**Use-Cases Draft**

Case 01: Suggest latest 10 datasets used by other user's based on an industry domain (Tag given to a dataset)

INP: Name of the Tag: Finance/Healthcare/Retail/Tourism etc.

OUT: Name of the Dataset, Data-size, no. of Rows, columns etc.

Operation: Get the tag-Id corresponding to the tag name from the Tag\_Map table. Find the matching details pertaining to Name, Data-size, no. of rows and columns etc. from the Dataset\_Metadata table corresponding to the tag-Id and display the top 10 rsults filtered on the basis of descending Starttime ()

Case 02: Suggest latest 10 datasets based on Algorithm(Regression/Classification) and the model-type used ( Linear/Logistic/GLM etc.)

INP: Model\_Type\_Name, Algorithm

OUT: Name of the Dataset, Data-size, no. of Rows, columns etc.

Operation:

Case 03: Details of the conditions under which model was tested

INP: Model\_ID

OUT: Run\_ID, Duration of run, Model\_Name,Presence in StackedEnsemble BestOfFamily

Operation: Get the matching Run\_Id from the ModelRun table for the input Model\_ID. Get the other details related to this model\_ID from the Leaderboard\_Metadata such as Start and End time of run. Append a calculated field to find duration (Start-End time). Get the model\_name and presence in Stackedensemble from the ID\_Map table for this Model\_Id

Case 04: Get average no. of models generated for a particular run\_time and datasets with the same tag having almost similar datsize in terms of no. of rows and columns

INP: Tag\_Name, Run\_time, No. of Rows, Columns in user's dataset

Out: Average no. of models generated

Operation: From the tags table get the Ids corresponding to the input tag name. Get the Dataset Ids for these tag\_Ids from the Tag\_Map table. Select only those dataset Ids for which the  the no. of rows and columns for all these datasets and the input run-time from the leaderboard\_Metadata table ( input values of rows and column lying within a particular range, lets say +/- 10). Group the data as per Run\_Id and calulate the total count of model of a particular type/total count of models in the run.

Case 05: For the datasets from a particular industry(tag), size(observations) and type of analysis to be done ( Regression/Classification/Time-series forecasting), which is the best model?

INP: Tag\_Name, No. of rows, columns, model\_species

OUT: Best Model's name and performance metric values ( RMSE for Regression/ AUC for Classification)

Operation: Get the dataset Ids for the Tag\_Name from the Tag\_Map table. Select only those dataset Ids for which the no. of rows and columns for all these datasets and the input run-time from the leaderboard\_Metadata table ( input values of rows and column lying within a particular range, lets say +/- 10). From the Data\_Map table find the Run\_Id for each of these Dataset\_Ids .

Get all the Model\_IDs for these Run\_Ids from the Model\_Run table and filter only the model\_Ids associated with the input model type from the ID\_Map table. Compare the performance values for all these model Ids

Case 06: For regression model, enlist the no. of linear, logistic, GBM and GLM models generated for a particular runtime.

INP: Runtime

OUT: No. of models in each Model\_Type

Opeartion: Get run\_Ids of all models for the input runtime. Filter the run\_Ids matching with

{the Model Ids for the particular model\_type Id(Regression in this case)} and calculate the count of model\_IDs grouped by each model\_type

Case 07: Which is the most commonly applied best model-type across run-times for a given dataset?

INP: Model-type\_ID, Dataset\_ID

OUT: Dataset Name, count(Model\_Type\_Id)

Operation: Filter Dataset\_Ids

Case 08: Display the hyperparameter values of all the models ( of same type) for a paticular dataset  & run-time. (Applicable for analysing cases where 2-3 GBM models are created during a 300/500 sec runtime)

INP: Dataset\_Id, Run-time, Model\_Type\_ID

OUT: Hyperparameter ID, values

Operation: Filter the Run\_Ids in the Data\_Map table based on the (Run-Ids matching input run-time and model\_species from leaderboard\_Metadat table and Dataset Id of input dataset). For each of those run\_Ids find the model-ids from the Model\_Run table and Get the hyperparam\_Id and value from the hyperparameter\_values table.

Case 09: Display the hyperparameter values of all the models ( of same type) for a paticular dataset  & across all run-times.

INP: Dataset\_Id, Model\_Type\_ID

OUT: Hyperparameter ID, values

Operation: Filter the Run\_Ids in the Data\_Map table based on the (Run-Ids matching model\_species from leaderboard\_Metadata table and Dataset Id of input dataset). For each of those run\_Ids find the model-ids from the Model\_Run table and Get the hyperparam\_Id and value from the hyperparameter\_values table.

For each model\_Id also display the RMSE, MAE,MAPE etc.. by joining with leaderboard table

Case 10: Display the default values of the hyperparameters for a particular model\_Type

INP: Model\_Type\_Name

OUT: List of Hyperparameter names and their values

Operation : Get the model\_Type\_Id for the model\_type name from the Model\_Map table and join with hyperparameter\_Default table to display the required values.

Case 11 : Calculate change in hyperparameter values (denote incresae/decrease using if/case statement) for a particular Model\_Type and run-time

INP: Model\_Type\_Id, Run-time

OUT:  Model\_Name, change in value

Operation: For a particular model-type and run-time get all the Run-Ids from Leaderboard\_Metadata table. For each run id get the model-Ids and join with Hyerparameter\_value and hyperparameter\_default value to calculate increase/decrease in the value compared to default.