

Normalization:

Checking the first Normal Form

No Repeating groups

No Data Values are atomic

No Each field has different name

Has a primary key

Checking the Second Normal Form

It is in the First Normal Form

All non key attributes are dependent on ALL parts of primary key

Each Field has a unique name

Has a primary key

Checking Third Normal Form

It should be in second normal form

All non-key attribute are not dependent on other non key attribute

Each field has a unique name

Has a primary key

## 5 Use Cases:

Case 1: Compare the metrics and also find the common Hyperparameter for like models such GBM, RANDOM FOREST, XRT

The screenshot shows the Microsoft SQL Server Management Studio interface. The query window contains the following SQL code:

```
--Compare the metrics and also find the common Hyperparameter for like models such GBM, RANDOM FOREST, XRT
USE [Hyperparamter_PK]
GO
SELECT
    ld.model_id, auc, logloss, rmse, gb.ntrees, rf.ntrees, xr.ntrees, gb.fold_assignment, xr.fold_assignment
FROM
    dbo.leaderboard_id
        FULL JOIN
    dbo.gbm_hyp gb ON ld.model_id = gb.model_id Full JOIN
    dbo.randomForest_hyp rf ON ld.model_id = rf.model_id Full JOIN
    dbo.xrt_hyp xr ON ld.model_id = xr.model_id
```

The results grid shows 27 rows of data, each representing a different model configuration. The columns include model\_id, auc, logloss, rmse, ntrees, fold\_assignment, and various hyperparameters. The properties pane on the right shows connection details for the current session.

model_id	auc	logloss	rmse	gb_ntrees	rf_ntrees	xr_ntrees	gb_fold_assignment	xr_fold_assignment
GBM_2_AutoML_20190420_162320	0.916771305320544	0.19911294705987	0.2082740					
GBM_grid_1_AutoML_20190420_185144_model_1	0.899784051120119	0.322867214132937	0.3241499					
GBM_grid_1_AutoML_20190420_181756_model_1	0.899784051120119	0.322867214132937	0.3241499					
StackedEnsemble_BestOfFamily_AutoML_20190420_195144	0.8998850228742	0.21418924732348	0.3244480					
StackedEnsemble_BestOfFamily_AutoML_20190420_195176	0.914434601545767	0.31426696039281	0.3115983					
StackedEnsemble_BestOfFamily_AutoML_20190420_162320	0.919206063779868	0.215655794158603	0.2123164					
XRT_1_AutoML_20190420_185144	0.902222178159801	0.3335150038808049	0.3191690					
XRT_1_AutoML_20190420_185144	0.902222178159801	0.3335150038808049	0.3191690					
XRT_1_AutoML_20190420_185176	0.907472517734447	0.3140582375585	0.3157924					
XRT_1_AutoML_20190420_162320	0.909568070606538	0.31657142666167	0.3170336					
XRT_1_AutoML_20190420_162320	0.909568070606538	0.31657142666167	0.3170336					

Case 2: What are the metrics for stacked ensemble model where base models are present for every model id

```
-- what are the metrics for stacked ensembling model where base models are present for every model_id
SELECT
    dbo.leaderboard.model_id, dbo.leaderboard.auc, dbo.leaderboard.logloss, dbo.leaderboard.mean_per_class_error, dbo.leaderboard.rmse, dbo.leaderboard
    FROM
        dbo.stackedEnsemble_hyp.blending_frame, dbo.stackedEnsemble_hyp.export_checkpoints_dir, dbo.stackedEnsemble_hyp.response_column, dbo.stack
        dbo.stackedEnsemble_hyp.metalearnr_params, dbo.stackedEnsemble_hyp.metalearnr_nfolds, dbo.stackedEnsemble_hyp.metalearnr_fold_column,
        dbo.stackedEnsemble_hyp.metalearnr_algorithm, dbo.stackedEnsemble_hyp.keep_levelone_frame, dbo.stackedEnsemble_hyp.validation_frame
        dbo.stackedEnsemble_hyp INNER JOIN
        dbo.leaderboard ON dbo.stackedEnsemble_hyp.model_id = dbo.leaderboard.model_id
        WHERE training_frame is not null
        AND metalearnr_nfolds >= 0
        AND metalearnr_algorithm in ('AUTO', 'glm', 'gbm', 'rf', 'deeplearning')
        AND base_models is not null
```

Results:

model_id	auc	logloss	mean_per_class_error	rmse
StackedEnsemble_BernOffFamily_AutoML_20190420_129144	0.91658893871	0.3141248611009	0.179809310144589	0.3111
StackedEnsemble_BernOffFamily_AutoML_20190420_129144	0.916588938271	0.3141248611009	0.179809310144589	0.3111
StackedEnsemble_BernOffFamily_AutoML_20190420_151756	0.916424601545767	0.314268699029281	0.180125407362666	0.3111
StackedEnsemble_BernOffFamily_AutoML_20190420_151756	0.916424601545767	0.314268699029281	0.180125407362666	0.3111
StackedEnsemble_BernOffFamily_AutoML_20190420_162320	0.915920603777968	0.315655796155804	0.189881537012969	0.3121
StackedEnsemble_BernOffFamily_AutoML_20190420_162320	0.915920603777968	0.315655796155804	0.189881537012969	0.3121

(6 rows affected)

Query executed successfully.

Case3: Find the Hyperparameters from GBM where number of trees are greater than 50 and distribution is either Bernoulli or multinomial min\_split\_improvement is between 1e-10 and 1e-3

```
-- Find the Hyperparameters from GBM where number of trees are greater than 50 and distribution is either bernoulli or multinomial
-- min_split_improvement between 1e-10 and 1e-3 and
SELECT
    dbo.gbm_hyp.[checkpoint], dbo.gbm_hyp.class_sampling_factors, dbo.gbm_hyp.col_sample_rate,
    dbo.gbm_hyp.col_sample_rate_change_per_level, dbo.gbm_hyp.custom_metric_func,
    dbo.gbm_hyp.col_sample_rate_per_tree,
    dbo.gbm_hyp.distribution, dbo.gbm_hyp.export_checkpoints_dir, dbo.gbm_hyp.fold_assignment,
    dbo.gbm_hyp.fold_column, dbo.gbm_hyp.histogram_type, dbo.gbm_hyp.huber_alpha,
    dbo.gbm_hyp.ignore_const_cols,
    dbo.gbm_hyp.learn_rate,
    dbo.schema_Hyperparameter_PK.dbo.gbm_hyp.max_abs_leafnode_pred,
    dbo.gbm_hyp.max_nunique_sets, dbo.gbm_hyp.min_rows, dbo.gbm_hyp.min_split_improvement,
    dbo.gbm_hyp.model_id, dbo.gbm_hyp.monotone_constraints, dbo.gbm_hyp.nbins, dbo.gbm_hyp.nbins_cats,
    dbo.gbm_hyp.nbins_top_level, dbo.gbm_hyp.nfolds, dbo.gbm_hyp.ntrees, dbo.gbm_hyp.offset_column,
    dbo.gbm_hyp.pred_noise_bandwidth, dbo.gbm_hyp.quantile_alpha, dbo.gbm_hyp.rz_stopping,
    dbo.gbm_hyp.response_column, dbo.gbm_hyp.sample_rate, dbo.gbm_hyp.sample_rate_per_class, dbo.gbm_hyp.score_each_iterat
    dbo.gbm_hyp.weight_column, dbo.leaderboard.model_id AS Expr1
    FROM
        dbo.leaderboard ON dbo.gbm_hyp.model_id = dbo.leaderboard.model_id
        WHERE ntrees > 50
        AND distribution in ('bernoulli', 'multinomial')
        AND min_split_improvement >= 1e-10 and min_split_improvement <= 1e-3
```

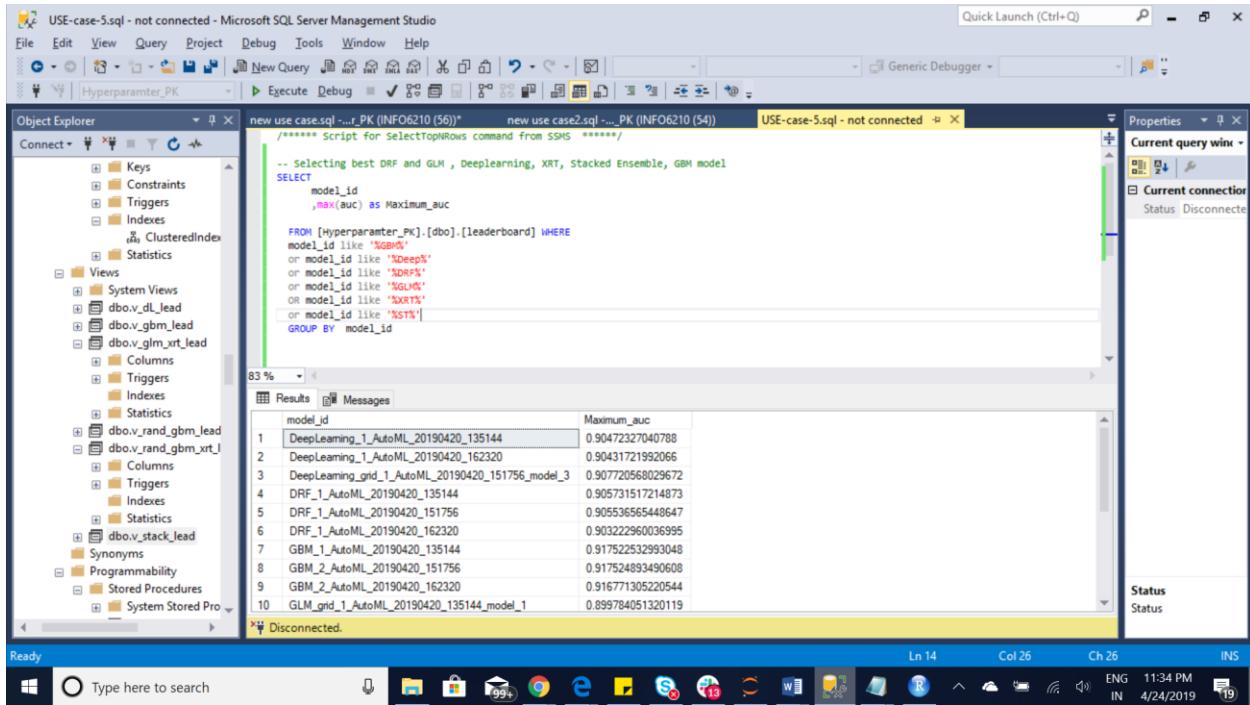
Results:

checkpoint	class_sampling_factors	col_sample_rate	col_sample_rate_change_per_level	custom_metric_func	col_sample_rate_per_tree	distribution
NULL	0.8	1	0.8	0.8	0.8	bernoulli
NULL	0.8	1	0.8	0.8	0.8	bernoulli

(2 rows affected)

Query executed successfully.

#### Case 4: Selecting best DRF and GLM , Deeplearning, XRT, Stacked Ensemble, GBM model



The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists various database objects like Keys, Constraints, Triggers, Indexes, Views, Columns, Triggers, and Statistics. The central pane displays a T-SQL query:

```
-- Selecting best DRF and GLM , Deeplearning, XRT, Stacked Ensemble, GBM model
SELECT
    model_id
    ,max(auc) as Maximum_auc
FROM [Hyperparameter_PK].[dbo].[leaderboard] WHERE
model_id like 'GBM%'
or model_id like 'Deep%'
or model_id like 'DRF%'
or model_id like 'GLM'
or model_id like 'XRT'
or model_id like 'SRT'
GROUP BY model_id
```

The Results pane shows the output of the query:

	model_id	Maximum_auc
1	DeepLearning_1_AutoML_20190420_135144	0.90472327040788
2	DeepLearning_1_AutoML_20190420_162320	0.90431721992066
3	DeepLearning_grid_1_AutoML_20190420_151756_model_3	0.907720568029672
4	DRF_1_AutoML_20190420_135144	0.905731517214873
5	DRF_1_AutoML_20190420_151756	0.905536565448647
6	DRF_1_AutoML_20190420_162320	0.903222960036995
7	GBM_1_AutoML_20190420_135144	0.917522532993048
8	GBM_2_AutoML_20190420_151756	0.917524893490608
9	GBM_2_AutoML_20190420_162320	0.916771305220544
10	GLM_grid_1_AutoML_20190420_135144_model_1	0.899784051320119

#### Case5 : Range of Hyperparameter and othe parameters for GBM

Usercase-2.sql - not connected" - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK

Object Explorer

Connect

Results

```

new use case2.sql ~..._PK (INFO6210 (54))   USE-case-5.sql - not connected    Usercase-2.sql - not connected" # SQLQuery158.sql - not connected
--range of hyperparameter and other parameters for gbm
--SELECT DISTINCT auc, gb.model_id, logloss, ntrees , min_split_improvement , col_sample_rate, distribution, fold_assignment,
keep_cross_validation_models, keep_cross_validation_predictions, max_depth, min_rows, nbins, nbins_cats,
nfolds, quantile_alpha, stopping_metric, stopping_rounds, stopping_tolerance, tweedie_power
FROM
    dbo.gbm_hyp gb INNER JOIN
    dbo.leaderboard ld ON gb.model_id = ld.model_id

```

Properties

Current query win

auc	model_id	logloss	ntrees	min_split_improvement	col_sample_rate	distribution
0.917522832993048	GBM_1_AutoML_20190420_135144	0.287859095153619	50	1E-05	1	AUTO
0.917522832993048	GBM_1_AutoML_20190420_135144	0.2979852085153619	60	1E-05	0.8	berнулли
0.91752404904608	GBM_2_AutoML_20190420_151756	0.2979852085153619	50	1E-05	1	AUTO
0.91752404904608	GBM_2_AutoML_20190420_151756	0.29797609185315	60	1E-05	0.8	berнулли
0.916771055220544	GBM_2_AutoML_20190420_162920	0.29851294705997	50	1E-05	1	AUTO
0.916771055220544	GBM_2_AutoML_20190420_162920	0.29851294705997	64	1E-05	0.8	berнулли

(6 rows affected)

Disconnected.

Ready

Type here to search

LN 12 Col 1 Ch 1 INS

ENG 11:26 PM IN 4/24/2019 19

## 5 Views:

### View1:

#### v\_rnd\_-gbm\_xrt\_lead

SQLQuery177.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (57)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK

SQLQuery177.sql ~...\_PK (INFO6210 (57)) # new use case.sql ~...\_PK (INFO6210 (56)) new use case2.sql ~...\_PK (INFO6210 (54)) USE-case-5.sql - not connected

```

ALTER VIEW [dbo].[v_rnd_gbm_xrt_lead]
AS
SELECT
    dbo.randomForest_hyp.max_depth AS rf_max_depth, dbo.randomForest_hyp.max_hit_ratio_k AS rf_max_hit_ratio_k, dbo.randomForest_hyp.max_runtime_secs AS rf_max_runtime_
    dbo.randomForest_hyp.min_rows AS rf_min_rows, dbo.randomForest_hyp.model_id AS rf_model_id, dbo.randomForest_hyp.mtries AS rf_nbins, dbo
    dbo.randomForest_hyp.nbins_top_level AS rf_nbins_top_level, dbo.randomForest_hyp.nfolds AS rf_nfolds, dbo.randomForest_hyp.ntrees AS rf_ntrees, dbo.randomForest_hyp
    nrof_random_trees AS rf_nrof_random_trees, dbo.xrt_hyp.build_tree_type AS xr_build_tree_type, dbo.xrt_hyp.checkpoint AS xr_build_tree_o
    dbo.xrt_hyp.balance_classes AS xr_balance_classes, dbo.xrt_hyp.max_depth AS xr_depth, dbo.xrt_hyp.max_leaves AS xr_max_leaves, dbo.xrt_hyp.max_node AS xr_max_node
    dbo.xrt_hyp.max_runtime AS xr_max_runtime, dbo.xrt_hyp.nbins AS xr_nbins, dbo.xrt_hyp.nbins_cats AS xr_nbins_cats, dbo.xrt_hyp.nbins_top_
    dbo.xrt_hyp.ntrees AS xr_ntrees, dbo.xrt_hyp.offset_column AS xr_offset_column, dbo.xrt_hyp.r2_stopping AS xr_r2_stopping, dbo.xrt_hyp.response_column AS xr_respons
    dbo.xrt_hyp.validation_frame AS xr_validation_frame, dbo.xrt_hyp.stoping_rounds AS xr_stoping_rounds, dbo.xrt_hyp.score_tree_interval AS xr_score_tree_interval, dbo.xrt_hyp.score_each_
    dbo.xrt_hyp.sample_rate AS xr_sample_rate

FROM
    dbo.leaderboard Full JOIN
    dbo.gbm_hyp ON dbo.leaderboard.model_id = dbo.gbm_hyp.model_id Full JOIN
    dbo.randomForest_hyp ON dbo.leaderboard.model_id = dbo.randomForest_hyp.model_id full JOIN
    dbo.xrt_hyp ON dbo.leaderboard.model_id = dbo.xrt_hyp.model_id
    WHERE dbo.leaderboard.model_id like '%GBM%' OR dbo.leaderboard.model_id like '%XRT%'
    OR dbo.leaderboard.model_id like '%SDF%'
    AND dbo.gbm_hyp.nbins_cats >= 0 AND dbo.gbm_hyp.nbins_cats <=2000
    AND dbo.xrt_hyp.ntrees >= 20
    AND dbo.xrt_hyp.min_split_improvement >= 1e-10 AND dbo.xrt_hyp.min_split_improvement <= 1e-3
    AND dbo.randomFor
    column min_split_improvement(float, not null) GO

```

Connected. (1/1)

is-swang01.ischool.uw.edu (... | INFO6210 (57) | Hyperparameter\_PK | 00:00:00 | 0 rows

Ready

Type here to search

LN 16 Col 14 Ch 14 INS

ENG 11:38 PM IN 4/24/2019 19

SQLQuery178.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (57)) - Microsoft SQL Server Management Studio

```

SELECT TOP 1000 [balance_classes]
      ,[build_tree_one_node]
      ,[check_constant_response]
      ,[checkpoint]
      ,[calibrate_mode1]
      ,[calibration_frame]
      ,[categorical_encoding]
      ,[export_checkpoints_dir]
      ,[fold_assignment]
      ,[fold_column]
      ,[histogram_type]
      ,[huber_alpha]
      ,[ignore_const_cols]
      ,[ignored_columns]
      ,[keep_cross_validation_fold_assignment]
      ,[learn_rate_annealing]
      ,[learn_rate]
      ,[max_abs_leafnode_pred]

```

Results

	score_each_iteration	score_tree_interval	tweedie_power	weights_column	validation_frame	id_model_id	auc	logloss
1	0	0	1.5	NULL	NULL	GBM_1_AutoML_20190420_135144	0.917522532993048	0.29785
2	0	5	1.5	NULL	NULL	GBM_1_AutoML_20190420_135144	0.917522532993048	0.29785
3	0	0	1.5	NULL	NULL	GBM_2_AutoML_20190420_151756	0.917524893490608	0.29776
4	0	5	1.5	NULL	NULL	GBM_2_AutoML_20190420_151756	0.917524893490608	0.29776
5	0	0	1.5	NULL	NULL	GBM_2_AutoML_20190420_162320	0.916771305220544	0.29881
6	0	5	1.5	NULL	NULL	GBM_2_AutoML_20190420_162320	0.916771305220544	0.29881
7	NULL	NULL	NULL	NULL	NULL	XRT_1_AutoML_20190420_135144	0.902222178159601	0.33351
8	NILL	NILL	NILL	NILL	NILL	XRT_1_AutoML_20190420_135144	0.902222178159601	0.33351

Query executed successfully.

## View2: [v\_dl\_lead]

SQLQuery179.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (58)) - Microsoft SQL Server Management Studio

```

ALTER VIEW [dbo].[v_dl_lead]
AS
SELECT      dbo.leaderboard.*, dbo.deepLearning_hyp.activation, dbo.deepLearning_hyp.adaptive_rate, dbo.deepLearning_hyp.alsencoder, dbo.deepLearning_hyp.categorical_encoding, dbo.deepLearning_hyp.[checkpoint], dbo.deepLearning_hyp.class_sampling_factor, dbo.deepLearning_hyp.elasticnet, dbo.deepLearning_hyp.elasticnet_alpha, dbo.deepLearning_hyp.elasticnet_lambda, dbo.deepLearning_hyp.early_stopping, dbo.deepLearning_hyp.eigs, dbo.deepLearning_hyp.epochs, dbo.deepLearning_hyp.epsilon, dbo.deepLearning_hyp.fold_assignment, dbo.deepLearning_hyp.fold_column, dbo.deepLearning_hyp.fold_size, dbo.deepLearning_hyp.fool_assignment, dbo.deepLearning_hyp.fool_size, dbo.deepLearning_hyp.fool_type, dbo.deepLearning_hyp.ignored_columns, dbo.deepLearning_hyp.initial_weight, dbo.deepLearning_hyp.initial_weight_scale, dbo.deepLearning_hyp.initial_bias, dbo.deepLearning_hyp.ignored_columns, dbo.deepLearning_hyp.initial_weight, dbo.deepLearning_hyp.hids, dbo.deepLearning_hyp.hidden_dropout_ratio, dbo.deepLearning_hyp.initial_weight_distribution, dbo.deepLearning_hyp.hids, dbo.deepLearning_hyp.max_runtime_secs, dbo.deepLearning_hyp.max_hit_ratio_k, dbo.deepLearning_hyp.mini_batch_size, dbo.deepLearning_hyp.momentum_gamma, dbo.deepLearning_hyp.momentum_stable, dbo.deepLearning_hyp.momentum_start, dbo.deepLearning_hyp.offset_column, dbo.deepLearning_hyp.overwrite_with_best_model, dbo.deepLearning_hyp.pretrained_attributes, dbo.deepLearning_hyp.rate_annealing, dbo.deepLearning_hyp.rate_decay, dbo.deepLearning_hyp.regression_stop, dbo.deepLearning_hyp.response_column, dbo.deepLearning_hyp.rho, dbo.deepLearning_hyp.score_duty_cycle, dbo.deepLearning_hyp.validation_frame, dbo.deepLearning_hyp.use_all_factor_levels, dbo.deepLearning_hyp.tweedie_power, dbo.deepLearning_hyp.train_samples_per_iteration, dbo.deepLearning_hyp.training_frame, dbo.deepLearning_hyp.target_ratio, dbo.deepLearning_hyp.stopping_rounds, dbo.deepLearning_hyp.stopping_metric, dbo.deepLearning_hyp.standardize, dbo.deepLearning_hyp.score_validation_samples, dbo.deepLearning_hyp.single_node_mode, dbo.deepLearning_hyp.sparse, dbo.deepLearning_hyp.threads, dbo.deepLearning_hyp.weight_initializer
FROM        dbo.deepLearning_hyp INNER JOIN
                      dbo.leaderboard ON dbo.deepLearning_hyp.model_id = dbo.leaderboard.model_id
WHERE        activation in ('Tanh', 'TanhWithDropout', 'Rectifier', 'RectifierWithDropout',
                           'Maxout', 'MaxoutWithDropout')
AND epoch >= 9
AND epsilon >= 1.0e-10 AND epsilon <= 1.0e-4
AND quantile_alpha between 0 AND 1
AND initial_weight_distribution in ('Uniform', 'UniformAdaptive', 'Normal')
GO

```

Connected. (1/1)

SQLQuery180.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (58)) - Microsoft SQL Server Management Studio

```

SELECT TOP (1000) [column_1]
,[model_id]
,[auc]
,[logloss]
,[mean_per_class_error]
,[mse]
,[mse]
,[activation]
,[adaptive_rate]
,[autoencoder]
,[average_activation]
,[balance_classes]
,[categorical_encoding]
,[checkpoint]
,[class_sampling_factors]

```

column_1	Unnamed_0	model_id	auc	logloss	mean_per_class_error	mse
1	3	DeepLearning_1_AutoML_20190420_135144	0.90472327040788	0.319802261454825	0.188899955976145	0.319805145480323
2	3	DeepLearning_1_AutoML_20190420_135144	0.90472327040788	0.319802261454825	0.188899955976145	0.319805145480323

Query executed successfully.

### View3: v\_gbm\_lead

SQLQuery181.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (59)) - Microsoft SQL Server Management Studio

```

USE [Hyperparameter_PK]
GO

SET QUOTED_IDENTIFIER ON
GO

ALTER VIEW [dbo].[v_gbm_lead]
AS
SELECT      dbo.gbm_hyp.balance_classes, dbo.gbm_hyp.build_tree_one_node, dbo.gbm_hyp.calibrate_model, dbo.gbm_hyp.calibration_frame, dbo.gbm_hyp.[checkpoint], dbo.gbm_hyp.class_sampling_factors, dbo.gbm_hyp.col_sample_rate, dbo.gbm_hyp.col_sample_rate,
            dbo.gbm_hyp.distribution, dbo.gbm_hyp.export_checkpoints_dir, dbo.gbm_hyp.fold_assignment, dbo.gbm_hyp.fold_column, dbo.gbm_hyp.ignored_columns, dbo.gbm_hyp.keep_cross_validation_fold_assignment, dbo.gbm_hyp.keep_cross_validation_model,
            dbo.gbm_hyp.learn_rate_annealing, dbo.gbm_hyp.max_abs_leafnode_pred, dbo.gbm_hyp.max_after_balance_size, dbo.gbm_hyp.max_runtime_secs, dbo.gbm_hyp.min_rows, dbo.gbm_hyp.min_split_improvement, dbo.gbm_hyp.model_id, dbo.gbm_hyp.nbins_top_level, dbo.gbm_hyp.nfolds, dbo.gbm_hyp.ntrees, dbo.gbm_hyp.offset_column, dbo.gbm_hyp.pred_noise,
            dbo.gbm_hyp.response_column, dbo.gbm_hyp.sample_rate, dbo.gbm_hyp.sample_rate_per_class, dbo.gbm_hyp.score_each_iteration, dbo.gbm_hyp.weights_column, dbo.leaderboard.model_id AS Expr1, dbo.leaderboard.auc, dbo.leaderboard.logloss, dbo.leaderboard.r2
FROM        dbo.gbm_hyp INNER JOIN
                    dbo.leaderboard ON dbo.gbm_hyp.model_id = dbo.leaderboard.model_id
WHERE        ntrees >= 50
            AND distribution IN ('bernoulli', 'quasibinomial', 'multinomial', 'poisson',
            'laplace', 'tweedie', 'gaussian', 'huber', 'gamma', 'quantile')
            AND categorical_encoding IN ('AUTO', 'Enum', 'EnumUnlimited', 'OneHotExplicit',
            'Eigen', 'LabelEncoder', 'SortByResponse')
            AND min_split_improvement >= 1e-10 AND min_split_improvement <= 1e-3
GO

```

SQLQuery182.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (59)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK | Execute Debug | Quick Launch (Ctrl+Q)

Object Explorer

```

SELECT TOP 1000 [balance_classes]
      ,[build_tree_one_node]
      ,[calibrate_model]
      ,[calibration_frame]
      ,[categorical_encoding]
      ,[check_constant_response]
      ,[checkpoint]
      ,[class_sampling_factors]
      ,[col_sample_rate]
      ,[col_sample_rate_change_per_level]
      ,[custom_metric_func]
      ,[col_sample_rate_per_tree]
      ,[distribution]
      ,[export_checkpoints_dir]
      ,[fold_assignment]
      ,[fold_column]

```

Results | Messages

balance_classes	build_tree_one_node	calibrate_model	calibration_frame	categorical_encoding	check_constant_response	checkpoint	class_sampling_factors
1	0	0	NULL	AUTO	1	NULL	NULL
2	0	0	NULL	AUTO	1	NULL	NULL
3	0	0	NULL	AUTO	1	NULL	NULL

Query executed successfully.

is-swang01.ischool.uw.edu ... INFO6210 (59) Hyperparameter\_PK 00:00:00 | 3 rows

Ln 1 Col 1 Ch 1 INS

Type here to search

#### View 4: v\_glm\_xrt\_lead

SQLQuery183.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (60)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK | Execute Debug | Quick Launch (Ctrl+Q)

Object Explorer

```

ALTER VIEW [dbo].[v_glm_xrt_lead]
AS
SELECT      dbo.glm_hyp.alpha, dbo.glm_hyp.balance_classes, dbo.glm_hyp.beta_constraints, dbo.glm_hyp.beta_epsilon, dbo.glm_hyp.class_sampling
            dbo.glm_hyp.early_stopping, dbo.glm_hyp.export_checkpoints_dir, dbo.glm_hyp.family, dbo.glm_hyp.fold_assignment, dbo.glm_hyp.ignored_columns, dbo.glm_hyp.interactions, dbo.glm_hyp.intercept, dbo.glm_hyp.max_active_variables, dbo.glm_hyp.max_after_balance_size, dbo.glm_hyp.max_confusion_matrix_size, dbo.glm_hyp.max_depth, dbo.glm_hyp.max_iter, dbo.glm_hyp.max_leaves, dbo.glm_hyp.max_depth_ne, dbo.glm_hyp.max_pruned_leaves, dbo.glm_hyp.max_training_values, dbo.glm_hyp.mse, dbo.glm_hyp.n_folds, dbo.glm_hyp.n_classes, dbo.glm_hyp.n_leaves, dbo.glm_hyp.n_trees, dbo.glm_hyp.non_parallel_ne, dbo.glm_hyp.tweedie_variance_power, dbo.glm_hyp.validation_frame, dbo.glm_hyp.weights.columns, dbo.glm_hyp.standardize, dbo.randomForest_hyp.binomial_double_trees, dbo.randomForest_hyp.build_tree_one_node, dbo.randomForest_hyp.calibrate_model, dbo.randomForest_hyp.check_constant_response, dbo.randomForest_hyp.checkpoint, dbo.randomForest_hyp.class_sampling_factor, dbo.randomForest_hyp.col_sample_rate_per_tree, dbo.randomForest_hyp.custom_metric_func, dbo.randomForest_hyp.fold_assignment AS rf_fold_assignment, dbo.randomForest_hyp.histogram_type, dbo.randomForest_hyp.f, dbo.randomForest_hyp.ignored_columns AS rf_ignored_columns, dbo.randomForest_hyp.keep_cross_validation_fold_assignment, dbo.randomForest_hyp.max_after_balance_size AS rf_max_after_balance_size, dbo.randomForest_hyp.max_after_bal, dbo.randomForest_hyp.max_hit_ratio_k AS rf_max_hit_ratio_k, dbo.randomForest_hyp.max_runtime_secs AS rf_max_runtime_sec, dbo.randomForest_hyp.model_id AS rf_model_id, dbo.randomForest_hyp.mtries, dbo.randomForest_hyp.nbins, dbo.randomForest_hyp.ntrees, dbo.randomForest_hyp.offset_column AS rf_offset_column, dbo.randomForest_hyp.r2_stopping, dbo.randomForest_hyp.sample_rate_per_class, dbo.randomForest_hyp.score_each_iteration AS rf_score_each_iteration, dbo.randomForest_hyp.stopping_rounds, dbo.randomForest_hyp.stop_training, dbo.randomForest_hyp.training_frame AS rf_training_frame, dbo.randomForest_hyp.weights_column AS rf_weights_column, dbo.leaderboard.model_id AS lr_model_id, dbo.leaderboard.auc, dbo.leaderboard.mse
FROM        dbo.glm_hyp INNER JOIN
                    dbo.leaderboard ON dbo.glm_hyp.model_id = dbo.leaderboard.model_id full join
                    dbo.randomForest_hyp ON dbo.leaderboard.model_id = dbo.randomForest_hyp.model_id
                    WHERE mtries between -1 and 7
                    AND histogram_type like '%AUTOS' or histogram_type like '%Uniform%' or histogram_type like '%Random%' or histogram_type like '%Global%' or histogram_type like '%RoundRobin%' AND theta > 0 AND solver in ('AUTO', 'IRLSH', 'LBFGS', 'COORDINATE_DESCENT_NAIVE', 'COORDINATE_DESCENT', 'GRADIENT_DESCENT_LH') AND logloss >= 0.3 AND mse >= 0.1

```

Connected. (1/1)

is-swang01.ischool.uw.edu ... INFO6210 (60) Hyperparameter\_PK 00:00:00 | 0 rows

Ln 50 Col 1 Ch 1 INS

Type here to search

SQLQuery184.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (61)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparamter\_PK | Execute Debug | Quick Launch (Ctrl+Q)

Object Explorer

```
----- script for SelectTopNRows command from SSMS -----
SELECT TOP 1000 [alpha]
,[balance_classes]
,[beta_constraints]
,[beta_epsilon]
,[class_sampling_factors]
,[compute_p_values]
,[custom_metric_func]
,[early_stopping]
,[export_checkpoints_dir]
,[family]
,[fold_assignment]
,[fold_column]
,[gradient_epsilon]
,[ignore_const_cols]
,[ignored_columns]
,[interaction_pairs]
```

Results | Messages

	distribution	rf_export_checkpoints_dir	rf_fold_assignment	histogram_type	rf_fold_column	rf_ignore_const_cols	rf_ignored_columns	keep
1	multinomial	NULL	Modulo	AUTO	NULL	1	[hours-per-week', 'gender', 'capital-loss]	0
2	AUTO	NULL	AUTO	AUTO	NULL	1	NULL	0
3	multinomial	NULL	Modulo	AUTO	NULL	1	[education', 'hours-per-week', 'capital-loss]	0
4	AUTO	NULL	AUTO	AUTO	NULL	1	NULL	0
5	multinomial	NULL	Modulo	AUTO	NULL	1	[education', 'hours-per-week', 'capital-loss]	0
6	AUTO	NULL	AUTO	AUTO	NULL	1	NULL	0

Query executed successfully.

is-swang01.ischool.uw.edu (INFO6210 (61) | Hyperparamter\_PK | 00:00:00 | 6 rows)

Ready | is-swang01.ischool.uw.edu (11.0 RTM) | Ch 1 | INS | ENG 11:50 PM IN 4/24/2019 | 19

Type here to search

## View5: v\_stack\_lead

SQLQuery186.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (64)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparamter\_PK | Execute Debug | Quick Launch (Ctrl+Q)

Object Explorer

```
----- Object: View [dbo].[v_stack_lead] Script Date: 4/24/2019 11:50:57 PM -----
USE [Hyperparamter_PK]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

ALTER VIEW [dbo].[v_stack_lead]
AS
SELECT      dbo.leaderboard.model_id, dbo.leaderboard.auc, dbo.leaderboard.logloss, dbo.leaderboard.mean_per_class_error, dbo.leaderboard.rmse
            ,dbo.stackedEnsemble_hyp.blending_frame, dbo.stackedEnsemble_hyp.export_checkpoints_dir, dbo.stackedEnsemble_hyp.response
            ,dbo.stackedEnsemble_hyp.metalearner_params, dbo.stackedEnsemble_hyp.metalearner_nfolds, dbo.stackedEnsemble_hyp.metalearner_algorithm, dbo.stackedEnsemble_hyp.keep_levelone_frame, dbo.stackedEnsemble_hyp.validation
            ,dbo.stackedEnsemble_hyp.training_nfolds
            ,dbo.leaderboard ON dbo.stackedEnsemble_hyp.model_id = dbo.leaderboard.model_id
WHERE training_frame is not null
AND metalearner_nfolds >= 0
AND metalearner_algorithm in ('AUTO', 'glm', 'gbm', 'drf', 'deeplearning')
AND base_models is not null
GO
```

Connected. (1/1) | is-swang01.ischool.uw.edu (INFO6210 (64) | Hyperparamter\_PK | 00:00:00 | 0 rows)

Ready | is-swang01.ischool.uw.edu (11.0 RTM) | Ch 1 | INS | LN 32 | Col 1 | Ch 1 | ENG 11:51 PM IN 4/24/2019 | 19

Type here to search

SQLQuery185.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (62)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK Execute Debug

Object Explorer

SQLQuery185.sql - ...\_PK (INFO6210 (62)) SQLQuery184.sql - ...\_PK (INFO6210 (61)) SQLQuery183.sql - ...\_PK (INFO6210 (60))

\*\*\*\*\* script for SelectTopNRows command from SSMS \*\*\*\*\*

```
SELECT TOP (1000) [model_id]
,[auc]
,[logloss]
,[mean_per_class_error]
,[rmse]
,[mse]
,[base_models]
,[blending_frame]
,[export_checkpoints_dir]
,[response_column]
,[se_model_id]
,[metalearner_params]
,[metalearner_nfolds]
,[metalearner_fold_column]
,[metalearner_fold_assignment]
,[metalearner_algorithm]
```

Results Messages

model_id	auc	logloss	mean_per_class_error	rmse	mse
1	StackedEnsemble_BestOfFamily_AutoML_20190420_135...	0.9165588028671	0.31410496211009	0.17909310144358	0.311726541590557
2	StackedEnsemble_BestOfFamily_AutoML_20190420_135...	0.9165588028671	0.31410496211009	0.17909310144358	0.311726541590557
3	StackedEnsemble_BestOfFamily_AutoML_20190420_151...	0.916434601545767	0.314268699029281	0.180125407362866	0.31189831533566
4	StackedEnsemble_BestOfFamily_AutoML_20190420_151...	0.916434601545767	0.314268699029281	0.180125407362866	0.31189831533566
5	StackedEnsemble_BestOfFamily_AutoML_20190420_162...	0.915920603777968	0.315655796158604	0.18388153701369	0.312316401275908
6	StackedEnsemble_BestOfFamily_AutoML_20190420_162...	0.915920603777968	0.315655796158604	0.18388153701369	0.312316401275908

Query executed successfully.

is-swang01.ischool.uw.edu (INFO6210 (62)) Hyperparameter\_PK 00:00:00 6 rows

Ln 1 Col 1 Ch 1 INS

Ready Type here to search

Windows Taskbar: File Explorer, Edge, Chrome, File Manager, Task View, Taskbar Icons, Start button, Search bar, System tray (Time, Date, Network)

## 5 Functions:

### Functions 1

#### f\_gbm\_lead

SQLQuery187.sql - is-swang01.ischool.uw.edu.Hyperparameter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio

File Edit View Query Project Debug Tools Window Help

Hyperparameter\_PK Execute Debug

Object Explorer

SQLQuery187.sql - ...\_PK (INFO6210 (54))

```
-- Author: <Author>,<Name>
-- Create date: <Create Date, >
-- Description: <Description, >
-----
```

```
ALTER FUNCTION [dbo].[f_gbm_lead]
(
    -- Add the parameters for the function here
    @learn_rate float
)
RETURNS int
AS
BEGIN
    -- Declare the return variable here
    DECLARE @auc int

    -- Add the T-SQL statements to compute the return value here
    SELECT @auc = max(Id.auc)
    FROM   dbo.gbm_gb gb INNER JOIN
           dbo.leaderboard ld ON gb.model_id = ld.model_id
    WHERE distribution in ('bernoulli', 'quasibinomial', 'multinomial', 'poisson',
                           'laplace', 'tweedie', 'gaussian', 'huber', 'gamma', 'quantile')
      AND min_split_improvement >= 1e-10 AND min_split_improvement <= 1e-3
      AND learn_rate = @learn_rate
      Group by ld.auc
      IF (@auc IS NULL)
      SET @auc = 0;

    -- Return the result of the function
    RETURN @auc
END
GO
```

Connected. (1/1) is-swang01.ischool.uw.edu (INFO6210 (54)) Hyperparameter\_PK 00:00:00 0 rows

Ln 46 Col 1 Ch 1 INS

Ready Type here to search

Windows Taskbar: File Explorer, Edge, Chrome, File Manager, Task View, Taskbar Icons, Start button, Search bar, System tray (Time, Date, Network)

## Function2: f\_dl\_hidden

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery189.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio". The main window displays the T-SQL code for the function:

```
-- Author: <Author>,<Name>
-- Create date: <Create Date> ,>
-- Description: <Description> ,>
-----
ALTER FUNCTION [dbo].[f_dl_hidden]
(
    -- Add the parameters for the function here
    @epochs float
)
RETURNS nvarchar(50)
AS
BEGIN
    -- Declare the return variable here
    DECLARE @hidden nvarchar(50)

    -- Add the T-SQL statements to compute the return value here
    SELECT @hidden = hidden
    FROM dbo.deepLearning_hyp dl INNER JOIN
         dbo.leaderboard ld ON dl.model_id = ld.model_id
    WHERE quantile_alpha between 0 AND 1
        AND initial_weight_distribution IN ('Uniform', 'UniformAdaptive', 'Normal')
        AND epochs = @epochs
    IF (@hidden IS NULL)
        SET @hidden = 0

    -- Return the result of the function
    RETURN @hidden
END
GO
```

The status bar at the bottom shows "Connected. (1/1)" and "is-swang01.ischool.uw.edu (... | INFO6210 (54) | Hyperparamter\_PK | 00:00:00 | 0 rows". The bottom right corner shows "Ln 44 Col 1 Ch 1 INS" and "ENG 11:54 PM IN 4/24/2019 19".

## Function3: f\_dt\_com

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery190.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio". The main window displays the T-SQL code for the function:

```
-- Create date: <Create Date>,>
-- Description: <Description>,>
-----
ALTER FUNCTION [dbo].[f_dt_com]
(
    -- Add the parameters for the function here
    @rmse int,
    @auc int
)
RETURNS TABLE
AS
RETURN

    -- Fill the table variable with the rows for your result set
    SELECT gb.categorical_encoding, gb.ignored_columns, rf.build_tree_one_node, rf.fold_assignment,
           xr.mtries, xr.bnins_cats, xr.nntrees
    FROM
        dbo.leaderboard ld Full JOIN
        dbo.gbm_hyp gb ON ld.model_id = gb.model_id Full JOIN
        dbo.randomForest_hyp rf ON ld.model_id = rf.model_id full JOIN
        dbo.xrt_hyp xr ON ld.model_id = xr.model_id

    WHERE ld.rmse = @rmse
    AND ld.auc = @auc
    AND gb.categorical_encoding like '%AUTOS'
    OR
    rf.distribution in ('multinomial', 'Auto')

GO
```

The status bar at the bottom shows "Connected. (1/1)" and "is-swang01.ischool.uw.edu (... | INFO6210 (54) | Hyperparamter\_PK | 00:00:00 | 0 rows". The bottom right corner shows "Ln 47 Col 1 Ch 1 INS" and "ENG 11:55 PM IN 4/24/2019 19".

#### Function4: f\_pred\_model

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists various database objects like System Stored Procedures, Functions, and Types. The central query editor window displays the T-SQL code for the function f\_pred\_model. The code includes comments for author, create date, and description, followed by an ALTER FUNCTION definition. It adds parameters for relationship and education, returns a table, and fills a table variable with rows from the dbo.predictions table based on workclass, predict, income, and hours\_per\_week. A GO statement is at the end.

```
--> <Author>,<Name>
--> <Create Date>,<>
--> <Description>,<>
=====
ALTER FUNCTION [dbo].[f_pred_model]
-- Add the parameters for the function here
@relationship nvarchar(50),
@education nvarchar(50)
)

RETURNS TABLE
AS
RETURN
(
-- Fill the table variable with the rows for your result set
SELECT workclass, predict, income, SUM(hours_per_week) AS Total_hours
FROM dbo.predictions
WHERE education = @education
AND relationship = @relationship
GROUP BY workclass, predict, income, hours_per_week
HAVING hours_per_week >= 20
)
GO
```

#### Function5: f\_se\_lead

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists various database objects. The central query editor window displays the T-SQL code for the function f\_se\_lead. It starts with USE [Hyperparameter\_PK] and GO. It includes comments for author, create date, and description, followed by an ALTER FUNCTION definition. It adds parameters for model\_id and metalearner\_algorithm. It returns a table and fills a table variable with rows from the dbo.stackedsensemble\_hyp table, joining it with the dbo.leadboard table on se.model\_id = ld.model\_id, where se.model\_id = @model\_id and metalearner\_algorithm = @metalearner\_algorithm. The results are grouped by base\_models, logloss, and metalearner\_nfolds. A GO statement is at the end.

```
USE [Hyperparameter_PK]
GO

===== Object: UserDefinedFunction [dbo].[f_se_lead] Script Date: 4/24/2019 11:57:08 PM =====
SET ANSI_NULLS ON
GO

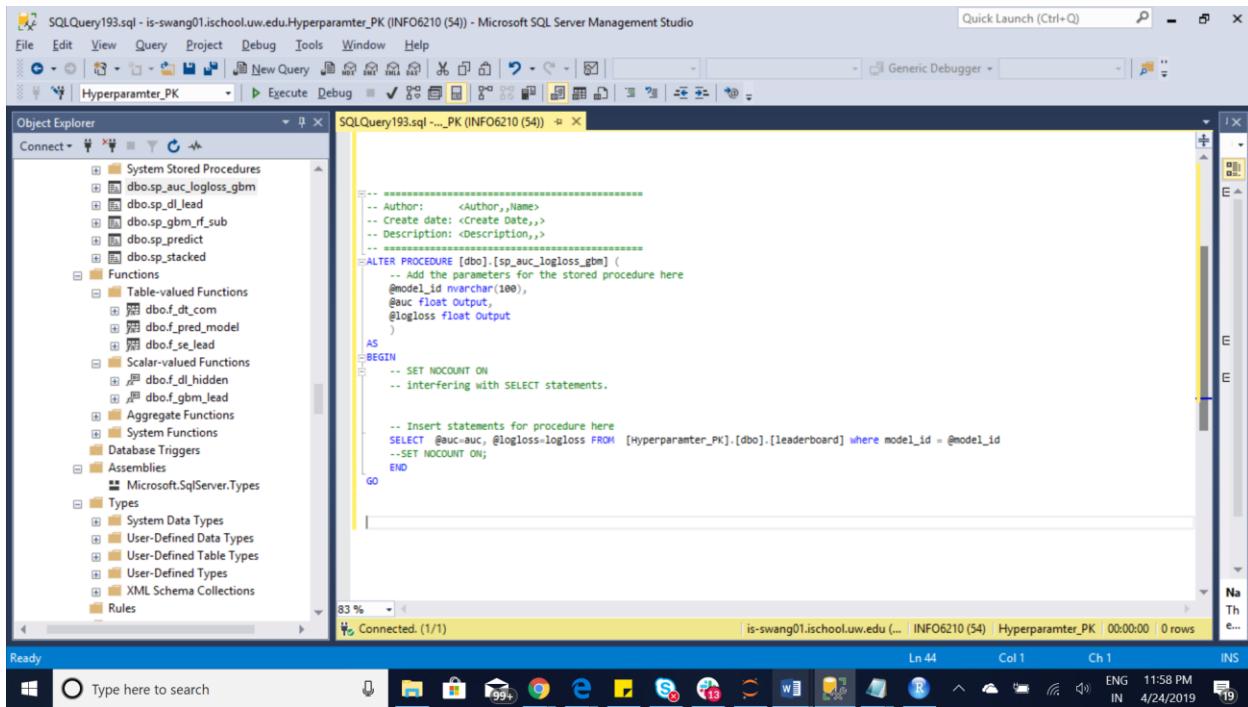
SET QUOTED_IDENTIFIER ON
GO

--> <Author>,<Name>
--> <Create Date>,<>
--> <Description>,<>
=====

ALTER FUNCTION [dbo].[f_se_lead]
(
-- Add the parameters for the function here
@model_id nvarchar(100),
@metalearner_algorithm nvarchar(50)
)
RETURNS TABLE
AS
RETURN
(
-- Fill the table variable with the rows for your result set
SELECT MAX(mse) AS Maximum_MSE, logloss, metalearner_nfolds, base_models
FROM dbo.stackedsensemble_hyp se INNER JOIN
dbo.leadboard ld ON se.model_id = ld.model_id
WHERE se.model_id = @model_id
AND metalearner_algorithm = @metalearner_algorithm
GROUP BY base_models, logloss, metalearner_nfolds
)
GO
```

## 5 stored Procedures

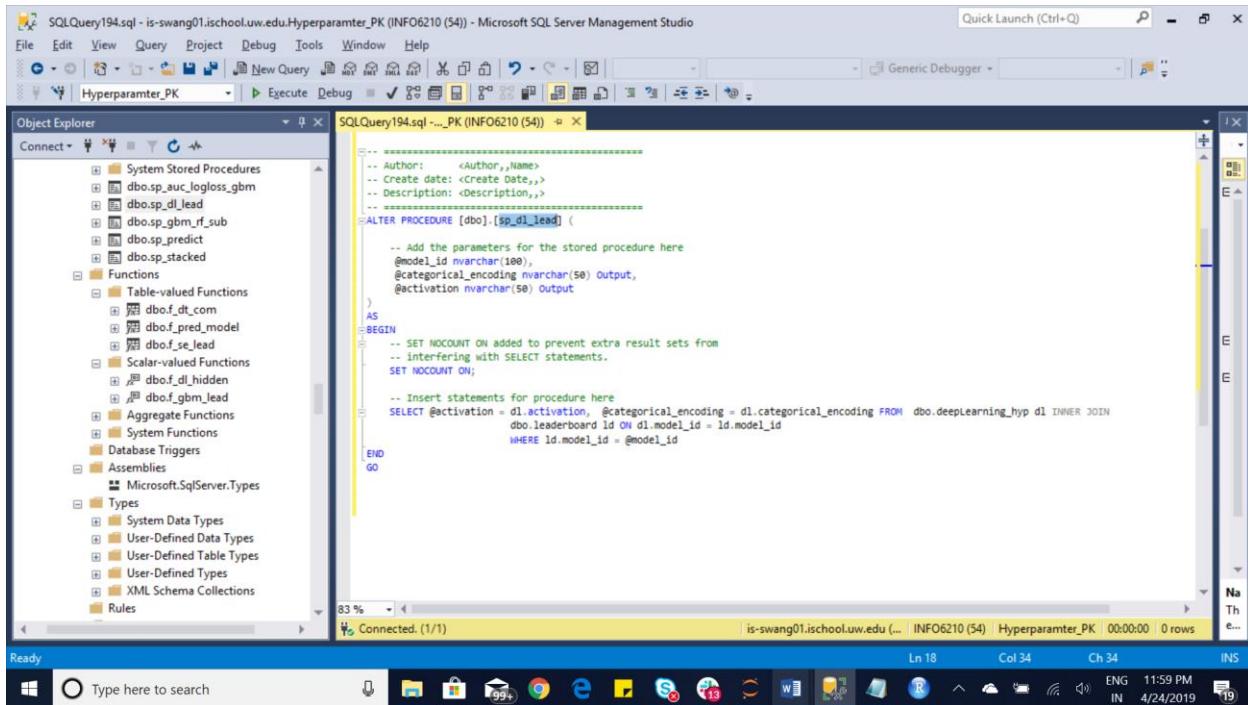
### Stored Procedure 1: sp\_auc\_logloss\_gbm



```
-- Author: <Author,,Name>
-- Create date: <Create Date,,>
-- Description: <Description,,>
=====
ALTER PROCEDURE [dbo].[sp_auc_logloss_gbm] (
    -- Add the parameters for the stored procedure here
    @model_id nvarchar(100),
    @auc float output,
    @logloss float output
)
AS
BEGIN
    -- SET NOCOUNT ON
    -- interfering with SELECT statements.

    -- Insert statements for procedure here
    SELECT @auc=auc, @logloss=logloss FROM [Hyperparamter_PK].[dbo].[leaderboard] where model_id = @model_id
    --SET NOCOUNT ON;
END
GO
```

### S. Procedure2: sp\_dl\_lead



```
-- Author: <Author,,Name>
-- Create date: <Create Date,,>
-- Description: <Description,,>
=====
ALTER PROCEDURE [dbo].[sp_dl_lead] (
    -- Add the parameters for the stored procedure here
    @model_id nvarchar(100),
    @categorical_encoding nvarchar(50) Output,
    @activation nvarchar(50) Output
)
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    -- Insert statements for procedure here
    SELECT @activation = dl.activation, @categorical_encoding = dl.categorical_encoding FROM dbo.deepLearning_hyp dl INNER JOIN
    dbo.leaderboard ld ON dl.model_id = ld.model_id
    WHERE ld.model_id = @model_id
END
GO
```

### S.Procedure3: sp\_gbm\_rf\_sub

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery195.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio". The Object Explorer pane on the left lists various database objects including System Stored Procedures, Functions, Types, and Assemblies. The central query editor window contains the T-SQL code for the stored procedure sp\_gbm\_rf\_sub. The code includes comments for author, create date, and description, followed by the ALTER PROCEDURE definition. It adds parameters for distribution, nfold\_assignment, response\_column, and ignored\_columns. The body of the procedure sets NOCOUNT ON, inserts statements for procedure logic, and ends with GO.

```
-- Author: <Author>,<Name>
-- Create date: <Create Date>,>
-- Description: <Description>,>
=====
ALTER PROCEDURE [dbo].[sp_gbm_rf_sub]
    -- Add the parameters for the stored procedure here
    @distribution nvarchar(50),
    @fold_assignment nvarchar(50),
    @response_column nvarchar(50) output,
    @ignored_columns nvarchar(50) output
)
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    -- Insert statements for procedure here
    SELECT @response_column = response_column , @ignored_columns = ignored_columns FROM [Hyperparamter_PK].[dbo].[gbm_hyp]
    WHERE distribution = @distribution
    AND min_split_improvement >= 1E-10 AND min_split_improvement <= 1E-05
    AND nFolds IN (select nFolds FROM [Hyperparamter_PK].[dbo].[randomForest_hyp] WHERE ntrees >= 50)
    AND distribution = @distribution
    AND fold_assignment = @fold_assignment
END
GO
```

### S.Procedure4: sp\_predict

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery196.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio". The Object Explorer pane on the left lists various database objects. The central query editor window contains the T-SQL code for the stored procedure sp\_predict. The code includes comments for author, create date, and description, followed by the ALTER PROCEDURE definition. It adds parameters for occupation and relationship. The body of the procedure sets QUOTED\_IDENTIFIER ON, sets NOCOUNT ON, inserts statements for procedure logic, and ends with GO.

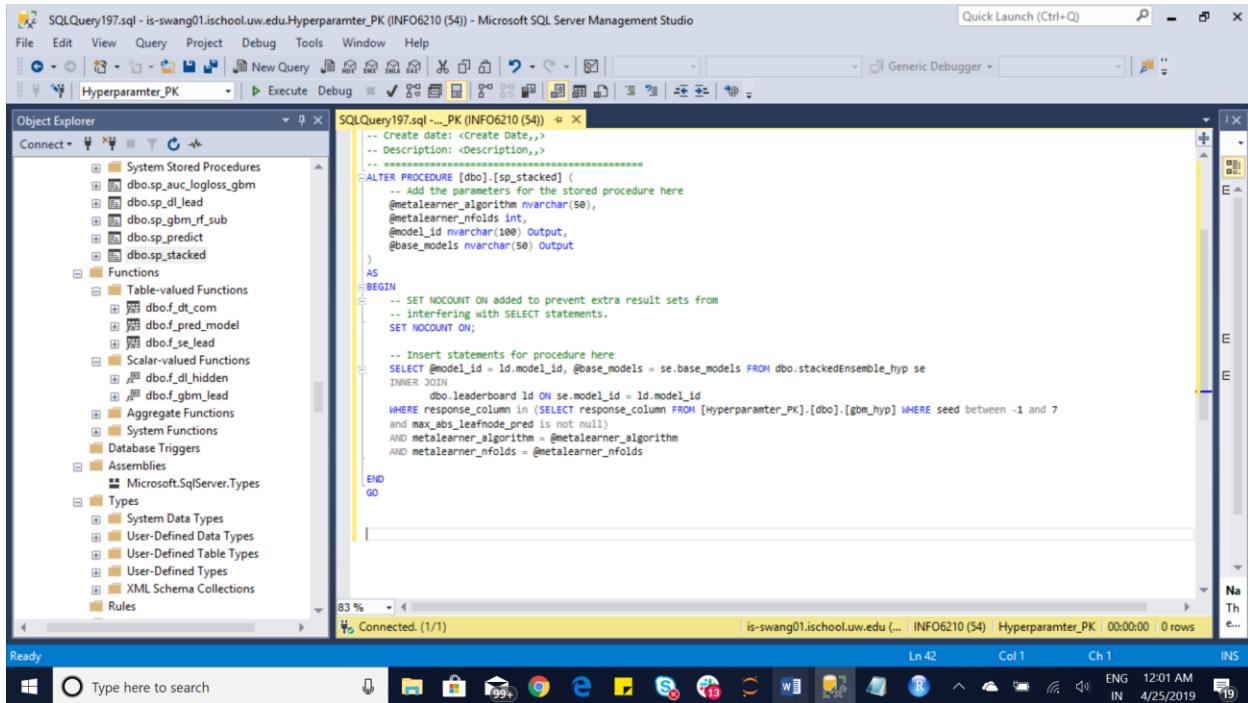
```
SET QUOTED_IDENTIFIER ON
GO

=====
-- Author: <Author>,<Name>
-- Create date: <Create Date>,>
-- Description: <Description>,>
=====
ALTER PROCEDURE [dbo].[sp_predict]
    -- Add the parameters for the stored procedure here
    @occupation nvarchar(50),
    @relationship nvarchar(50)

)
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    -- Insert statements for procedure here
    SELECT age,workclass , predict , count(capital_gain)
    FROM [Hyperparamter_PK].[dbo].[predictions] WHERE occupation = @occupation and relationship = @relationship
    group by age,workclass , capital_gain , predict
END
GO
```

## S. Procedure 5: sp\_stacked



The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery197.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54)) - Microsoft SQL Server Management Studio". The Object Explorer pane on the left lists various database objects: System Stored Procedures (dbo.sp\_auc\_logloss\_gbm, dbo.sp\_dl\_lead, dbo.sp\_gbm\_rf\_sub, dbo.sp\_predict, dbo.sp\_stacked), Functions (Table-valued Functions: dbo.f\_dt\_com, dbo.f\_pred\_model, dbo.f\_se\_lead, Scalar-valued Functions: dbo.f\_dl\_hidden, dbo.f\_gbm\_lead, Aggregate Functions, System Functions, Database Triggers, Assemblies, Microsoft.SqlServer.Types), Types (System Data Types, User-Defined Data Types, User-Defined Table Types, User-Defined Types, XML Schema Collections, Rules). The main results pane displays the script for the stored procedure sp\_stacked:

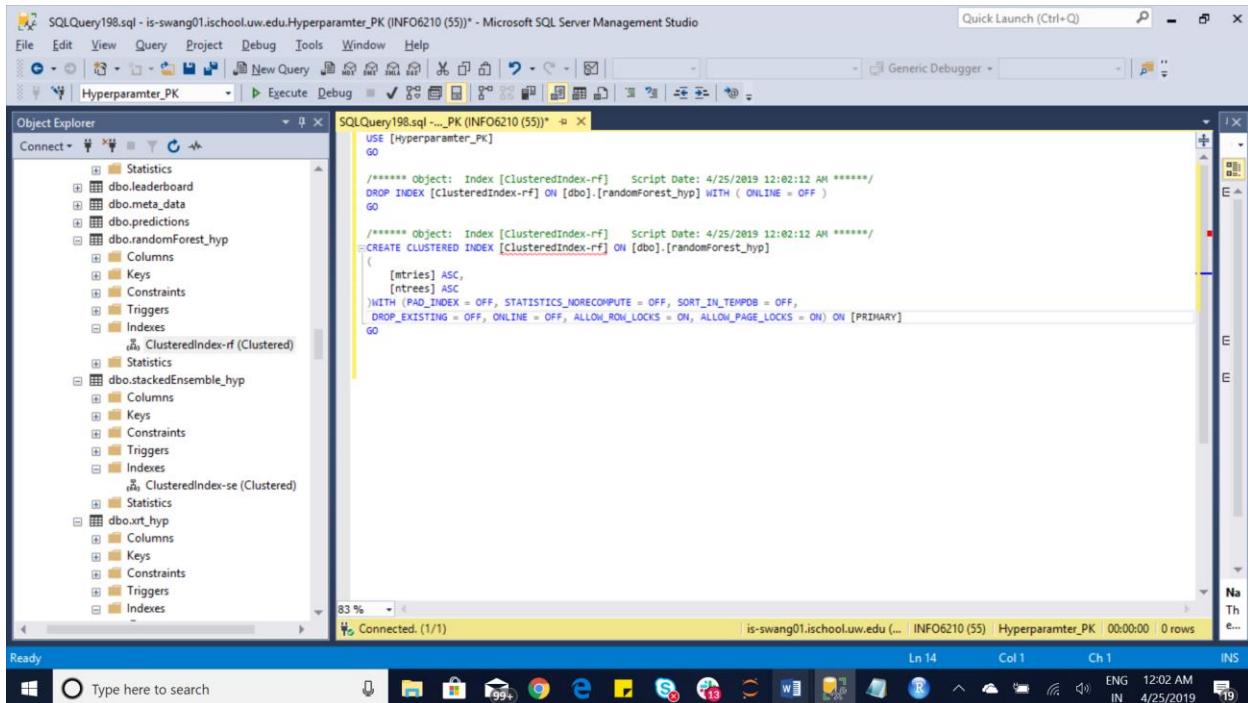
```
-- Create date: <create Date>
-- Description: <Description>
=====
ALTER PROCEDURE [dbo].[sp_stacked] (
    -- Add the parameters for the stored procedure here
    @metalearnern_algorithm nvarchar(50),
    @metalearnern_nfolds int,
    @model_id nvarchar(100) Output,
    @base_models nvarchar(50) Output
)
AS
BEGIN
    -- SET NOCOUNT ON added to prevent extra result sets from
    -- interfering with SELECT statements.
    SET NOCOUNT ON;

    -- Insert statements for procedure here
    SELECT @model_id = ld.model_id, @base_models = se.base_models
    FROM dbo.stackedEnsemble_hyp se
    INNER JOIN
        dbo.leaderboard ld ON se.model_id = ld.model_id
    WHERE response_column IN (SELECT response_column FROM [Hyperparamter_PK].[dbo].[gbm_hyp] WHERE seed between -1 and 7
    and max_abs_leafnode_pred is not null)
    AND metalearnern_algorithm = @metalearnern_algorithm
    AND metalearnern_nfolds = @metalearnern_nfolds

END
GO
```

## 5 Indexes:

### Index 1: ClusteredIndex-rf



The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery198.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (55)) - Microsoft SQL Server Management Studio". The Object Explorer pane on the left lists various database objects: Statistics (dbo.leaderboard, dbo.meta\_data, dbo.predictions, dbo.randomForest\_hyp, dbo.stackEnsemble\_hyp, dbo.stackEnsemble\_rf, dbo.stackEnsemble\_se), Keys, Constraints, Triggers, Indexes (ClusteredIndex-rf (Clustered), ClusteredIndex-se (Clustered)), Statistics (dbo.stackEnsemble\_hyp, dbo.stackEnsemble\_rf, dbo.stackEnsemble\_se), dbo.art\_hyp, dbo.gbm\_hyp, dbo.rf\_hyp. The main results pane displays the script for creating the ClusteredIndex-rf:

```
USE [Hyperparamter_PK]
GO

===== Object: Index [ClusteredIndex-rf] Script Date: 4/25/2019 12:02:12 AM =====/
DROP INDEX [ClusteredIndex-rf] ON [dbo].[randomForest_hyp] WITH ( ONLINE = OFF )
GO

===== Object: Index [ClusteredIndex-rf] Script Date: 4/25/2019 12:02:12 AM =====/
CREATE CLUSTERED INDEX [ClusteredIndex-rf] ON [dbo].[randomForest_hyp]
(
    [mtries] ASC,
    [ntrees] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
    DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
```

## Index2: ClusteredIndex-se

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer window on the left displays the database structure for 'Hyperparamter\_PK'. The 'Indexes' node under 'dbo.stackedEnsemble\_hyp' has two entries: 'ClusteredIndex-rf (Clustered)' and 'ClusteredIndex-se (Clustered)'. The 'ClusteredIndex-se (Clustered)' entry is selected. The central query editor window contains the T-SQL script for creating this index:

```
USE [Hyperparamter_PK]
GO

/*===== Object: Index [ClusteredIndex-se] Script Date: 4/25/2019 12:02:56 AM =====*/
DROP INDEX [ClusteredIndex-se] ON [dbo].[stackedEnsemble_hyp] WITH ( ONLINE = OFF )
GO

SET ANSI_PADDING ON
GO

/*===== Object: Index [ClusteredIndex-se] Script Date: 4/25/2019 12:02:56 AM =====*/
CREATE CLUSTERED INDEX [ClusteredIndex-se] ON [dbo].[stackedEnsemble_hyp]
(
    [base_models] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF, ONLINE = OFF,
ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
```

The status bar at the bottom indicates the connection is 'Connected. (1/1)', the server is 'is-swang01.ischool.uw.edu', the database is 'INFO6210 (55)', the schema is 'Hyperparamter\_PK', and there are '0 rows' affected.

## Index3: ClusteredIndex-xr

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer window on the left displays the database structure for 'Hyperparamter\_PK'. The 'Indexes' node under 'dbo.xrt\_hyp' has one entry: 'ClusteredIndex-xr (Clustered)'. This entry is selected. The central query editor window contains the T-SQL script for creating this index:

```
USE [Hyperparamter_PK]
GO

/*===== Object: Index [ClusteredIndex-xr] Script Date: 4/25/2019 12:03:49 AM =====*/
DROP INDEX [ClusteredIndex-xr] ON [dbo].[xrt_hyp] WITH ( ONLINE = OFF )
GO

/*===== Object: Index [ClusteredIndex-xr] Script Date: 4/25/2019 12:03:49 AM =====*/
CREATE CLUSTERED INDEX [ClusteredIndex-xr] ON [dbo].[xrt_hyp]
(
    [build_tree_one_node] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF,
DROP_EXISTING = OFF, ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
```

The status bar at the bottom indicates the connection is 'Connected. (1/1)', the server is 'is-swang01.ischool.uw.edu', the database is 'INFO6210 (55)', the schema is 'Hyperparamter\_PK', and there are '0 rows' affected.

## Index4: [ClusteredIndex-md]

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery201.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (55))\* - Microsoft SQL Server Management Studio". The Object Explorer on the left shows a database structure with several tables like gbm\_hyp, glm\_hyp, and randomForest\_hyp, each containing columns, keys, constraints, triggers, and indexes. The central query editor window displays the following T-SQL script:

```
USE [Hyperparamter_PK]
GO

/*===== Object: Index [ClusteredIndex-md] Script Date: 4/25/2019 12:04:46 AM =====*/
DROP INDEX [ClusteredIndex-md] ON [dbo].[gbm_hyp] WITH ( ONLINE = OFF )
GO

/*===== Object: Index [ClusteredIndex-md] Script Date: 4/25/2019 12:04:46 AM =====*/
CREATE CLUSTERED INDEX [ClusteredIndex-md] ON [dbo].[gbm_hyp]
(
    [ntrees] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF,
    ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
```

The status bar at the bottom indicates "Connected. (1/1)" and "is-swang01.ischool.uw.edu (... | INFO6210 (55) | Hyperparamter\_PK | 00:00:00 | 0 rows". The taskbar at the bottom right shows the date and time as "4/25/2019 12:04 AM".

## Index5: ClusteredIndex-gl

The screenshot shows the Microsoft SQL Server Management Studio interface. The title bar reads "SQLQuery202.sql - is-swang01.ischool.uw.edu.Hyperparamter\_PK (INFO6210 (54))\* - Microsoft SQL Server Management Studio". The Object Explorer on the left shows a similar database structure to the previous screenshot. The central query editor window displays the following T-SQL script:

```
USE [Hyperparamter_PK]
GO

/*===== Object: Index [ClusteredIndex-gl] Script Date: 4/25/2019 12:05:30 AM =====*/
DROP INDEX [ClusteredIndex-gl] ON [dbo].[glm_hyp] WITH ( ONLINE = OFF )
GO

SET ANSI_PADDING ON
GO

/*===== Object: Index [ClusteredIndex-gl] Script Date: 4/25/2019 12:05:30 AM =====*/
CREATE CLUSTERED INDEX [ClusteredIndex-gl] ON [dbo].[glm_hyp]
(
    [alpha] ASC
)WITH (PAD_INDEX = OFF, STATISTICS_NORECOMPUTE = OFF, SORT_IN_TEMPDB = OFF, DROP_EXISTING = OFF,
    ONLINE = OFF, ALLOW_ROW_LOCKS = ON, ALLOW_PAGE_LOCKS = ON) ON [PRIMARY]
GO
```

The status bar at the bottom indicates "Connected. (1/1)" and "is-swang01.ischool.uw.edu (... | INFO6210 (54) | Hyperparamter\_PK | 00:00:00 | 0 rows". The taskbar at the bottom right shows the date and time as "4/25/2019 12:05 AM".