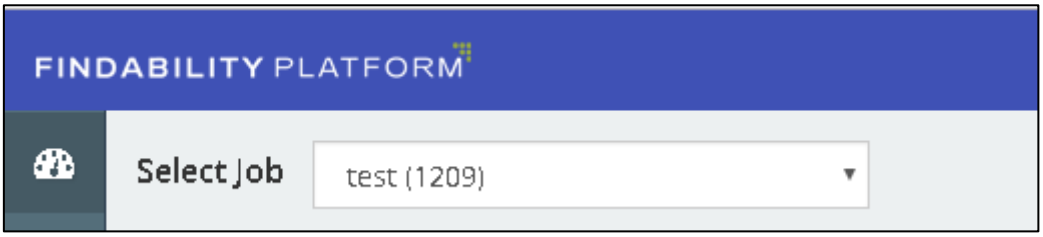


Figure 32 - Select Job Dropdown

On selection, system will show the description of the job, which user has given on the job creation screen.

Figure 33 - Job Description

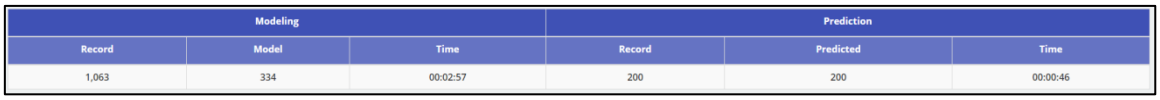
2.9.1 Prediction Job

For Prediction Jobs the system will provide the analytics as follows:

2.6.1.1 Job Summary

This provides a summary of the job as following:

1. Number of records in Training File
2. Number of models created
3. Time taken for creating models
4. Number of records in Prediction File
5. Number of records Predicted
6. Time taken for Prediction



Modeling			Prediction		
Record	Model	Time	Record	Predicted	Time
1,063	334	00:02:57	200	200	00:00:46

Figure 34 - Job Summary

2.6.1.2 Prediction vs Actual

This chart provides the comparison of predicted vs actual values. This will only be generated in case actual values of the target variable are present in the prediction data set.

The chart generated can be of following types based on the target variable:

1. **Binary and Single value Prediction**: In case User predicts Single value of target variable by choosing Less Frequent Value or More Frequent Value option. The chart will be showing the hit percentage and lift for every decile in the result.

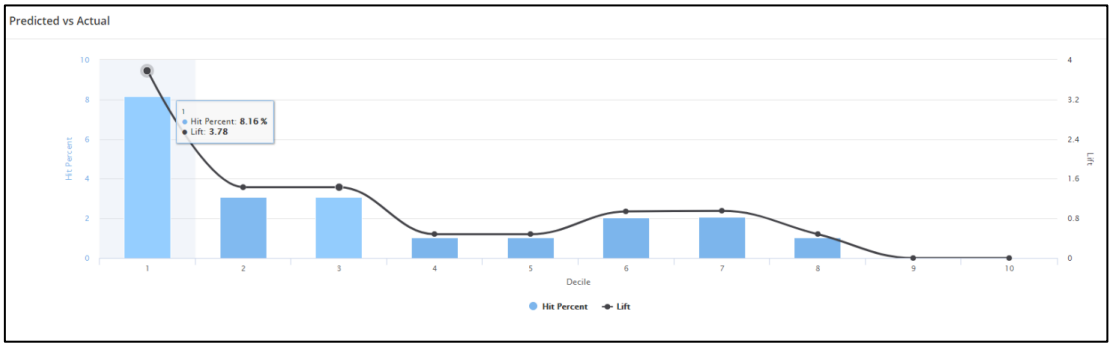


Figure 35 - Predicted vs Actual - Binary and Single value Prediction

On clicking on one of the deciles, system provides the drill down in form of the records associated with the decile.

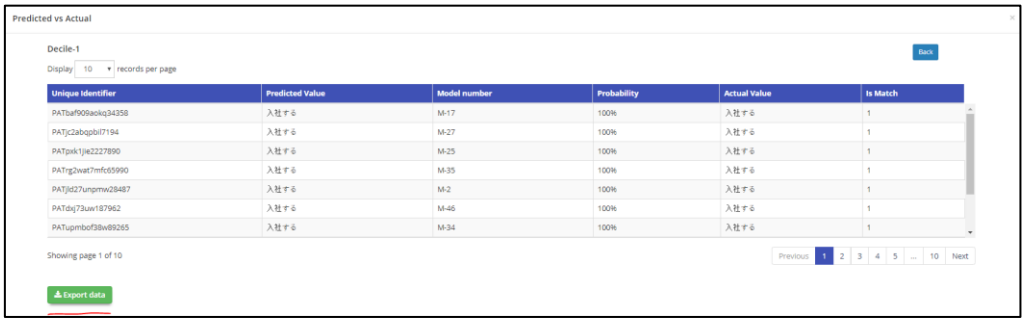


Figure 36 - Predicted vs Actual - Binary and Single Prediction Drilldown

Clicking on export data allows the users to download the results associated with the given decile.

	A	B	C	D	E	F
2	Unique Identifier	Predicted Value	Model Number	Probability		
3	PATbaf909aokq3435	入社する	M-17	100.00%		
4	PATjc2abqpbil7194	入社する	M-27	100.00%		
5	PATpxk1jie2227890	入社する	M-25	100.00%		
6	PATrg2wat7mfc6599	入社する	M-35	100.00%		
7	PATjld27unpmw284	入社する	M-2	100.00%		
8	PATdxj73uw187962	入社する	M-46	100.00%		
9	PATupmbof38w892	入社する	M-34	100.00%		
10	PAT6gura87fh37388	入社する	M-46	100.00%		
11	PATdpev7d98c06908	入社する	M-46	100.00%		
12	PATqjhrepesus1404	入社する	M-44	100.00%		
13	PATuw1bimk983833	入社する	M-44	100.00%		
14	PATs6j7b183qe8655	入社する	M-45	100.00%		
15	PATbuavk8gkfg5707	入社する	M-5	100.00%		
16	PATs34b6crsp65297	入社する	M-22	100.00%		
17	PATpiflriq2770109	入社する	M-5	100.00%		
18	PATugp6sgux4r6768	入社する	M-22	100.00%		

Figure 37 - Decile Export

Table description

SL	Column Name	Column Definition	Comment

1	Unique Identifier	Unique value which uniquely identified the records	
2	Predicted value	Job description given while creating a new job.	
3	Model Number	Represents model used to predict the row	
4	Probability	Probability associated with the prediction	

Table 3 - Decile Export Description

2. **Binary, both value prediction**: In case the target variable is binary with user choosing "both values" option, the comparison is shown as a heat map as follows

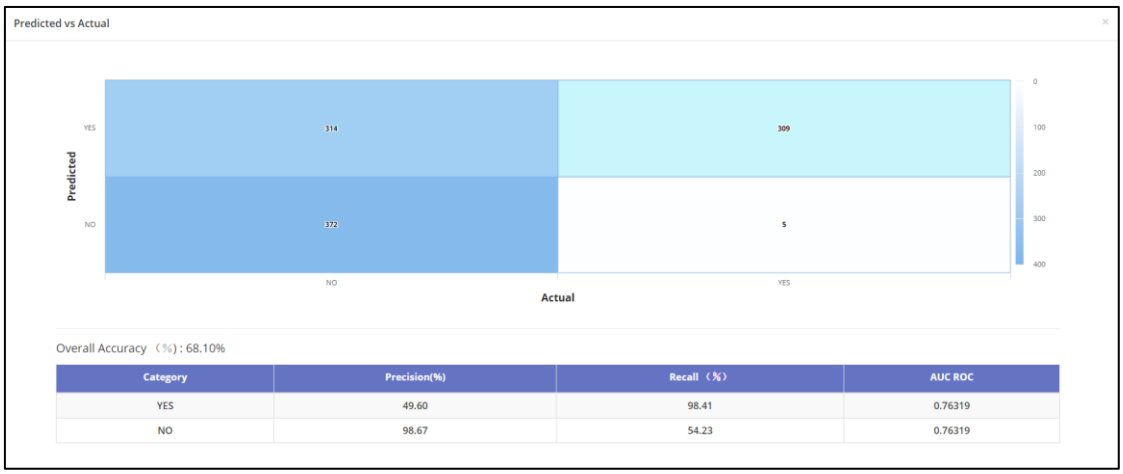


Figure 38 - Predicted vs Actual - Binary, both value prediction - Categorical

Along with heatmap, system also shows following features regarding the results:

a. **Precision**: It is measure of how often it is correct, when it predicts for a specific category.
It is the ratio of correctly predicted records to total number of records predicted for that category.

b. **Recall:** It is measure of how often it predicts the category correctly for those records which belong to that category.

It is the ratio of correctly predicted records to total number of records belonging to that category.

c. **Overall Accuracy**: Overall accuracy is the ratio of correct predictions to total predictions made.

d. **AUC ROC**: AUC stands for Area under the ROC (Receiver Operating Characteristic) Curve. AUC measures the entire two-dimensional area underneath the entire ROC curve from (0,0) to (1,1). ROC Curve tells us about how good the system can distinguish between two categories.
This score gives us a good idea of how well the prediction results are, higher the value of AUC, better is the accuracy.

The definitions are available on the UI on hovering over the terms.

3. **Multi Value - Numerical**: In case the target variable has more than 2 unique values in training dataset, and all of them are numbers, the comparison is shown as a scatter chart as follows:

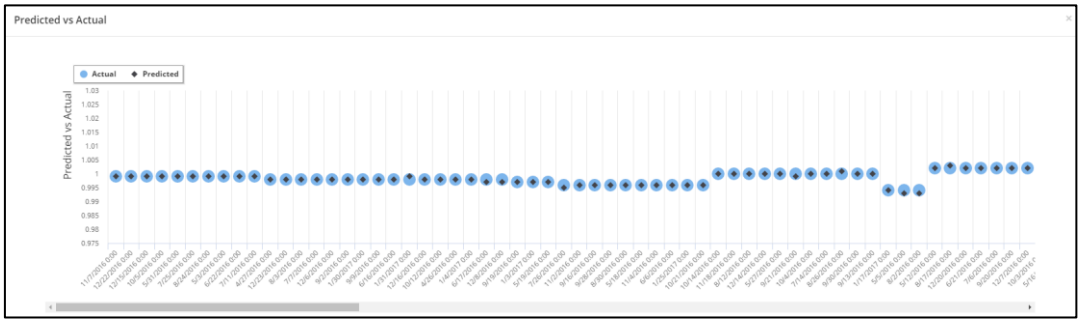
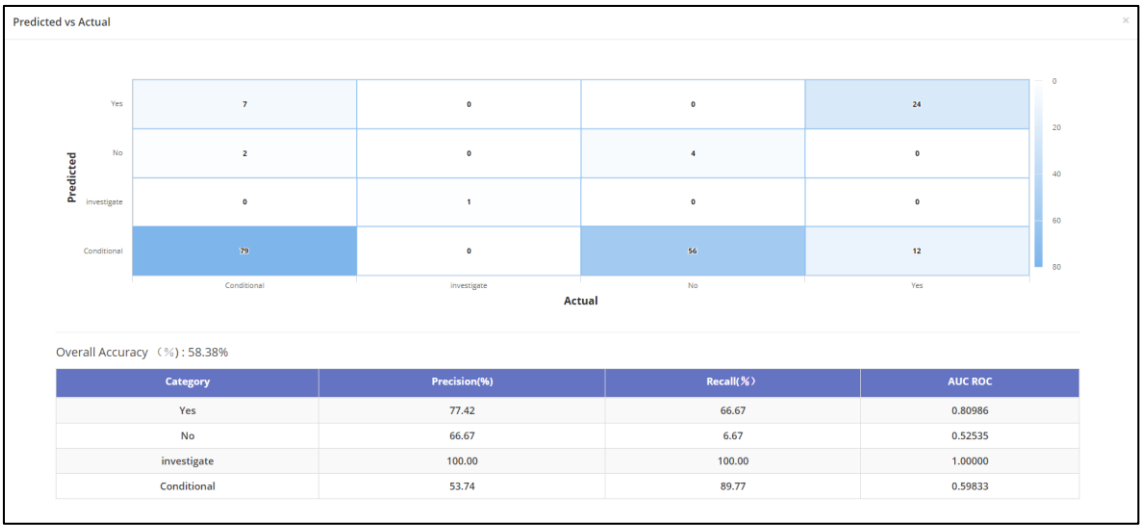


Figure 39 - Predicted vs Actual - Multi Value - Numerical

4. **Multi Value - Categorical**: In case the target variable has more than 2 unique values in training dataset, and at least one of them is non-numeric, the comparison is shown as a heat map as follows



Category	Precision(%)	Recall(%)	AUC ROC
Yes	77.42	66.67	0.80986
No	66.67	6.67	0.52535
investigate	100.00	100.00	1.00000
Conditional	53.74	89.77	0.59833

Figure 40 - Predicted vs Actual - Multi Value - Categorical

Along with heatmap, system also shows Precision, Recall, Overall accuracy and AUC. These terms are defined in section 2.6.1.2. The definitions are available on the UI on hovering over the terms.

2.6.1.3 Models

This bar chart provides a comparison of models and the number of records predicted using that model.

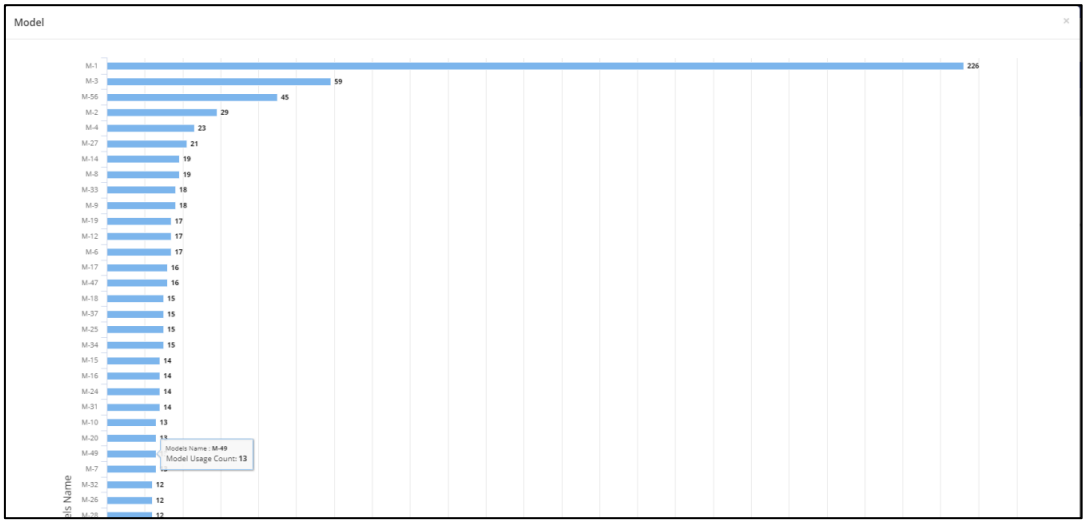


Figure 41 - Models

2.6.1.4 Variables

This bar chart provides a comparison of variables and the number of models in which they are present.

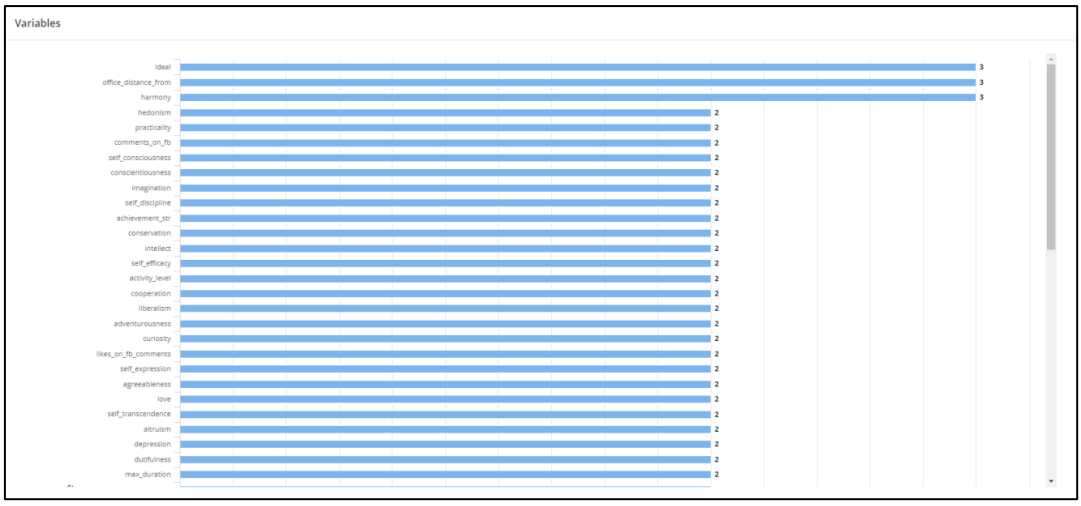


Figure 42 - Variables

2.6.1.5 Variables of Models

This tables provides the details of all models used for prediction and the variables involved. For CTV jobs the system shows the variables in the order of their importance from highest to lowest. For DTV jobs, the all the variables have same importance.

Variables of Models

Model Number	Variables1	Variables2	Variables3
M-2	emotionality	office_distance_from	agreeableness
M-3	emotionality	harmony	office_distance_from
M-4	harmony	ideal	office_distance_from
M-56	cooperation	anger	per_of_time_travel_req
M-1	current_ctc	agreeableness	-
M-5	harmony	ideal	-
M-6	ideal	self_expression	-
M-7	orderliness	self_expression	-
M-8	orderliness	challenge	-
M-9	challenge	practicality	-
M-10	practicality	excitement_seeking	-
M-11	excitement_seeking	hedonism	-
M-12	hedonism	self_transcendence	-
M-14	self_enhancement	comments_on_fb	-
M-15	comments_on_fb	depression	-
M-16	depression	structure	-
M-17	senti_of_posts_liked_on_fb	structure	-

Figure 43 - Variables of Models

2.6.1.6 Download Results

Allows users to download the results in Excel Format.

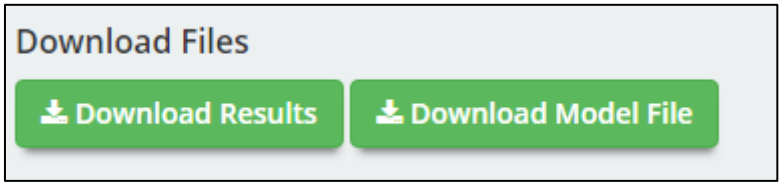


Figure 44 - Download Results

2.6.1.7 Download Model File

Allows users to download the models file. This will not be available for Prediction only jobs.

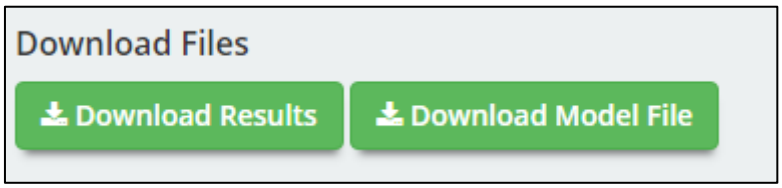


Figure 45 - Download Models

2.9.2 Forecast Job

For Forecast Jobs analytics provided by system, changes whether subgroups were identified by the system during job processing or not. Therefore, the analytics provided are as following:

2.6.2.1 With subgroups

In case there are subgroups recognized during job processing, system provides following analytics for the forecast job:

2.9.2.1.1 Overall Summary

This will show overall summary of the jobs, showing details like Total number of records, Groups Identified, Models created, Total processing time etc. Along with this, system will also show Variables impact and Overall variables associated with every model created.

This section contains following subsection:

2.6.2.1.1.1 Job Summary

This provides a summary of the job as following:

1. Number of records in Training File
2. Number of subgroups detected
3. Total number of models created
4. Time taken for creating models
5. Number of records in Prediction File
6. Number of subgroups forecast
7. Number of models applied for forecast
8. Time taken for Forecast

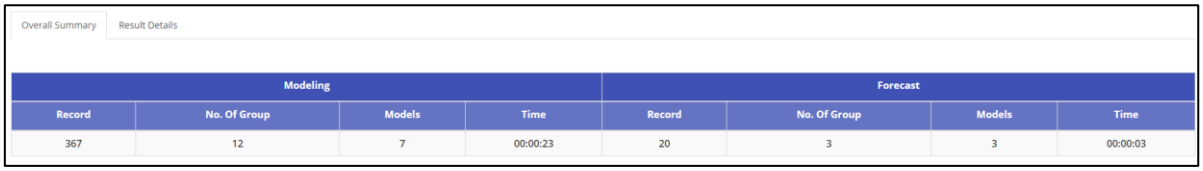


Figure 46 - MultiModel Forecast - Job Summary

2.6.2.1.1.2 Models

This section provides the contribution of the variable in the models that were used for forecast.

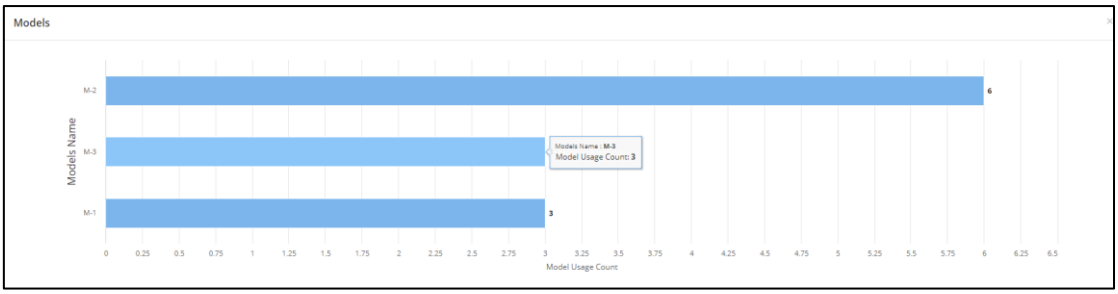


Figure 47 - MultiModel Forecast – Models

2.6.2.1.1.3 Variables

This section provides the contribution of the variable in the models that were used for forecast.

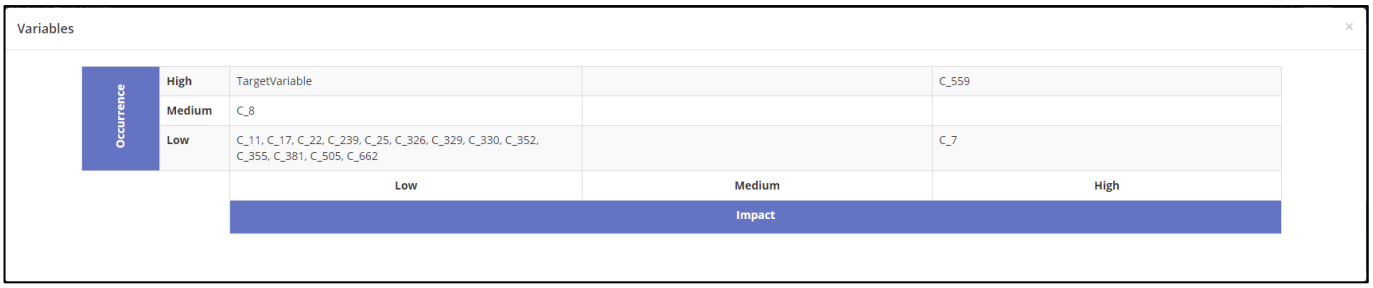


Figure 48 - MultiModel Forecast - Variable Contribution

Here, the contribution is shown on based on two factors:

a. Occurrence: This factor represents the number of models where the variable is present.
b. Impact: This factor represents the impact (Weight) of the variable in the models where the variable is present.

Also, the contribution has been classified on three scales, which are

a. Low
b. Medium
c. High

For Eg: Target variable in the above example, has high contribution when considering the occurrence, and Variable "C_7" has high contribution considering impact on model. Similarly, C_559 has high contribution considering both high impact and high occurrence.

2.6.2.1.1.4 Variables of Models

This section provides the view of variables associated with a model and the group where model was applied, in a tabular view.

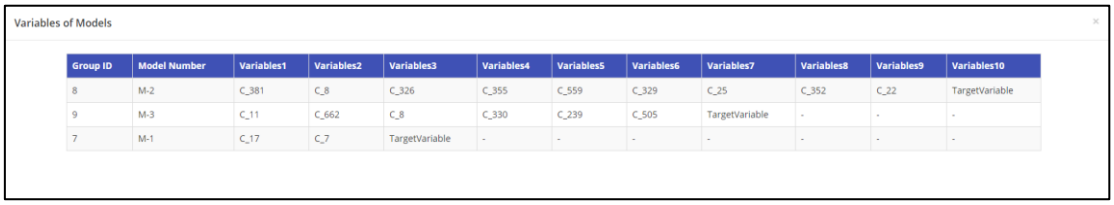


Figure 49 - MultiModel Forecast - Variables of Models

2.9.2.1.2 Result Details

This section provides the results of forecast of individual subgroups identified in the forecast dataset. On selecting a group id from the dropdown, the system shows the individual results for the subgroup.

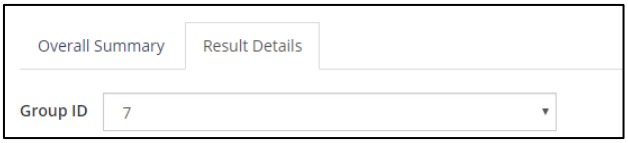


Figure 50 - Multimodel Forecast - Group ID Selection

After selecting the group, results are shown in following subsections as follows:

2.6.2.1.2.1 Forecast Results

This section shows the results of the forecast for the subgroup, with following details:

1. **Timestamp**: Shows the timestamp associated with the future periods present in the prediction file provided during job creation.
2. **Forecast Values**: The values that the system has forecast for each period.
3. **Actual Values**: Actual values provided by the user. The values will not be present, if actual values associated with the period are not uploaded or not present in the prediction file under target variable column.
4. **Deviations**: In case the actual values are present, system will also show deviations of predicted values from actual values.

Timestamp	Forecast Values	Actual Values	Deviation Values
20-06-2017	1.002	1.008	0.006
21-06-2017	1.004	1.000	-0.004
22-06-2017	1.000	0.999	-0.001

Figure 51 - MultiModel Forecast - Forecast Results

2.6.2.1.2.2 Forecast Trend

Here, System provides a graphical view of the trend present in the training data set and the forecast values. This is the same view that is shown in the Job summary. System provides facility to zoom out the view by clicking on the zoom out button.

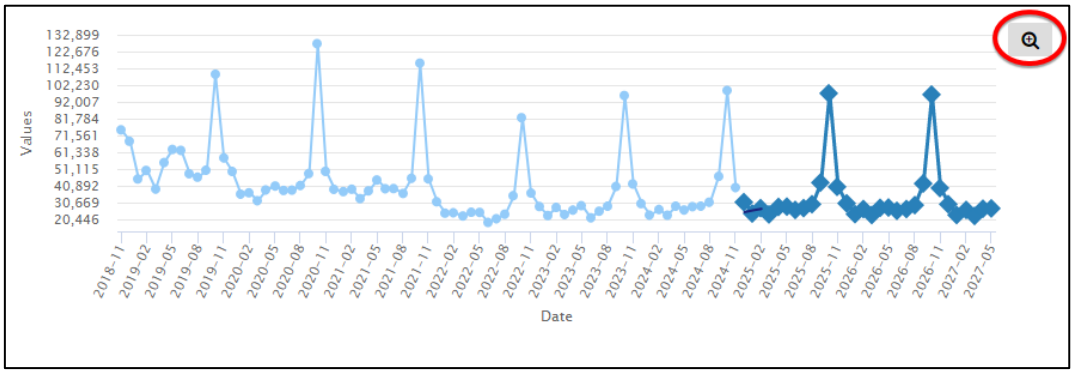


Figure 52 - MultiModel Forecast - Forecast Trend

2.6.2.1.2.3 Model Details

This section provides following details for the model used for forecast for the selected subgroup:

1. **Model Number**: This represents the model used for forecast for the group.
2. **Number of Important Variables**: The count of variables (columns/headers) used to create the model, which was used for forecasting process for the group.
3. **Important Variable(s)**: A comma separated list of variables used for creating the model, which was used for forecasting process for the group.

2.6.2.1.2.4 Modeling Metrics

During modeling phase of FPForecast, system generates following metrics, which measure the performance of the models. These values are based on actuals available in training file and the forecasts generated during modelling as part of the modelling process:

a. **Mean Error (ME)**
   The Mean Error (ME) is the average of all the Forecast errors in the data set.
   A “Forecast Error” is the difference between Forecast value and Actual value. It is mean of (Actual Value – Forecast Value).

b. **Root Mean Squared Error (RMSE)**
The Root Mean Square Error (RMSE) refers to standard deviation of the residuals (Forecast errors). The RMSE is a frequently used measure of the differences between Forecast value and the Actual value.
RMSE, being square root of variance, can be interpreted as the standard deviation of the unexplained variance and has the useful property of being in the same units as the response variable.
RMSE is measure of accuracy to compare errors of different models for the same dataset and lower values of RMSE indicate better model.

c. **Mean Absolute Error (MAE)**
The Mean Absolute Error (MAE) is the average of absolute values of Forecast errors.
MAE measures the average magnitude of the errors in a set of Forecasts, without considering their direction or sign. It is mean of |Actual Value – Forecast Value|.

d. **Mean Percentage Error (MPE)**
The Mean Percentage Error (MPE) is the average of percent Forecast errors in the data set. Percent forecast error is actual value minus forecast value divided by actual value multiplied by 100.
In statistics, MPE is the computed average of percentage errors by which forecasts of a model differ from actual values of the quantity being forecast. It is mean of 100*(Actual Value – Forecast Value)/Actual Value.

e. **Mean Absolute Percentage Error (MAPE)**
The Mean Absolute Percentage Error (MAPE) is the average of absolute percent Forecast errors in the data set. Absolute percent forecast error is absolute of actual value minus forecast value divided by actual value multiplied by 100.
It is mean of 100*|Actual Value – Forecast Value|/Actual Value.

f. **Mean Absolute Scaled Error (MASE)**
The Mean Absolute Scaled Error (MASE) is independent of the scale of the data, so can be used to compare forecasts across data sets with different scales.

These definitions are available on the UI as well after mouseover the metric.

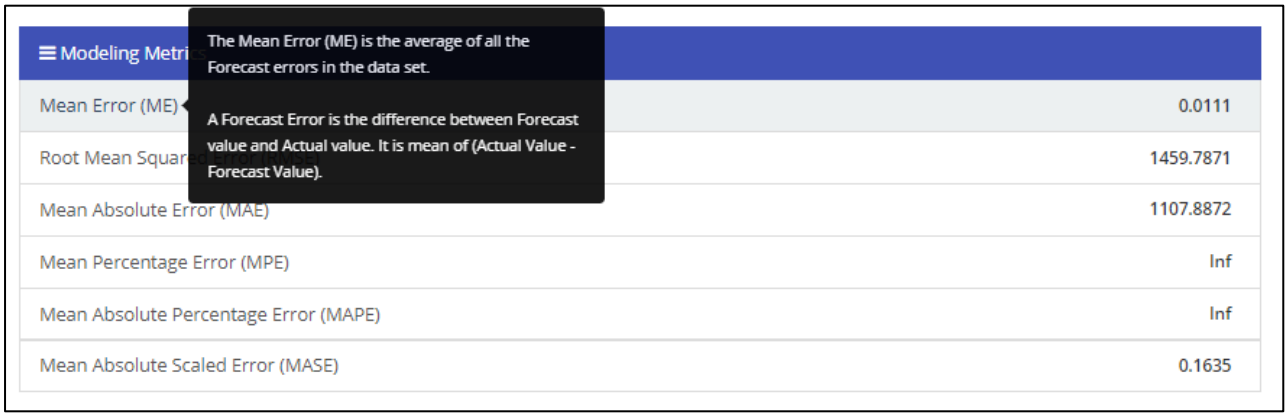


Figure 53 - Multimodel Forecast - Modeling Metrics

The lower the errors, the better is the performance.

2.6.2.1.2.5 Forecast Metrics

During forecast phase of FPForecast, system generates following metrics, which measure the performance of the forecast. These values are based on actuals available in forecast file and the forecasts generated during modelling as part of the forecast process.

The definitions of the metrics are present in the section 2.6.2.1.2.4 . The lower the errors, the better is the performance.

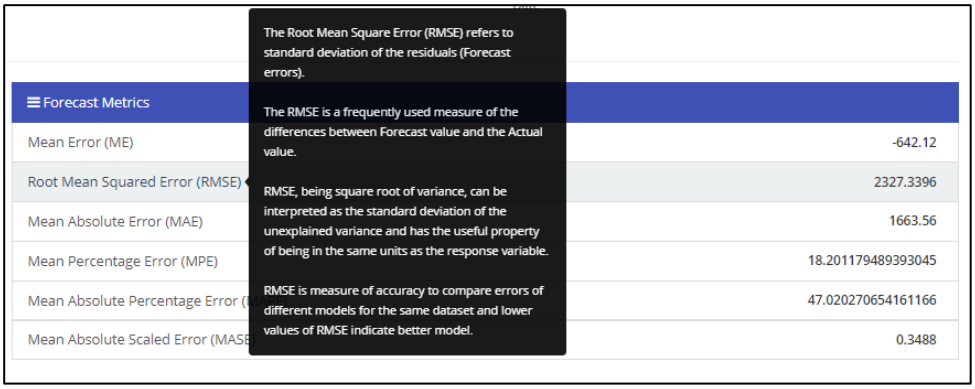


Figure 54 - Multimodel Forecast - Forecast Metrics

2.6.2.1.2.6 Model Details

This section provides the following information regarding the model used to forecast the group

- Model Number: Model used to forecast the group
- Number of variables: Number of variables used in the model
- Important variables: List of variables used in the model.

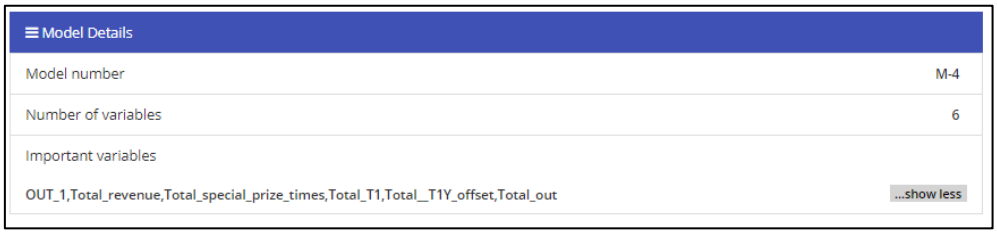


Figure 55 - Multimodel Forecast - Model Details

Clicking on Show more & Show less toggles full listing of variables.

2.6.2.1.2.7 Download Template and Upload Actuals

System Allows the users to upload the actual values associated with the time stamp. These values then are used as a method of comparison in Results and Trends.

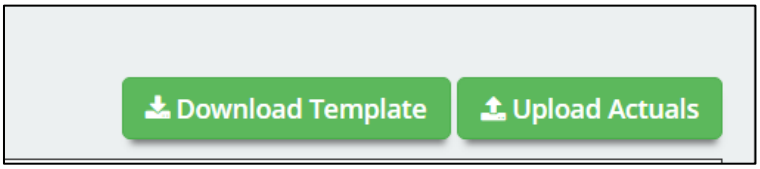


Figure 56 - Download Template & Upload Actuals

Users can download the template by clicking the “Download Template” button. In case users have provided actual values in the forecast file, the downloaded template will contain those actual values, which can be edited by users. Users should not edit the time stamp values.

	A	B	C
1		FPForecast Report	
2	Group ID	TimeStamp	Actual Value
3	7	2017-06-20	
4	7	2017-06-21	
5	7	2017-06-22	
6	8	2017-06-24	
7	8	2017-06-25	
8	8	2017-06-26	
9	8	2017-06-27	
10	8	2017-06-28	
11	8	2017-06-29	
12	8	2017-06-30	
13	8	2017-07-01	
14	8	2017-07-02	
15	8	2017-07-03	
16	9	2017-07-13	
17	9	2017-07-14	
18	9	2017-07-15	
19	9	2017-07-16	
20	9	2017-07-17	
21	9	2017-07-18	
22	9	2017-07-19	

Figure 57 - Upload Actuals Template - Without No Actuals

	A	B	C
1		FPForecast Report	
2	Group ID	TimeStamp	Actual Value
3	7	2017-06-20	1.008
4	7	2017-06-21	
5	7	2017-06-22	
6	8	2017-06-24	1.002
7	8	2017-06-25	0.995
8	8	2017-06-26	
9	8	2017-06-27	
10	8	2017-06-28	
11	8	2017-06-29	
12	8	2017-06-30	
13	8	2017-07-01	
14	8	2017-07-02	
15	8	2017-07-03	1.008
16	9	2017-07-13	1.002
17	9	2017-07-14	
18	9	2017-07-15	
19	9	2017-07-16	
20	9	2017-07-17	
21	9	2017-07-18	0.985
22	9	2017-07-19	0.986
23			
24			
25			

Figure 58 - Uploads Actuals Template - With Some Actual Values

Users then can upload the filled in template using “Upload Actuals” button.

2.6.2.1.2.8 Download Results

Allows users to download the Forecast Results in Excel Format by clicking on the “Download Results” button.

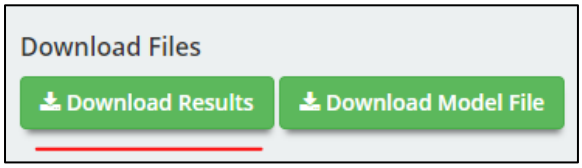


Figure 59 - Download Forecast Results - FPForecast

2.6.2.1.2.9 Download Models

This allows users to download the models generated for the job so that they can be reused in prediction only Jobs. The button will be not available for prediction only jobs.

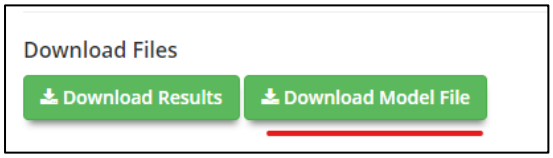


Figure 60 - Download Models - FPForecast

2.6.2.2 Without subgroups

In case there are no subgroups recognized, system provided following analytics for the forecast job.

2.6.2.2.1 Forecast Results

This section shows the results of the forecast job, with following details:

1. **Timestamp**: Shows the timestamp associated with the future periods present in the prediction file provided during job creation.
2. **Forecast Values**: The values that the system has forecast for each period.
3. **Actual Values**: Actual values provided by the user. The values will not be present, if actual values associated with the period are not uploaded or not present in the prediction file under target variable column.
4. **Deviations**: In case the actual values are present, system will also show deviations of predicted values from actual values.

Timestamp	Forecast Values	Actual Values	Deviation Values
2018-10-13	12,356.0000	11,258.5600	-1,097.4400
2018-10-14	694.0000	56,895.0000	56,201.0000
2018-10-15	694.0000	65,895.0000	65,201.0000
2018-10-16	694.0000	23,569.0000	22,875.0000

Figure 61 - Forecast Results

2.6.2.2.2 Forecast Trend

Here, System provides a graphical view of the trend present in the training data set and the forecast values. This is the same view that is shown in the Job summary. System provides facility to zoom out the view by clicking on the zoom out button.

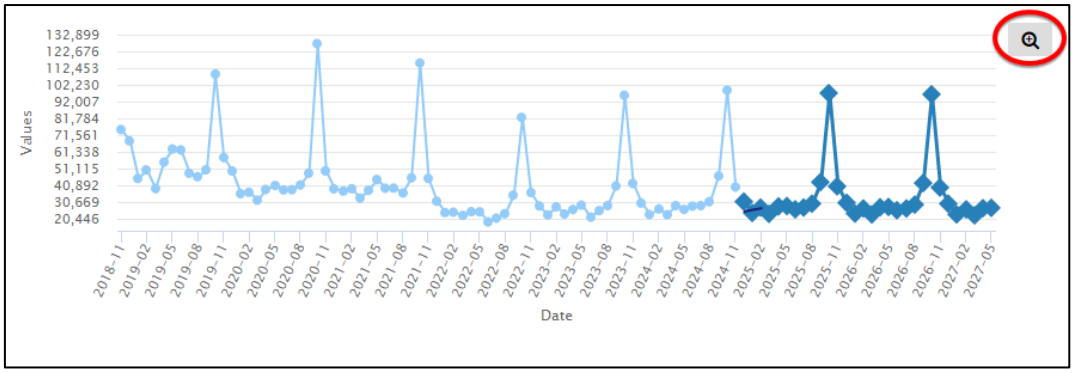


Figure 62 - Forecast Trend - Zoom

2.6.2.2.3 Modeling Metrics

During modeling phase of FPForecast, system generates following metrics, which measure the performance of the models. These values are based on actuals available in training file and the forecasts generated during modelling as part of the modelling process

The definitions of the metrics are present in the section 2.6.2.1.2.4 . The lower the errors, the better is the performance.

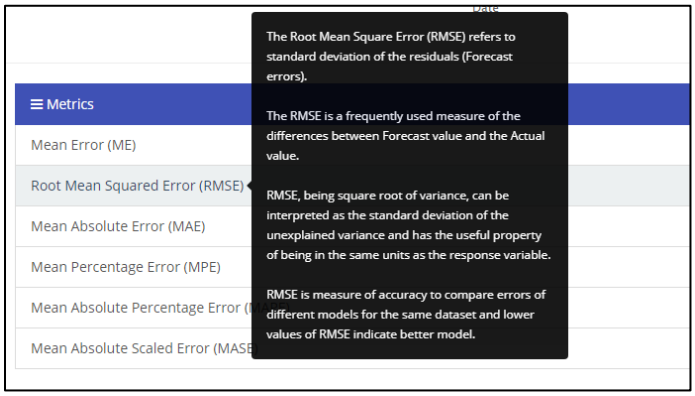


Figure 63 - Metric Definition

The lower the errors, the better is the performance.

2.6.2.2.4 Forecast Metrics

During forecast phase of FPForecast, system generates following metrics, which measure the performance of the forecast. These values are based on actuals available in forecast file and the forecasts generated during modelling as part of the forecast process.

The definitions of the metrics are present in the section 2.6.2.1.2.4 . The lower the errors, the better is the performance.

2.6.2.2.5 Logs

In this Section, system gives a summary of some of the details about the job as follows:

a. **Job Type**: The Job type selected while running the forecast job.
b. **Total records in training file**: Number of records in training file. (Not available for forecast only job)
c. **Total records in forecast file**: Number of records in forecast file. (Not available for modeling only job)
d. **Total modeling time**: Total time system required to create models. (Not available for forecast only job)
e. **Total forecast time**: Total time system required to forecast results based on models created. (Not available for modeling only job)
f. **Total processing time**: Total time system required to complete the job.
g. **Number of important variables**: The variables (columns/headers) from training file used to create the models. This will be sum of both system generated as well as variables present in the dataset.

2.6.2.2.6 Important variable(s)

This section lists the variables (columns/headers) used to create the models. Along with it, system also shows number of System Generated variables using the target variable. The system shows the variables in the order of their importance from highest to lowest.

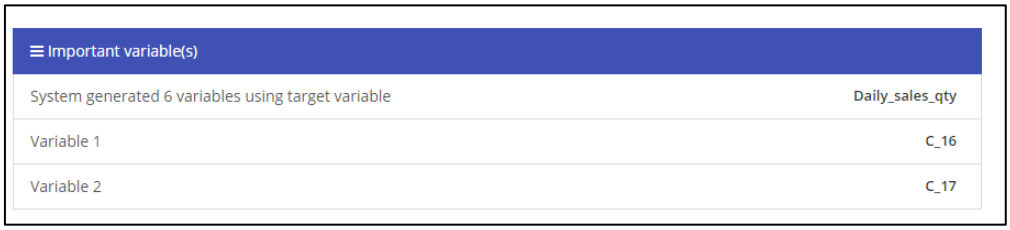


Figure 64 - Important Variable(s)

2.6.2.2.7 Download Template and Upload Actuals

System Allows the users to upload the actual values associated with the time stamp. These values then are used as a method of comparison in Results and Trends.

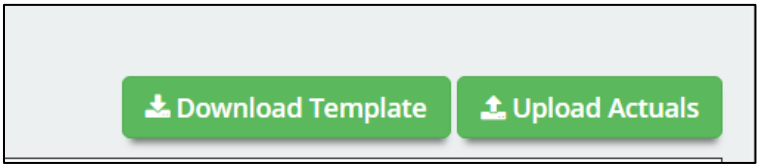


Figure 65 - Download Template & Upload Actuals

Users can download the template by clicking the "Download Template" button. In case users have provided actual values in the forecast file, the downloaded template will contain those actual values, which can be edited by users. Users should not edit the time stamp values.

	A	B
1	FPForecast Report	
2	TimeStamp	Actual Value
3	2018-09-28	
4	2018-09-29	
5	2018-09-30	
6	2018-10-01	

Figure 66 - Upload Actuals Template - Without No Actuals

	A	B
1	FPForecast Report	
2	TimeStamp	Actual Value
3	2018-02-12	
4	2018-02-13	0.3
5	2018-02-14	
6	2018-02-15	0.2
7	2018-02-16	
8	2018-02-17	-0.2
9	2018-02-18	
10	2018-02-19	-1.4
11	2018-02-20	0.2
12	2018-02-21	-0.5
13	2018-02-22	1.8

Figure 67 - Uploads Actuals Template - With Some Actual Values

Users then can upload the filled in template using "Upload Actuals" button.

2.6.2.2.8 Download Results

Allows users to download the Forecast Results in Excel Format by clicking on the "Download Results" button.

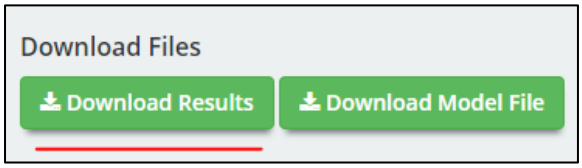


Figure 68 - Download Forecast Results - FPForecast

2.6.2.2.9 Download Models

This allows users to download the models generated for the job so that they can be reused in prediction only Jobs. The button will be not available for prediction only jobs.

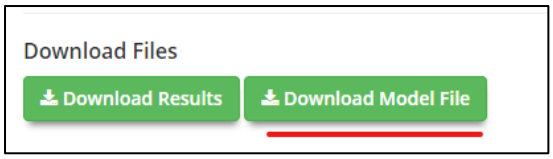


Figure 69 - Download Models - FPForecast

2.10 Understanding Download Result File

This file can be downloaded by clicking the “Download Results” button on the analytics page.

The files have following sheets, based on the type of job:

2.10.1 FPPredict

The result file includes following sheets.

1. **Prediction Info**
2. **Prediction Result**
3. **Predicted Vs Actual** (in case actual values of the target variable are present in the prediction data set, else this sheet will be not be present)
4. **Decile Wise Performance** (This sheet will not be available in case actual values of the target variable are not present in the prediction data set, if the target was binary and prediction was done for single value or the job is categorized as CTV)
5. **Modeling Metrics** (Only available when job is categorized as CTV)
6. **Prediction Metrics** (Only available when job is categorized as CTV)
7. **Model Created**
8. **Model Performance**
9. **Critical Variables**

2.7.1.1 Prediction Info

This sheet provides general information about the job.

	A	B	C
1		Prediction Analysis	
2	Job Name	Loan Prediction	
3	Job Description	Loan Prediction for 20180901 to 20180930	
4	Job Type	modeling + prediction	
5	Modeling mode	Multiple Values	
6	Job Time	00:01:20	
7	Number of models generated	248	
8	Number of models satisfy criteria	247	
9	Number of scenarios of models satisfy criteria	3474	
10	Traning File Name	train_loan 2.csv	
11	Training file number of Record	429	
12	Training file number of Variable	67	
13	Target Variable	loan	
14	Modeling Time	00:00:32	
15	Prediction File Name	predict_loan 2.csv	
16	Prediction file number of Records	185	
17	Prediction file number of Variables	67	
18	Scoring Time	00:00:33	

Figure 70 - Results - Prediction Info - FPPredict

Each column of “prediction info” sheet is explained in table below:

Table description

SL	Column Name	Column Definition	Comment
1	Job Name	Job name given while creating a new job.	
2	Job Description	Job description given while creating a new job.	
3	Job Type	Task which was selected while creating a new job.	
4	Modeling mode	“Less frequent value”: in case that the target variable is binary and selected “Less frequent value” during job creation. “More frequent value”: in case that the target variable is binary and selected “More frequent value” during job creation. Multiple value: In case that the target variable is binary and selected “both value” during job creation or the target variable is not binary.	
5	Job Time	Total time required for the job.	
6	Number of models generated	Number of all models generated on the job.	
8	Training File Name	Name of the training file uploaded during job creation.	Only available for Modeling and Prediction
9	Training file number of Record	Number of records in the training file	
10	Training file number of Variable	Number of variables in the training file.	
11	Target Variable	The variable present in the training file and selected during job creation.	Only available for Modeling and Prediction
12	Modeling Time	Total time taken for modeling	
13	Prediction File Name	Name of prediction file uploaded during job creation.	
14	Prediction file number of Records	Number of records in the prediction file.	
15	Prediction file number of Variables	Number of variables in the prediction file.	
16	Scoring Time	Total time taken for the prediction.	

Table 4 - Results - Prediction Info - FPPredict

2.7.1.2 Prediction Result

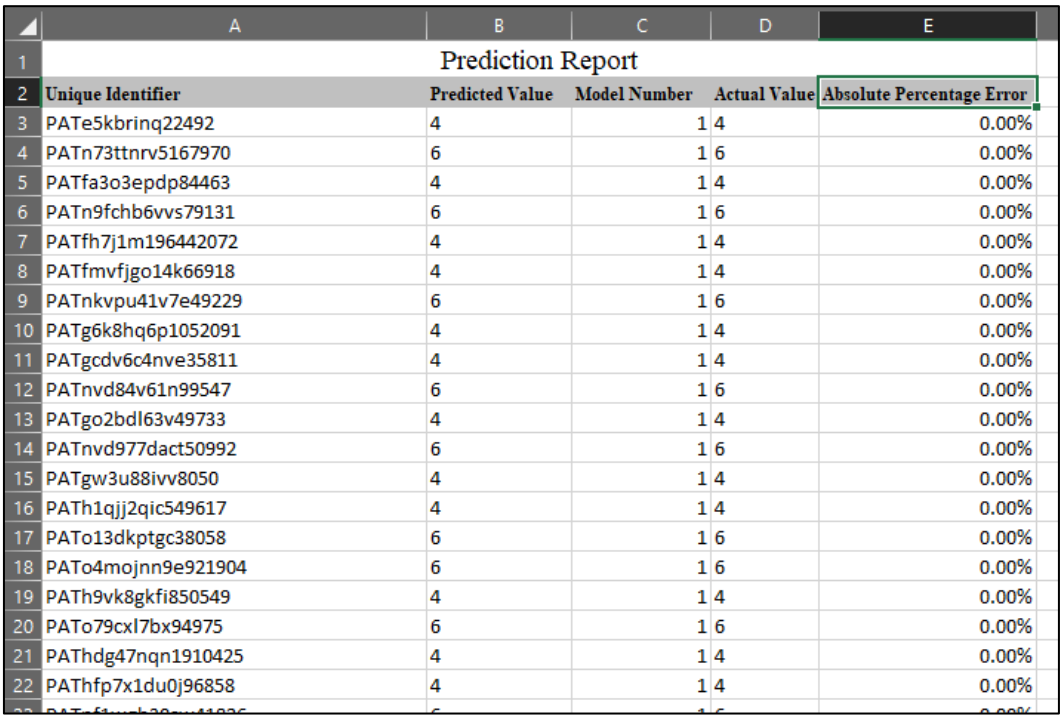
This sheet provides the table of predicted value, model number and probability per “Unique Identifier”. In case actual values of the target variable are present in the prediction data set, “Actual Value” and matching result will also be part of this sheet. (0: is not matched, 1: is matched)

	A	B	C	D	G
1	Prediction Report				
2	Unique Identifier	Predicted Value	Model Number	Probability	Decile
3	PAT92jhtnt30r3408	入社する	9	100.00%	1
4	PATgdhc87at7l44383	入社する	1	100.00%	1
5	PATpu57mgeo6t680	入社する	1	100.00%	1
6	PATiq4dlr20xl25718	入社する	1	100.00%	1
7	PATr6ml5w1ha2278	入社する	1	100.00%	1
8	PATnkvpu41v7e492	入社する	1	100.00%	1
9	PATk6x5x8fer61532	入社する	47	100.00%	1
10	PAT6gura87fh37388	入社する	46	100.00%	1
11	PATfnbof4a4r1506	入社する	1	100.00%	1
12	PATtkgi6fklam52752	入社する	1	100.00%	1
13	PATkxvld12a9g3149	入社する	55	100.00%	1
14	PATc16r82ev587458	入社する	1	100.00%	1
15	PATg2warl2h992928	入社する	1	100.00%	1
16	PATs6j7b183qe8655	入社する	45	100.00%	1
17	PATnvd84v61n99547	入社する	42	100.00%	1

Figure 71 - Results - Prediction Result – FPPredict – Binary Single Value Prediction

	A	B	C	D	E	F
1	Prediction Report					
2	Unique Identifier	Predicted Value	Model Number	Probability	Actual Value	Is Match
3	LP001030	Yes	209	99.97%	Yes	1
4	LP001325	Yes	209	99.97%	Yes	1
5	LP002734	Yes	201	86.4%	Yes	1
6	LP002731	Yes	201	86.4%	Yes	1
7	LP002587	Yes	201	86.4%	Yes	1
8	LP002622	Yes	201	86.4%	Yes	1
9	LP002637	Yes	201	86.4%	Yes	1
10	LP002652	Yes	201	86.4%	Yes	1
11	LP002684	Yes	201	86.4%	Yes	1
12	LP002720	Yes	201	86.4%	Yes	1
13	LP002729	Yes	201	86.21%	Yes	1
14	LP002588	Yes	201	86.21%	Yes	1
15	LP001977	Conditional	10	66.67%	Conditional	1
16	LP002068	Conditional	10	66.67%	Conditional	1
17	LP002103	Conditional	10	66.67%	Conditional	1
18	LP001963	Conditional	10	66.67%	Conditional	1

Figure 72 - Results - Prediction Result - FPPredict - Binary Both Value and Multivalue Prediction DTV



	A	B	C	D	E
1	Prediction Report				
2	Unique Identifier	Predicted Value	Model Number	Actual Value	Absolute Percentage Error
3	PATe5kbrinq22492	4	1	4	0.00%
4	PATn73ttnrv5167970	6	1	6	0.00%
5	PATfa3o3epdp84463	4	1	4	0.00%
6	PATn9fchb6vvs79131	6	1	6	0.00%
7	PATfh7j1m196442072	4	1	4	0.00%
8	PATfmvfjgo14k66918	4	1	4	0.00%
9	PATnkvpu41v7e49229	6	1	6	0.00%
10	PATg6k8hq6p1052091	4	1	4	0.00%
11	PATgcdv6c4nve35811	4	1	4	0.00%
12	PATnvd84v61n99547	6	1	6	0.00%
13	PATgo2bdl63v49733	4	1	4	0.00%
14	PATnvd977dact50992	6	1	6	0.00%
15	PATgw3u88ivv8050	4	1	4	0.00%
16	PATh1qjj2qic549617	4	1	4	0.00%
17	PATo13dkptgc38058	6	1	6	0.00%
18	PATo4mojnn9e921904	6	1	6	0.00%
19	PATh9vk8gkfi850549	4	1	4	0.00%
20	PATo79cxl7bx94975	6	1	6	0.00%
21	PAThdg47nqn1910425	4	1	4	0.00%
22	PAThfp7x1du0j96858	4	1	4	0.00%

Figure 73 - Results - Prediction Result - FPPredict - Multivalue Prediction CTV

Table description

SL	Column Name	Column Definition
1	Unique Identifier	Unique value to identify each record in prediction file and in the column selected as Unique Identifier during job creation.
2	Predicted Value	Predicted value for each identifier.
3	Model Number	An identifier of a model generated by a modeling job and used for prediction for each identifier.
4	Probability	Probability provided from system for each predicted value in case the job is processed as DTV.
5	Actual Value	Actual value for each identifier. (Will not be available for Binary Single Value Prediction)
6	Is Match	Results whether predicted and actual is matched (1) or not (0). (Will not be available for Binary Single Value Prediction)
7	Absolute Percentage Error	Represents the absolute percentage error of the predicted value over actual value. Calculated as ABS((Predicted value - Actual Value) / Actual Value * 100). This value is only available for DTV Use cases.

Table 5 - Results - Prediction Result - FPPredict

At the end of the Sheet, system provides matches (DTV) and Mean Absolute Percentage Errors(CTV).

2.7.1.3 Predicted Vs Actual

This sheet provides comparison of predicted vs actual values. This sheet is available in case actual values of the target variable are present in the prediction data set. The detail will change as following type per target.

1. **Binary and Single value:** in case the target variable is binary and "Less frequent value" or "More frequent value" in selected during job creation, this sheet contains the table with details (details below) along with chart same as defined in section 2.6.1.2.

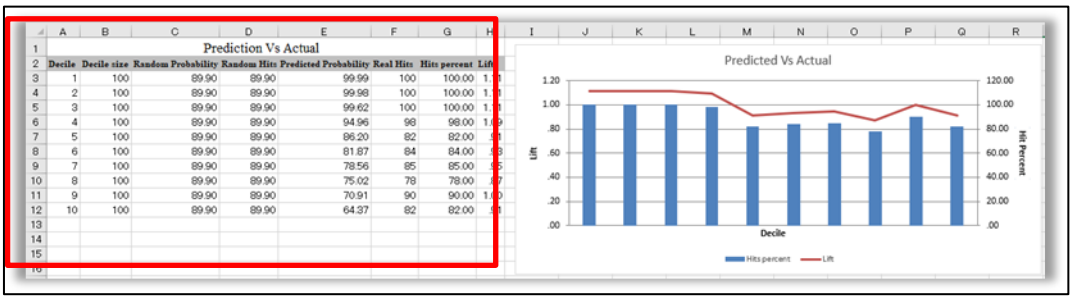


Figure 74 - Results - Prediction vs Actual sheet for Binary and Single value prediction- FPPredict

Table description

SL	Column Name	Column Definition
1	Decile	Decile number. Decile is created by sorting the result by probability in descending order and divide the results in equal 10 parts.
2	Decile Size	Number of records in this decile.
3	Random Probability	Probability of finding the match records in this decile without any prediction.
4	Random Hits	Number of records matching in this decile randomly without any prediction.
5	Predicted Probability	Average probability for this decile using Findability Platform®
6	Real Hits	Number of records matched in this decile using Findability Platform®
7	Hit Percent	% of records matched in this decile using Findability Platform®
8	Lift	Improvement of % of record matched using Findability Platform® vs Random hits.

Table 6 - Results - Prediction vs Actual sheet for Binary and Single value prediction - FPPredict

2. **Numerical:** In case the target variable is numerical and has more than 2 unique values, scatter chart and the table of predicted value, actual value, matching result are provided. (0: is not matched, 1: is matched) This sheet contains the table with details (details below) along with chart same as defined in section 2.6.1.2.

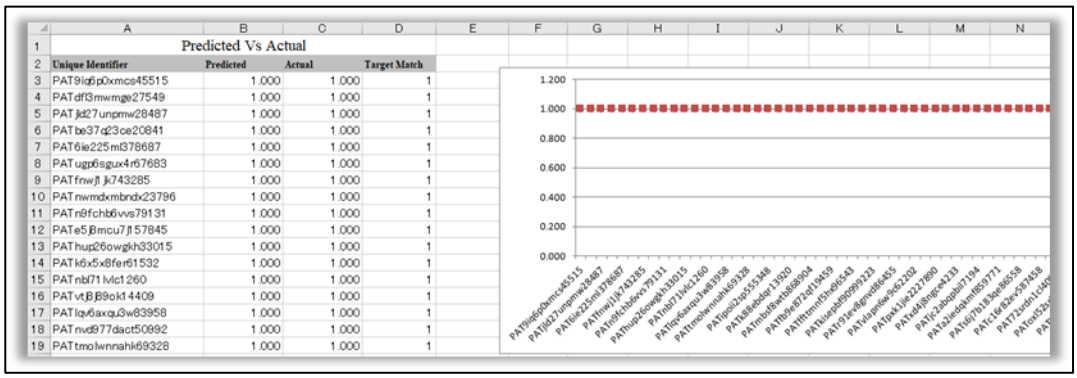


Figure 75 - Results - Prediction Vs Actual (Scatter chart) - FPPredict

Table description

SL	Column Name	Column Definition
1	Unique Identifier	Unique value to identify each record in prediction file and in the column selected as Unique Identifier during job creation.
2	Predicted	Predicted value for each identifier.
3	Actual	Actual value for each identifier.
4	Target Match	Results whether predicted and actual is matched (1) or not (0).

Table 7 - Results - Prediction Vs Actual (Scatter chart) - FPPredict

3. **Categorical or Binary (Both Values):** in case the target variable has 2 or more alphabetical values or the job is run as Binary with both values , this sheet provides cross-tabulation table for actual vs predicted as same result of heat map explained in section 2.6.1.2. This table shows matched number of predicted vs actual for each target variable along with Precision, Recall, Overall Accuracy and AUC.

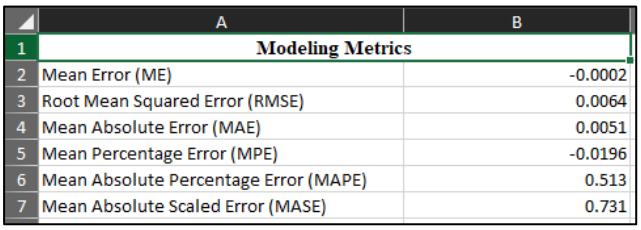
Predicted Vs Actual							
		Actual					
		Conditional	investigate	No	Yes	Precision(%)	AUC ROC
Predicted	Yes	7	0	0	24	77.42	0.80986
	No	2	0	4	0	66.67	0.52535
	investigate	0	1	0	0	100	1
	Conditional	79	0	56	12	53.74	0.59833
	Recall （%）	89.77	100	6.67	66.67		
Overall Accuracy （%）	58.38						

Figure 76 - Results - Predicted Vs Actual (Cross-tabulation) - FPPredict

2.7.1.4 Modeling Metrics

This sheet provides metrics generated during the modeling phase of the job. These metrics measure the performance of the models, the lower the value, the better performing are the models. The descriptions of these metrics are provided in section 2.6.2.1.2.4

This sheet is only available when the job is categorized as a CTV job



	A	B
1	**Modeling Metrics**	
2	Mean Error (ME)	-0.0002
3	Root Mean Squared Error (RMSE)	0.0064
4	Mean Absolute Error (MAE)	0.0051
5	Mean Percentage Error (MPE)	-0.0196
6	Mean Absolute Percentage Error (MAPE)	0.513
7	Mean Absolute Scaled Error (MASE)	0.731

Figure 77 - Modeling Metrics sheet for CTV prediction - FPPredict

Table description

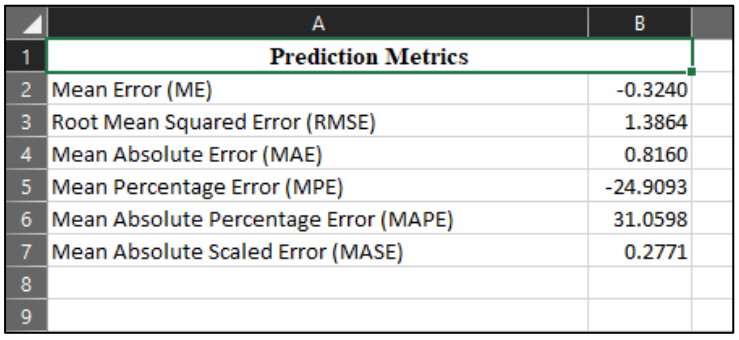
SL	Column Name	Column Definition
1	Mean Error (ME)	ME modeling metric value.
2	Root Mean Squared Error (RMSE)	RMSE modeling metric value.
3	Mean Absolute Error (MAE)	RMSE modeling metric value.
4	Mean Percentage Error (MAE)	MAE modeling metric value.
5	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value.
6	Mean Absolute Scaled Error (MASE)	MASE modeling metric value.

Table 8 - Results - Modeling Metrics sheet for CTV prediction - FPPredict

2.7.1.5 Prediction Metrics

This sheet provides metrics generated during the prediction phase of the job. These metrics measure the performance of the prediction, the lower the value, the better performing is the prediction. The descriptions of these metrics are provided in section 2.6.2.1.2.4 .

This sheet is only available when the job is categorized as a CTV job



	A	B
1	**Prediction Metrics**	
2	Mean Error (ME)	-0.3240
3	Root Mean Squared Error (RMSE)	1.3864
4	Mean Absolute Error (MAE)	0.8160
5	Mean Percentage Error (MPE)	-24.9093
6	Mean Absolute Percentage Error (MAPE)	31.0598
7	Mean Absolute Scaled Error (MASE)	0.2771
8		
9		

Figure 78 - Prediction Metrics sheet for CTV prediction - FPPredict

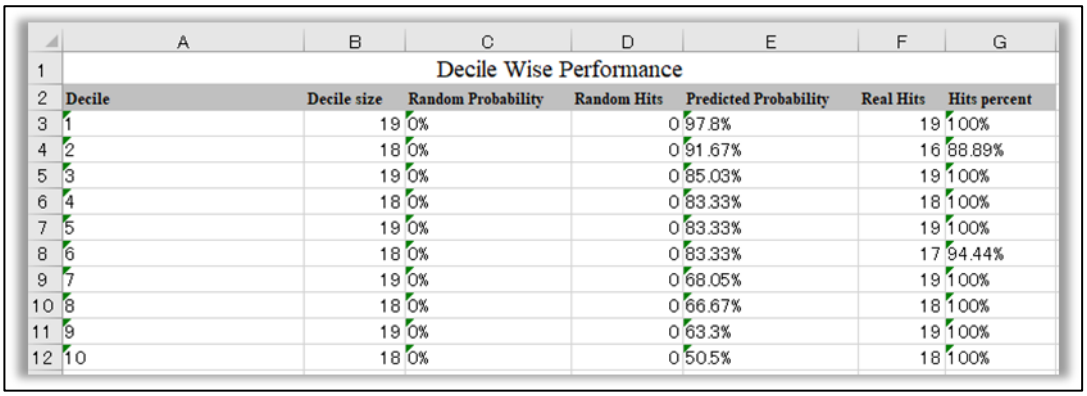
Table description

SL	Column Name	Column Definition
1	Mean Error (ME)	ME modeling metric value.
2	Root Mean Squared Error (RMSE)	RMSE modeling metric value.
3	Mean Absolute Error (MAE)	RMSE modeling metric value.
4	Mean Percentage Error (MAE)	MAE modeling metric value.
5	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value.
6	Mean Absolute Scaled Error (MASE)	MASE modeling metric value.

Table 9 - Results - Prediction Metrics sheet for CTV prediction - FPPredict

2.7.1.6 Decile Wise Performance

This sheet provides decile wise performance with the table of random probability, random hits, predicted probability, real hits, hit in case the job is processed as DTV job. This sheet is available in case actual values of the target variable are present in the prediction data set for a DTV job. This will not be available for jobs processed as CTV jobs.



	A	B	C	D	E	F	G
1	Decile Wise Performance						
2	Decile	Decile size	Random Probability	Random Hits	Predicted Probability	Real Hits	Hits percent
3	1	19	0%	0	97.8%	19	100%
4	2	18	0%	0	91.67%	16	88.89%
5	3	19	0%	0	85.03%	19	100%
6	4	18	0%	0	83.33%	18	100%
7	5	19	0%	0	83.33%	19	100%
8	6	18	0%	0	83.33%	17	94.44%
9	7	19	0%	0	68.05%	19	100%
10	8	18	0%	0	66.67%	18	100%
11	9	19	0%	0	63.3%	19	100%
12	10	18	0%	0	50.5%	18	100%

Figure 79 - Results - Decile Wise Performance- FPPredict

Table description

SL	Column Name	Column Definition
1	Decile	Decile number. Decile is created by sorting the result by probability in descending order and divide the results in equal 10 parts.
2	Decile Size	Number of records in this decile.
3	Random Probability	Probability of finding the match records in this decile without any prediction.
4	Random Hits	Number of records matching in this decile randomly without any prediction.
5	Predicted Probability	Average probability for this decile using Findability Platform®
6	Real Hits	Number of records matched in this decile using Findability Platform®
7	Hit Percent	% of records matched in this decile using Findability Platform®

Table 10 - Results - Decile Wise Performance - FPPredict

2.7.1.7 Models

This sheet provides table and bar chart which is same as defined in section 2.6.1.3.

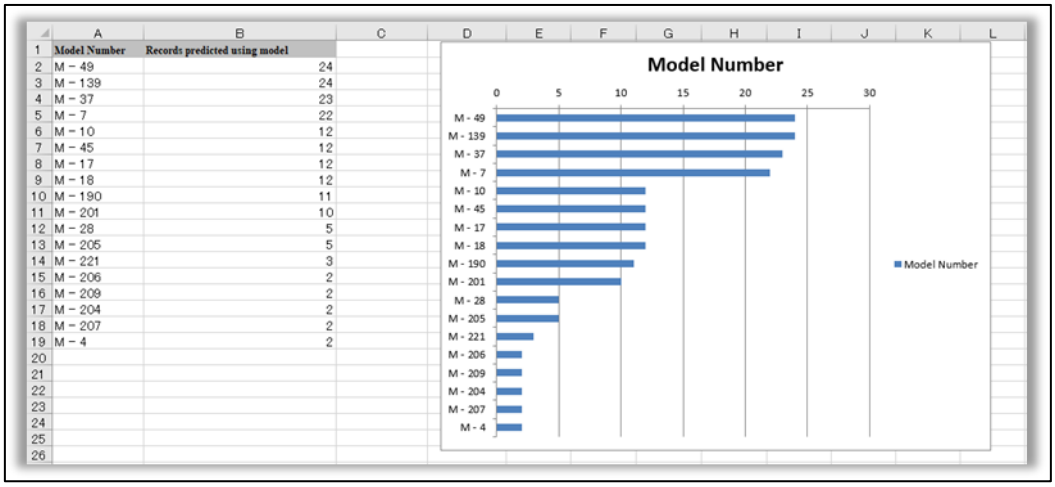


Figure 80 - Results - Model Created - FPPredict

Table description

SL	Column Name	Column Definition
1	Model Number	An identifier of a model generated by a modeling job and used for prediction.
2	Records predicted using model	Number of records predicted using this model.

Table 11 - Results - Model Created - FPPredict

2.7.1.8 Variables

This sheet provides table and bar chart which is same as defined in section 2.6.1.4.

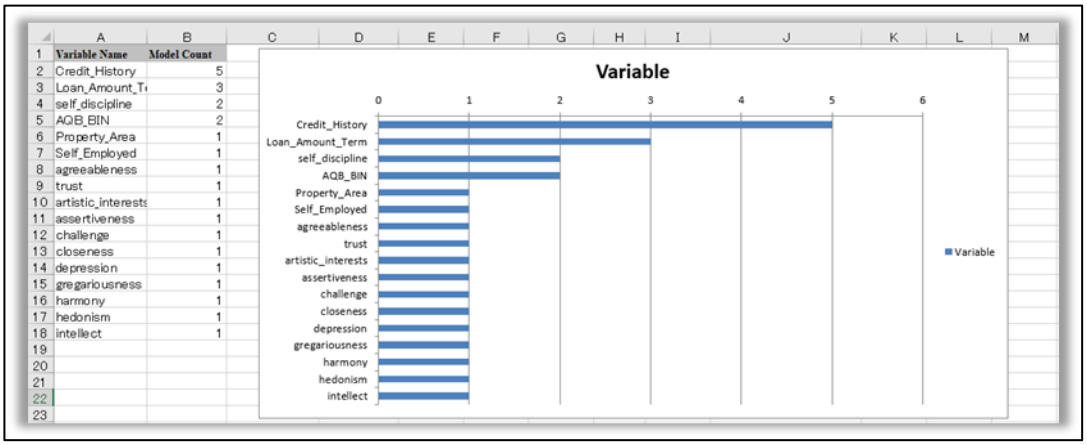


Figure 81 - Results - Model Performance - FPPredict

Table description

SL	Column Name	Column Definition

1	Variable Name	Name of variable (column name) which used for prediction.
2	Model Count	Number of models used per variable name.

Table 12 - Results - Model Performance - FPPredict

2.7.1.9 Variables of Models

This sheet provides table which is same as defined in section 2.6.1.5. For CTV jobs the system shows the variables in the order of their importance from highest to lowest. For DTV jobs, the all the variables have same importance.

	A	B	C	D	E
1	**Model Number**	**Variable 1**	**Variable 2**	**Variable 3**	**Variable 4**
2	204	agreeableness	AQB_BIN	Credit_History	Loan_Amount_Term
3	206	Credit_History	Loan_Amount_Term	Property_Area	
4	201	Credit_History	harmony		
5	207	AQB_BIN	Credit_History		
6	4	hedonism			
7	7	trust			
8	10	challenge			
9	17	gregariousness			
10	18	self_discipline			
11	28	self_discipline			
12	37	artistic_interests			
13	45	closeness			
14	49	assertiveness			
15	139	depression			
16	190	intellect			
17	205	Credit_History			
18	209	Loan_Amount_Term			
19	221	Self_Employed			

Figure 82 - Results - Critical Variables - FPPredict

Table description

SL	Column Name	Column Definition
1	Model Number	An identifier of a model generated by a modeling job and used for prediction.
2	Variable 1 - N	Name of variables involved in this model.

Table 13 - Results - Critical Variable - FPPredict

2.10.2 FPForecast

The result file includes following sheets.

1. **Forecast Info**
2. **Forecast Result**
3. **Forecast Vs Actual** (in case actual values of the target variable are present either in prediction dataset or were uploaded via upload actuals functionality, this sheet will be present in the report)
4. **Modeling Metrics**
5. **Forecast Metrics**
6. **Important variables** (in case there are no variables used besides target variable, this sheet will be present in the report)
7. **Variables** (Only in case system detects subgroups in system)
8. **Variables of Models** (Only in case system detects subgroups in system)

2.7.2.1 Forecast Info

This sheet provides general information about the forecast job. In case there are no subgroups detected the format will be as follows:

	A	B	C
1	Forecast Analysis		
2	Job Name	testp	
3	Job Description	testp	
4	Job Type	Forecast	
5	Training File Name	-	
6	Total Records in Training File	-	
7	Total Modeling Time	-	
8	Forecast File Name	daily_sales_qty_prediction (1).csv	
9	Total records in Forecast file	685	
10	Total Forecast Time	00:00:02	
11	Total Processing Time	00:00:18	
12	Number of Important Variable(s)	11	
13			

Figure 83 - Forecast Info sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Job name	Job name provided during creation of the job
2	Job Description	Job description provided during creation the job
3	Job Type	Type of the FPForecast job. This can be one of the following: 1. “Modeling and Forecast” 2. “Modeling” 3. “Forecast”
4	Training File name	Name of the file that was used as the training dataset
5	Total Records in the Training File	Number of the rows present in the training file except the headers.
6	Total Modeling Time	Total time taken by the system to create the models. This is shown in “hh:mm:ss” format.
7	Forecast File name	Name of the file that was used as the forecast dataset
8	Total Records in the Forecast File	Number of the rows present in the forecast file except the headers.
9	Total Forecast Time	Total time taken by the system for forecast. This is shown in “hh:mm:ss” format.

10	Total Processing time	Total time taken by the system to complete the job. This is shown in “hh:mm:ss” format.
11	Number of Important variables	Shows the number of variables besides the target variable which were used for creating the models as well as number of System Generated Variables generated using target variable, if available.

Table 14 - Forecast Info Sheet Description – FPForecast

On the other hand, if system detects subgroups, the format of the sheet is as follows

	A	B	C
1	Forecast Analysis		
2	Job Name	Sample forecast	
3	Job Description	test	
4	Job Type	Modeling And Forecast	
5	Training File Name	sample_training_file.csv	
6	Total Records in Training File	367	
7	Total Modeling Time	00:00:23	
8	Forecast File Name	sample_forecast_file.csv	
9	Total records in Forecast file	20	
10	Total Forecast Time	00:00:03	
11	Total Processing Time	00:04:47	
12	Total number of models created	7	
13	Number of groups identified in training file	12	
14	Number of models applied	3	
15	Number of groups forecast in forecast file	3	
16			

Figure 84 - Forecast Analysis - MultiModel Forecast

Table description

SL	Column Name	Column Definition
1	Job name	Job name provided during creation of the job
2	Job Description	Job description provided during creation the job
3	Job Type	Type of the FPForecast job. This can be one of the following: 1. “Modeling and Forecast” 2. “Modeling” 3. “Forecast”
4	Training File name	Name of the file that was used as the training dataset
5	Total Records in the Training File	Number of the rows present in the training file except the headers.
6	Total Modeling Time	Total time taken by the system to create the models. This is shown in “hh:mm:ss” format.
7	Forecast File name	Name of the file that was used as the forecast dataset
8	Total Records in the Forecast File	Number of the rows present in the forecast file except the headers.
9	Total Forecast Time	Total time taken by the system for forecast. This is shown in “hh:mm:ss” format.
10	Total Processing time	Total time taken by the system to complete the job. This is shown in “hh:mm:ss” format.
11	Total number of models created	Shows the number of model created from the training data received.
12	Number of groups identified in training file	Represents the number of groups identified based on the group identifier provided during job submission

13	Number of models applied	The number of models that were used for forecast.
14	Number of groups forecast in forecast file	The number of groups present in forecast dataset whose forecast process completed successfully.

Table 15 - Forecast Info Sheet Description - MultiModel Forecast

2.7.2.2 Forecast Result

This sheet provides results of the forecast job, providing forecast value corresponding to each interval present in the forecast file. There is visual presentation available as well in format of line chart showing the trend between values of target variable present in training set and forecast values, in case when system doesn't detect subgroups.

In case system doesn't recognizes subgroups in the dataset, the sheet is shown as below.

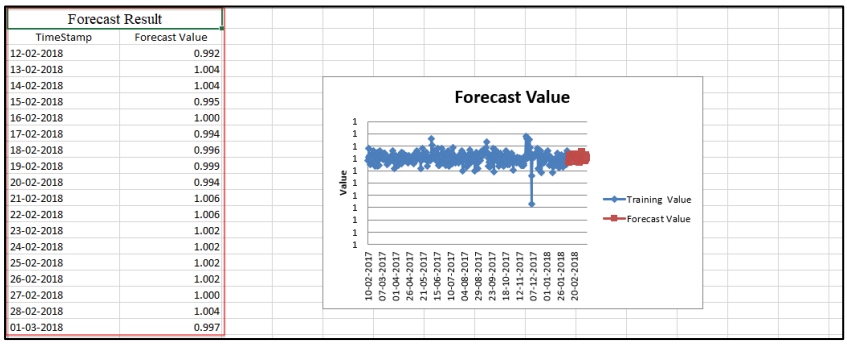


Figure 85 - Forecast Result sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Timestamp	The values present timestamp column of the forecast dataset, sorted in ascending order.
2	Forecast Value	Forecast values for each interval present in the forecast dataset.

Table 16 - Forecast Result Sheet Description - FPForecast

In case system recognizes subgroups in the dataset, the sheet is shown as below.

	A	B	C	D
1	Forecast Result			
2	Group ID	TimeStamp	Model Number	Forecast Value
3	10	01-08-2019	1	37,777
4	10	02-08-2019	1	42,715
5	10	03-08-2019	1	17,588
6	10	04-08-2019	1	9,446
7	10	05-08-2019	1	50,460
8	10	06-08-2019	1	31,880
9	10	07-08-2019	1	41,888
10	10	08-08-2019	1	46,555
11	10	09-08-2019	1	37,879
12	10	10-08-2019	1	25,110
13	10	11-08-2019	1	14,631
14	10	12-08-2019	1	42,784
15	10	13-08-2019	1	21,813
16	10	14-08-2019	1	39,633
17	10	15-08-2019	1	29,940
18	10	16-08-2019	1	38,519
19	10	17-08-2019	1	27,376
20	10	18-08-2019	1	16,399
21	10	19-08-2019	1	40,932
22	10	20-08-2019	1	27,363
23	10	21-08-2019	1	41,759
24	10	22-08-2019	1	36,587
25	10	23-08-2019	1	39,602

Figure 86 - Forecast Result sheet - MultiModel Forecast

Table description

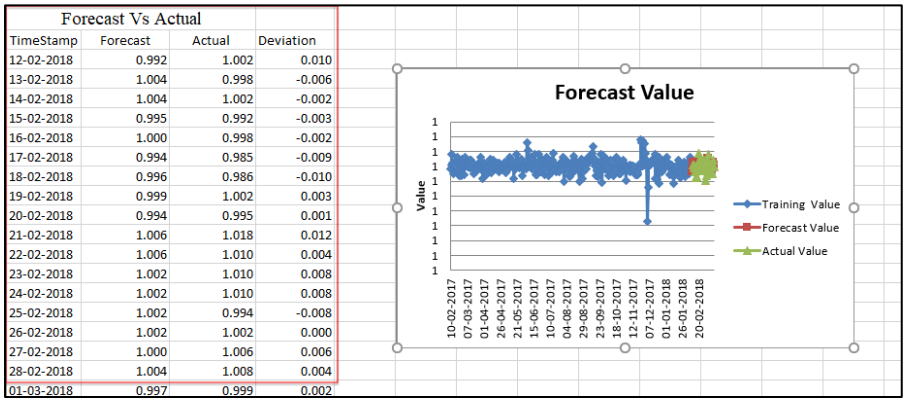
SL	Column Name	Column Definition
1	Group ID	Value present under column identified as “Group Identifier” during job submission.
2	Timestamp	The values present timestamp column of the forecast dataset, sorted in ascending order.
3	Model Number	Represents the model which was used to forecast the value associated with the record
4	Forecast Value	Forecast values for each interval present in the forecast dataset.

Table 17 - Forecast Result Sheet Description - MultiModel Forecast

2.7.2.3 Forecast vs Actual

In the case there were actual values present in the forecast dataset or user upload actuals using “Upload Actuals” option in the UI, this sheet will provide a comparison between forecast and actual values, along with deviation with the actual value. There is visual presentation available as well in format of line chart showing the trend between values of target variable present in training set, forecast values and actual values, in case when system doesn’t detect subgroups.

When system doesn’t detect subgroups, the representation is as follows



Forecast Vs Actual

TimeStamp	Forecast	Actual	Deviation
12-02-2018	0.992	1.002	0.010
13-02-2018	1.004	0.998	-0.006
14-02-2018	1.004	1.002	-0.002
15-02-2018	0.995	0.992	-0.003
16-02-2018	1.000	0.998	-0.002
17-02-2018	0.994	0.985	-0.009
18-02-2018	0.996	0.986	-0.010
19-02-2018	0.999	1.002	0.003
20-02-2018	0.994	0.995	0.001
21-02-2018	1.006	1.018	0.012
22-02-2018	1.006	1.010	0.004
23-02-2018	1.002	1.010	0.008
24-02-2018	1.002	1.010	0.008
25-02-2018	1.002	0.994	-0.008
26-02-2018	1.002	1.002	0.000
27-02-2018	1.000	1.006	0.006
28-02-2018	1.004	1.008	0.004
01-03-2018	0.997	0.999	0.002

Figure 87 - Forecast vs Actual sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Timestamp	The values present timestamp column of the forecast dataset, sorted in ascending order.
2	Forecast	Forecast values for each interval present in the forecast dataset.
3	Actual	Actual value provided by the users for the interval
4	Deviation	Difference of the forecast value from the actual value (Actual Value – Forecast Value).

Table 18 - Forecast vs Actual Description - FPForecast

When system detects subgroups, the representation is as follows

	A	B	C	D	E	F
1	Forecast Vs Actual					
2	Group ID	TimeStamp	Forecast	Actual	Deviation	
3	7	20-06-2017	1.003	1.008	0.005	
4	7	21-06-2017	1.005	1.000	-0.005	
5	7	22-06-2017	1.000	0.999	-0.001	
6	8	24-06-2017	1.002	1.002	0.000	
7	8	25-06-2017	1.008	0.995	-0.013	
8	8	26-06-2017	1.005	1.018	0.013	
9	8	27-06-2017	1.004	1.010	0.006	
10	8	28-06-2017	1.004	1.010	0.006	
11	8	29-06-2017	1.000	1.010	0.010	
12	8	30-06-2017	0.997	0.994	-0.003	
13	8	01-07-2017	0.994	1.002	0.008	
14	8	02-07-2017	1.000	1.006	0.006	
15	8	03-07-2017	0.997	1.008	0.011	
16	9	13-07-2017	0.998	1.002	0.004	
17	9	14-07-2017	1.000	0.998	-0.002	
18	9	15-07-2017	0.991	1.002	0.011	
19	9	16-07-2017	0.992	0.992	0.000	
20	9	17-07-2017	1.001	0.998	-0.003	
21	9	18-07-2017	0.994	0.985	-0.009	
22	9	19-07-2017	0.992	0.986	-0.006	
23						

Figure 88 - Forecast vs Actual sheet - MultiModel Forecast

Table description

SL	Column Name	Column Definition
1	Group ID	Value present under column identified as “Group Identifier” during job submission.
2	Timestamp	The values present timestamp column of the forecast dataset, sorted in ascending order.
3	Forecast	Forecast values for each interval present in the forecast dataset.
4	Actual	Actual value provided by the users for the interval
5	Deviation	Difference of the forecast value from the actual value (Actual Value – Forecast Value).

Table 19 - Forecast vs Actual Description - MultiModel Forecast

2.7.2.4 Modeling Metrics

This sheet provides metrics generated during the modeling phase of the job. These metrics measure the performance of the models, the lower the value, the better performing are the models. The descriptions of these metrics are provided in section 2.6.2.1.2.4

When system doesn’t detect subgroups, the representation is as follows

	A	B
1	**Modeling Metrics**	
2	Mean Error (ME)	-0.0002
3	Root Mean Squared Error (RMSE)	0.0064
4	Mean Absolute Error (MAE)	0.0051
5	Mean Percentage Error (MPE)	-0.0196
6	Mean Absolute Percentage Error (MAPE)	0.513
7	Mean Absolute Scaled Error (MASE)	0.731

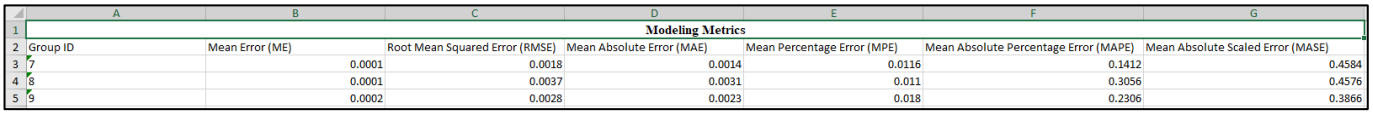
Figure 89 - Metrics sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Mean Error (ME)	ME modeling metric value.
2	Root Mean Squared Error (RMSE)	RMSE modeling metric value.
3	Mean Absolute Error (MAE)	RMSE modeling metric value.
4	Mean Percentage Error (MAE)	MAE modeling metric value.
5	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value.
6	Mean Absolute Scaled Error (MASE)	MASE modeling metric value.

Table 20 - Modeling Metrics Sheet Description - FPForecast

When system detect subgroups, the representation is as follows



	A	B	C	D	E	F	G
1	Modeling Metrics						
2	Group ID	Mean Error (ME)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	Mean Percentage Error (MPE)	Mean Absolute Percentage Error (MAPE)	Mean Absolute Scaled Error (MASE)
3	7	0.0001	0.0018	0.0014	0.0116	0.1412	0.4584
4	8	0.0001	0.0037	0.0031	0.011	0.3056	0.4576
5	9	0.0002	0.0028	0.0023	0.018	0.2306	0.3866

Figure 90 - Modeling Metrics sheet - MultiModel Forecast

Table description

SL	Column Name	Column Definition
1	Group ID	Value present under column identified as “Group Identifier” during job submission.
2	Mean Error (ME)	ME modeling metric value for the model used for forecasting the subgroup associated with group ID.
3	Root Mean Squared Error (RMSE)	RMSE modeling metric value for the model used for forecasting the subgroup associated with group ID.
4	Mean Absolute Error (MAE)	RMSE modeling metric value for the model used for forecasting the subgroup associated with group ID.
5	Mean Percentage Error (MAE)	MAE modeling metric value for the model used for forecasting the subgroup associated with group ID.
6	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value for the model used for forecasting the subgroup associated with group ID.
7	Mean Absolute Scaled Error (MASE)	MASE modeling metric value for the model used for forecasting the subgroup associated with group ID.

Table 21 - Metric Sheet Description - MultiModel Forecast

2.7.2.5 Forecast Metrics

This sheet provides metrics generated during the forecast phase of the job. These metrics measure the performance of the models, the lower the value, the better performing are the models. The descriptions of these metrics are provided in section 2.6.2.1.2.4

When system doesn't detect subgroups, the representation is as follows

	A	B
1	**Forecast Metrics**	
2	Mean Error (ME)	-0.0015
3	Root Mean Squared Error (RMSE)	0.0061
4	Mean Absolute Error (MAE)	0.0047
5	Mean Percentage Error (MPE)	-0.1576
6	Mean Absolute Percentage Error (MAPE)	0.4671
7	Mean Absolute Scaled Error (MASE)	0.6402

Figure 91 - Forecast Metrics sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Mean Error (ME)	ME modeling metric value.
2	Root Mean Squared Error (RMSE)	RMSE modeling metric value.
3	Mean Absolute Error (MAE)	RMSE modeling metric value.
4	Mean Percentage Error (MAE)	MAE modeling metric value.
5	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value.
6	Mean Absolute Scaled Error (MASE)	MASE modeling metric value.

Table 22 - Modeling Metrics Sheet Description - FPForecast

When system detect subgroups, the representation is as follows

	A	B	C	D	E	F	G
1	**Modeling Metrics**						
2	Group ID	Mean Error (ME)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)	Mean Percentage Error (MPE)	Mean Absolute Percentage Error (MAPE)	Mean Absolute Scaled Error (MASE)
3	7	0.0001	0.0018	0.0014	0.0116	0.1412	0.4584
4	8	0.0001	0.0037	0.0031	0.011	0.3056	0.4576
5	9	0.0002	0.0028	0.0023	0.018	0.2306	0.3866

Figure 92 - Metrics sheet - MultiModel Forecast

Table description

SL	Column Name	Column Definition
1	Group ID	Value present under column identified as “Group Identifier” during job submission.
2	Mean Error (ME)	ME modeling metric value for the model used for forecasting the subgroup associated with group ID.
3	Root Mean Squared Error (RMSE)	RMSE modeling metric value for the model used for forecasting the subgroup associated with group ID.
4	Mean Absolute Error (MAE)	RMSE modeling metric value for the model used for forecasting the subgroup associated with group ID.
5	Mean Percentage Error (MAE)	MAE modeling metric value for the model used for forecasting the subgroup associated with group ID.
6	Mean Absolute Percentage Error (MAPE)	MAPE modeling metric value for the model used for forecasting the subgroup associated with group ID.
7	Mean Absolute Scaled Error (MASE)	MASE modeling metric value for the model used for forecasting the subgroup associated with group ID.

Table 23 - Metric Sheet Description - MultiModel Forecast

2.7.2.6 Important Variable(s)

This sheet provides a list of the variables present in the training set which were used to create the models besides the target variable. The system shows the variables in the order of their importance from highest to lowest.

This sheet is only available when system doesn’t detect any subgroups in the training dataset.

	A	B	C
1	Important Variables		
2	Serial No	Variable	
3	1	System generated 6 variables using target variable	
4	2	ar1	
5	3	ar2	
6	4	ar3	
7	5	ar4	
8	6	ar5	
9			
10			

Figure 93 - Important Variable(s) sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Serial No	Serial number
2	Variable	Variable names present in the training set as well as number of System Generated Variables generated using target variable.

Table 24 - Important Variable(s) Description - FPForecast

2.7.2.7 Models

This sheet provides a models used for forecast along with the number of records it was used for forecast.

This sheet is only available when system detects subgroups in the training dataset.

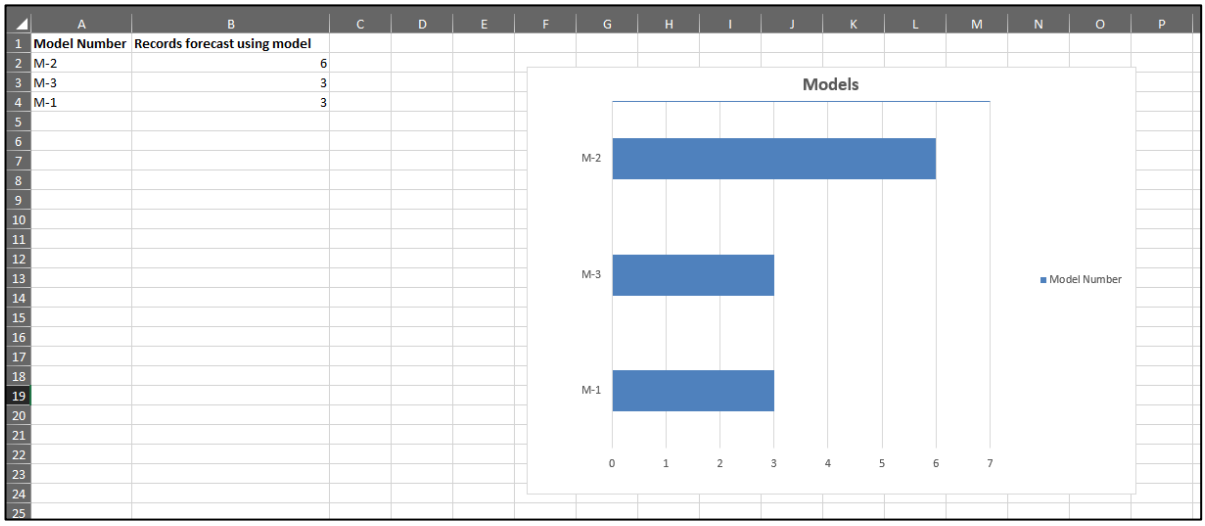


Figure 94 - Models sheet - FPForecast

Table description

SL	Column Name	Column Definition
1	Model Number	
2	Records Forecast using model	Number of records predicted using the model

Table 25 - Models Description - FPForecast

2.7.2.8 Variables

This sheet represents the contribution of the variables in the applicable models for MultiModel forecast. The results are shown in 3 X 3 Matrix, with Rows representing the number of times the variable is part of the models and Columns represent Impact (Weight) of the variable in the models. These factored are classified into 3 categories of Low, Medium and High.

This is same as described in section 2.6.2.1.1.3

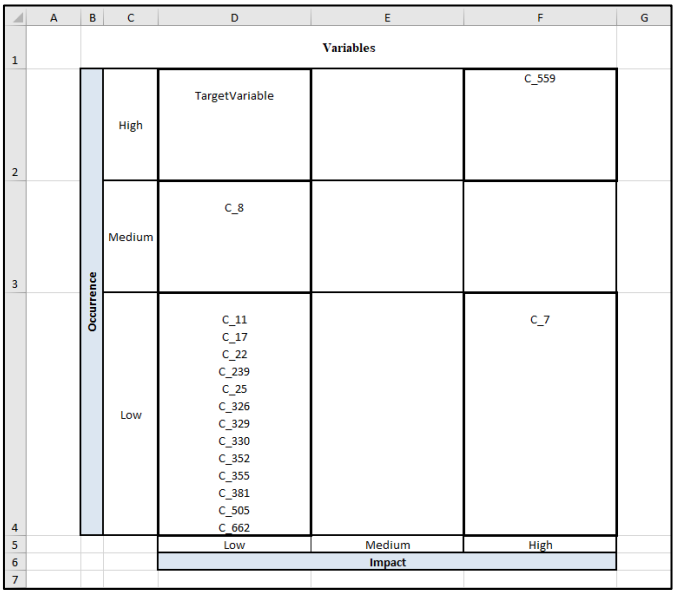


Figure 95 - Variables Sheet

2.7.2.9 Variables of Models

This sheet provides information regarding the variables associated with each applied model. This sheet will only be available in case of MultiModel Forecast.

The sheet will be shown as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Group ID	Model Number	Variable1	Variable2	Variable3	Variable4	Variable5	Variable6	Variable7	Variable8	Variable9	Variable10	
2	8	M-2	C_381	C_8	C_326	C_355	C_559	C_329	C_25	C_352	C_22	TargetVariable	
3	9	M-3	C_11	C_662	C_8	C_330	C_239	C_505	TargetVariable				
4	7	M-1	C_17	C_7	TargetVariable								
5													

Figure 96 - Variables of Models - MultiModel Forecast

Table description

SL	Column Name	Column Definition
1	Group ID	Value present under column identified as “Group Identifier” during job submission.
1	Model Number	Identifier of the applied model used for the Forecast of the group.
2	Variable1 .. VariableN	Variable associated with the given model.

Table 26 - Important Variable(s) Description - MultiModel Forecast

2.11 Dataset Management

System allows users to manage the datasets to be used for FPPredict and FPForecast jobs using this module. The datasets uploaded here are available under option “Dataset location cloud”, when starting a new job. The features of this module are as follows:

2.11.1 Add Files

System allows users to upload datasets and models archive using “+ Files” button present at the top of the screen.

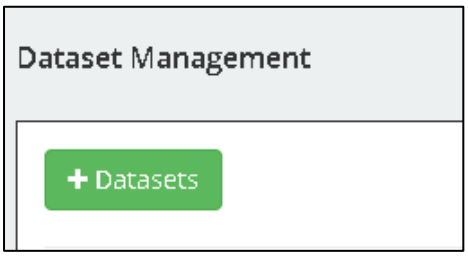


Figure 97 - Add Datasets

On clicking the screen, system provides users with a screen where users can upload datasets which are to be used while submitting jobs. The datasets can be uploaded by

1. Dragging and Dropping files
2. Click Browse and selecting files.

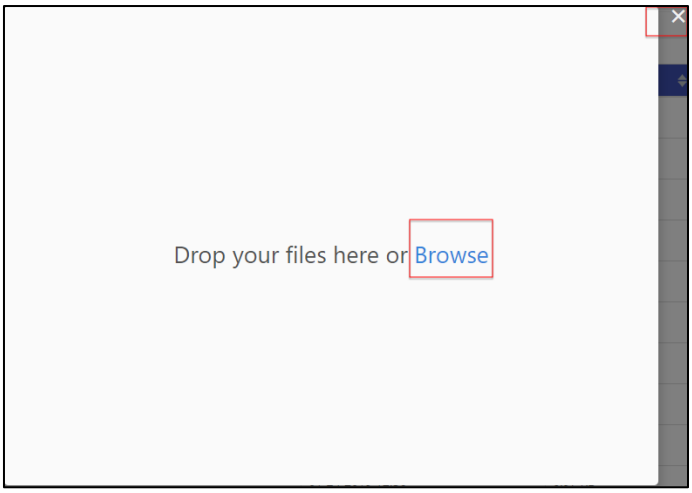


Figure 98 - Upload Files Screen

Multiple files can be uploaded at a time. Only files of the extensions “.csv” and “.models” can be uploaded.

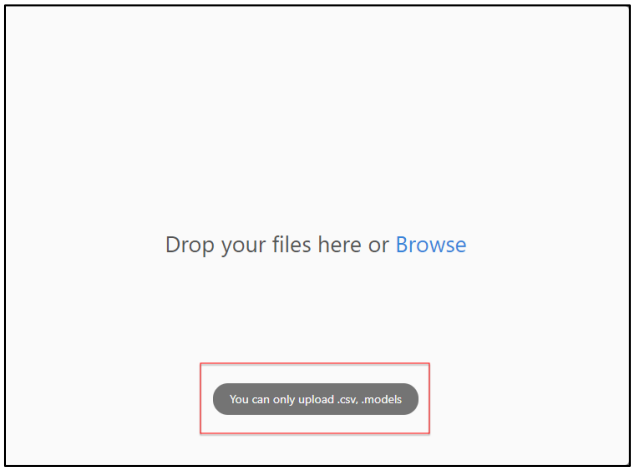


Figure 99 - File type error

During upload system shows progress of the upload to the users.

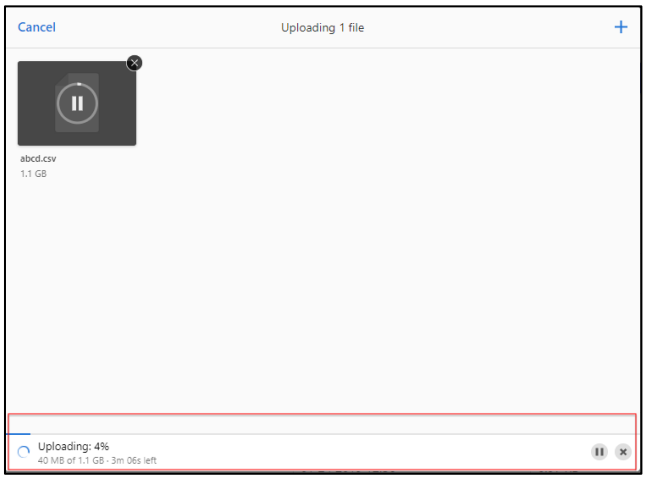


Figure 100 - Upload Progress

In case the file upload fails due to internet connection or browser crash, the file upload resumes from part where upload failed.

2.11.2 Manage Files

2.8.2.1 View datasets

System allows users to view the datasets and models archive uploaded. The information shown is as follows:

1. File name
2. Upload date and time
3. File size

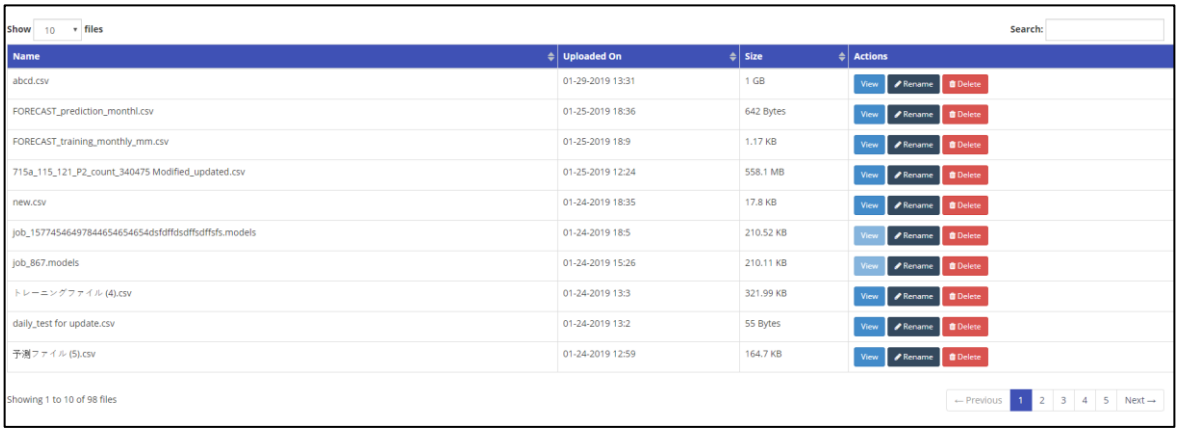


Figure 101 - Datasets

2.8.2.2 Sort dataset list

System allows users to sort the list based on File name, Upload date & time and File size.

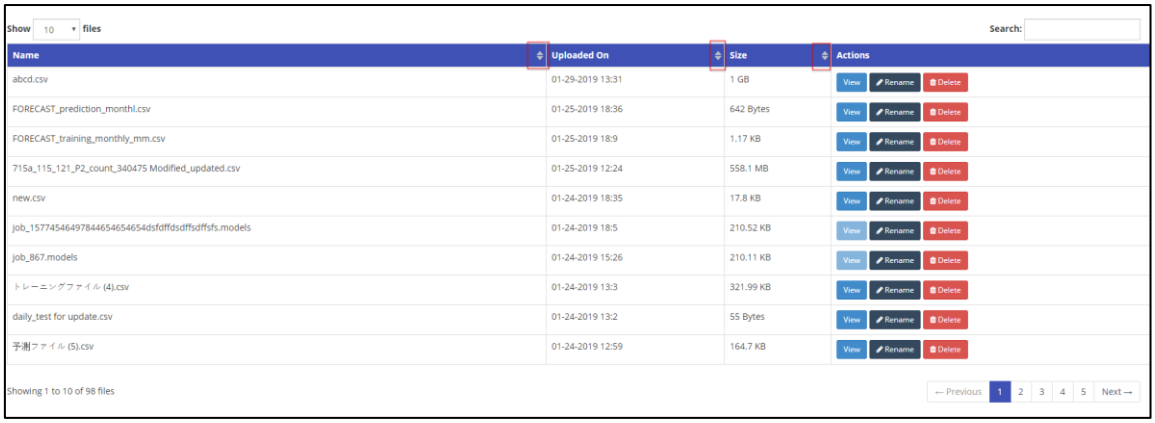


Figure 102 - Sort Datasets

2.8.2.3 Search list

System allows users to search through the list based on file name, upload date and time, and file size.

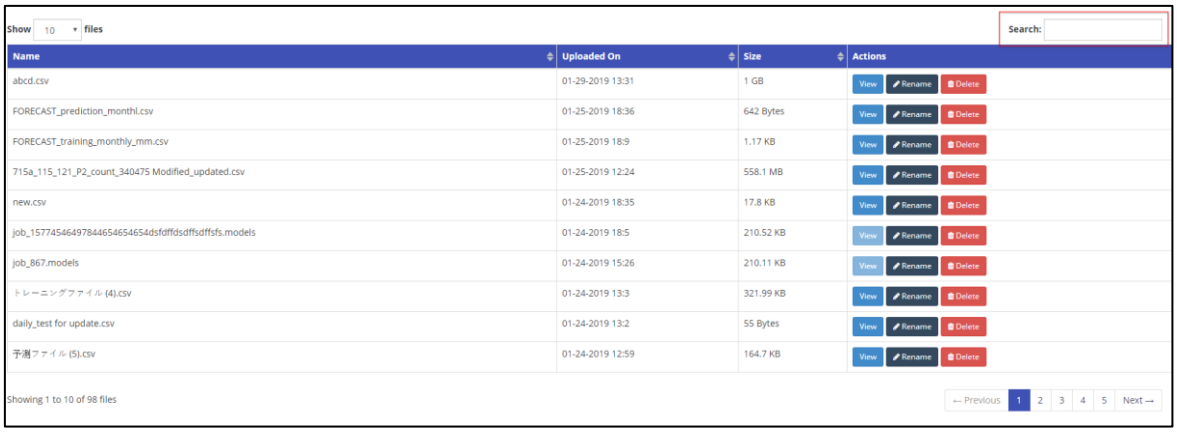


Figure 103 - Search Datasets

2.8.2.4 Pagination

In case the total number of files uploaded by the users is greater than files shown in a single screen, system shows the list in multiple pages. The pages are accessible on the control present at the bottom right side of the table.

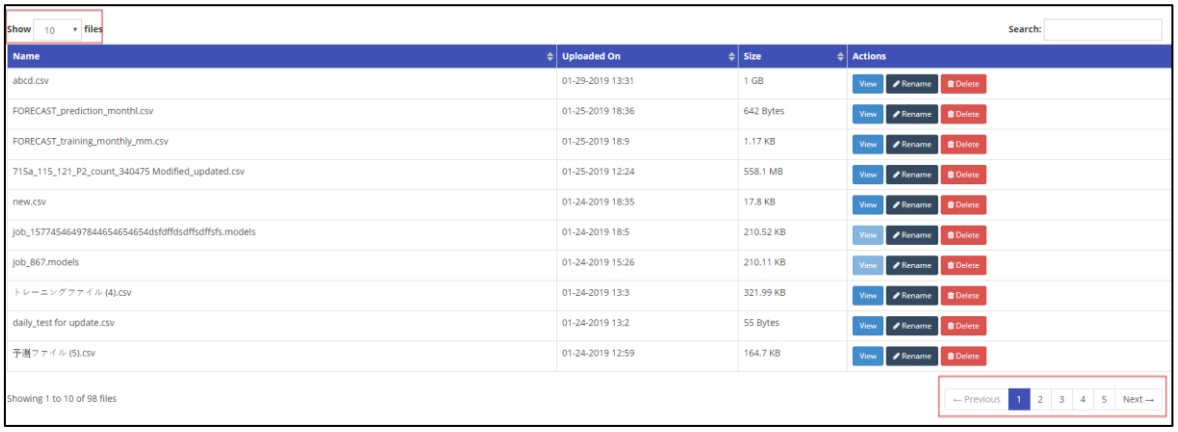


Figure 104 - Dataset Pagination

2.8.2.5 View dataset contents

System allows users to view the top 10 rows of the dataset by clicking the “View” button. This option however will not be available for model archive file.

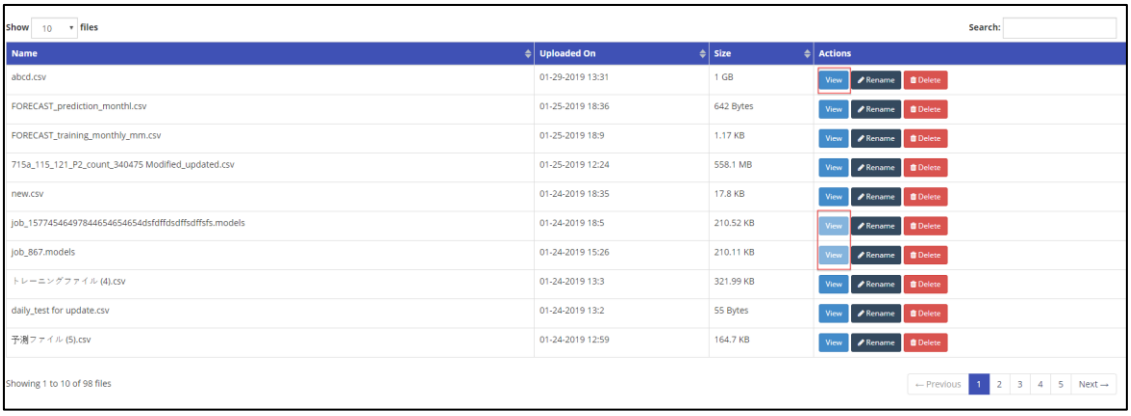
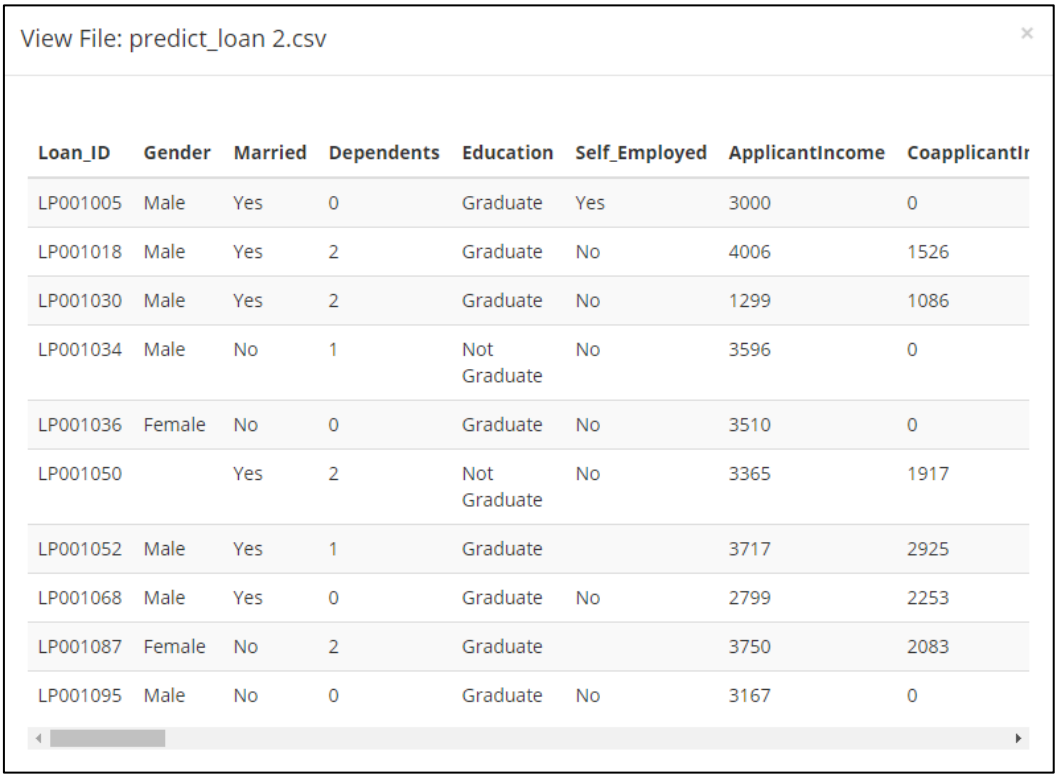


Figure 105 - View Dataset contents



Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIr
LP001005	Male	Yes	0	Graduate	Yes	3000	0
LP001018	Male	Yes	2	Graduate	No	4006	1526
LP001030	Male	Yes	2	Graduate	No	1299	1086
LP001034	Male	No	1	Not Graduate	No	3596	0
LP001036	Female	No	0	Graduate	No	3510	0
LP001050		Yes	2	Not Graduate	No	3365	1917
LP001052	Male	Yes	1	Graduate		3717	2925
LP001068	Male	Yes	0	Graduate	No	2799	2253
LP001087	Female	No	2	Graduate		3750	2083
LP001095	Male	No	0	Graduate	No	3167	0

Figure 106 - Dataset file content

2.8.2.6 Rename files

System allows users rename files. In case the file with the updated name is already present, system notifies user if the file should be overwritten.

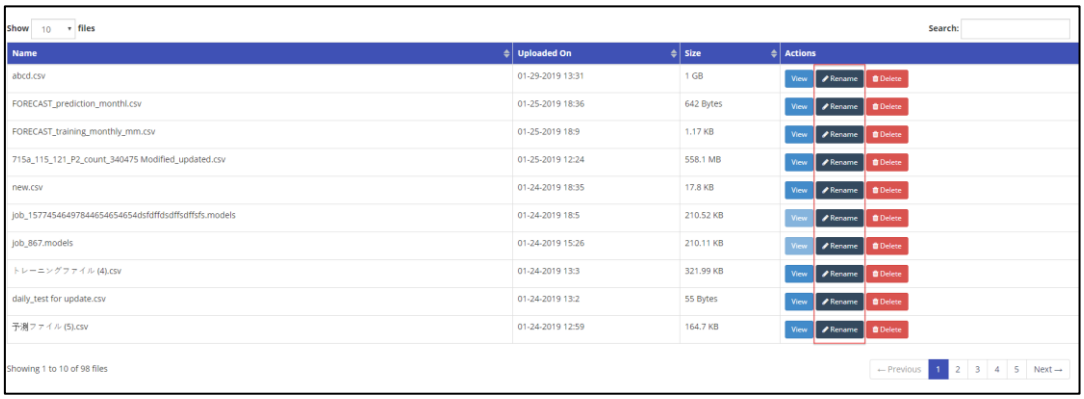


Figure 107 - Rename dataset

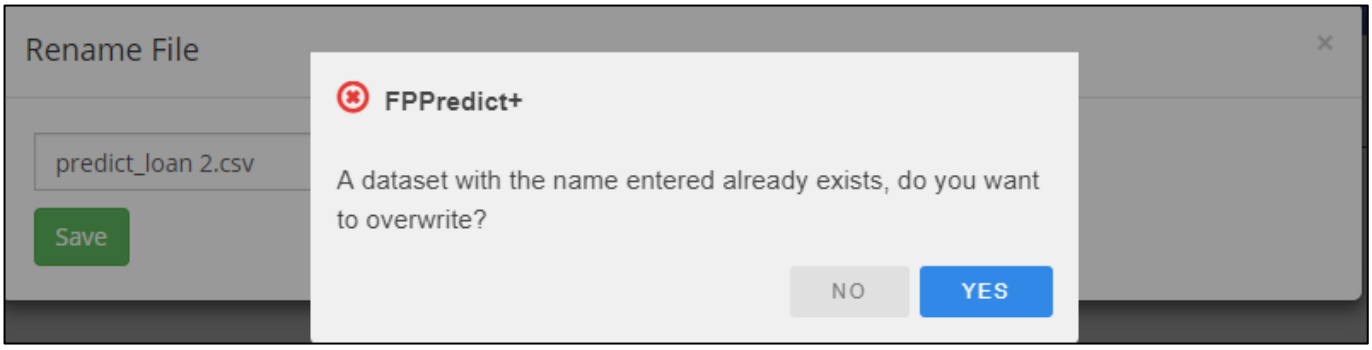


Figure 108 - Rename dataset - Confirmation

2.8.2.7 Delete files

System allows users delete files. On clicking delete, system asks users confirmation for deleting the file, after which it gets delete on confirmation.

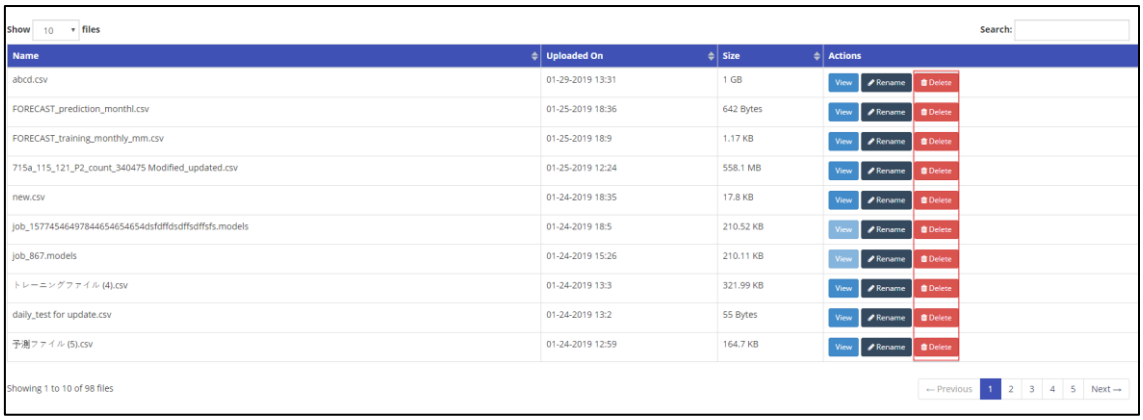


Figure 109 - Delete dataset

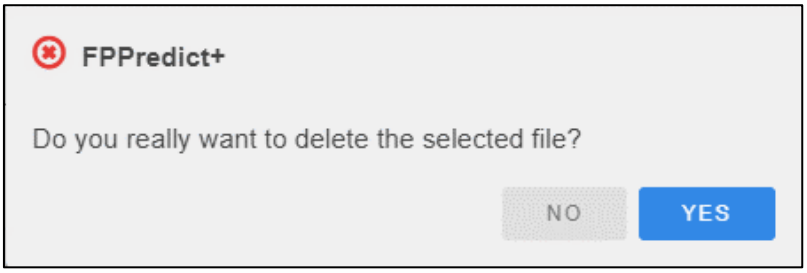


Figure 110 - Delete dataset - confirmation

2.12 License Information

This page provides information regarding the license that has been used to register that system. This page also allows users to create the API key used for submitting jobs using REST APIs. Following are the main sections of this page.

2.12.1 License Information

This section shows following information to the user:

- **License Status**: A pictorial representation of license status, showing if the license is currently valid or not
- **Register:** This button will redirect users to register instance page, allowing them to reregister the instance.
  This button will only visible in case the license is invalid.
- **License Information:** Th user will be presented with the license information including license from date, licence to date, Customer ID and license key in masked format.

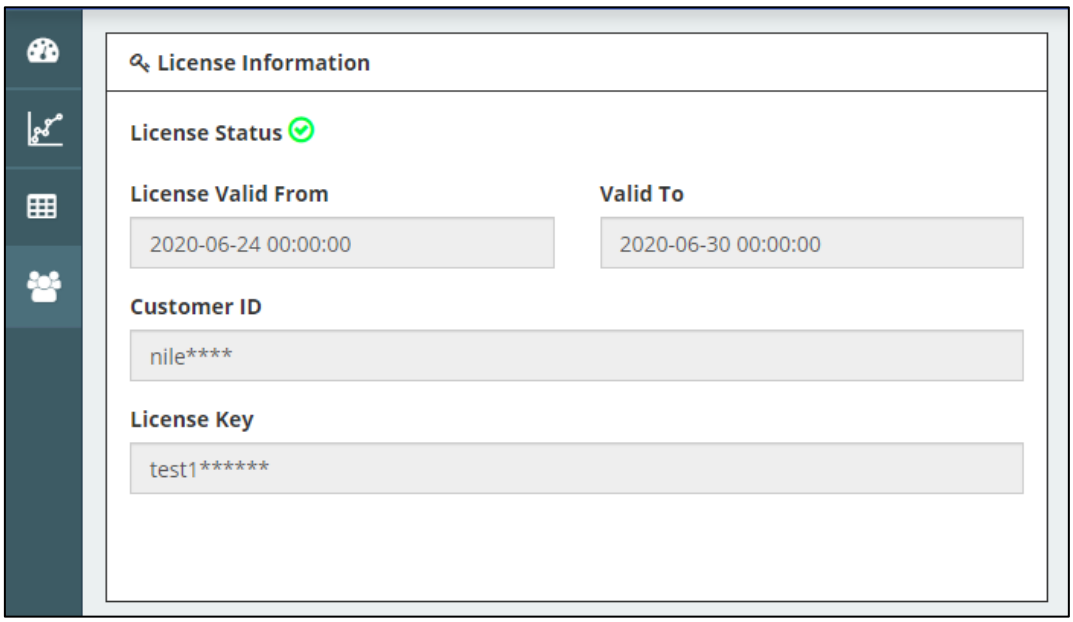


Figure 111 - License Information

2.12.2 Modules Information

This section shows the components available to the user, based on the license applied. The available components will be shown with “green tick mark”.

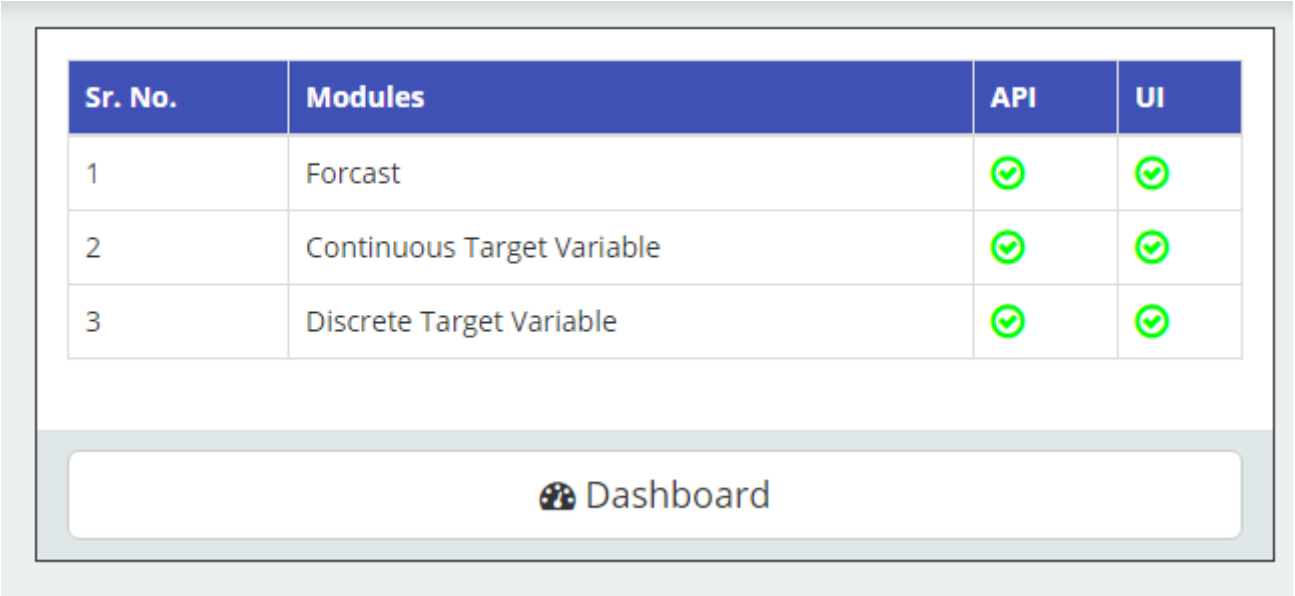


Figure 112 - Available Components

Clicking on the button named “Dashboard”. Will redirect users to dashboard page.

2.12.3 API Keys

This section allows the users to create API keys and view the generate API keys.

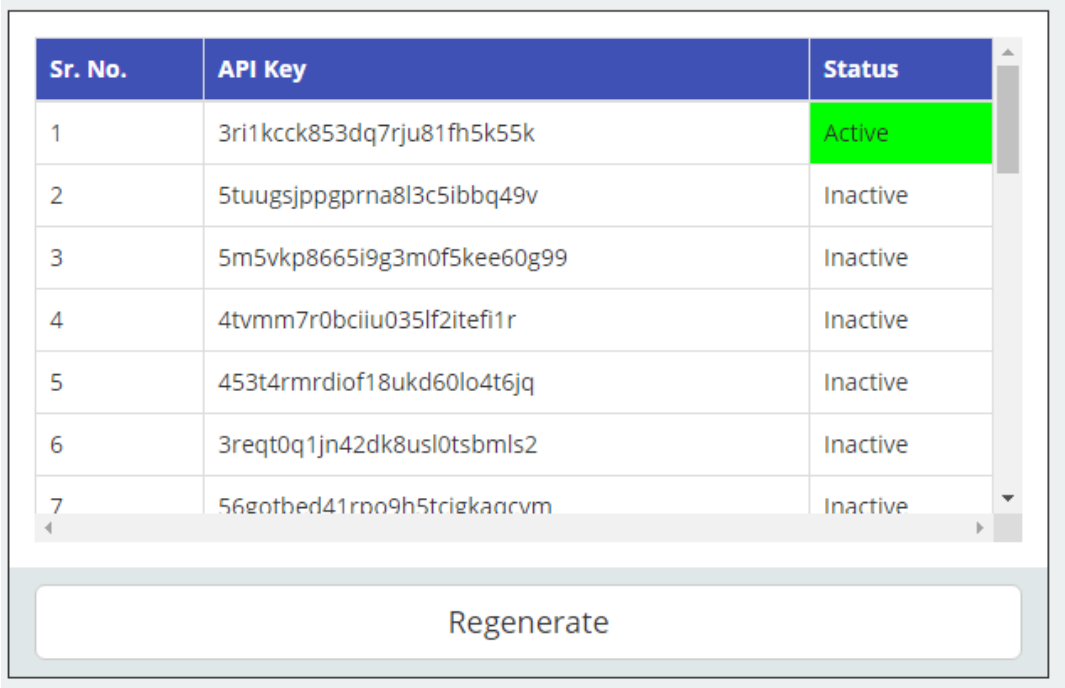


Figure 113 - API KEYs

In case the API key was already generated, system will allow users to regenerate the API key. Only one active API key will be present in the system at a given time. The active API keys can be used with REST APIs to submit jobs.

2.13 Support

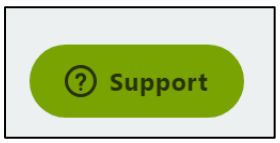


Figure 114 - Support Button

System facilitates users to get more value from the product and ask support for the product by clicking the “Support” button.

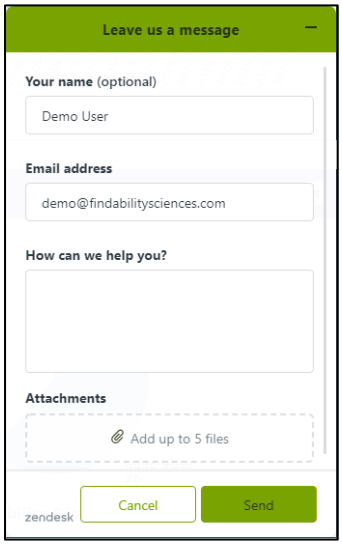


Figure 115 - Support Form

After clicking, users can ask their queries, providing the Username and Email Address.

Appendix 1 – Guidelines for Data Sets for Prediction

Following are the set of guidelines for the data preparation. If not mentioned specifically, the guidelines apply for both training and prediction sets.

1. Headers

a. Column header names should be maximum 30 characters in length.
b. Column header names should always start with alphabets (a-z) or underscore (_) character for files containing ASCII data only Non-ASCII data set does not have this restriction.
c. Column header names should only consist of alphabets (a-z), numbers (0-9), underscore (_) and hyphen (-) character for files containing ASCII data only Non-ASCII data set does not have this restriction.
d. Column header names should be same in name and order in both the datasets.
e. Datasets should not have duplicate column header names.
a. Datasets should not have any empty column header names.

2. Dataset

a. System only accepts datasets in CSV format with following specification:
    - Field Delimiter: Comma (,) character
    - Row Delimiter: New Line character
    - Encoding format: UTF-8 Without BOM
b. Total number of columns should be same in all the rows and should be equal to the number of headers.
c. There is an EOF character present at the end of both training and prediction sets.
d. The column chosen as target variable should have minimum 2 unique values.
e. The column chosen as target variable shouldn't have any empty values.
f. Values present in the column chosen as ID variable should have unique values, which can be used to identify records in the dataset.
g. The column chosen as ID variable shouldn't have any empty value.
h. The data representing same values, should be present in the data in the same format. System will treat 4229 and 4,229 as two different values.
i. Files should not contain any empty rows at the end of the file.

Appendix 2 – Guidelines for Data Sets for Forecast

Following are the set of guidelines for the data preparation. If not mentioned specifically, the guidelines apply for both training and Forecast sets.

1. Headers

a. Column header names should be maximum 30 characters in length.
b. Column header names should not start with numbers.
c. Column header names should be same in name and order in both the datasets
d. Datasets should not have duplicate column header names.
e. Datasets should not have any empty column header names.

2. Dataset

a. System only accepts datasets in CSV format with following specification:
    - Field Delimiter: Comma (,) character
    - Row Delimiter: New Line character
    - Encoding format: UTF-8 Without BOM
b. The column chosen as target variable should only contain numeric values.
c. The column chosen as target variable shouldn't have any empty values.
d. Values present in the column chosen as Timestamp variable should have timestamp values conforming to the format provided during job submission.
e. Values present in the column chosen as Timestamp variable should have unique values, which can be used to identify records in the dataset.
f. No more than 10% of the total number of values present in the datasets should be empty.
g. Files should not contain any empty rows at the end of the file.
h. Number of records in training dataset should be at least twice the number of records in forecast dataset.

Appendix 3 – Binary Target Prediction

In case the target variable has only 2 unique values in training file, FPPredict+ provides following options to the users

A. **Less Frequent Value Prediction:** In this case system predicts all the records with target value as the less frequent value with the models optimized to provide overall best results for the less frequent value only with the highest probability possible.

B. **More Frequent Value Prediction:** In this case system predicts all the records with target value as the more frequent value with the models optimized to provide overall best results for the more frequent value only with the highest probability possible.

C. **Both Values Prediction:** In this case system predicts for both the possible target values, with the models optimized for each of the target values. Here, the models created are not necessarily the same as those got created while running the jobs with any of the options “Less Frequent value” or “More frequent value”.

An exception for both value prediction is the case where the less frequent target value is very infrequent, e.g. 5%, the more frequent target value may not be predicted unless there are patterns that show significance greater than random probability of occurrence of more frequent value. In such cases, system will predict only for less frequent value.

Appendix 4 – Tracker Variable

User may select any of the independent variables as tracker variable. There can only be one tracker variable for a job.

Using the tracker variable is equivalent to splitting the training and prediction files each into multiple files – one for every unique value of the tracker variable. Each of these “tracker files” contains all the records from the original file with the same tracker value and all tracker files combined contain all training or prediction records from the original file. In a nutshell, each model gets associated with a unique tracker value along with a unique target value.

Selecting a tracker variable gives you the option to force the modeling and predictions to always have the same tracker variable value. For example, you want to predict employment income and “education” is one of your independent variables. By selecting “education” as tracker variable you force that records are predicted using only models created from training records with the same value for “education”.

A record with a tracker value that was not part of any model will not be predicted.

Appendix 5 – DTV Jobs Service URL

In case the system gives error “Required component is unreachable. It can be due to internet connection or invalid configuration”, when running DTV jobs, Make sure the following URL is accessible from the OpenShift Cluster:

“http://169.61.73.205”

In case it is not, you will not be able to run DTV jobs.

Appendix 6 – Interaction with Cloud Pak for data

For users of cloud pak for data, the interaction is primarily available in form UI as well as APIs. A general architectural solution is going to look as follows:

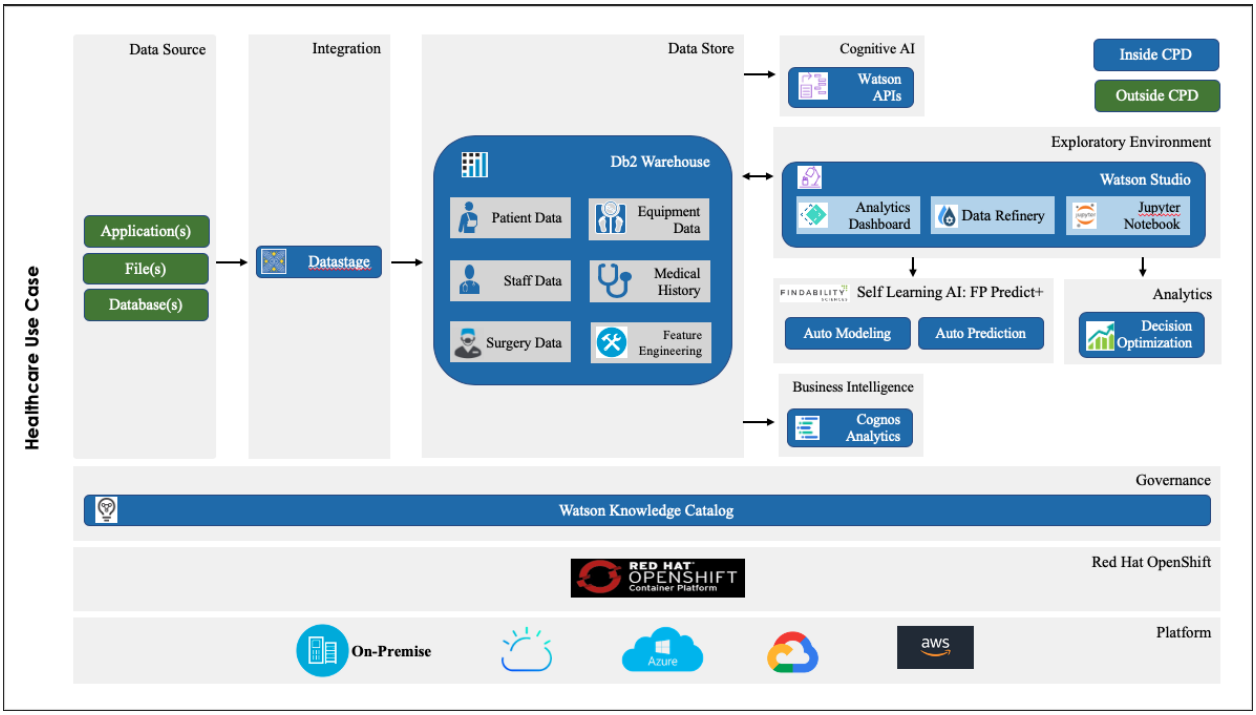


Figure 116 - Architecture - CPD and FP-Predict+

The general steps for interaction will be as follows:

- Navigate to the Application URL, got after the deployment of the Application.
- Enter the provided User Id and Password in the below page. Please also note the URL to the below page as it will be required for calling FS Predict+ API's

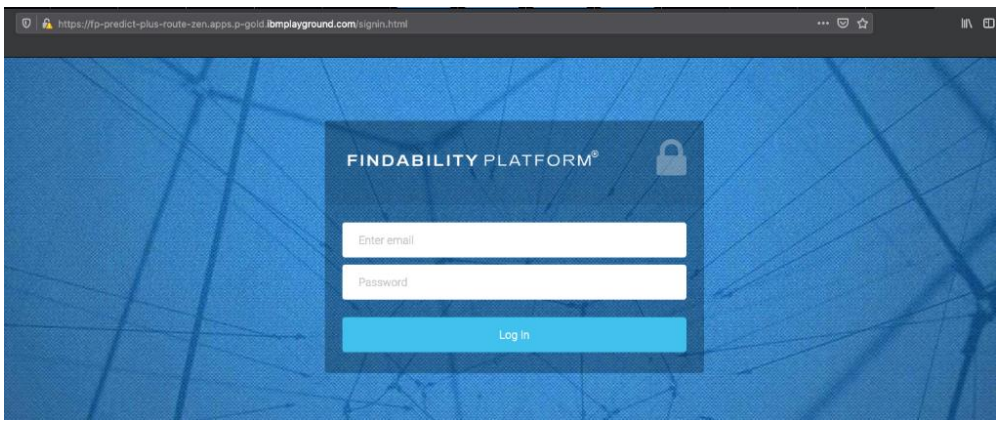


Figure 117 - Login Page

- Once you login, register the instance and click on License Information

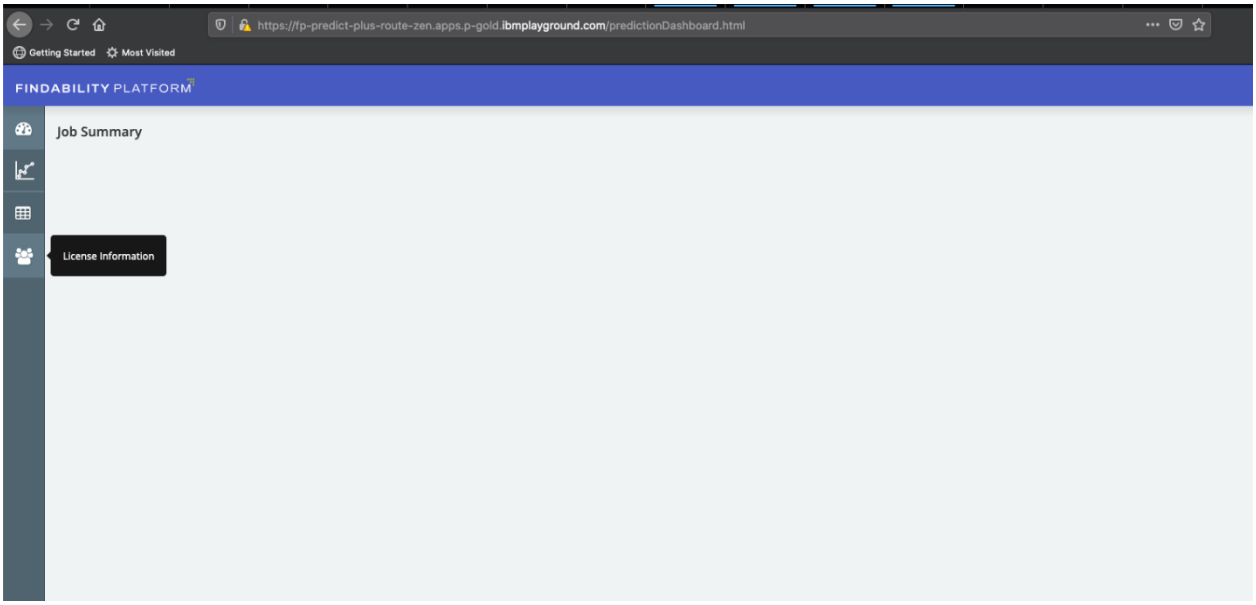


Figure 118 - Licence information navigation

- Copy the API key from the page. First time users will see a 'Generate' option to generate the API Key for the first time.

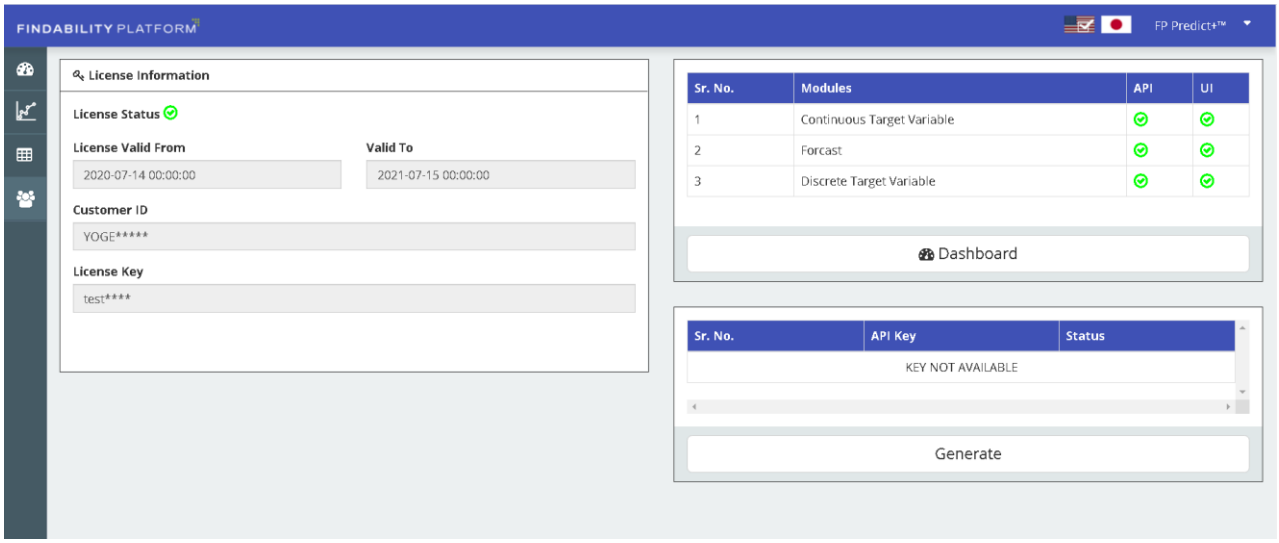


Figure 119 - Generate License

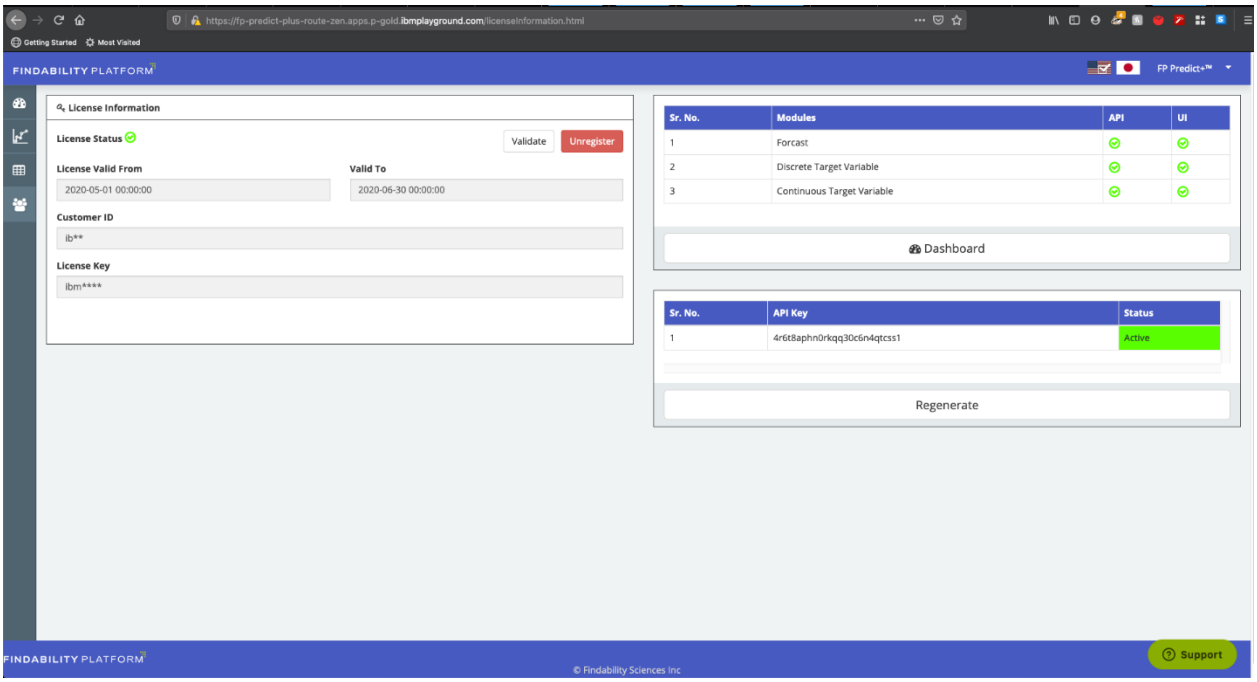


Figure 120 - Generated License

- The URL and the API Key can be used to call FS Predict + APIs. In the below example, the APIs are called from Jupyter Notebook in CPD. Set up the required variables in Jupyter Notebook

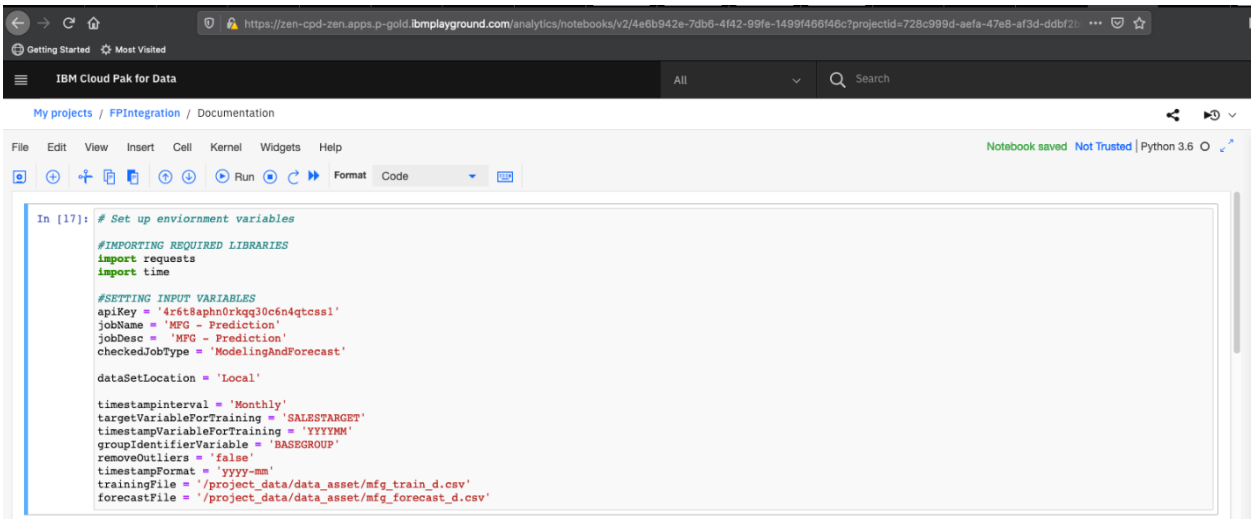


Figure 121 - Jupyter Setup parameters

- Call API to submit a prediction job

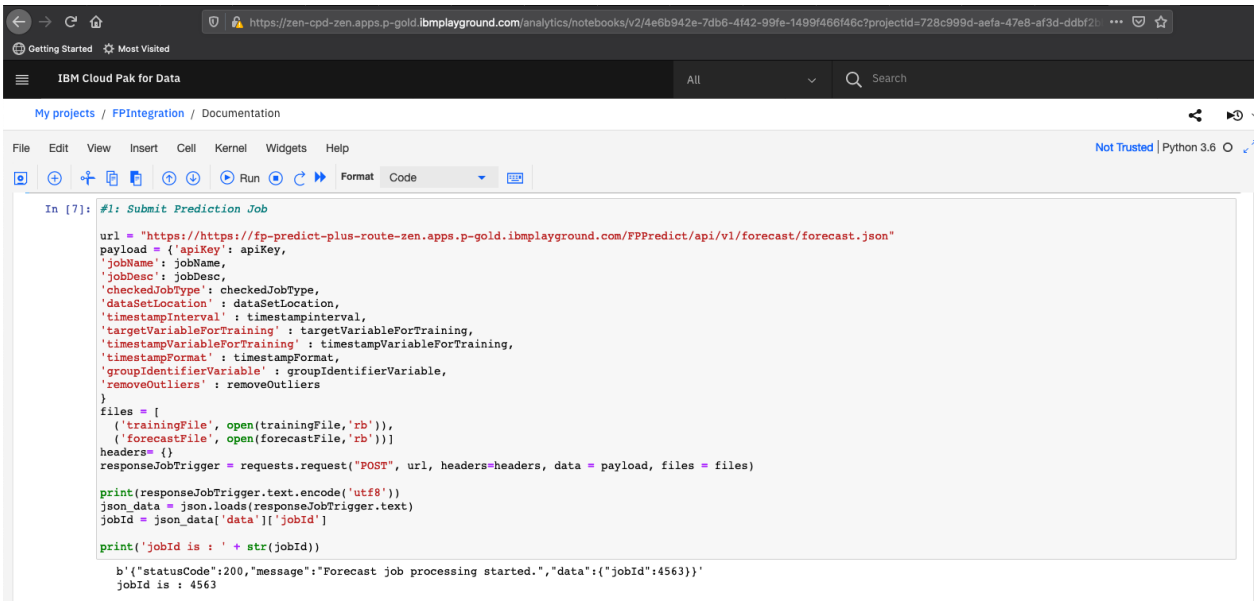


Figure 122 - Submit job via API

- Call API to check the status of a forecasting job

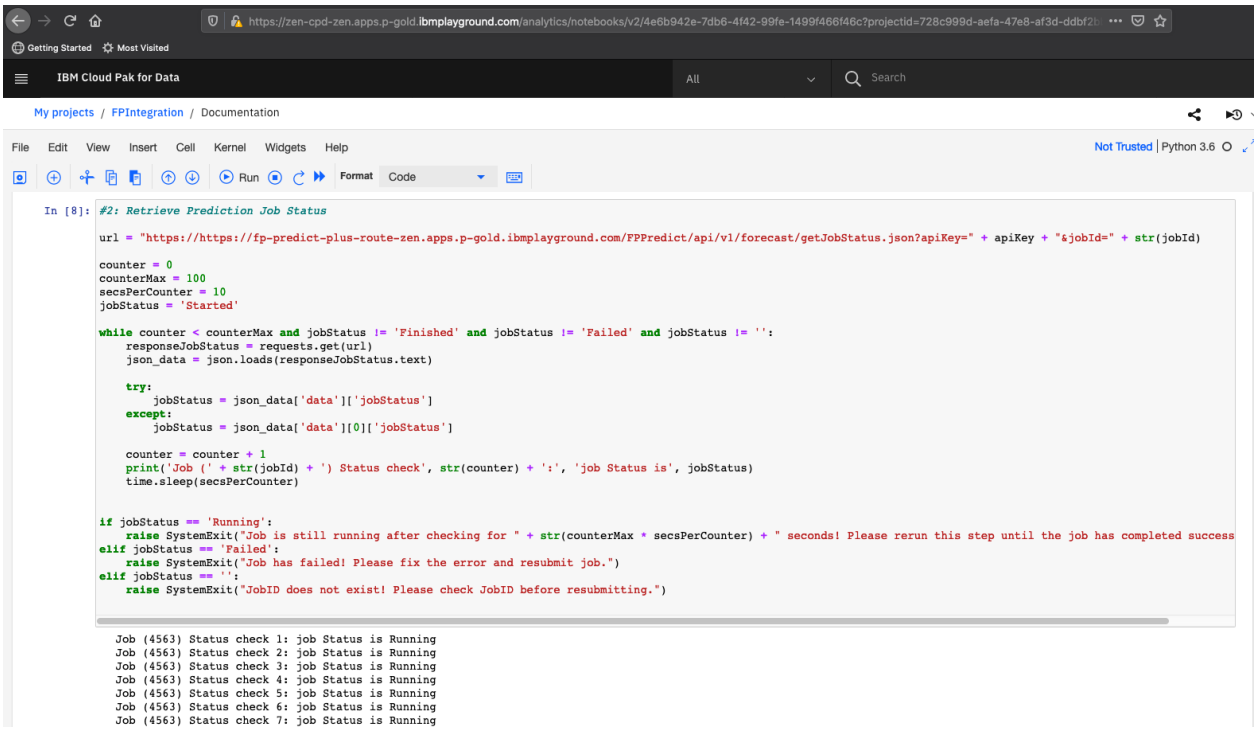


Figure 123 - Status Check

- Once the job finishes, Call API to retrieve results of a forecasting job

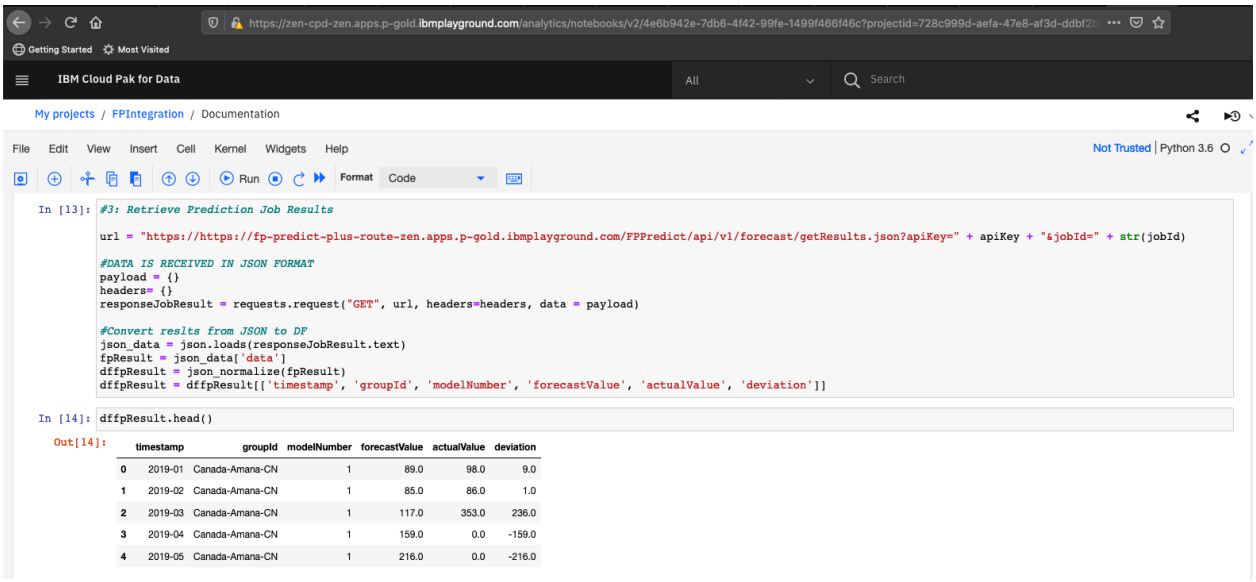


Figure 124 - View Results

The results saved here then can be used with Cognos and Db2 to save the results and create Dashboards and Reporting.