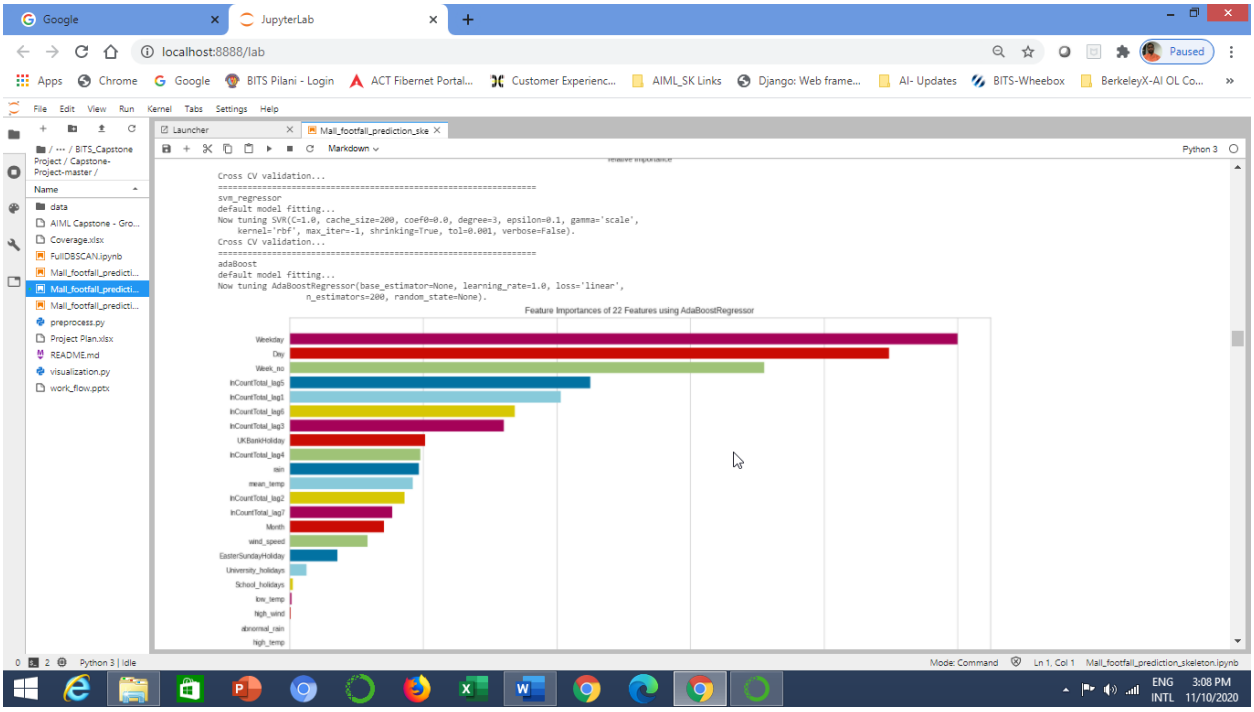
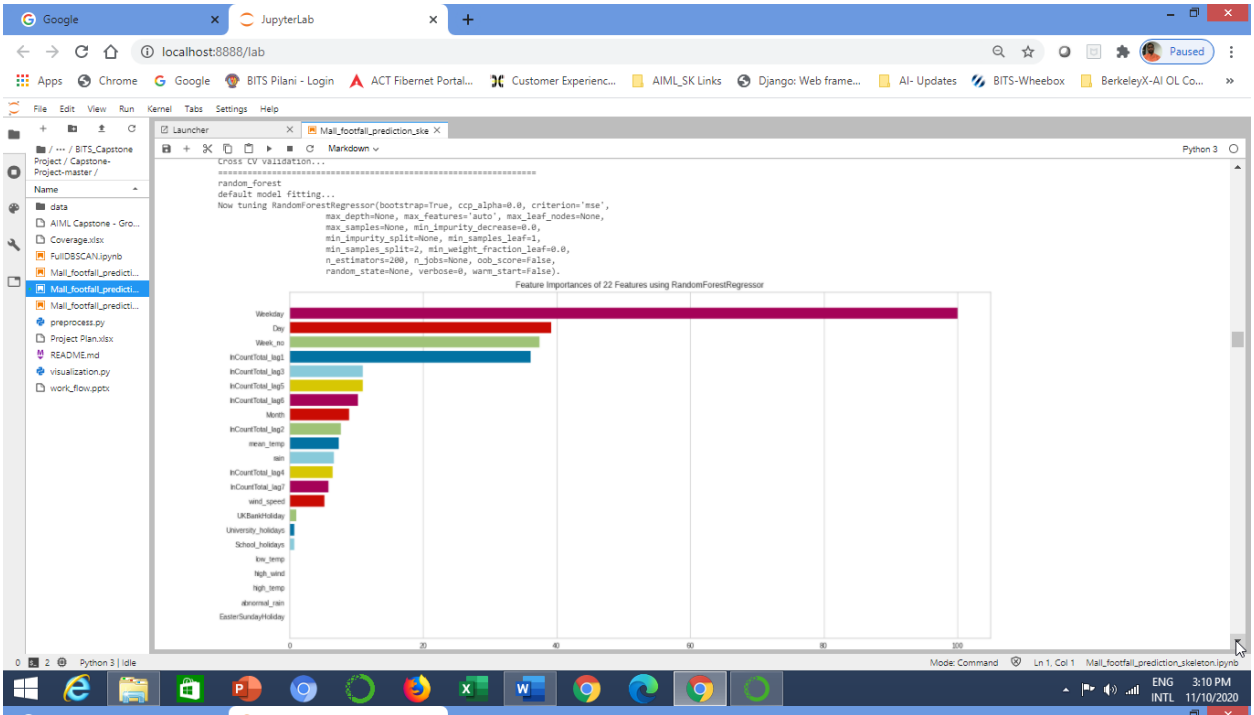
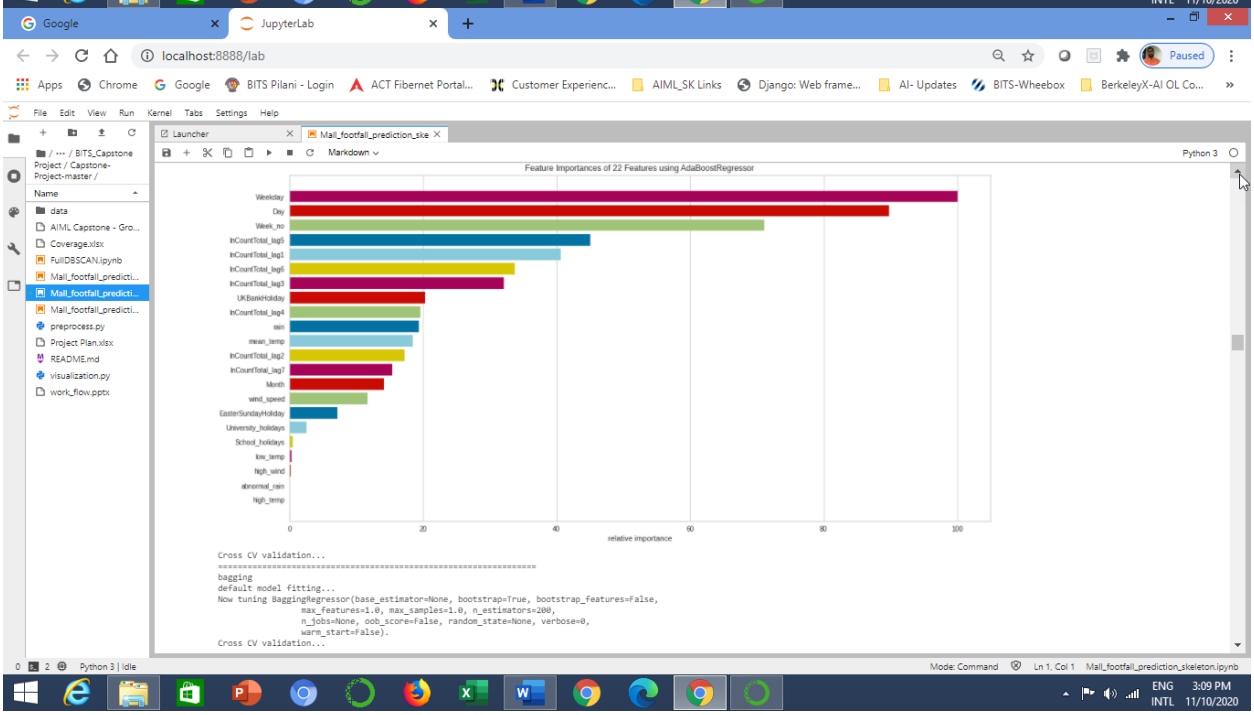
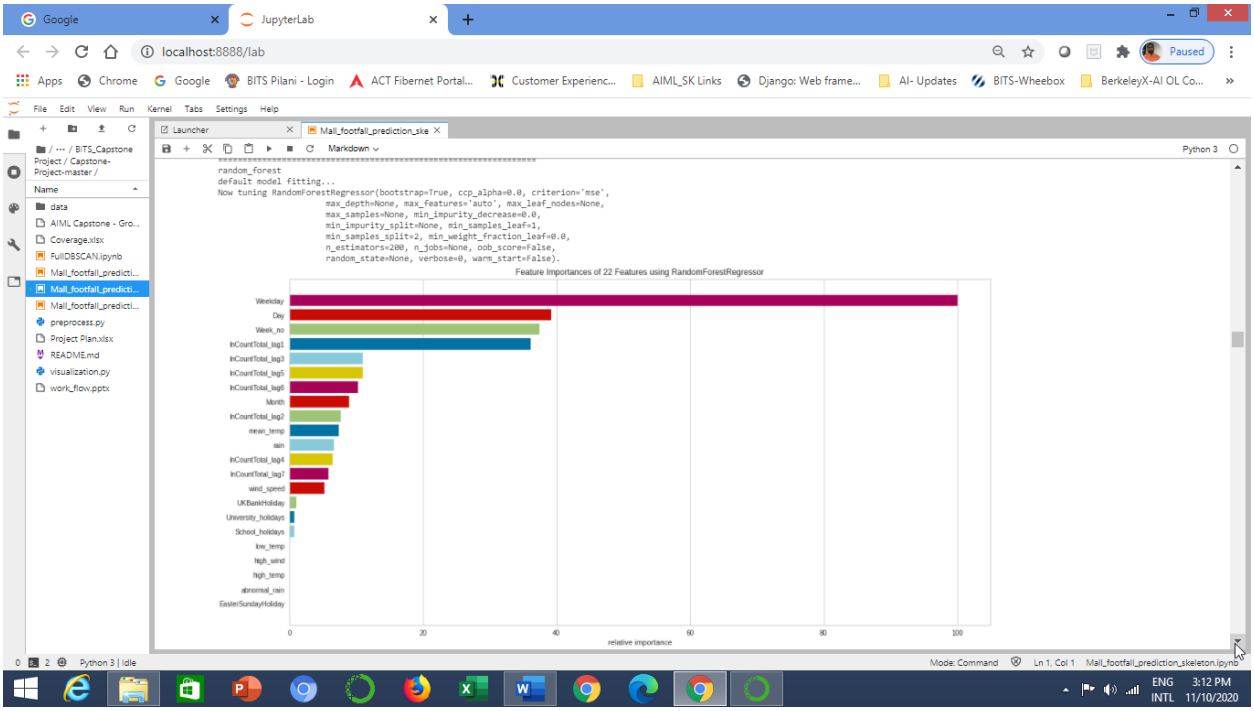
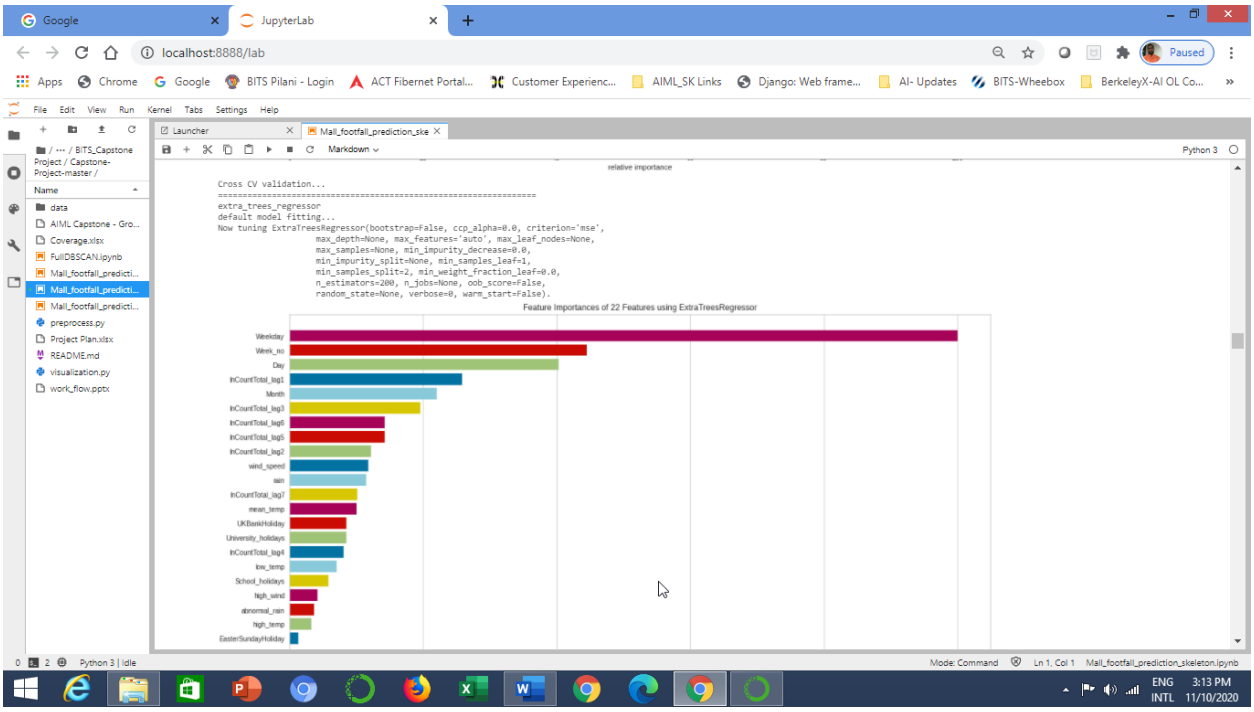
Code SNIPPETS_PART-2

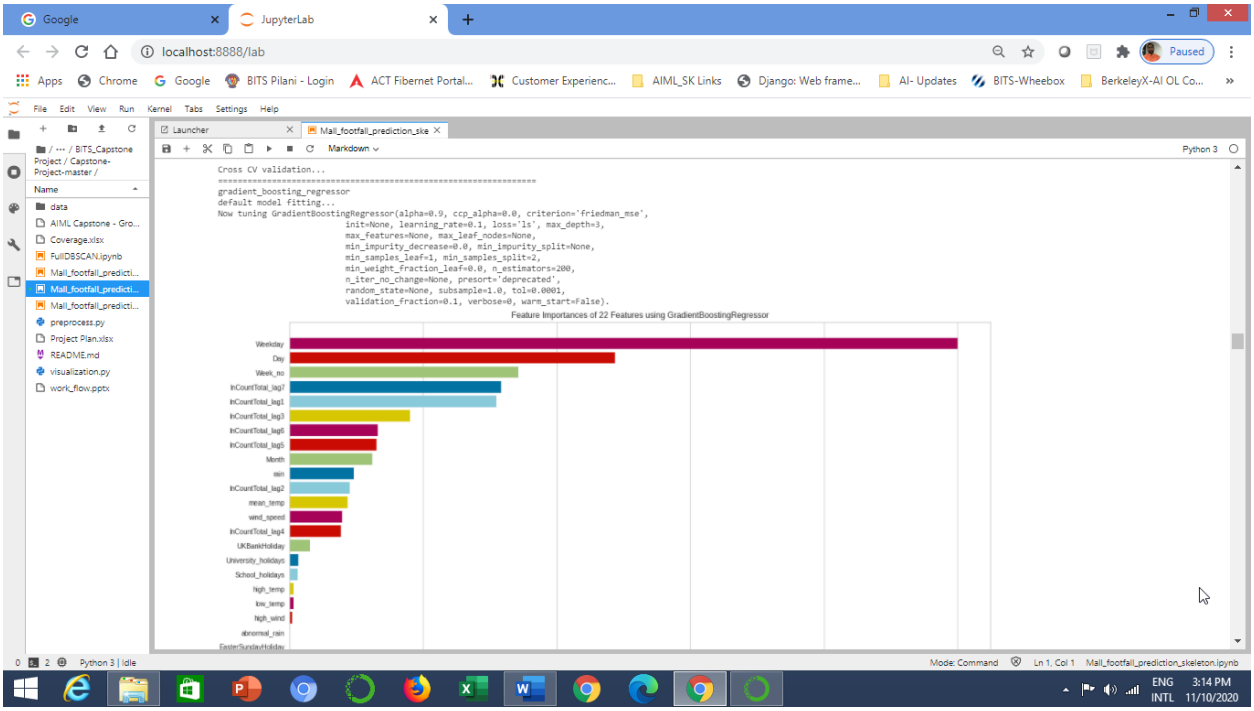


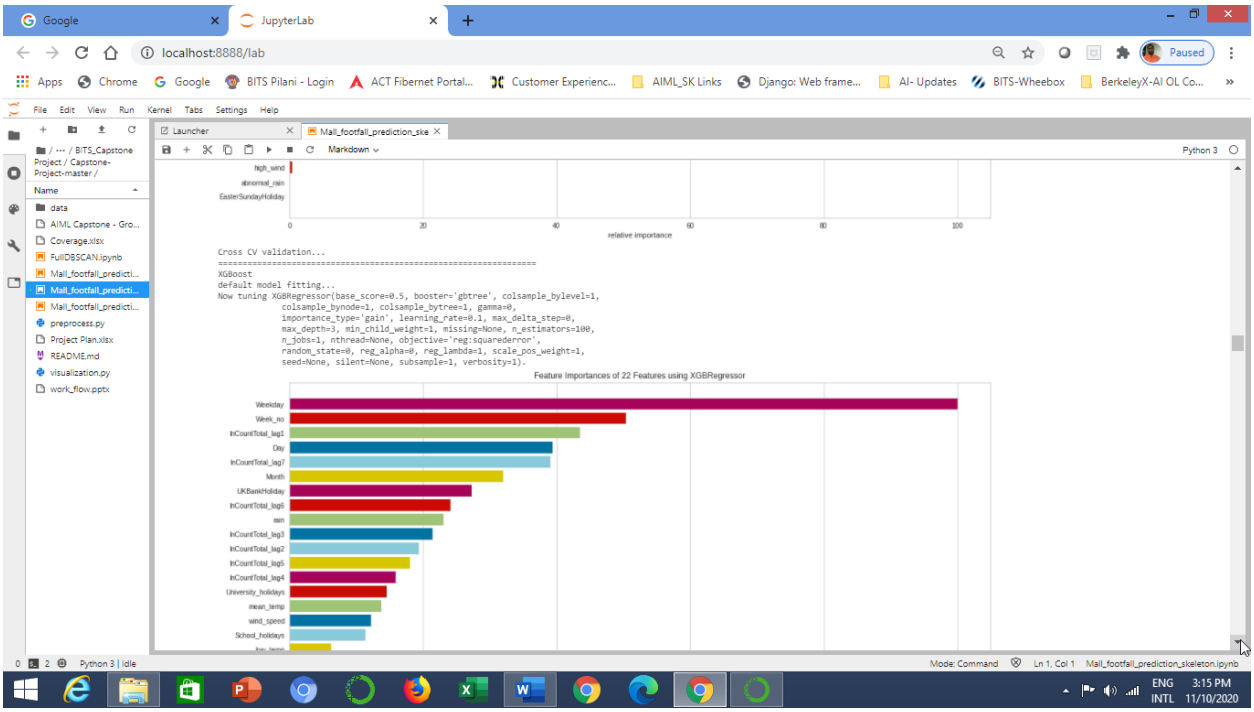

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Cross CV validation...
==========
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag6
Month
InCountTotal_lag2
mean_temp
rain
InCountTotal_lag4
InCountTotal_lag7
wind_speed
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
high_temp
abnormal_rain
EasterSundayHoliday
0
20
40
60
80
100
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:10 PM
11/10/2020

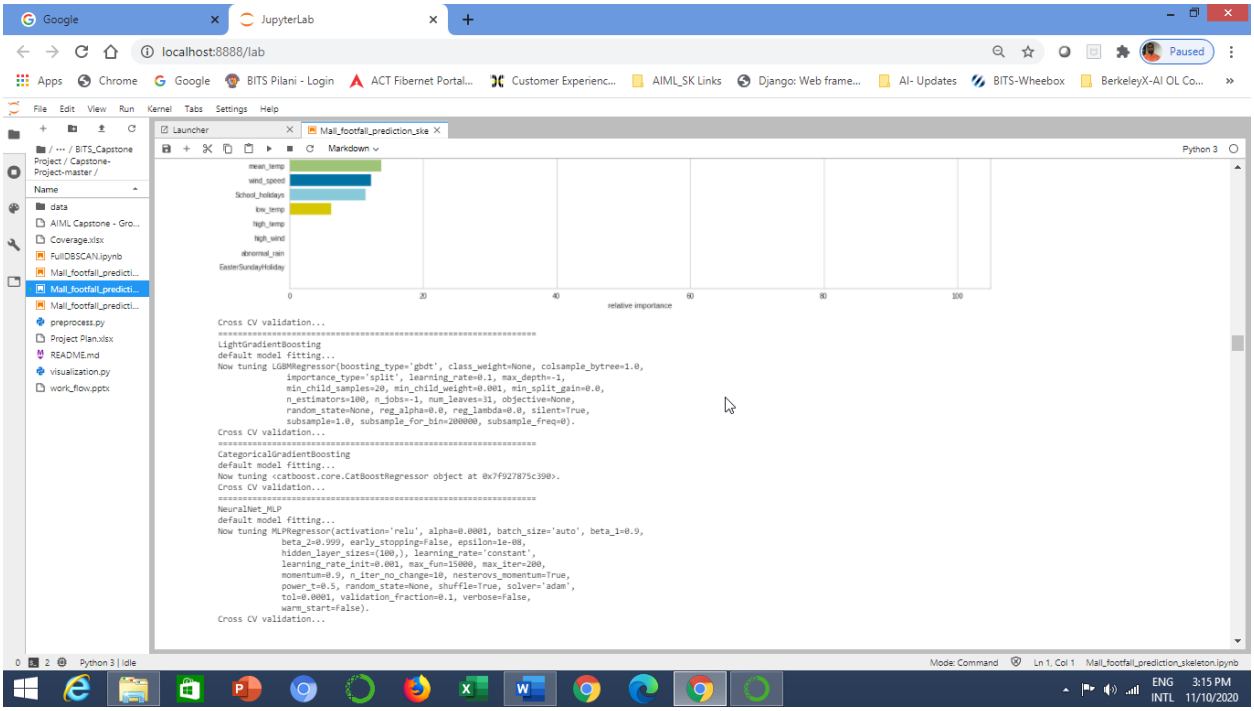

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Feature Importances of 22 Features using AdaBoostRegressor
Weekday
Day
Week_no
InCountTotal_lag5
InCountTotal_lag1
InCountTotal_lag6
InCountTotal_lag3
UKBankHoliday
InCountTotal_lag4
rain
mean_temp
InCountTotal_lag2
InCountTotal_lag7
Month
wind_speed
EasterSundayHoliday
University_holidays
School_holidays
low_temp
high_wind
abnormal_rain
high_temp
0
20
40
60
80
100
relative importance
Cross CV validation...
==========
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:09 PM
11/10/2020

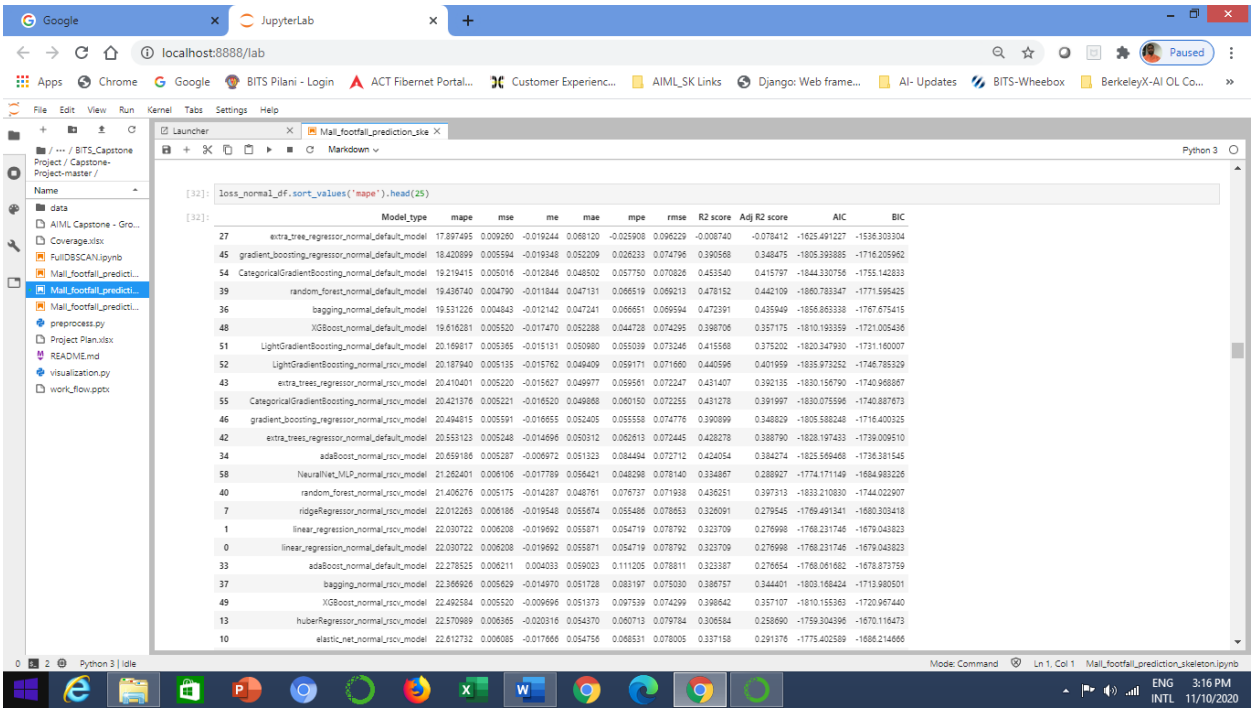

random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Weekday
Day
Week_no
lnCountTotal_lag1
lnCountTotal_lag3
lnCountTotal_lag5
lnCountTotal_lag6
Month
lnCountTotal_lag2
mean_temp
rain
lnCountTotal_lag4
lnCountTotal_lag7
wind_speed
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
high_temp
abnormal_rain
EasterSundayHoliday
0
20
40
60
80
100
relative importance

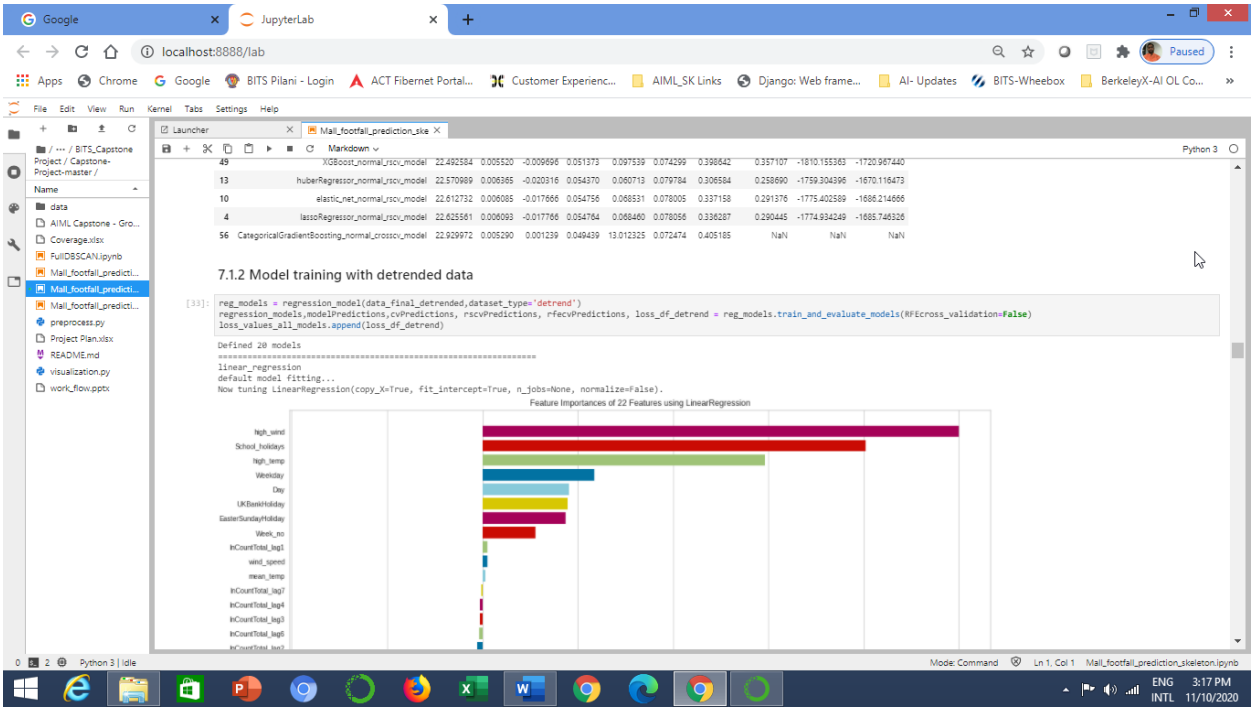

relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using ExtraTreesRegressor
Weekday
Week_no
Day
lnCountTotal_lag1
Month
lnCountTotal_lag3
lnCountTotal_lag6
lnCountTotal_lag5
lnCountTotal_lag2
wind_speed
rain
lnCountTotal_lag7
mean_temp
UKBankHoliday
University_holidays
lnCountTotal_lag4
low_temp
School_holidays
high_wind
abnormal_rain
high_temp
EasterSundayHoliday

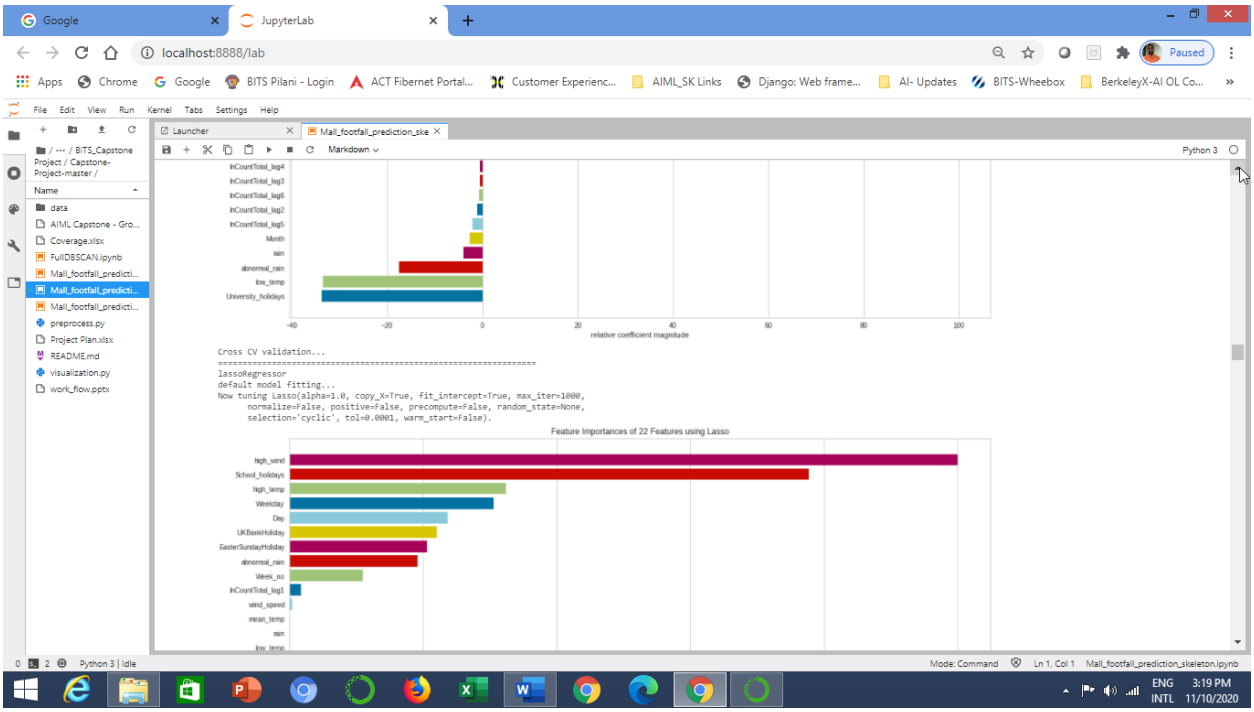

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 22 Features using GradientBoostingRegressor
Weekday
Day
Week_no
InCountTotal_lag7
InCountTotal_lag1
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag5
Month
min
InCountTotal_lag2
mean_temp
wind_speed
InCountTotal_lag4
UKBankHoliday
University_holidays
School_holidays
high_temp
low_temp
high_wind
abnormal_rain
EasterSundayHoliday
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:14 PM
11/10/2020

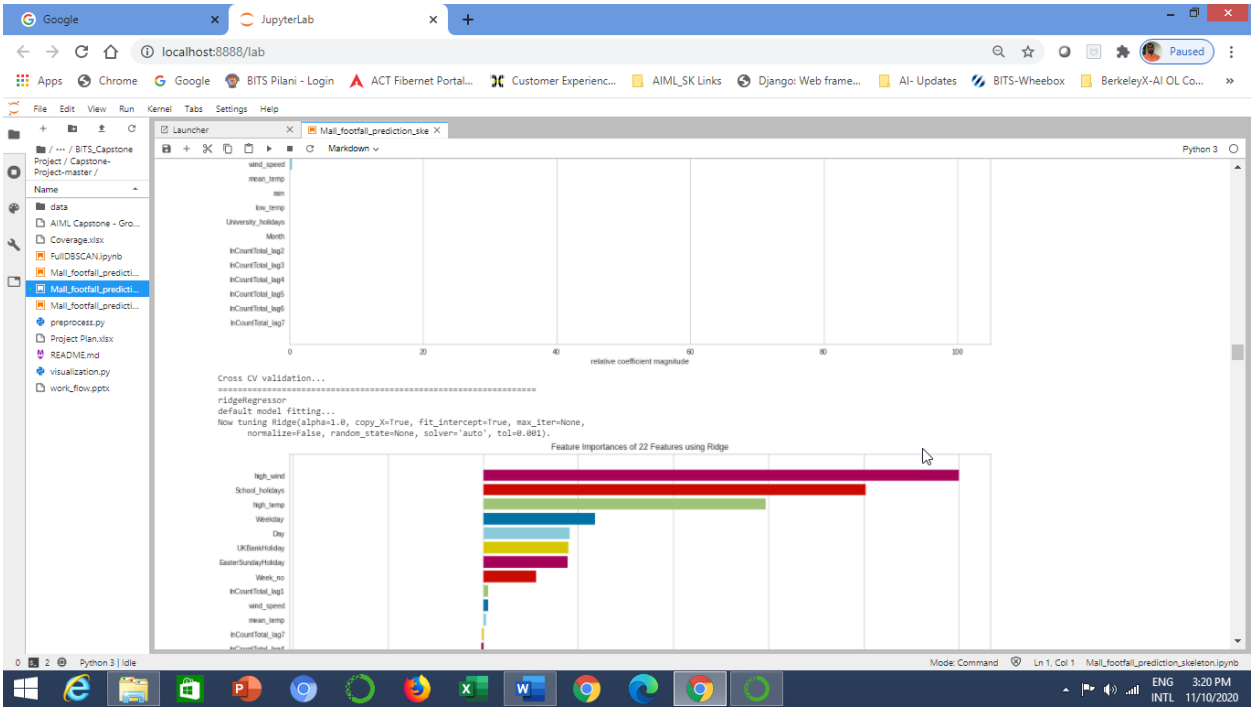

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
high_wind
abnormal_rain
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 22 Features using XGBRegressor
Weekday
Week_no
InCountTotal_lag1
Day
InCountTotal_lag7
Month
UKBankHoliday
InCountTotal_lag6
min
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag5
InCountTotal_lag4
University_holidays
mean_temp
wind_speed
School_holidays
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:15 PM
11/10/2020

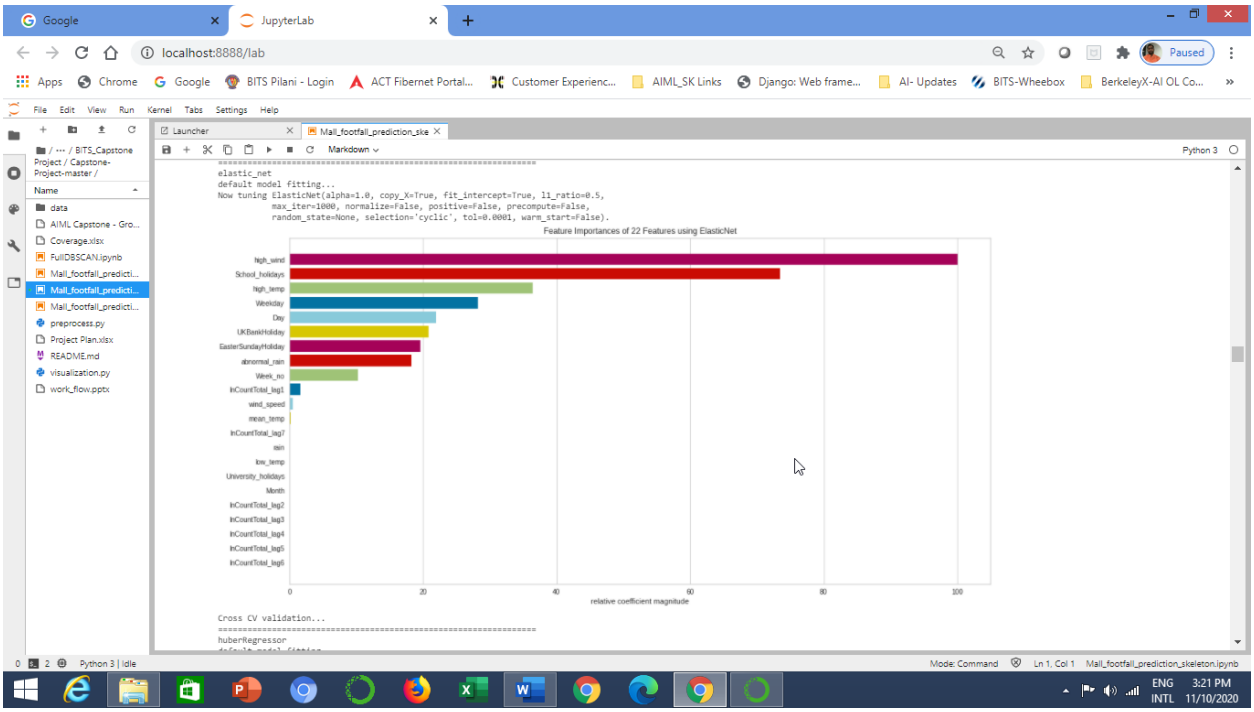

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
mean_temp
wind_speed
School_holidays
low_temp
high_temp
high_wind
abnormal_rain
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f927875c390>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
learning_rate_init=0.001, max_fun=15000, max_iter=200,
momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
power_t=0.5, random_state=None, shuffle=True, solver='adam',
tol=0.0001, validation_fraction=0.1, verbose=False,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:15 PM
11/10/2020

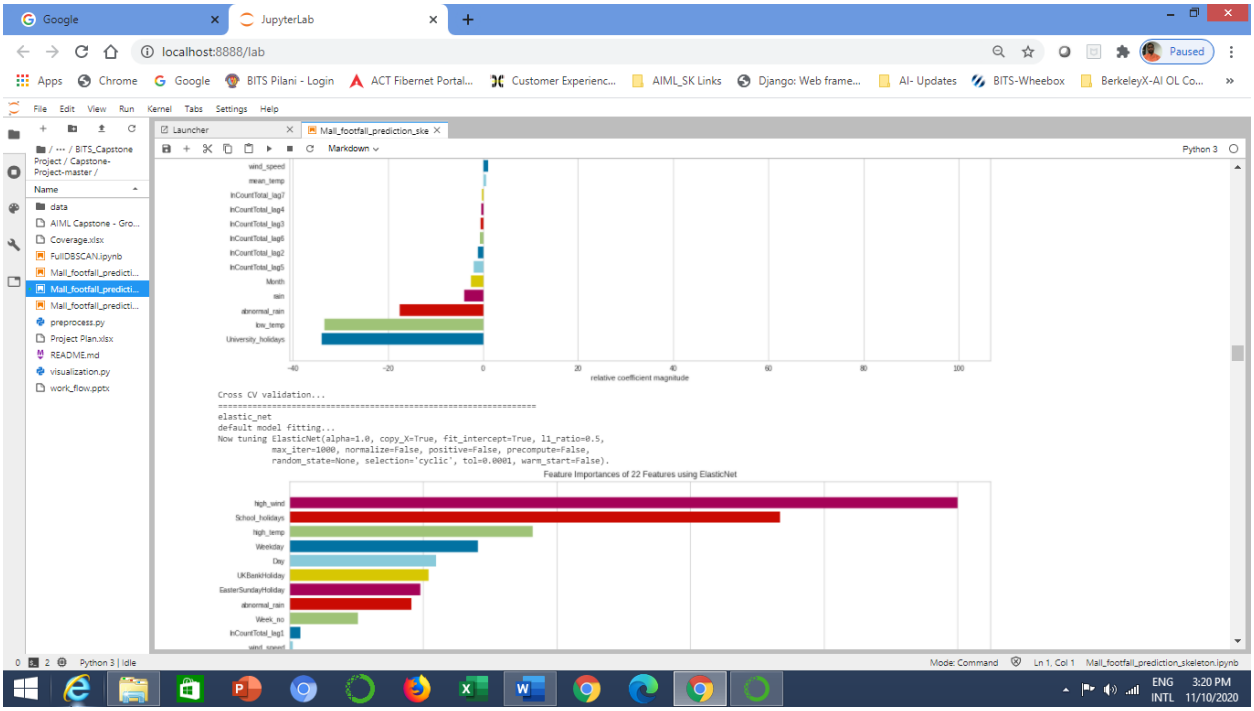

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
[32]: loss_normal_df.sort_values('mape').head(25)
[32]:
Model_type mape mse me mae mpe rmse R2 score Adj R2 score AIC BIC
27 extra_tree_regressor_normal_default_model 17.897495 0.009260 -0.019244 0.068120 -0.025908 0.096229 -0.008740 -0.078412 -1625.491227 -1536.303304
45 gradient_boosting_regressor_normal_default_model 18.420899 0.005594 -0.019348 0.052209 0.026233 0.074796 0.390568 0.348475 -1805.393885 -1716.205962
54 CategoricalGradientBoosting_normal_default_model 19.219415 0.005016 -0.012846 0.048502 0.057750 0.070826 0.453340 0.415797 -1844.330756 -1755.142833
39 random_forest_normal_default_model 19.436740 0.004790 -0.011844 0.047131 0.066519 0.069213 0.478152 0.442109 -1860.783347 -1771.595425
36 bagging_normal_default_model 19.531226 0.004843 -0.012142 0.047241 0.066651 0.069594 0.472391 0.435949 -1856.863338 -1767.675415
48 XGBoost_normal_default_model 19.616281 0.005520 -0.017470 0.052288 0.044728 0.074295 0.398706 0.357175 -1810.193359 -1721.005436
51 LightGradientBoosting_normal_default_model 20.169817 0.005365 -0.015131 0.050980 0.055039 0.073246 0.415568 0.375202 -1820.347930 -1731.160007
52 LightGradientBoosting_normal_rscv_model 20.187940 0.005135 -0.015762 0.049409 0.059171 0.071660 0.440596 0.401959 -1835.973252 -1746.785329
43 extra_trees_regressor_normal_rscv_model 20.410401 0.005220 -0.015627 0.049977 0.059561 0.072247 0.431407 0.392135 -1830.156790 -1740.968867
55 CategoricalGradientBoosting_normal_rscv_model 20.421376 0.005221 -0.016520 0.049868 0.060150 0.072255 0.431278 0.391997 -1830.075596 -1740.887673
46 gradient_boosting_regressor_normal_rscv_model 20.494815 0.005591 -0.016655 0.052405 0.055558 0.074776 0.390899 0.348829 -1805.588248 -1716.400325
42 extra_trees_regressor_normal_default_model 20.553123 0.005248 -0.014696 0.050312 0.062613 0.072445 0.428278 0.388790 -1828.197433 -1739.009510
34 adaBoost_normal_rscv_model 20.659186 0.005287 -0.006972 0.051323 0.084494 0.072712 0.424054 0.384274 -1825.569468 -1736.381545
58 NeuralNet_MLP_normal_rscv_model 21.262401 0.006106 -0.017789 0.056421 0.048298 0.078140 0.334867 0.288927 -1774.171149 -1684.983226
40 random_forest_normal_rscv_model 21.406276 0.005175 -0.014287 0.048761 0.076737 0.071938 0.436251 0.397313 -1833.210830 -1744.022907
7 ridgeRegressor_normal_rscv_model 22.012263 0.006186 -0.019548 0.055674 0.055486 0.078653 0.326091 0.279545 -1769.491341 -1680.303418
1 linear_regression_normal_rscv_model 22.030722 0.006208 -0.019692 0.055871 0.054719 0.078792 0.323709 0.276998 -1768.231746 -1679.043823
0 linear_regression_normal_default_model 22.030722 0.006208 -0.019692 0.055871 0.054719 0.078792 0.323709 0.276998 -1768.231746 -1679.043823
33 adaBoost_normal_default_model 22.278525 0.006211 0.004033 0.059023 0.111205 0.078811 0.323387 0.276654 -1768.061682 -1678.873759
37 bagging_normal_rscv_model 22.366926 0.005629 -0.014970 0.051728 0.083197 0.075030 0.386757 0.344401 -1803.168424 -1713.980501
49 XGBoost_normal_rscv_model 22.492584 0.005520 -0.009696 0.051373 0.097539 0.074299 0.398642 0.357107 -1810.155363 -1720.967440
13 huberRegressor_normal_rscv_model 22.570989 0.006365 -0.020316 0.054370 0.060713 0.079784 0.306584 0.258690 -1759.304396 -1670.116473
10 elastic_net_normal_rscv_model 22.612732 0.006085 -0.017666 0.054756 0.068531 0.078005 0.337158 0.291376 -1775.402589 -1686.214666
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:16 PM
11/10/2020

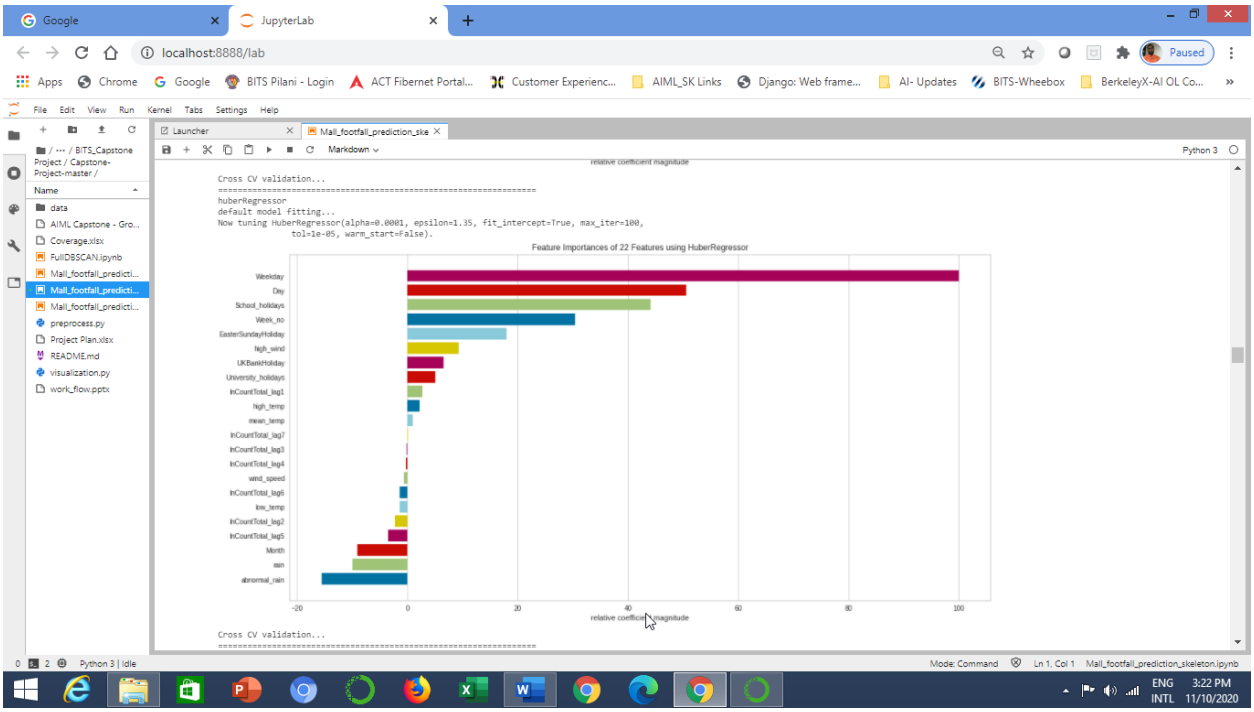

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
7.1.2 Model training with detrended data
[33]:
reg_models = regression_model(data_final_detrended,dataset_type='detrend')
regression_models,modelPredictions,cvPredictions, rscvPredictions, rfecvPredictions, loss_df_detrend = reg_models.train_and_evaluate_models(RFEcross_validation=False)
loss_values_all_models.append(loss_df_detrend)
Defined 20 models
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
Feature Importances of 22 Features using LinearRegression
high_wind
School_holidays
high_temp
Weekday
Day
UKBankHoliday
EasterSundayHoliday
Week_no
InCountTotal_lag1
wind_speed
mean_temp
InCountTotal_lag7
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag6
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:17 PM
11/10/2020

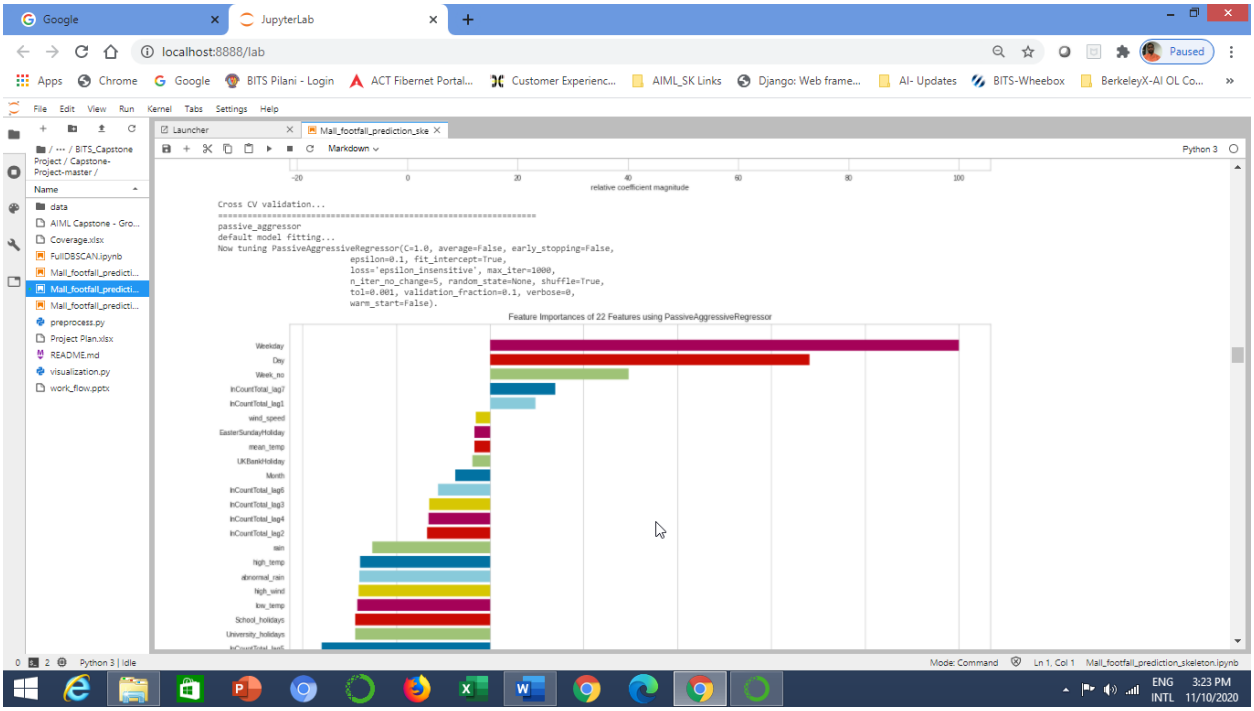

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag2
InCountTotal_lag5
Month
rain
abnormal_rain
low_temp
University_holidays
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using Lasso
high_wind
School_holidays
high_temp
Weekday
Day
UKBankHoliday
EasterSundayHoliday
abnormal_rain
Week_no
InCountTotal_lag1
wind_speed
mean_temp
rain
low_temp
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:19 PM
11/10/2020

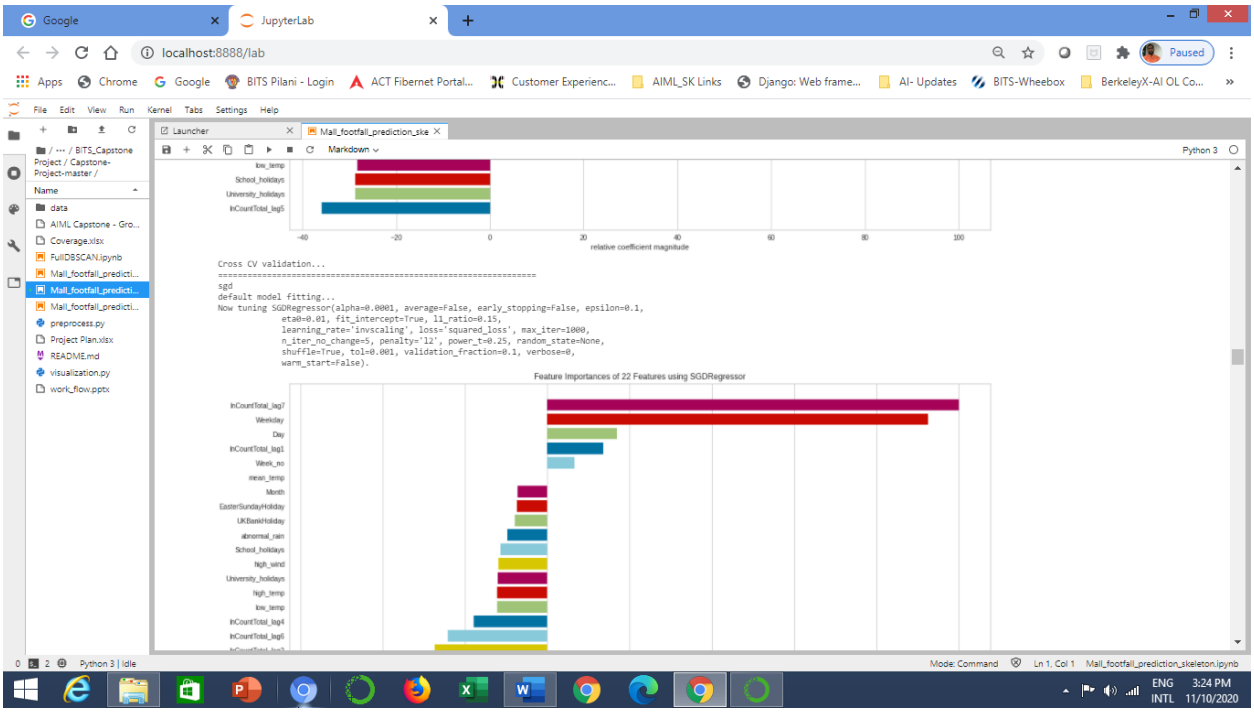

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
wind_speed
mean_temp
rain
low_temp
University_holidays
Month
InCountTotal_lag2
InCountTotal_lag3
InCountTotal_lag4
InCountTotal_lag5
InCountTotal_lag6
InCountTotal_lag7
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 22 Features using Ridge
high_wind
School_holidays
high_temp
Weekday
Day
UKBankHoliday
EasterSundayHoliday
Week_no
InCountTotal_lag1
wind_speed
mean_temp
InCountTotal_lag7
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:20 PM
11/10/2020

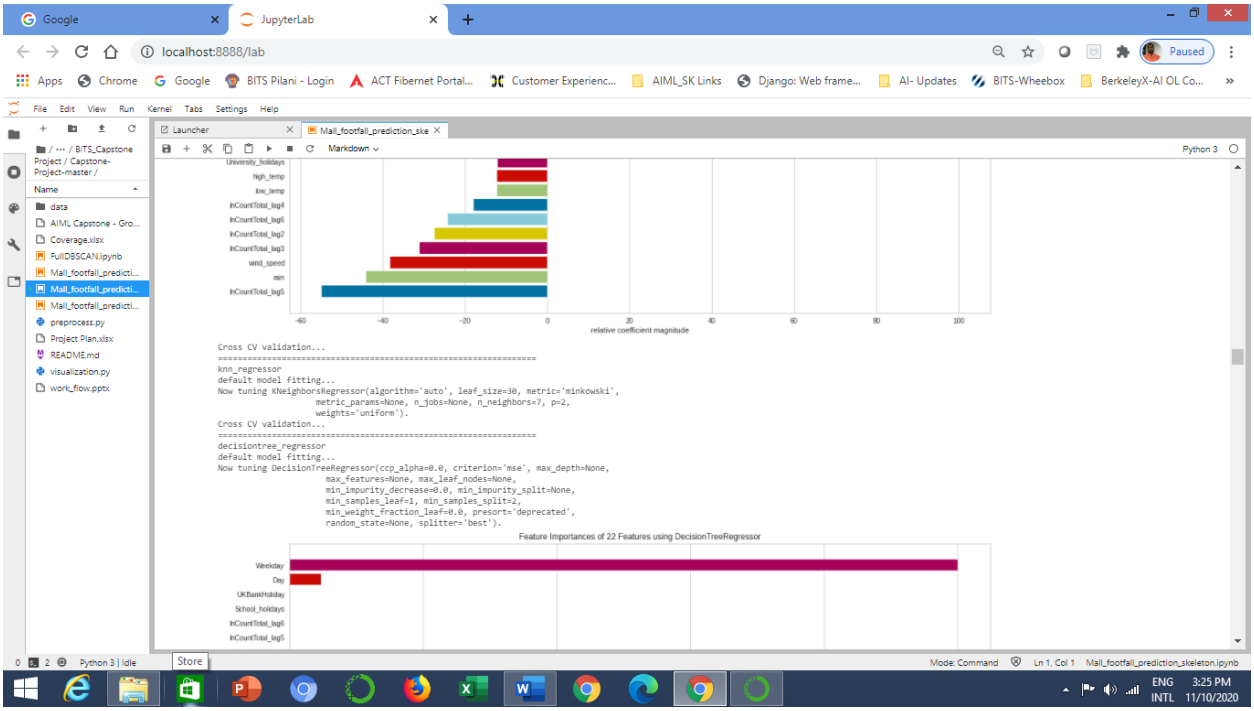

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using ElasticNet
high_wind
School_holidays
high_temp
Weekday
Day
UKBankHoliday
EasterSundayHoliday
abnormal_rain
Week_no
InCountTotal_lag1
wind_speed
mean_temp
InCountTotal_lag7
rain
low_temp
University_holidays
Month
InCountTotal_lag2
InCountTotal_lag3
InCountTotal_lag4
InCountTotal_lag5
InCountTotal_lag6
relative coefficient magnitude
Cross CV validation...
huberRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:21 PM
11/10/2020

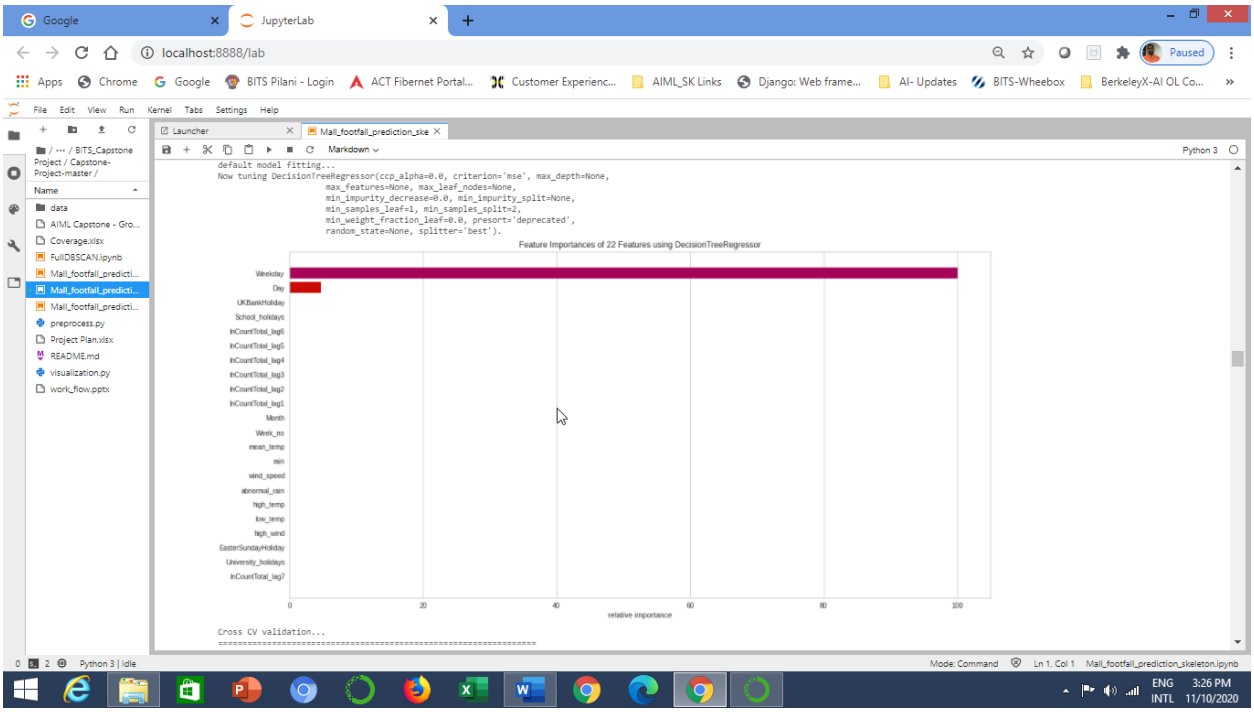

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
wind_speed
mean_temp
InCountTotal_lag7
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag2
InCountTotal_lag5
Month
rain
abnormal_rain
low_temp
University_holidays
-40
-20
0
20
40
60
80
100
relative coefficient magnitude
Cross CV validation...
elastic_net
deFault model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using ElasticNet
high_wind
School_holidays
high_temp
Weekday
Day
UKBankHoliday
EasterSundayHoliday
abnormal_rain
Week_no
InCountTotal_lag1
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:20 PM
11/10/2020

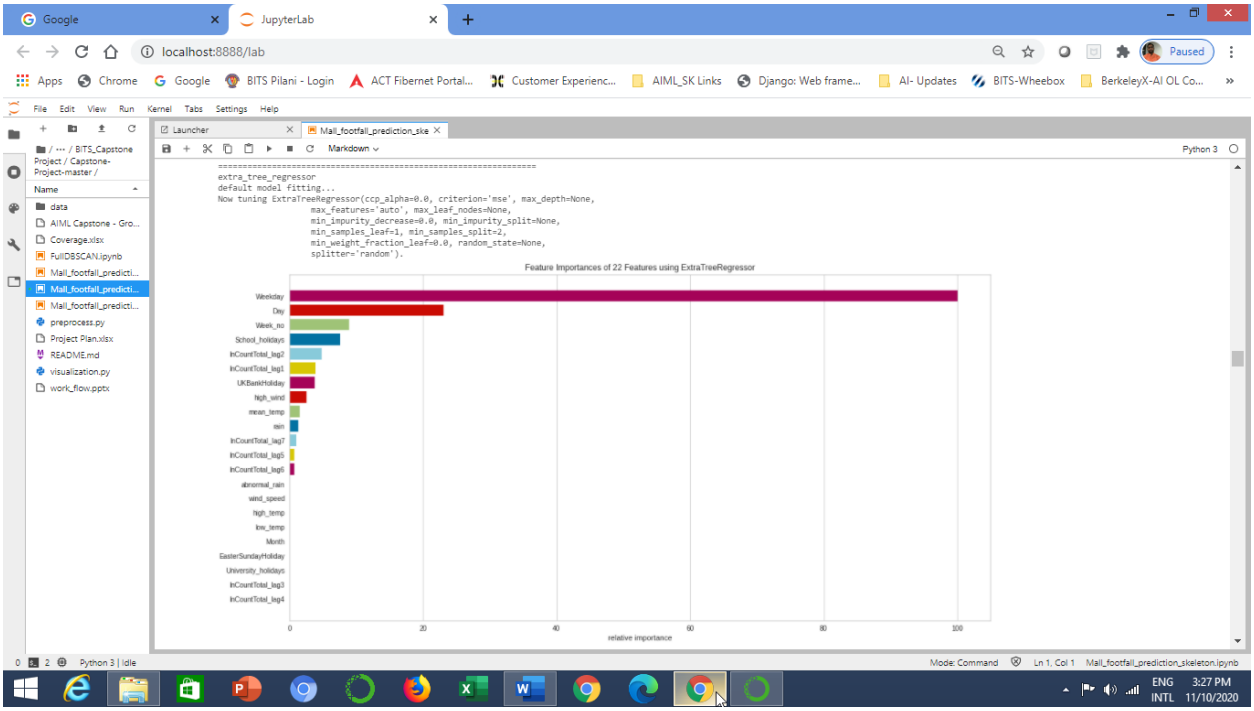

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
huberRegressor
deFault model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 22 Features using HuberRegressor
Weekday
Day
School_holidays
Week_no
EasterSundayHoliday
high_wind
UKBankHoliday
University_holidays
InCountTotal_lag1
high_temp
mean_temp
InCountTotal_lag7
InCountTotal_lag3
InCountTotal_lag4
wind_speed
InCountTotal_lag6
low_temp
InCountTotal_lag2
InCountTotal_lag5
Month
rain
abnormal_rain
-20
0
20
40
60
80
100
relative coefficient magnitude
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:22 PM
11/10/2020

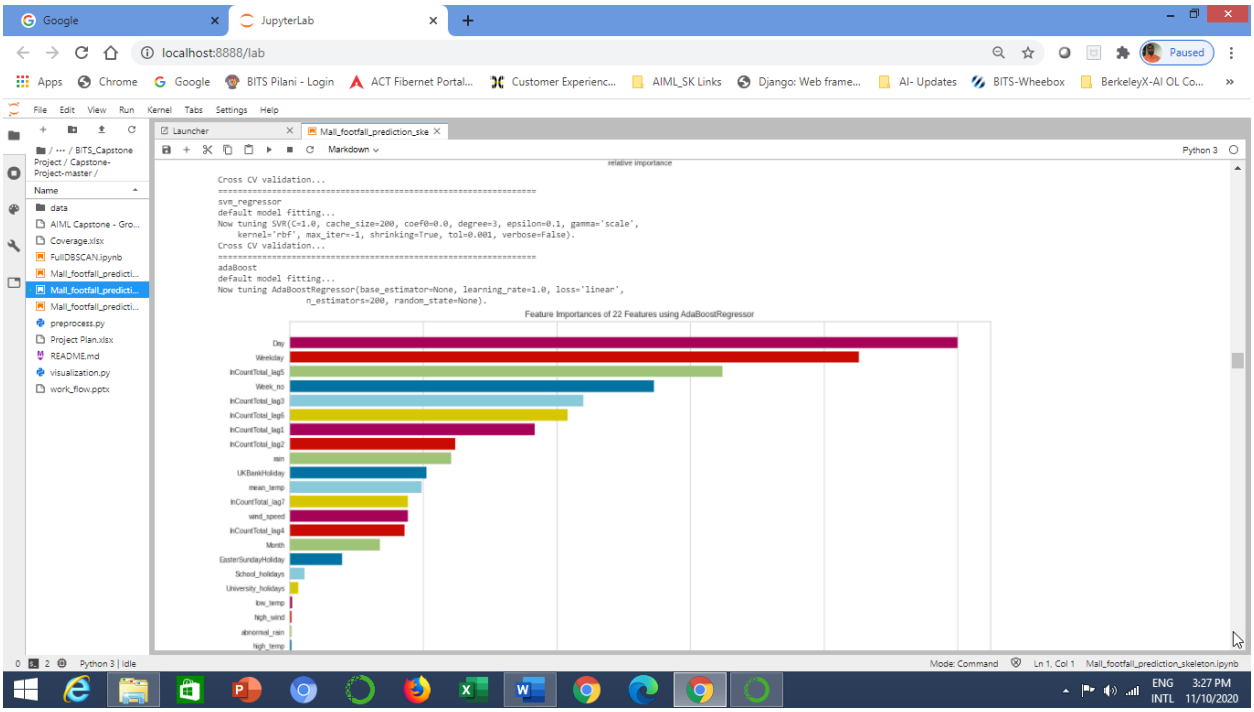

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using PassiveAggressiveRegressor
Weekday
Day
Week_no
InCountTotal_lag7
InCountTotal_lag1
wind_speed
EasterSundayHoliday
mean_temp
UKBankHoliday
Month
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag4
InCountTotal_lag2
rain
high_temp
abnormal_rain
high_wind
low_temp
School_holidays
University_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:23 PM
11/10/2020

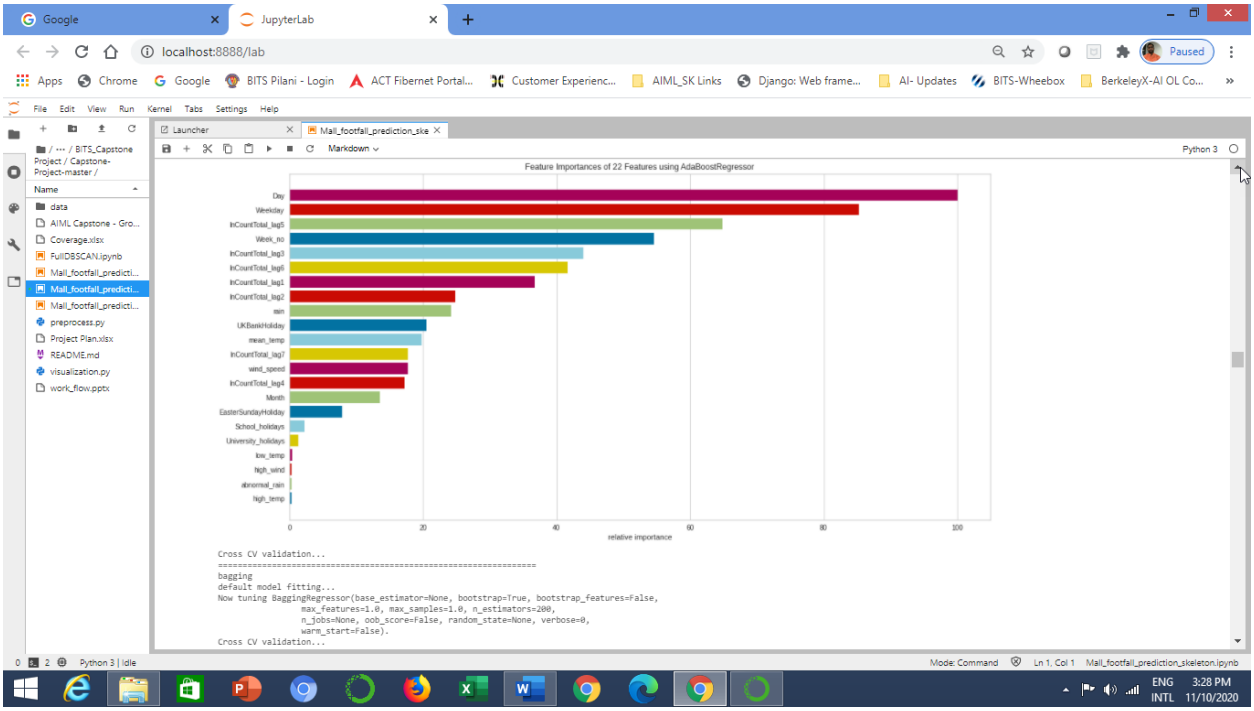

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
low_temp
School_holidays
University_holidays
InCountTotal_lag5
relative coefficient magnitude
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using SGDRegressor
InCountTotal_lag7
Weekday
Day
InCountTotal_lag1
Week_no
mean_temp
Month
EasterSundayHoliday
UKBankHoliday
abnormal_rain
School_holidays
high_wind
University_holidays
high_temp
low_temp
InCountTotal_lag4
InCountTotal_lag6
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:24 PM
11/10/2020

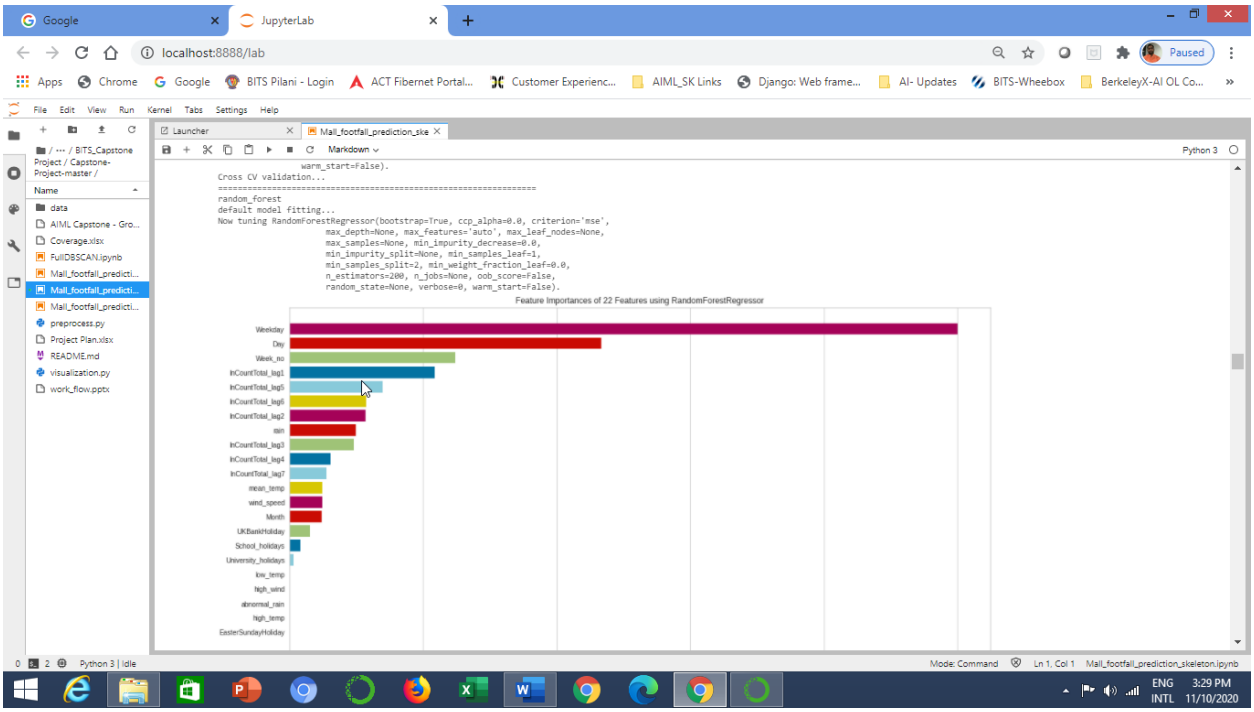

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
University_holidays
high_temp
low_temp
InCountTotal_lag4
InCountTotal_lag6
InCountTotal_lag2
InCountTotal_lag3
wind_speed
rain
InCountTotal_lag5
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 22 Features using DecisionTreeRegressor
Weekday
Day
UKBankHoliday
School_holidays
InCountTotal_lag6
InCountTotal_lag5
Store
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:25 PM
11/10/2020

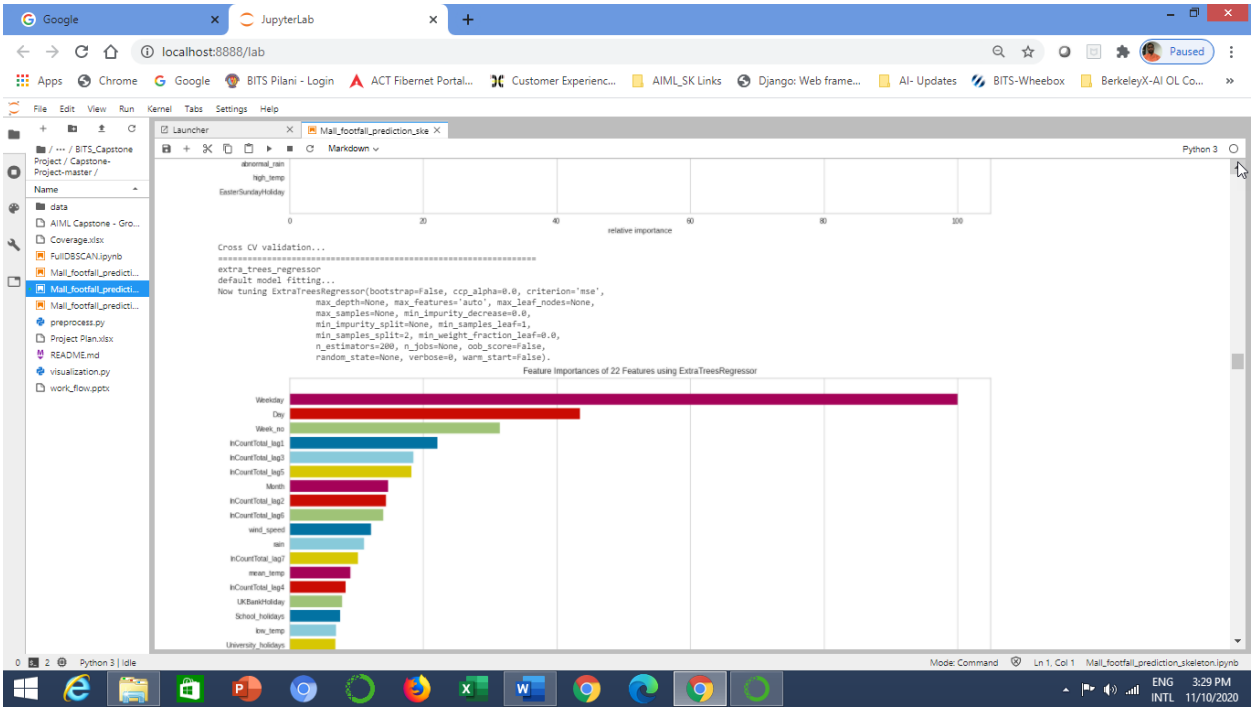

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 22 Features using DecisionTreeRegressor
Weekday
Day
UKBankHoliday
School_holidays
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag1
Month
Week_no
mean_temp
rain
wind_speed
abnormal_rain
high_temp
low_temp
high_wind
EasterSundayHoliday
University_holidays
InCountTotal_lag7
relative importance
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:26 PM
11/10/2020

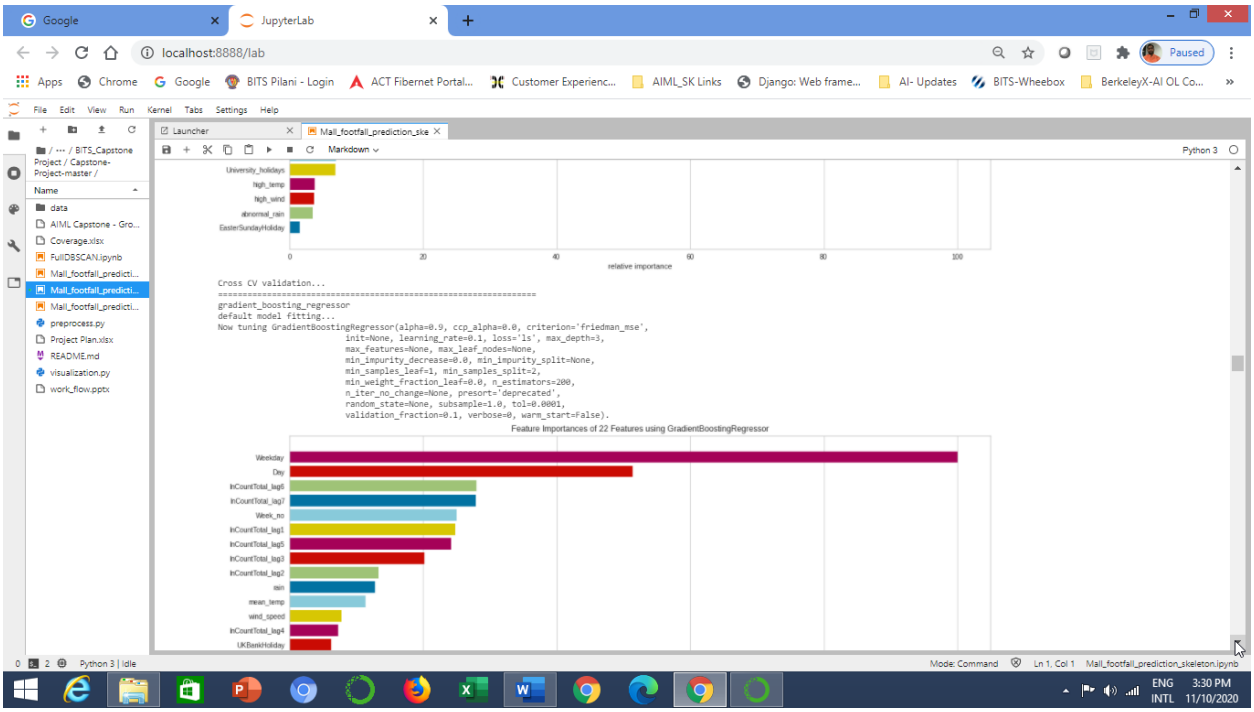

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').
Feature Importances of 22 Features using ExtraTreeRegressor
Weekday
Day
Week_no
School_holidays
InCountTotal_lag2
InCountTotal_lag1
UKBankHoliday
high_wind
mean_temp
rain
InCountTotal_lag7
InCountTotal_lag5
InCountTotal_lag6
abnormal_rain
wind_speed
high_temp
low_temp
Month
EasterSundayHoliday
University_holidays
InCountTotal_lag3
InCountTotal_lag4
0
20
40
60
80
100
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:27 PM
11/10/2020

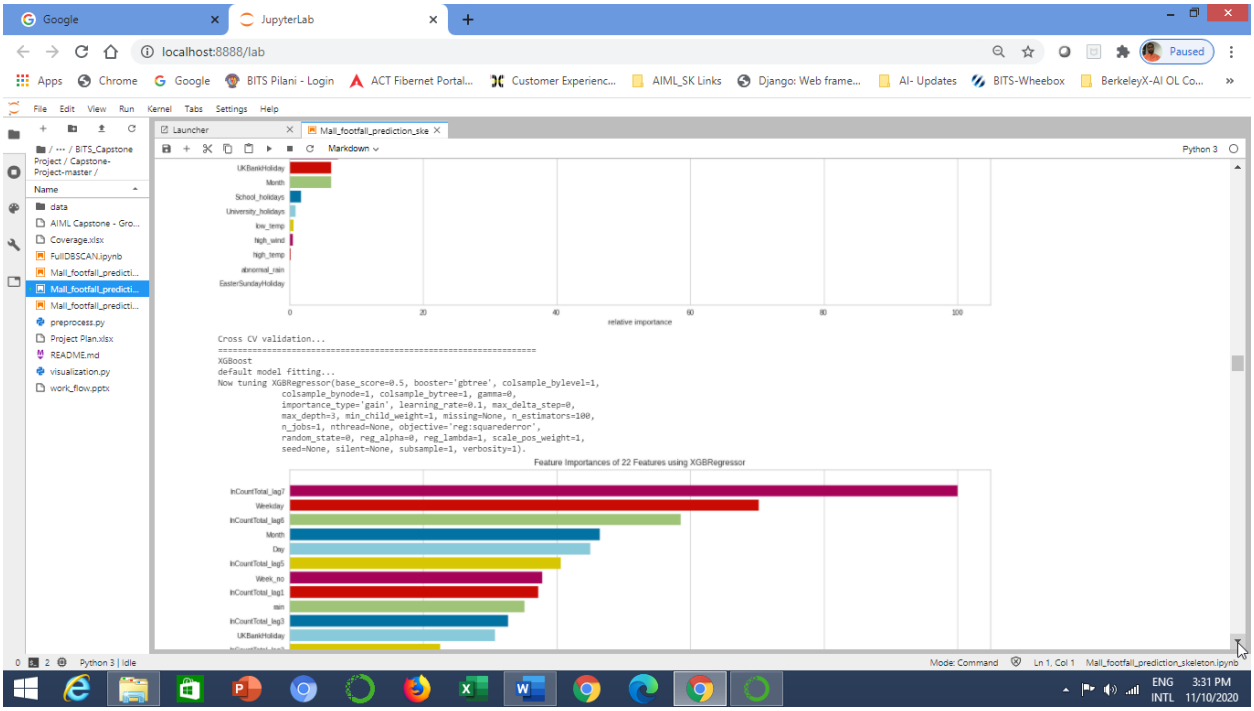

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative importance
Cross CV validation...
svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 22 Features using AdaBoostRegressor
Day
Weekday
InCountTotal_lag5
Week_no
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag1
InCountTotal_lag2
rain
UKBankHoliday
mean_temp
InCountTotal_lag7
wind_speed
InCountTotal_lag4
Month
EasterSundayHoliday
School_holidays
University_holidays
low_temp
high_wind
abnormal_rain
high_temp
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:27 PM
11/10/2020

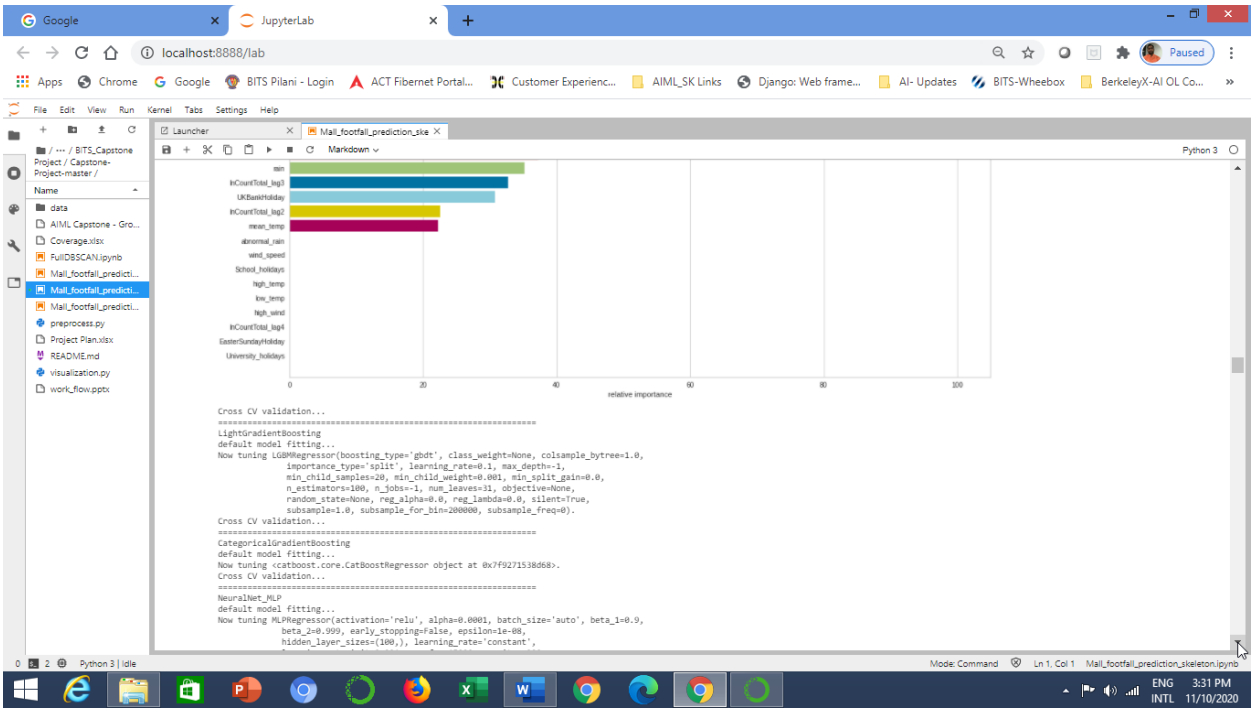

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Feature Importances of 22 Features using AdaBoostRegressor
Day
Weekday
lnCountTotal_lag5
Week_no
lnCountTotal_lag3
lnCountTotal_lag6
lnCountTotal_lag1
lnCountTotal_lag2
min
UKBankHoliday
mean_temp
lnCountTotal_lag7
wind_speed
lnCountTotal_lag4
Month
EasterSundayHoliday
School_holidays
University_holidays
low_temp
high_wind
abnormal_rain
high_temp
0
20
40
60
80
100
relative importance
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:28 PM
11/10/2020

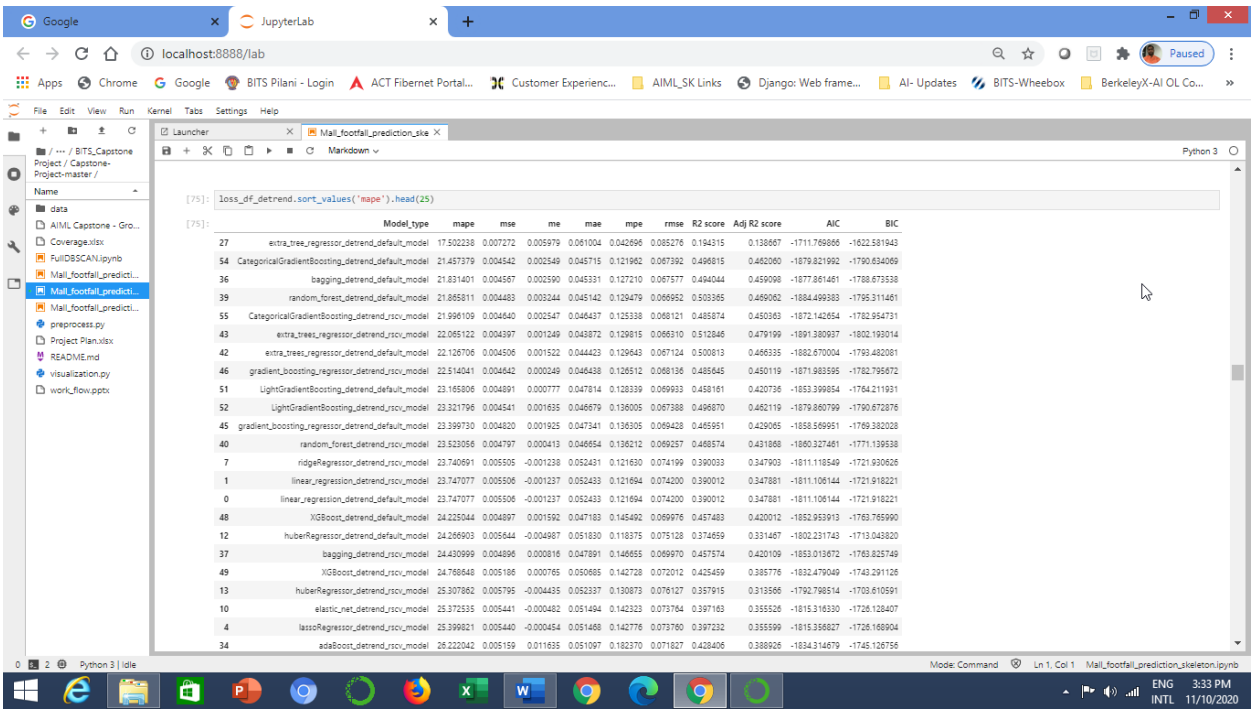

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Weekday
Day
Week_no
lnCountTotal_lag1
lnCountTotal_lag5
lnCountTotal_lag6
lnCountTotal_lag2
min
lnCountTotal_lag3
lnCountTotal_lag4
lnCountTotal_lag7
mean_temp
wind_speed
Month
UKBankHoliday
School_holidays
University_holidays
low_temp
high_wind
abnormal_rain
high_temp
EasterSundayHoliday
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:29 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
abnormal_rain
high_temp
EasterSundayHoliday
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using ExtraTreesRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag3
InCountTotal_lag5
Month
InCountTotal_lag2
InCountTotal_lag6
wind_speed
rain
InCountTotal_lag7
mean_temp
InCountTotal_lag4
UKBankHoliday
School_holidays
low_temp
University_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:29 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
University_holidays
high_temp
high_wind
abnormal_rain
EasterSundayHoliday
relative importance
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 22 Features using GradientBoostingRegressor
Weekday
Day
InCountTotal_lag6
InCountTotal_lag7
Week_no
InCountTotal_lag1
InCountTotal_lag5
InCountTotal_lag3
InCountTotal_lag2
rain
mean_temp
wind_speed
InCountTotal_lag4
UKBankHoliday
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:30 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
UKBankHoliday
Month
School_holidays
University_holidays
low_temp
high_wind
high_temp
abnormal_rain
EasterSundayHoliday
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 22 Features using XGBRegressor
InCountTotal_lag7
Weekday
InCountTotal_lag6
Month
Day
InCountTotal_lag5
Week_no
InCountTotal_lag1
min
InCountTotal_lag3
UKBankHoliday
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:31 PM
11/10/2020

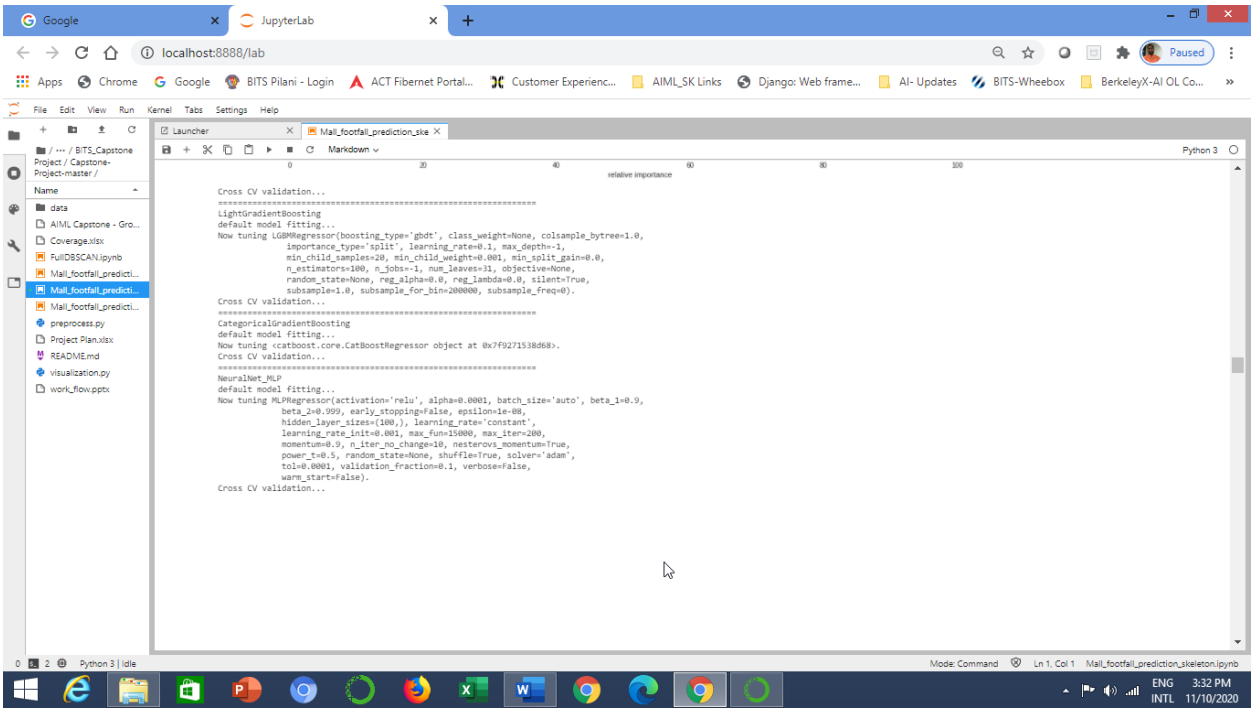

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
min
InCountTotal_lag3
UKBankHoliday
InCountTotal_lag2
mean_temp
abnormal_rain
wind_speed
School_holidays
high_temp
low_temp
high_wind
InCountTotal_lag4
EasterSundayHoliday
University_holidays
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f9271538d68>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:31 PM
11/10/2020



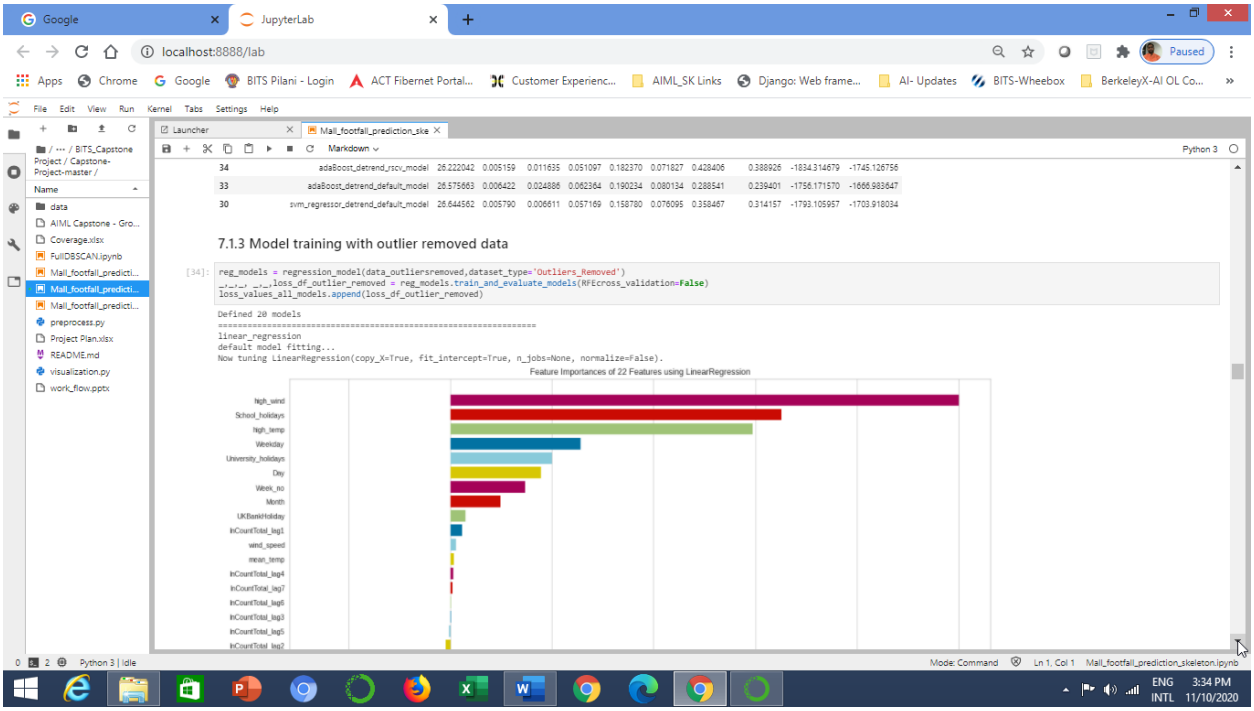
```
loss_df_detrend.sort_values('mape').head(25)
```

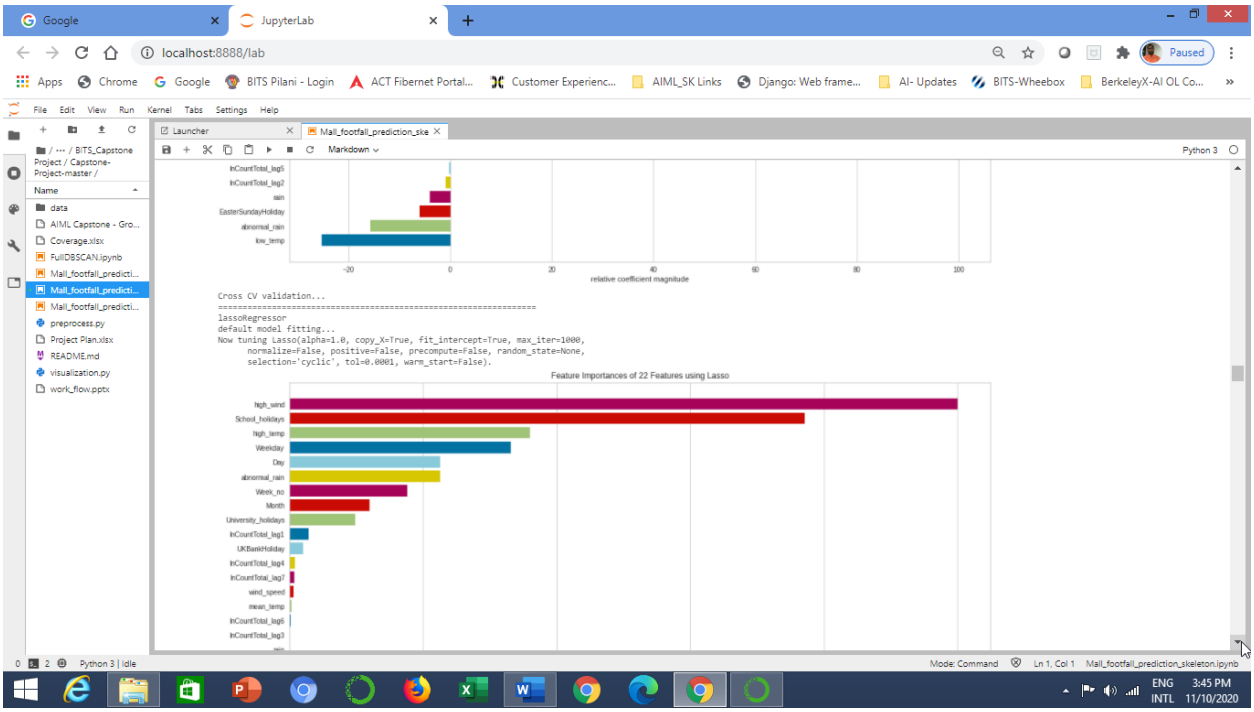
[75]:

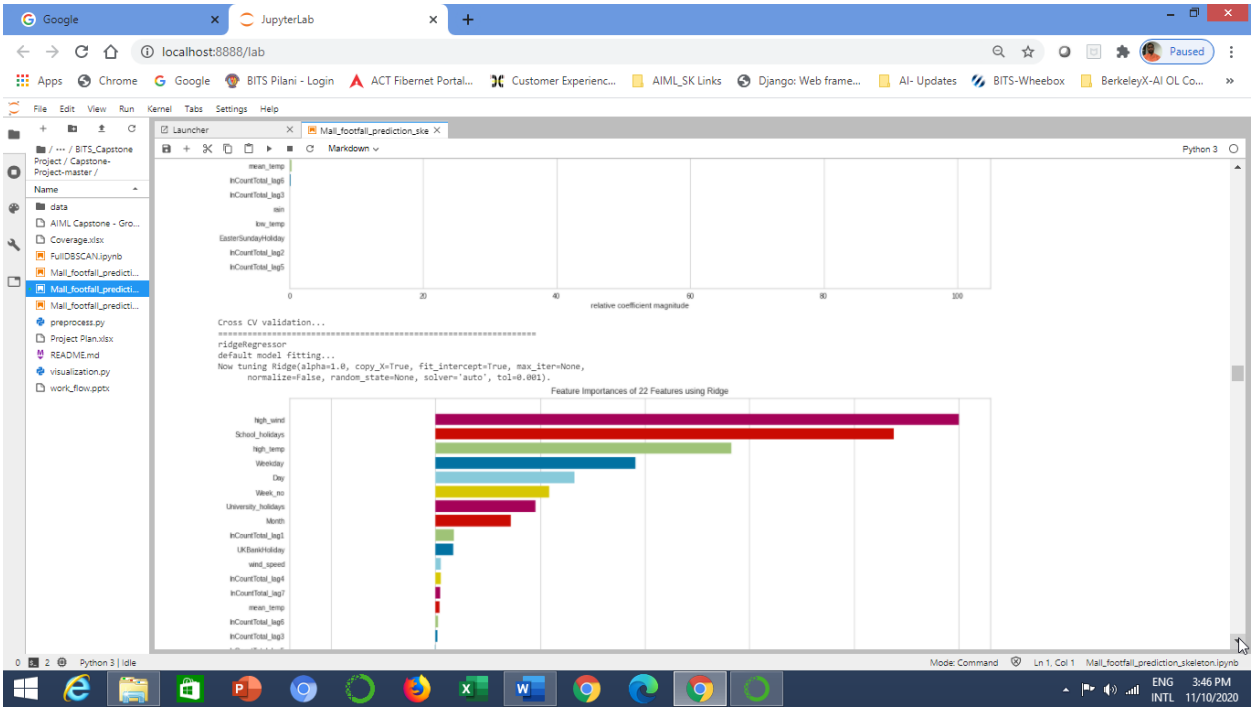
	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
27	extra_tree_regressor_detrend_default_model	17.502238	0.007272	0.005979	0.061004	0.042696	0.085276	0.194315	0.138667	-1711.769866	-1622.581943
54	CategoricalGradientBoosting_detrend_default_model	21.457379	0.004542	0.002549	0.045715	0.121962	0.067392	0.496815	0.462060	-1879.821992	-1790.634069
36	bagging_detrend_default_model	21.831401	0.004567	0.002590	0.045331	0.127210	0.067577	0.494044	0.459098	-1877.861461	-1788.673538
39	random_forest_detrend_default_model	21.865811	0.004483	0.003244	0.045142	0.129479	0.066952	0.503365	0.469062	-1884.499383	-1795.311461
55	CategoricalGradientBoosting_detrend_rscv_model	21.996109	0.004640	0.002547	0.046437	0.125338	0.068121	0.485874	0.450363	-1872.142654	-1782.954731
43	extra_trees_regressor_detrend_rscv_model	22.065122	0.004397	0.001249	0.043872	0.129815	0.066310	0.512846	0.479199	-1891.380937	-1802.193014
42	extra_trees_regressor_detrend_default_model	22.126706	0.004506	0.001522	0.044423	0.129643	0.067124	0.500813	0.466335	-1882.670004	-1793.482081
46	gradient_boosting_regressor_detrend_rscv_model	22.514041	0.004642	0.000249	0.046438	0.126512	0.068136	0.485645	0.450119	-1871.983595	-1782.795672
51	LightGradientBoosting_detrend_default_model	23.165806	0.004891	0.000777	0.047814	0.128339	0.069933	0.458161	0.420736	-1853.399854	-1764.211931
52	LightGradientBoosting_detrend_rscv_model	23.321796	0.004541	0.001635	0.046679	0.136005	0.067388	0.496870	0.462119	-1879.860799	-1790.672876
45	gradient_boosting_regressor_detrend_default_model	23.399730	0.004820	0.001925	0.047341	0.136305	0.069428	0.465951	0.429065	-1858.569951	-1769.382028
40	random_forest_detrend_rscv_model	23.523056	0.004797	0.000413	0.046654	0.136212	0.069257	0.468574	0.431868	-1860.327461	-1771.139538
7	ridgeRegressor_detrend_rscv_model	23.740691	0.005505	-0.001238	0.052431	0.121630	0.074199	0.390033	0.347903	-1811.118549	-1721.930626
1	linear_regression_detrend_rscv_model	23.747077	0.005506	-0.001237	0.052433	0.121694	0.074200	0.390012	0.347881	-1811.106144	-1721.918221
0	linear_regression_detrend_default_model	23.747077	0.005506	-0.001237	0.052433	0.121694	0.074200	0.390012	0.347881	-1811.106144	-1721.918221
48	XGBoost_detrend_default_model	24.225044	0.004897	0.001592	0.047183	0.145492	0.069976	0.457483	0.420012	-1852.953913	-1763.765990
12	huberRegressor_detrend_default_model	24.266903	0.005644	-0.004987	0.051830	0.118375	0.075128	0.374659	0.331467	-1802.231743	-1713.043820
37	bagging_detrend_rscv_model	24.430999	0.004896	0.000816	0.047891	0.146655	0.069970	0.457574	0.420109	-1853.013672	-1763.825749
49	XGBoost_detrend_rscv_model	24.768648	0.005186	0.000765	0.050685	0.142728	0.072012	0.425459	0.385776	-1832.479049	-1743.291126
13	huberRegressor_detrend_rscv_model	25.307862	0.005795	-0.004435	0.052337	0.130873	0.076127	0.357915	0.313566	-1792.798514	-1703.610591
10	elastic_net_detrend_rscv_model	25.372535	0.005441	-0.000482	0.051494	0.142323	0.073764	0.397163	0.355526	-1815.316330	-1726.128407
4	lassoRegressor_detrend_rscv_model	25.399821	0.005440	-0.000454	0.051468	0.142776	0.073760	0.397232	0.355599	-1815.356827	-1726.168904
34	adaBoost_detrend_rscv_model	26.222042	0.005159	0.011635	0.051097	0.182370	0.071827	0.428406	0.388926	-1834.314679	-1745.126756

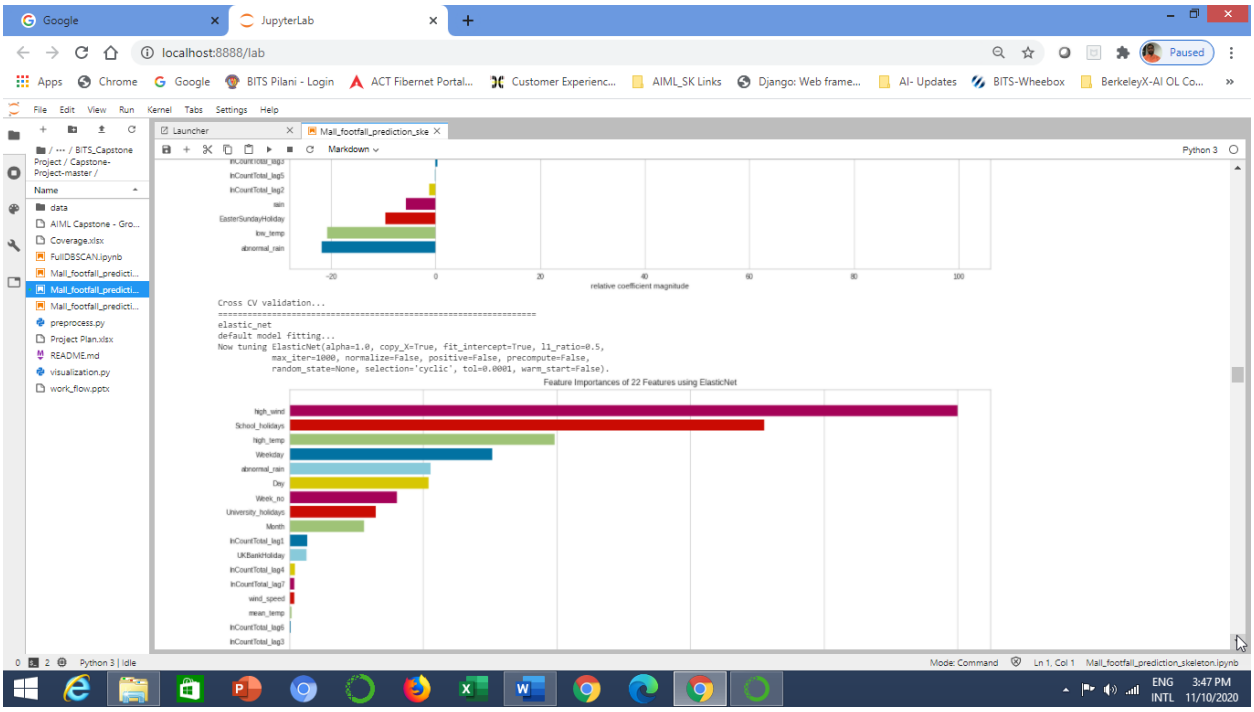


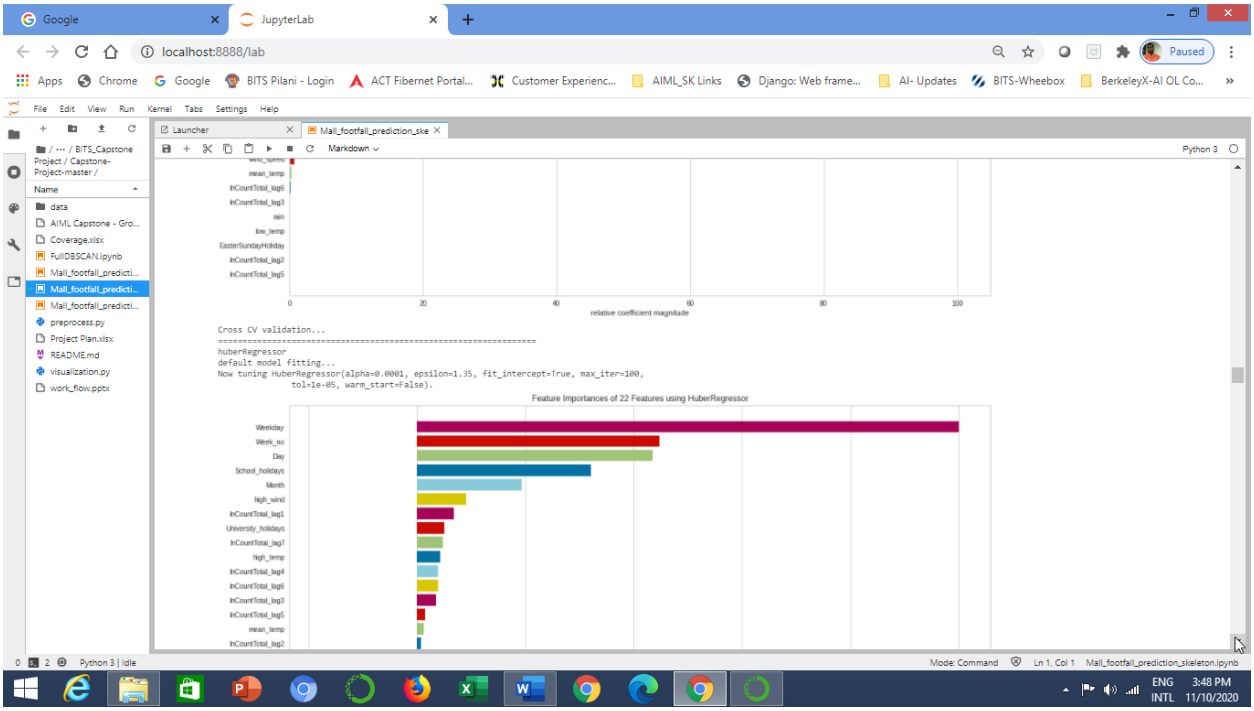
```
relative importance

Cross CV validation...
============================================================
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
               importance_type='split', learning_rate=0.1, max_depth=-1,
               min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
               n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
               random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
               subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
============================================================
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f9271538d68>.
Cross CV validation...
============================================================
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=(100,), learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=200,
              momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
              power_t=0.5, random_state=None, shuffle=True, solver='adam',
              tol=0.0001, validation_fraction=0.1, verbose=False,
              warm_start=False).
Cross CV validation...
```

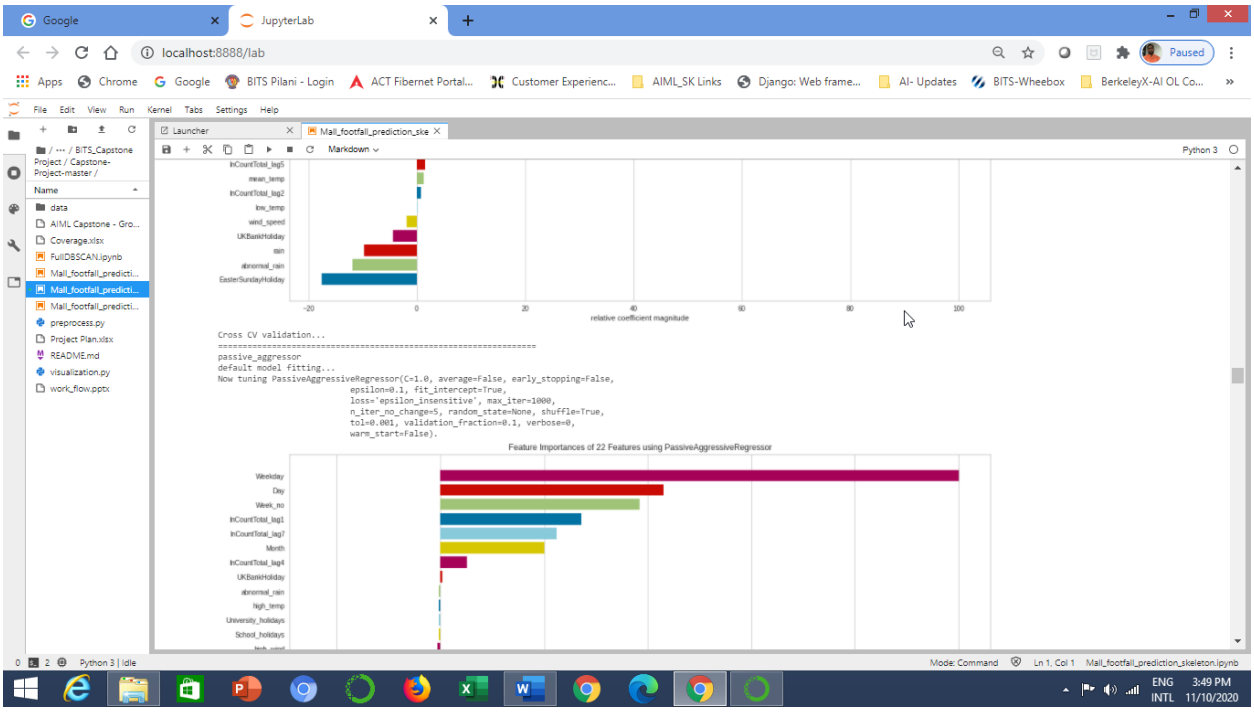

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
34 adaBoost_detrend_rscv_model 26.222042 0.005159 0.011635 0.051097 0.182370 0.071827 0.428406 0.388926 -1834.314679 -1745.126756
33 adaBoost_detrend_default_model 26.575663 0.006422 0.024886 0.062364 0.190234 0.080134 0.288541 0.239401 -1756.171570 -1666.983647
30 svm_regressor_detrend_default_model 26.644562 0.005790 0.006611 0.057169 0.158780 0.076095 0.358467 0.314157 -1793.105957 -1703.918034
7.1.3 Model training with outlier removed data
[34]: reg_models = regression_model(data_outliersremoved,dataset_type='Outliers_Removed')
loss_values_all_models.append(loss_df_outlier_removed)
Defined 20 models
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
Feature Importances of 22 Features using LinearRegression
high_wind
School_holidays
high_temp
Weekday
University_holidays
Day
Week_no
Month
UKBankHoliday
InCountTotal_lag1
wind_speed
mean_temp
InCountTotal_lag4
InCountTotal_lag7
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag2
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 3:34 PM 11/10/2020

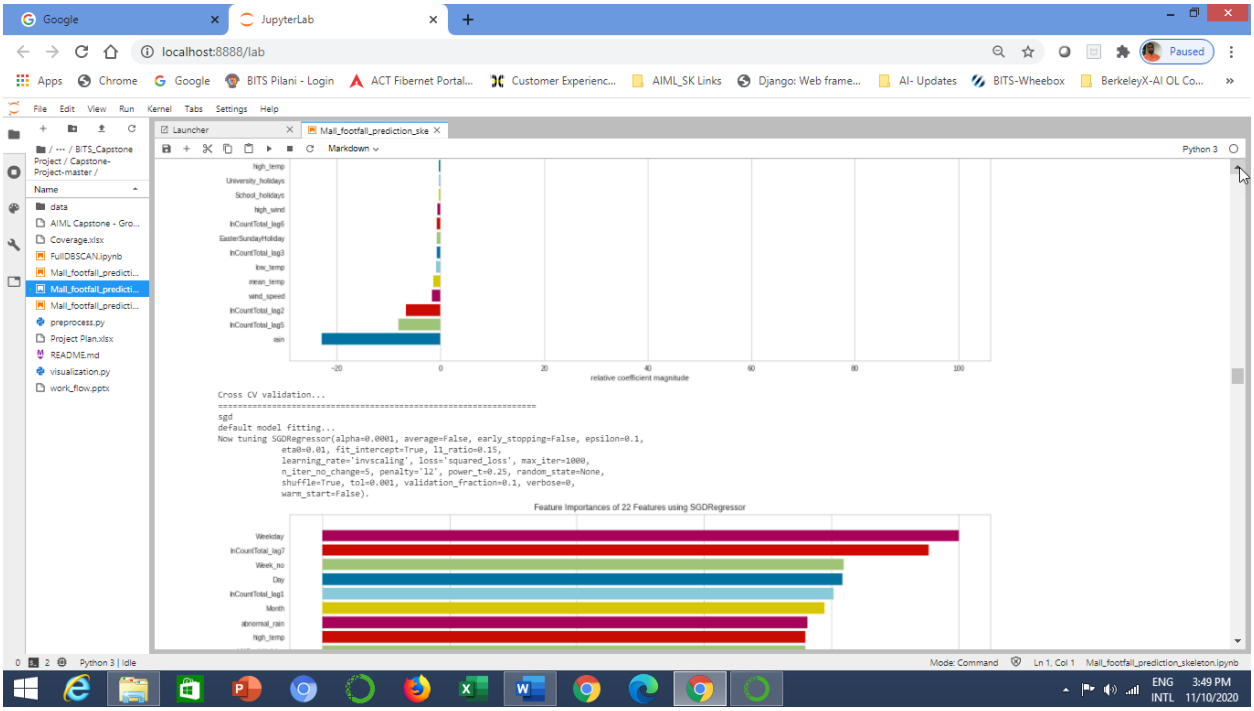

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag5
InCountTotal_lag2
rain
EasterSundayHoliday
abnormal_rain
low_temp
-20
0
20
40
60
80
100
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using Lasso
high_wind
School_holidays
high_temp
Weekday
Day
abnormal_rain
Week_no
Month
University_holidays
InCountTotal_lag1
UKBankHoliday
InCountTotal_lag4
InCountTotal_lag7
wind_speed
mean_temp
InCountTotal_lag6
InCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 3:45 PM 11/10/2020

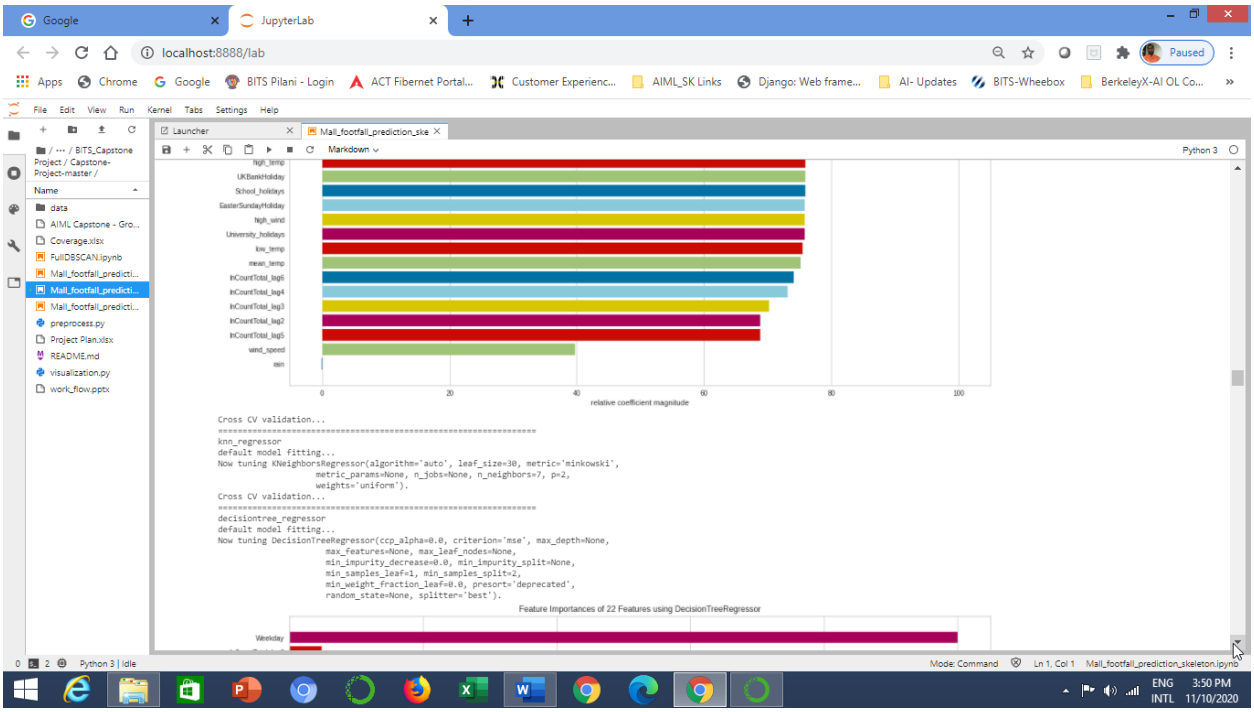

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 22 Features using Ridge
high_wind
School_holidays
high_temp
Weekday
Day
Week_no
University_holidays
Month
InCountTotal_lag1
UKBankHoliday
wind_speed
InCountTotal_lag4
InCountTotal_lag7
mean_temp
InCountTotal_lag6
InCountTotal_lag3
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:46 PM
11/10/2020

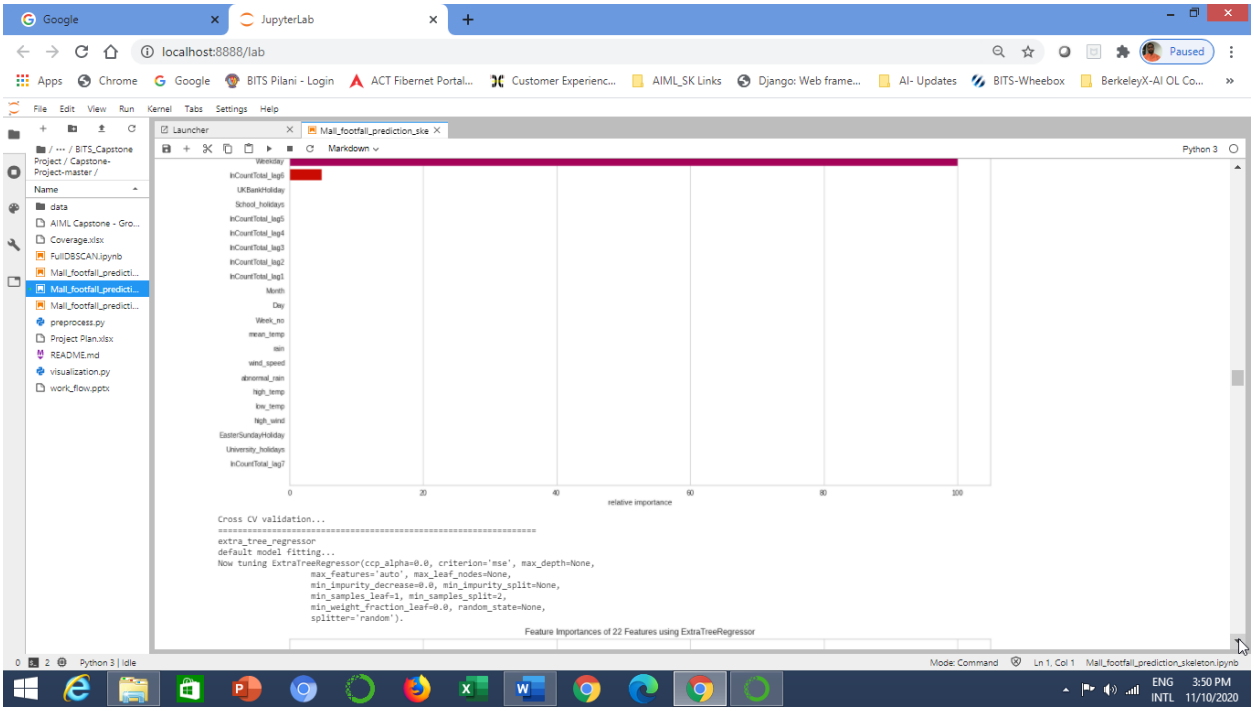

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using ElasticNet
high_wind
School_holidays
high_temp
Weekday
abnormal_rain
Day
Week_no
University_holidays
Month
InCountTotal_lag1
UKBankHoliday
InCountTotal_lag4
InCountTotal_lag7
wind_speed
mean_temp
InCountTotal_lag6
InCountTotal_lag3
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:47 PM
11/10/2020

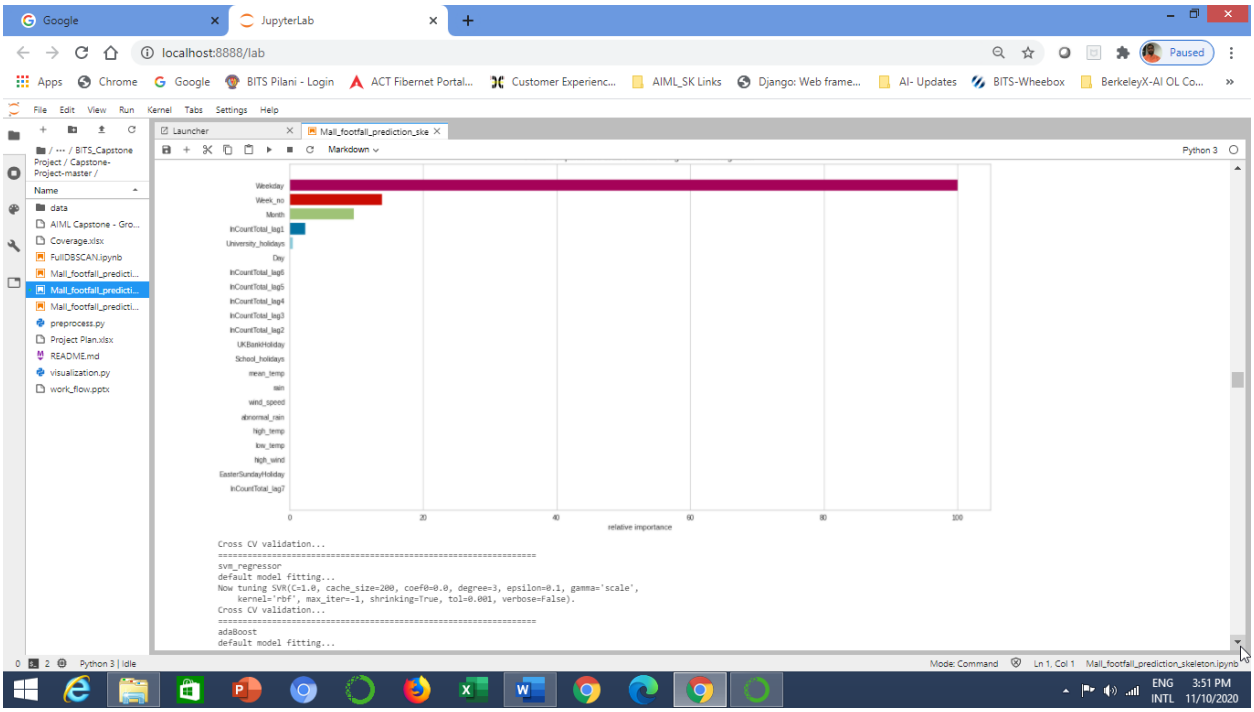

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
huberRegressor
default model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 22 Features using HuberRegressor
Weekday
Week_no
Day
School_holidays
Month
high_wind
InCountTotal_lag1
University_holidays
InCountTotal_lag7
high_temp
InCountTotal_lag4
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
mean_temp
InCountTotal_lag2
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:48 PM
11/10/2020

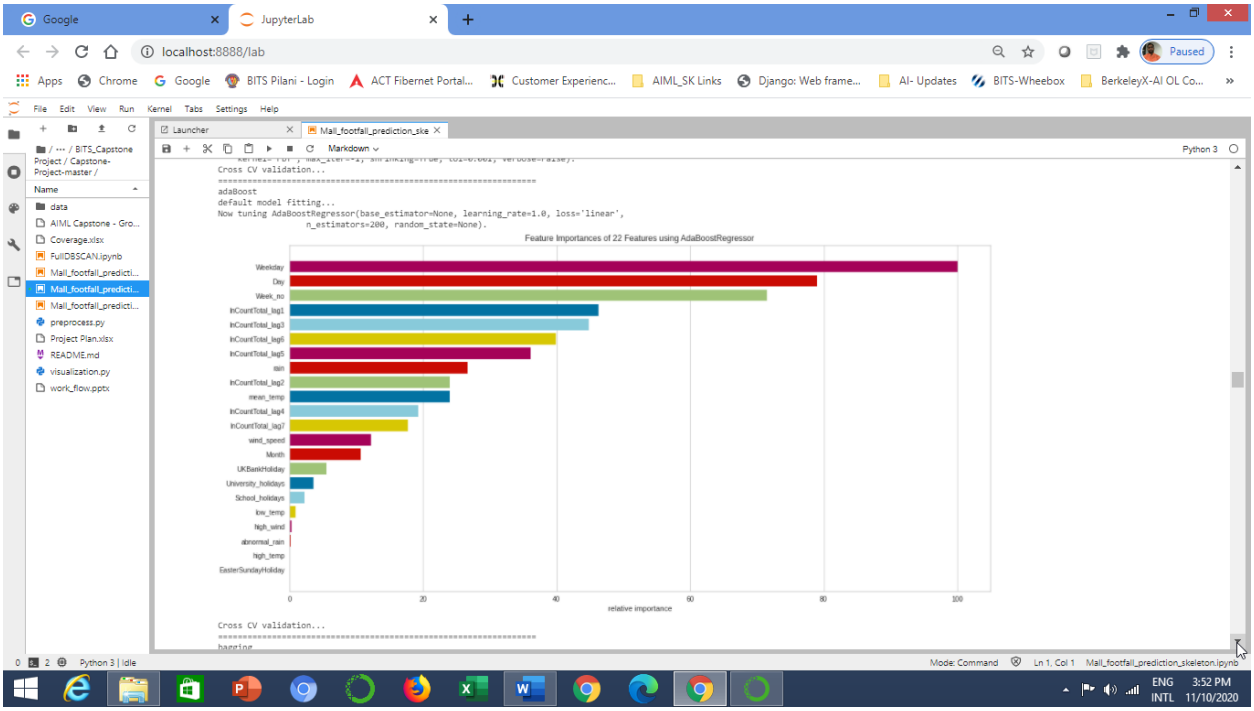

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using PassiveAggressiveRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag7
Month
InCountTotal_lag4
UKBankHoliday
abnormal_rain
high_temp
University_holidays
School_holidays
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:49 PM
11/10/2020

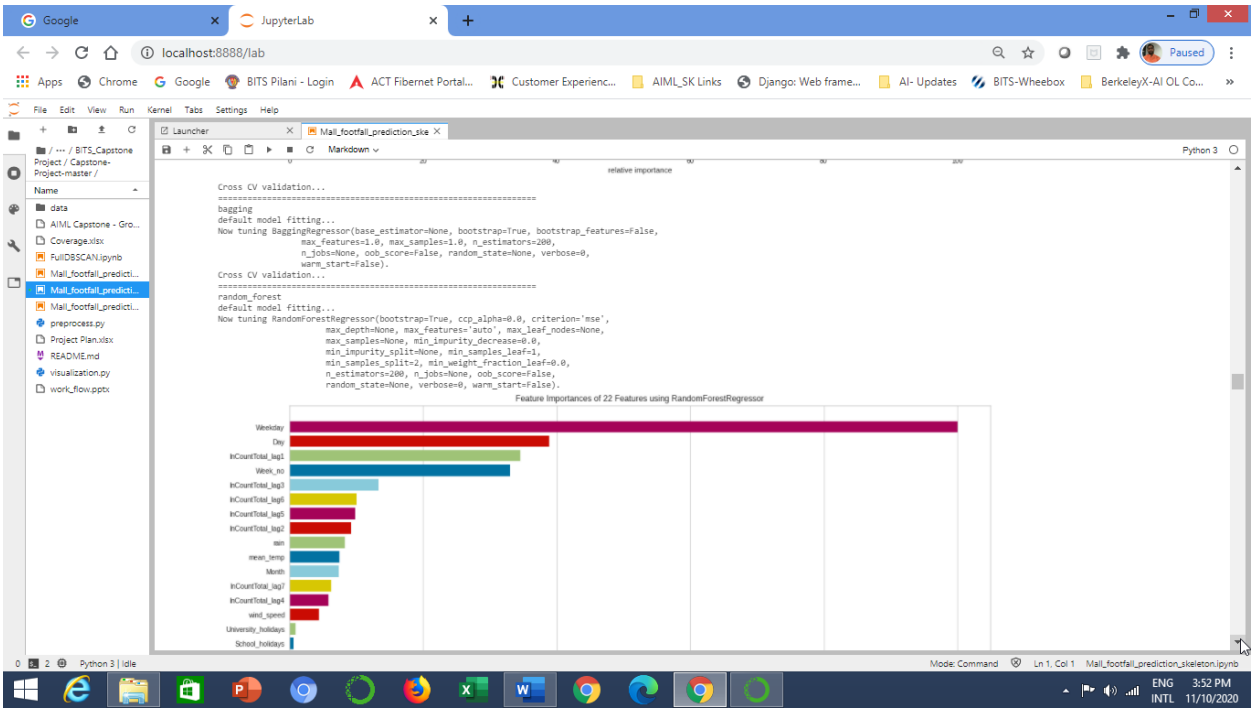

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
high_temp
University_holidays
School_holidays
high_wind
InCountTotal_lag6
EasterSundayHoliday
InCountTotal_lag3
low_temp
mean_temp
wind_speed
InCountTotal_lag2
InCountTotal_lag5
rain
relative coefficient magnitude
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using SGDRegressor
Weekday
InCountTotal_lag7
Week_no
Day
InCountTotal_lag1
Month
abnormal_rain
high_temp
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:49 PM
11/10/2020

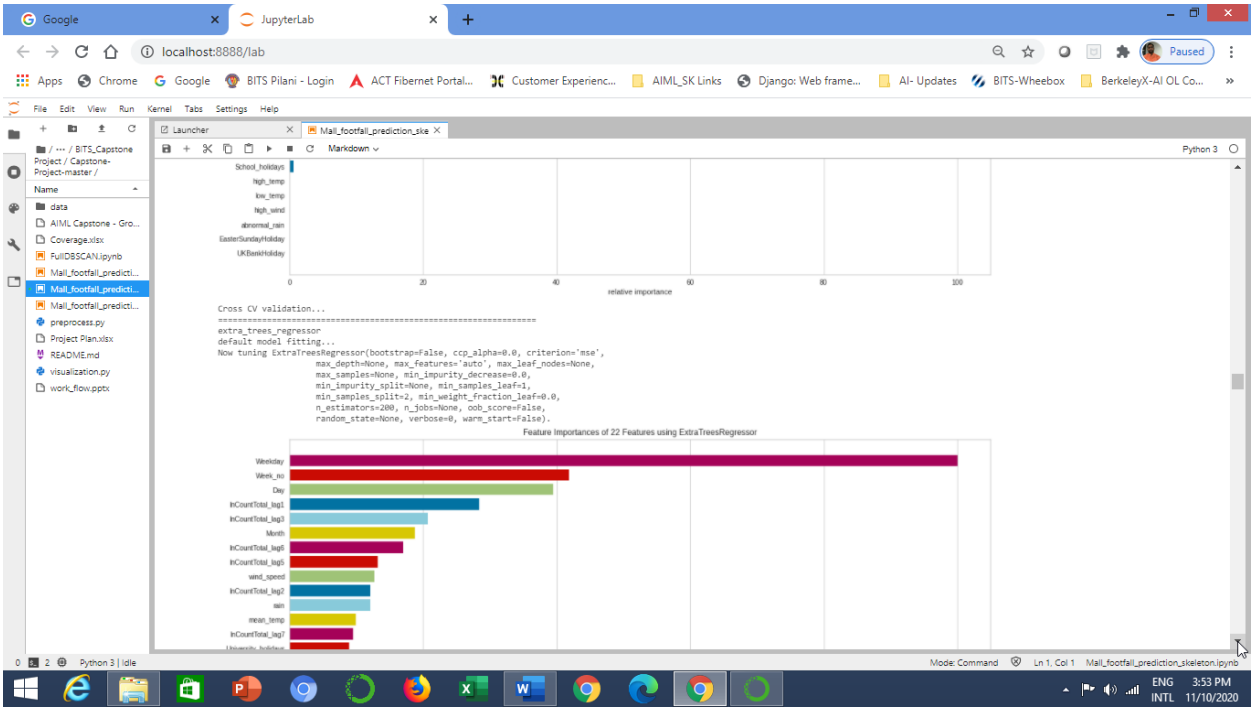

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
high_temp
UKBankHoliday
School_holidays
EasterSundayHoliday
high_wind
University_holidays
low_temp
mean_temp
InCountTotal_lag6
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag5
wind_speed
rain
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 22 Features using DecisionTreeRegressor
Weekday
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:50 PM
11/10/2020

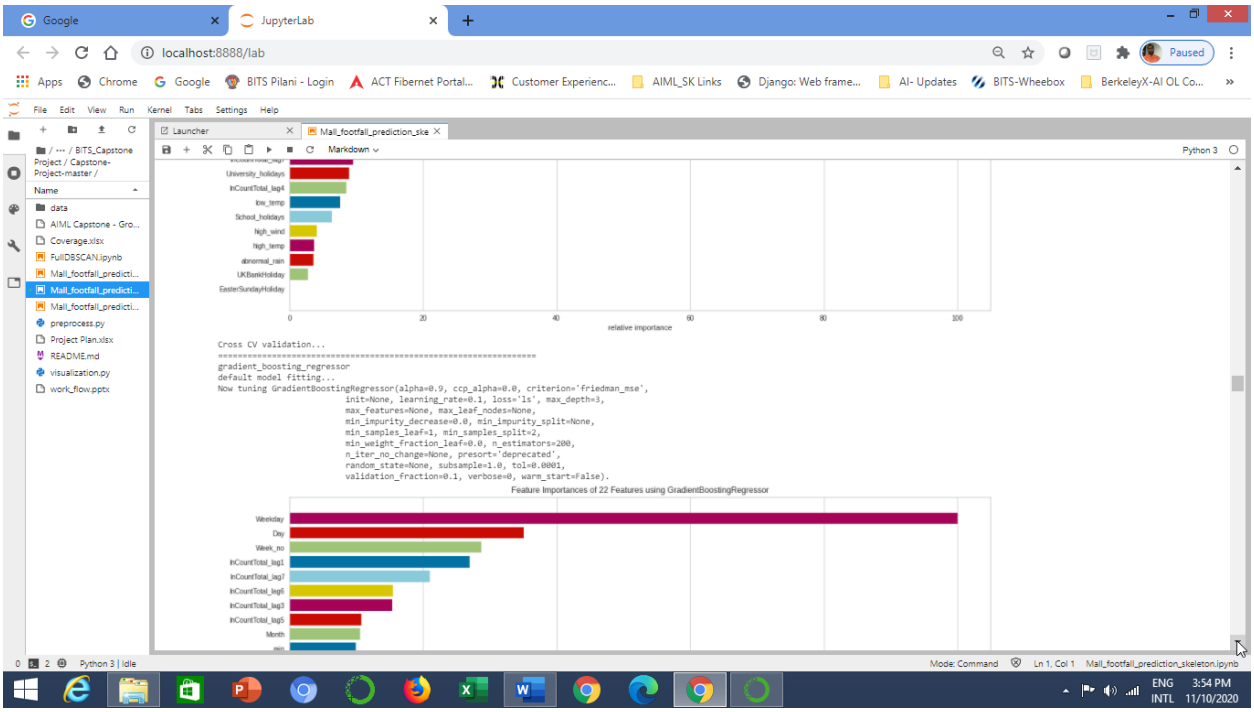

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Weekday
InCountTotal_lag6
UKBankHoliday
School_holidays
InCountTotal_lag5
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag1
Month
Day
Week_no
mean_temp
rain
wind_speed
abnormal_rain
high_temp
low_temp
high_wind
EasterSundayHoliday
University_holidays
InCountTotal_lag7
0
20
40
60
80
100
relative importance
Cross CV validation...
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').
Feature Importances of 22 Features using ExtraTreeRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:50 PM
11/10/2020

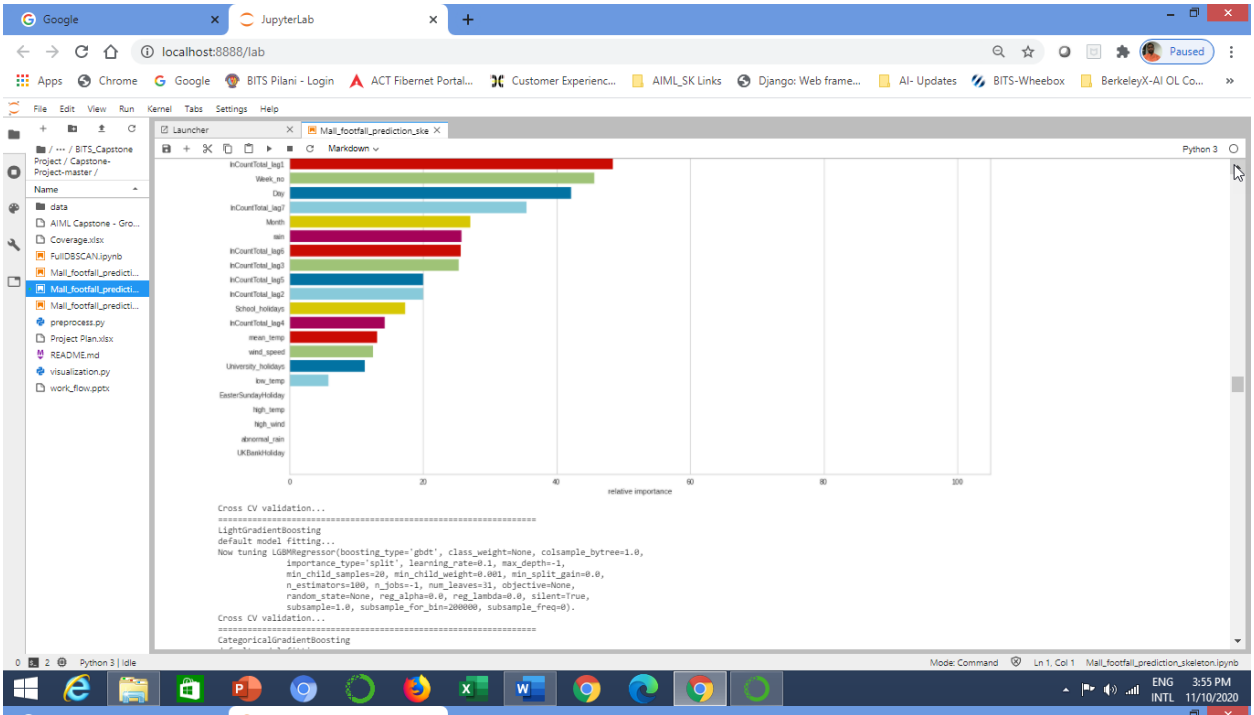

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Weekday
Week_no
Month
InCountTotal_lag1
University_holidays
Day
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
UKBankHoliday
School_holidays
mean_temp
rain
wind_speed
abnormal_rain
high_temp
low_temp
high_wind
EasterSundayHoliday
InCountTotal_lag7
0
20
40
60
80
100
relative importance
Cross CV validation...
svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:51 PM
11/10/2020

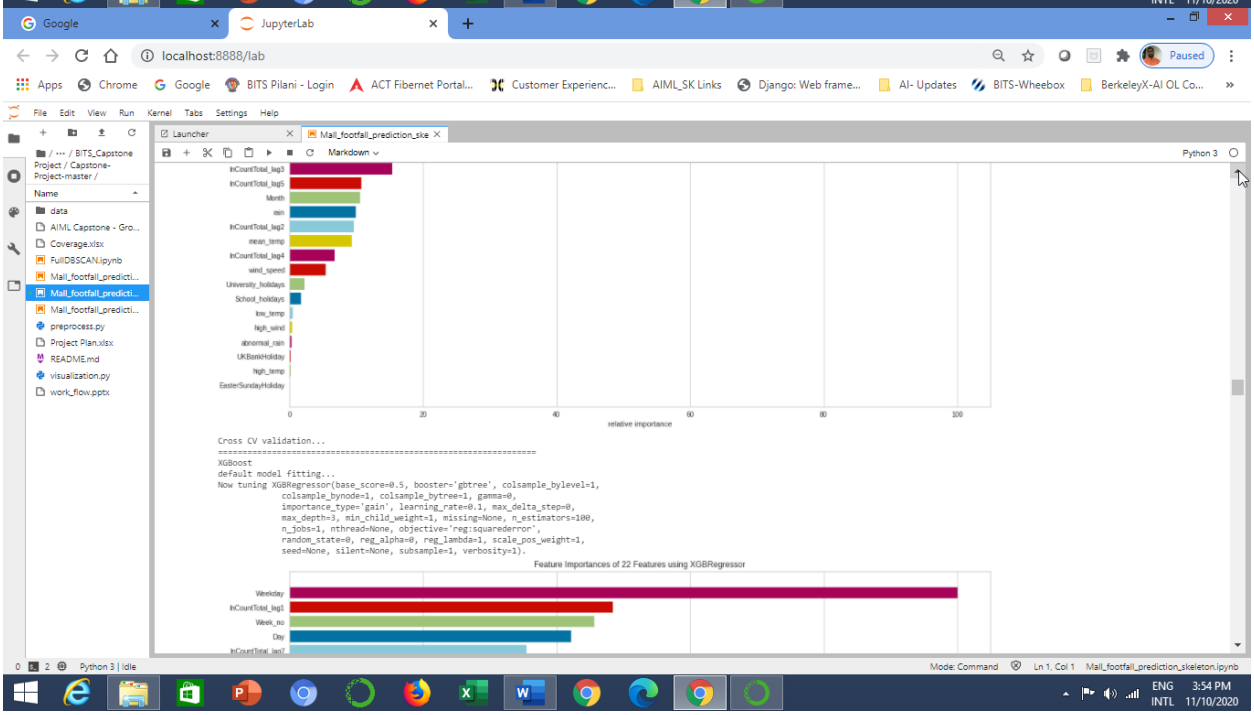

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 22 Features using AdaBoostRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag5
rain
InCountTotal_lag2
mean_temp
InCountTotal_lag4
InCountTotal_lag7
wind_speed
Month
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
abnormal_rain
high_temp
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
bagging
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:52 PM
11/10/2020

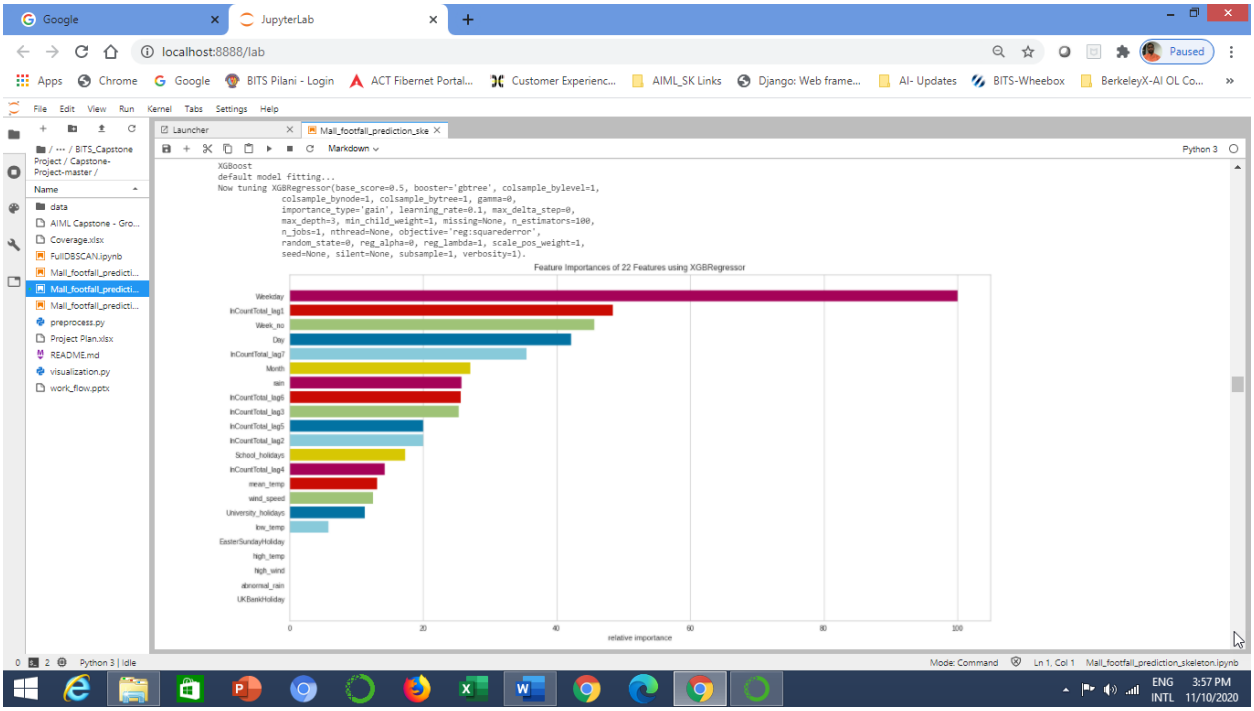

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Weekday
Day
InCountTotal_lag1
Week_no
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag2
rain
mean_temp
Month
InCountTotal_lag7
InCountTotal_lag4
wind_speed
University_holidays
School_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:52 PM
11/10/2020

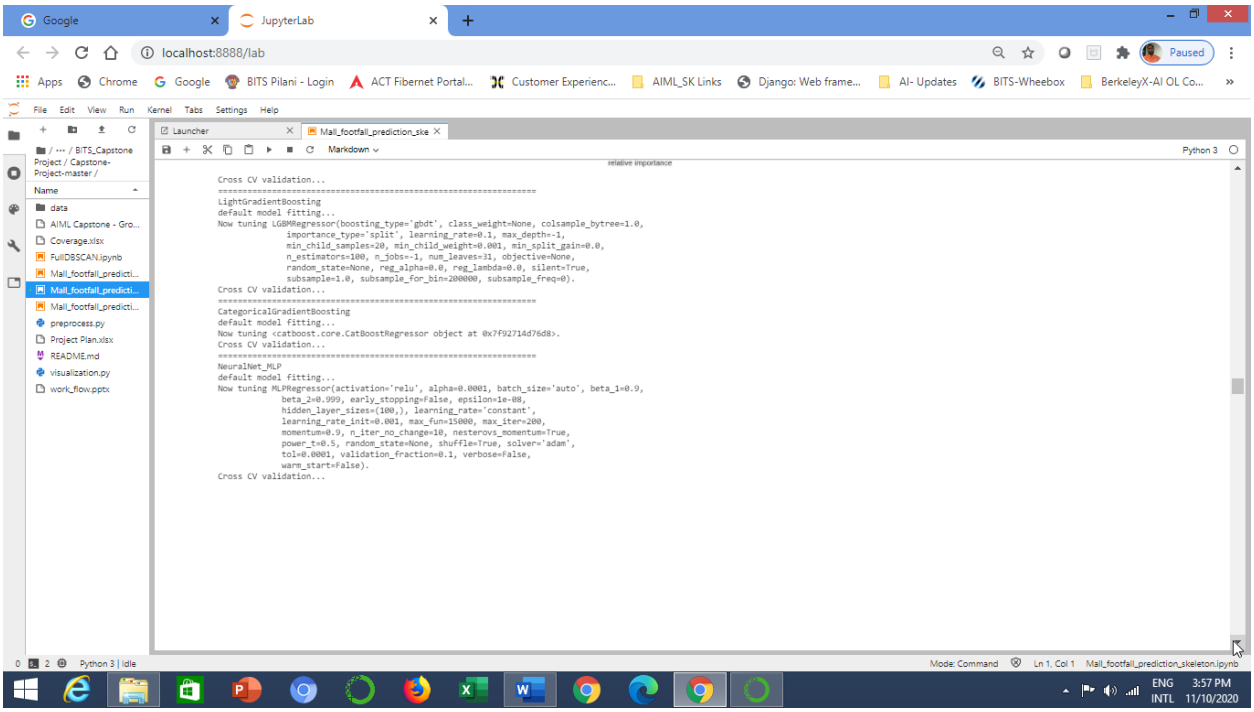

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using ExtraTreesRegressor
Weekday
Week_no
Day
InCountTotal_lag1
InCountTotal_lag3
Month
InCountTotal_lag6
InCountTotal_lag5
wind_speed
InCountTotal_lag2
rain
mean_temp
InCountTotal_lag7
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:53 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 22 Features using GradientBoostingRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag7
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
Month
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:54 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag1
Week_no
Day
InCountTotal_lag7
Month
rain
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag2
School_holidays
InCountTotal_lag4
mean_temp
wind_speed
University_holidays
low_temp
EasterSundayHoliday
high_temp
high_wind
abnormal_rain
UKBankHoliday
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:55 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag3
InCountTotal_lag5
Month
rain
InCountTotal_lag2
mean_temp
InCountTotal_lag4
wind_speed
University_holidays
School_holidays
low_temp
high_wind
abnormal_rain
UKBankHoliday
high_temp
EasterSundayHoliday
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 22 Features using XGBRegressor
Weekday
InCountTotal_lag1
Week_no
Day
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
3:54 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 22 Features using XGBRegressor
Weekday
InCountTotal_lag1
Week_no
Day
InCountTotal_lag7
Month
sin
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag2
School_holidays
InCountTotal_lag4
mean_temp
wind_speed
University_holidays
low_temp
EasterSundayHoliday
high_temp
high_wind
abnormal_rain
UKBankHoliday
0
20
40
60
80
100
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:57 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f92714d76d8>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
learning_rate_init=0.001, max_fun=15000, max_iter=200,
momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
power_t=0.5, random_state=None, shuffle=True, solver='adam',
tol=0.0001, validation_fraction=0.1, verbose=False,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
3:57 PM
11/10/2020

Google | JupyterLab | localhost:8888/lab

Launcher | Mall_footfall_prediction_ske

```
[35]: loss_df_outlier_removed.sort_values('mape').head(25)
```

[35]:

	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
36	bagging_Outliers_Removed_default_model	11.815070	0.004205	-0.013330	0.044784	-0.009810	0.064849	0.472342	0.435897	-1907.281280	-1818.093357
39	random_forest_Outliers_Removed_default_model	11.878525	0.004131	-0.012771	0.044793	-0.007844	0.064274	0.481667	0.445866	-1913.646660	-1824.458737
54	CategoricalGradientBoosting_Outliers_Removed_d...	11.991476	0.004348	-0.012474	0.045386	-0.007304	0.065937	0.454493	0.416815	-1895.404784	-1806.216861
40	random_forest_Outliers_Removed_rscv_model	12.102028	0.004409	-0.015046	0.045848	-0.012178	0.066401	0.446783	0.408573	-1890.394464	-1801.206541
45	gradient_boosting_regressor_Outliers_Removed_d...	12.205760	0.004710	-0.017622	0.047270	-0.021842	0.068632	0.408985	0.368164	-1866.800007	-1777.612084
52	LightGradientBoosting_Outliers_Removed_rscv_model	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
42	extra_trees_regressor_Outliers_Removed_default...	12.529236	0.004402	-0.015246	0.047466	-0.014811	0.066346	0.447709	0.409563	-1890.992657	-1801.804734
43	extra_trees_regressor_Outliers_Removed_rscv_model	12.538173	0.004435	-0.015710	0.047578	-0.015819	0.066594	0.443570	0.405138	-1888.327207	-1799.139284
48	XGBoost_Outliers_Removed_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708
46	gradient_boosting_regressor_Outliers_Removed_r...	12.674935	0.004782	-0.017924	0.048787	-0.018662	0.069151	0.400024	0.358584	-1861.427696	-1772.239773
55	CategoricalGradientBoosting_Outliers_Removed_r...	12.691450	0.004432	-0.013857	0.047479	-0.008073	0.066570	0.443965	0.405561	-1888.580893	-1799.392970
51	LightGradientBoosting_Outliers_Removed_default...	12.709911	0.004599	-0.014983	0.048127	-0.014091	0.067815	0.422976	0.383121	-1875.352644	-1786.164721
44	extra_trees_regressor_Outliers_Removed_crosscv...	12.759066	0.004284	0.001634	0.045323	3.553093	0.065373	0.424719	NaN	NaN	NaN
37	bagging_Outliers_Removed_rscv_model	13.075546	0.004862	-0.016499	0.049595	-0.009537	0.069731	0.389914	0.347776	-1855.462495	-1766.274572
34	adaBoost_Outliers_Removed_rscv_model	13.114304	0.004669	-0.010261	0.048754	0.005294	0.068328	0.414209	0.373749	-1869.969828	-1780.781905
49	XGBoost_Outliers_Removed_rscv_model	13.225608	0.004730	-0.010576	0.049062	0.006588	0.068771	0.406586	0.365599	-1865.353668	-1776.165745
41	random_forest_Outliers_Removed_crosscv_model	13.306124	0.004530	0.001448	0.047006	3.838474	0.067111	0.399623	NaN	NaN	NaN
47	gradient_boosting_regressor_Outliers_Removed_c...	13.352104	0.004582	0.000457	0.047676	3.300756	0.067519	0.384451	NaN	NaN	NaN
56	CategoricalGradientBoosting_Outliers_Removed_c...	13.626440	0.004750	0.000092	0.048497	3.750387	0.068608	0.378603	NaN	NaN	NaN
35	adaBoost_Outliers_Removed_crosscv_model	13.852529	0.004794	0.001173	0.049539	4.010600	0.068853	0.373306	NaN	NaN	NaN
12	huberRegressor_Outliers_Removed_default_model	13.974179	0.005750	-0.018115	0.052593	-0.016486	0.075828	0.278562	0.228733	-1795.612351	-1706.424428
13	huberRegressor_Outliers_Removed_rscv_model	14.059795	0.005811	-0.018731	0.053074	-0.017528	0.076230	0.270892	0.220533	-1791.837026	-1702.649103
38	bagging_Outliers_Removed_crosscv_model	14.094448	0.004852	0.000288	0.050122	3.985995	0.069406	0.363602	NaN	NaN	NaN

Python 3 | Idle — Mode: Command — Ln 1, Col 1 — Mall_footfall_prediction_skeleton.ipynb

ENG INTL 3:58 PM 11/10/2020

Google | JupyterLab | localhost:8888/lab

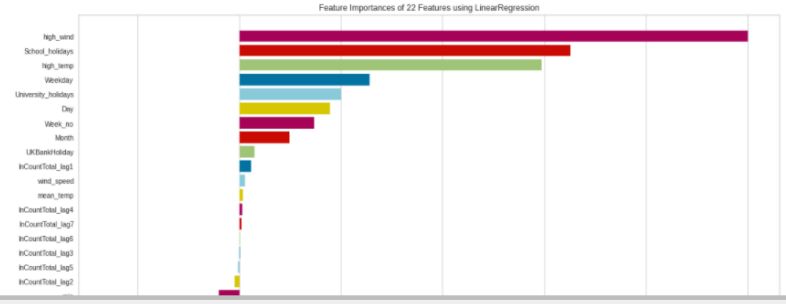
Launcher | Mall_footfall_prediction_ske

13	huberRegressor_Outliers_Removed_rscv_model	14.059795	0.005811	-0.018731	0.053074	-0.017528	0.076230	0.270892	0.220533	-1791.837026	-1702.649103
38	bagging_Outliers_Removed_crosscv_model	14.094448	0.004852	0.000288	0.050122	3.985995	0.069406	0.363602	NaN	NaN	NaN
10	elastic_net_Outliers_Removed_rscv_model	14.331859	0.005650	-0.018112	0.053930	-0.014345	0.075165	0.291117	0.242155	-1801.880079	-1712.692156
7	ridgeRegressor_Outliers_Removed_rscv_model	14.332307	0.005696	-0.019093	0.053998	-0.017202	0.075473	0.285294	0.235930	-1798.959332	-1709.771409

```
[36]: reg_models = regression_model(data_outliersremoved,dataset_type='Outliers_Removed_shuffled')
      _,_,_,_,_,loss_df_outlier_removed = reg_models.train_and_evaluate_models(RFEcross_validation=False, shuffle=True)
      loss_values_all_models.append(loss_df_outlier_removed)
```

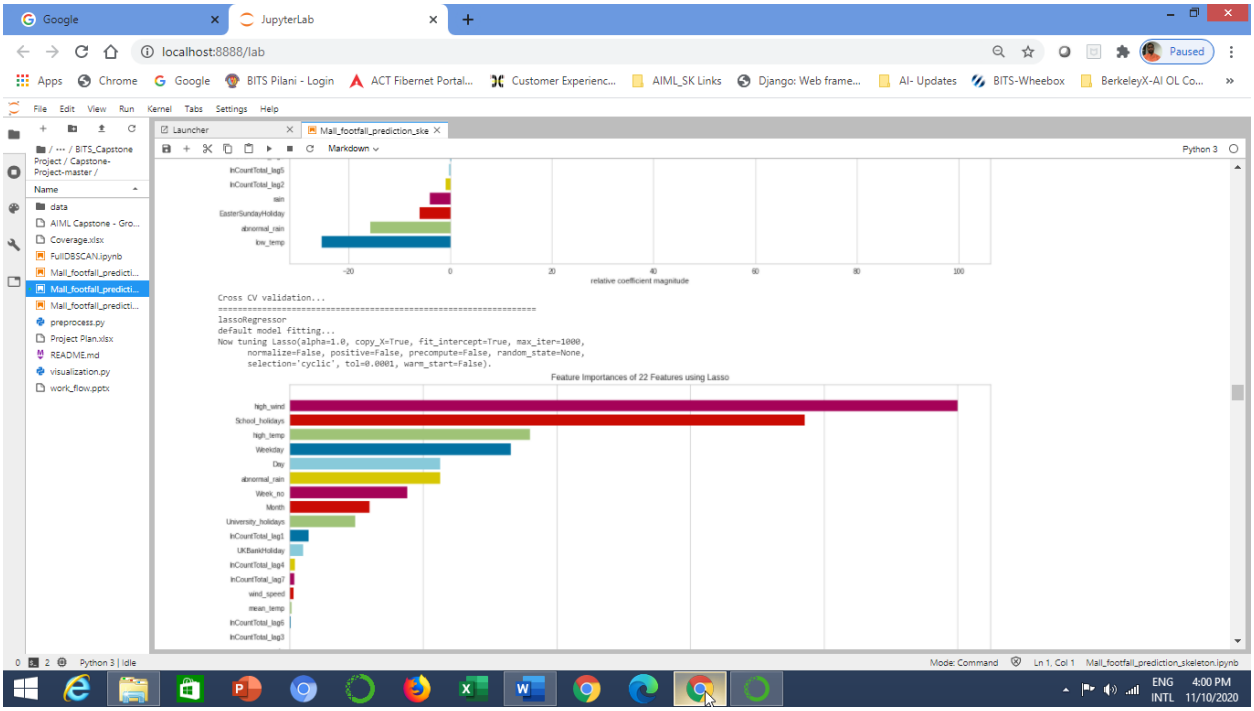
```
Defined 20 models
======================================================================
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
```

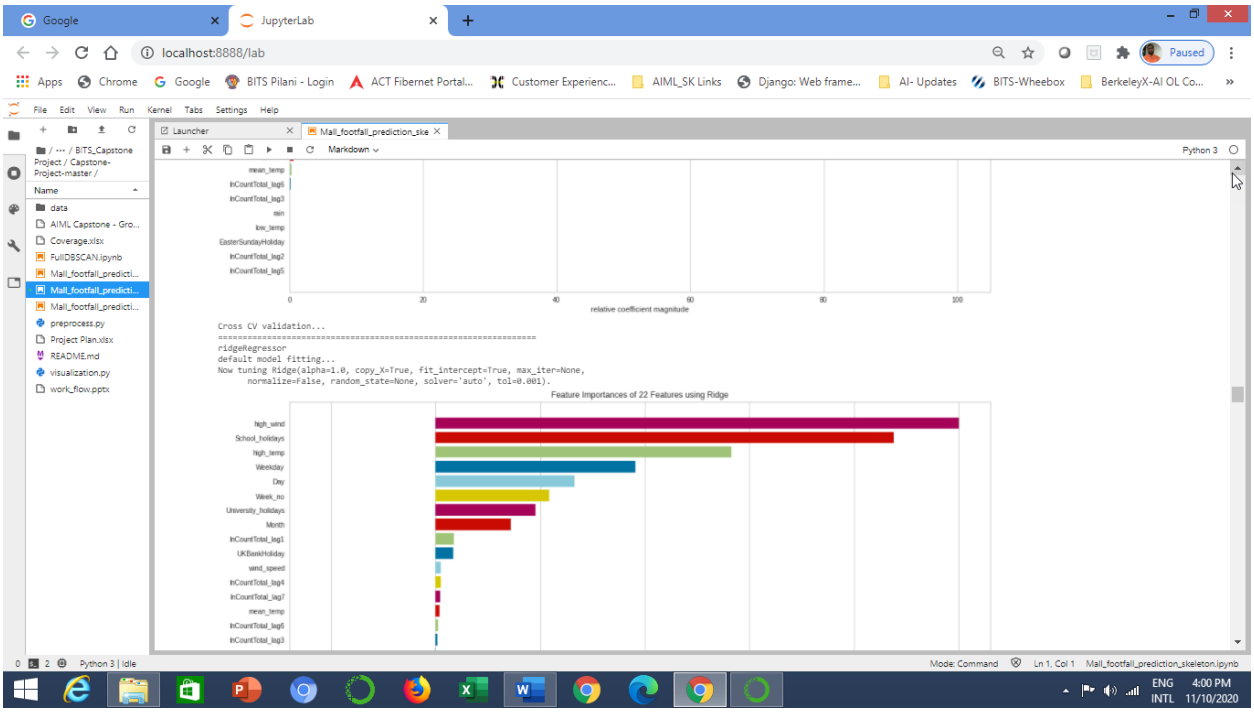
Feature Importances of 22 Features using LinearRegression

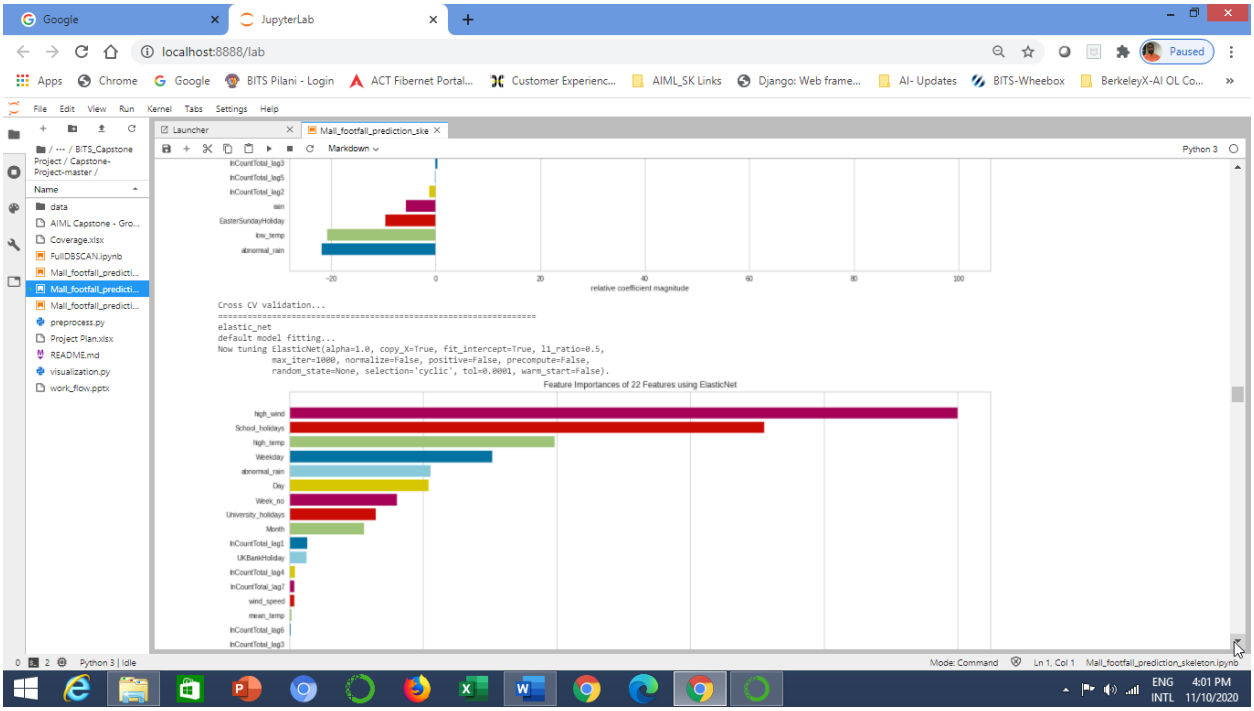


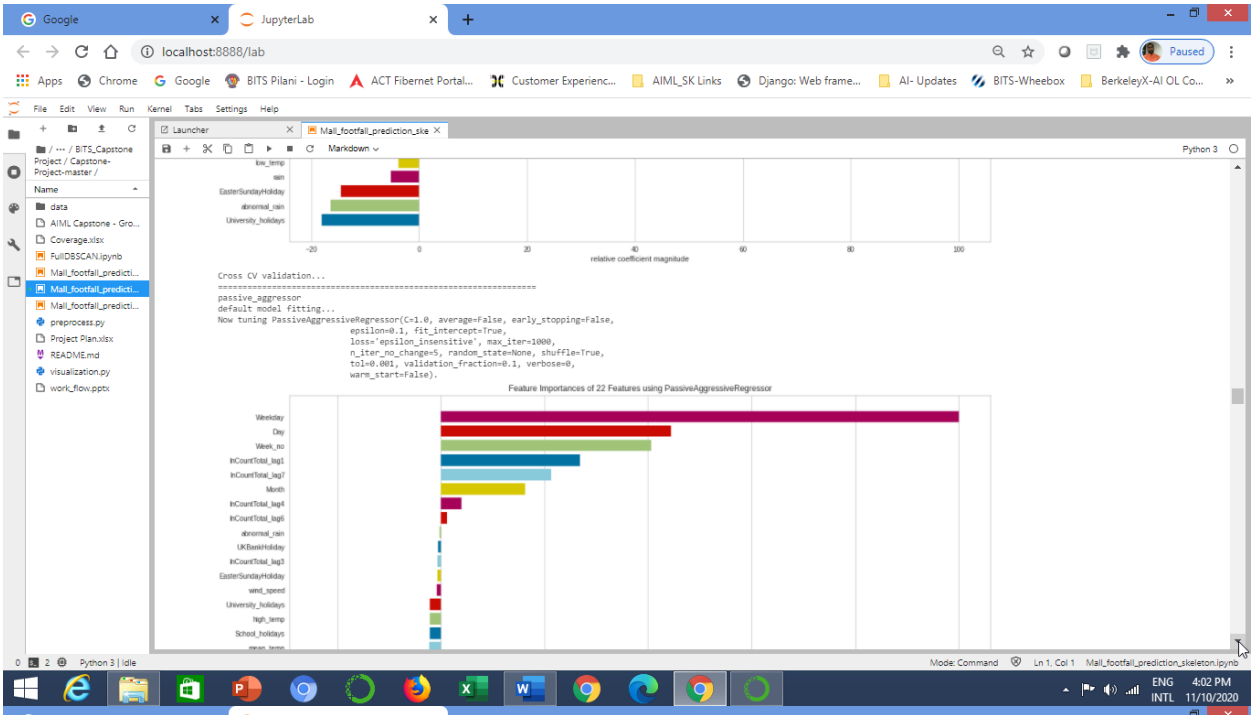
Python 3 | Idle — Mode: Command — Ln 1, Col 1 — Mall_footfall_prediction_skeleton.ipynb

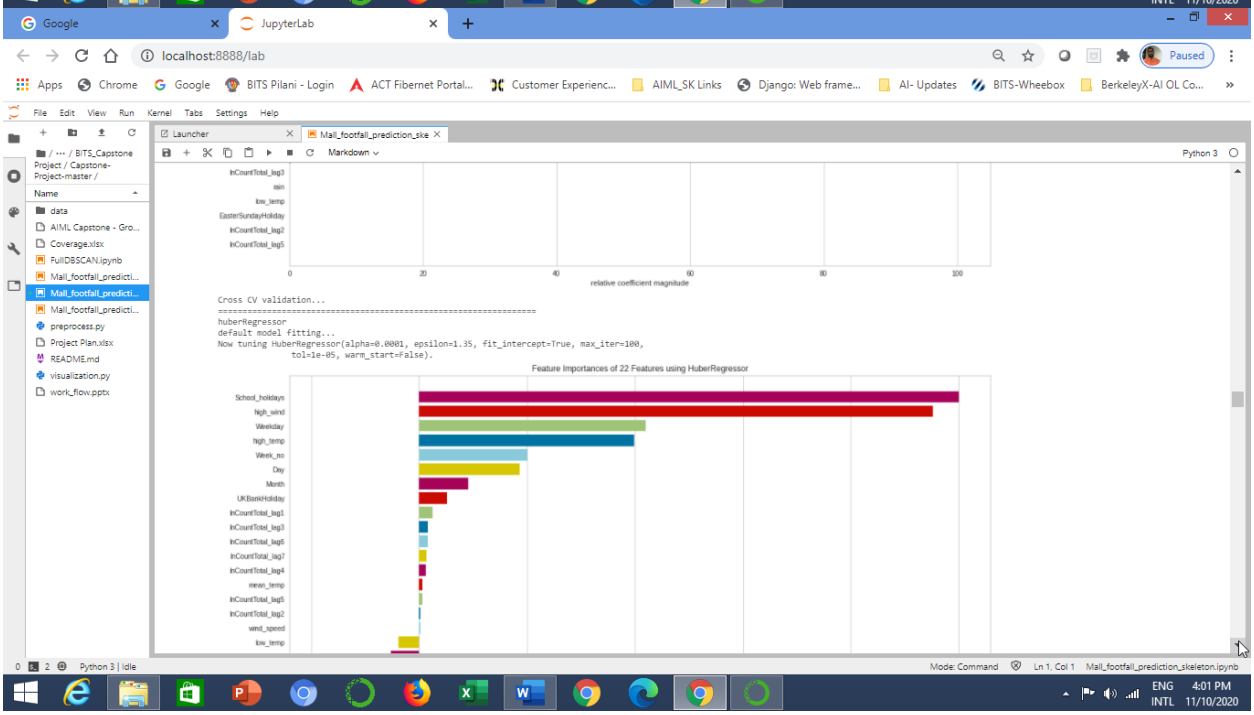
ENG INTL 3:59 PM 11/10/2020

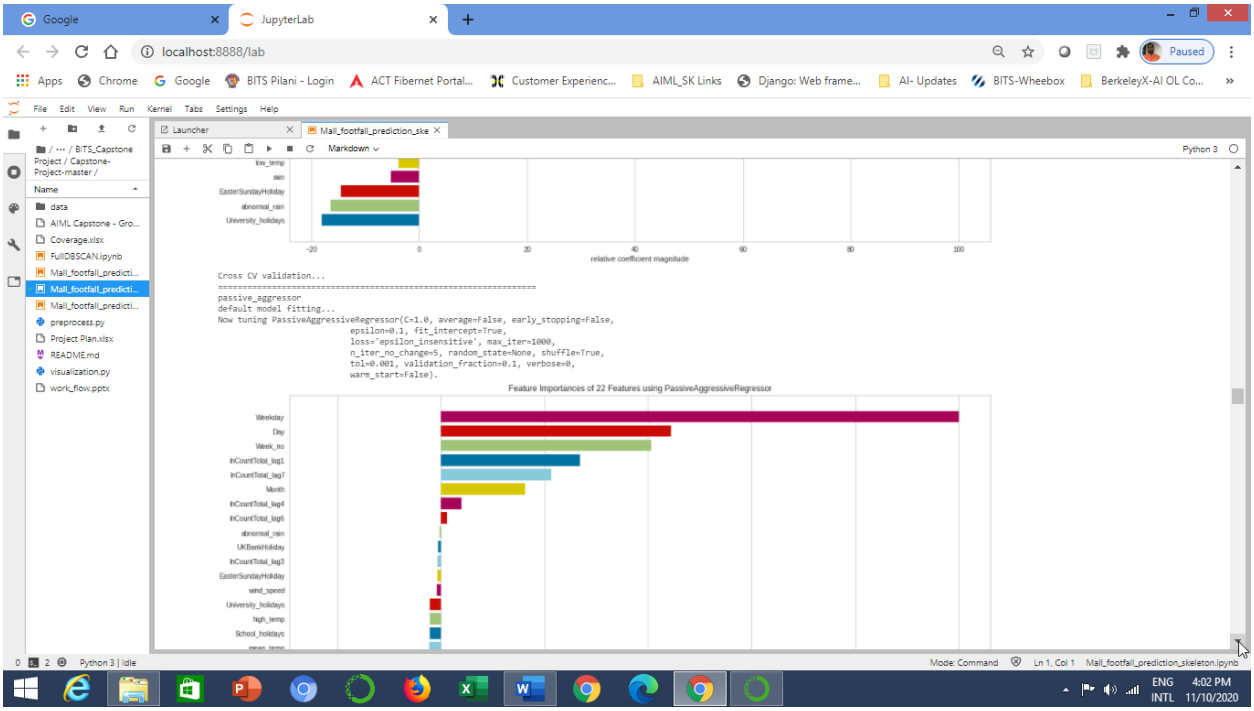

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using Lasso
high_wind
School_holidays
high_temp
Weekday
Day
abnormal_rain
Week_no
Month
University_holidays
InCountTotal_lag1
UKBankHoliday
InCountTotal_lag4
InCountTotal_lag7
wind_speed
mean_temp
InCountTotal_lag6
InCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:00 PM
11/10/2020

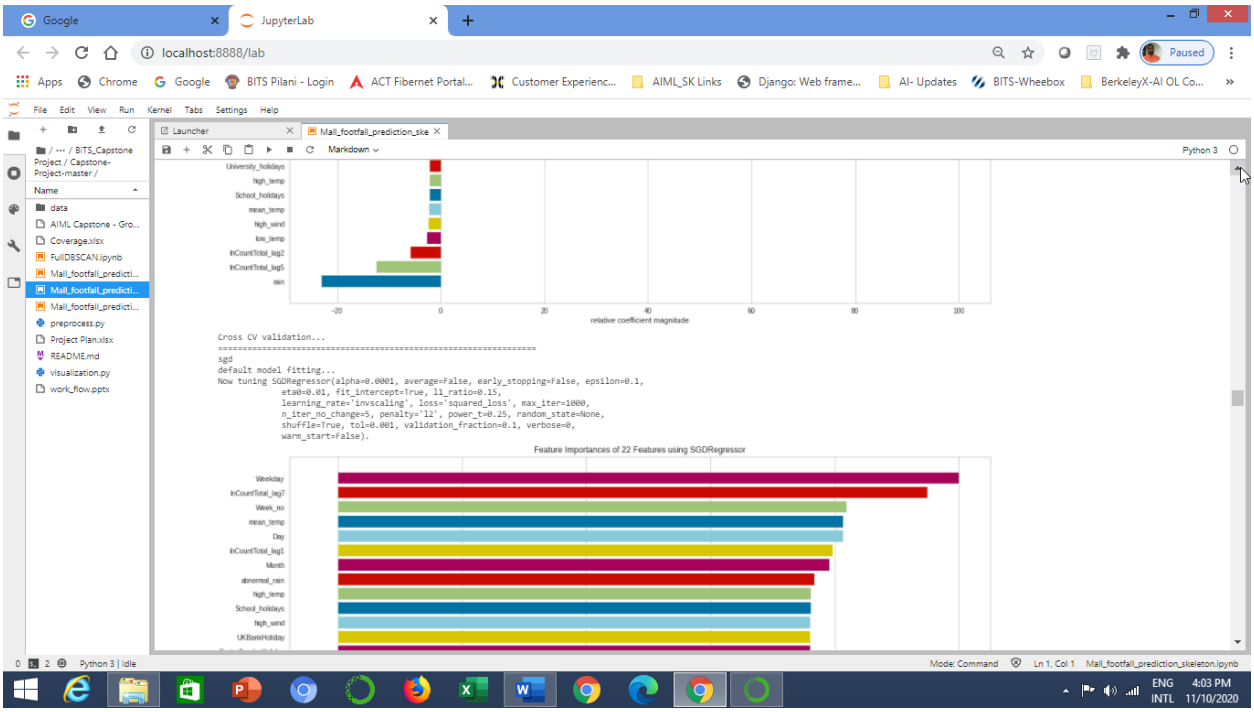

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 22 Features using Ridge
high_wind
School_holidays
high_temp
Weekday
Day
Week_no
University_holidays
Month
InCountTotal_lag1
UKBankHoliday
wind_speed
InCountTotal_lag4
InCountTotal_lag7
mean_temp
InCountTotal_lag6
InCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:00 PM
11/10/2020

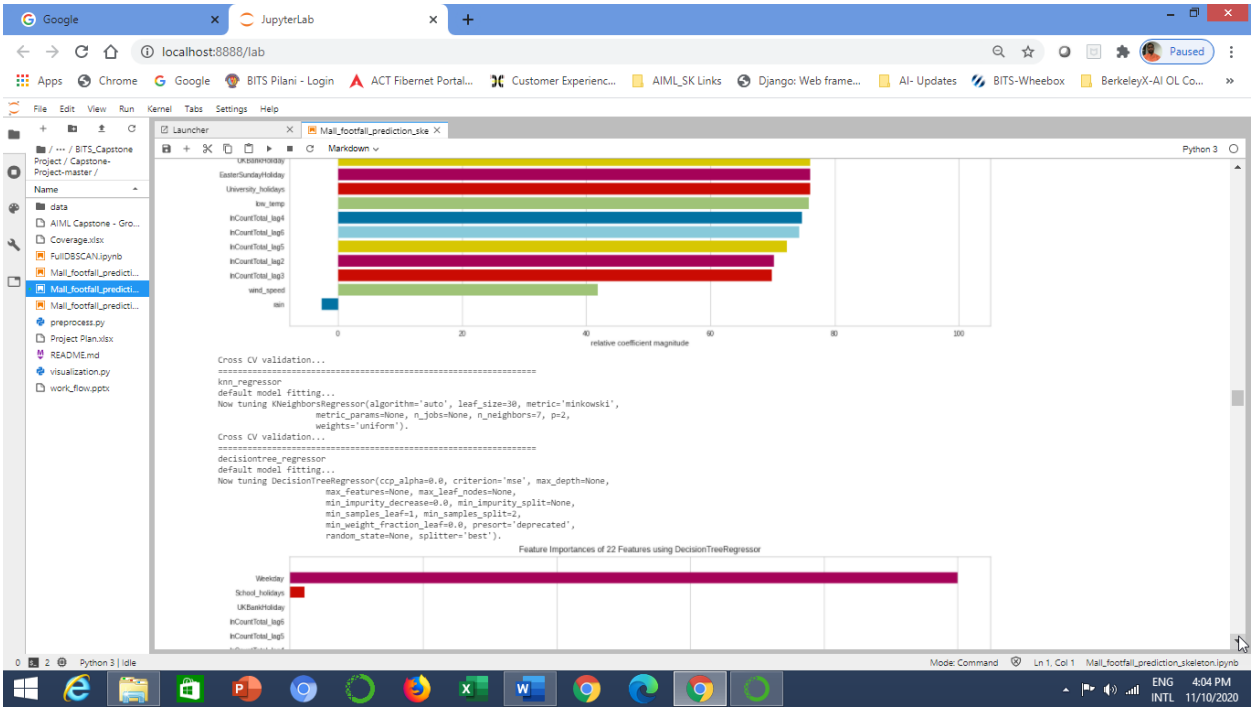

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 22 Features using ElasticNet
high_wind
School_holidays
high_temp
Weekday
abnormal_rain
Day
Week_no
University_holidays
Month
UKBankHoliday
wind_speed
mean_temp
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:01 PM
11/10/2020

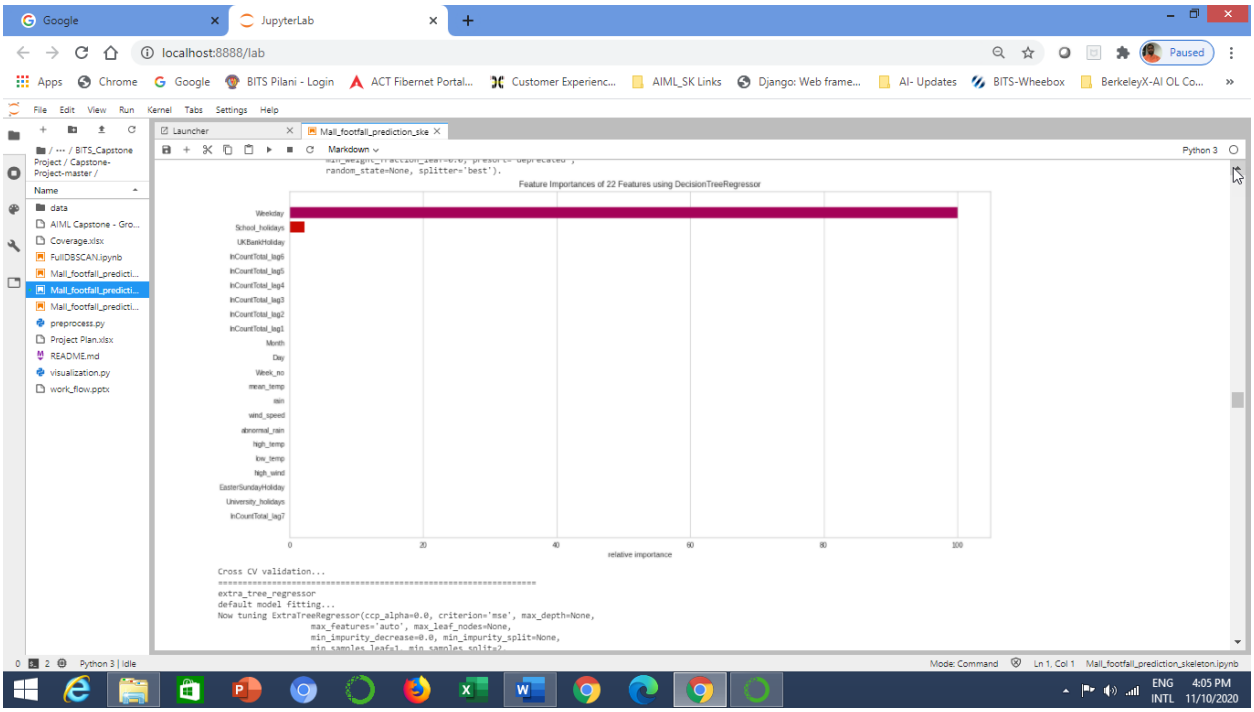

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
low_temp
rain
EasterSundayHoliday
abnormal_rain
University_holidays
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using PassiveAggressiveRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag7
Month
InCountTotal_lag4
InCountTotal_lag6
abnormal_rain
UKBankHoliday
InCountTotal_lag3
EasterSundayHoliday
wind_speed
University_holidays
high_temp
School_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:02 PM
11/10/2020

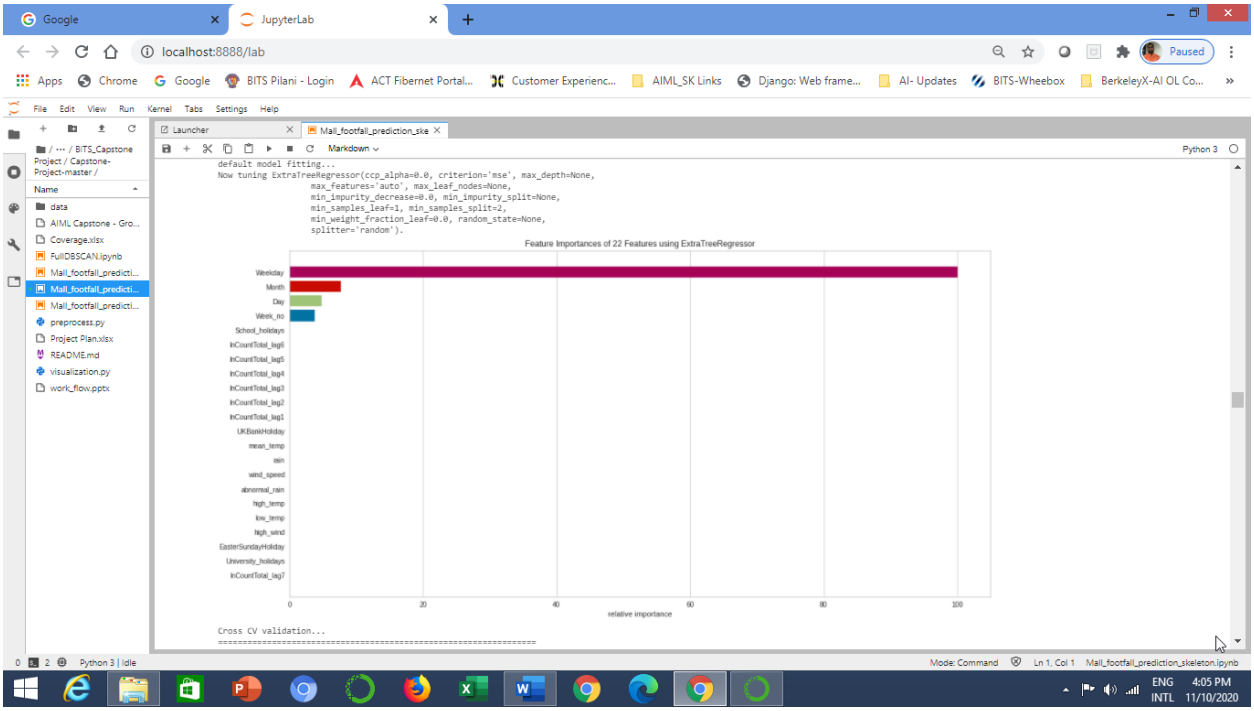

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag3
rain
low_temp
EasterSundayHoliday
InCountTotal_lag2
InCountTotal_lag5
relative coefficient magnitude
Cross CV validation...
huberRegressor
default model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 22 Features using HuberRegressor
School_holidays
high_wind
Weekday
high_temp
Week_no
Day
Month
UKBankHoliday
InCountTotal_lag1
InCountTotal_lag3
InCountTotal_lag6
InCountTotal_lag7
InCountTotal_lag4
mean_temp
InCountTotal_lag5
InCountTotal_lag2
wind_speed
low_temp
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:01 PM
11/10/2020

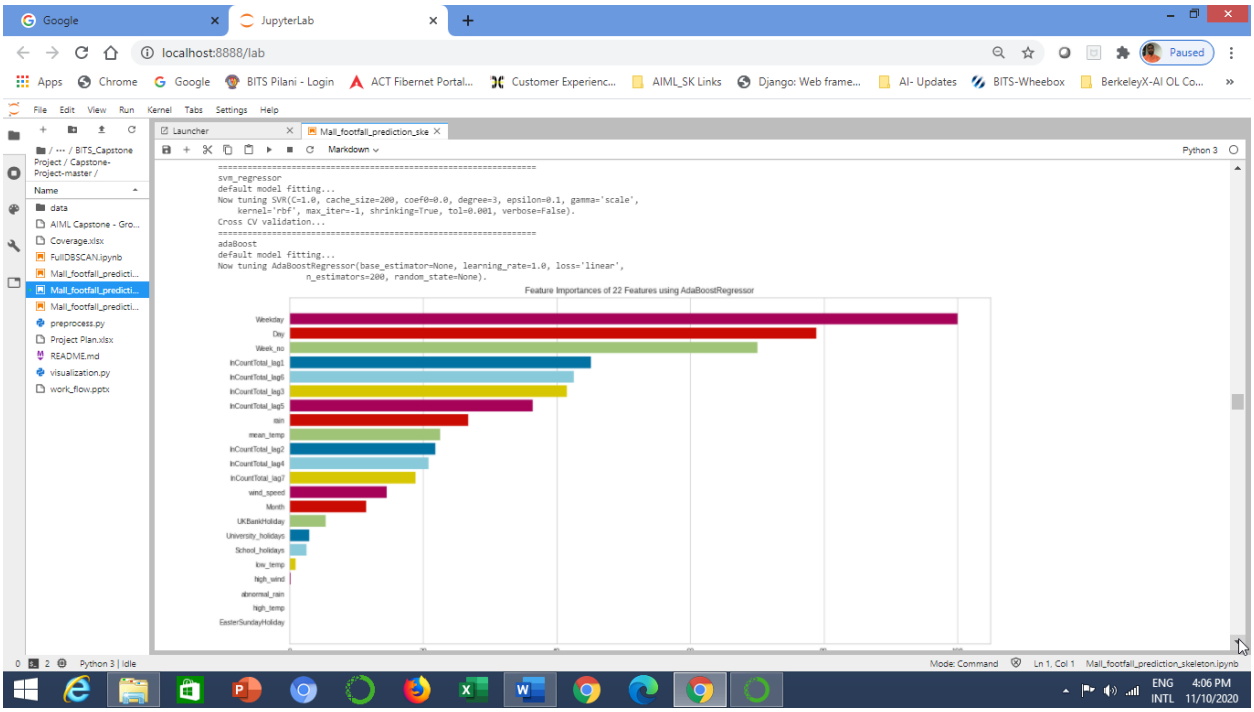

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
low_temp
rain
EasterSundayHoliday
abnormal_rain
University_holidays
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using PassiveAggressiveRegressor
Weekday
Day
Week_no
lnCountTotal_lag1
lnCountTotal_lag7
Month
lnCountTotal_lag4
lnCountTotal_lag6
abnormal_rain
UKBankHoliday
lnCountTotal_lag3
EasterSundayHoliday
wind_speed
University_holidays
high_temp
School_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:02 PM
11/10/2020

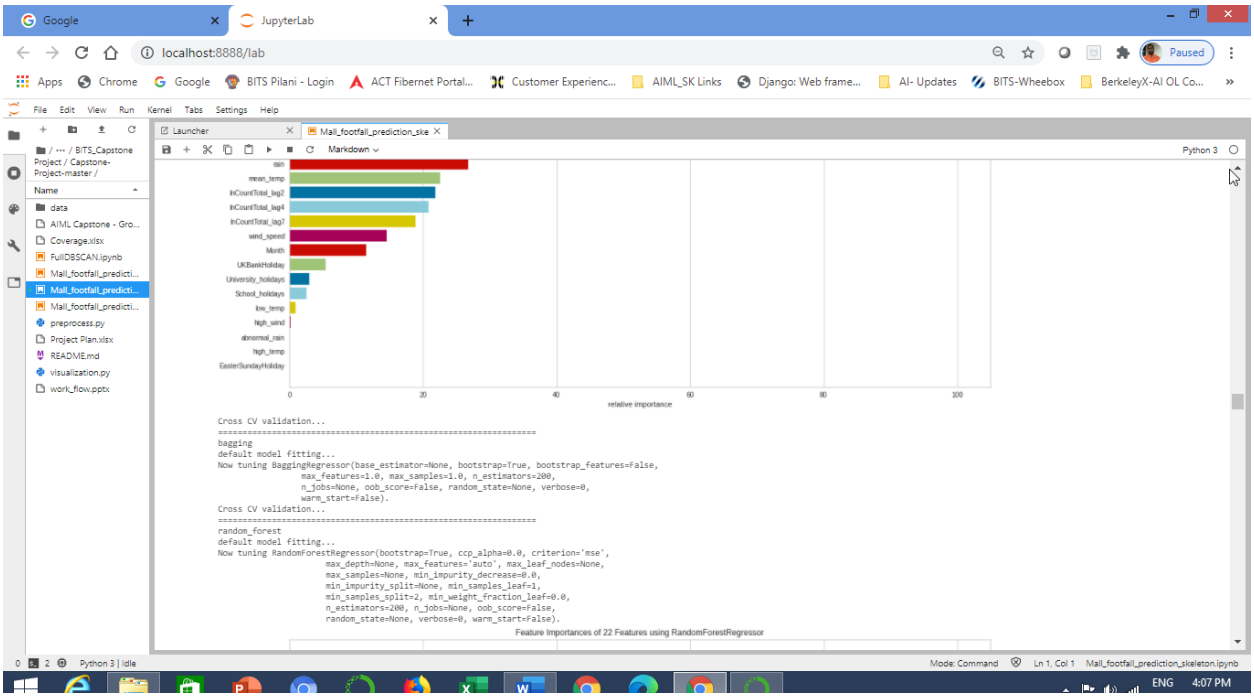

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
University_holidays
high_temp
School_holidays
mean_temp
high_wind
low_temp
lnCountTotal_lag2
lnCountTotal_lag5
rain
relative coefficient magnitude
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 22 Features using SGDRegressor
Weekday
lnCountTotal_lag7
Week_no
mean_temp
Day
lnCountTotal_lag1
Month
abnormal_rain
high_temp
School_holidays
high_wind
UKBankHoliday
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:03 PM
11/10/2020

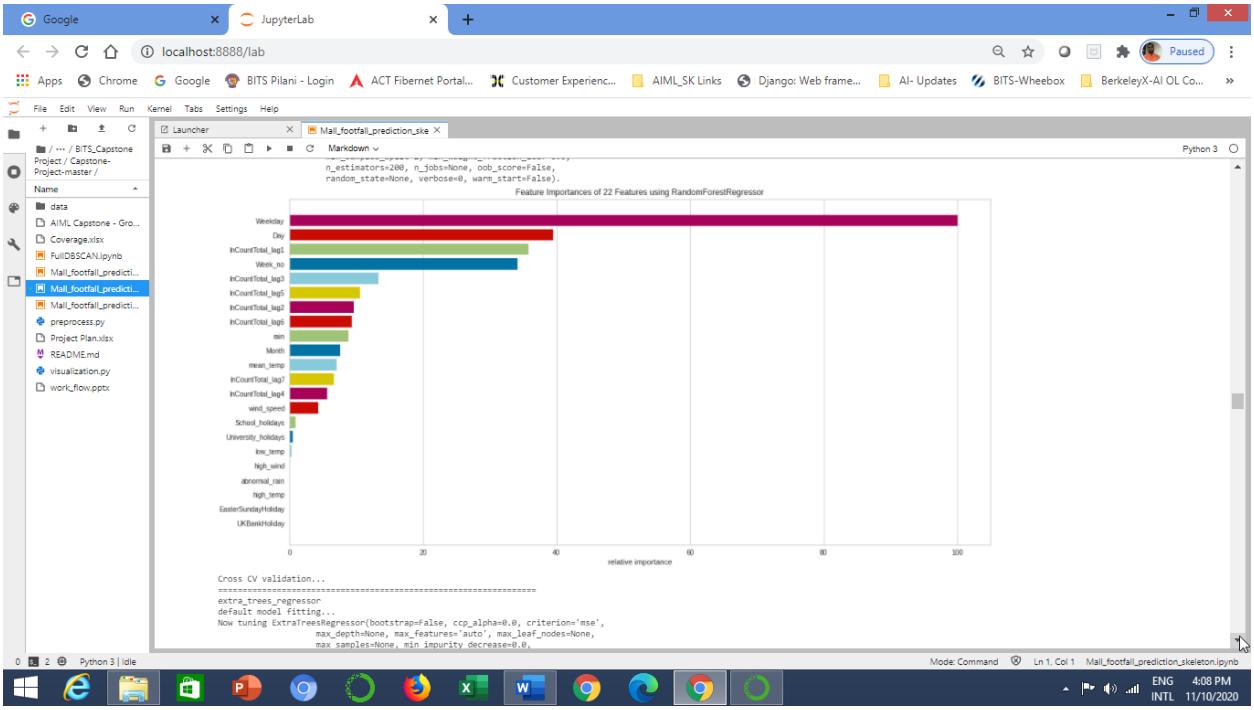

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
EasterSundayHoliday
University_holidays
low_temp
InCountTotal_lag4
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag2
InCountTotal_lag3
wind_speed
rain
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 22 Features using DecisionTreeRegressor
Weekday
School_holidays
UKBankHoliday
InCountTotal_lag6
InCountTotal_lag5
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:04 PM
11/10/2020

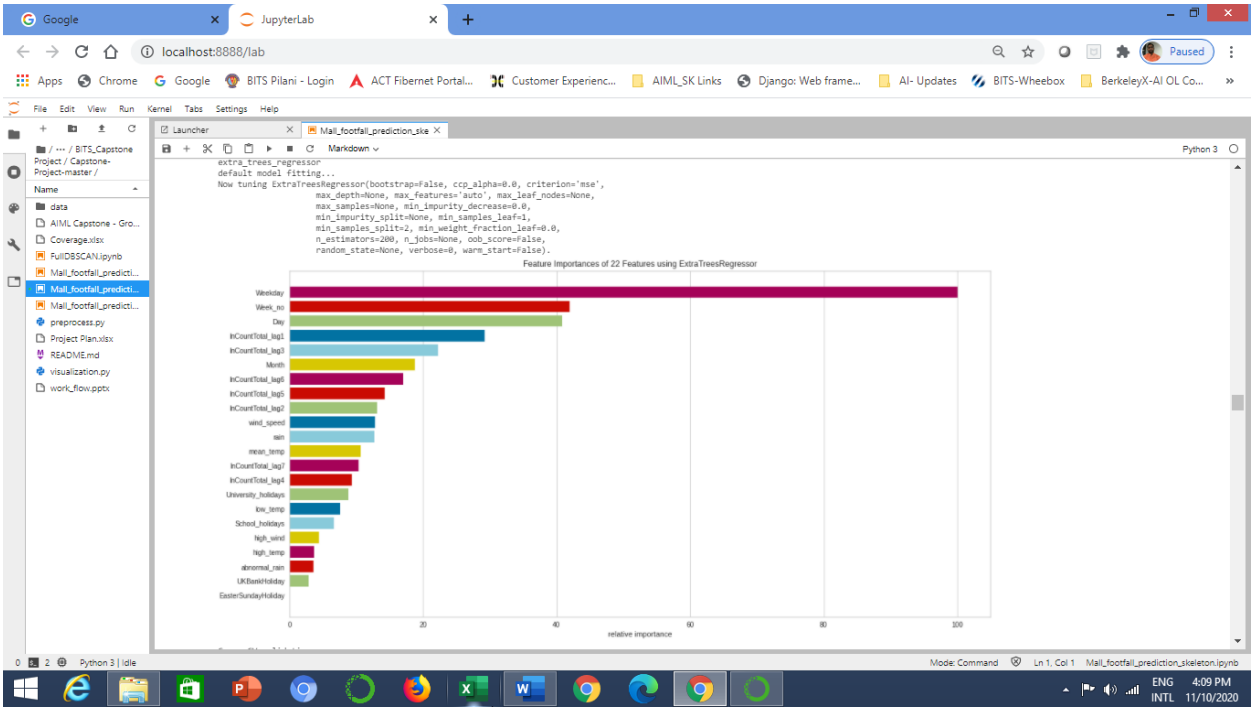

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
random_state=None, splitter='best').
Feature Importances of 22 Features using DecisionTreeRegressor
Weekday
School_holidays
UKBankHoliday
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag1
Month
Day
Week_no
mean_temp
rain
wind_speed
abnormal_rain
high_temp
low_temp
high_wind
EasterSundayHoliday
University_holidays
InCountTotal_lag7
relative importance
Cross CV validation...
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:05 PM
11/10/2020

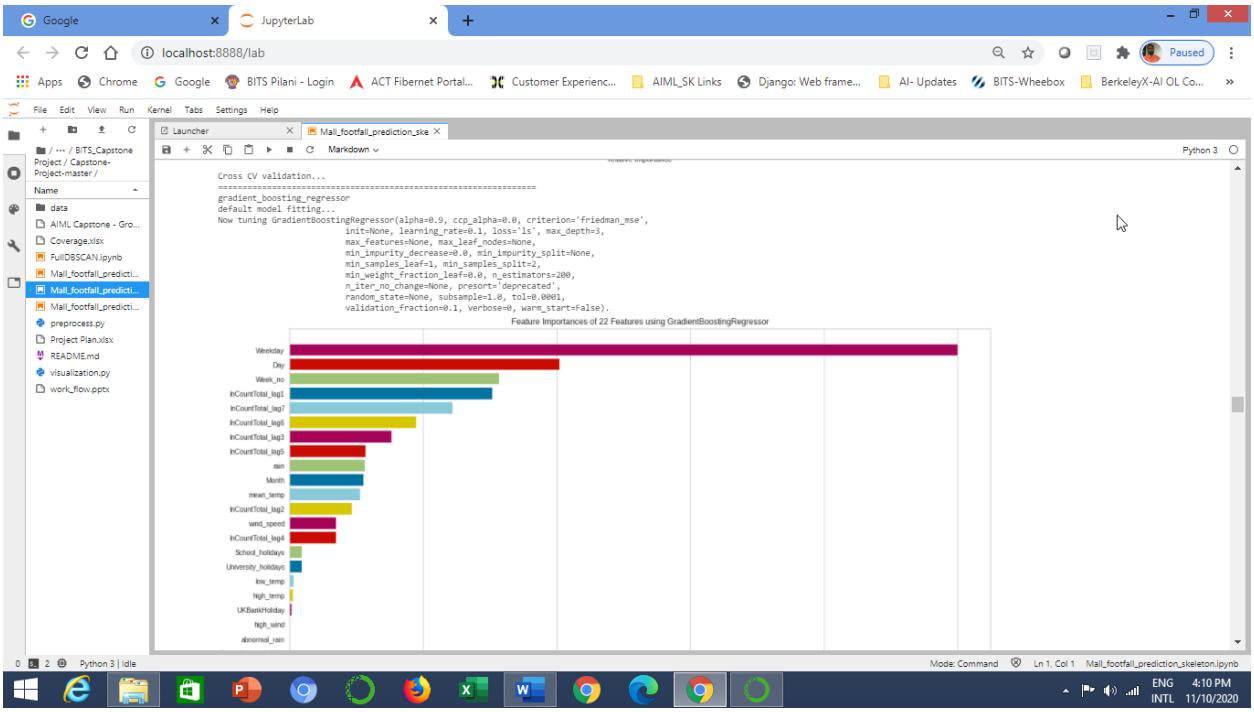

default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').
Feature Importances of 22 Features using ExtraTreeRegressor
Weekday
Month
Day
Week_no
School_holidays
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag4
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag1
UKBankHoliday
mean_temp
rain
wind_speed
abnormal_rain
high_temp
low_temp
high_wind
EasterSundayHoliday
University_holidays
InCountTotal_lag7
0
20
40
60
80
100
relative importance
Cross CV validation...

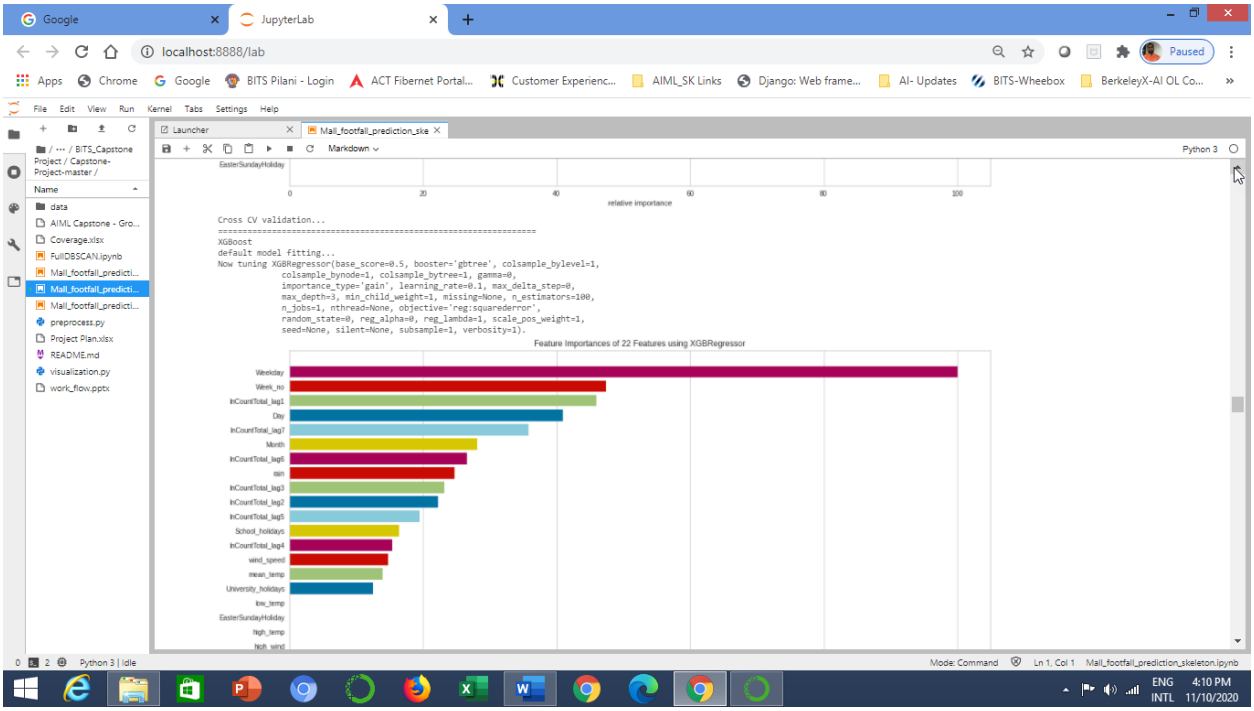

svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 22 Features using AdaBoostRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
rain
mean_temp
InCountTotal_lag2
InCountTotal_lag4
InCountTotal_lag7
wind_speed
Month
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
abnormal_rain
high_temp
EasterSundayHoliday

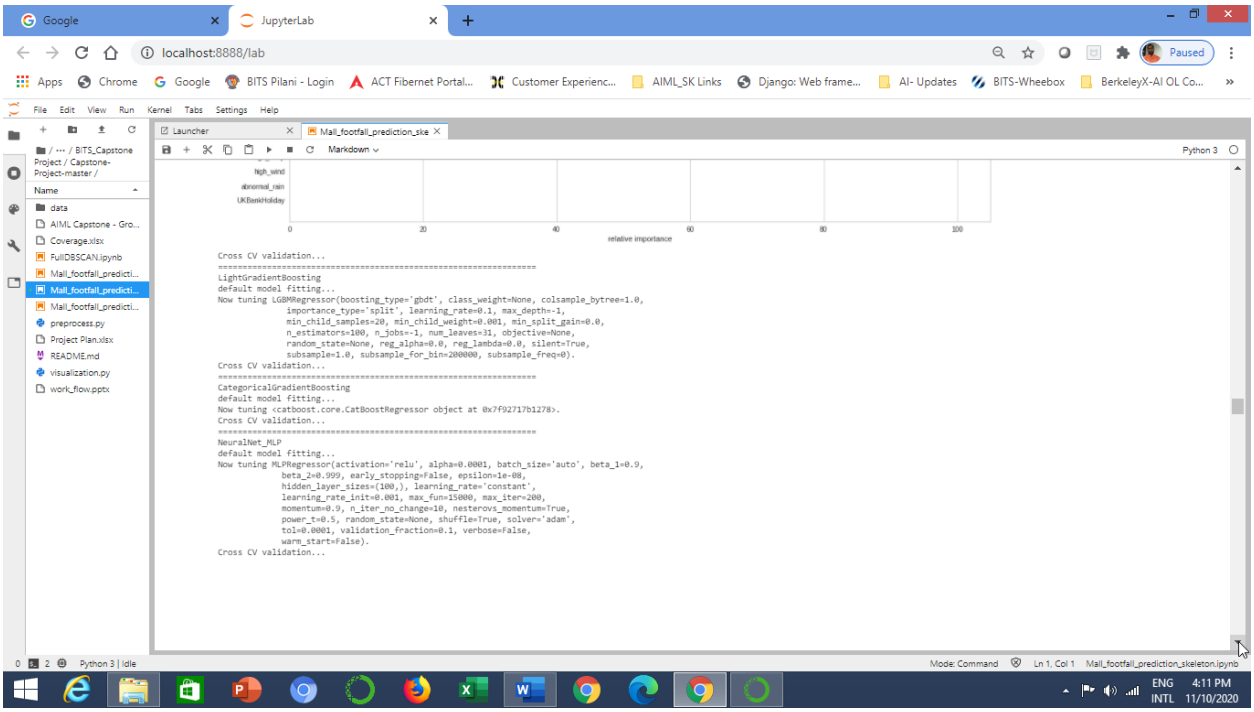

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
rain
mean_temp
lnCountTotal_lag2
lnCountTotal_lag4
lnCountTotal_lag7
wind_speed
Month
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
abnormal_rain
high_temp
EasterSundayHoliday
relative importance
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
4:07 PM


Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using RandomForestRegressor
Weekday
Day
lnCountTotal_lag1
Week_no
lnCountTotal_lag3
lnCountTotal_lag5
lnCountTotal_lag2
lnCountTotal_lag6
rain
Month
mean_temp
lnCountTotal_lag7
lnCountTotal_lag4
wind_speed
School_holidays
University_holidays
low_temp
high_wind
abnormal_rain
high_temp
EasterSundayHoliday
UKBankHoliday
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:08 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 22 Features using ExtraTreesRegressor
Weekday
Week_no
Day
InCountTotal_lag1
InCountTotal_lag3
Month
InCountTotal_lag6
InCountTotal_lag5
InCountTotal_lag2
wind_speed
rain
mean_temp
InCountTotal_lag7
InCountTotal_lag4
University_holidays
low_temp
School_holidays
high_wind
high_temp
abnormal_rain
UKBankHoliday
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:09 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 22 Features using GradientBoostingRegressor
Weekday
Day
Week_no
InCountTotal_lag1
InCountTotal_lag7
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag5
rain
Month
mean_temp
InCountTotal_lag2
wind_speed
InCountTotal_lag4
School_holidays
University_holidays
low_temp
high_temp
UKBankHoliday
high_wind
abnormal_rain
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:10 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 22 Features using XGBRegressor
Weekday
Week_no
InCountTotal_lag1
Day
InCountTotal_lag7
Month
InCountTotal_lag6
min
InCountTotal_lag3
InCountTotal_lag2
InCountTotal_lag5
School_holidays
InCountTotal_lag4
wind_speed
mean_temp
University_holidays
low_temp
EasterSundayHoliday
high_temp
high_wind
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:10 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
high_wind
abnormal_rain
UKBankHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f92717b1278>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
learning_rate_init=0.001, max_fun=15000, max_iter=200,
momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
power_t=0.5, random_state=None, shuffle=True, solver='adam',
tol=0.0001, validation_fraction=0.1, verbose=False,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:11 PM
11/10/2020

Google | JupyterLab | localhost:8888/lab

```
loss_df_outlier_removed.sort_values('mape').head(25)
```

[37]:

	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
39	random_forest_Outliers_Removed_shuffled_defaul...	11.697544	0.004136	-0.013269	0.044251	-0.009544	0.064315	0.481009	0.445163	-1913.193796	-1824.005873
36	bagging_Outliers_Removed_shuffled_default_model	11.759516	0.004164	-0.013640	0.044665	-0.010641	0.064529	0.477538	0.441452	-1910.814239	-1821.626316
45	gradient_boosting_regressor_Outliers_Removed_s...	12.213195	0.004682	-0.017509	0.047274	-0.021555	0.068422	0.412601	0.372030	-1868.991295	-1779.803372
52	LightGradientBoosting_Outliers_Removed_shuffle...	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
40	random_forest_Outliers_Removed_shuffled_rscv_m...	12.332504	0.004447	-0.013993	0.046429	-0.008571	0.066687	0.442008	0.403468	-1887.326676	-1798.138753
54	CategoricalGradientBoosting_Outliers_Removed_s...	12.358471	0.004391	-0.013838	0.046761	-0.010404	0.066263	0.449089	0.411038	-1891.885690	-1802.697767
55	CategoricalGradientBoosting_Outliers_Removed_s...	12.382344	0.004613	-0.016361	0.047490	-0.016135	0.067916	0.421260	0.381287	-1874.292851	-1785.104928
43	extra_trees_regressor_Outliers_Removed_shuffle...	12.415550	0.004390	-0.015855	0.047120	-0.016406	0.066259	0.449149	0.411103	-1891.924883	-1802.736960
42	extra_trees_regressor_Outliers_Removed_shuffle...	12.455072	0.004412	-0.015203	0.047198	-0.014596	0.066420	0.446477	0.408246	-1890.197171	-1801.009248
48	XGBoost_Outliers_Removed_shuffled_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708
46	gradient_boosting_regressor_Outliers_Removed_s...	12.632322	0.004775	-0.016352	0.048288	-0.014328	0.069103	0.400854	0.359472	-1861.922079	-1772.734156
44	extra_trees_regressor_Outliers_Removed_shuffle...	12.641993	0.004270	0.001854	0.044996	3.583380	0.065291	0.425421	NaN	NaN	NaN
51	LightGradientBoosting_Outliers_Removed_shuffle...	12.709911	0.004599	-0.014983	0.048127	-0.014091	0.067815	0.422976	0.383121	-1875.352644	-1786.164721
37	bagging_Outliers_Removed_shuffled_rscv_model	13.049147	0.004863	-0.016209	0.049521	-0.008672	0.069732	0.389888	0.347748	-1855.446934	-1766.259011
56	CategoricalGradientBoosting_Outliers_Removed_s...	13.132737	0.004550	-0.000036	0.047296	3.027100	0.067242	0.385288	NaN	NaN	NaN
41	random_forest_Outliers_Removed_shuffled_crossc...	13.233996	0.004534	0.001688	0.046830	3.916720	0.067086	0.398068	NaN	NaN	NaN
34	adaBoost_Outliers_Removed_shuffled_rscv_model	13.259735	0.004769	-0.010869	0.049260	0.003878	0.069060	0.401605	0.360274	-1862.369705	-1773.181782
49	XGBoost_Outliers_Removed_shuffled_rscv_model	13.259941	0.004772	-0.010815	0.049262	0.006114	0.069082	0.401215	0.359857	-1862.137055	-1772.949132
47	gradient_boosting_regressor_Outliers_Removed_s...	13.445203	0.004540	0.001265	0.047983	3.554040	0.067213	0.393039	NaN	NaN	NaN
35	adaBoost_Outliers_Removed_shuffled_crossrcv_model	13.917584	0.004873	0.000281	0.049795	3.822424	0.069452	0.361506	NaN	NaN	NaN
12	huberRegressor_Outliers_Removed_shuffled_defau...	13.974179	0.005750	-0.018115	0.052593	-0.016486	0.075828	0.278562	0.228733	-1795.612475	-1706.424552
13	huberRegressor_Outliers_Removed_shuffled_rscv_...	14.023577	0.005841	-0.019958	0.053099	-0.021893	0.076425	0.267163	0.216546	-1790.015758	-1700.827835
38	bagging_Outliers_Removed_shuffled_crossrcv_model	14.217424	0.004912	0.000906	0.050443	4.198927	0.069815	0.356890	NaN	NaN	NaN
10	elastic_net_Outliers_Removed_shuffled_rscv_model	14.331859	0.005650	-0.018112	0.053930	-0.014345	0.075165	0.291117	0.242155	-1801.880079	-1712.692156

Google | JupyterLab | localhost:8888/lab

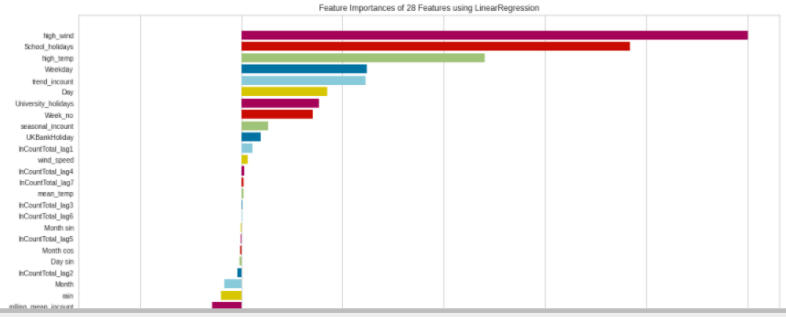
10	elastic_net_Outliers_Removed_shuffled_rscv_model	14.331859	0.005650	-0.018112	0.053930	-0.014345	0.075165	0.291117	0.242155	-1801.880079	-1712.692156
4	lassoRegressor_Outliers_Removed_shuffled_rscv_...	14.334564	0.005657	-0.018216	0.053958	-0.014579	0.075212	0.290229	0.241205	-1801.432826	-1712.244903

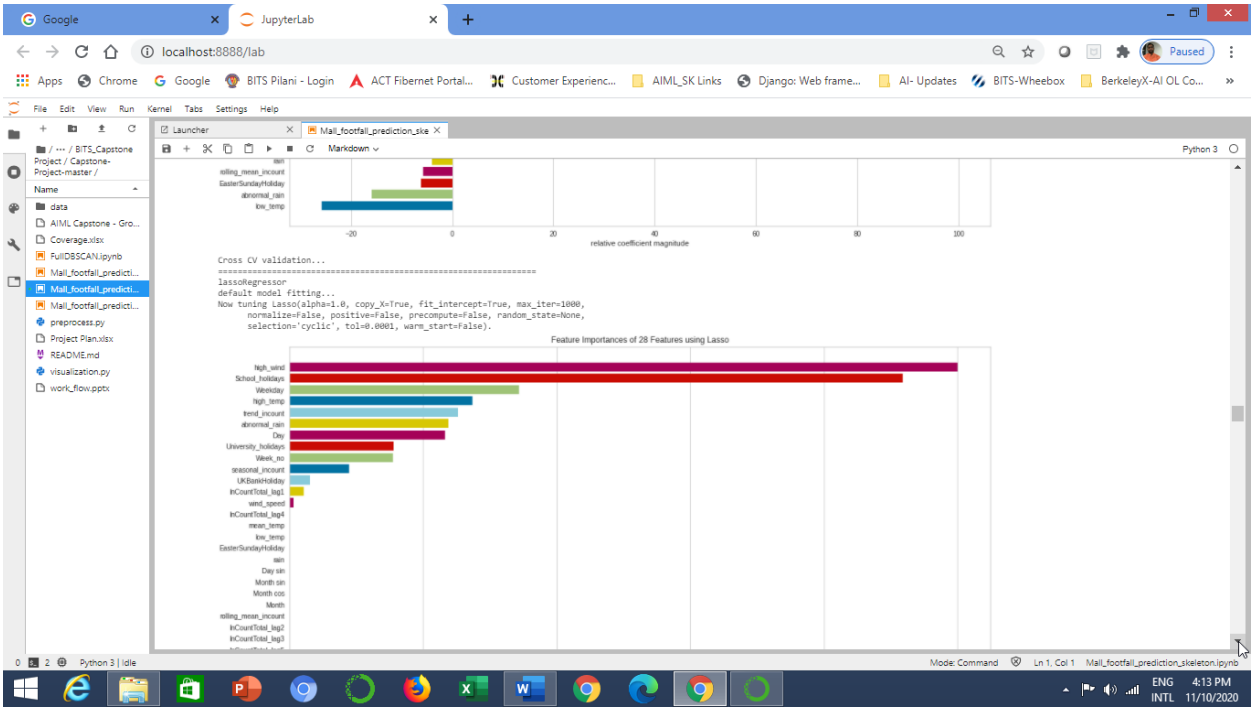
7.1.4 Model training with extra features

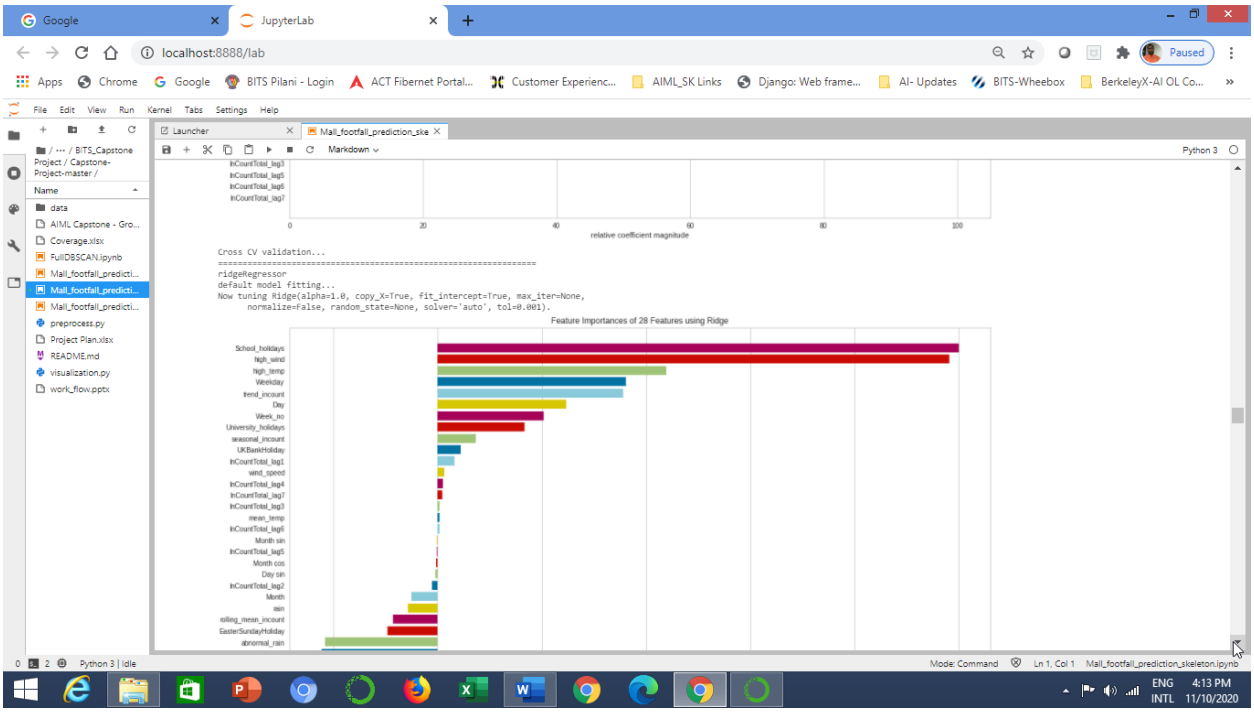
```
reg_models = regression_model(data_outliersremoved,dataset_type='OutliersRemoved_extra_features_shuffled')
_,_,_, _, loss_normal_df = reg_models.train_and_evaluate_models(RFEcross_validation=False,shuffle=True, extra_input_features=True)
loss_values_all_models.append(loss_normal_df)
```

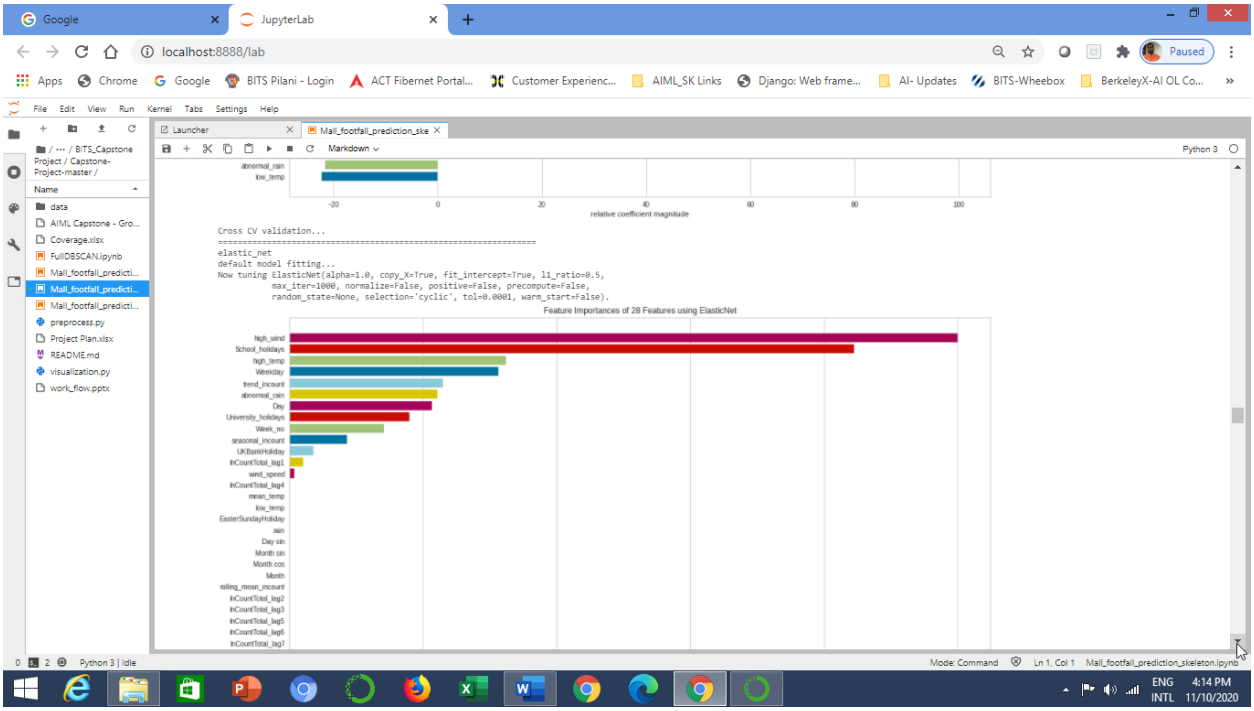
[38]:

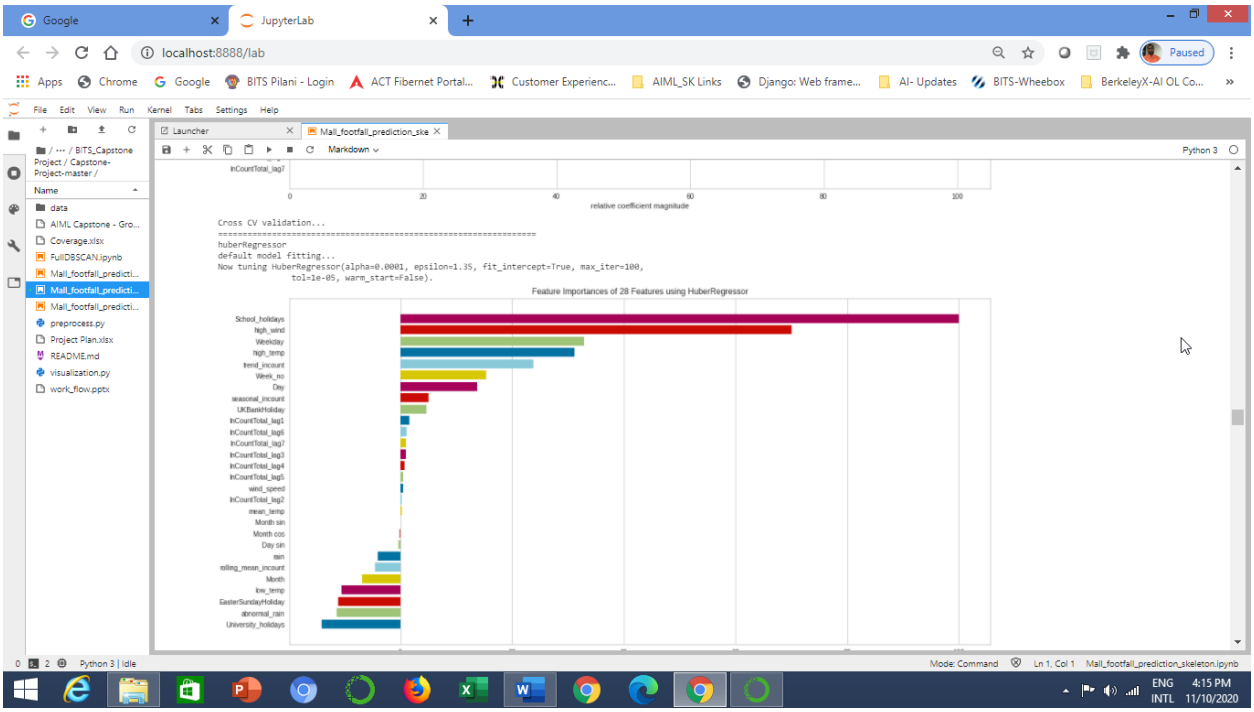
```
Defined 20 models
================================================================
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
```

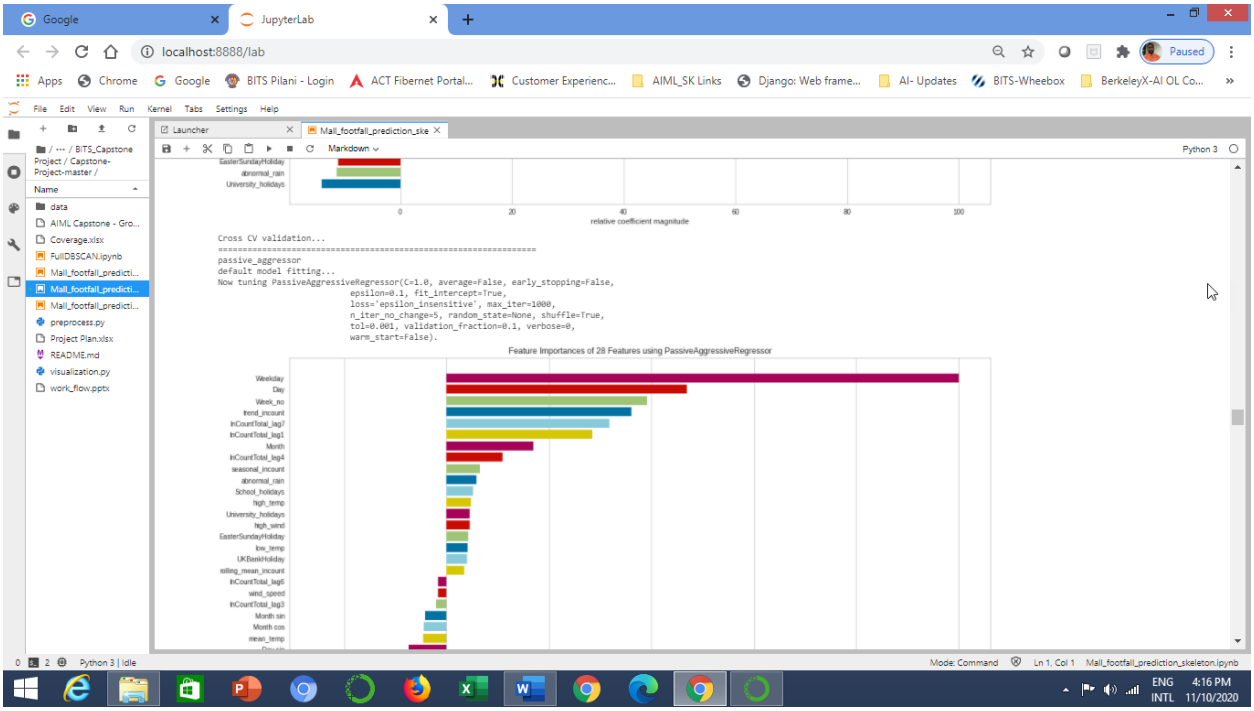


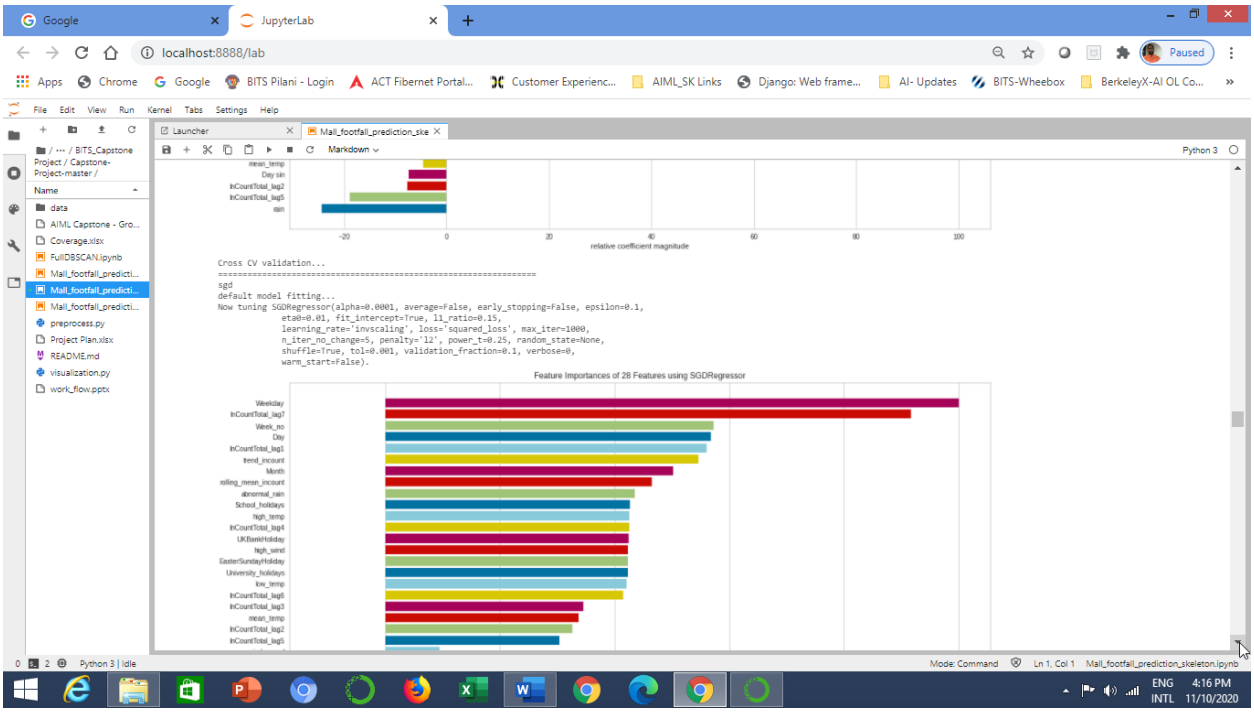

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
rolling_mean_incount
EasterSundayHoliday
abnormal_rain
low_temp
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 28 Features using Lasso
high_wind
School_holidays
Weekday
high_temp
trend_incount
abnormal_rain
Day
University_holidays
Week_no
seasonal_incount
UKBankHolidays
InCountTotal_lag1
wind_speed
InCountTotal_lag4
mean_temp
low_temp
EasterSundayHoliday
rain
Day sin
Month sin
Month cos
Month
rolling_mean_incount
InCountTotal_lag2
InCountTotal_lag3
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:13 PM
11/10/2020

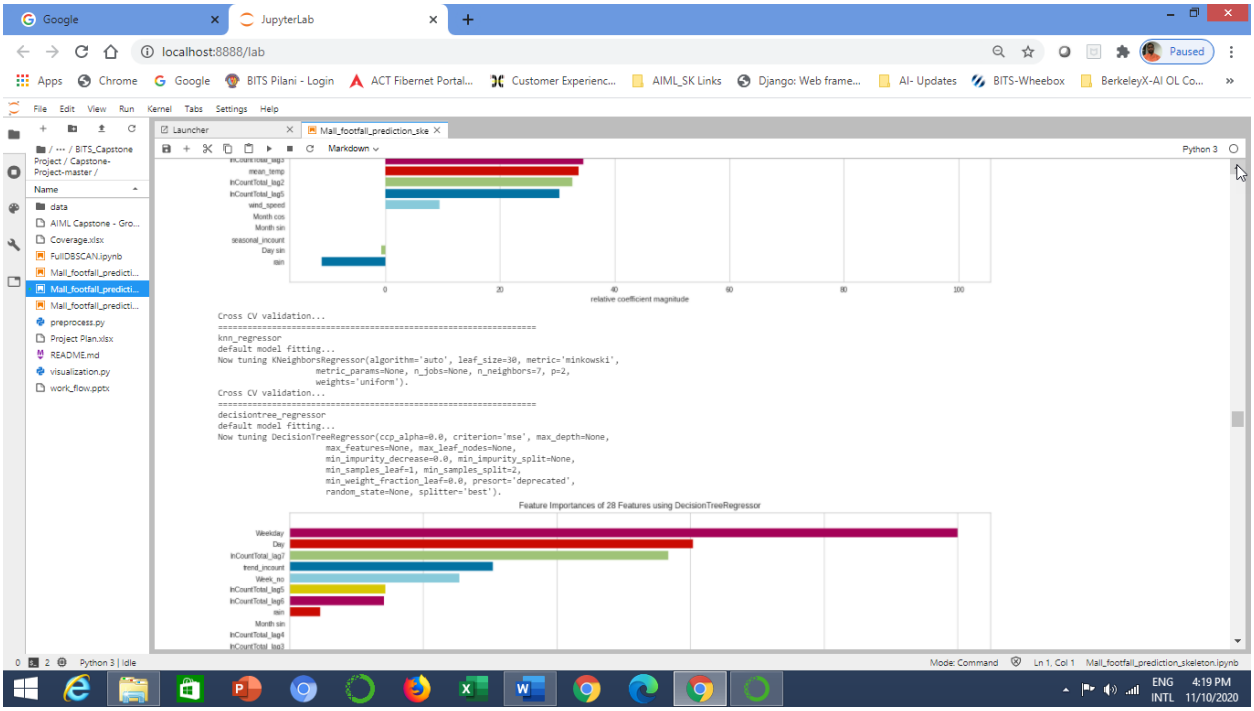

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag6
InCountTotal_lag7
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 28 Features using Ridge
School_holidays
high_wind
high_temp
Weekday
trend_incount
Day
Week_no
University_holidays
seasonal_incount
UKBankHolidays
InCountTotal_lag1
wind_speed
InCountTotal_lag4
InCountTotal_lag7
InCountTotal_lag3
mean_temp
InCountTotal_lag6
Month sin
InCountTotal_lag5
Month cos
Day sin
InCountTotal_lag2
Month
rain
rolling_mean_incount
EasterSundayHoliday
abnormal_rain
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:13 PM
11/10/2020

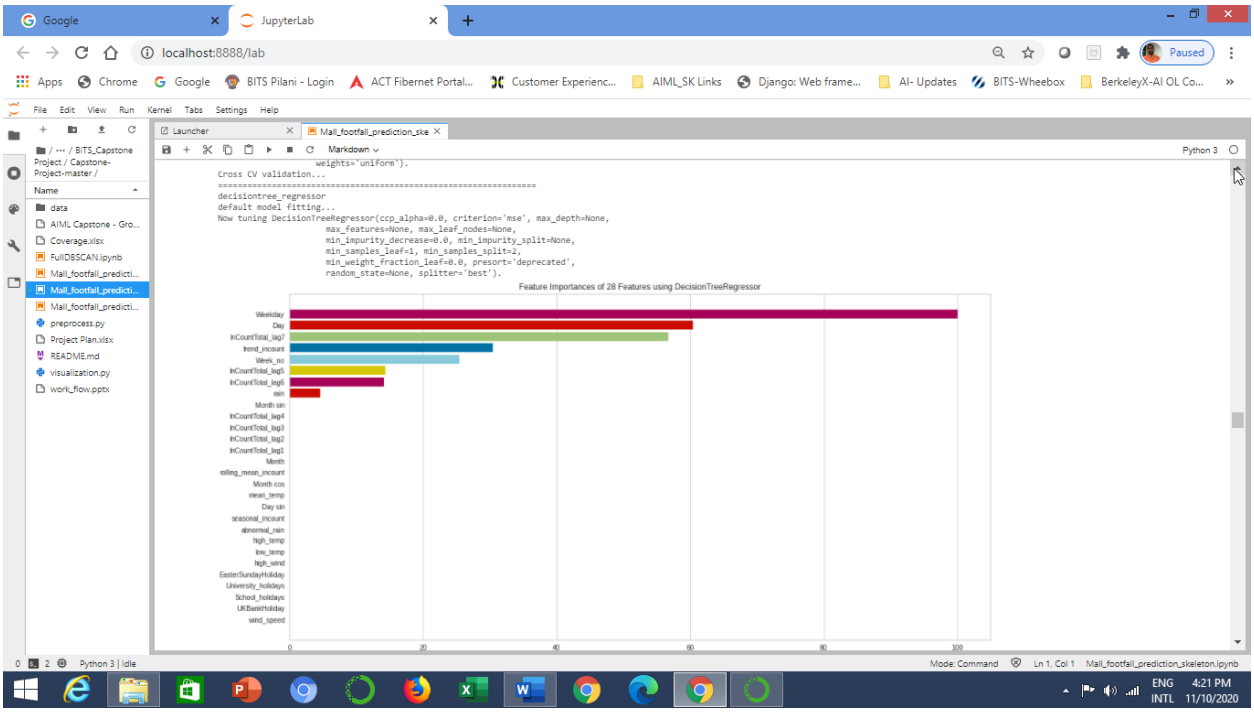

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 28 Features using ElasticNet
high_wind
School_holidays
high_temp
Weekday
trend_incount
abnormal_rain
Day
University_holidays
Week_no
seasonal_incount
UKBankHoliday
InCountTotal_lag1
wind_speed
InCountTotal_lag4
mean_temp
low_temp
EasterSundayHoliday
rain
Day sin
Month sin
Month cos
Month
rolling_mean_incount
InCountTotal_lag2
InCountTotal_lag3
InCountTotal_lag5
InCountTotal_lag6
InCountTotal_lag7
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:14 PM
11/10/2020

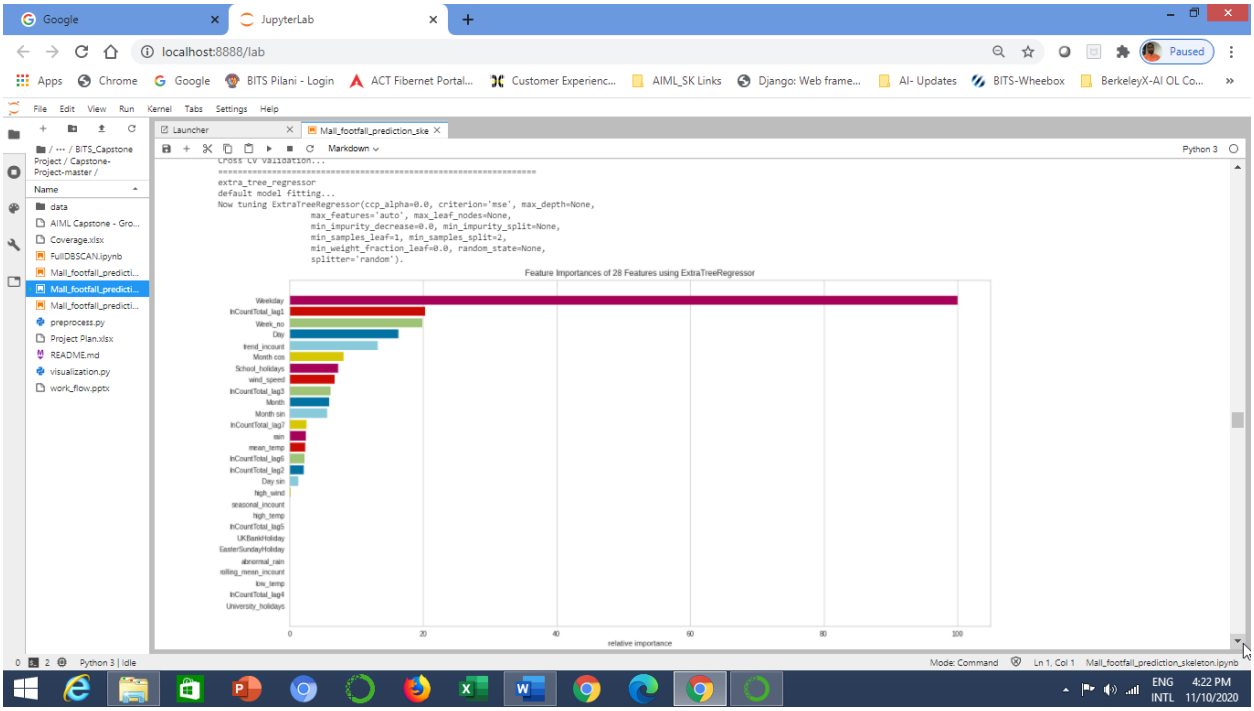

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
huberRegressor
default model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 28 Features using HuberRegressor
School_holidays
high_wind
Weekday
high_temp
trend_incount
Week_no
Day
seasonal_incount
UKBankHoliday
InCountTotal_lag1
InCountTotal_lag6
InCountTotal_lag7
InCountTotal_lag3
InCountTotal_lag4
InCountTotal_lag5
wind_speed
InCountTotal_lag2
mean_temp
Month sin
Month cos
Day sin
rain
rolling_mean_incount
Month
low_temp
EasterSundayHoliday
abnormal_rain
University_holidays
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:15 PM
11/10/2020

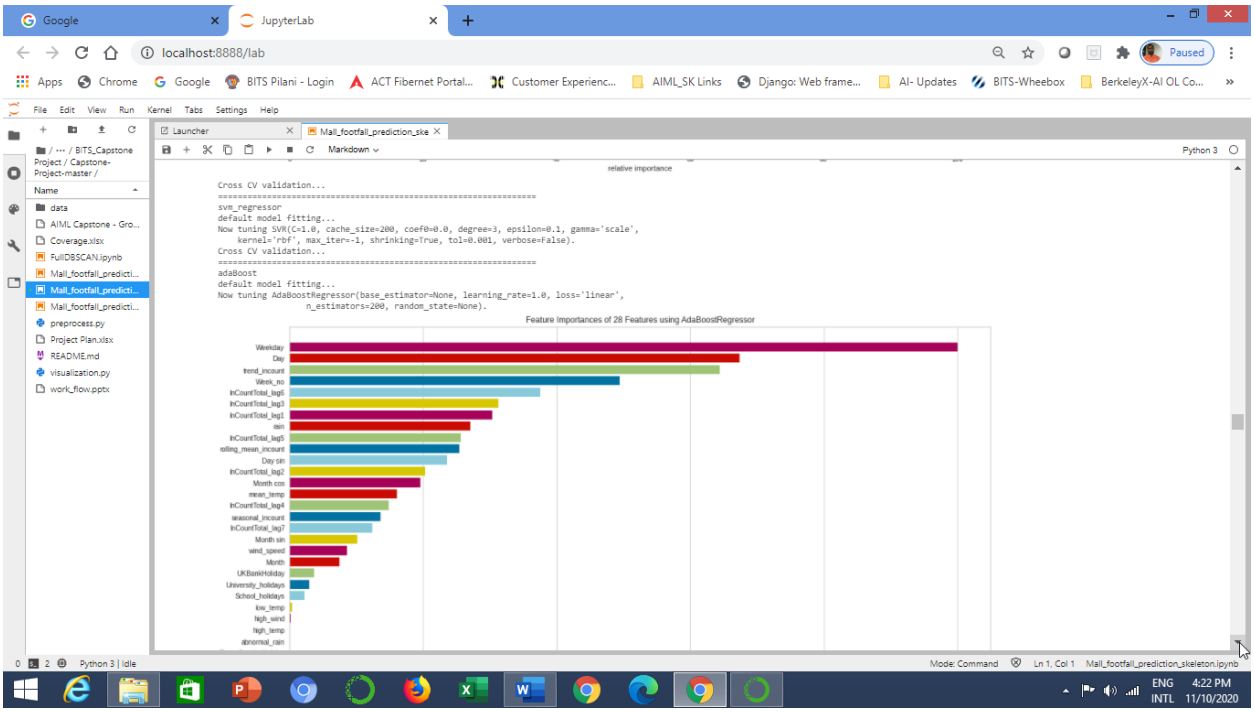

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 28 Features using PassiveAggressiveRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:16 PM
11/10/2020

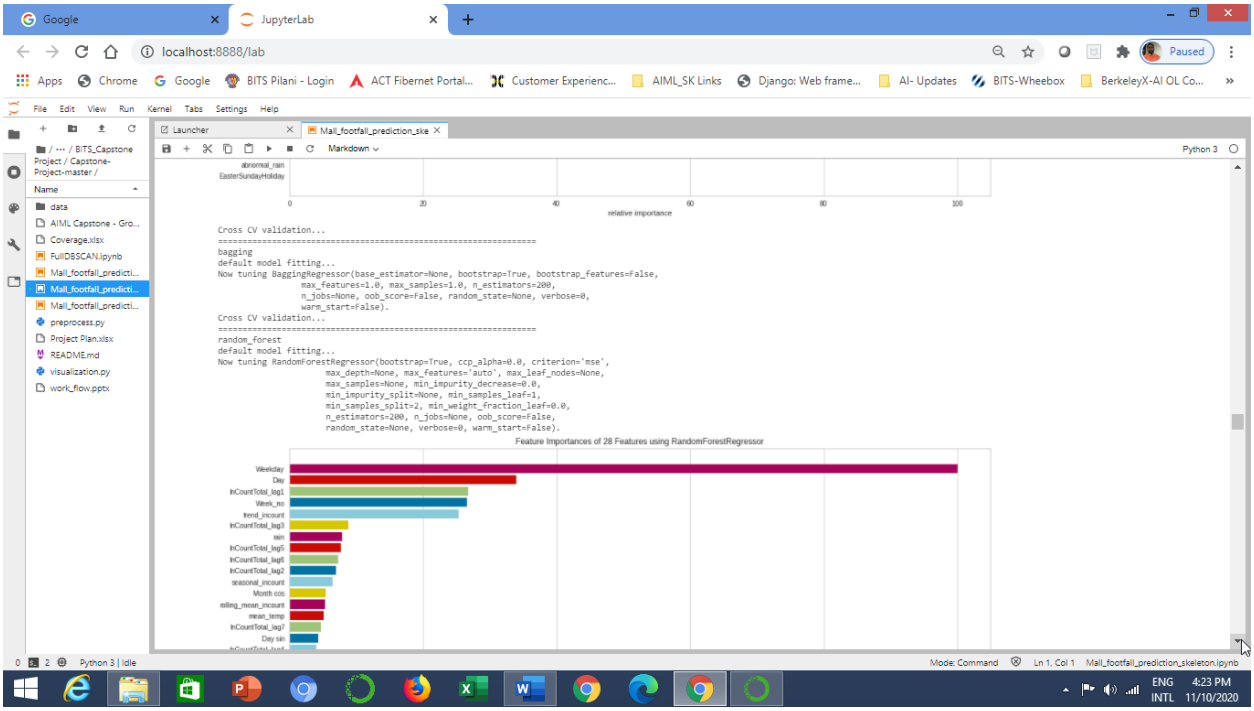

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative coefficient magnitude
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 28 Features using SGDRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:16 PM
11/10/2020

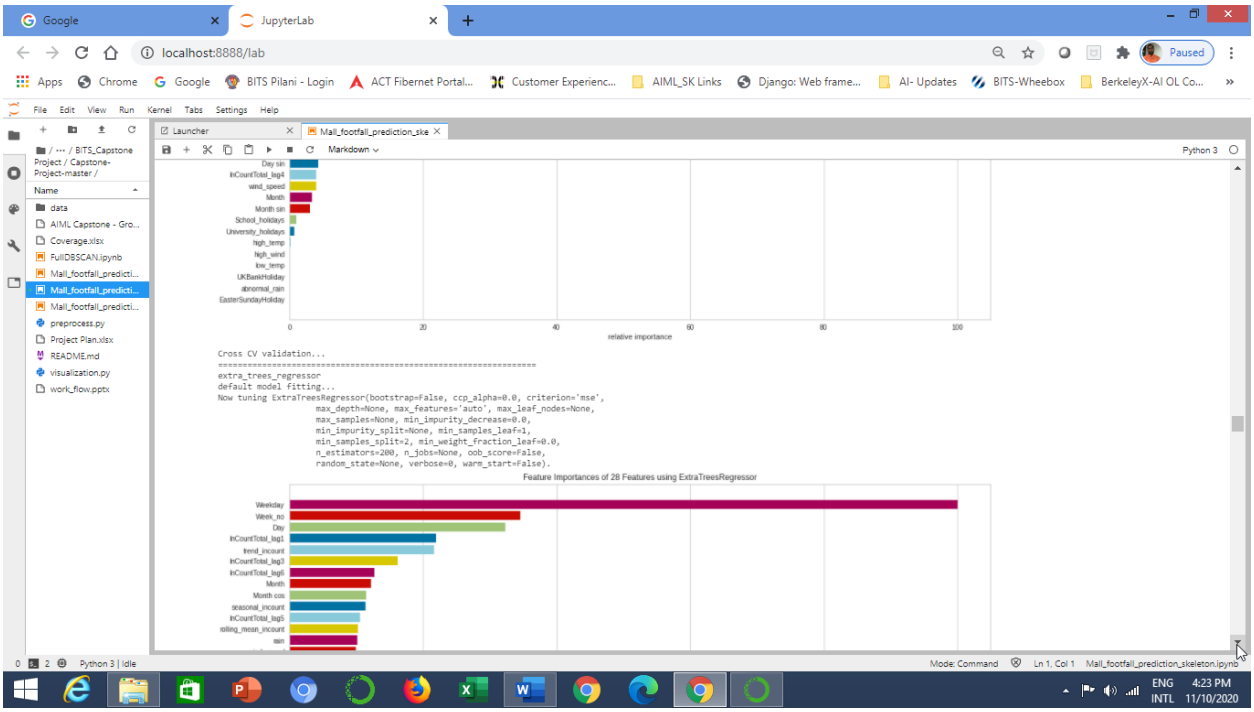

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 28 Features using DecisionTreeRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:19 PM
11/10/2020

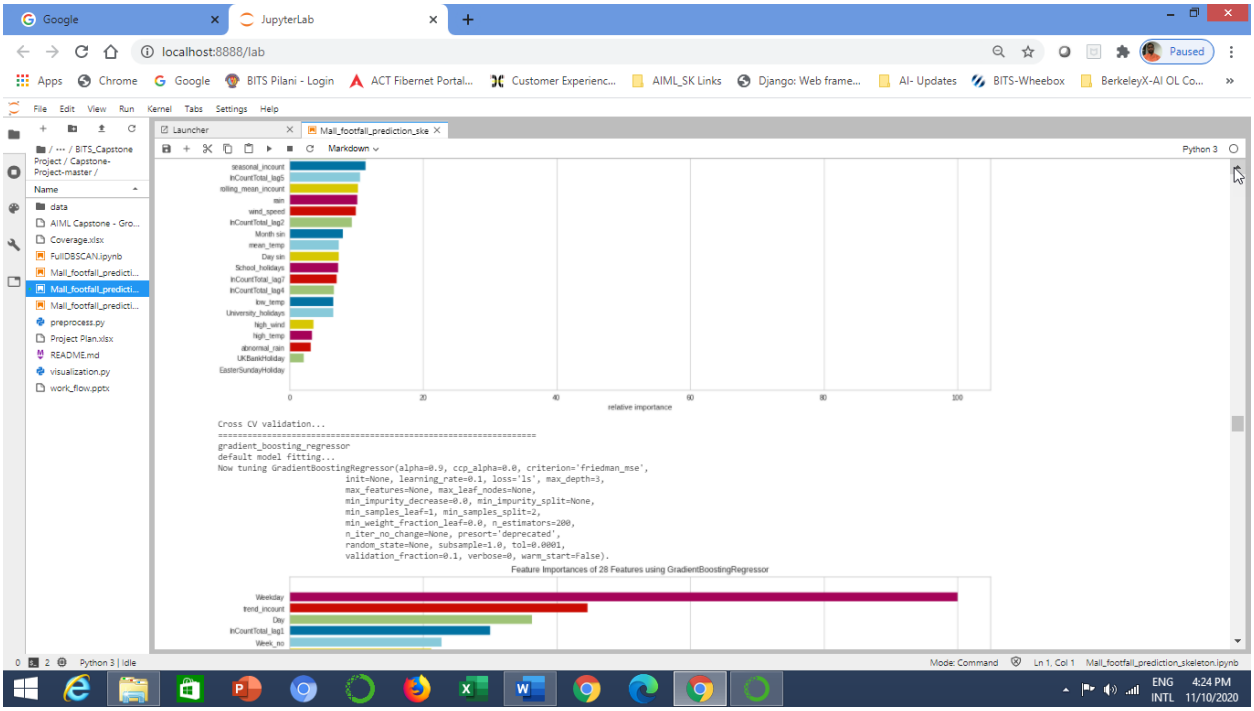

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 28 Features using DecisionTreeRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:21 PM
11/10/2020

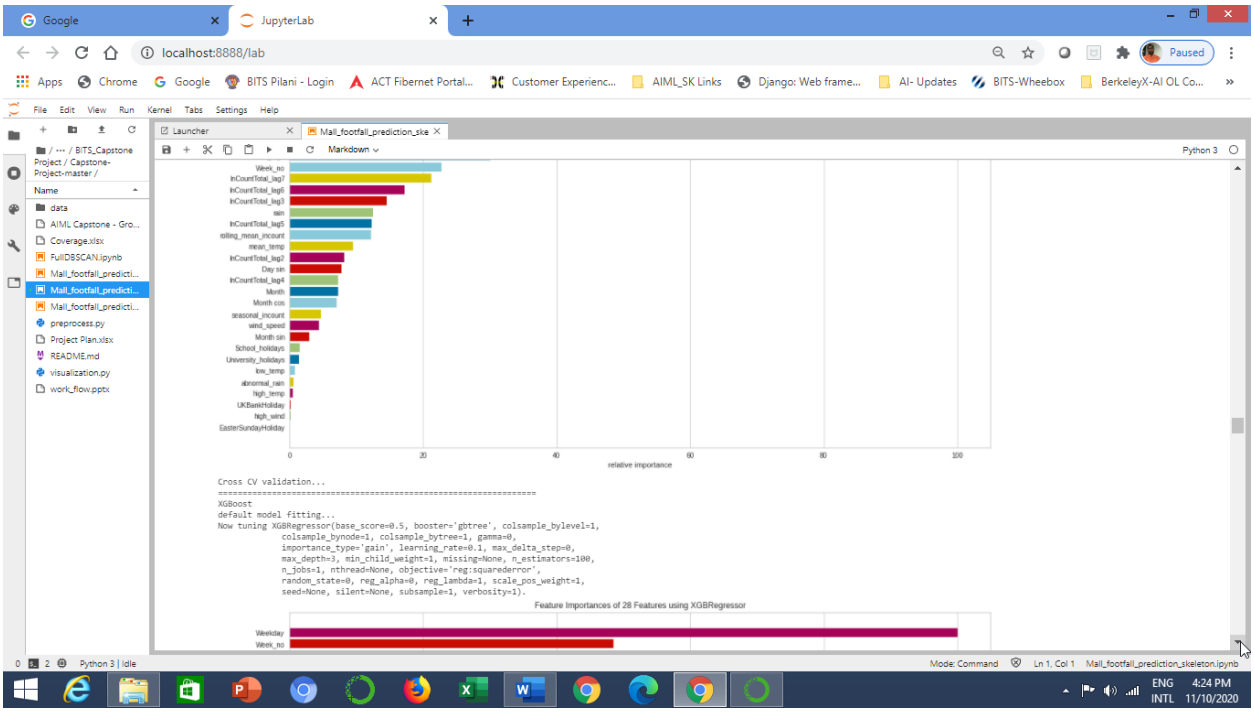

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Cross CV validation...
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').
Feature Importances of 28 Features using ExtraTreeRegressor
Weekday
InCountTotal_lag1
Week_no
Day
trend_incount
Month cos
School_holidays
wind_speed
InCountTotal_lag3
Month
Month sin
InCountTotal_lag7
min
mean_temp
InCountTotal_lag6
InCountTotal_lag2
Day sin
high_wind
seasonal_incount
high_temp
InCountTotal_lag5
UKBankHoliday
EasterSundayHoliday
abnormal_rain
rolling_mean_incount
low_temp
InCountTotal_lag4
University_holidays
0
20
40
60
80
100
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:22 PM
11/10/2020

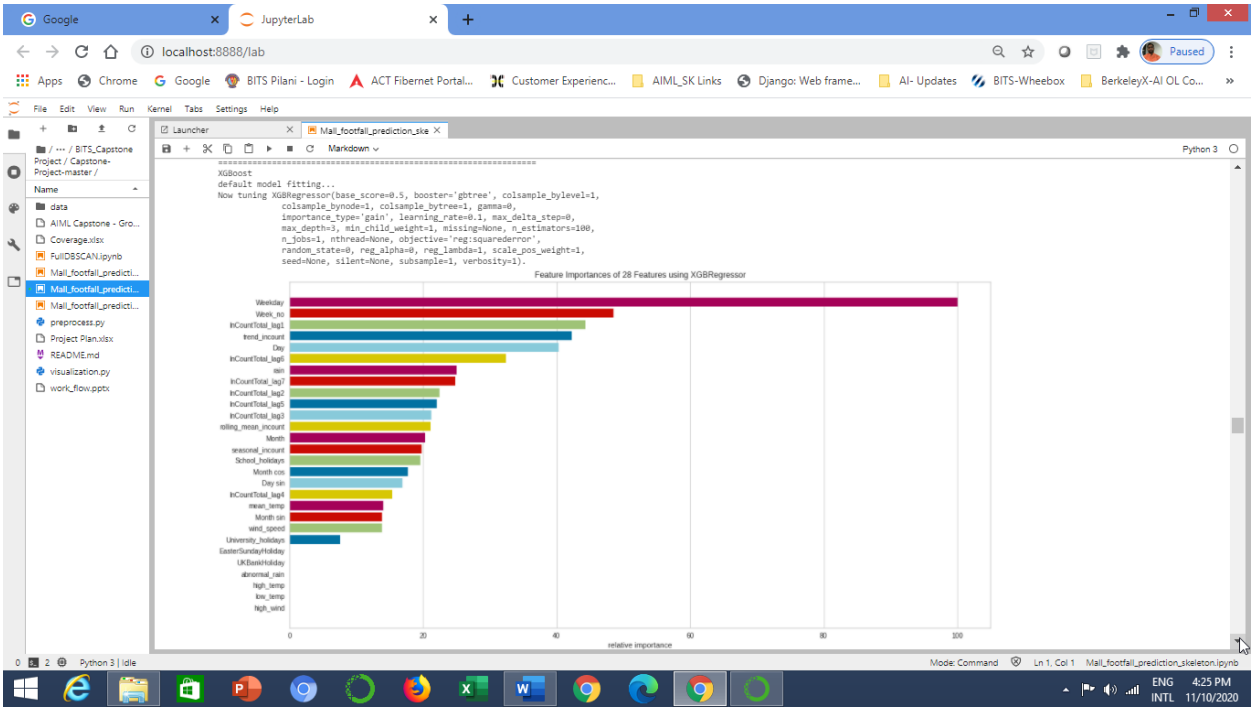

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 28 Features using AdaBoostRegressor
Weekday
Day
trend_incount
Week_no
InCountTotal_lag6
InCountTotal_lag3
InCountTotal_lag1
min
InCountTotal_lag5
rolling_mean_incount
Day sin
InCountTotal_lag2
Month cos
mean_temp
InCountTotal_lag4
seasonal_incount
InCountTotal_lag7
Month sin
wind_speed
Month
UKBankHoliday
University_holidays
School_holidays
low_temp
high_wind
high_temp
abnormal_rain
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:22 PM
11/10/2020

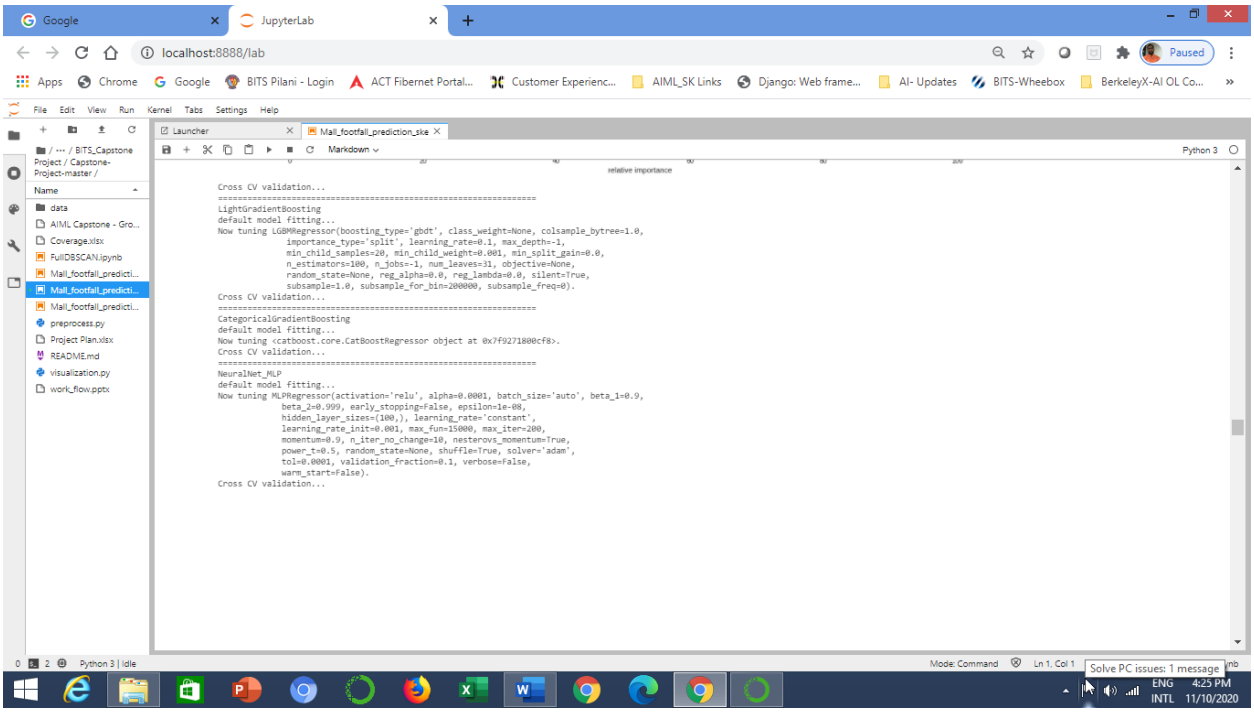

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 28 Features using RandomForestRegressor
Weekday
Day
InCountTotal_lag1
Week_no
trend_incount
InCountTotal_lag3
min
InCountTotal_lag5
InCountTotal_lag6
InCountTotal_lag2
seasonal_incount
Month cos
rolling_mean_incount
mean_temp
InCountTotal_lag7
Day sin
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:23 PM
11/10/2020

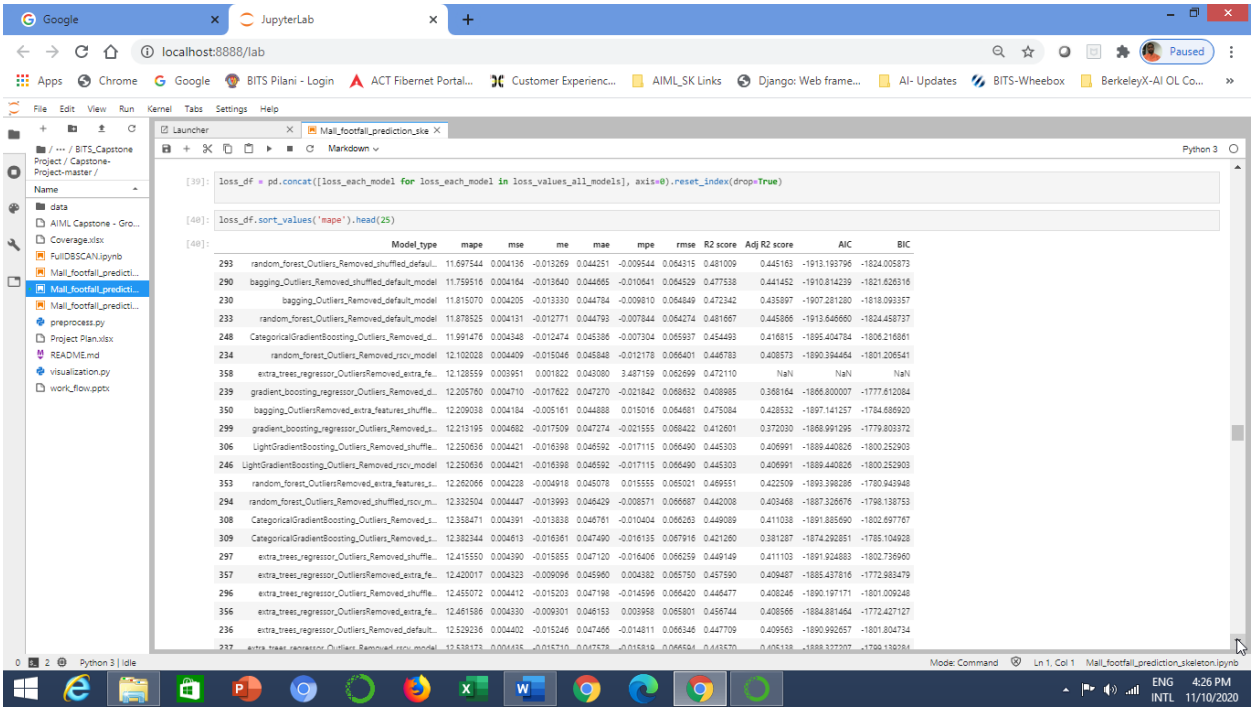

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 28 Features using ExtraTreesRegressor
Weekday
Week_no
Day
InCountTotal_lag1
trend_incount
InCountTotal_lag3
InCountTotal_lag6
Month
Month cos
seasonal_incount
InCountTotal_lag5
rolling_mean_incount
min
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:23 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
seasonal_incount
lnCountTotal_lag5
rolling_mean_incount
min
wind_speed
lnCountTotal_lag2
Month sin
mean_temp
Day sin
School_holidays
lnCountTotal_lag7
lnCountTotal_lag4
low_temp
University_holidays
high_wind
high_temp
abnormal_rain
UKBankHoliday
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 28 Features using GradientBoostingRegressor
Weekday
trend_incount
Day
lnCountTotal_lag1
Week_no
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
4:24 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Week_no
lnCountTotal_lag7
lnCountTotal_lag6
lnCountTotal_lag3
min
lnCountTotal_lag5
rolling_mean_incount
mean_temp
lnCountTotal_lag2
Day sin
lnCountTotal_lag4
Month
Month cos
seasonal_incount
wind_speed
Month sin
School_holidays
University_holidays
low_temp
abnormal_rain
high_temp
UKBankHoliday
high_wind
EasterSundayHoliday
0
20
40
60
80
100
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 28 Features using XGBRegressor
Weekday
Week_no
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
4:24 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 28 Features using XGBRegressor
Weekday
Week_no
InCountTotal_lag1
trend_incount
Day
InCountTotal_lag6
ssn
InCountTotal_lag7
InCountTotal_lag2
InCountTotal_lag5
InCountTotal_lag3
rolling_mean_incount
Month
seasonal_incount
School_holidays
Month cos
Day sin
InCountTotal_lag4
mean_temp
Month sin
wind_speed
University_holidays
EasterSundayHolday
UKBankHolday
abnormal_rain
high_temp
low_temp
high_wind
0
20
40
60
80
100
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:25 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f9271800cf8>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
learning_rate_init=0.001, max_fun=15000, max_iter=200,
momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
power_t=0.5, random_state=None, shuffle=True, solver='adam',
tol=0.0001, validation_fraction=0.1, verbose=False,
warm_start=False).
Cross CV validation...
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Solve PC issues: 1 message
ENG
INTL
4:25 PM
11/10/2020

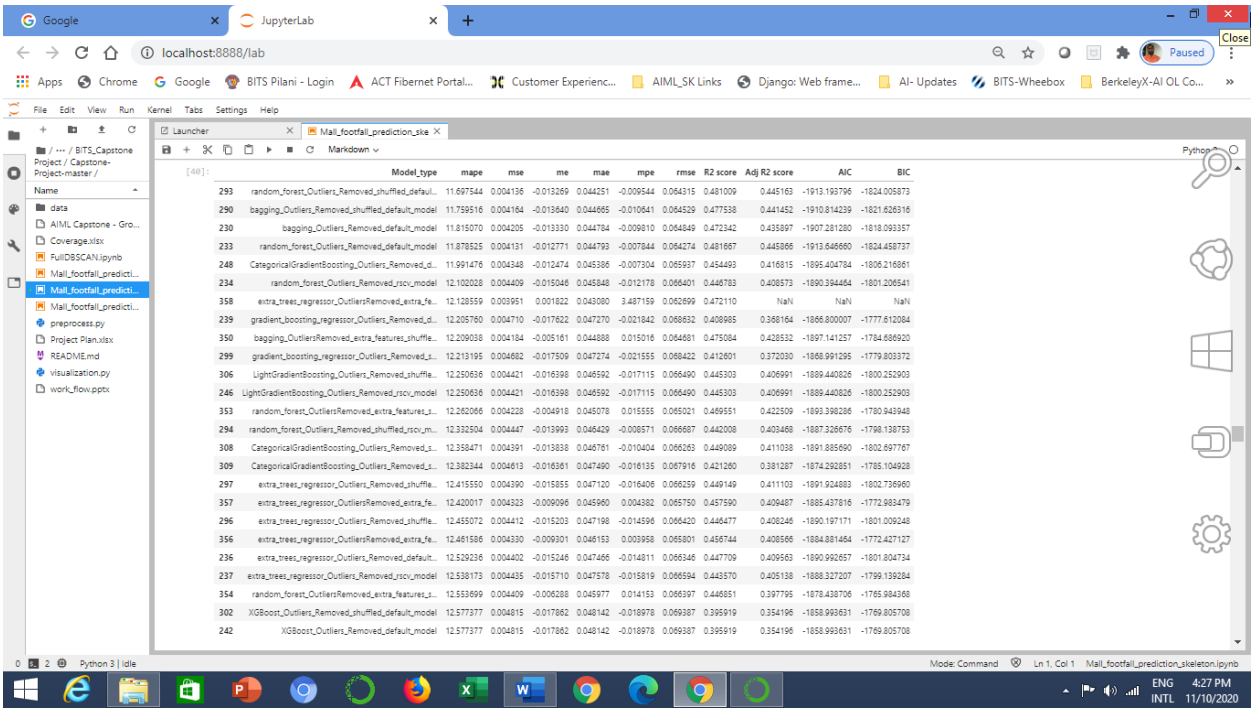


Google | JupyterLab | localhost:8888/lab

```
[39]: loss_df = pd.concat([loss_each_model for loss_each_model in loss_values_all_models], axis=0).reset_index(drop=True)

[40]: loss_df.sort_values('mape').head(25)
```

[40]:

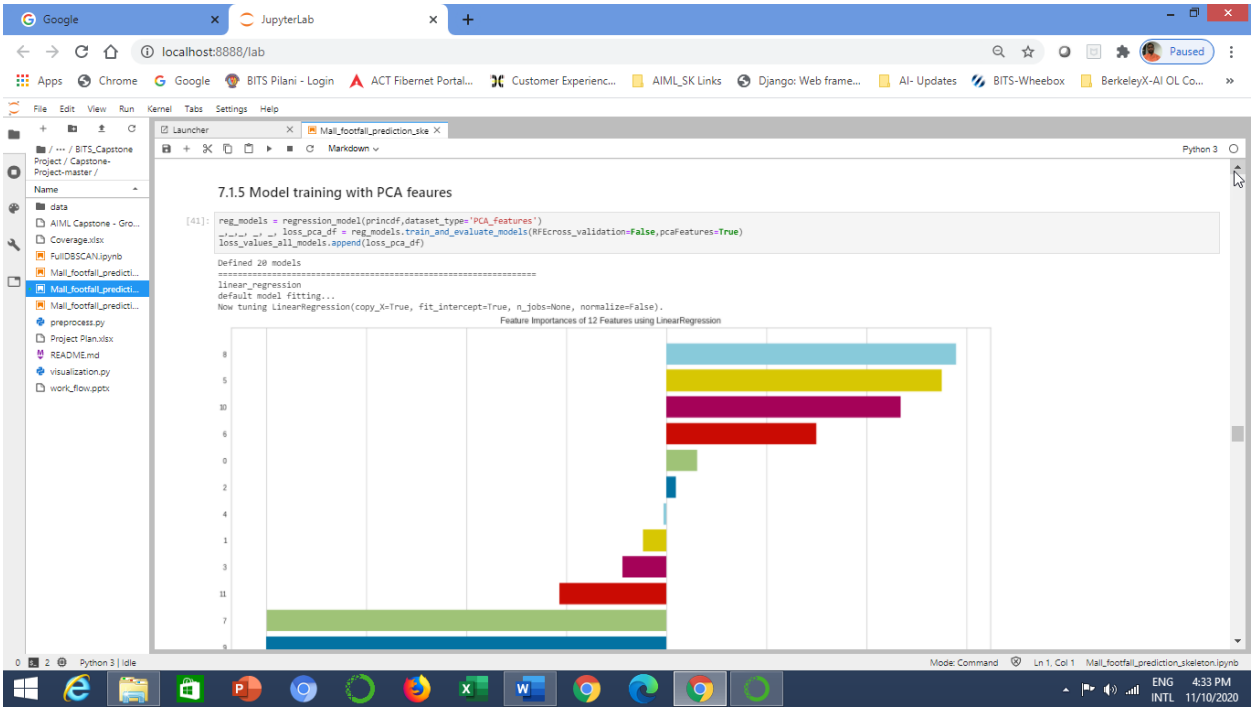
	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
293	random_forest_Outliers_Removed_shuffled_defaul...	11.697544	0.004136	-0.013269	0.044251	-0.009544	0.064315	0.481009	0.445163	-1913.193796	-1824.005873
290	bagging_Outliers_Removed_shuffled_default_model	11.759516	0.004164	-0.013640	0.044665	-0.010641	0.064529	0.477538	0.441452	-1910.814239	-1821.626316
230	bagging_Outliers_Removed_default_model	11.815070	0.004205	-0.013330	0.044784	-0.009810	0.064849	0.472342	0.435897	-1907.281280	-1818.093357
233	random_forest_Outliers_Removed_default_model	11.878525	0.004131	-0.012771	0.044793	-0.007844	0.064274	0.481667	0.445866	-1913.646660	-1824.458737
248	CategoricalGradientBoosting_Outliers_Removed_d...	11.991476	0.004348	-0.012474	0.045386	-0.007304	0.065937	0.454493	0.416815	-1895.404784	-1806.216861
234	random_forest_Outliers_Removed_rscv_model	12.102028	0.004409	-0.015046	0.045848	-0.012178	0.066401	0.446783	0.408573	-1890.394464	-1801.206541
358	extra_trees_regressor_OutliersRemoved_extra_fe...	12.128559	0.003951	0.001822	0.043080	3.487159	0.062699	0.472110	NaN	NaN	NaN
239	gradient_boosting_regressor_Outliers_Removed_d...	12.205760	0.004710	-0.017622	0.047270	-0.021842	0.068632	0.408985	0.368164	-1866.800007	-1777.612084
350	bagging_OutliersRemoved_extra_features_shuffle...	12.209038	0.004184	-0.005161	0.044888	0.015016	0.064681	0.475084	0.428532	-1897.141257	-1784.686920
299	gradient_boosting_regressor_Outliers_Removed_s...	12.213195	0.004682	-0.017509	0.047274	-0.021555	0.068422	0.412601	0.372030	-1868.991295	-1779.803372
306	LightGradientBoosting_Outliers_Removed_shuffle...	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
246	LightGradientBoosting_Outliers_Removed_rscv_model	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
353	random_forest_OutliersRemoved_extra_features_s...	12.262066	0.004228	-0.004918	0.045078	0.015555	0.065021	0.469551	0.422509	-1893.398286	-1780.943948
294	random_forest_Outliers_Removed_shuffled_rscv_m...	12.332504	0.004447	-0.013993	0.046429	-0.008571	0.066687	0.442008	0.403468	-1887.326676	-1798.138753
308	CategoricalGradientBoosting_Outliers_Removed_s...	12.358471	0.004391	-0.013838	0.046761	-0.010404	0.066263	0.449089	0.411038	-1891.885690	-1802.697767
309	CategoricalGradientBoosting_Outliers_Removed_s...	12.382344	0.004613	-0.016361	0.047490	-0.016135	0.067916	0.421260	0.381287	-1874.292851	-1785.104928
297	extra_trees_regressor_Outliers_Removed_shuffle...	12.415550	0.004390	-0.015855	0.047120	-0.016406	0.066259	0.449149	0.411103	-1891.924883	-1802.736960
357	extra_trees_regressor_OutliersRemoved_extra_fe...	12.420017	0.004323	-0.009096	0.045960	0.004382	0.065750	0.457590	0.409487	-1885.437816	-1772.983479
296	extra_trees_regressor_Outliers_Removed_shuffle...	12.455072	0.004412	-0.015203	0.047198	-0.014596	0.066420	0.446477	0.408246	-1890.197171	-1801.009248
356	extra_trees_regressor_OutliersRemoved_extra_fe...	12.461586	0.004330	-0.009301	0.046153	0.003958	0.065801	0.456744	0.408566	-1884.881464	-1772.427127
236	extra_trees_regressor_Outliers_Removed_default...	12.529236	0.004402	-0.015246	0.047466	-0.014811	0.066346	0.447709	0.409563	-1890.992657	-1801.804734
237	extra_trees_regressor_Outliers_Removed_rscv_model	12.538173	0.004435	-0.015710	0.047578	-0.015819	0.066594	0.443570	0.405138	-1888.327207	-1799.139284
354	random_forest_OutliersRemoved_extra_features_s...	12.553699	0.004409	-0.006288	0.045977	0.014153	0.066397	0.446851	0.397795	-1878.438706	-1765.984368
302	XGBoost_Outliers_Removed_shuffled_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708
242	XGBoost_Outliers_Removed_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708

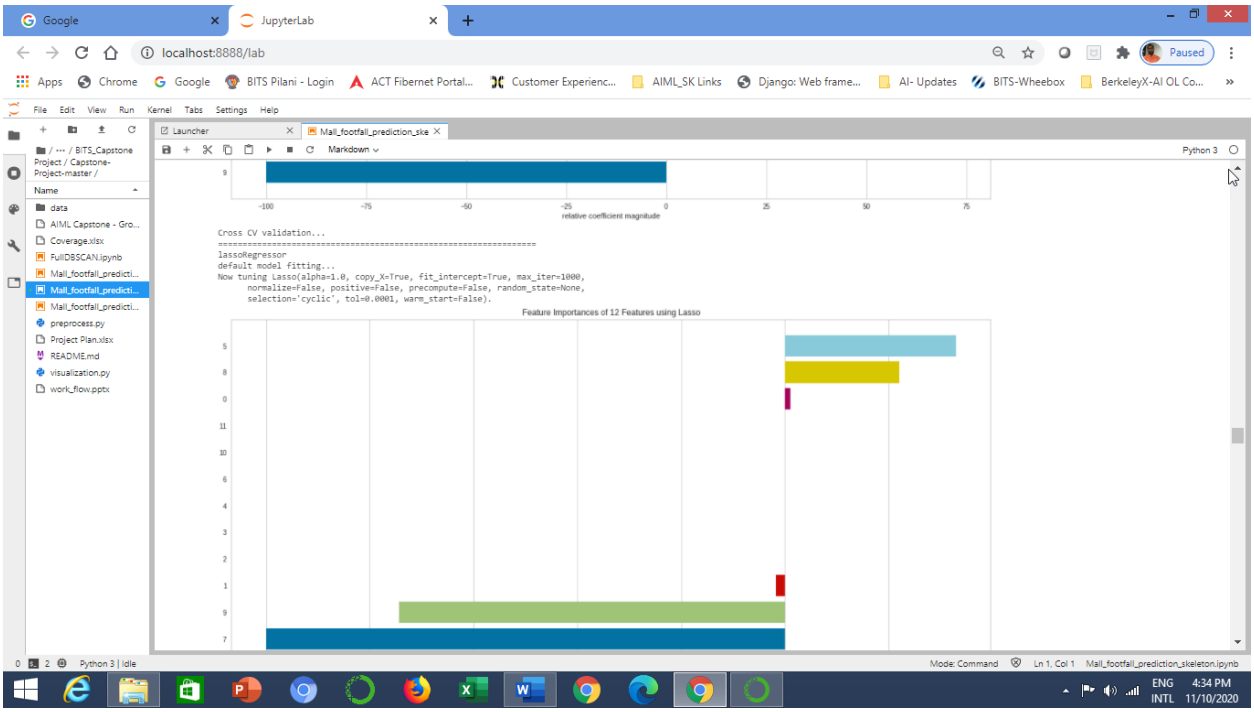


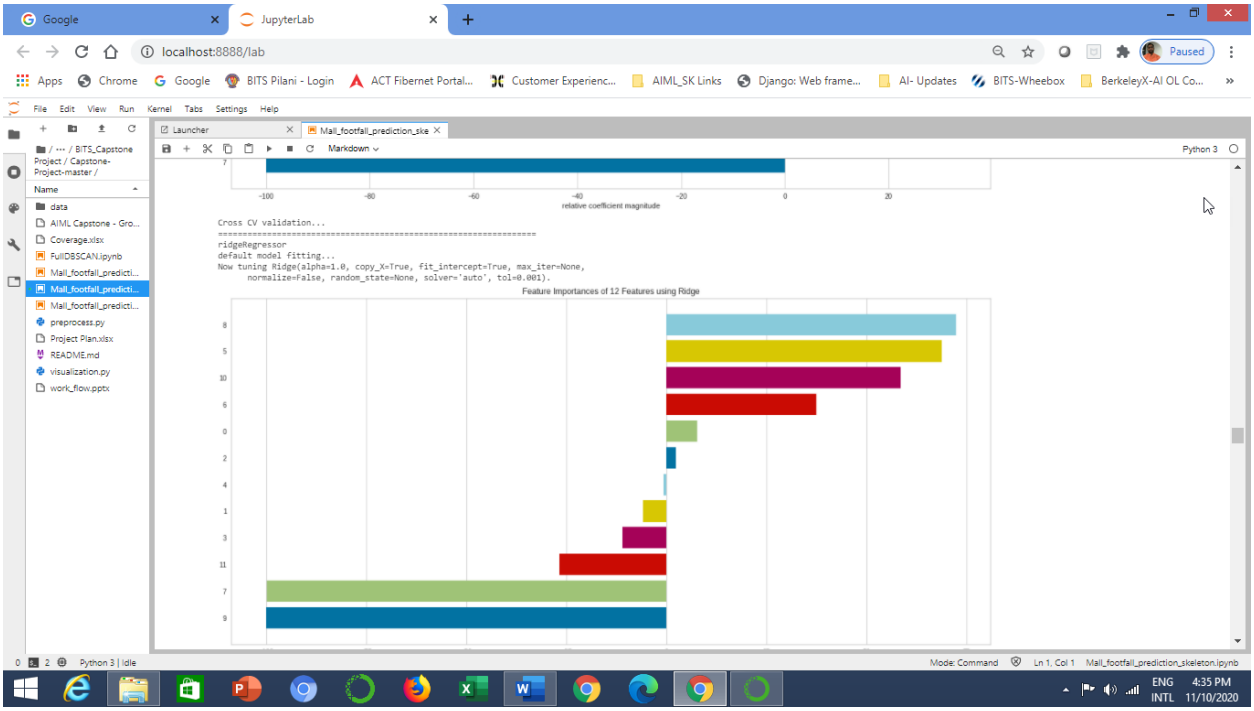
Google | JupyterLab | localhost:8888/lab

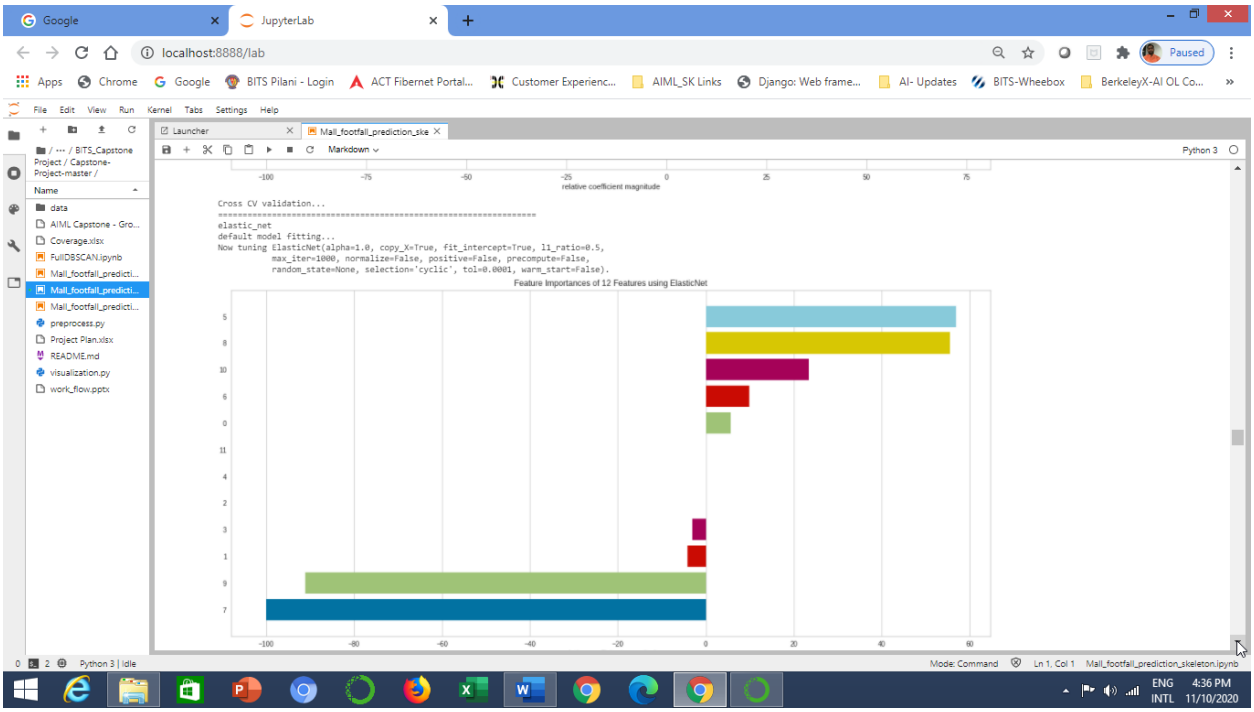
[40]:

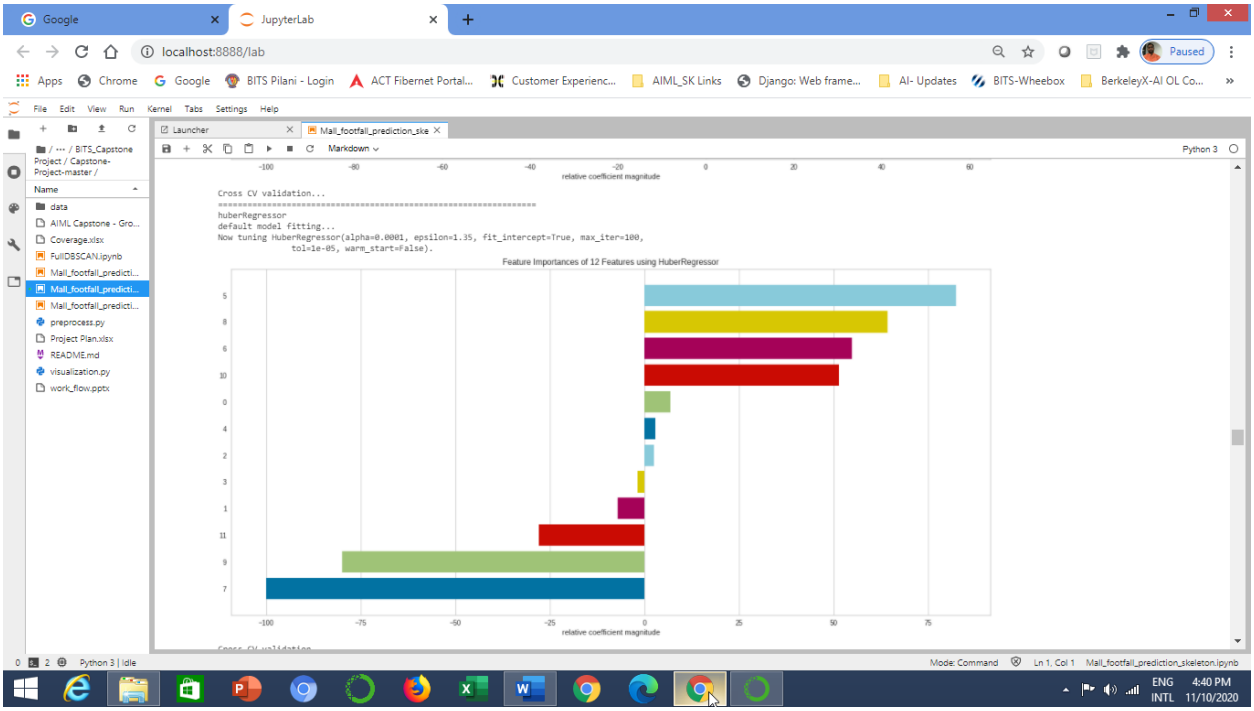
	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
293	random_forest_Outliers_Removed_shuffled_defaul...	11.697544	0.004136	-0.013269	0.044251	-0.009544	0.064315	0.481009	0.445163	-1913.193796	-1824.005873
290	bagging_Outliers_Removed_shuffled_default_model	11.759516	0.004164	-0.013640	0.044665	-0.010641	0.064529	0.477538	0.441452	-1910.814239	-1821.626316
230	bagging_Outliers_Removed_default_model	11.815070	0.004205	-0.013330	0.044784	-0.009810	0.064849	0.472342	0.435897	-1907.281280	-1818.093357
233	random_forest_Outliers_Removed_default_model	11.878525	0.004131	-0.012771	0.044793	-0.007844	0.064274	0.481667	0.445866	-1913.646660	-1824.458737
248	CategoricalGradientBoosting_Outliers_Removed_d...	11.991476	0.004348	-0.012474	0.045386	-0.007304	0.065937	0.454493	0.416815	-1895.404784	-1806.216861
234	random_forest_Outliers_Removed_rscv_model	12.102028	0.004409	-0.015046	0.045848	-0.012178	0.066401	0.446783	0.408573	-1890.394464	-1801.206541
358	extra_trees_regressor_OutliersRemoved_extra_fe...	12.128559	0.003951	0.001822	0.043080	3.487159	0.062699	0.472110	NaN	NaN	NaN
239	gradient_boosting_regressor_Outliers_Removed_d...	12.205760	0.004710	-0.017622	0.047270	-0.021842	0.068632	0.408985	0.368164	-1866.800007	-1777.612084
350	bagging_OutliersRemoved_extra_features_shuffle...	12.209038	0.004184	-0.005161	0.044888	0.015016	0.064681	0.475084	0.428532	-1897.141257	-1784.686920
299	gradient_boosting_regressor_Outliers_Removed_s...	12.213195	0.004682	-0.017509	0.047274	-0.021555	0.068422	0.412601	0.372030	-1868.991295	-1779.803372
306	LightGradientBoosting_Outliers_Removed_shuffle...	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
246	LightGradientBoosting_Outliers_Removed_rscv_model	12.250636	0.004421	-0.016398	0.046592	-0.017115	0.066490	0.445303	0.406991	-1889.440826	-1800.252903
353	random_forest_OutliersRemoved_extra_features_s...	12.262066	0.004228	-0.004918	0.045078	0.015555	0.065021	0.469551	0.422509	-1893.398286	-1780.943948
294	random_forest_Outliers_Removed_shuffled_rscv_m...	12.332504	0.004447	-0.013993	0.046429	-0.008571	0.066687	0.442008	0.403468	-1887.326676	-1798.138753
308	CategoricalGradientBoosting_Outliers_Removed_s...	12.358471	0.004391	-0.013838	0.046761	-0.010404	0.066263	0.449089	0.411038	-1891.885690	-1802.697767
309	CategoricalGradientBoosting_Outliers_Removed_s...	12.382344	0.004613	-0.016361	0.047490	-0.016135	0.067916	0.421260	0.381287	-1874.292851	-1785.104928
297	extra_trees_regressor_Outliers_Removed_shuffle...	12.415550	0.004390	-0.015855	0.047120	-0.016406	0.066259	0.449149	0.411103	-1891.924883	-1802.736960
357	extra_trees_regressor_OutliersRemoved_extra_fe...	12.420017	0.004323	-0.009096	0.045960	0.004382	0.065750	0.457590	0.409487	-1885.437816	-1772.983479
296	extra_trees_regressor_Outliers_Removed_shuffle...	12.455072	0.004412	-0.015203	0.047198	-0.014596	0.066420	0.446477	0.408246	-1890.197171	-1801.009248
356	extra_trees_regressor_OutliersRemoved_extra_fe...	12.461586	0.004330	-0.009301	0.046153	0.003958	0.065801	0.456744	0.408566	-1884.881464	-1772.427127
236	extra_trees_regressor_Outliers_Removed_default...	12.529236	0.004402	-0.015246	0.047466	-0.014811	0.066346	0.447709	0.409563	-1890.992657	-1801.804734
237	extra_trees_regressor_Outliers_Removed_rscv_model	12.538173	0.004435	-0.015710	0.047578	-0.015819	0.066594	0.443570	0.405138	-1888.327207	-1799.139284
354	random_forest_OutliersRemoved_extra_features_s...	12.553699	0.004409	-0.006288	0.045977	0.014153	0.066397	0.446851	0.397795	-1878.438706	-1765.984368
302	XGBoost_Outliers_Removed_shuffled_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708
242	XGBoost_Outliers_Removed_default_model	12.577377	0.004815	-0.017862	0.048142	-0.018978	0.069387	0.395919	0.354196	-1858.993631	-1769.805708

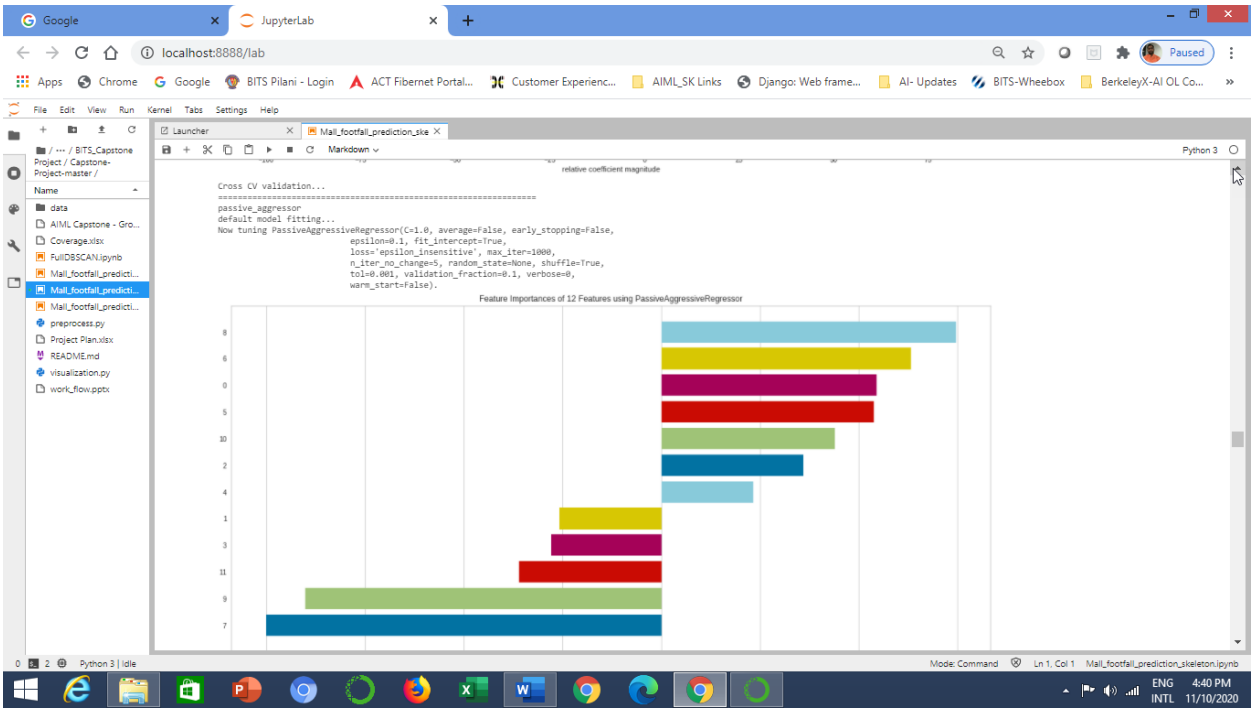

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
7.1.5 Model training with PCA feaures
[41]:
reg_models = regression_model(princdf,dataset_type='PCA_features')
_,_,_, _, _, loss_pca_df = reg_models.train_and_evaluate_models(RFEcross_validation=False,pcaFeatures=True)
loss_values_all_models.append(loss_pca_df)
Defined 20 models
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
Feature Importances of 12 Features using LinearRegression
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:33 PM
11/10/2020

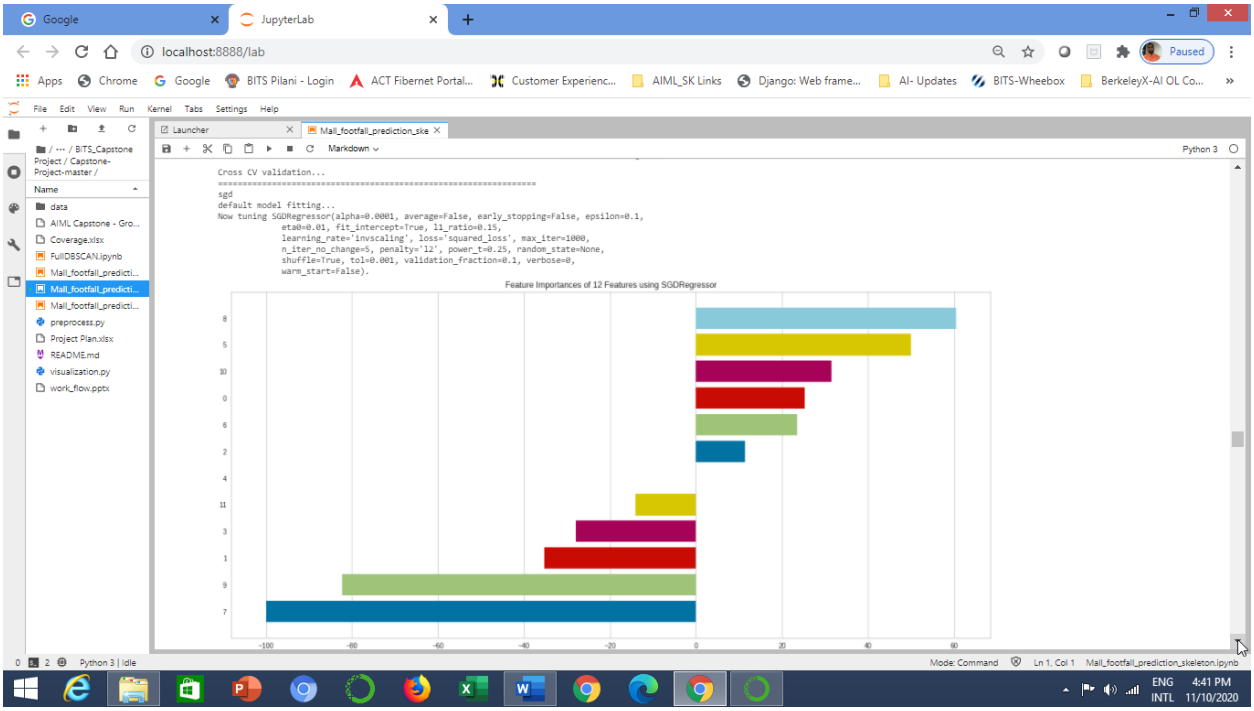

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 12 Features using Lasso
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:34 PM
11/10/2020

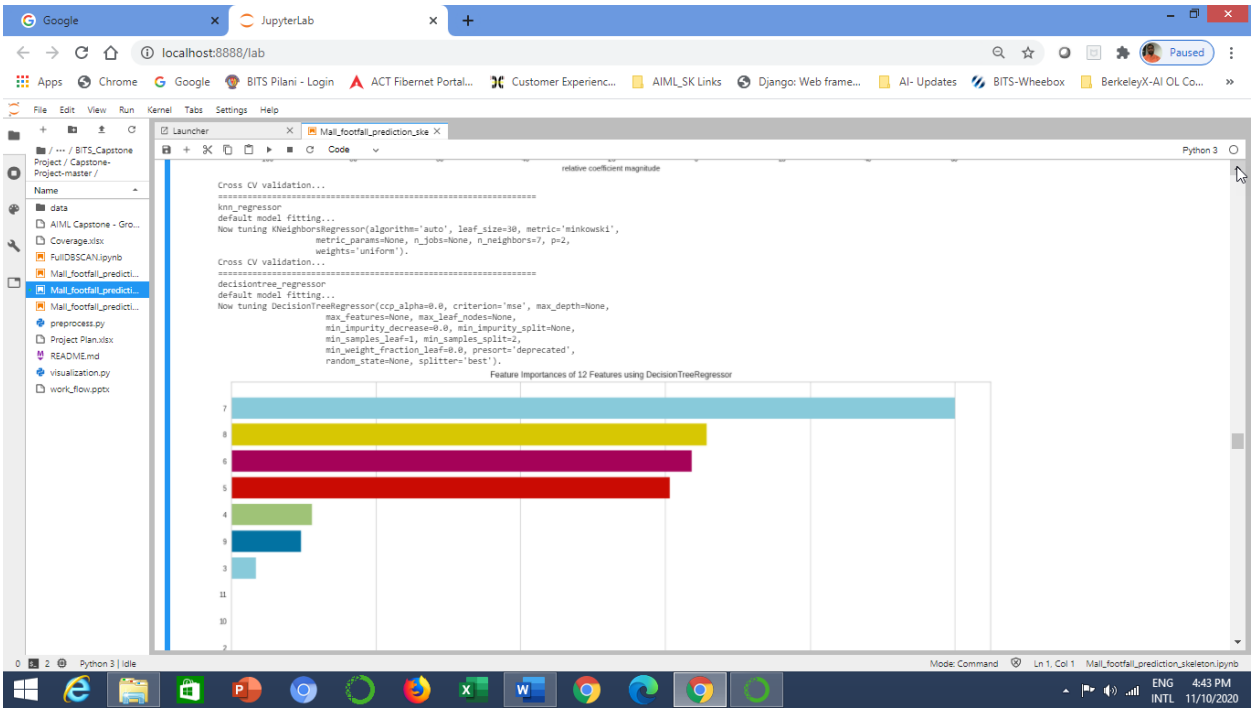

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 12 Features using Ridge
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:35 PM
11/10/2020

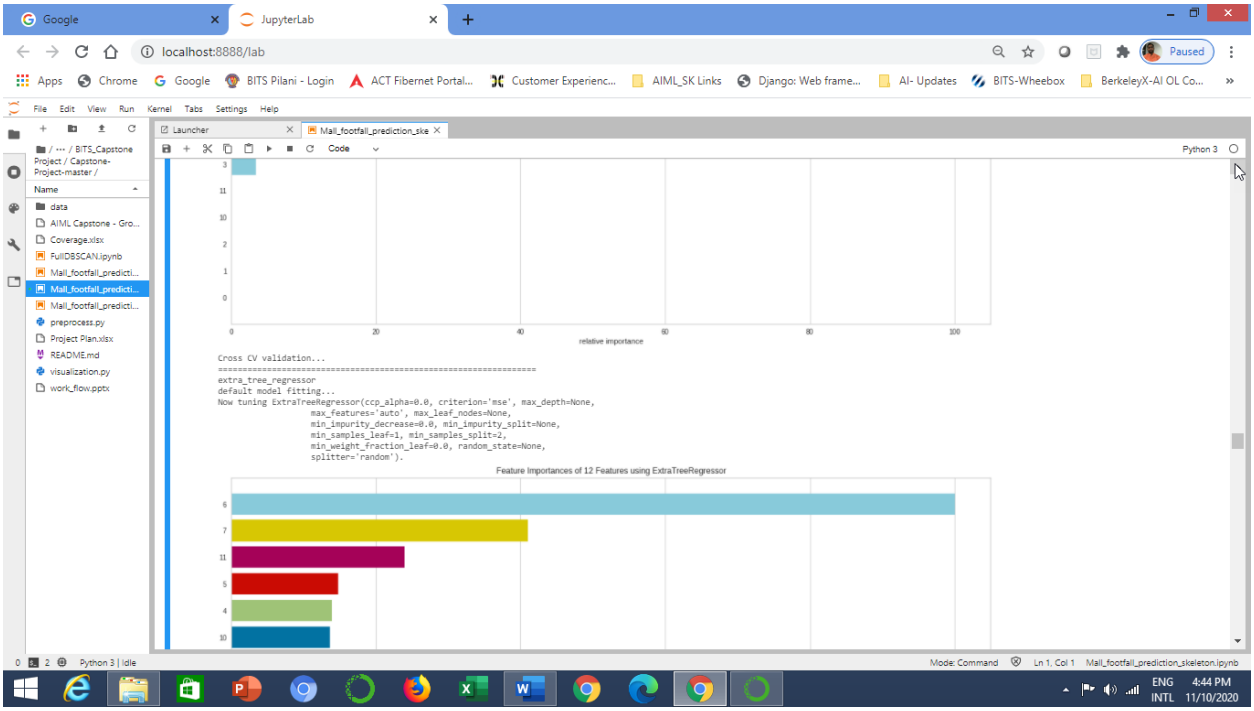

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 12 Features using ElasticNet
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:36 PM
11/10/2020

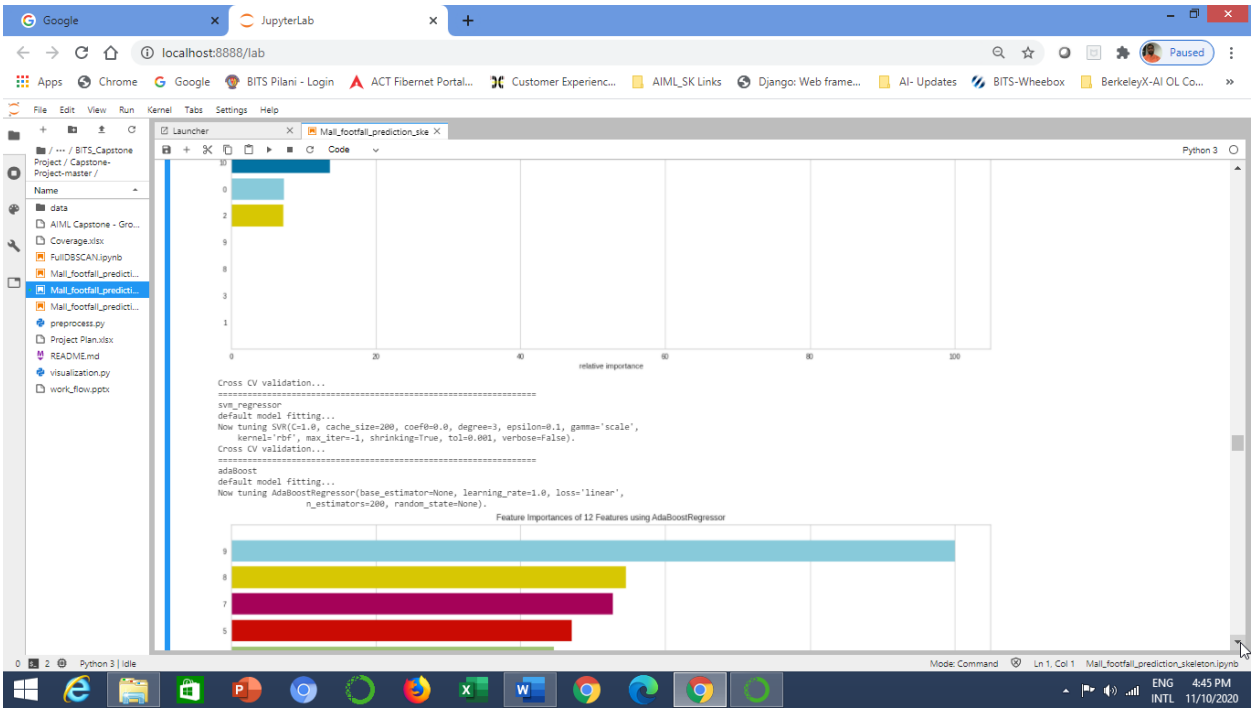

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
huberRegressor
default model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 12 Features using HuberRegressor
relative coefficient magnitude
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:40 PM
11/10/2020

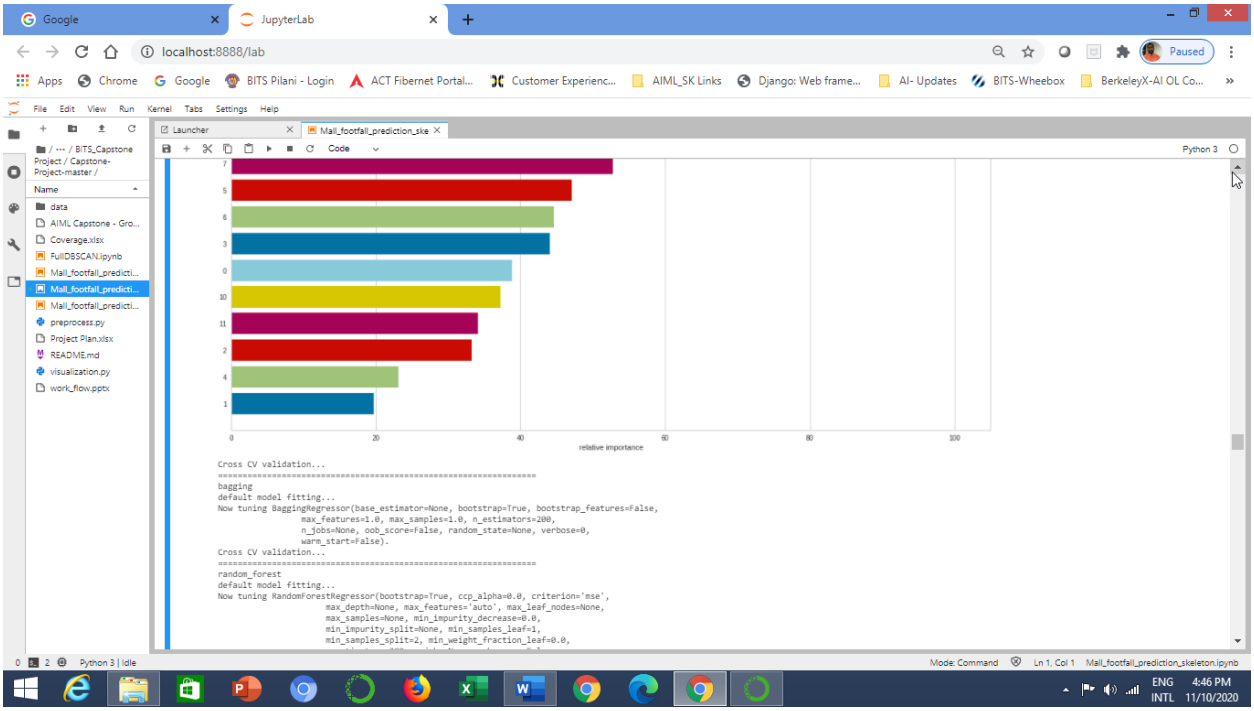

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 12 Features using PassiveAggressiveRegressor
relative coefficient magnitude
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:40 PM
11/10/2020

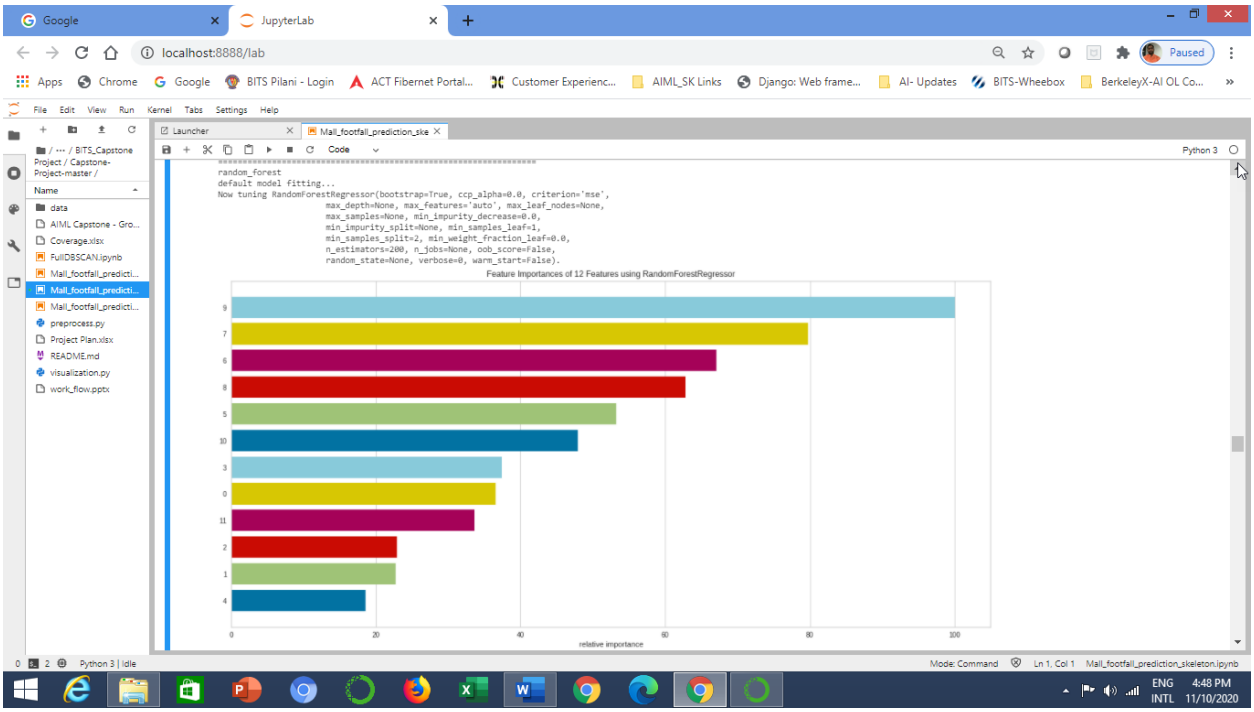

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Markdown
Python 3
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 12 Features using SGDRegressor
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:41 PM
11/10/2020

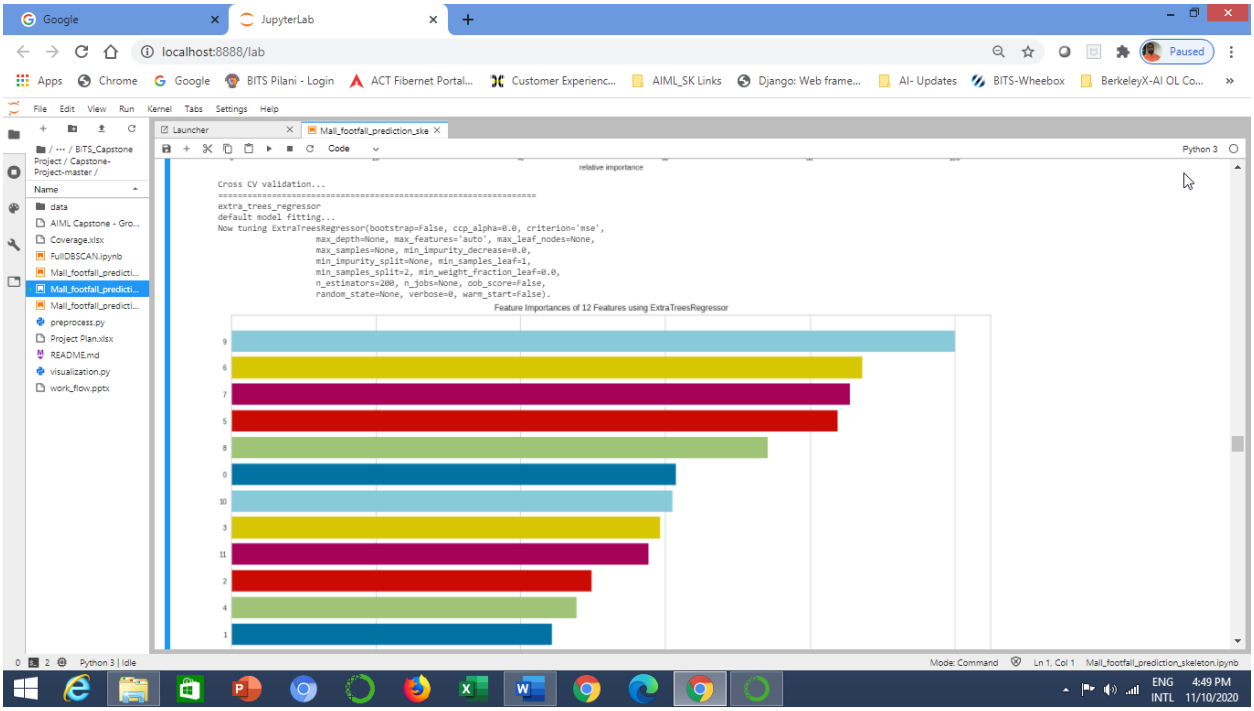

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 12 Features using DecisionTreeRegressor
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:43 PM
11/10/2020

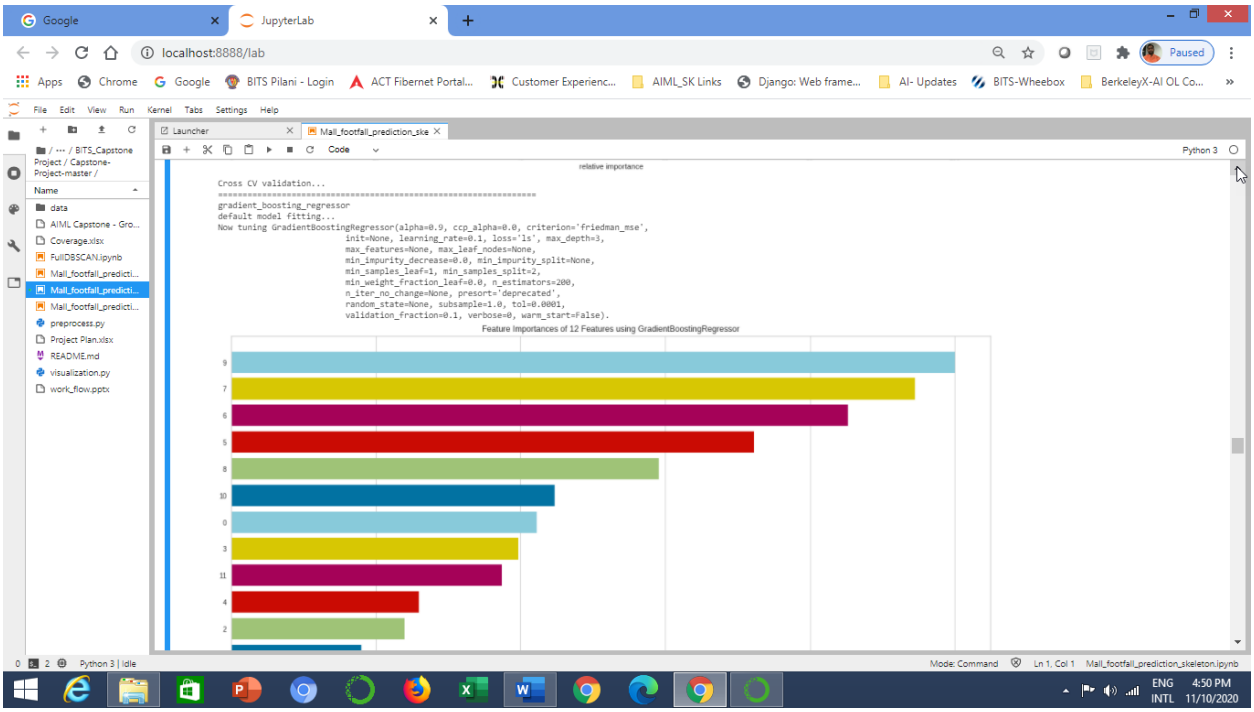

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').
Feature Importances of 12 Features using ExtraTreeRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:44 PM
11/10/2020

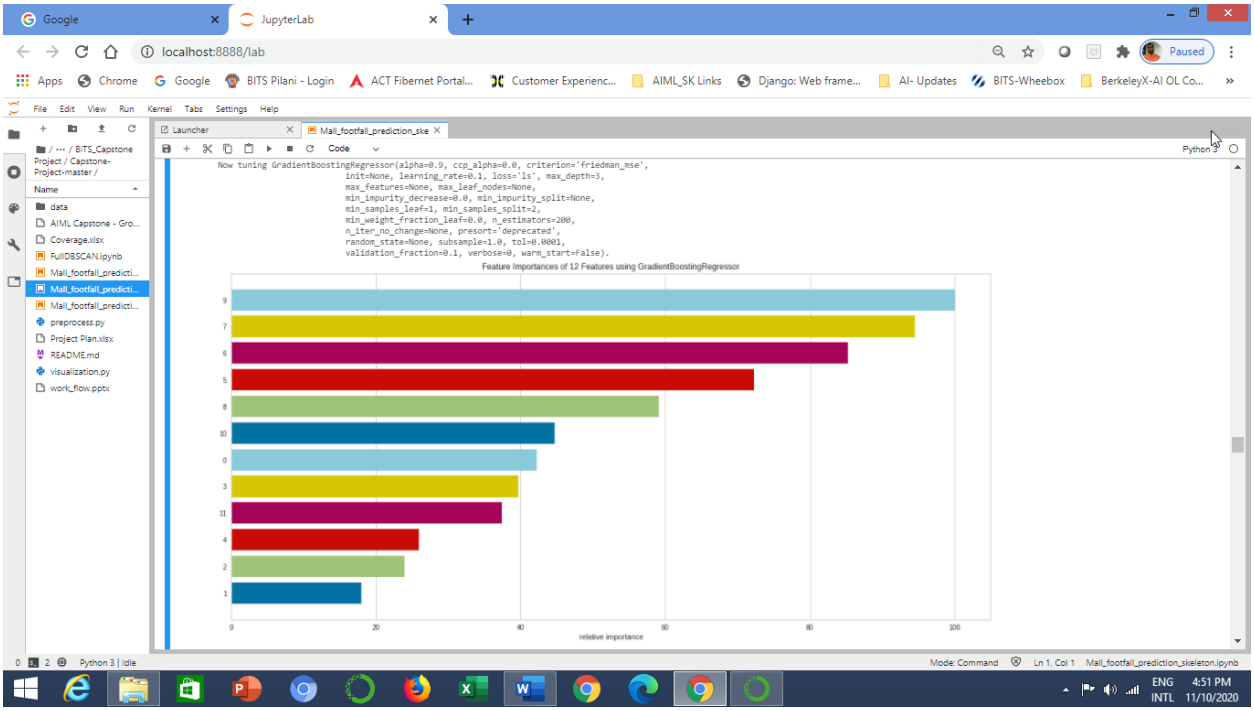

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 12 Features using AdaBoostRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:45 PM
11/10/2020

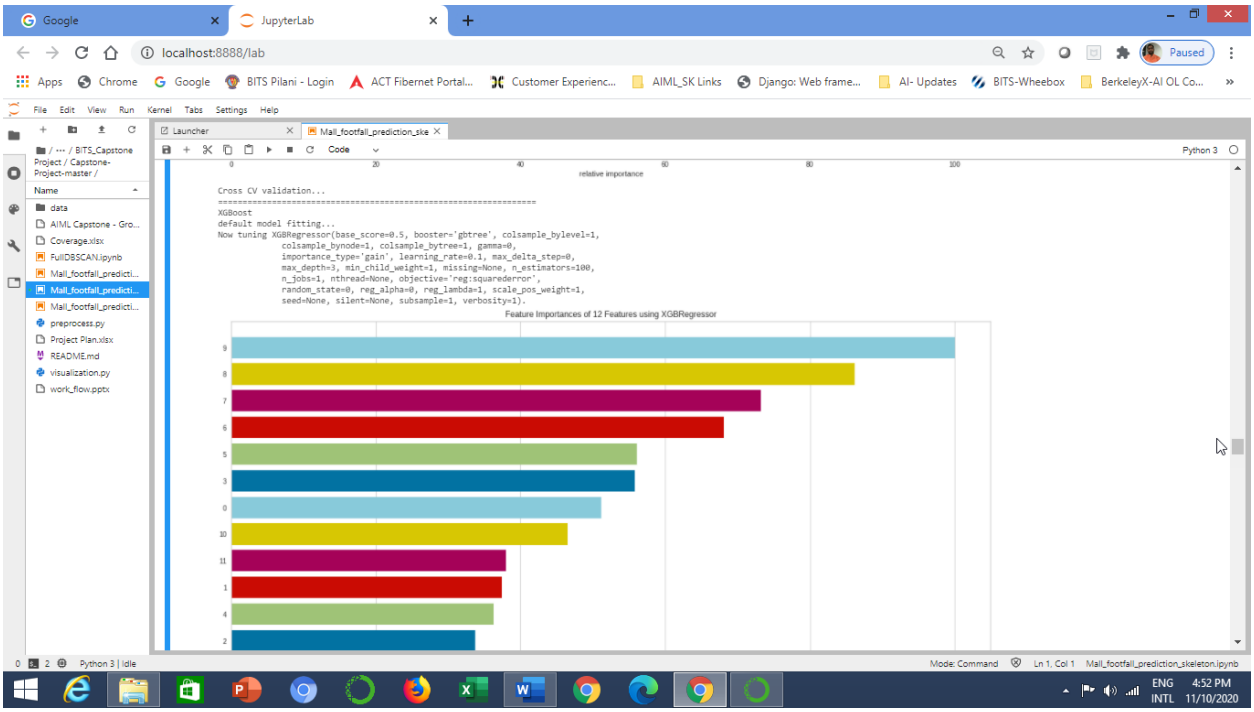

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:46 PM
11/10/2020

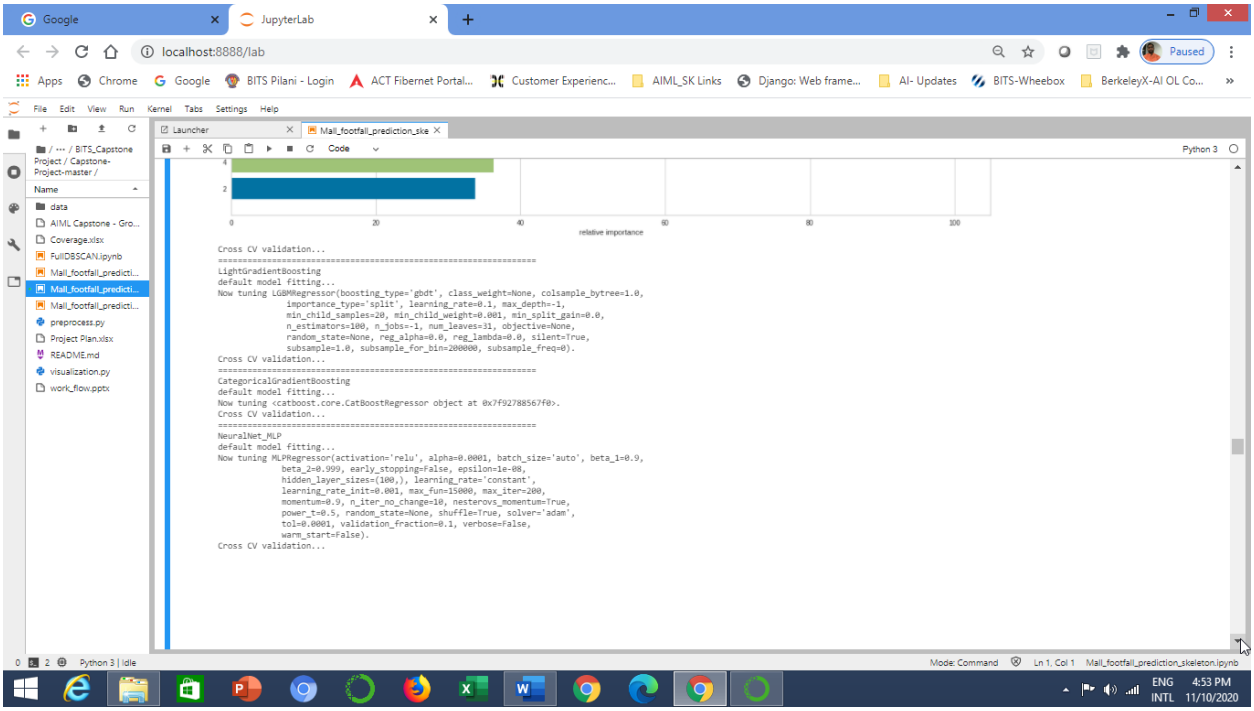

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 12 Features using RandomForestRegressor
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:48 PM
11/10/2020

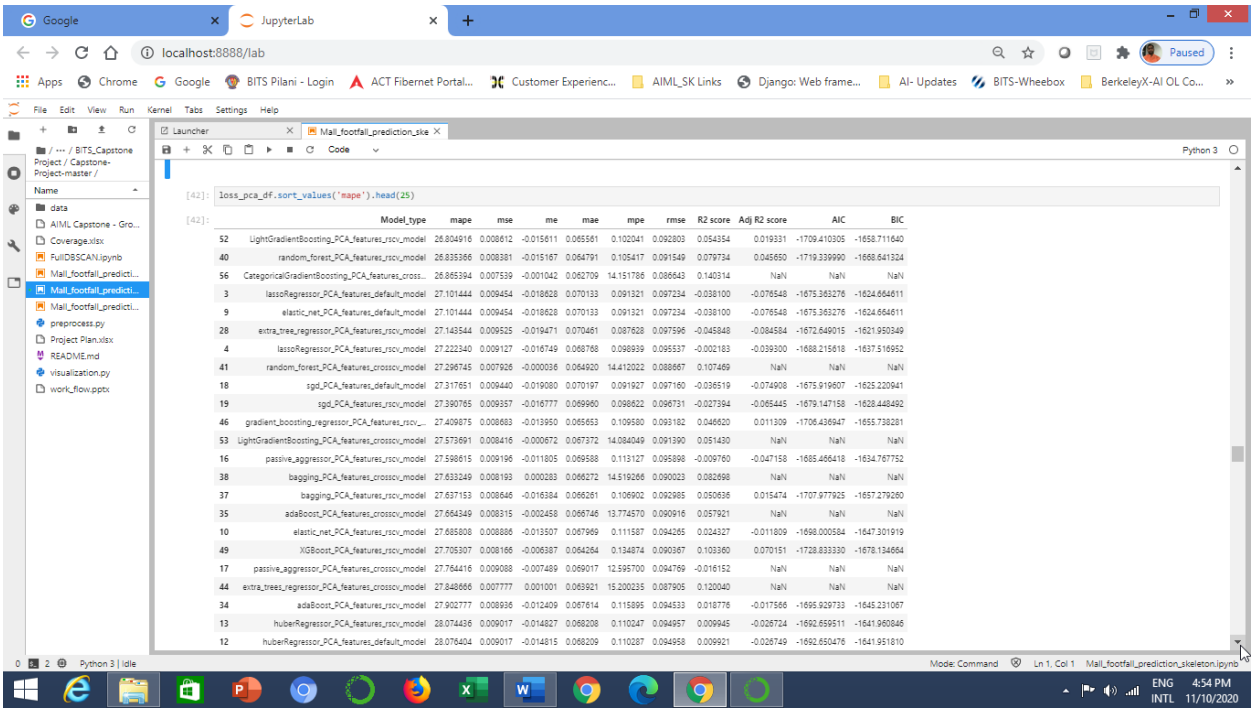

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 12 Features using ExtraTreesRegressor
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:49 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 12 Features using GradientBoostingRegressor
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:50 PM
11/10/2020

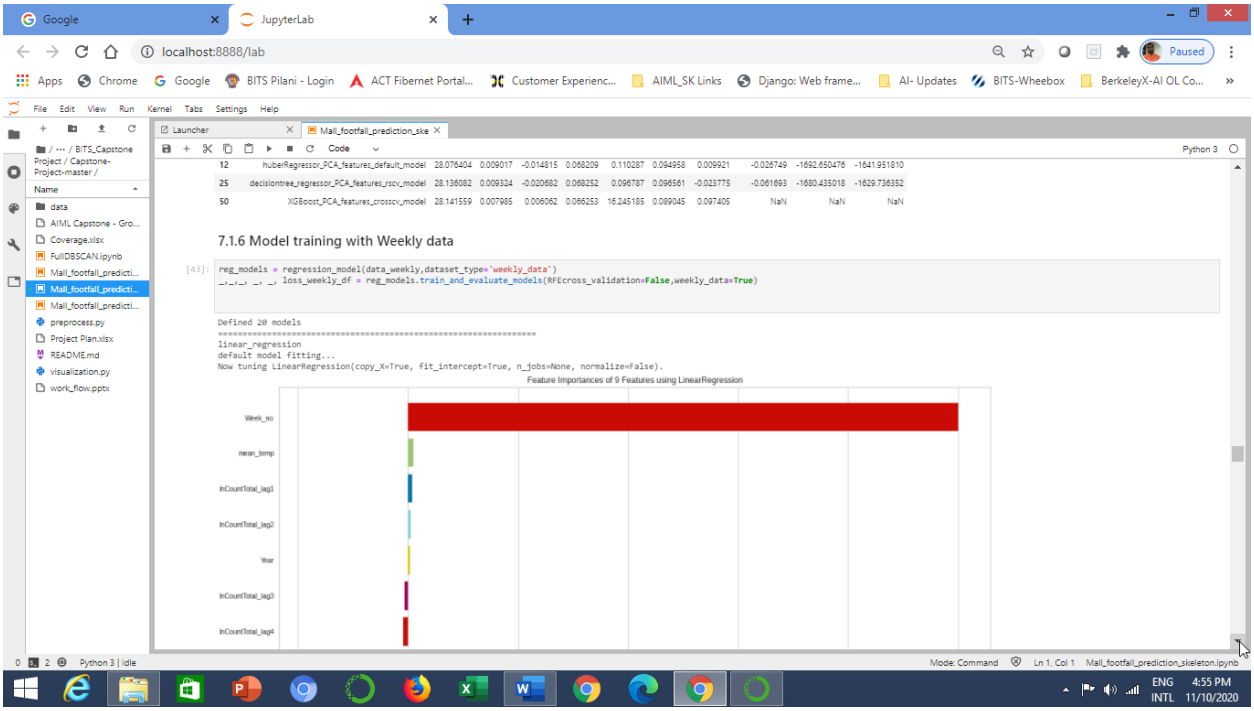

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 12 Features using GradientBoostingRegressor
relative importance
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:51 PM
11/10/2020

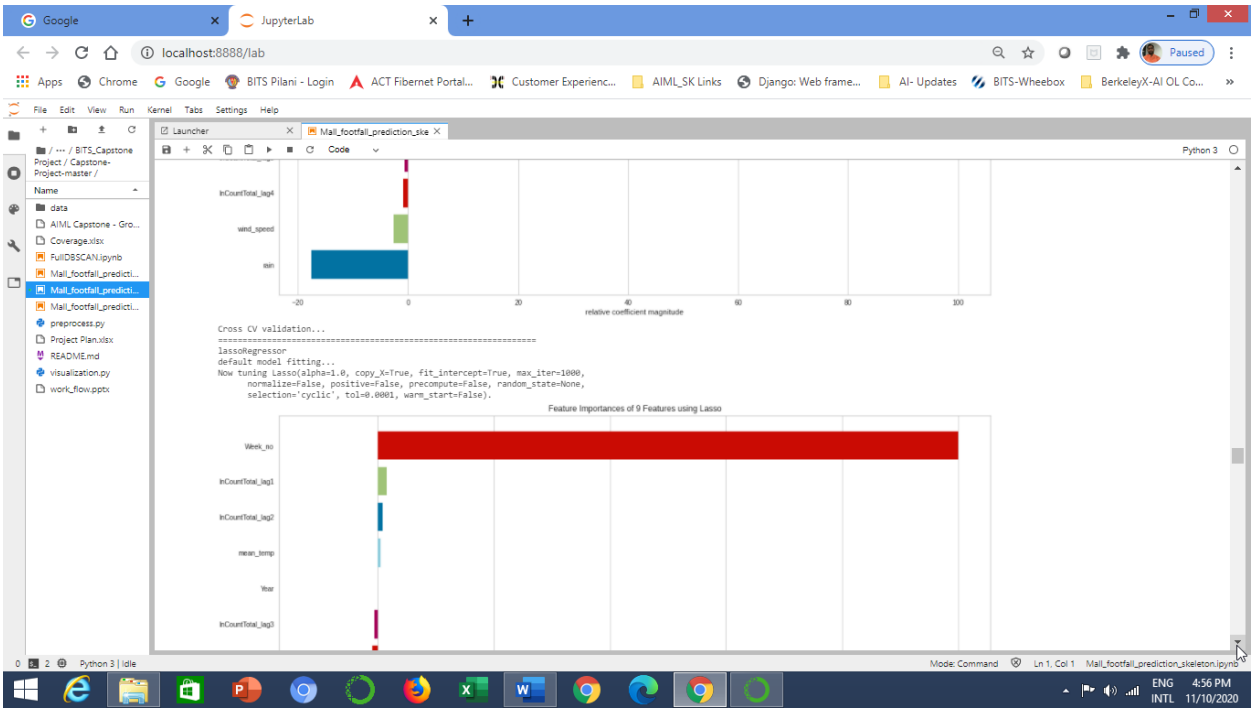

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative importance
Cross CV validation...
XGBoost
default model fitting...
Now tuning XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
colsample_bynode=1, colsample_bytree=1, gamma=0,
importance_type='gain', learning_rate=0.1, max_delta_step=0,
max_depth=3, min_child_weight=1, missing=None, n_estimators=100,
n_jobs=-1, nthread=None, objective='reg:squarederror',
random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
seed=None, silent=None, subsample=1, verbosity=1).
Feature Importances of 12 Features using XGBRegressor
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:52 PM
11/10/2020

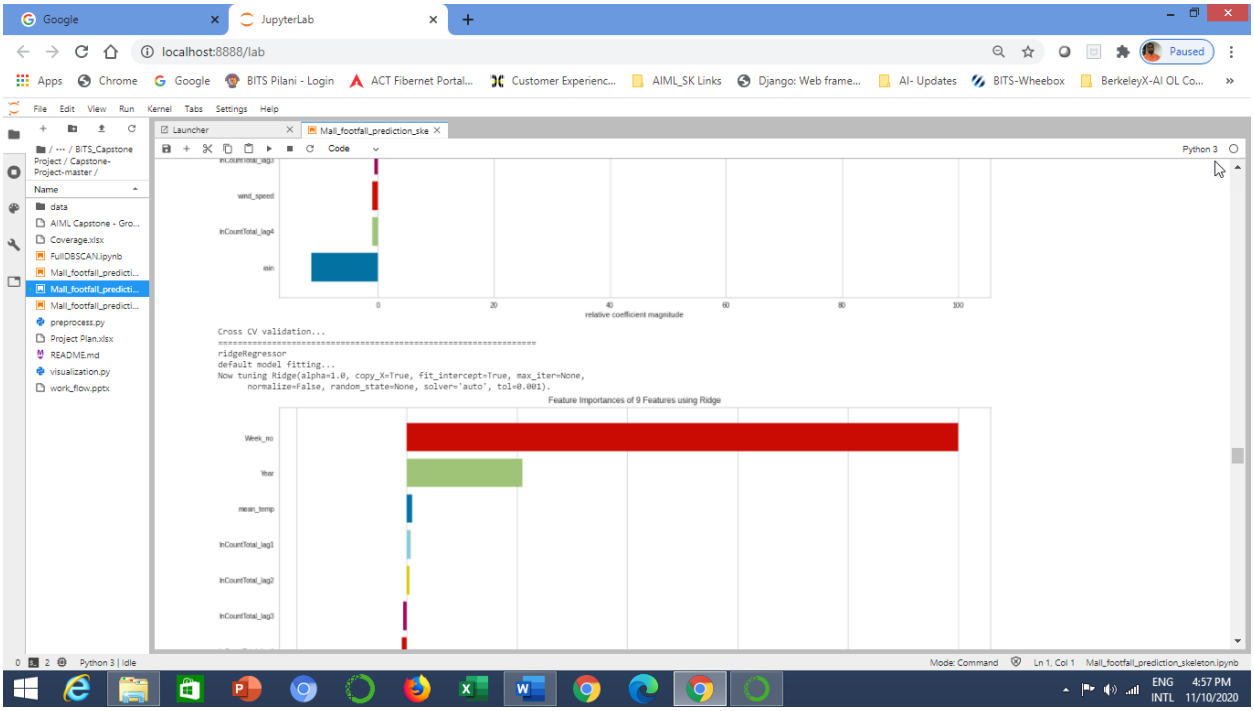


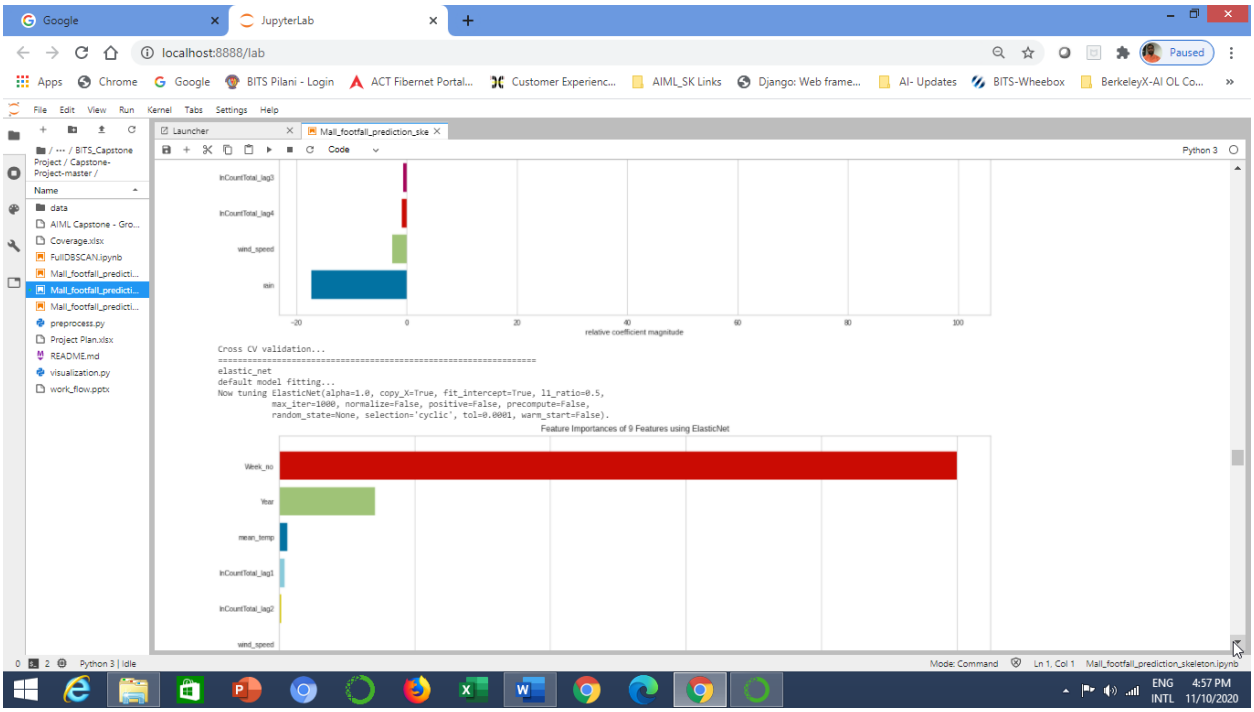


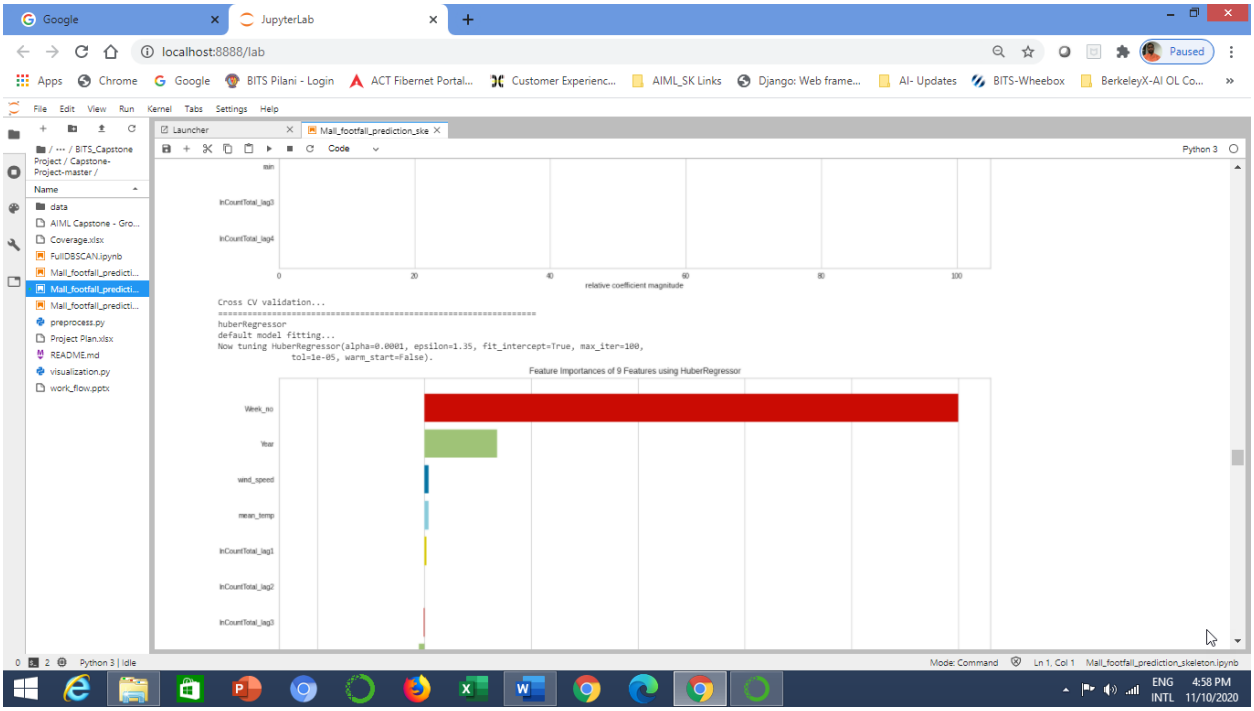
	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
52	LightGradientBoosting_PCA_features_rscv_model	26.804916	0.008612	-0.015611	0.065561	0.102041	0.092803	0.054354	0.019331	-1709.410305	-1658.711640
40	random_forest_PCA_features_rscv_model	26.835366	0.008381	-0.015167	0.064791	0.105417	0.091549	0.079734	0.045650	-1719.339990	-1668.641324
56	CategoricalGradientBoosting_PCA_features_cross...	26.865394	0.007539	-0.001042	0.062709	14.151786	0.086643	0.140314	NaN	NaN	NaN
3	lassoRegressor_PCA_features_default_model	27.101444	0.009454	-0.018628	0.070133	0.091321	0.097234	-0.038100	-0.076548	-1675.363276	-1624.664611
9	elastic_net_PCA_features_default_model	27.101444	0.009454	-0.018628	0.070133	0.091321	0.097234	-0.038100	-0.076548	-1675.363276	-1624.664611
28	extra_tree_regressor_PCA_features_rscv_model	27.143544	0.009525	-0.019471	0.070461	0.087628	0.097596	-0.045848	-0.084584	-1672.649015	-1621.950349
4	lassoRegressor_PCA_features_rscv_model	27.222340	0.009127	-0.016749	0.068768	0.098939	0.095537	-0.002183	-0.039300	-1688.215618	-1637.516952
41	random_forest_PCA_features_crossrcv_model	27.296745	0.007926	-0.000036	0.064920	14.412022	0.088667	0.107469	NaN	NaN	NaN
18	sgd_PCA_features_default_model	27.317651	0.009440	-0.019080	0.070197	0.091927	0.097160	-0.036519	-0.074908	-1675.919607	-1625.220941
19	sgd_PCA_features_rscv_model	27.390765	0.009357	-0.016777	0.069960	0.098822	0.096731	-0.027394	-0.065445	-1679.147158	-1628.448492
46	gradient_boosting_regressor_PCA_features_rscv_...	27.409875	0.008683	-0.013950	0.065653	0.109580	0.093182	0.046620	0.011309	-1706.436947	-1655.738281
53	LightGradientBoosting_PCA_features_crossrcv_model	27.573691	0.008416	-0.000672	0.067372	14.084049	0.091390	0.051430	NaN	NaN	NaN
16	passive_aggressor_PCA_features_rscv_model	27.598615	0.009196	-0.011805	0.069588	0.113127	0.095898	-0.009760	-0.047158	-1685.466418	-1634.767752
38	bagging_PCA_features_crossrcv_model	27.633249	0.008193	0.000283	0.066272	14.519266	0.090023	0.082698	NaN	NaN	NaN
37	bagging_PCA_features_rscv_model	27.637153	0.008646	-0.016384	0.066261	0.106902	0.092985	0.050636	0.015474	-1707.977925	-1657.279260
35	adaBoost_PCA_features_crossrcv_model	27.664349	0.008315	-0.002458	0.066746	13.774570	0.090916	0.057921	NaN	NaN	NaN
10	elastic_net_PCA_features_rscv_model	27.685808	0.008886	-0.013507	0.067969	0.111587	0.094265	0.024327	-0.011809	-1698.000584	-1647.301919
49	XGBoost_PCA_features_rscv_model	27.705307	0.008166	-0.006387	0.064264	0.134874	0.090367	0.103360	0.070151	-1728.833330	-1678.134664
17	passive_aggressor_PCA_features_crossrcv_model	27.764416	0.009088	-0.007489	0.069017	12.595700	0.094769	-0.016152	NaN	NaN	NaN
44	extra_trees_regressor_PCA_features_crossrcv_model	27.848666	0.007777	0.001001	0.063921	15.200235	0.087905	0.120040	NaN	NaN	NaN
34	adaBoost_PCA_features_rscv_model	27.902777	0.008936	-0.012409	0.067614	0.115895	0.094533	0.018776	-0.017566	-1695.929733	-1645.231067
13	huberRegressor_PCA_features_rscv_model	28.074436	0.009017	-0.014827	0.068208	0.110247	0.094957	0.009945	-0.026724	-1692.659511	-1641.960846
12	huberRegressor_PCA_features_default_model	28.076404	0.009017	-0.014815	0.068209	0.110287	0.094958	0.009921	-0.026749	-1692.650476	-1641.951810

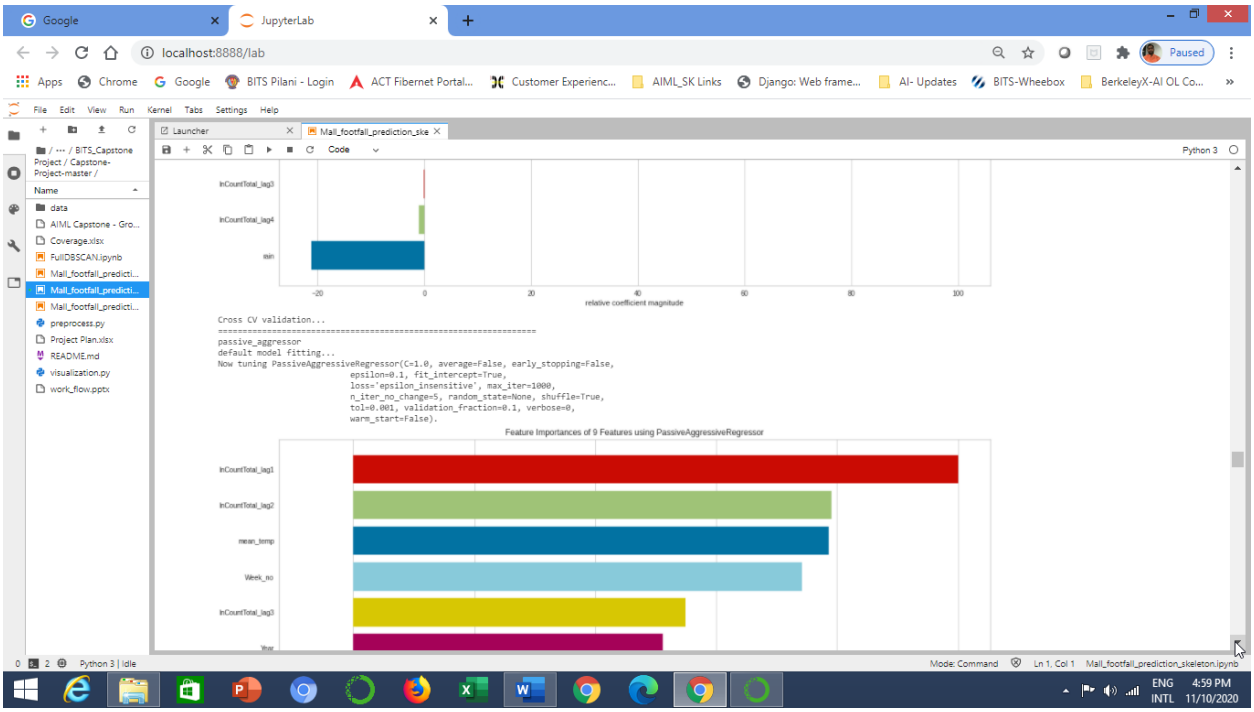

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
12 huberRegressor_PCA_features_default_model 28.076404 0.009017 -0.014815 0.068209 0.110287 0.094958 0.009921 -0.026749 -1692.650476 -1641.951810
25 decisiontree_regressor_PCA_features_rscv_model 28.136082 0.009324 -0.020682 0.068252 0.096787 0.096561 -0.023775 -0.061693 -1680.435018 -1629.736352
50 XGBoost_PCA_features_crosscv_model 28.141559 0.007985 0.006062 0.066253 16.245185 0.089045 0.097405 NaN NaN NaN
7.1.6 Model training with Weekly data
[43]:
reg_models = regression_model(data_weekly,dataset_type='weekly_data')
loss_weekly_df = reg_models.train_and_evaluate_models(RFEcross_validation=False,weekly_data=True)
Defined 20 models
linear_regression
default model fitting...
Now tuning LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False).
Feature Importances of 9 Features using LinearRegression
Week_no
mean_temp
lnCountTotal_lag1
lnCountTotal_lag2
Year
lnCountTotal_lag3
lnCountTotal_lag4
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 4:55 PM 11/10/2020

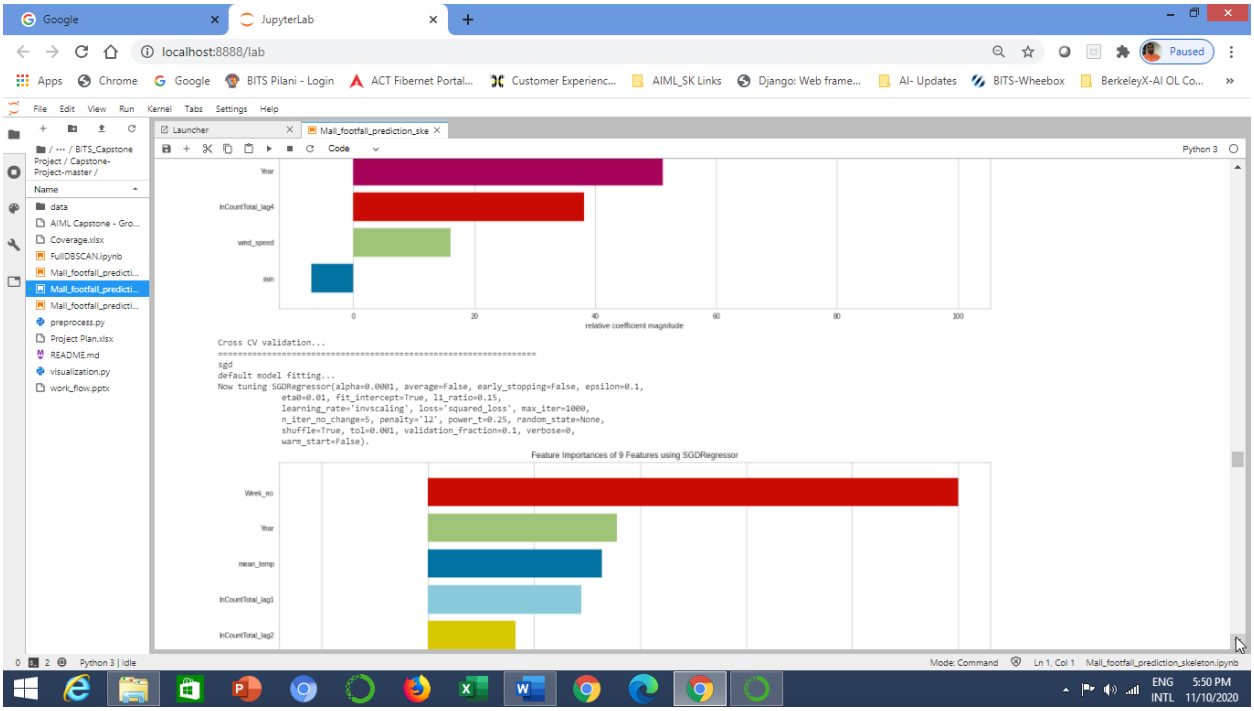

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
lnCountTotal_lag4
wind_speed
min
-20
0
20
40
60
80
100
relative coefficient magnitude
Cross CV validation...
lassoRegressor
default model fitting...
Now tuning Lasso(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=1000,
normalize=False, positive=False, precompute=False, random_state=None,
selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 9 Features using Lasso
Week_no
lnCountTotal_lag1
lnCountTotal_lag2
mean_temp
Year
lnCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 4:56 PM 11/10/2020

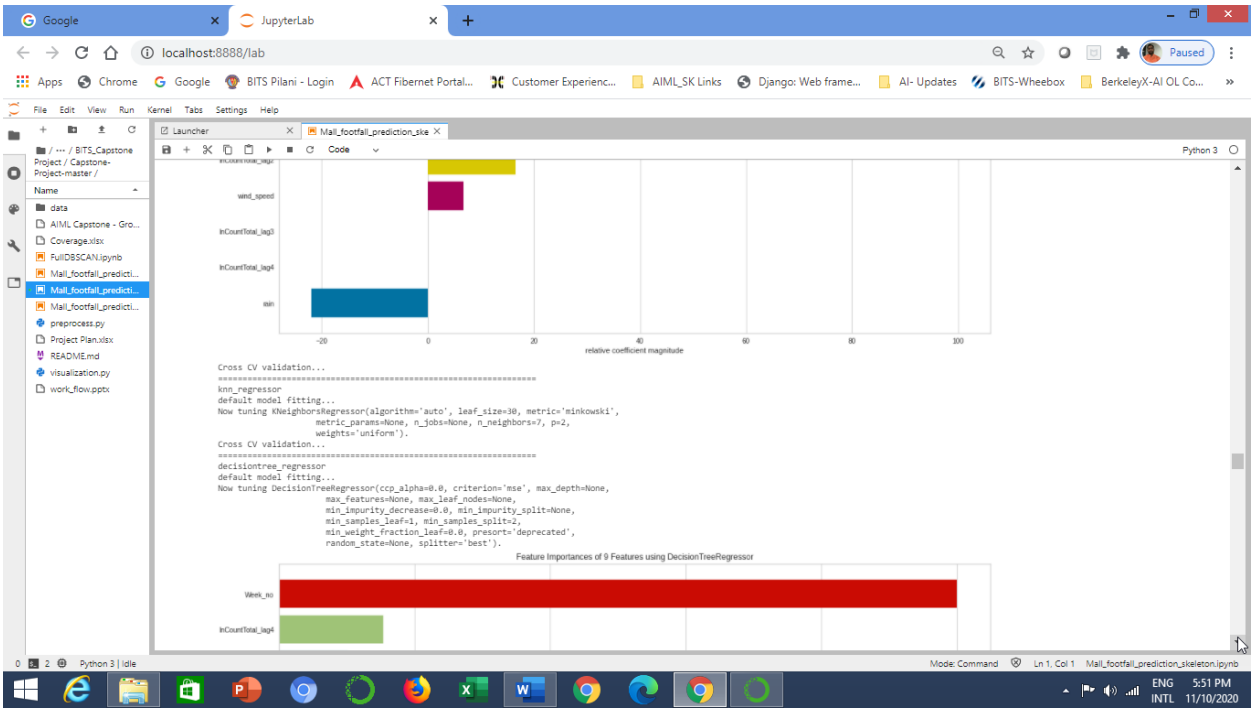

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
wind_speed
lnCountTotal_lag4
min
relative coefficient magnitude
Cross CV validation...
ridgeRegressor
default model fitting...
Now tuning Ridge(alpha=1.0, copy_X=True, fit_intercept=True, max_iter=None,
normalize=False, random_state=None, solver='auto', tol=0.001).
Feature Importances of 9 Features using Ridge
Week_no
Year
mean_temp
lnCountTotal_lag1
lnCountTotal_lag2
lnCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:57 PM
11/10/2020

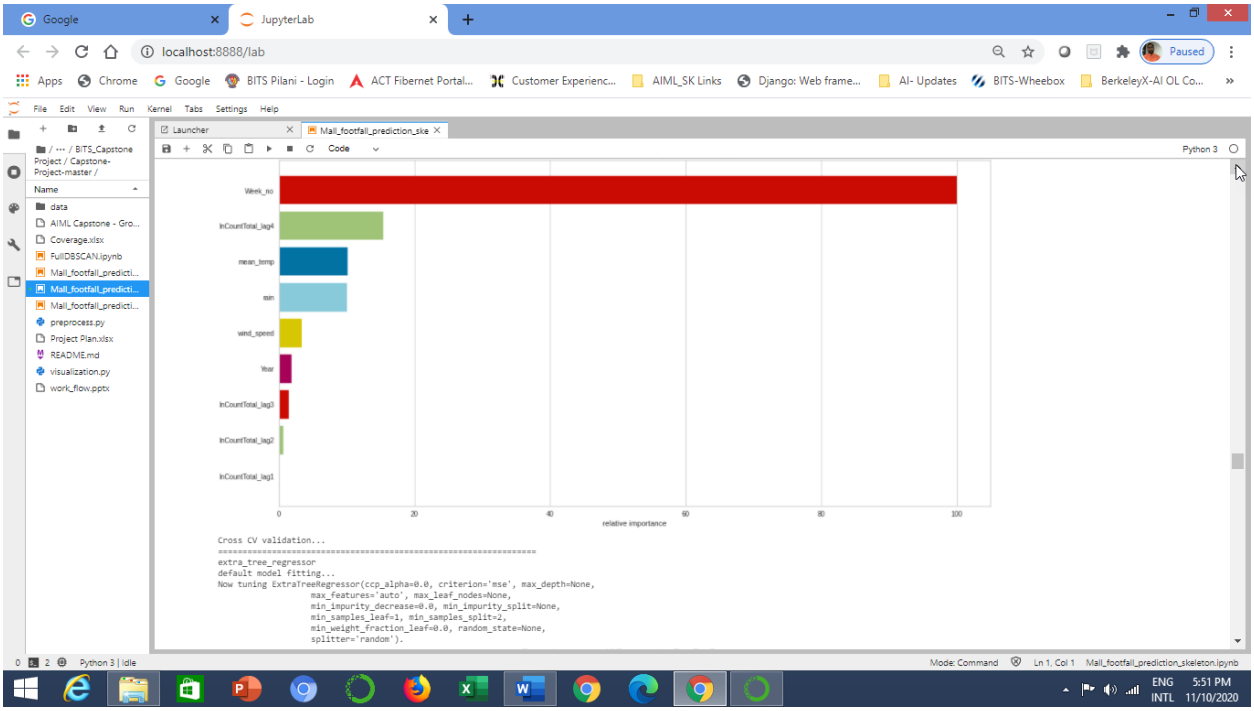

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
lnCountTotal_lag3
lnCountTotal_lag4
wind_speed
min
relative coefficient magnitude
Cross CV validation...
elastic_net
default model fitting...
Now tuning ElasticNet(alpha=1.0, copy_X=True, fit_intercept=True, l1_ratio=0.5,
max_iter=1000, normalize=False, positive=False, precompute=False,
random_state=None, selection='cyclic', tol=0.0001, warm_start=False).
Feature Importances of 9 Features using ElasticNet
Week_no
Year
mean_temp
lnCountTotal_lag1
lnCountTotal_lag2
wind_speed
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
4:57 PM
11/10/2020

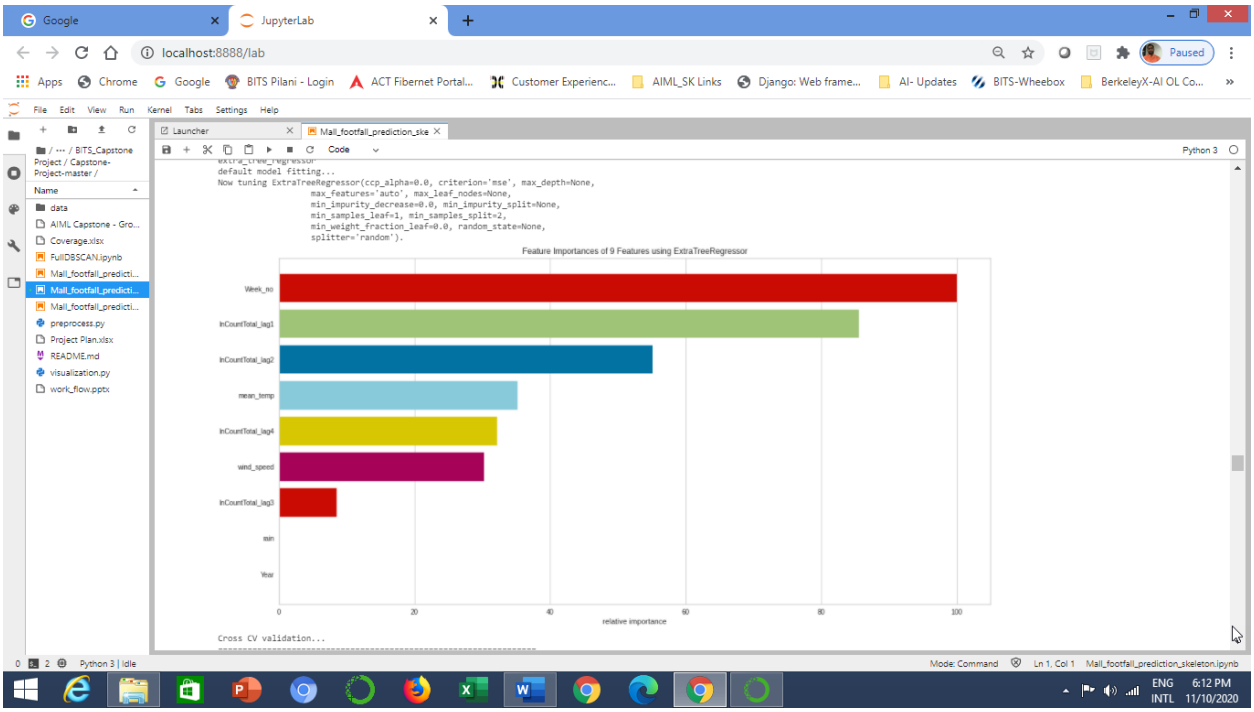

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative coefficient magnitude
Cross CV validation...
huberRegressor
default model fitting...
Now tuning HuberRegressor(alpha=0.0001, epsilon=1.35, fit_intercept=True, max_iter=100,
tol=1e-05, warm_start=False).
Feature Importances of 9 Features using HuberRegressor
Week_no
Year
wind_speed
mean_temp
lnCountTotal_lag1
lnCountTotal_lag2
lnCountTotal_lag3
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
4:58 PM
11/10/2020

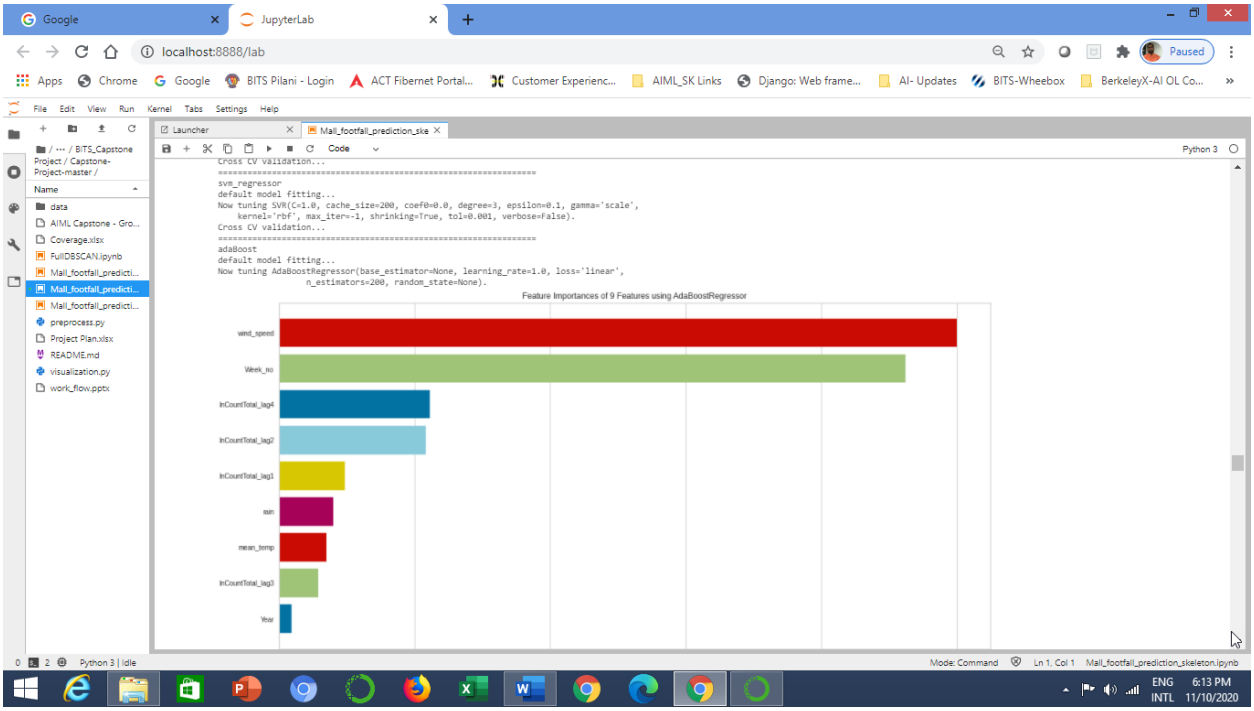

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
relative coefficient magnitude
Cross CV validation...
passive_aggressor
default model fitting...
Now tuning PassiveAggressiveRegressor(C=1.0, average=False, early_stopping=False,
epsilon=0.1, fit_intercept=True,
loss='epsilon_insensitive', max_iter=1000,
n_iter_no_change=5, random_state=None, shuffle=True,
tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 9 Features using PassiveAggressiveRegressor
lnCountTotal_lag1
lnCountTotal_lag2
mean_temp
Week_no
lnCountTotal_lag3
Year
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
4:59 PM
11/10/2020

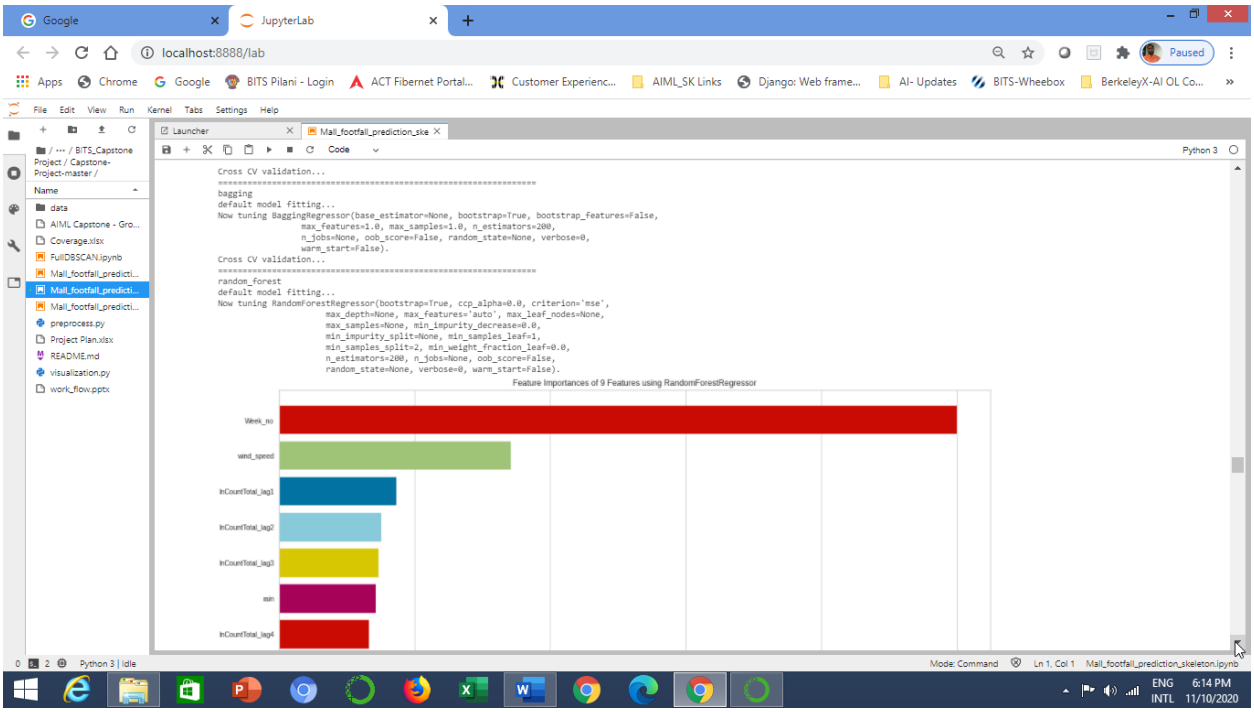

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Year
lnCountTotal_lag4
wind_speed
min
relative coefficient magnitude
Cross CV validation...
sgd
default model fitting...
Now tuning SGDRegressor(alpha=0.0001, average=False, early_stopping=False, epsilon=0.1,
eta0=0.01, fit_intercept=True, l1_ratio=0.15,
learning_rate='invscaling', loss='squared_loss', max_iter=1000,
n_iter_no_change=5, penalty='l2', power_t=0.25, random_state=None,
shuffle=True, tol=0.001, validation_fraction=0.1, verbose=0,
warm_start=False).
Feature Importances of 9 Features using SGDRegressor
Week_no
Year
mean_temp
lnCountTotal_lag1
lnCountTotal_lag2
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
5:50 PM
11/10/2020

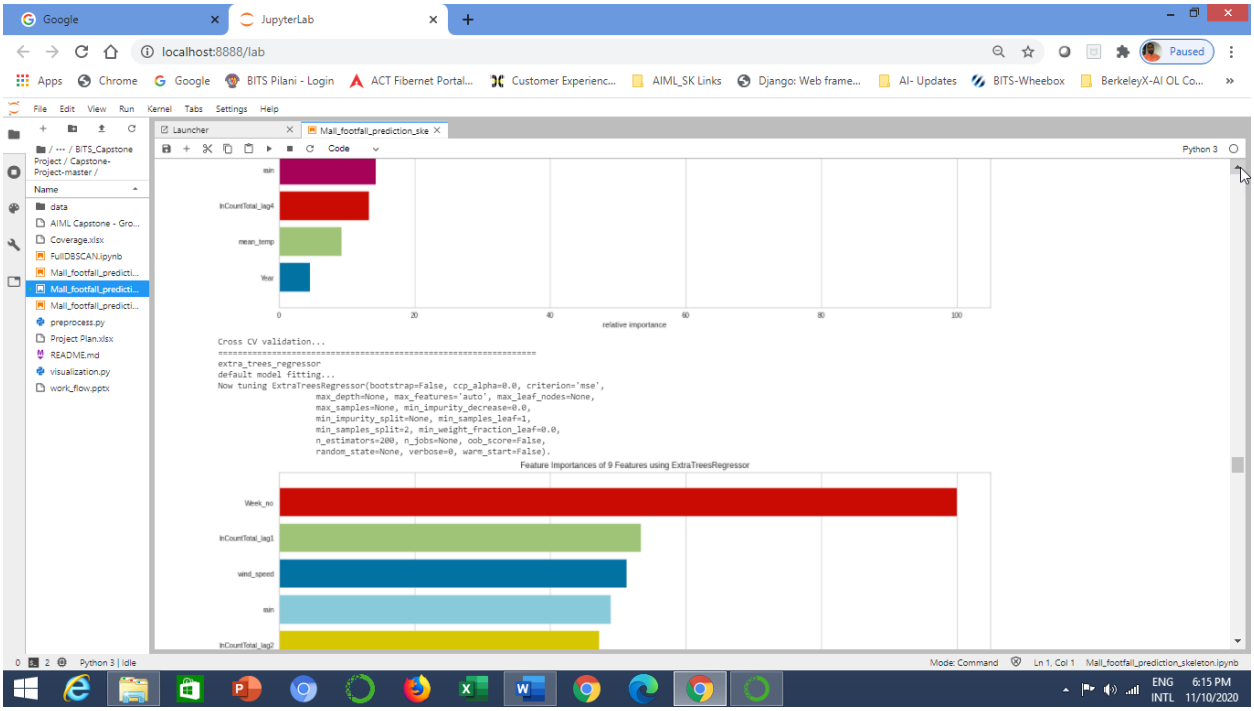

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
wind_speed
lnCountTotal_lag3
lnCountTotal_lag4
min
relative coefficient magnitude
Cross CV validation...
knn_regressor
default model fitting...
Now tuning KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
metric_params=None, n_jobs=None, n_neighbors=7, p=2,
weights='uniform').
Cross CV validation...
decisiontree_regressor
default model fitting...
Now tuning DecisionTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, presort='deprecated',
random_state=None, splitter='best').
Feature Importances of 9 Features using DecisionTreeRegressor
Week_no
lnCountTotal_lag4
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
5:51 PM
11/10/2020

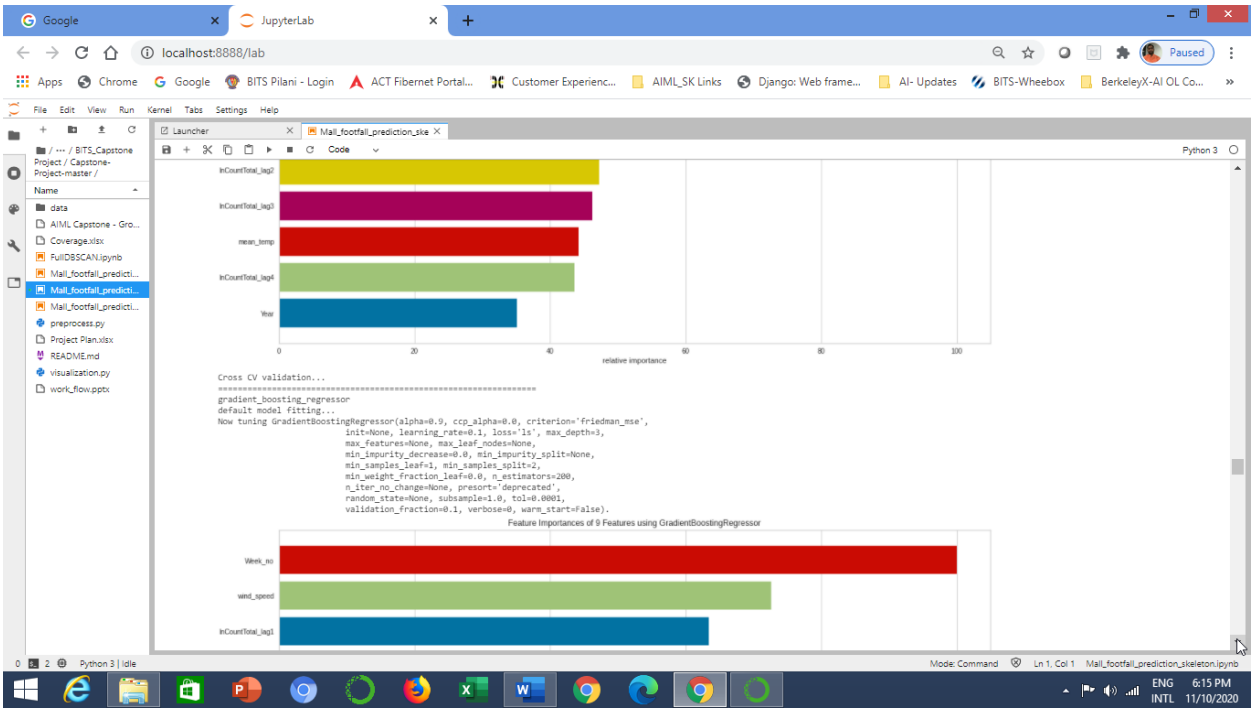

Week_no
lnCountTotal_lag4
mean_temp
min
wind_speed
Year
lnCountTotal_lag3
lnCountTotal_lag2
lnCountTotal_lag1
0
20
40
60
80
100
relative importance
Cross CV validation...
extra_tree_regressor
default model fitting...
Now tuning ExtraTreeRegressor(ccp_alpha=0.0, criterion='mse', max_depth=None,
max_features='auto', max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, random_state=None,
splitter='random').

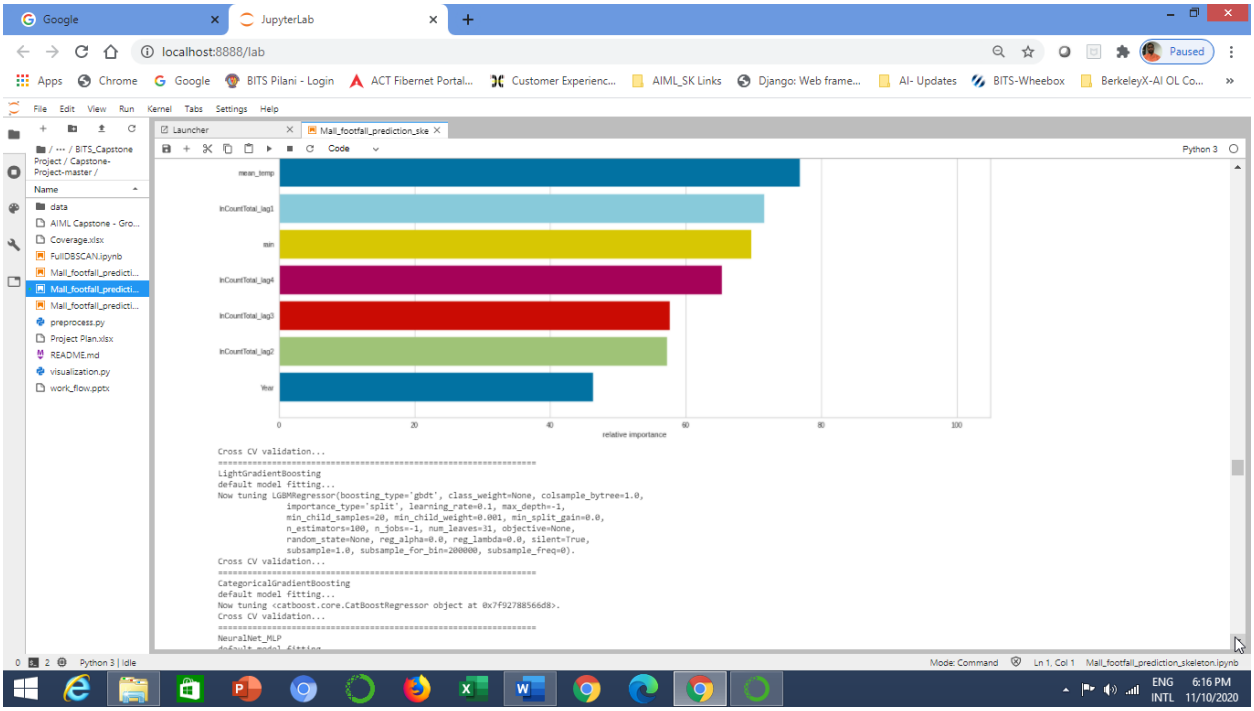

Feature Importances of 9 Features using ExtraTreeRegressor
Week_no
lnCountTotal_lag1
lnCountTotal_lag2
mean_temp
lnCountTotal_lag4
wind_speed
lnCountTotal_lag3
min
Year
0
20
40
60
80
100
relative importance
Cross CV validation...

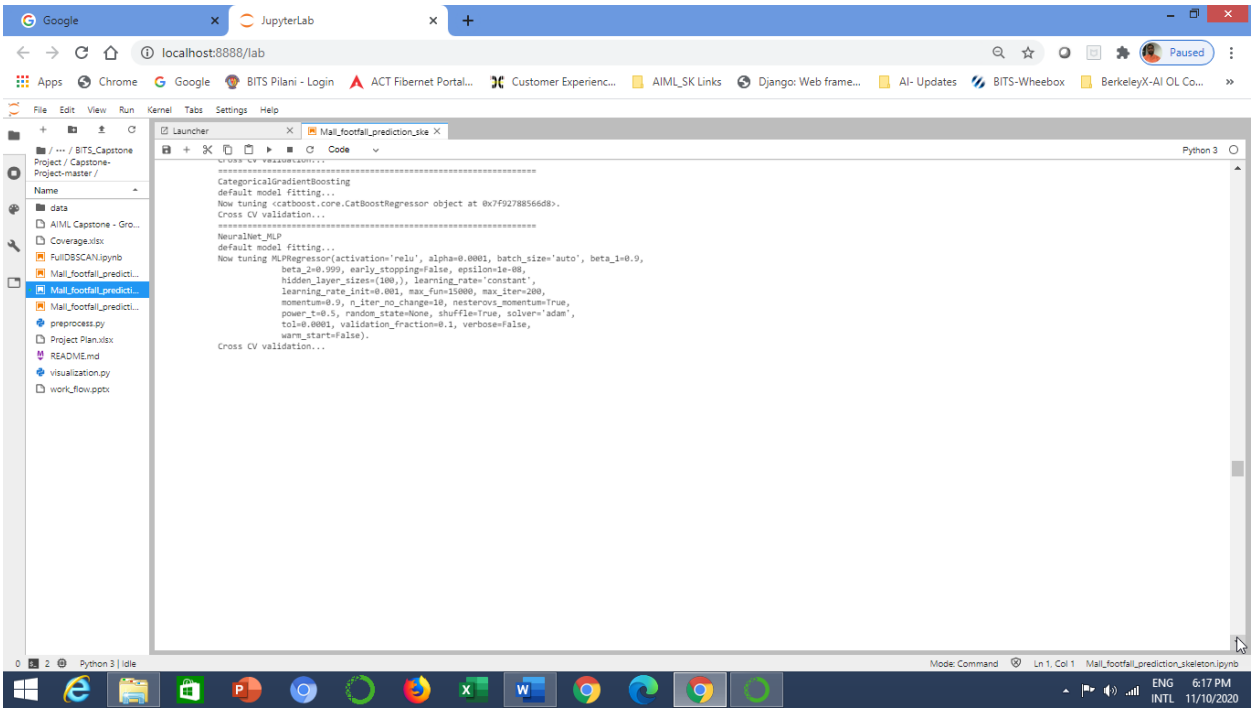

Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Cross CV validation...
svm_regressor
default model fitting...
Now tuning SVR(C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale',
kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False).
Cross CV validation...
adaBoost
default model fitting...
Now tuning AdaBoostRegressor(base_estimator=None, learning_rate=1.0, loss='linear',
n_estimators=200, random_state=None).
Feature Importances of 9 Features using AdaBoostRegressor
wind_speed
Week_no
lnCountTotal_lag4
lnCountTotal_lag2
lnCountTotal_lag1
min
mean_temp
lnCountTotal_lag3
Year
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:13 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Cross CV validation...
bagging
default model fitting...
Now tuning BaggingRegressor(base_estimator=None, bootstrap=True, bootstrap_features=False,
max_features=1.0, max_samples=1.0, n_estimators=200,
n_jobs=None, oob_score=False, random_state=None, verbose=0,
warm_start=False).
Cross CV validation...
random_forest
default model fitting...
Now tuning RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 9 Features using RandomForestRegressor
Week_no
wind_speed
lnCountTotal_lag1
lnCountTotal_lag2
lnCountTotal_lag3
min
lnCountTotal_lag4
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:14 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
min
lnCountTotal_lag4
mean_temp
Year
relative importance
Cross CV validation...
extra_trees_regressor
default model fitting...
Now tuning ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
max_depth=None, max_features='auto', max_leaf_nodes=None,
max_samples=None, min_impurity_decrease=0.0,
min_impurity_split=None, min_samples_leaf=1,
min_samples_split=2, min_weight_fraction_leaf=0.0,
n_estimators=200, n_jobs=None, oob_score=False,
random_state=None, verbose=0, warm_start=False).
Feature Importances of 9 Features using ExtraTreesRegressor
Week_no
lnCountTotal_lag1
wind_speed
min
lnCountTotal_lag2
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:15 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
lnCountTotal_lag2
lnCountTotal_lag3
mean_temp
lnCountTotal_lag4
Year
relative importance
Cross CV validation...
gradient_boosting_regressor
default model fitting...
Now tuning GradientBoostingRegressor(alpha=0.9, ccp_alpha=0.0, criterion='friedman_mse',
init=None, learning_rate=0.1, loss='ls', max_depth=3,
max_features=None, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=1, min_samples_split=2,
min_weight_fraction_leaf=0.0, n_estimators=200,
n_iter_no_change=None, presort='deprecated',
random_state=None, subsample=1.0, tol=0.0001,
validation_fraction=0.1, verbose=0, warm_start=False).
Feature Importances of 9 Features using GradientBoostingRegressor
Week_no
wind_speed
lnCountTotal_lag1
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:15 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
mean_temp
lnCountTotal_lag1
min
lnCountTotal_lag4
lnCountTotal_lag3
lnCountTotal_lag2
Year
0
20
40
60
80
100
relative importance
Cross CV validation...
LightGradientBoosting
default model fitting...
Now tuning LGBMRegressor(boosting_type='gbdt', class_weight=None, colsample_bytree=1.0,
importance_type='split', learning_rate=0.1, max_depth=-1,
min_child_samples=20, min_child_weight=0.001, min_split_gain=0.0,
n_estimators=100, n_jobs=-1, num_leaves=31, objective=None,
random_state=None, reg_alpha=0.0, reg_lambda=0.0, silent=True,
subsample=1.0, subsample_for_bin=200000, subsample_freq=0).
Cross CV validation...
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f92788566d8>.
Cross CV validation...
NeuralNet_MLP
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:16 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
CategoricalGradientBoosting
default model fitting...
Now tuning <catboost.core.CatBoostRegressor object at 0x7f92788566d8>.
Cross CV validation...
NeuralNet_MLP
default model fitting...
Now tuning MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
beta_2=0.999, early_stopping=False, epsilon=1e-08,
hidden_layer_sizes=(100,), learning_rate='constant',
learning_rate_init=0.001, max_fun=15000, max_iter=200,
momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
power_t=0.5, random_state=None, shuffle=True, solver='adam',
tol=0.0001, validation_fraction=0.1, verbose=False,
warm_start=False).
Cross CV validation...
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:17 PM
11/10/2020

Google | JupyterLab | localhost:8888/lab

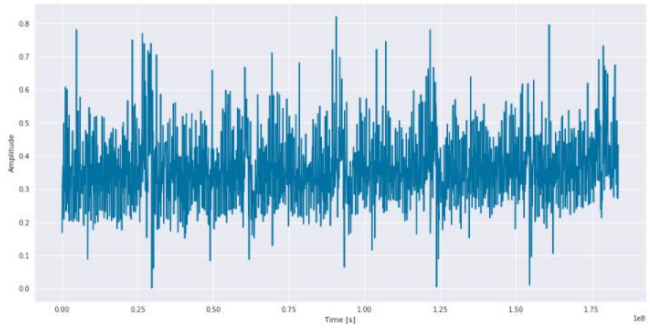
```
loss_weekly_df.sort_values('mape').head(15)
```

	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
16	passive_aggressor_weekly_data_rscv_model	7.422109	0.074722	-0.046541	0.204578	-0.008566	0.273353	0.386368	-0.051940	-44.849591	-32.660832
10	elastic_net_weekly_data_rscv_model	7.604363	0.086880	-0.095026	0.216313	-0.024186	0.294754	0.286523	-0.223103	-41.080706	-28.891947
7	ridgeRegressor_weekly_data_rscv_model	7.716040	0.087065	-0.082116	0.218258	-0.019288	0.295067	0.285005	-0.225705	-41.027590	-28.838831
18	sgd_weekly_data_default_model	7.726780	0.079596	-0.023588	0.210905	0.000700	0.282128	0.346339	-0.120563	-43.269722	-31.080963
0	linear_regression_weekly_data_default_model	7.728431	0.086534	-0.079074	0.218293	-0.018129	0.294167	0.289362	-0.218236	-41.180393	-28.991635
42	extra_trees_regressor_weekly_data_default_model	7.781591	0.088191	-0.066643	0.217247	-0.014674	0.296970	0.275757	-0.241559	-40.706287	-28.517529
1	linear_regression_weekly_data_rscv_model	7.813878	0.095666	-0.119821	0.223899	-0.033625	0.309299	0.214370	-0.346795	-38.672300	-26.483541
12	huberRegressor_weekly_data_default_model	7.971273	0.099953	-0.109871	0.227959	-0.029165	0.316154	0.179162	-0.407150	-37.576330	-25.387572
4	lassoRegressor_weekly_data_rscv_model	7.988315	0.102159	-0.129113	0.229825	-0.036573	0.319624	0.161045	-0.438208	-37.030541	-24.841783
13	huberRegressor_weekly_data_rscv_model	8.028735	0.100692	-0.107223	0.229333	-0.028012	0.317320	0.173096	-0.417551	-37.392231	-25.203473
19	sgd_weekly_data_rscv_model	8.085729	0.092049	-0.116343	0.230060	-0.032379	0.303395	0.244075	-0.295871	-39.635916	-27.447158
55	CategoricalGradientBoosting_weekly_data_rscv_m...	8.115489	0.110251	-0.123155	0.233457	-0.034995	0.332041	0.094594	-0.552125	-35.124865	-22.936106
58	NeuralNet_MLP_weekly_data_rscv_model	8.129184	0.092107	-0.127029	0.229737	-0.037574	0.303491	0.243600	-0.296686	-39.620193	-27.431435
34	adaBoost_weekly_data_rscv_model	8.185287	0.104819	-0.114693	0.234434	-0.031501	0.323757	0.139207	-0.475646	-36.388095	-24.199337
31	svm_regressor_weekly_data_rscv_model	8.306406	0.104254	-0.110780	0.235685	-0.029748	0.322885	0.143840	-0.467703	-36.523021	-24.334262

```
loss_weekly_df.sort_values('mape').to_csv("weekly_data_model_loss_values.csv")
```

7.2 Model training with LSTM

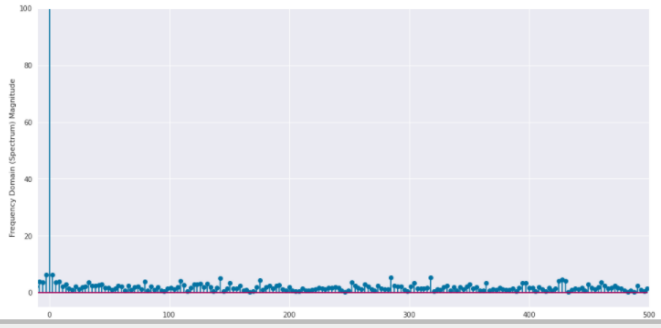
```
t = np.linspace(0,data_final.shape[0]*24*60*60, data_final.shape[0])
s = data_final['InCountTotal'].values
sns.set_style("darkgrid")
plt.ylabel("Amplitude")
plt.xlabel("Time [s]")
plt.plot(t, s)
plt.show()
```




```
#Applying Fourier Transform
```

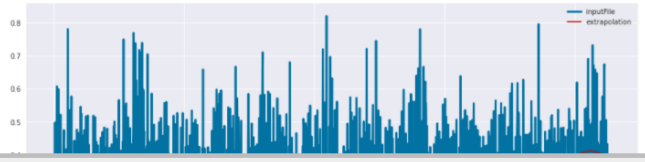
```
#Applying Fourier Transform
fft = fftpack.fft(s[0:365])

f_s = 1000
freqs = fftpack.fftfreq(len(s[0:365])) * f_s
fig, ax = plt.subplots()

ax.stem(freqs, np.abs(fft))
ax.set_xlabel('Frequency in Hertz [Hz]')
ax.set_ylabel('Frequency Domain (Spectrum) Magnitude')
ax.set_xlim(-10, f_s / 2)
ax.set_ylim(-5, 100)
```

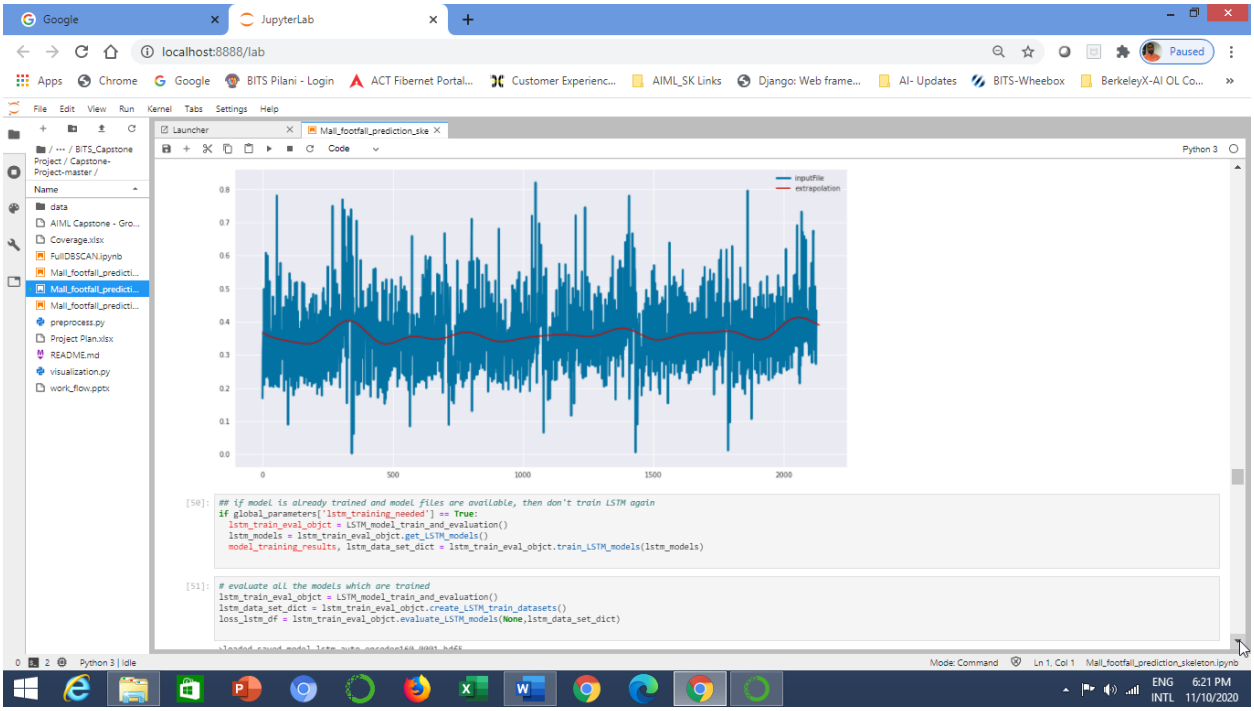
```
(-5.0, 100.0)
```

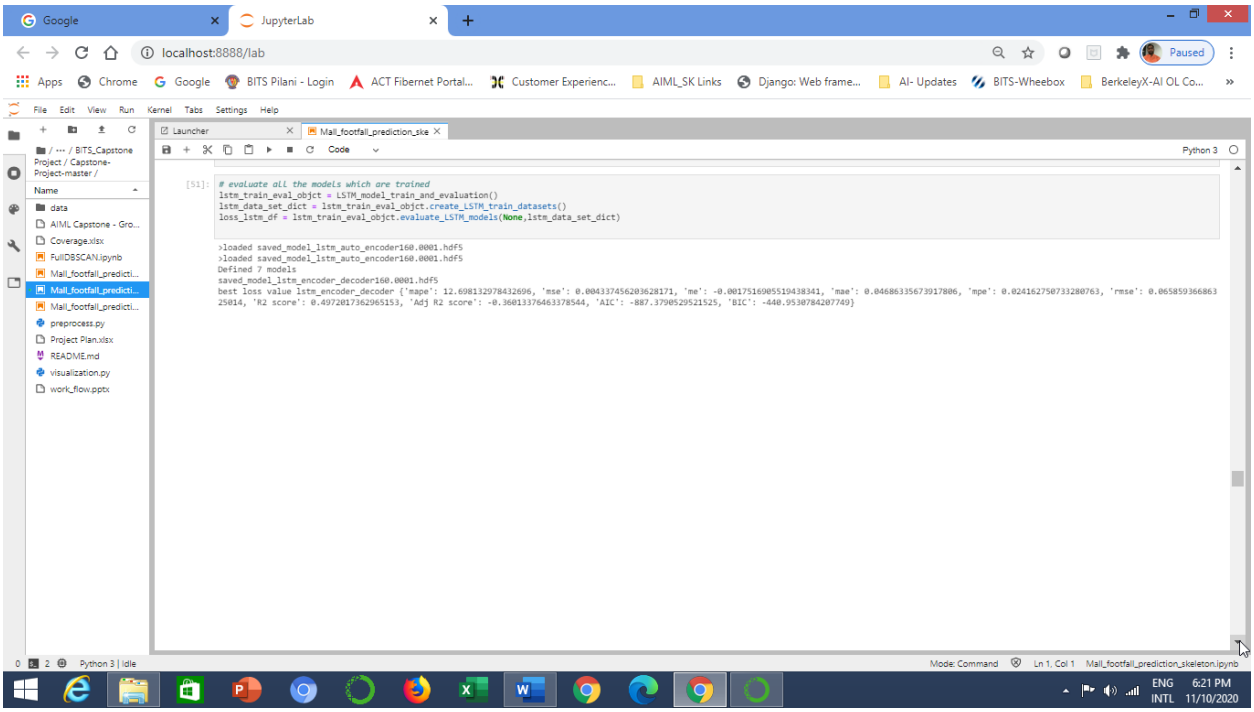


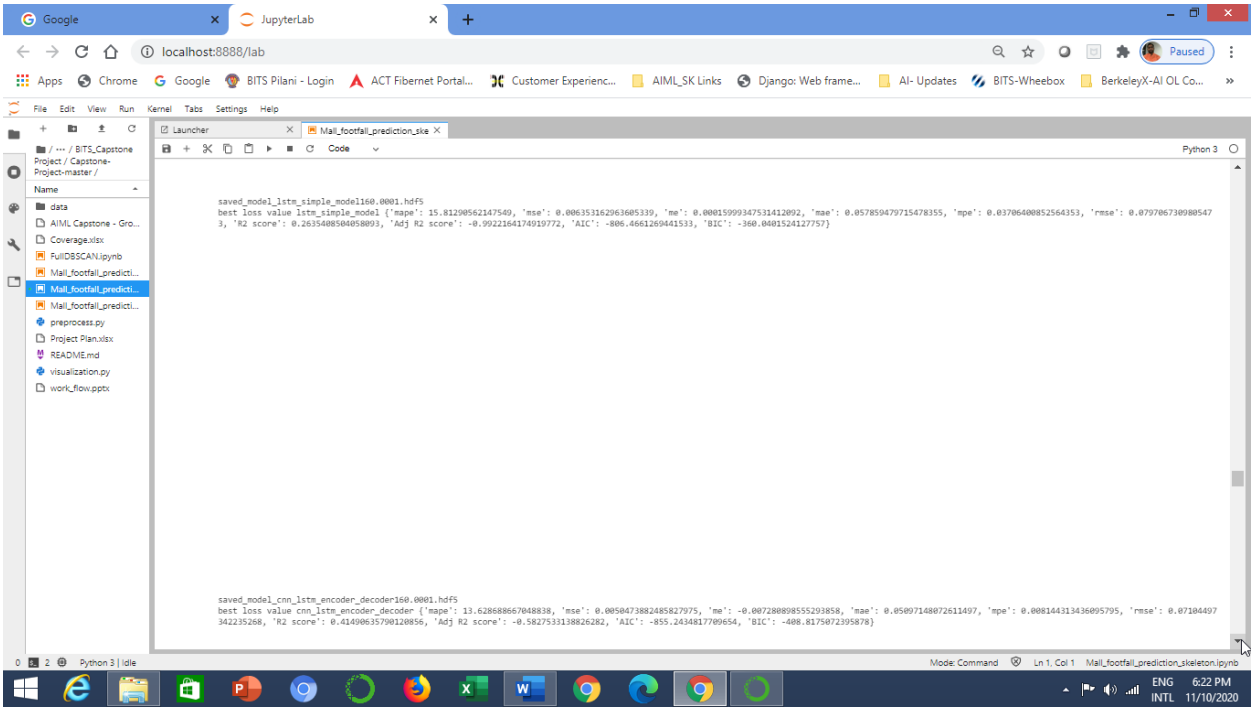

```
def fourierExtrapolation(x, n_predict):
    n = x.size
    n_harm = 10                     # number of harmonics in model
    t = np.arange(0, n)
    p = np.polyfit(t, x, 1)         # find linear trend in x
    x_notrend = x - p[0] * t        # detrended x
    x_freqdom = fftpack.fft(x_notrend)  # detrended x in frequency domain
    f = fftpack.fftfreq(n)              # frequencies
    indexes = list(range(n))
    # sort indexes by frequency, lower -> higher
    indexes.sort(key = lambda i: np.abs(f[i]))
    t = np.arange(0, n + n_predict)
    restored_sig = np.zeros(t.size)
    for i in indexes[:1 + n_harm * 2]:
        ampli = np.absolute(x_freqdom[i]) / n   # amplitude
        phase = np.angle(x_freqdom[i])          # phase
        restored_sig += ampli * np.cos(2 * np.pi * f[i] * t + phase)
    return restored_sig + p[0] * t

x = data_final['InCountTotal'].values
extrapolation = fourierExtrapolation(x, 10)
plt.plot(np.arange(0, x.size), x, 'b', label = 'inputFile', linewidth = 3)
plt.plot(np.arange(0, extrapolation.size), extrapolation, 'r', label = 'extrapolation')
plt.legend()
```

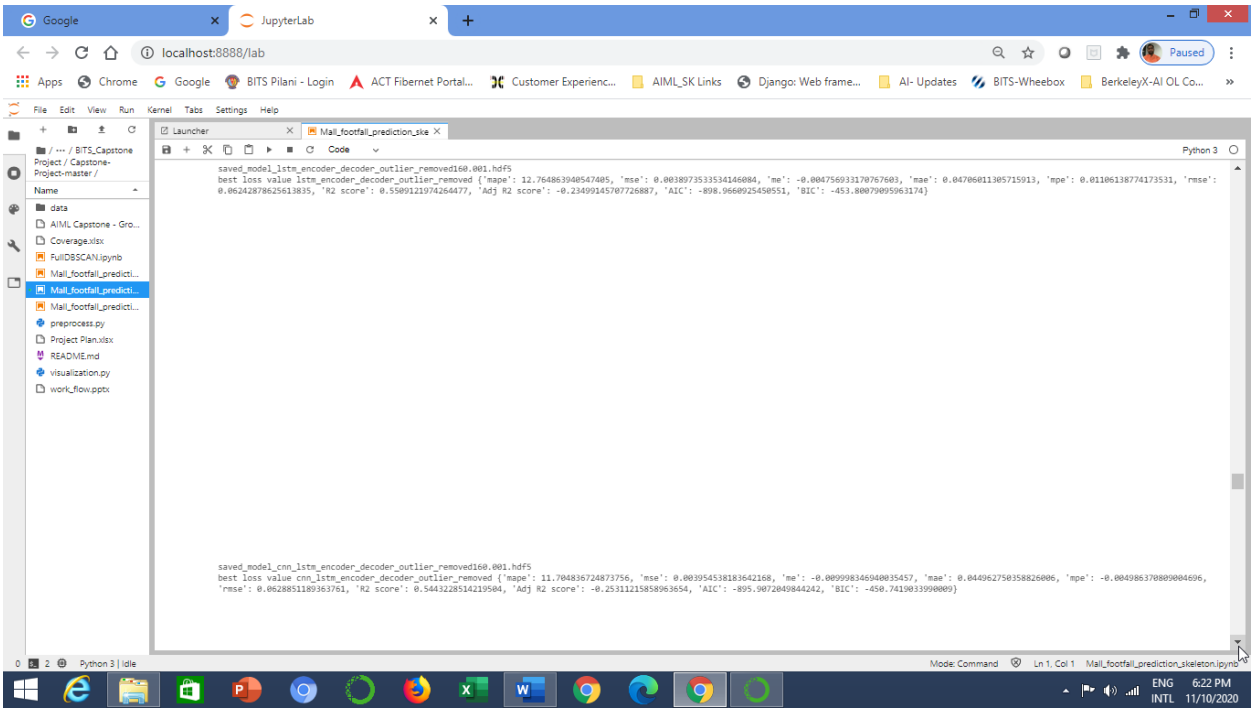
```
<matplotlib.legend.Legend at 0x7f9279e6c710>
```

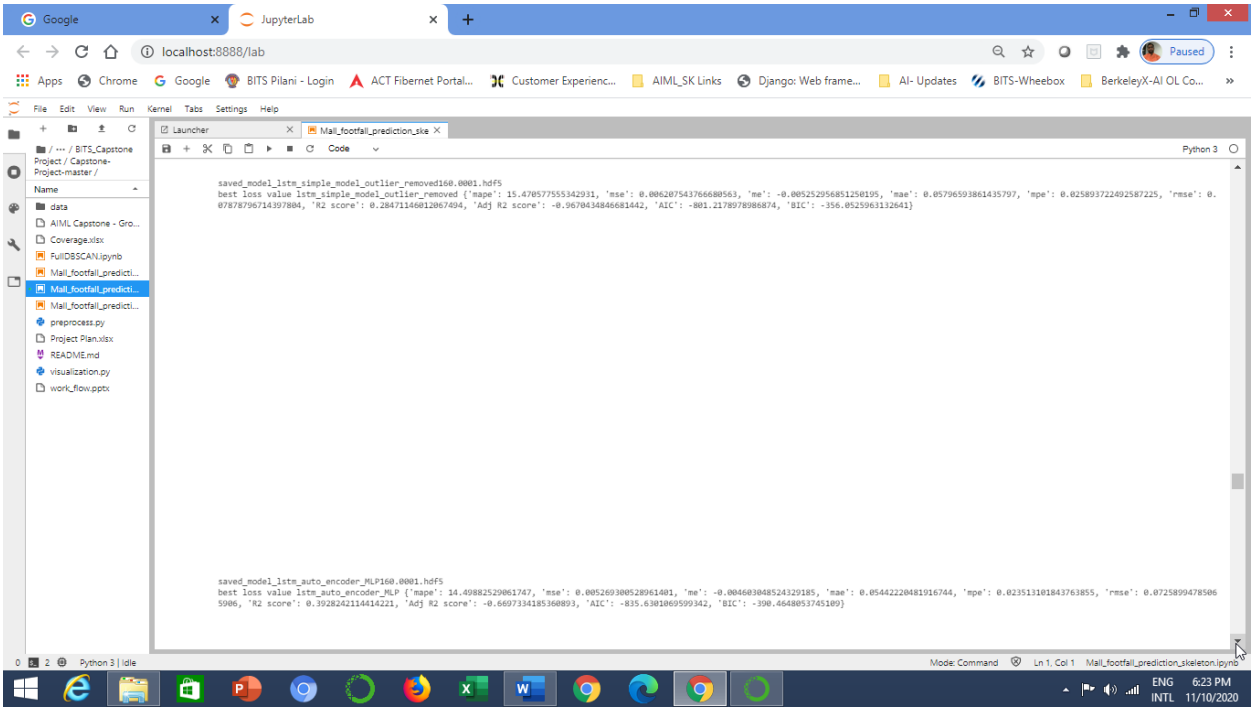


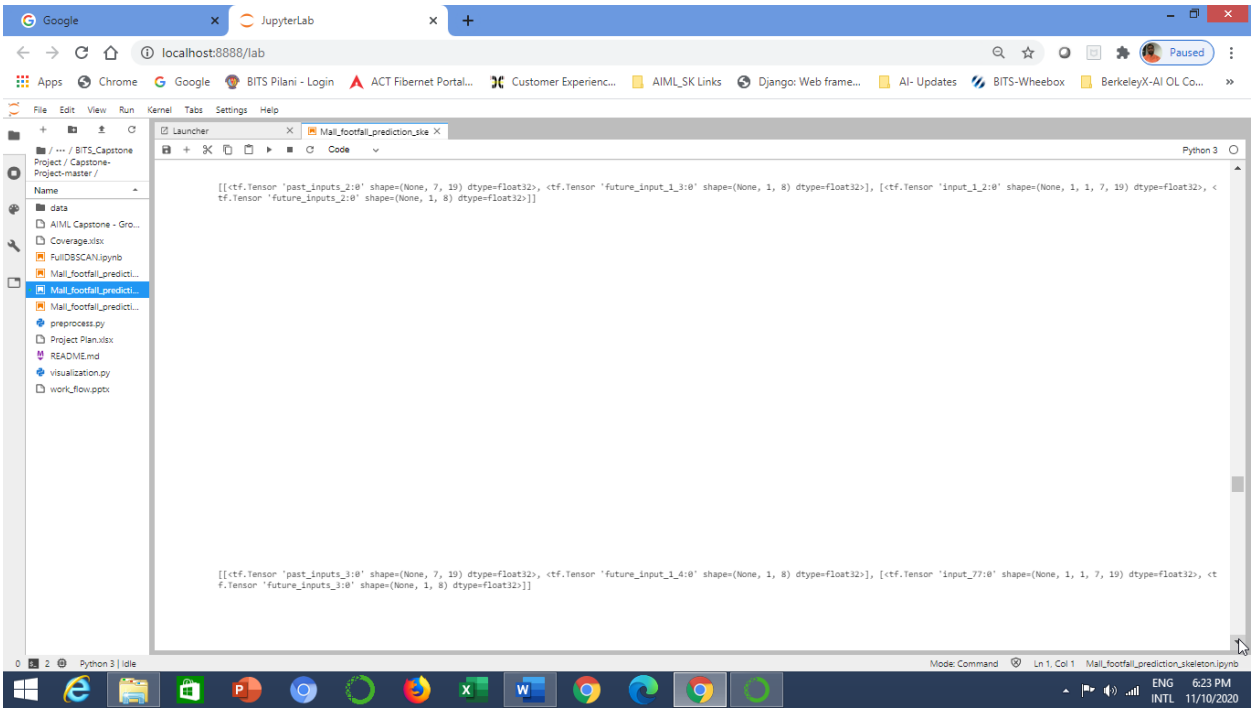

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
inputFile
extrapolation
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
500
1000
1500
2000
[50]:
if model is already trained and model files are available, then don't train LSTM again
if global_parameters['lstm_training_needed'] == True:
lstm_train_eval_objct = LSTM_model_train_and_evaluation()
lstm_models = lstm_train_eval_objct.get_LSTM_models()
model_training_results, lstm_data_set_dict = lstm_train_eval_objct.train_LSTM_models(lstm_models)
[51]:
# evaluate all the models which are trained
lstm_train_eval_objct = LSTM_model_train_and_evaluation()
lstm_data_set_dict = lstm_train_eval_objct.create_LSTM_train_datasets()
loss_lstm_df = lstm_train_eval_objct.evaluate_LSTM_models(None,lstm_data_set_dict)
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:21 PM
11/10/2020

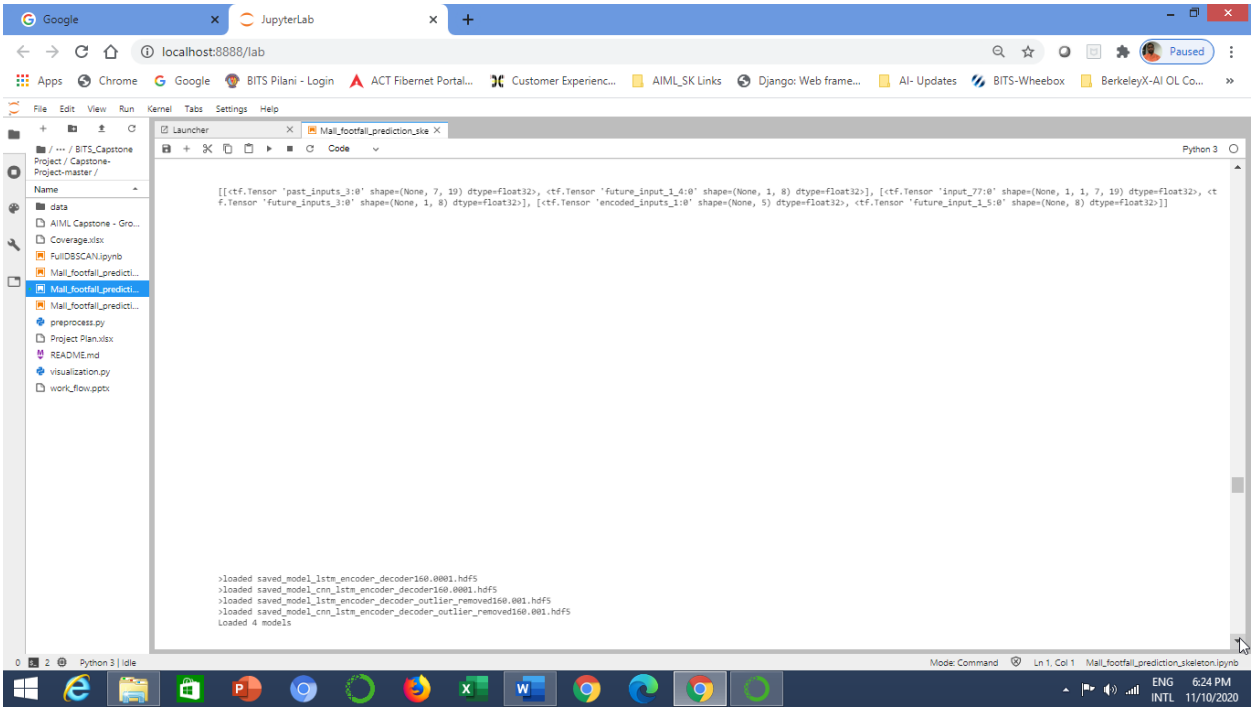

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
[51]:
# evaluate all the models which are trained
lstm_train_eval_objct = LSTM_model_train_and_evaluation()
lstm_data_set_dict = lstm_train_eval_objct.create_LSTM_train_datasets()
loss_lstm_df = lstm_train_eval_objct.evaluate_LSTM_models(None,lstm_data_set_dict)
>loaded saved_model_lstm_auto_encoder160.0001.hdf5
>loaded saved_model_lstm_auto_encoder160.0001.hdf5
Defined 7 models
saved_model_lstm_encoder_decoder160.0001.hdf5
best loss value lstm_encoder_decoder {'mape': 12.698132978432696, 'mse': 0.004337456203628171, 'me': -0.0017516905519438341, 'mae': 0.04686335673917806, 'mpe': 0.024162750733280763, 'rmse': 0.065859366863
25014, 'R2 score': 0.4972017362965153, 'Adj R2 score': -0.36013376463378544, 'AIC': -887.3790529521525, 'BIC': -440.9530784207749}
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:21 PM
11/10/2020

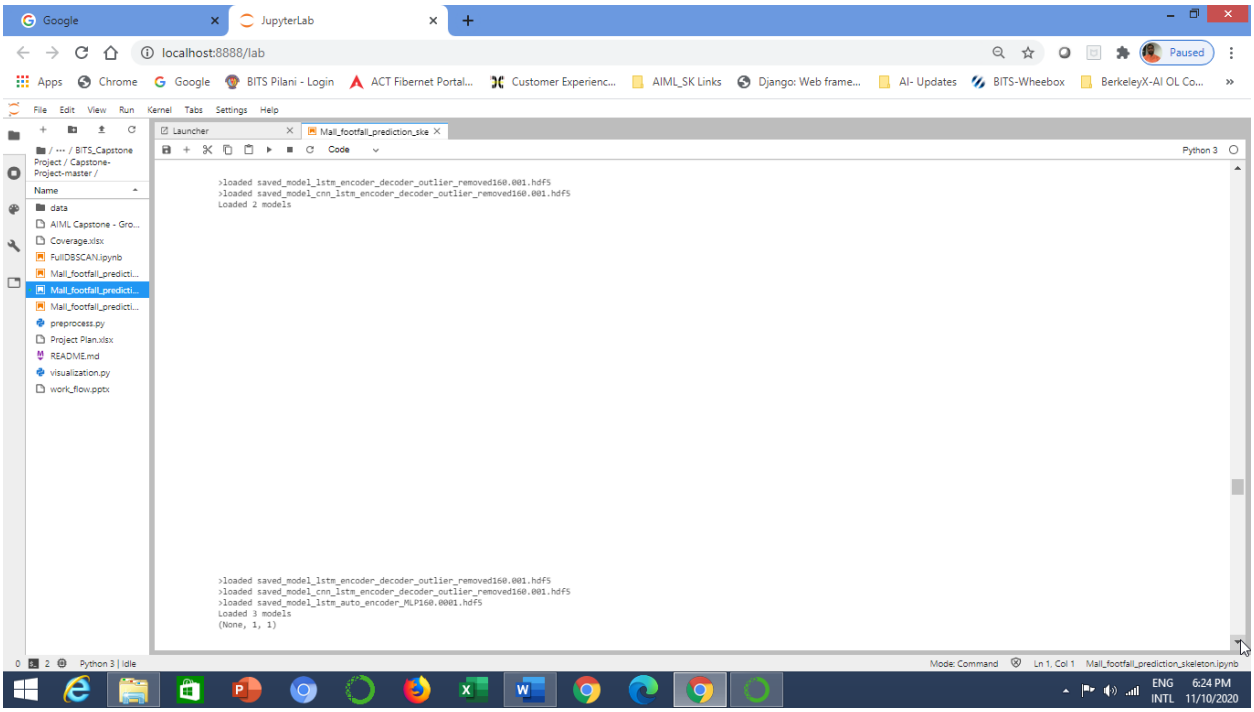

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
saved_model_lstm_simple_model160.0001.hdf5
best loss value lstm_simple_model {'mape': 15.81290562147549, 'mse': 0.006353162963605339, 'me': 0.0001599934753141209 2, 'mae': 0.057859479715478355, 'mpe': 0.03706400852564353, 'rmse': 0.079706730980547 3, 'R2 score': 0.2635408504058093, 'Adj R2 score': -0.9922164174919772, 'AIC': -806.4661269441533, 'BIC': -360.040152412775 7}
saved_model_cnn_lstm_encoder_decoder160.0001.hdf5
best loss value cnn_lstm_encoder_decoder {'mape': 13.628688667048838, 'mse': 0.005047388248582797 5, 'me': -0.007280898555293858, 'mae': 0.05097148072611497, 'mpe': 0.008144313436095795, 'rmse': 0.07104497 342235268, 'R2 score': 0.41490635790120856, 'Adj R2 score': -0.5827533138826282, 'AIC': -855.2434817709654, 'BIC': -408.8175072395878}
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:22 PM 11/10/2020

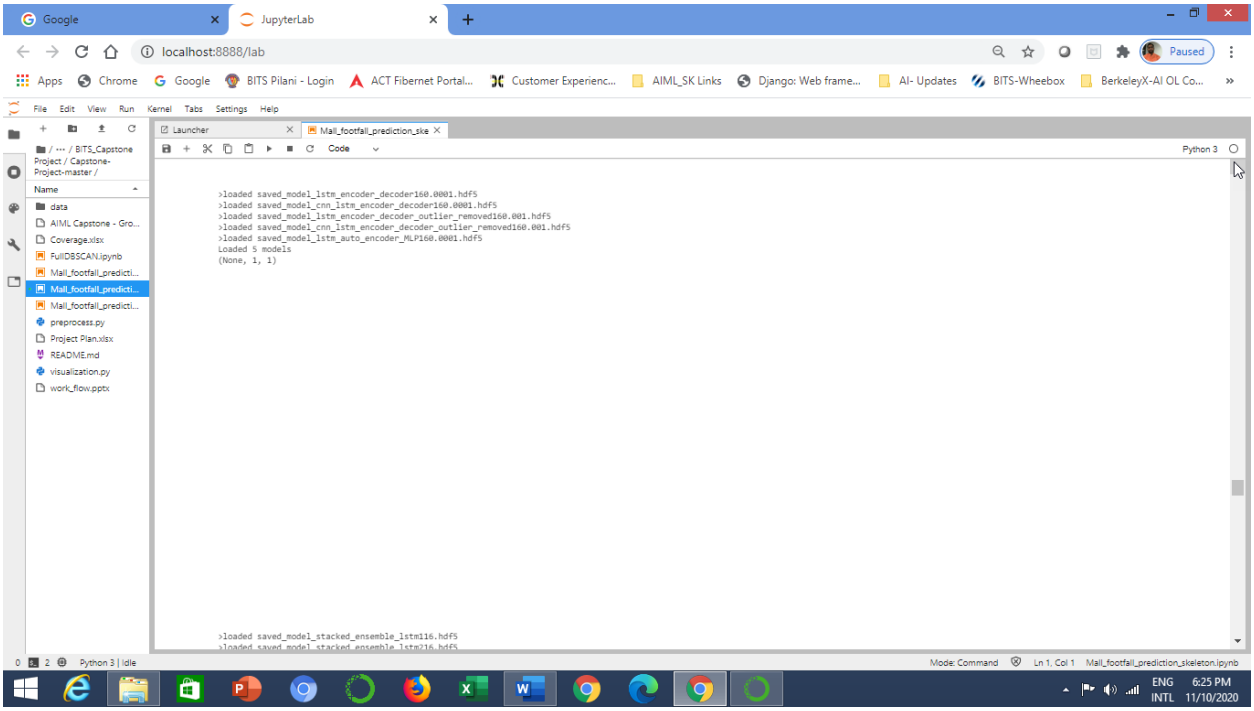

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
saved_model_lstm_encoder_decoder_outlier_removed160.001.hdf5
best loss value lstm_encoder_decoder_outlier_removed {'mape': 12.764863940547405, 'mse': 0.003897353353414608 4, 'me': -0.004756933170767603, 'mae': 0.04706011305715913, 'mpe': 0.011061387747 3531, 'rmse': 0.06242878625613835, 'R2 score': 0.5509121974264477, 'Adj R2 score': -0.2349914570772688 7, 'AIC': -898.9660925450551, 'BIC': -453.80079095963174}
saved_model_cnn_lstm_encoder_decoder_outlier_removed160.001.hdf5
best loss value cnn_lstm_encoder_decoder_outlier_removed {'mape': 11.704836724873756, 'mse': 0.003954538183642168, 'me': -0.009998346940035457, 'mae': 0.04496275035882600 6, 'mpe': -0.004986370809004696, 'rmse': 0.06288511189363761, 'R2 score': 0.5443228514219584, 'Adj R2 score': -0.2531121585896365 4, 'AIC': -895.9072049844242, 'BIC': -450.7419033990009}
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:22 PM 11/10/2020

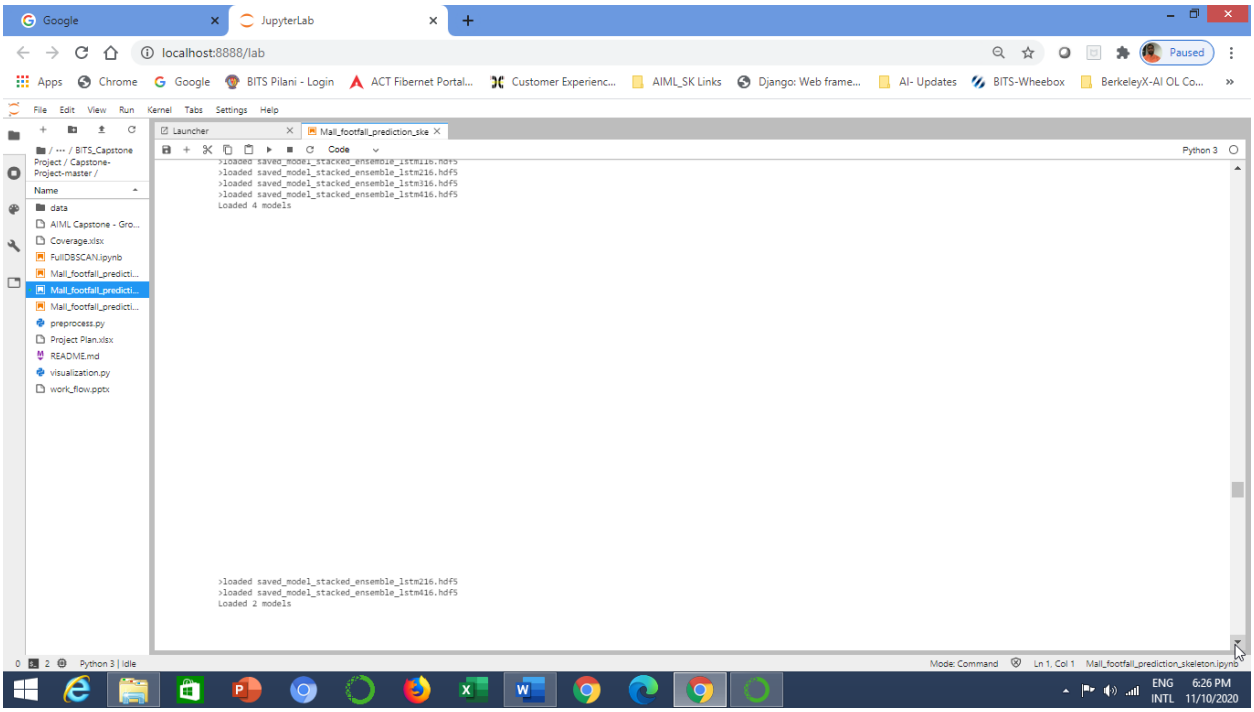

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
saved_model_lstm_simple_model_outlier_removed160.0001.hdf5
best loss value lstm_simple_model_outlier_removed {'mape': 15.470577555342931, 'mse': 0.006207543766680563, 'me': -0.00525295685125019 5, 'mae': 0.05796593861435797, 'mpe': 0.025893722492587225, 'rmse': 0.07878796714397804, 'R2 score': 0.2847114601206749 4, 'Adj R2 score': -0.9670434846681442, 'AIC': -801.2178978986874, 'BIC': -356.0525963132641}
saved_model_lstm_auto_encoder_MLP160.0001.hdf5
best loss value lstm_auto_encoder_MLP {'mape': 14.49882529061747, 'mse': 0.005260300528961401, 'me': -0.004603048524329185, 'mae': 0.05442220481916744, 'mpe': 0.023513101843763855, 'rmse': 0.07258994785065906, 'R2 score': 0.3928242114414221, 'Adj R2 score': -0.6697334185360893, 'AIC': -835.6301069599342, 'BIC': -390.4648053745109}
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:23 PM 11/10/2020

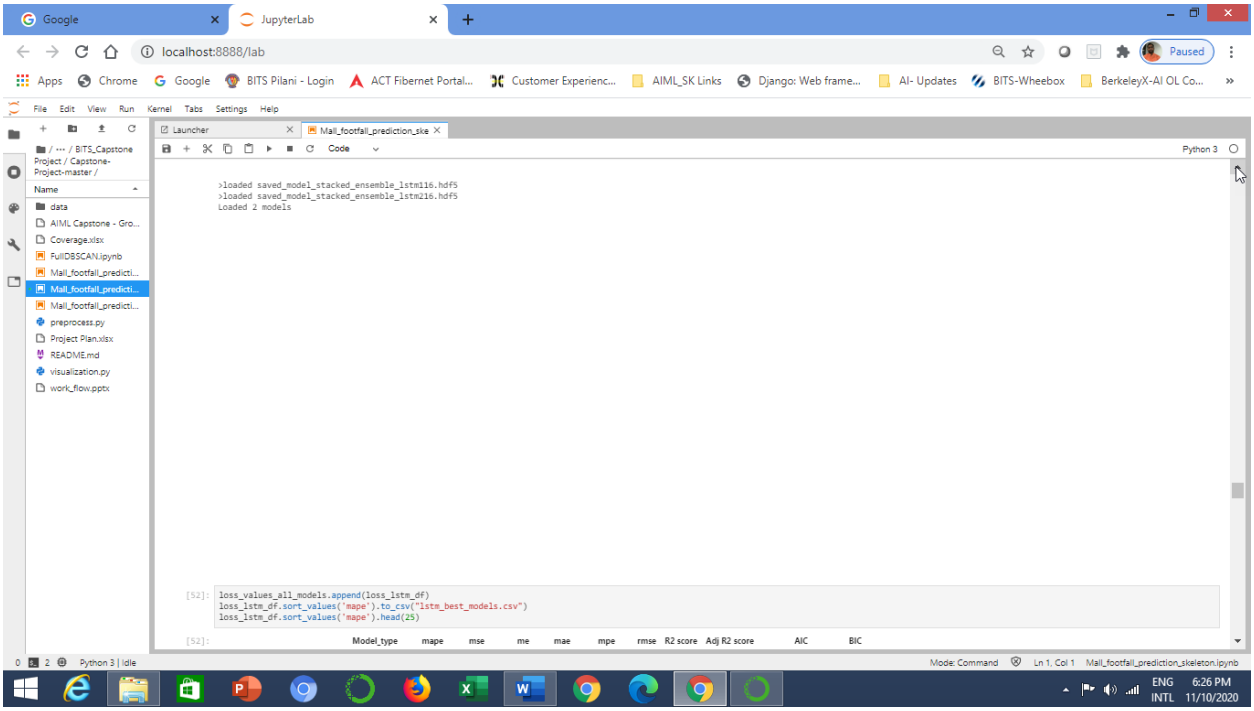

Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
[[<tf.Tensor 'past_inputs_2:0' shape=(None, 7, 19) dtype=float32>, <tf.Tensor 'future_input_1_3:0' shape=(None, 1, 8) dtype=float32>], [<tf.Tensor 'input_1_2:0' shape=(None, 1, 1, 7, 19) dtype=float32>, <tf.Tensor 'future_inputs_2:0' shape=(None, 1, 8) dtype=float32>]]
[[<tf.Tensor 'past_inputs_3:0' shape=(None, 7, 19) dtype=float32>, <tf.Tensor 'future_input_1_4:0' shape=(None, 1, 8) dtype=float32>], [<tf.Tensor 'input_77:0' shape=(None, 1, 1, 7, 19) dtype=float32>, <tf.Tensor 'future_inputs_3:0' shape=(None, 1, 8) dtype=float32>]]
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:23 PM 11/10/2020

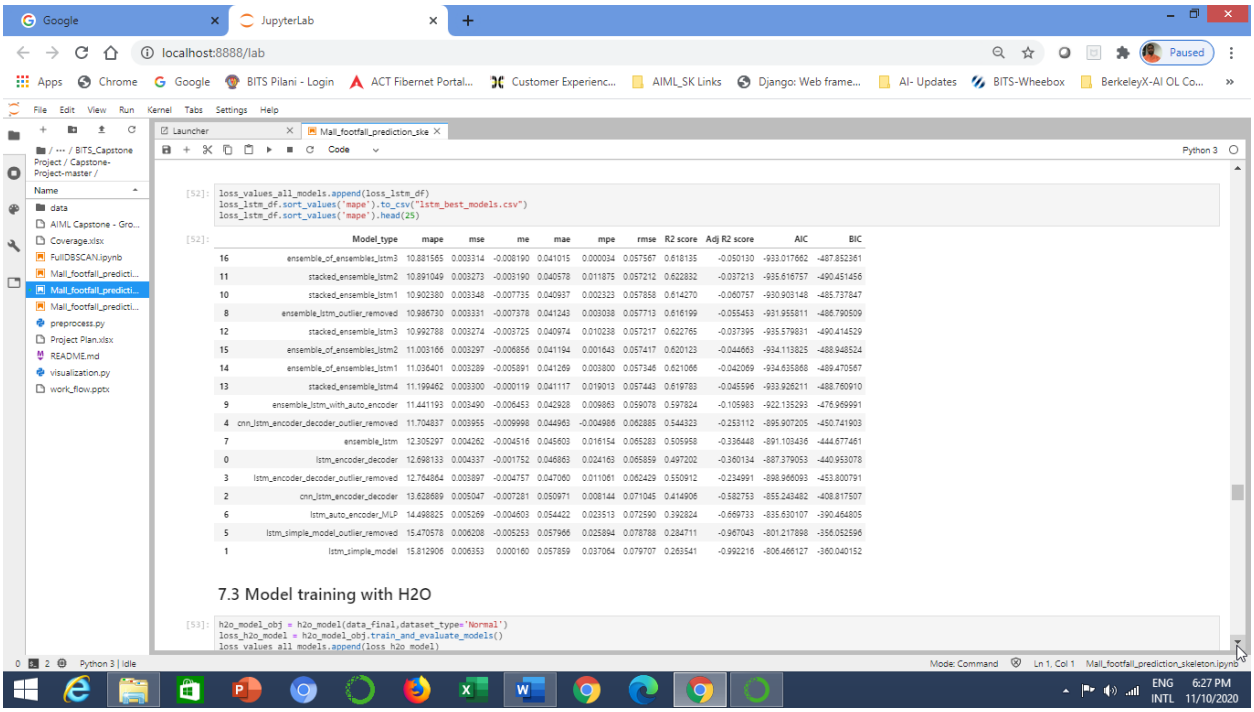

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
[[<tf.Tensor 'past_inputs_3:0' shape=(None, 7, 19) dtype=float32>, <tf.Tensor 'future_input_1_4:0' shape=(None, 1, 8) dtype=float32>], [<tf.Tensor 'input_77:0' shape=(None, 1, 1, 7, 19) dtype=float32>, <tf.Tensor 'future_inputs_3:0' shape=(None, 1, 8) dtype=float32>], [<tf.Tensor 'encoded_inputs_1:0' shape=(None, 5) dtype=float32>, <tf.Tensor 'future_input_1_5:0' shape=(None, 8) dtype=float32>]]
>loaded saved_model_lstm_encoder_decoder160.0001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder160.0001.hdf5
>loaded saved_model_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder_outlier_removed160.001.hdf5
Loaded 4 models
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:24 PM
11/10/2020

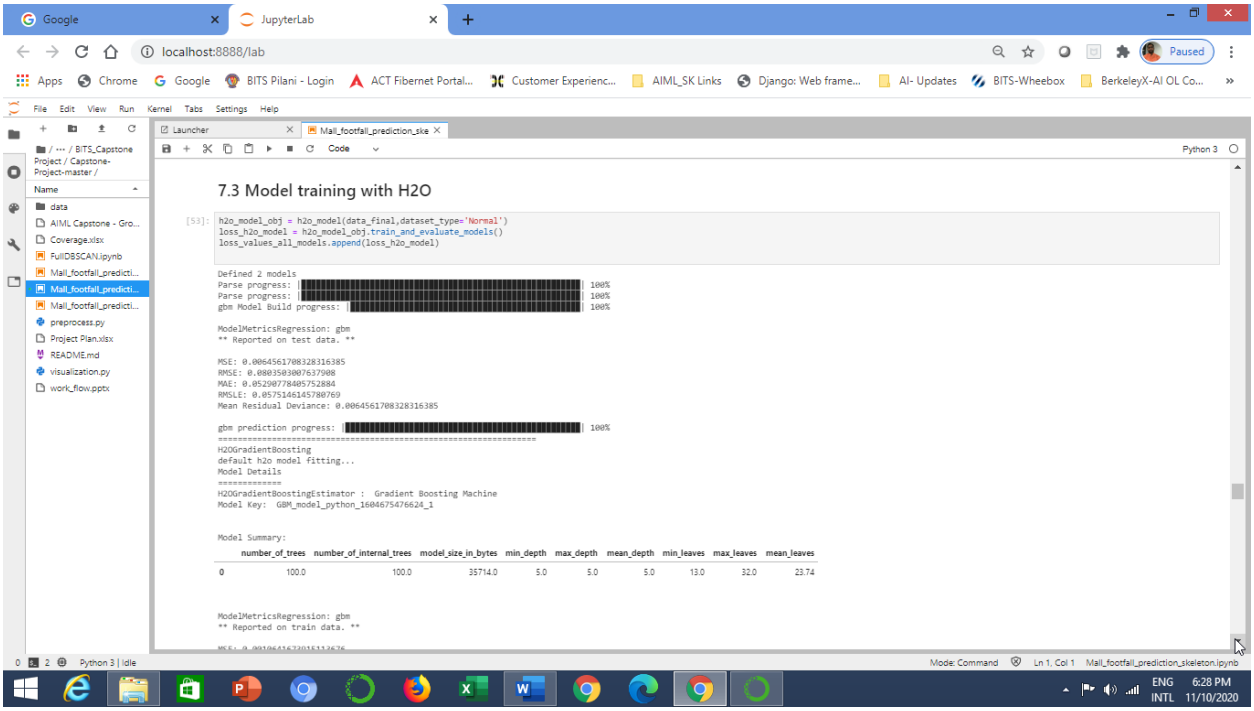

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
>loaded saved_model_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder_outlier_removed160.001.hdf5
Loaded 2 models
>loaded saved_model_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_lstm_auto_encoder_MLP160.0001.hdf5
Loaded 3 models
(None, 1, 1)
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL
6:24 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
>loaded saved_model_lstm_encoder_decoder160.0001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder160.0001.hdf5
>loaded saved_model_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_cnn_lstm_encoder_decoder_outlier_removed160.001.hdf5
>loaded saved_model_lstm_auto_encoder_MLP160.0001.hdf5
Loaded 5 models
(None, 1, 1)
>loaded saved_model_stacked_ensemble_lstm116.hdf5
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:25 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
>loaded saved_model_stacked_ensemble_lstm216.hdf5
>loaded saved_model_stacked_ensemble_lstm316.hdf5
>loaded saved_model_stacked_ensemble_lstm416.hdf5
Loaded 4 models
>loaded saved_model_stacked_ensemble_lstm216.hdf5
>loaded saved_model_stacked_ensemble_lstm416.hdf5
Loaded 2 models
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:26 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
>loaded saved_model_stacked_ensemble_lstm116.hdf5
>loaded saved_model_stacked_ensemble_lstm216.hdf5
Loaded 2 models
[52]:
loss_values_all_models.append(loss_lstm_df)
loss_lstm_df.sort_values('mape').to_csv("lstm_best_models.csv")
loss_lstm_df.sort_values('mape').head(25)
[52]:
Model_type mape mse me mae mpe rmse R2 score Adj R2 score AIC BIC
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:26 PM 11/10/2020


Google
JupyterLab
localhost:8888/lab
Launcher
Mall_footfall_prediction_ske
Code
Python 3
[52]:
loss_values_all_models.append(loss_lstm_df)
loss_lstm_df.sort_values('mape').to_csv("lstm_best_models.csv")
loss_lstm_df.sort_values('mape').head(25)
[52]:
Model_type mape mse me mae mpe rmse R2 score Adj R2 score AIC BIC
16 ensemble_of_ensembles_lstm3 10.881565 0.003314 -0.008190 0.041015 0.000034 0.057567 0.618135 -0.050130 -933.017662 -487.852361
11 stacked_ensemble_lstm2 10.891049 0.003273 -0.003190 0.040578 0.011875 0.057212 0.622832 -0.037213 -935.616757 -490.451456
10 stacked_ensemble_lstm1 10.902380 0.003348 -0.007735 0.040937 0.002323 0.057858 0.614270 -0.060757 -930.903148 -485.737847
8 ensemble_lstm_outlier_removed 10.986730 0.003331 -0.007378 0.041243 0.003038 0.057713 0.616199 -0.055453 -931.955811 -486.790509
12 stacked_ensemble_lstm3 10.992788 0.003274 -0.003725 0.040974 0.010238 0.057217 0.622765 -0.037395 -935.579831 -490.414529
15 ensemble_of_ensembles_lstm2 11.003166 0.003297 -0.006856 0.041194 0.001643 0.057417 0.620123 -0.044663 -934.113825 -488.948524
14 ensemble_of_ensembles_lstm1 11.036401 0.003289 -0.005891 0.041269 0.003800 0.057346 0.621066 -0.042069 -934.635868 -489.470567
13 stacked_ensemble_lstm4 11.199462 0.003300 -0.000119 0.041117 0.019013 0.057443 0.619783 -0.045596 -933.926211 -488.760910
9 ensemble_lstm_with_auto_encoder 11.441193 0.003490 -0.006453 0.042928 0.009863 0.059078 0.597824 -0.105983 -922.135293 -476.969991
4 cnn_lstm_encoder_decoder_outlier_removed 11.704837 0.003955 -0.009998 0.044963 -0.004986 0.062885 0.544323 -0.253112 -895.907205 -450.741903
7 ensemble_lstm 12.305297 0.004262 -0.004516 0.045603 0.016154 0.065283 0.505958 -0.336448 -891.103436 -444.677461
0 lstm_encoder_decoder 12.698133 0.004337 -0.001752 0.046863 0.024163 0.065859 0.497202 -0.360134 -887.379053 -440.953078
3 lstm_encoder_decoder_outlier_removed 12.764864 0.003897 -0.004757 0.047060 0.011061 0.062429 0.550912 -0.234991 -898.966093 -453.800791
2 cnn_lstm_encoder_decoder 13.628689 0.005047 -0.007281 0.050971 0.008144 0.071045 0.414906 -0.582753 -855.243482 -408.817507
6 lstm_auto_encoder_MLP 14.498825 0.005269 -0.004603 0.054422 0.023513 0.072590 0.392824 -0.669733 -835.630107 -390.464805
5 lstm_simple_model_outlier_removed 15.470578 0.006208 -0.005253 0.057966 0.025894 0.078788 0.284711 -0.967043 -801.217898 -356.052596
1 lstm_simple_model 15.812906 0.006353 0.000160 0.057859 0.037064 0.079707 0.263541 -0.992216 -806.466127 -360.040152
7.3 Model training with H2O
[53]:
h2o_model_obj = h2o_model(data_final,dataset_type='Normal')
loss_h2o_model = h2o_model_obj.train_and_evaluate_models()
loss_values_all_models.append(loss_h2o_model)
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:27 PM 11/10/2020



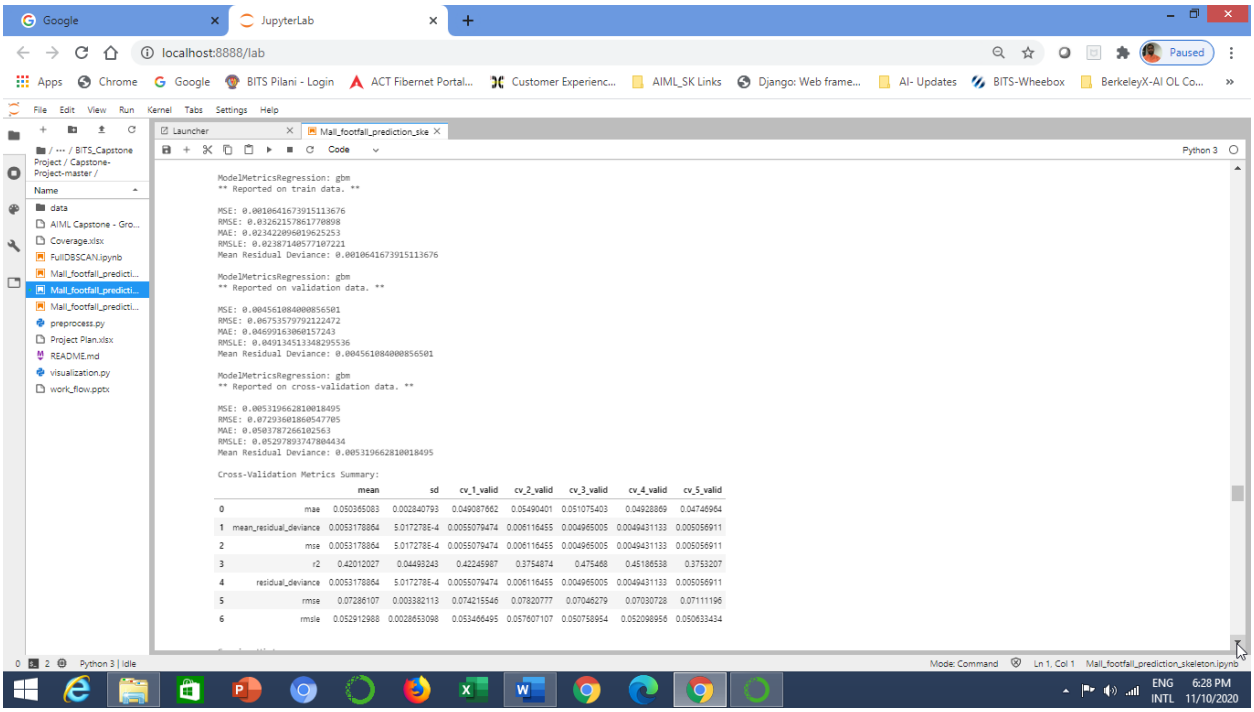
7.3 Model training with H2O

```
[53]: h2o_model_obj = h2o_model(data_final,dataset_type='Normal')
      loss_h2o_model = h2o_model_obj.train_and_evaluate_models()
      loss_values_all_models.append(loss_h2o_model)
```

```
Defined 2 models
Parse progress: |████████████████████████████████████████| 100%
Parse progress: |████████████████████████████████████████| 100%
gbm Model Build progress: |████████████████████████████████████████| 100%

ModelMetricsRegression: gbm
** Reported on test data. **

MSE: 0.0064561708328316385
RMSE: 0.080353007637908
MAE: 0.052907784857528844
RMSLE: 0.0575146145780769
Mean Residual Deviance: 0.0064561708328316385

gbm prediction progress: |████████████████████████████████████████| 100%
================================================================
H2OGradientBoosting
default h2o model fitting...
Model Details
=============
H2OGradientBoostingEstimator :  Gradient Boosting Machine
Model Key:  GBM_model_python_1604675476624_1


Model Summary:
```

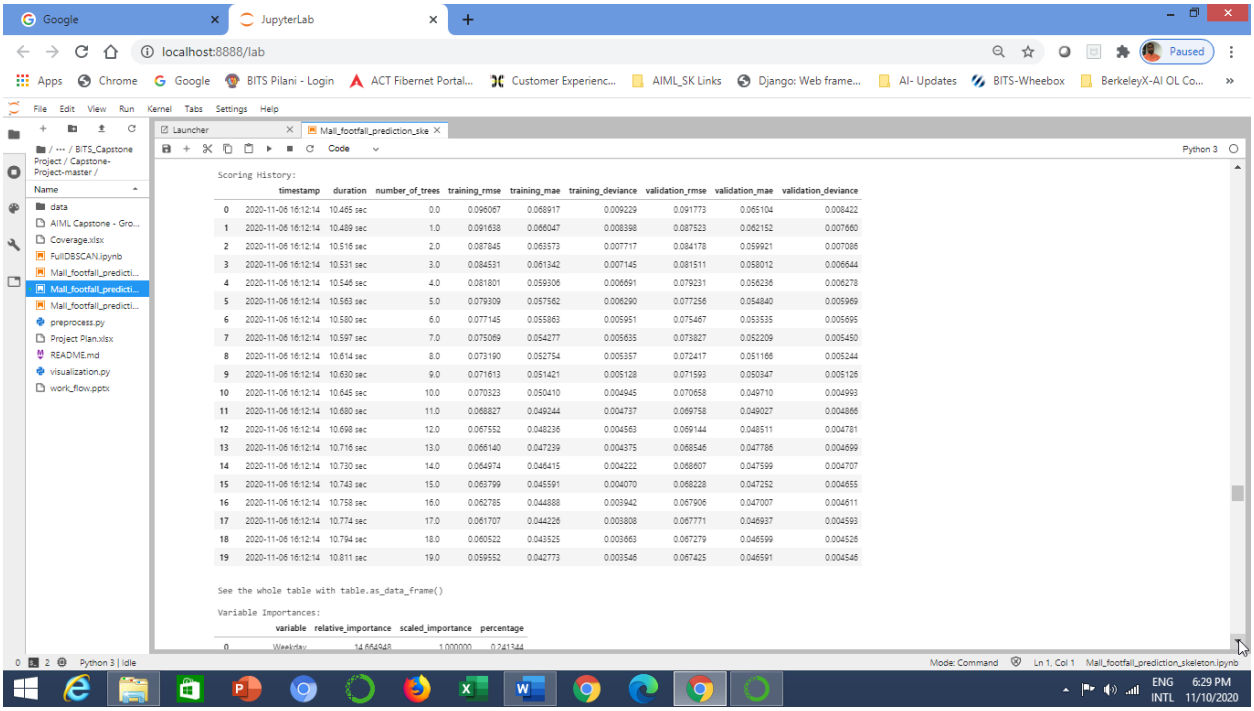
	number_of_trees	number_of_internal_trees	model_size_in_bytes	min_depth	max_depth	mean_depth	min_leaves	max_leaves	mean_leaves
0	100.0	100.0	35714.0	5.0	5.0	5.0	13.0	32.0	23.74

```
ModelMetricsRegression: gbm
** Reported on train data. **
```



```
ModelMetricsRegression: gbm
** Reported on train data. **

MSE: 0.0010641673915113676
RMSE: 0.032621578617708988
MAE: 0.023422096019625253
RMSLE: 0.02387140577107221
Mean Residual Deviance: 0.0010641673915113676

ModelMetricsRegression: gbm
** Reported on validation data. **

MSE: 0.004561084000856501
RMSE: 0.06753579792122472
MAE: 0.046991630610572 43
RMSLE: 0.04913451334829553 6
Mean Residual Deviance: 0.004561084000856501

ModelMetricsRegression: gbm
** Reported on cross-validation data. **

MSE: 0.005319662810018495
RMSE: 0.07293601860547705
MAE: 0.050378726610256 3
RMSLE: 0.05297893747804434
Mean Residual Deviance: 0.005319662810018495

Cross-Validation Metrics Summary:
```

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	cv_5_valid
0	mae	0.050365083	0.002840793	0.049087662	0.05490401	0.051075403	0.04928869	0.04746964
1	mean_residual_deviance	0.0053178864	5.017278E-4	0.0055079474	0.006116455	0.004965005	0.0049431133	0.005056911
2	mse	0.0053178864	5.017278E-4	0.0055079474	0.006116455	0.004965005	0.0049431133	0.005056911
3	r2	0.42012027	0.04493243	0.42245987	0.3754874	0.475468	0.45186538	0.3753207
4	residual_deviance	0.0053178864	5.017278E-4	0.0055079474	0.006116455	0.004965005	0.0049431133	0.005056911
5	rmse	0.07286107	0.003382113	0.074215546	0.07820777	0.07046279	0.07030728	0.07111196
6	rmsle	0.052912988	0.0028653098	0.053466495	0.057607107	0.050758954	0.052098956	0.050633434



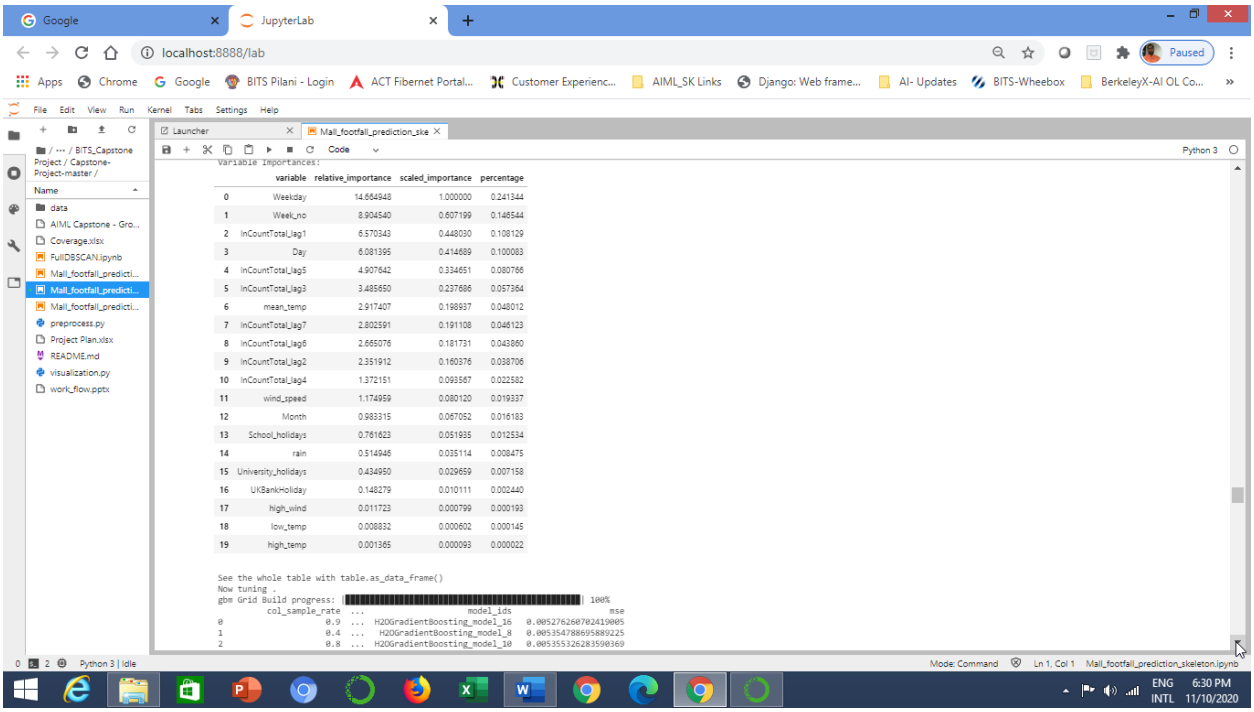
Scoring History:

	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
0	2020-11-06 16:12:14	10.465 sec	0.0	0.096067	0.068917	0.009229	0.091773	0.065104	0.008422
1	2020-11-06 16:12:14	10.489 sec	1.0	0.091638	0.066047	0.008398	0.087523	0.062152	0.007660
2	2020-11-06 16:12:14	10.516 sec	2.0	0.087845	0.063573	0.007717	0.084178	0.059921	0.007086
3	2020-11-06 16:12:14	10.531 sec	3.0	0.084531	0.061342	0.007145	0.081511	0.058012	0.006644
4	2020-11-06 16:12:14	10.546 sec	4.0	0.081801	0.059306	0.006691	0.079231	0.056236	0.006278
5	2020-11-06 16:12:14	10.563 sec	5.0	0.079309	0.057562	0.006290	0.077256	0.054840	0.005969
6	2020-11-06 16:12:14	10.580 sec	6.0	0.077145	0.055863	0.005951	0.075467	0.053535	0.005695
7	2020-11-06 16:12:14	10.597 sec	7.0	0.075069	0.054277	0.005635	0.073827	0.052209	0.005450
8	2020-11-06 16:12:14	10.614 sec	8.0	0.073190	0.052754	0.005357	0.072417	0.051166	0.005244
9	2020-11-06 16:12:14	10.630 sec	9.0	0.071613	0.051421	0.005128	0.071593	0.050347	0.005126
10	2020-11-06 16:12:14	10.645 sec	10.0	0.070323	0.050410	0.004945	0.070658	0.049710	0.004993
11	2020-11-06 16:12:14	10.680 sec	11.0	0.068827	0.049244	0.004737	0.069758	0.049027	0.004866
12	2020-11-06 16:12:14	10.698 sec	12.0	0.067552	0.048236	0.004563	0.069144	0.048511	0.004781
13	2020-11-06 16:12:14	10.716 sec	13.0	0.066140	0.047239	0.004375	0.068546	0.047786	0.004699
14	2020-11-06 16:12:14	10.730 sec	14.0	0.064974	0.046415	0.004222	0.068607	0.047599	0.004707
15	2020-11-06 16:12:14	10.743 sec	15.0	0.063799	0.045591	0.004070	0.068228	0.047252	0.004655
16	2020-11-06 16:12:14	10.758 sec	16.0	0.062785	0.044888	0.003942	0.067906	0.047007	0.004611
17	2020-11-06 16:12:14	10.774 sec	17.0	0.061707	0.044226	0.003808	0.067771	0.046937	0.004593
18	2020-11-06 16:12:14	10.794 sec	18.0	0.060522	0.043525	0.003663	0.067279	0.046599	0.004526
19	2020-11-06 16:12:14	10.811 sec	19.0	0.059552	0.042773	0.003546	0.067425	0.046591	0.004546

See the whole table with table.as_data_frame()

Variable Importances:

	variable	relative_importance	scaled_importance	percentage
0	Weekday	14.664948	1.000000	0.241344



Variable Importances:

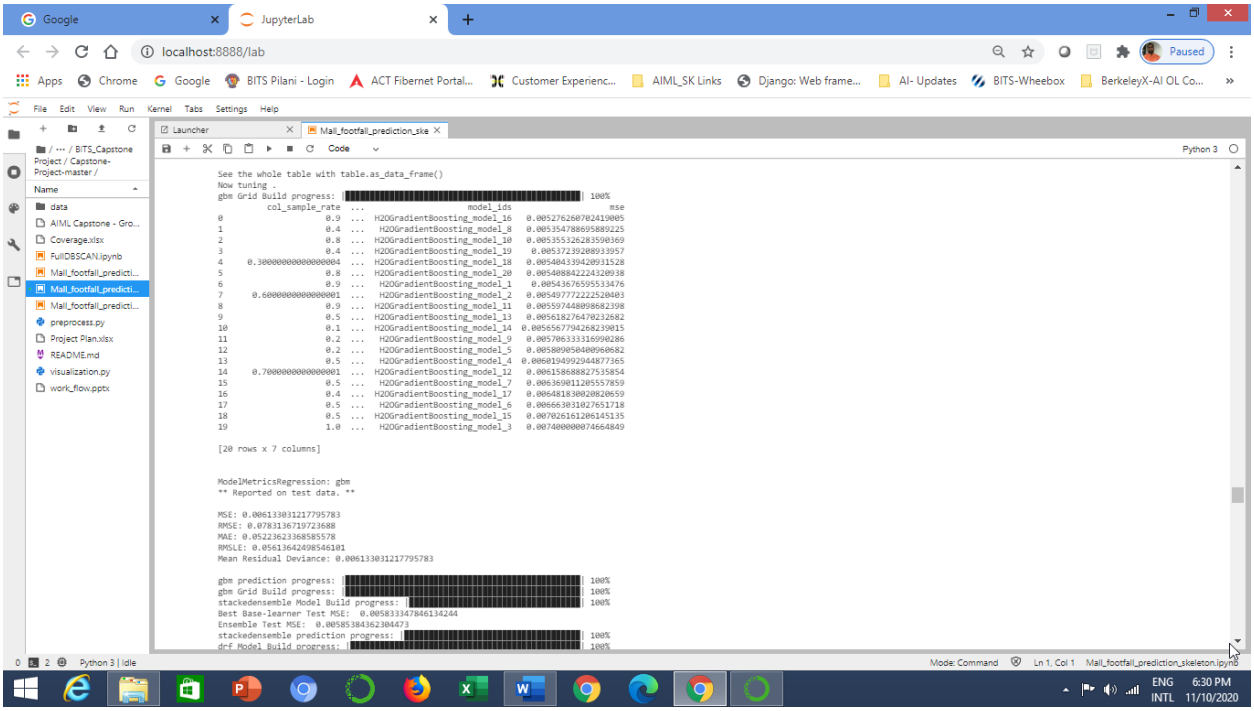
	variable	relative_importance	scaled_importance	percentage
0	Weekday	14.664948	1.000000	0.241344
1	Week_no	8.904540	0.607199	0.146544
2	lnCountTotal_lag1	6.570343	0.448030	0.108129
3	Day	6.081395	0.414689	0.100083
4	lnCountTotal_lag5	4.907642	0.334651	0.080766
5	lnCountTotal_lag3	3.485650	0.237686	0.057364
6	mean_temp	2.917407	0.198937	0.048012
7	lnCountTotal_lag7	2.802591	0.191108	0.046123
8	lnCountTotal_lag6	2.665076	0.181731	0.043860
9	lnCountTotal_lag2	2.351912	0.160376	0.038706
10	lnCountTotal_lag4	1.372151	0.093567	0.022582
11	wind_speed	1.174959	0.080120	0.019337
12	Month	0.983315	0.067052	0.016183
13	School_holidays	0.761623	0.051935	0.012534
14	rain	0.514946	0.035114	0.008475
15	University_holidays	0.434950	0.029659	0.007158
16	UKBankHoliday	0.148279	0.010111	0.002440
17	high_wind	0.011723	0.000799	0.000193
18	low_temp	0.008832	0.000602	0.000145
19	high_temp	0.001365	0.000093	0.000022

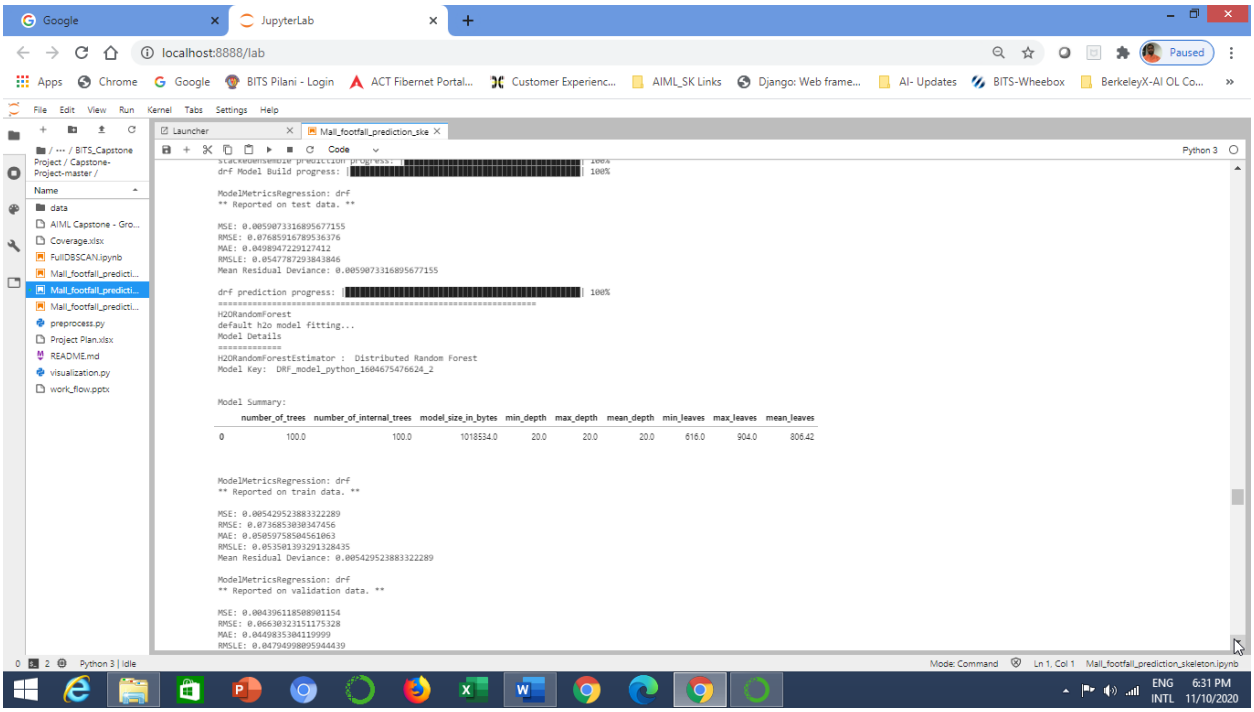
See the whole table with table.as_data_frame()

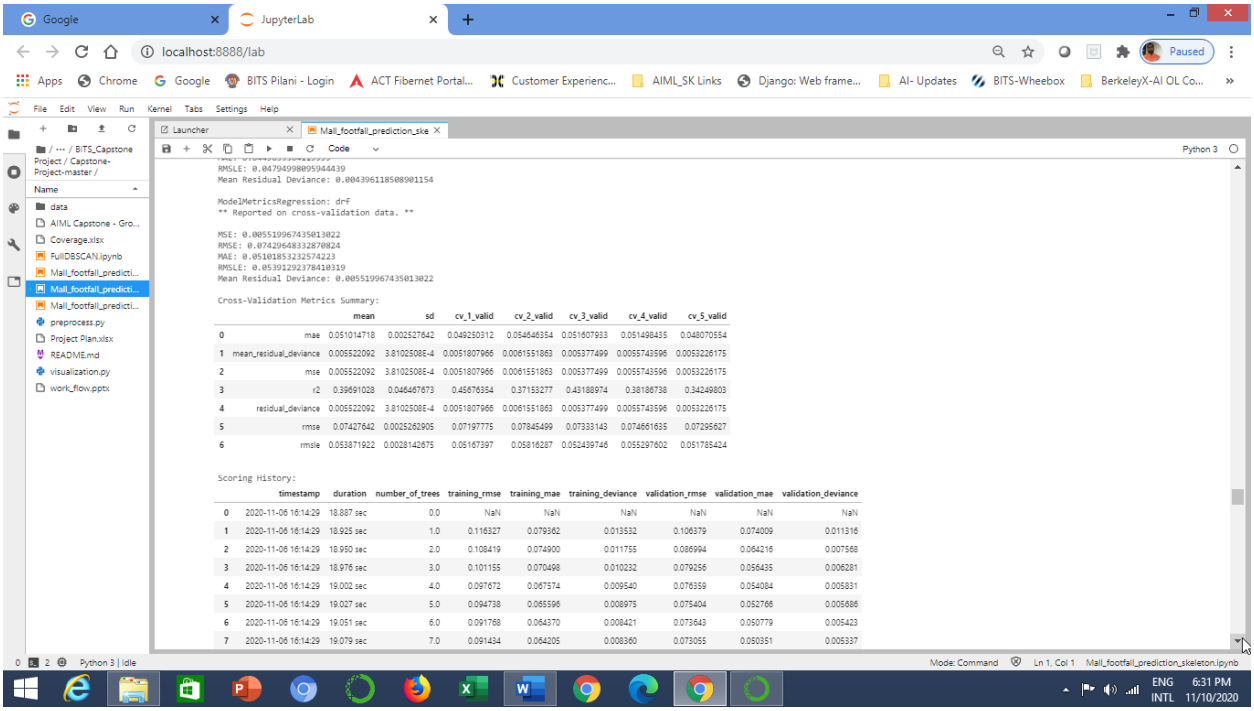
Now tuning ..

gbm Grid Build progress: |████████████████████████████████| 100%

	col_sample_rate	...	model_ids	mse
0	0.9	...	H2OGradientBoosting_model_16	0.005276260702419005
1	0.4	...	H2OGradientBoosting_model_8	0.005354788695889225
2	0.8	...	H2OGradientBoosting_model_10	0.005355326283590369


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
See the whole table with table.as_data_frame()
Now tuning .
gbm Grid Build progress: 100%
col_sample_rate ... model_ids mse
0 0.9 ... H2OGradientBoosting_model_16 0.005276260702419005
1 0.4 ... H2OGradientBoosting_model_8 0.005354788695889225
2 0.8 ... H2OGradientBoosting_model_10 0.005355326283590369
3 0.4 ... H2OGradientBoosting_model_19 0.00537239208933957
4 0.3000000000000004 ... H2OGradientBoosting_model_18 0.005404339420931528
5 0.8 ... H2OGradientBoosting_model_20 0.005408842224320938
6 0.9 ... H2OGradientBoosting_model_1 0.00543676595533476
7 0.6000000000000001 ... H2OGradientBoosting_model_2 0.005497772222520403
8 0.9 ... H2OGradientBoosting_model_11 0.005597448098682398
9 0.5 ... H2OGradientBoosting_model_13 0.005618276470232682
10 0.1 ... H2OGradientBoosting_model_14 0.005565779426829015
11 0.2 ... H2OGradientBoosting_model_9 0.005706333316990286
12 0.2 ... H2OGradientBoosting_model_5 0.005880950400960682
13 0.5 ... H2OGradientBoosting_model_4 0.006019499229448777365
14 0.7000000000000001 ... H2OGradientBoosting_model_12 0.006158688827535854
15 0.5 ... H2OGradientBoosting_model_7 0.006365911205557859
16 0.4 ... H2OGradientBoosting_model_17 0.006481830020820659
17 0.5 ... H2OGradientBoosting_model_6 0.006663031027651718
18 0.5 ... H2OGradientBoosting_model_15 0.007026161206145135
19 1.0 ... H2OGradientBoosting_model_3 0.007400000074664849
[20 rows x 7 columns]
ModelMetricsRegression: gbm
** Reported on test data. **
MSE: 0.006133031217795783
RMSE: 0.07831367197223688
MAE: 0.05223623368585578
RMSLE: 0.05613642498546101
Mean Residual Deviance: 0.006133031217795783
gbm prediction progress: 100%
gbm Grid Build progress: 100%
stackedensemble Model Build progress: 100%
Best Base-learner Test MSE: 0.005833347846134244
Ensemble Test MSE: 0.005853843623004473
stackedensemble prediction progress: 100%
drf Model Build progress: 100%
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:30 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
drf Model Build progress: 100%
ModelMetricsRegression: drf
** Reported on test data. **
MSE: 0.005907331689567715
RMSE: 0.07685916789536376
MAE: 0.0498947229127412
RMSLE: 0.05477872938438846
Mean Residual Deviance: 0.005907331689567715
drf prediction progress: 100%
H2ORandomForest
default h2o model fitting...
Model Details
H2ORandomForestEstimator : Distributed Random Forest
Model Key: DRF_model_python_1604675476624_2
Model Summary:
number_of_trees number_of_internal_trees model_size_in_bytes min_depth max_depth mean_depth min_leaves max_leaves mean_leaves
0 100.0 100.0 1018534.0 20.0 20.0 20.0 616.0 904.0 806.42
ModelMetricsRegression: drf
** Reported on train data. **
MSE: 0.005429523883322289
RMSE: 0.07368530830347456
MAE: 0.05059758504561063
RMSLE: 0.05350139329132843
Mean Residual Deviance: 0.005429523883322289
ModelMetricsRegression: drf
** Reported on validation data. **
MSE: 0.004396118508901154
RMSE: 0.06630323151175328
MAE: 0.04498353041199999
RMSLE: 0.04794998095944439
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:31 PM
11/10/2020



RMSLE: 0.04794998095944439
Mean Residual Deviance: 0.004396118508901154

ModelMetricsRegression: drf
** Reported on cross-validation data. **

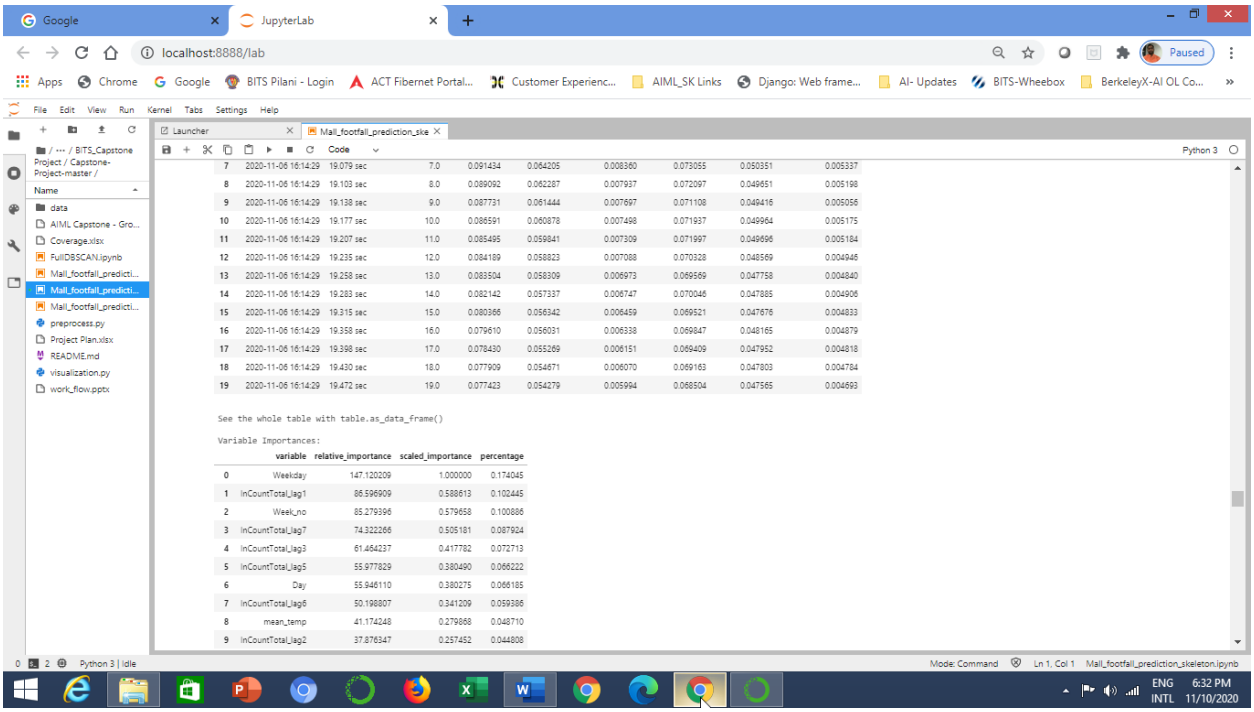
MSE: 0.005519967435013022
RMSE: 0.07429648332878824
MAE: 0.05101853232574223
RMSLE: 0.0539129237841031
Mean Residual Deviance: 0.005519967435013022

Cross-Validation Metrics Summary:

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	cv_5_valid
0	mae	0.051014718	0.002527642	0.049250312	0.054646354	0.051607933	0.051498435	0.048070554
1	mean_residual_deviance	0.005522092	3.8102508E-4	0.0051807966	0.0061551863	0.005377499	0.0055743596	0.0053226175
2	mse	0.005522092	3.8102508E-4	0.0051807966	0.0061551863	0.005377499	0.0055743596	0.0053226175
3	r2	0.39691028	0.046467673	0.45676354	0.37153277	0.43188974	0.38186738	0.34249803
4	residual_deviance	0.005522092	3.8102508E-4	0.0051807966	0.0061551863	0.005377499	0.0055743596	0.0053226175
5	rmse	0.07427642	0.0025262905	0.07197775	0.07845499	0.07333143	0.074661635	0.07295627
6	rmsle	0.053871922	0.0028142675	0.05167397	0.05816287	0.052439746	0.055297602	0.051785424

Scoring History:

	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
0	2020-11-06 16:14:29	18.887 sec	0.0	NaN	NaN	NaN	NaN	NaN	NaN
1	2020-11-06 16:14:29	18.925 sec	1.0	0.116327	0.079362	0.013532	0.106379	0.074009	0.011316
2	2020-11-06 16:14:29	18.950 sec	2.0	0.108419	0.074900	0.011755	0.086994	0.064216	0.007568
3	2020-11-06 16:14:29	18.976 sec	3.0	0.101155	0.070498	0.010232	0.079256	0.056435	0.006281
4	2020-11-06 16:14:29	19.002 sec	4.0	0.097672	0.067574	0.009540	0.076359	0.054084	0.005831
5	2020-11-06 16:14:29	19.027 sec	5.0	0.094738	0.065596	0.008975	0.075404	0.052766	0.005686
6	2020-11-06 16:14:29	19.051 sec	6.0	0.091768	0.064370	0.008421	0.073643	0.050779	0.005423
7	2020-11-06 16:14:29	19.079 sec	7.0	0.091434	0.064205	0.008360	0.073055	0.050351	0.005337

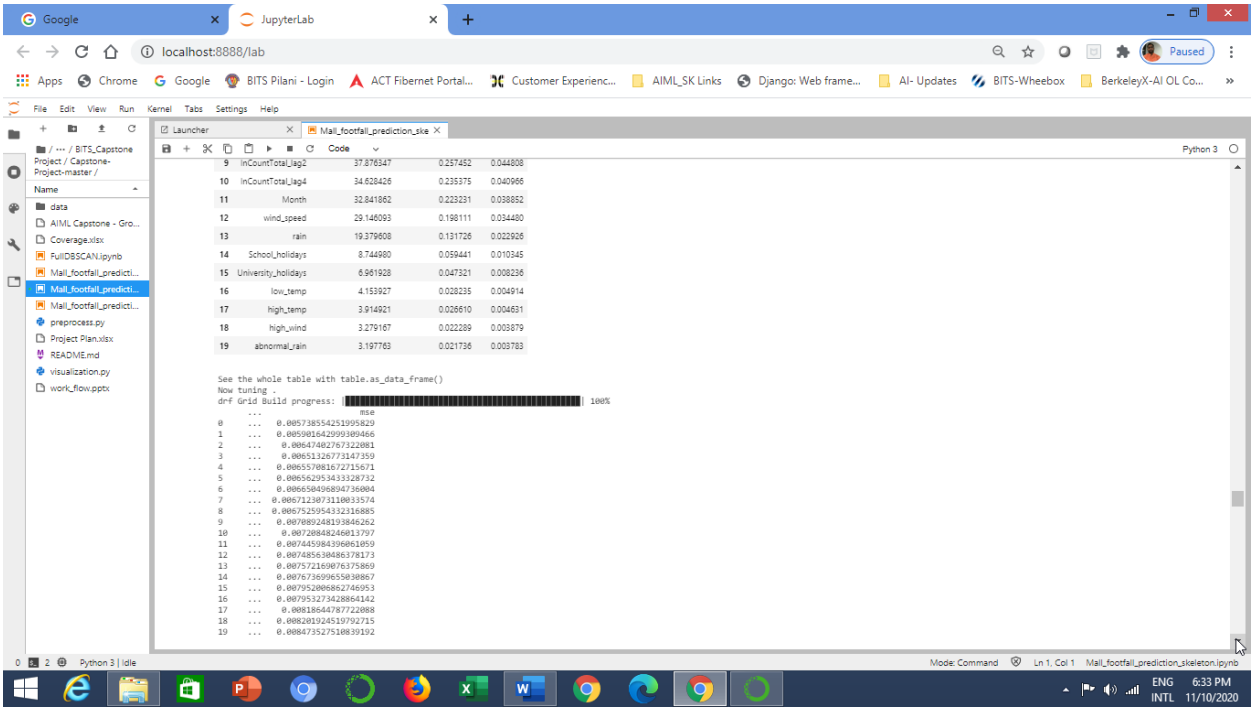


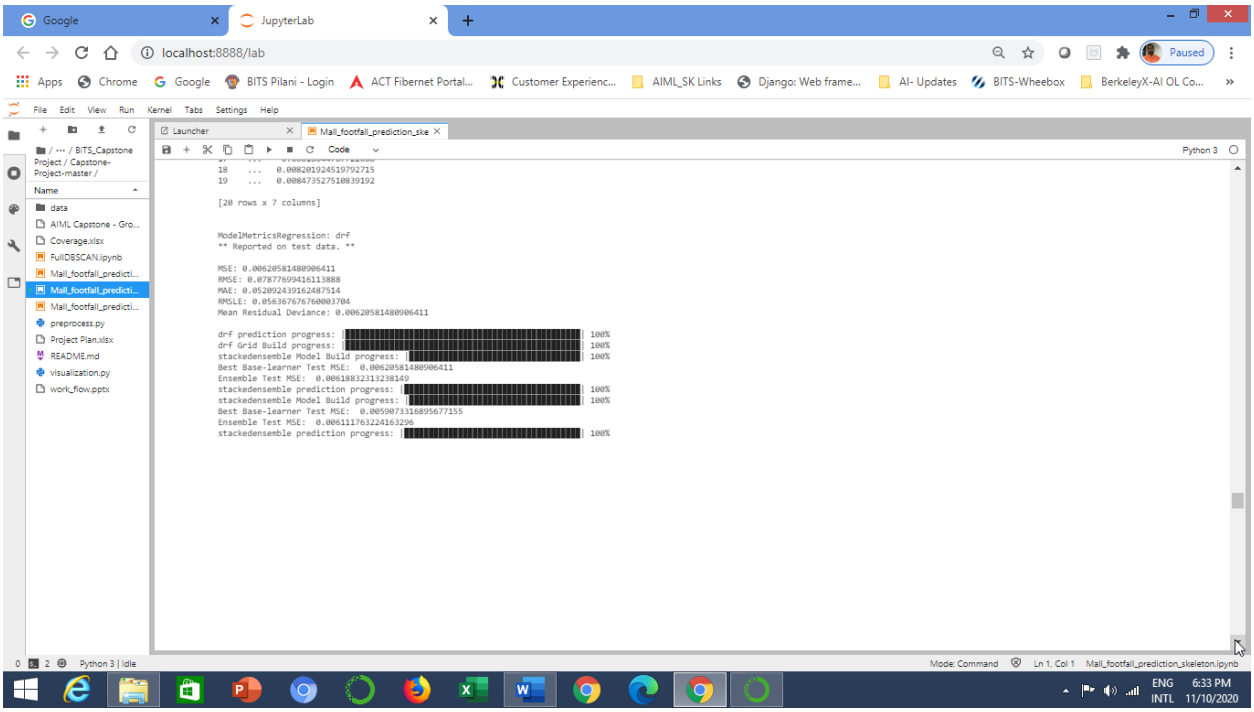
7	2020-11-06 16:14:29	19.079 sec	7.0	0.091434	0.064205	0.008360	0.073055	0.050351	0.005337
8	2020-11-06 16:14:29	19.103 sec	8.0	0.089092	0.062287	0.007937	0.072097	0.049651	0.005198
9	2020-11-06 16:14:29	19.138 sec	9.0	0.087731	0.061444	0.007697	0.071108	0.049416	0.005056
10	2020-11-06 16:14:29	19.177 sec	10.0	0.086591	0.060878	0.007498	0.071937	0.049964	0.005175
11	2020-11-06 16:14:29	19.207 sec	11.0	0.085495	0.059841	0.007309	0.071997	0.049696	0.005184
12	2020-11-06 16:14:29	19.235 sec	12.0	0.084189	0.058823	0.007088	0.070328	0.048569	0.004946
13	2020-11-06 16:14:29	19.258 sec	13.0	0.083504	0.058309	0.006973	0.069569	0.047758	0.004840
14	2020-11-06 16:14:29	19.283 sec	14.0	0.082142	0.057337	0.006747	0.070046	0.047885	0.004906
15	2020-11-06 16:14:29	19.315 sec	15.0	0.080366	0.056342	0.006459	0.069521	0.047676	0.004833
16	2020-11-06 16:14:29	19.358 sec	16.0	0.079610	0.056031	0.006338	0.069847	0.048165	0.004879
17	2020-11-06 16:14:29	19.398 sec	17.0	0.078430	0.055269	0.006151	0.069409	0.047952	0.004818
18	2020-11-06 16:14:29	19.430 sec	18.0	0.077909	0.054671	0.006070	0.069163	0.047803	0.004784
19	2020-11-06 16:14:29	19.472 sec	19.0	0.077423	0.054279	0.005994	0.068504	0.047565	0.004693

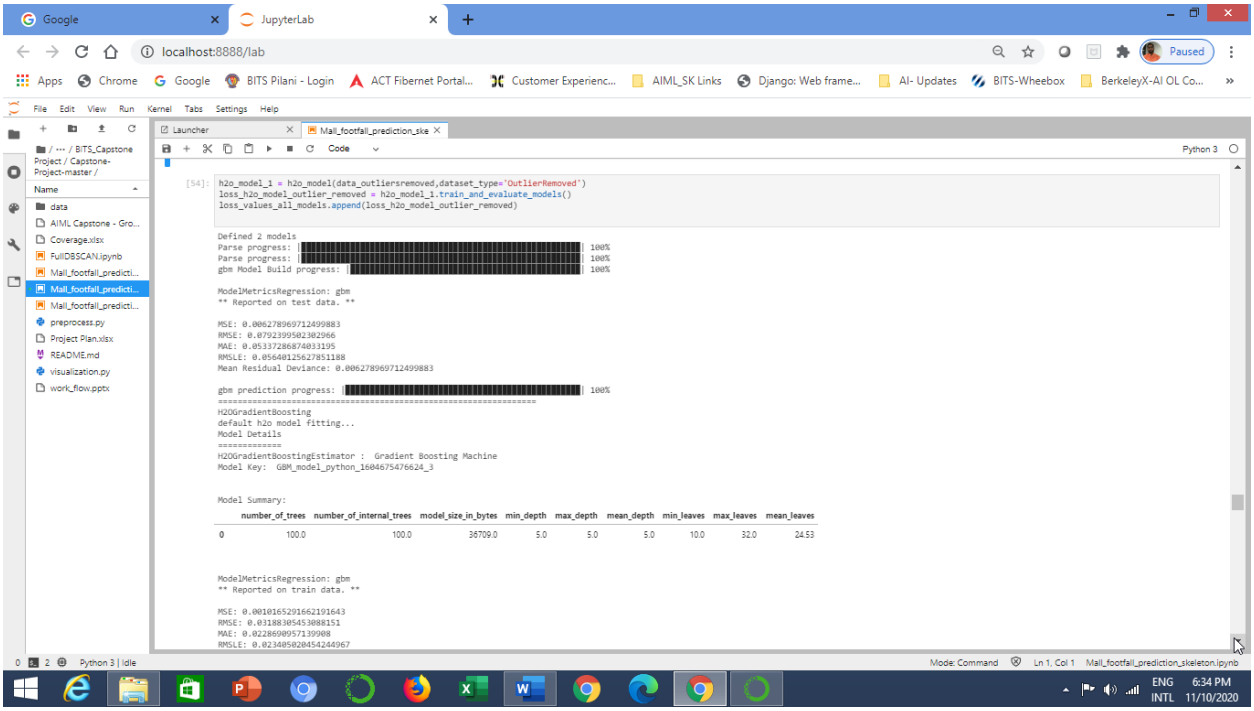
See the whole table with table.as_data_frame()

Variable Importances:

	variable	relative_importance	scaled_importance	percentage
0	Weekday	147.120209	1.000000	0.174045
1	InCountTotal_lag1	86.596909	0.588613	0.102445
2	Week_no	85.279396	0.579658	0.100886
3	InCountTotal_lag7	74.322266	0.505181	0.087924
4	InCountTotal_lag3	61.464237	0.417782	0.072713
5	InCountTotal_lag5	55.977829	0.380490	0.066222
6	Day	55.946110	0.380275	0.066185
7	InCountTotal_lag6	50.198807	0.341209	0.059386
8	mean_temp	41.174248	0.279868	0.048710
9	InCountTotal_lag2	37.876347	0.257452	0.044808


JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
9 InCountTotal_lag2 37.876347 0.257452 0.044808
10 InCountTotal_lag4 34.628426 0.235375 0.040966
11 Month 32.841862 0.223231 0.038852
12 wind_speed 29.146093 0.198111 0.034480
13 rain 19.379608 0.131726 0.022926
14 School_holidays 8.744980 0.059441 0.010345
15 University_holidays 6.961928 0.047321 0.008236
16 low_temp 4.153927 0.028235 0.004914
17 high_temp 3.914921 0.026610 0.004631
18 high_wind 3.279167 0.022289 0.003879
19 abnormal_rain 3.197763 0.021736 0.003783
See the whole table with table.as_data_frame()
Now tuning .
drf Grid Build progress: | 100%
... mse
0 ... 0.005738554251995829
1 ... 0.005901642999309466
2 ... 0.006474027673220 81
3 ... 0.006513267731473 59
4 ... 0.006557081672715671
5 ... 0.006562953433328732
6 ... 0.006650496894736004
7 ... 0.006712307311003 3574
8 ... 0.006752595432316885
9 ... 0.007089248193846262
10 ... 0.007208482460137 97
11 ... 0.007445984396061059
12 ... 0.007485630486378173
13 ... 0.007572169076375869
14 ... 0.007673699655030867
15 ... 0.007952006862746953
16 ... 0.007953273428864142
17 ... 0.008186447877220 88
18 ... 0.008201924519792715
19 ... 0.008473527510839192
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:33 PM 11/10/2020


JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
18 ... 0.008201924519792715
19 ... 0.008473527510839192
[20 rows x 7 columns]
ModelMetricsRegression: drf
** Reported on test data. **
MSE: 0.006205814809 06411
RMSE: 0.078776994161 13888
MAE: 0.052092439162487514
RMSLE: 0.056367676760003704
Mean Residual Deviance: 0.006205814809 06411
drf prediction progress: | 100%
drf Grid Build progress: | 100%
stackedensemble Model Build progress: | 100%
Best Base-learner Test MSE: 0.006205814809 06411
Ensemble Test MSE: 0.006188231123 8149
stackedensemble prediction progress: | 100%
stackedensemble Model Build progress: | 100%
Best Base-learner Test MSE: 0.005907331689 5677155
Ensemble Test MSE: 0.006111763224 163296
stackedensemble prediction progress: | 100%
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:33 PM 11/10/2020



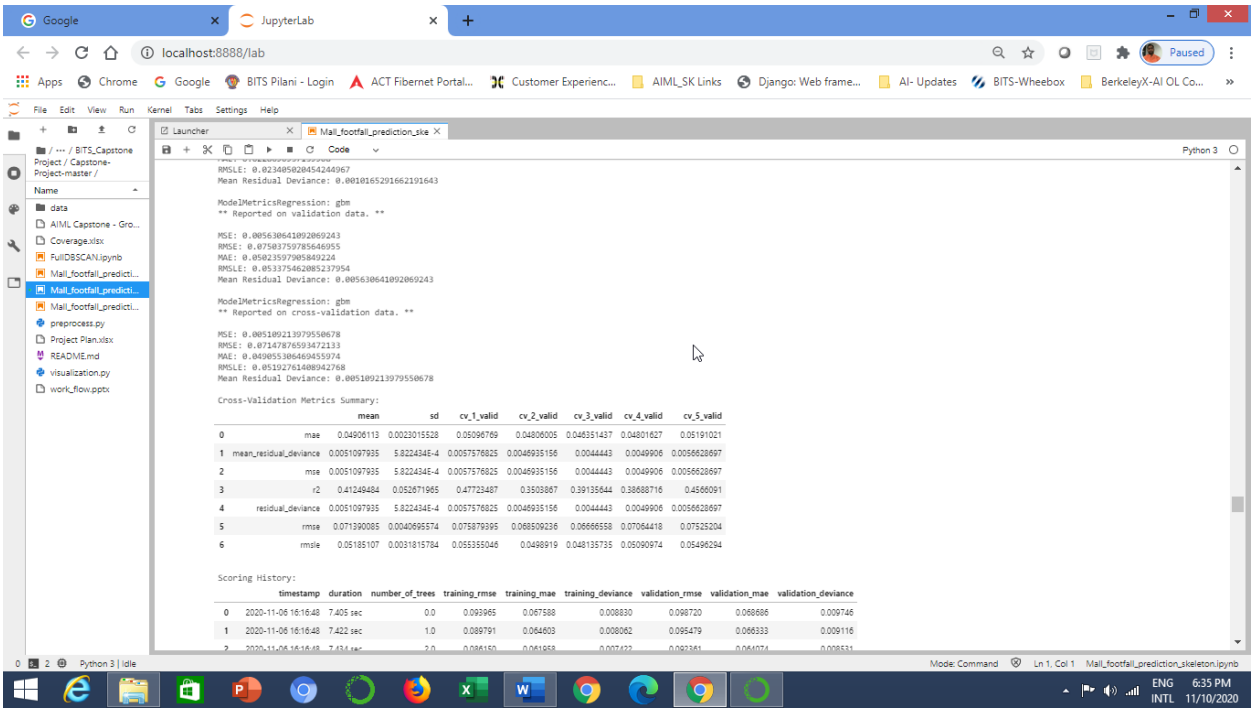
```
[54]: h2o_model_1 = h2o_model(data_outliersremoved,dataset_type='OutlierRemoved')
loss_h2o_model_outlier_removed = h2o_model_1.train_and_evaluate_models()
loss_values_all_models.append(loss_h2o_model_outlier_removed)
```

```
Defined 2 models
Parse progress: |████████████████████████████████████████| 100%
Parse progress: |████████████████████████████████████████| 100%
gbm Model Build progress: |████████████████████████████████████████| 100%

ModelMetricsRegression: gbm
** Reported on test data. **

MSE: 0.006278969712499883
RMSE: 0.07923995023029666
MAE: 0.05337286874033195
RMSLE: 0.05640125627851188
Mean Residual Deviance: 0.006278969712499883

gbm prediction progress: |████████████████████████████████████████| 100%
================================================================
H2OGradientBoosting
default h2o model fitting...
Model Details
=============
H2OGradientBoostingEstimator :  Gradient Boosting Machine
Model Key:  GBM_model_python_1604675476624_3
```

Model Summary:

	number_of_trees	number_of_internal_trees	model_size_in_bytes	min_depth	max_depth	mean_depth	min_leaves	max_leaves	mean_leaves
0	100.0	100.0	36709.0	5.0	5.0	5.0	10.0	32.0	24.53

```
ModelMetricsRegression: gbm
** Reported on train data. **

MSE: 0.0010165291662191643
RMSE: 0.031883054533088151
MAE: 0.022869095713990 8
RMSLE: 0.023405020454244967
```

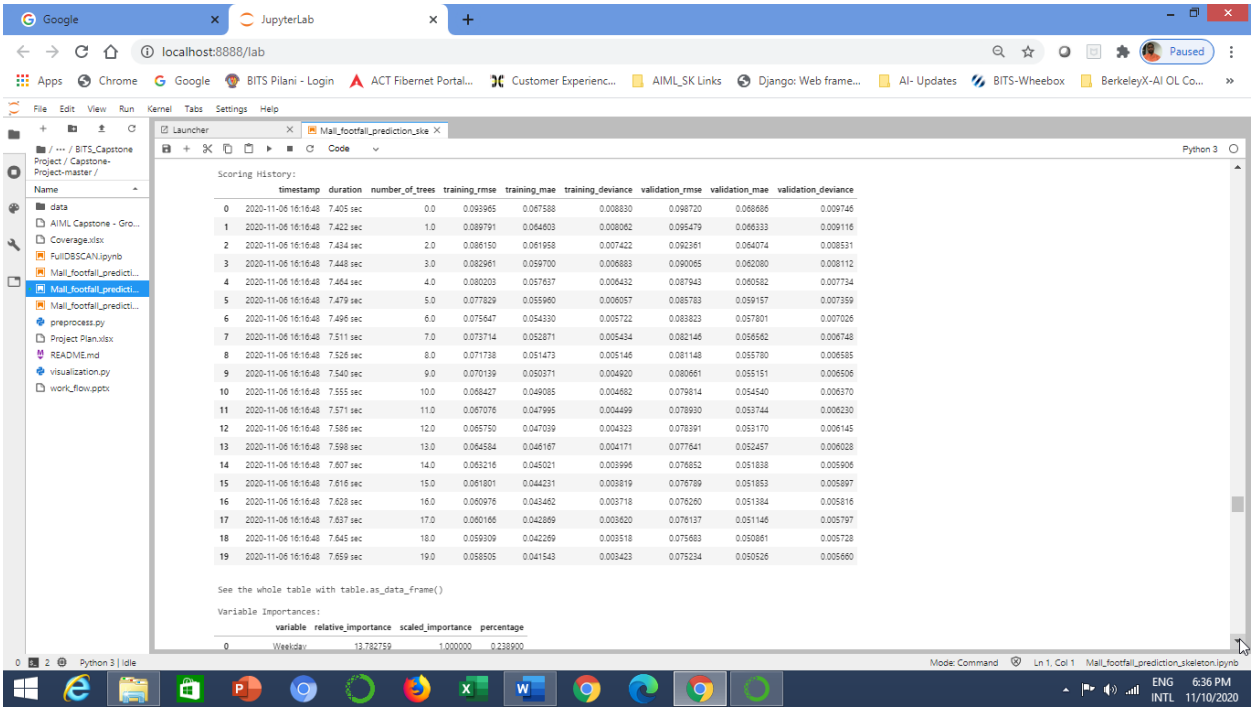


```
RMSLE: 0.023405020454244967
Mean Residual Deviance: 0.0010165291662191643

ModelMetricsRegression: gbm
** Reported on validation data. **

MSE: 0.005630641092069243
RMSE: 0.07503759785646955
MAE: 0.05023597905849224
RMSLE: 0.05337546208527954
Mean Residual Deviance: 0.005630641092069243

ModelMetricsRegression: gbm
** Reported on cross-validation data. **

MSE: 0.005109213979550678
RMSE: 0.07147876593472133
MAE: 0.04905530646945974
RMSLE: 0.05192761408942768
Mean Residual Deviance: 0.005109213979550678

Cross-Validation Metrics Summary:
```

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	cv_5_valid
0	mae	0.04906113	0.0023015528	0.05096769	0.04806005	0.046351437	0.04801627	0.05191021
1	mean_residual_deviance	0.0051097935	5.822434E-4	0.0057576825	0.0046935156	0.0044443	0.0049906	0.0056628697
2	mse	0.0051097935	5.822434E-4	0.0057576825	0.0046935156	0.0044443	0.0049906	0.0056628697
3	r2	0.41249484	0.052671965	0.47723487	0.3503867	0.39135644	0.38688716	0.4566091
4	residual_deviance	0.0051097935	5.822434E-4	0.0057576825	0.0046935156	0.0044443	0.0049906	0.0056628697
5	rmse	0.071390085	0.0040695574	0.075879395	0.068509236	0.06666558	0.07064418	0.07525204
6	rmsle	0.05185107	0.0031815784	0.055355046	0.0498919	0.048135735	0.05090974	0.05496294

Scoring History:

	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
0	2020-11-06 16:16:48	7.405 sec	0.0	0.093965	0.067588	0.008830	0.098720	0.068686	0.009746
1	2020-11-06 16:16:48	7.422 sec	1.0	0.089791	0.064603	0.008062	0.095479	0.066333	0.009116
2	2020-11-06 16:16:48	7.434 sec	2.0	0.086150	0.061958	0.007422	0.092361	0.064074	0.008531



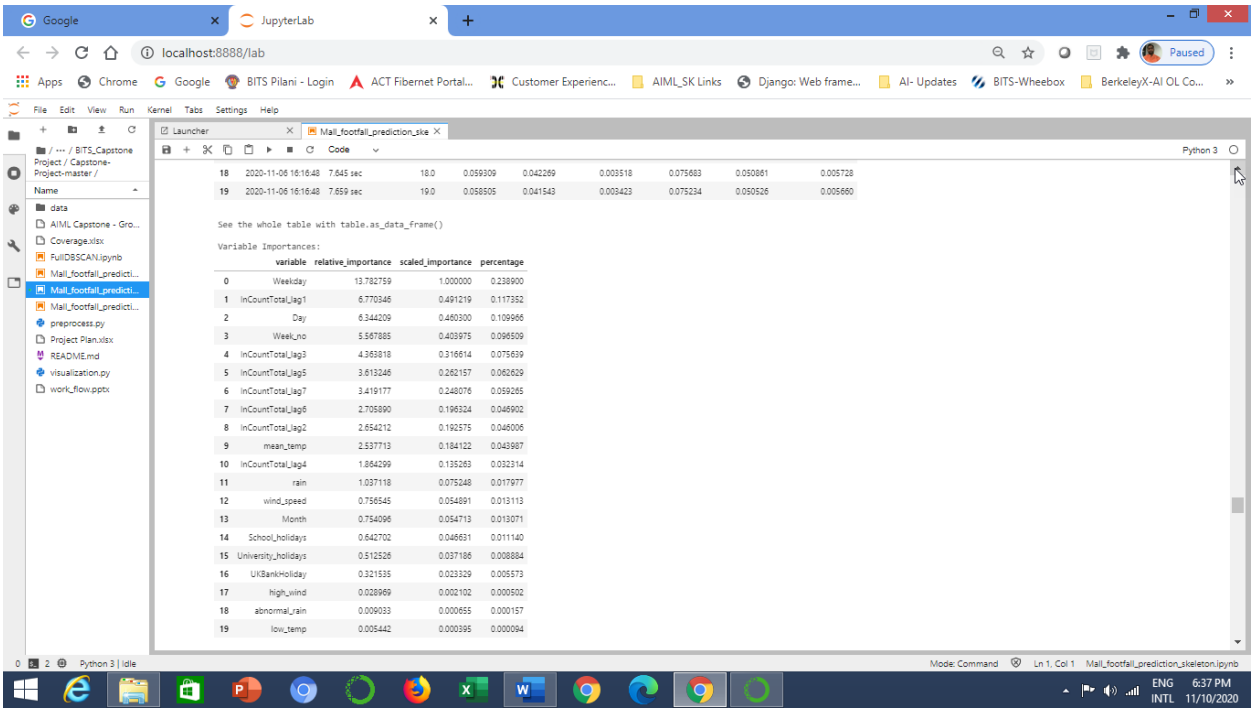
Scoring History:

	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
0	2020-11-06 16:16:48	7.405 sec	0.0	0.093965	0.067588	0.008830	0.098720	0.068686	0.009746
1	2020-11-06 16:16:48	7.422 sec	1.0	0.089791	0.064603	0.008062	0.095479	0.066333	0.009116
2	2020-11-06 16:16:48	7.434 sec	2.0	0.086150	0.061958	0.007422	0.092361	0.064074	0.008531
3	2020-11-06 16:16:48	7.448 sec	3.0	0.082961	0.059700	0.006883	0.090065	0.062080	0.008112
4	2020-11-06 16:16:48	7.464 sec	4.0	0.080203	0.057637	0.006432	0.087943	0.060582	0.007734
5	2020-11-06 16:16:48	7.479 sec	5.0	0.077829	0.055960	0.006057	0.085783	0.059157	0.007359
6	2020-11-06 16:16:48	7.496 sec	6.0	0.075647	0.054330	0.005722	0.083823	0.057801	0.007026
7	2020-11-06 16:16:48	7.511 sec	7.0	0.073714	0.052871	0.005434	0.082146	0.056562	0.006748
8	2020-11-06 16:16:48	7.526 sec	8.0	0.071738	0.051473	0.005146	0.081148	0.055780	0.006585
9	2020-11-06 16:16:48	7.540 sec	9.0	0.070139	0.050371	0.004920	0.080661	0.055151	0.006506
10	2020-11-06 16:16:48	7.555 sec	10.0	0.068427	0.049085	0.004682	0.079814	0.054540	0.006370
11	2020-11-06 16:16:48	7.571 sec	11.0	0.067076	0.047995	0.004499	0.078930	0.053744	0.006230
12	2020-11-06 16:16:48	7.586 sec	12.0	0.065750	0.047039	0.004323	0.078391	0.053170	0.006145
13	2020-11-06 16:16:48	7.598 sec	13.0	0.064584	0.046167	0.004171	0.077641	0.052457	0.006028
14	2020-11-06 16:16:48	7.607 sec	14.0	0.063216	0.045021	0.003996	0.076852	0.051838	0.005906
15	2020-11-06 16:16:48	7.616 sec	15.0	0.061801	0.044231	0.003819	0.076789	0.051853	0.005897
16	2020-11-06 16:16:48	7.628 sec	16.0	0.060976	0.043462	0.003718	0.076260	0.051384	0.005816
17	2020-11-06 16:16:48	7.637 sec	17.0	0.060166	0.042869	0.003620	0.076137	0.051146	0.005797
18	2020-11-06 16:16:48	7.645 sec	18.0	0.059309	0.042269	0.003518	0.075683	0.050861	0.005728
19	2020-11-06 16:16:48	7.659 sec	19.0	0.058505	0.041543	0.003423	0.075234	0.050526	0.005660

See the whole table with table.as_data_frame()

Variable Importances:

	variable	relative_importance	scaled_importance	percentage
0	Weekday	13.782759	1.000000	0.238900



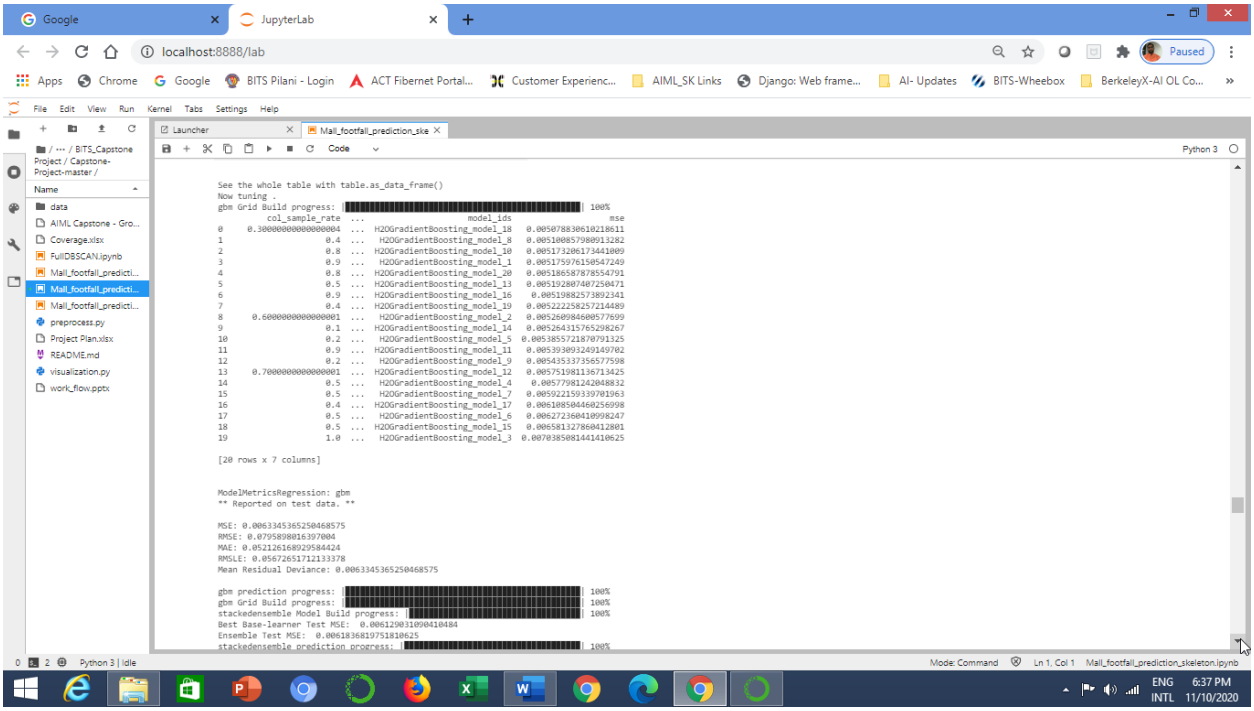
Scoring History:

	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
18	2020-11-06 16:16:48	7.645 sec	18.0	0.059309	0.042269	0.003518	0.075683	0.050861	0.005728
19	2020-11-06 16:16:48	7.659 sec	19.0	0.058505	0.041543	0.003423	0.075234	0.050526	0.005660

See the whole table with table.as_data_frame()

Variable Importances:

	variable	relative_importance	scaled_importance	percentage
0	Weekday	13.782759	1.000000	0.238900
1	lnCountTotal_lag1	6.770346	0.491219	0.117352
2	Day	6.344209	0.460300	0.109966
3	Week_no	5.567885	0.403975	0.096509
4	lnCountTotal_lag3	4.363818	0.316614	0.075639
5	lnCountTotal_lag5	3.613246	0.262157	0.062629
6	lnCountTotal_lag7	3.419177	0.248076	0.059265
7	lnCountTotal_lag6	2.705890	0.196324	0.046902
8	lnCountTotal_lag2	2.654212	0.192575	0.046006
9	mean_temp	2.537713	0.184122	0.043987
10	lnCountTotal_lag4	1.864299	0.135263	0.032314
11	rain	1.037118	0.075248	0.017977
12	wind_speed	0.756545	0.054891	0.013113
13	Month	0.754096	0.054713	0.013071
14	School_holidays	0.642702	0.046631	0.011140
15	University_holidays	0.512526	0.037186	0.008884
16	UKBankHoliday	0.321535	0.023329	0.005573
17	high_wind	0.028969	0.002102	0.000502
18	abnormal_rain	0.009033	0.000655	0.000157
19	low_temp	0.005442	0.000395	0.000094



See the whole table with table.as_data_frame()
Now tuning .
gbm Grid Build progress: | 100%

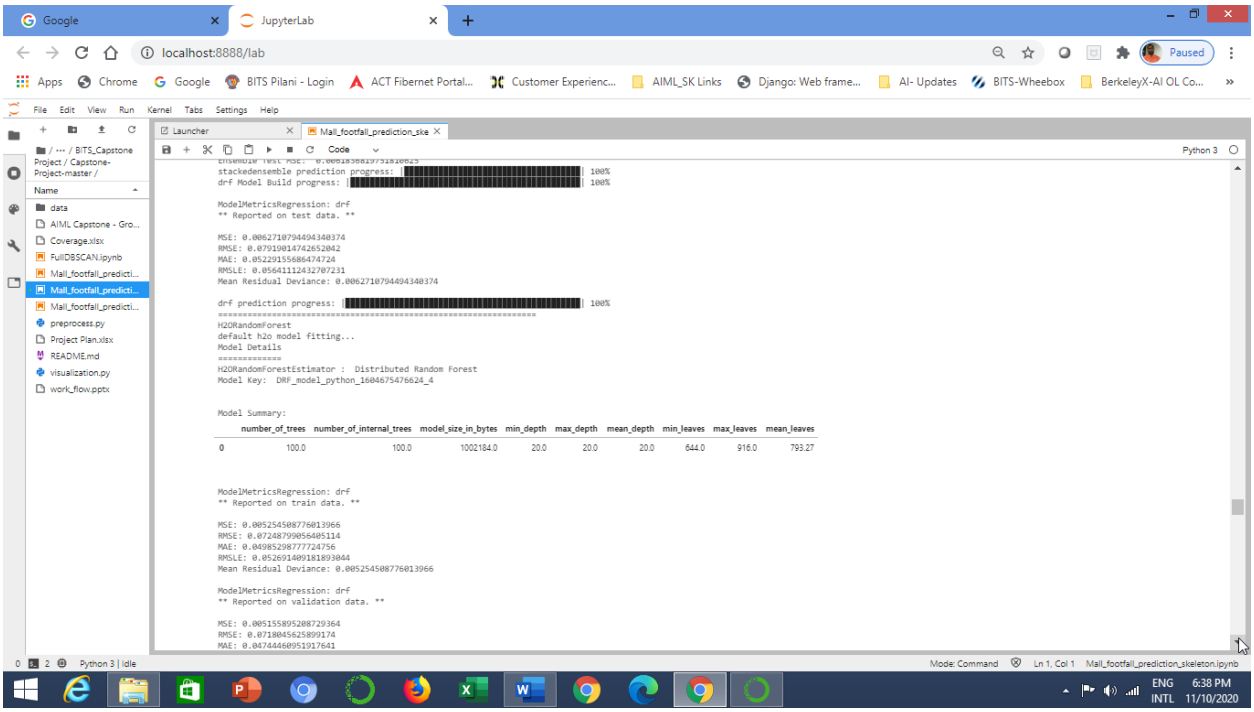
	col_sample_rate	...	model_ids	mse
0	0.3000000000000004	...	H2OGradientBoosting_model_18	0.005078830610218611
1	0.4	...	H2OGradientBoosting_model_8	0.005100857980913282
2	0.8	...	H2OGradientBoosting_model_10	0.005173206173441009
3	0.9	...	H2OGradientBoosting_model_1	0.005175976150547249
4	0.8	...	H2OGradientBoosting_model_20	0.005186587878554791
5	0.5	...	H2OGradientBoosting_model_13	0.005192807407250471
6	0.9	...	H2OGradientBoosting_model_16	0.00519882573892341
7	0.4	...	H2OGradientBoosting_model_19	0.005222258257214489
8	0.6000000000000001	...	H2OGradientBoosting_model_2	0.005260984600577699
9	0.1	...	H2OGradientBoosting_model_14	0.005264315765298267
10	0.2	...	H2OGradientBoosting_model_5	0.005385572187079132
11	0.9	...	H2OGradientBoosting_model_11	0.005393093249149702
12	0.2	...	H2OGradientBoosting_model_9	0.005435337356577598
13	0.7000000000000001	...	H2OGradientBoosting_model_12	0.005751081136713425
14	0.5	...	H2OGradientBoosting_model_4	0.00577981242048832
15	0.5	...	H2OGradientBoosting_model_7	0.005922159339701963
16	0.4	...	H2OGradientBoosting_model_17	0.006108504460256998
17	0.5	...	H2OGradientBoosting_model_6	0.006272360410998247
18	0.5	...	H2OGradientBoosting_model_15	0.006581327864128013
19	1.0	...	H2OGradientBoosting_model_3	0.007038508144141062 5

[20 rows x 7 columns]

ModelMetricsRegression: gbm
** Reported on test data. **

MSE: 0.006334536525046857 5
RMSE: 0.07958980163970044
MAE: 0.052126168929584424
RMSLE: 0.05672651712133378
Mean Residual Deviance: 0.006334536525046857 5

gbm prediction progress: | 100%
gbm Grid Build progress: | 100%
stackedensemble Model Build progress: | 100%
Best Base-learner Test MSE: 0.006129031090410484
Ensemble Test MSE: 0.006183681975181062 5
stackedensemble prediction progress: | 100%



stackedensemble prediction progress: | 100%
drf Model Build progress: | 100%

ModelMetricsRegression: drf
** Reported on test data. **

MSE: 0.006271079449434037 4
RMSE: 0.07919014742652042
MAE: 0.05229155686474724
RMSLE: 0.05641112432707231
Mean Residual Deviance: 0.006271079449434037 4

drf prediction progress: | 100%
H2ORandomForest
default h2o model fitting...
Model Details
H2ORandomForestEstimator : Distributed Random Forest
Model Key: DRF_model_python_1604675476624_4

Model Summary:

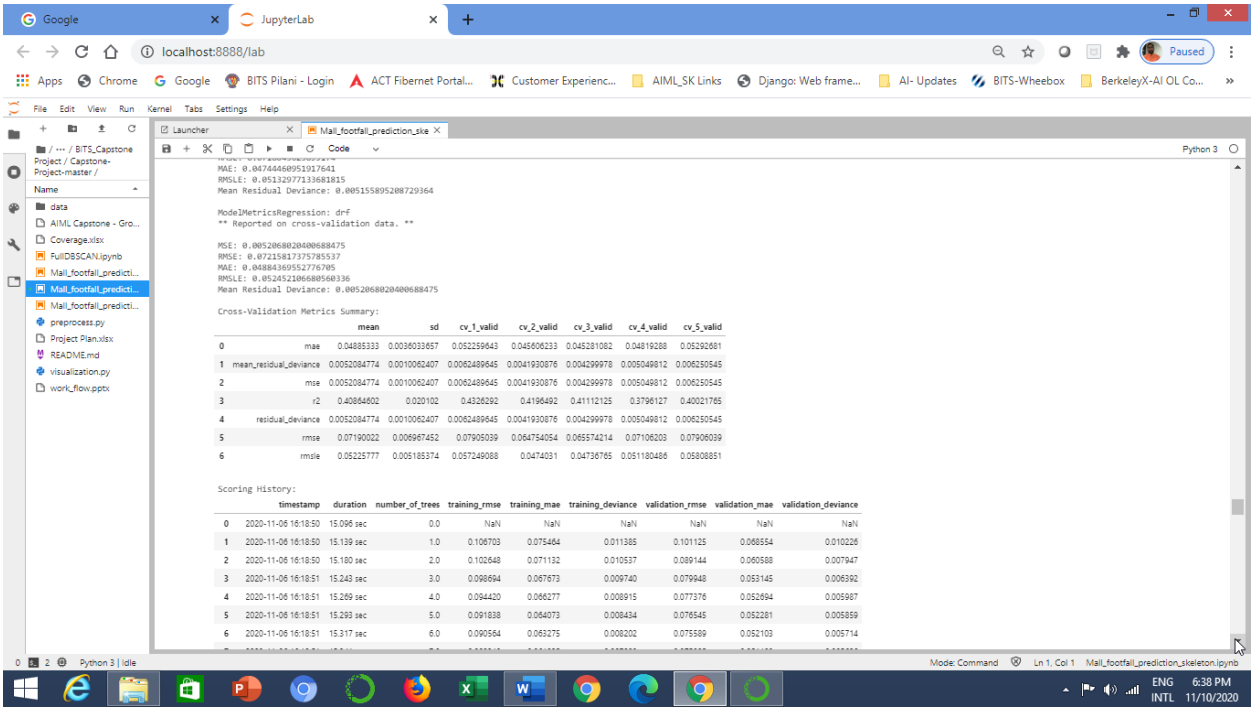
	number_of_trees	number_of_internal_trees	model_size_in_bytes	min_depth	max_depth	mean_depth	min_leaves	max_leaves	mean_leaves
0	100.0	100.0	1002184.0	20.0	20.0	20.0	644.0	916.0	793.27

ModelMetricsRegression: drf
** Reported on train data. **

MSE: 0.005254508776013966
RMSE: 0.07248799056405114
MAE: 0.04985298777724756
RMSLE: 0.05269140918189304 4
Mean Residual Deviance: 0.005254508776013966

ModelMetricsRegression: drf
** Reported on validation data. **

MSE: 0.005155895208729364
RMSE: 0.07180456259917 4
MAE: 0.04744468951917641

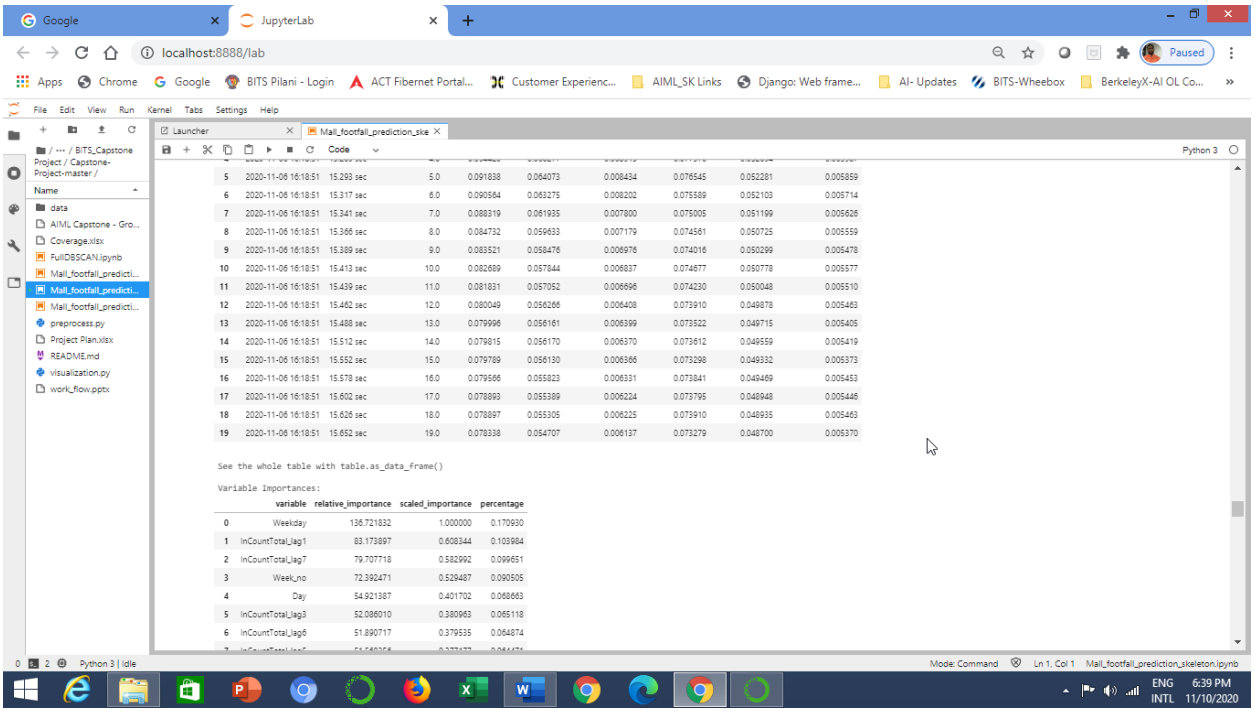


```
MAE: 0.04744460951917641
RMSLE: 0.05132977133681815
Mean Residual Deviance: 0.0051558952087293 64

ModelMetricsRegression: drf
** Reported on cross-validation data. **

MSE: 0.005206802040068847 5
RMSE: 0.07215817375785537
MAE: 0.04884369552776705
RMSLE: 0.05245210668056033 6
Mean Residual Deviance: 0.005206802040068847 5

Cross-Validation Metrics Summary:
```

		mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_valid	cv_5_valid
0	mae	0.04885333	0.0036033657	0.052259643	0.045606233	0.045281082	0.04819288	0.05292681
1	mean_residual_deviance	0.0052084774	0.0010062407	0.0062489645	0.0041930876	0.004299978	0.005049812	0.006250545
2	mse	0.0052084774	0.0010062407	0.0062489645	0.0041930876	0.004299978	0.005049812	0.006250545
3	r2	0.40864602	0.020102	0.4326292	0.4196492	0.41112125	0.3796127	0.40021765
4	residual_deviance	0.0052084774	0.0010062407	0.0062489645	0.0041930876	0.004299978	0.005049812	0.006250545
5	rmse	0.07190022	0.006967452	0.07905039	0.064754054	0.065574214	0.07106203	0.07906039
6	rmsle	0.05225777	0.005185374	0.057249088	0.0474031	0.04736765	0.051180486	0.05808851

```
Scoring History:
```

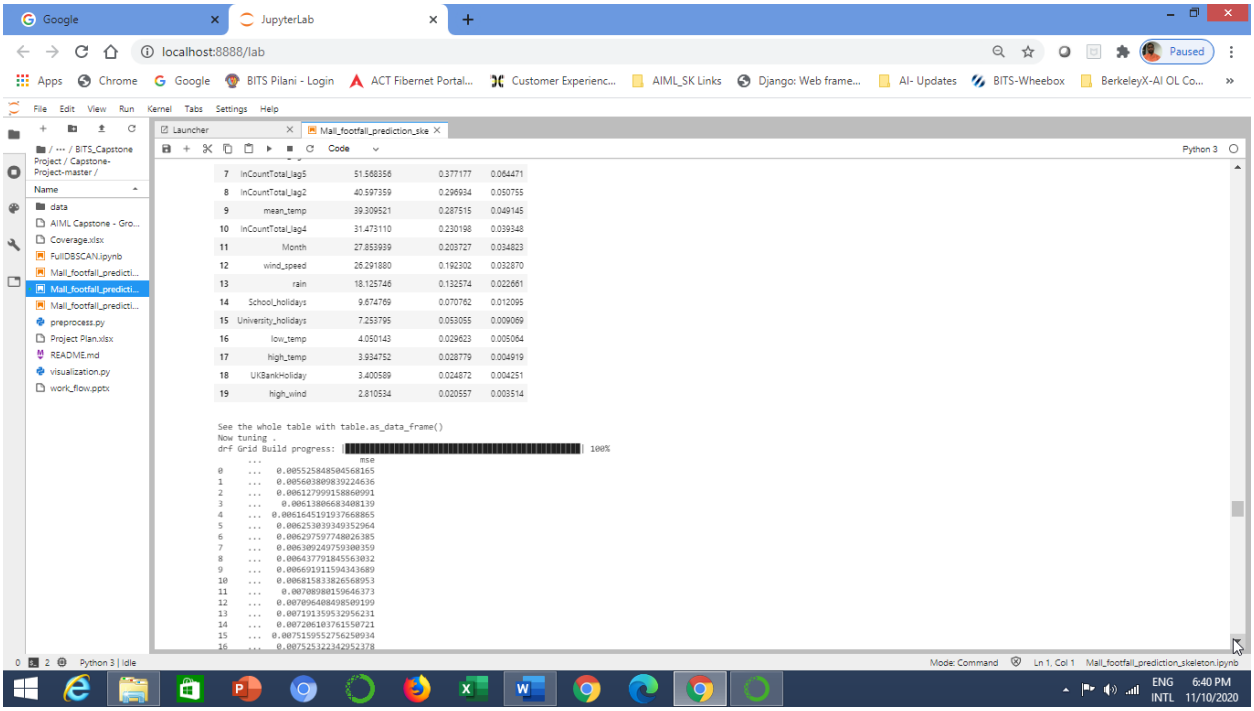
	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
0	2020-11-06 16:18:50	15.096 sec	0.0	NaN	NaN	NaN	NaN	NaN	NaN
1	2020-11-06 16:18:50	15.139 sec	1.0	0.106703	0.075464	0.011385	0.101125	0.068554	0.010226
2	2020-11-06 16:18:50	15.180 sec	2.0	0.102648	0.071132	0.010537	0.089144	0.060588	0.007947
3	2020-11-06 16:18:51	15.243 sec	3.0	0.098694	0.067673	0.009740	0.079948	0.053145	0.006392
4	2020-11-06 16:18:51	15.269 sec	4.0	0.094420	0.066277	0.008915	0.077376	0.052694	0.005987
5	2020-11-06 16:18:51	15.293 sec	5.0	0.091838	0.064073	0.008434	0.076545	0.052281	0.005859
6	2020-11-06 16:18:51	15.317 sec	6.0	0.090564	0.063275	0.008202	0.075589	0.052103	0.005714

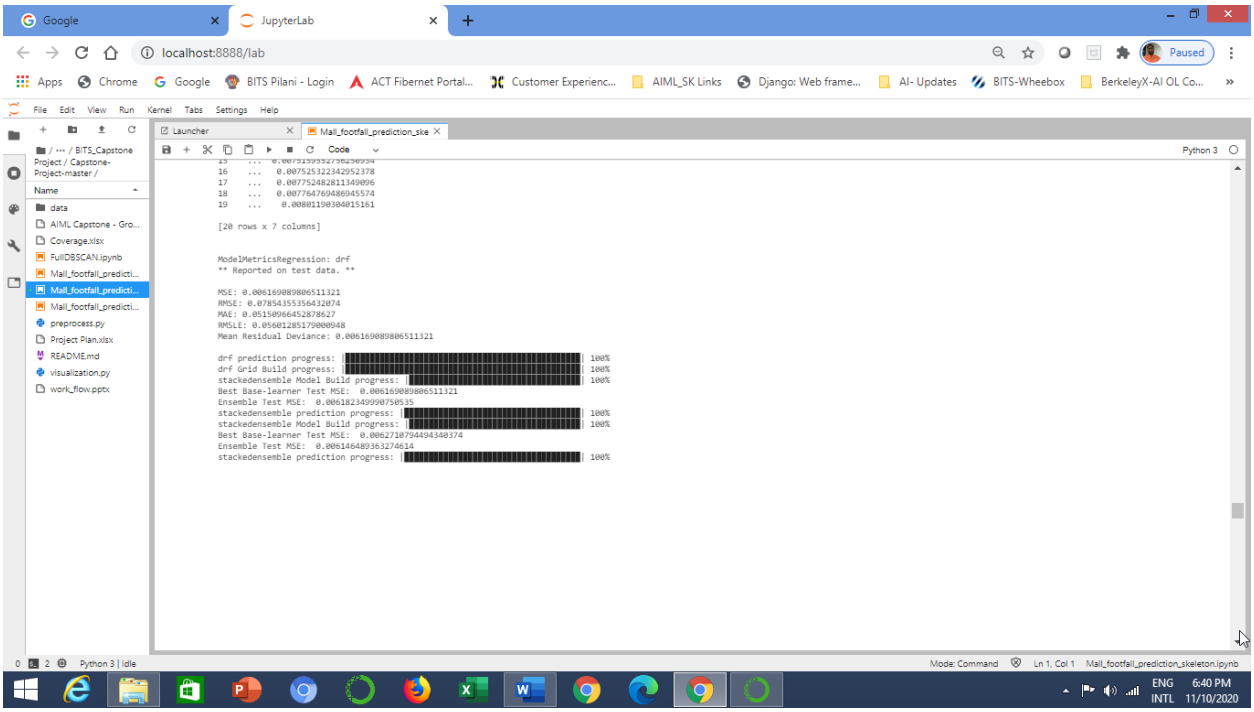


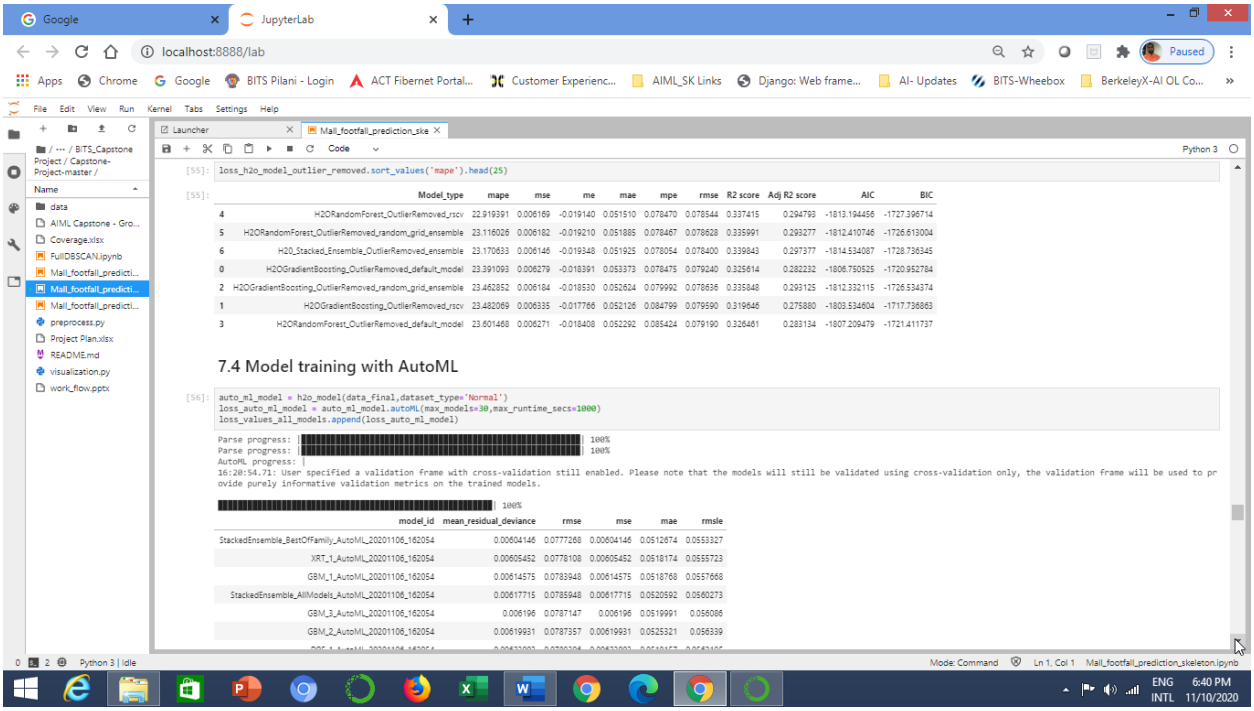
	timestamp	duration	number_of_trees	training_rmse	training_mae	training_deviance	validation_rmse	validation_mae	validation_deviance
5	2020-11-06 16:18:51	15.293 sec	5.0	0.091838	0.064073	0.008434	0.076545	0.052281	0.005859
6	2020-11-06 16:18:51	15.317 sec	6.0	0.090564	0.063275	0.008202	0.075589	0.052103	0.005714
7	2020-11-06 16:18:51	15.341 sec	7.0	0.088319	0.061935	0.007800	0.075005	0.051199	0.005626
8	2020-11-06 16:18:51	15.366 sec	8.0	0.084732	0.059633	0.007179	0.074561	0.050725	0.005559
9	2020-11-06 16:18:51	15.389 sec	9.0	0.083521	0.058476	0.006976	0.074016	0.050299	0.005478
10	2020-11-06 16:18:51	15.413 sec	10.0	0.082689	0.057844	0.006837	0.074677	0.050778	0.005577
11	2020-11-06 16:18:51	15.439 sec	11.0	0.081831	0.057052	0.006696	0.074230	0.050048	0.005510
12	2020-11-06 16:18:51	15.462 sec	12.0	0.080049	0.056266	0.006408	0.073910	0.049878	0.005463
13	2020-11-06 16:18:51	15.488 sec	13.0	0.079996	0.056161	0.006399	0.073522	0.049715	0.005405
14	2020-11-06 16:18:51	15.512 sec	14.0	0.079815	0.056170	0.006370	0.073612	0.049559	0.005419
15	2020-11-06 16:18:51	15.552 sec	15.0	0.079789	0.056130	0.006366	0.073298	0.049332	0.005373
16	2020-11-06 16:18:51	15.578 sec	16.0	0.079566	0.055823	0.006331	0.073841	0.049469	0.005453
17	2020-11-06 16:18:51	15.602 sec	17.0	0.078893	0.055389	0.006224	0.073795	0.048948	0.005446
18	2020-11-06 16:18:51	15.626 sec	18.0	0.078897	0.055305	0.006225	0.073910	0.048935	0.005463
19	2020-11-06 16:18:51	15.652 sec	19.0	0.078338	0.054707	0.006137	0.073279	0.048700	0.005370

```
See the whole table with table.as_data_frame()

Variable Importances:
```

	variable	relative_importance	scaled_importance	percentage
0	Weekday	136.721832	1.000000	0.170930
1	lnCountTotal_lag1	83.173897	0.608344	0.103984
2	lnCountTotal_lag7	79.707718	0.582992	0.099651
3	Week_no	72.392471	0.529487	0.090505
4	Day	54.921387	0.401702	0.068663
5	lnCountTotal_lag3	52.086010	0.380963	0.065118
6	lnCountTotal_lag6	51.890717	0.379535	0.064874


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
7 lnCountTotal_lag5 51.568356 0.377177 0.064471
8 lnCountTotal_lag2 40.597359 0.296934 0.050755
9 mean_temp 39.309521 0.287515 0.049145
10 lnCountTotal_lag4 31.473110 0.230198 0.039348
11 Month 27.853939 0.203727 0.034823
12 wind_speed 26.291880 0.192302 0.032870
13 rain 18.125746 0.132574 0.022661
14 School_holidays 9.674769 0.070762 0.012095
15 University_holidays 7.253795 0.053055 0.009069
16 low_temp 4.050143 0.029623 0.005064
17 high_temp 3.934752 0.028779 0.004919
18 UKBankHoliday 3.400589 0.024872 0.004251
19 high_wind 2.810534 0.020557 0.003514
See the whole table with table.as_data_frame()
Now tuning .
drf Grid Build progress: | 100%
... mse
0 ... 0.005525848504568165
1 ... 0.005603809839224636
2 ... 0.006127999158860991
3 ... 0.0061380668340813 9
4 ... 0.006164519193766886 5
5 ... 0.006253039349352964
6 ... 0.006297597748026385
7 ... 0.006309249759300359
8 ... 0.006437791845563032
9 ... 0.006691911594343689
10 ... 0.006815833826568953
11 ... 0.007080980159646373
12 ... 0.007096408498509199
13 ... 0.007191359532956231
14 ... 0.007206103761550721
15 ... 0.007515955276250934
16 ... 0.007525322342952378
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:40 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Launcher
Mall_footfall_prediction_ske
Code
Python 3
16 ... 0.007525322342952378
17 ... 0.007752482811340096
18 ... 0.007764769486945574
19 ... 0.008011903040151161
[20 rows x 7 columns]
ModelMetricsRegression: drf
** Reported on test data. **
MSE: 0.006169089806511321
RMSE: 0.07854355356432074
MAE: 0.05150966452878627
RMSLE: 0.05601285179000948
Mean Residual Deviance: 0.006169089806511321
drf prediction progress: | 100%
drf Grid Build progress: | 100%
stackedensemble Model Build progress: | 100%
Best Base-learner Test MSE: 0.006169089806511321
Ensemble Test MSE: 0.006182349990750535
stackedensemble prediction progress: | 100%
stackedensemble Model Build progress: | 100%
Best Base-learner Test MSE: 0.006271079494340374
Ensemble Test MSE: 0.006146489363274614
stackedensemble prediction progress: | 100%
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:40 PM
11/10/2020



```
loss_h2o_model_outlier_removed.sort_values('mape').head(25)
```

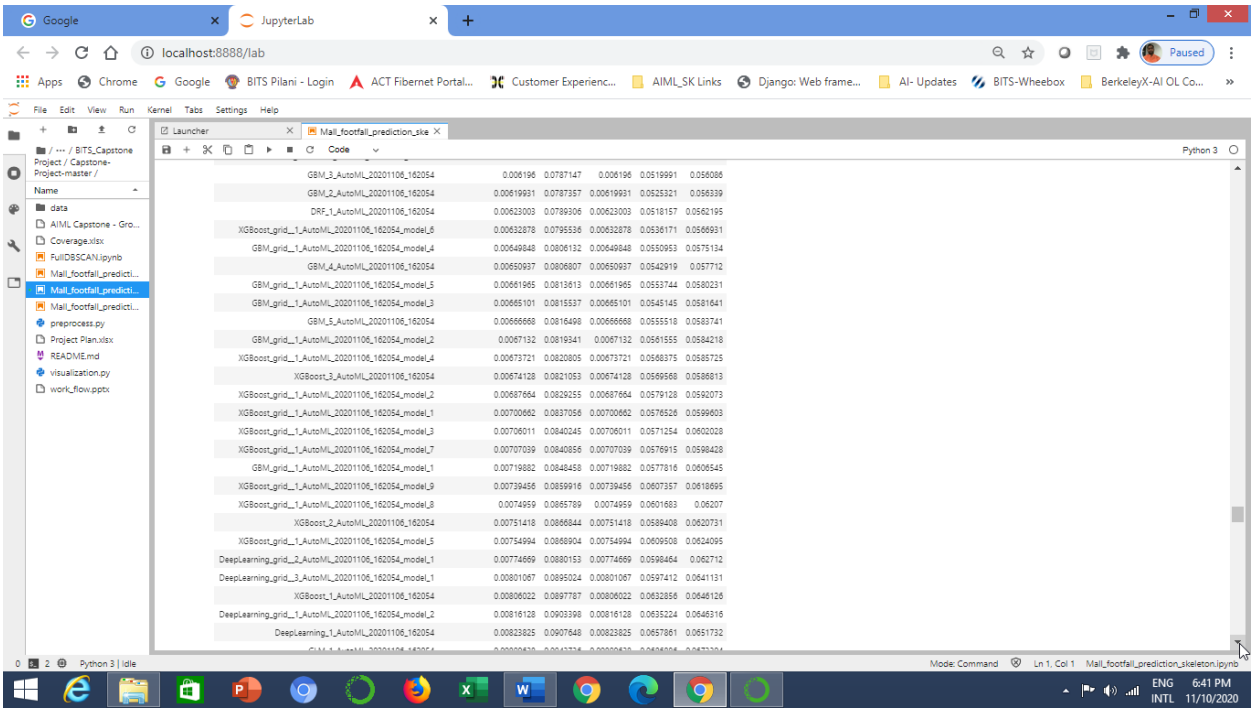
	Model_type	mape	mse	me	mae	mpe	rmse	R2 score	Adj R2 score	AIC	BIC
4	H2ORandomForest_OutlierRemoved_rscv	22.919391	0.006169	-0.019140	0.051510	0.078470	0.078544	0.337415	0.294793	-1813.194456	-1727.396714
5	H2ORandomForest_OutlierRemoved_random_grid_ensemble	23.116026	0.006182	-0.019210	0.051885	0.078467	0.078628	0.335991	0.293277	-1812.410746	-1726.613004
6	H2O_Stacked_Ensemble_OutlierRemoved_ensemble	23.170633	0.006146	-0.019348	0.051925	0.078054	0.078400	0.339843	0.297377	-1814.534087	-1728.736345
0	H2OGradientBoosting_OutlierRemoved_default_model	23.391093	0.006279	-0.018391	0.053373	0.078475	0.079240	0.325614	0.282232	-1806.750525	-1720.952784
2	H2OGradientBoosting_OutlierRemoved_random_grid_ensemble	23.462852	0.006184	-0.018530	0.052624	0.079992	0.078636	0.335848	0.293125	-1812.332115	-1726.534374
1	H2OGradientBoosting_OutlierRemoved_rscv	23.482069	0.006335	-0.017766	0.052126	0.084799	0.079590	0.319646	0.275880	-1803.534604	-1717.736863
3	H2ORandomForest_OutlierRemoved_default_model	23.601468	0.006271	-0.018408	0.052292	0.085424	0.079190	0.326461	0.283134	-1807.209479	-1721.411737

7.4 Model training with AutoML

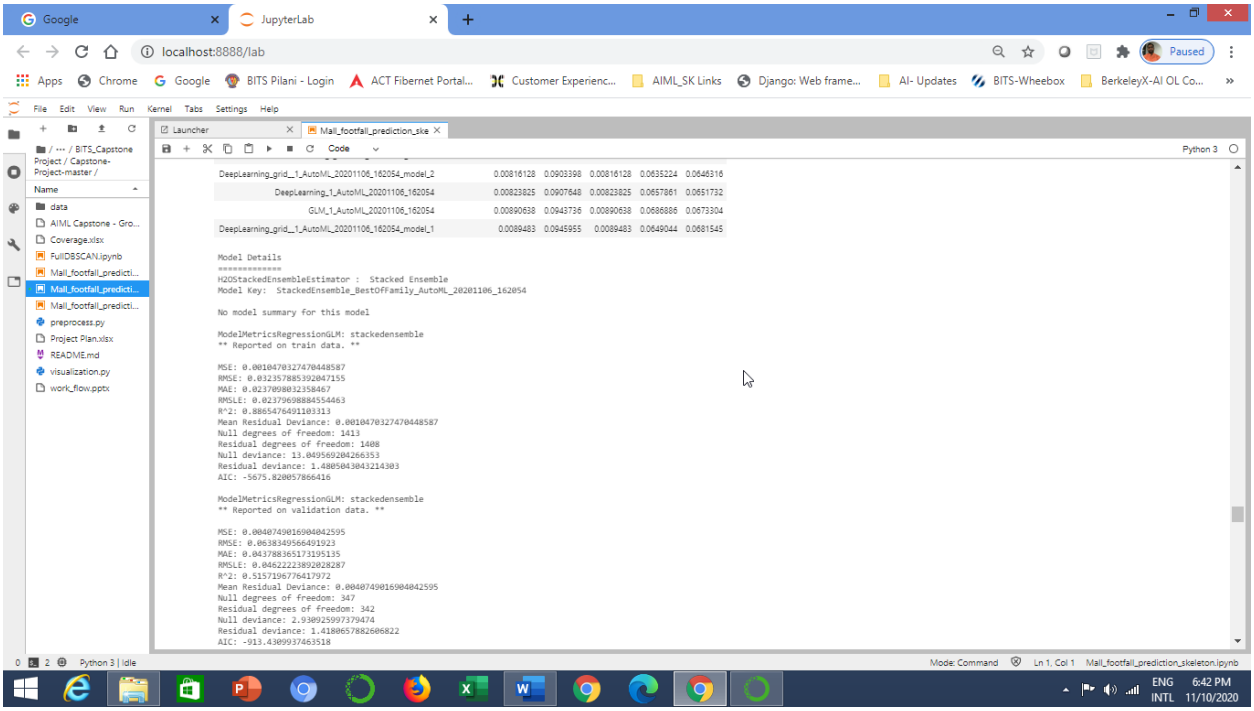
```
auto_ml_model = h2o_model(data_final,dataset_type='Normal')
loss_auto_ml_model = auto_ml_model.autoML(max_models=30,max_runtime_secs=1000)
loss_values_all_models.append(loss_auto_ml_model)
```

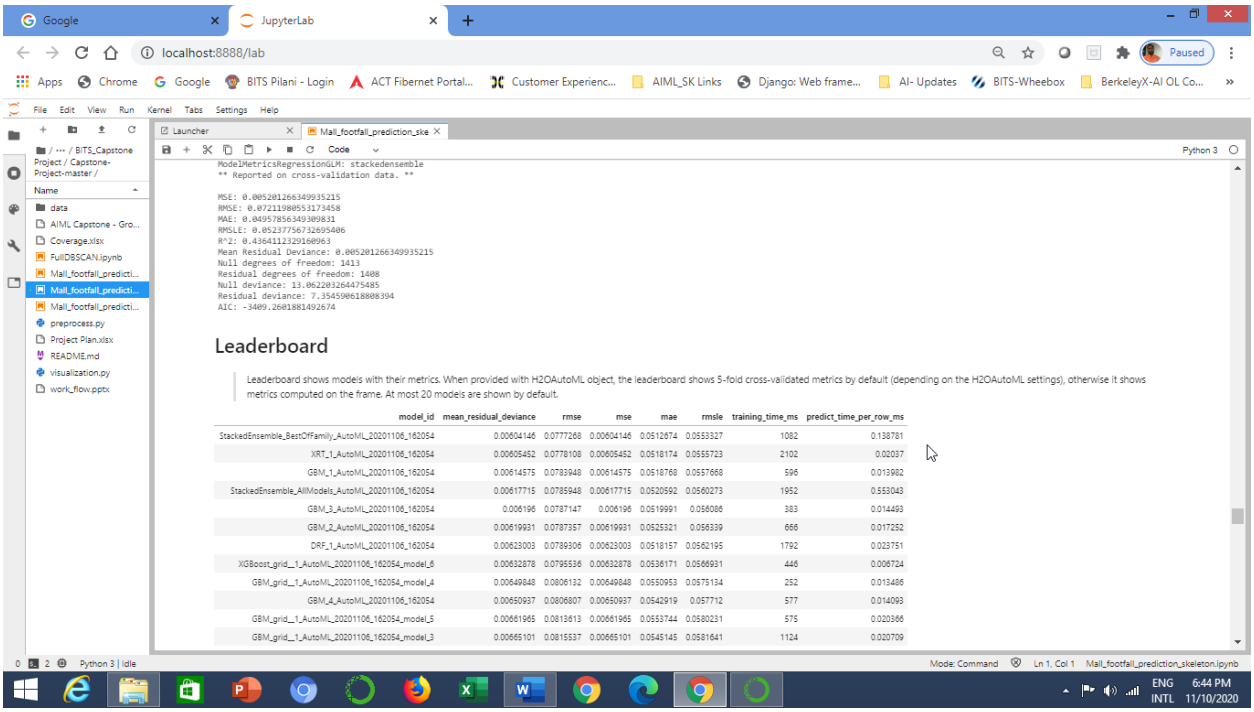
```
Parse progress: |████████████████████████████████████████| 100%
Parse progress: |████████████████████████████████████████| 100%
AutoML progress: |
16:20:54.71: User specified a validation frame with cross-validation still enabled. Please note that the models will still be validated using cross-validation only, the validation frame will be used to pr
ovide purely informative validation metrics on the trained models.

|████████████████████████████████████████| 100%
```

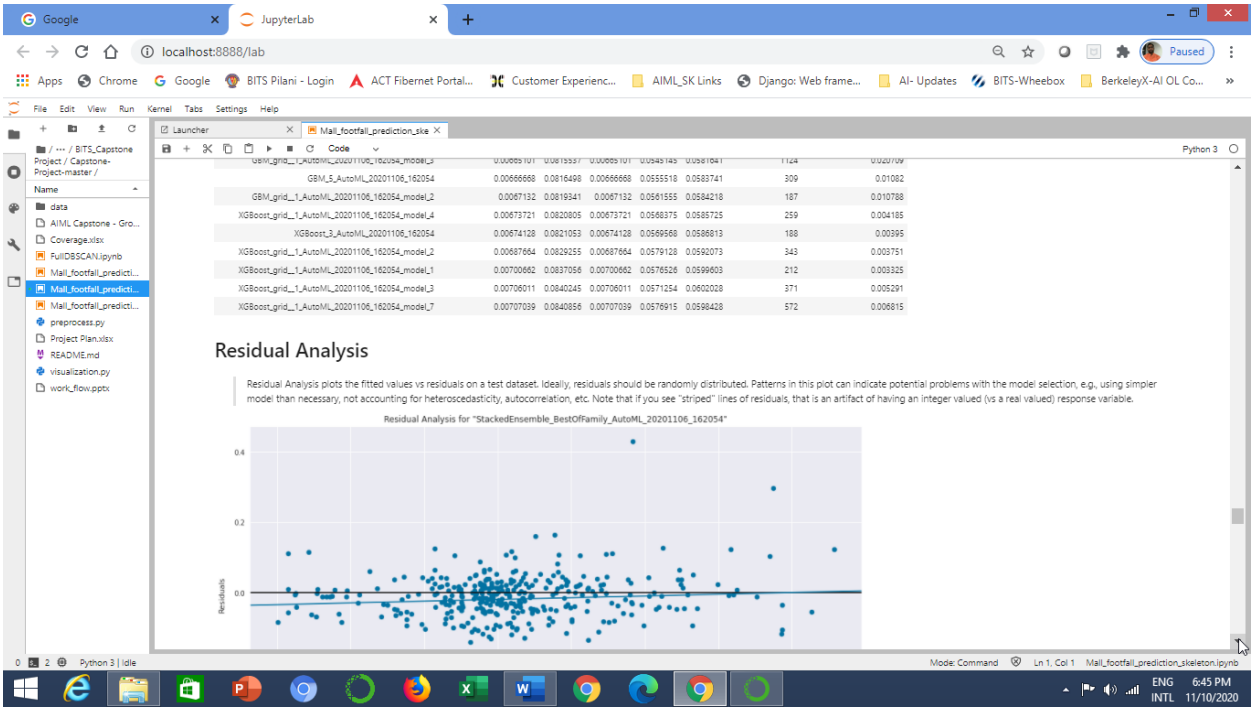
model_id	mean_residual_deviance	rmse	mse	mae	rmsle
StackedEnsemble_BestOfFamily_AutoML_20201106_162054	0.00604146	0.0777268	0.00604146	0.0512674	0.0553327
XRT_1_AutoML_20201106_162054	0.00605452	0.0778108	0.00605452	0.0518174	0.0555723
GBM_1_AutoML_20201106_162054	0.00614575	0.0783948	0.00614575	0.0518768	0.0557668
StackedEnsemble_AllModels_AutoML_20201106_162054	0.00617715	0.0785948	0.00617715	0.0520592	0.0560273
GBM_3_AutoML_20201106_162054	0.006196	0.0787147	0.006196	0.0519991	0.056086
GBM_2_AutoML_20201106_162054	0.00619931	0.0787357	0.00619931	0.0523521	0.056339

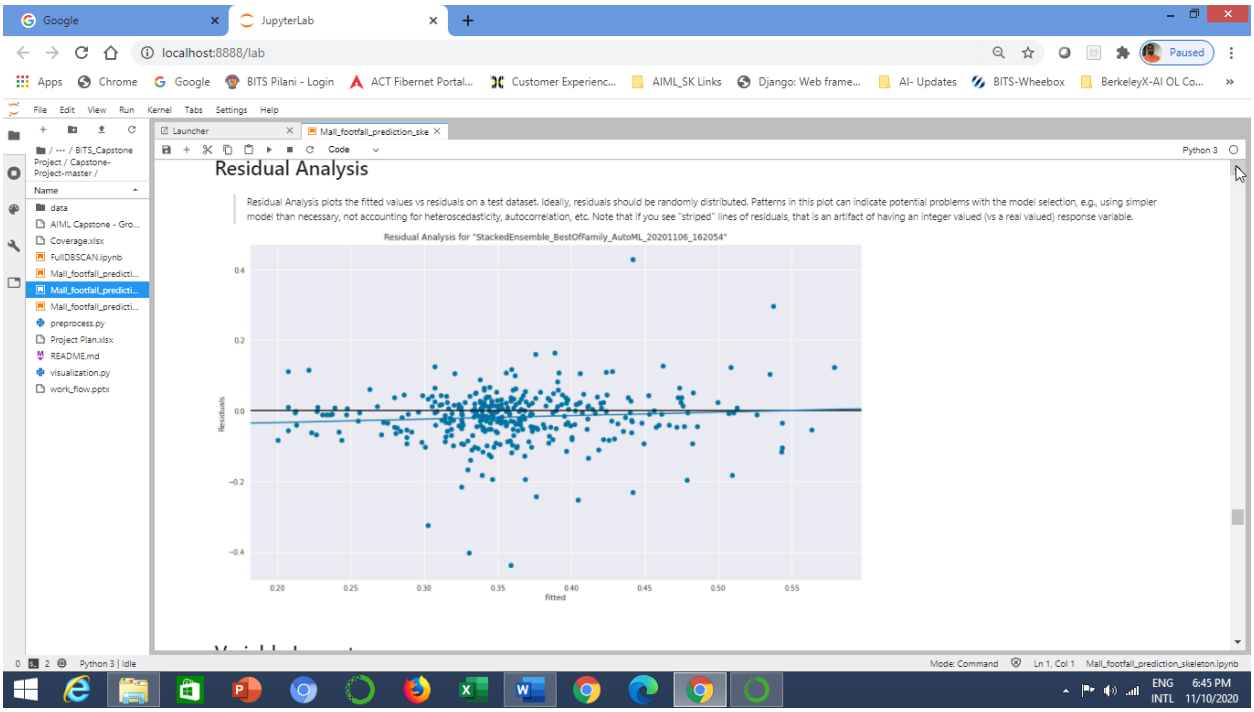


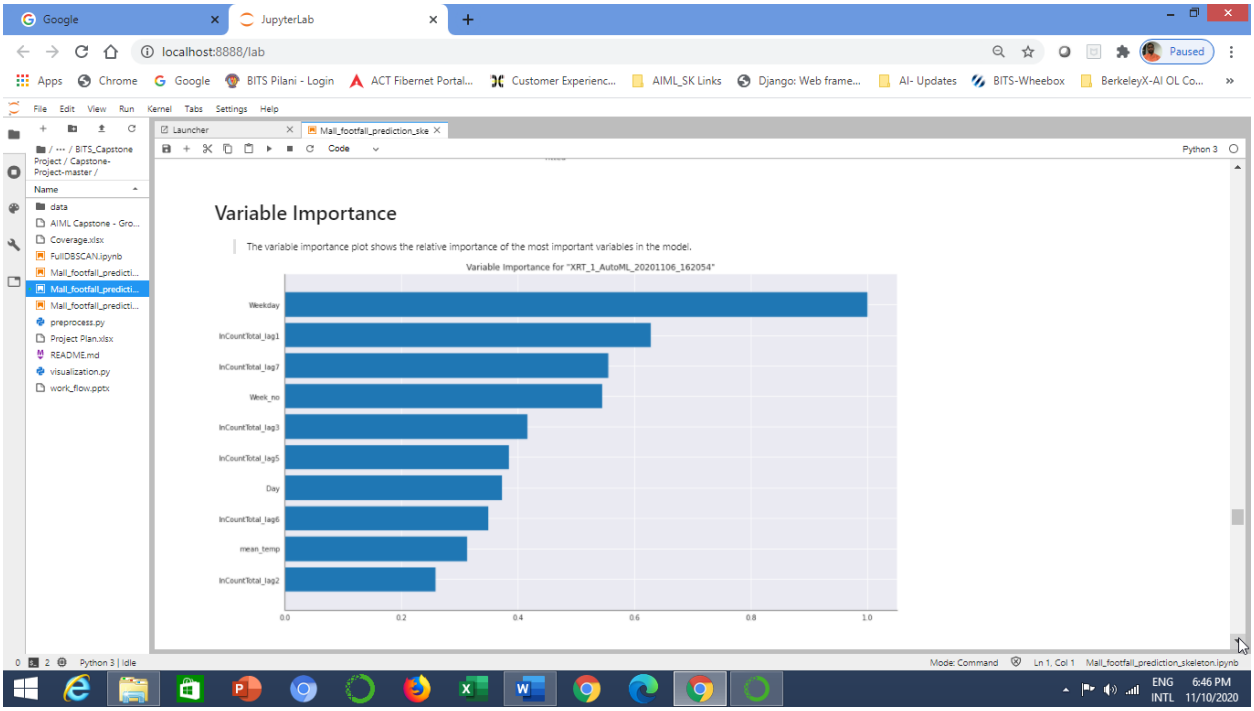
GBM_3_AutoML_20201106_162054	0.006196	0.0787147	0.006196	0.0519991	0.056086
GBM_2_AutoML_20201106_162054	0.00619931	0.0787357	0.00619931	0.0523521	0.056339
DRF_1_AutoML_20201106_162054	0.00623003	0.0789306	0.00623003	0.0518157	0.0562195
XGBoost_grid__1_AutoML_20201106_162054_model_6	0.00632878	0.0795536	0.00632878	0.0536171	0.0566931
GBM_grid__1_AutoML_20201106_162054_model_4	0.00649848	0.0806132	0.00649848	0.0550953	0.0575134
GBM_4_AutoML_20201106_162054	0.00650937	0.0806807	0.00650937	0.0542919	0.057712
GBM_grid__1_AutoML_20201106_162054_model_5	0.00661965	0.0813613	0.00661965	0.0553744	0.0580231
GBM_grid__1_AutoML_20201106_162054_model_3	0.00665101	0.0815537	0.00665101	0.0545145	0.0581641
GBM_5_AutoML_20201106_162054	0.00666668	0.0816498	0.00666668	0.0555518	0.0583741
GBM_grid__1_AutoML_20201106_162054_model_2	0.0067132	0.0819341	0.0067132	0.0561555	0.0584218
XGBoost_grid__1_AutoML_20201106_162054_model_4	0.00673721	0.0820805	0.00673721	0.0568375	0.0585725
XGBoost_3_AutoML_20201106_162054	0.00674128	0.0821053	0.00674128	0.0569568	0.0586813
XGBoost_grid__1_AutoML_20201106_162054_model_2	0.00687664	0.0829255	0.00687664	0.0579128	0.0592073
XGBoost_grid__1_AutoML_20201106_162054_model_1	0.00700662	0.0837056	0.00700662	0.0576526	0.0599603
XGBoost_grid__1_AutoML_20201106_162054_model_3	0.00706011	0.0840245	0.00706011	0.0571254	0.0602028
XGBoost_grid__1_AutoML_20201106_162054_model_7	0.00707039	0.0840856	0.00707039	0.0576915	0.0598428
GBM_grid__1_AutoML_20201106_162054_model_1	0.00719882	0.0848458	0.00719882	0.0577816	0.0606545
XGBoost_grid__1_AutoML_20201106_162054_model_9	0.00739456	0.0859916	0.00739456	0.0607357	0.0618695
XGBoost_grid__1_AutoML_20201106_162054_model_8	0.0074959	0.0865789	0.0074959	0.0601683	0.06207
XGBoost_2_AutoML_20201106_162054	0.00751418	0.0866844	0.00751418	0.0589408	0.0620731
XGBoost_grid__1_AutoML_20201106_162054_model_5	0.00754994	0.0868904	0.00754994	0.0609508	0.0624095
DeepLearning_grid__2_AutoML_20201106_162054_model_1	0.00774669	0.0880153	0.00774669	0.0598464	0.062712
DeepLearning_grid__3_AutoML_20201106_162054_model_1	0.00801067	0.0895024	0.00801067	0.0597412	0.0641131
XGBoost_1_AutoML_20201106_162054	0.00806022	0.0897787	0.00806022	0.0632856	0.0646126
DeepLearning_grid__1_AutoML_20201106_162054_model_2	0.00816128	0.0903398	0.00816128	0.0635224	0.0646316
DeepLearning_1_AutoML_20201106_162054	0.00823825	0.0907648	0.00823825	0.0657861	0.0651732

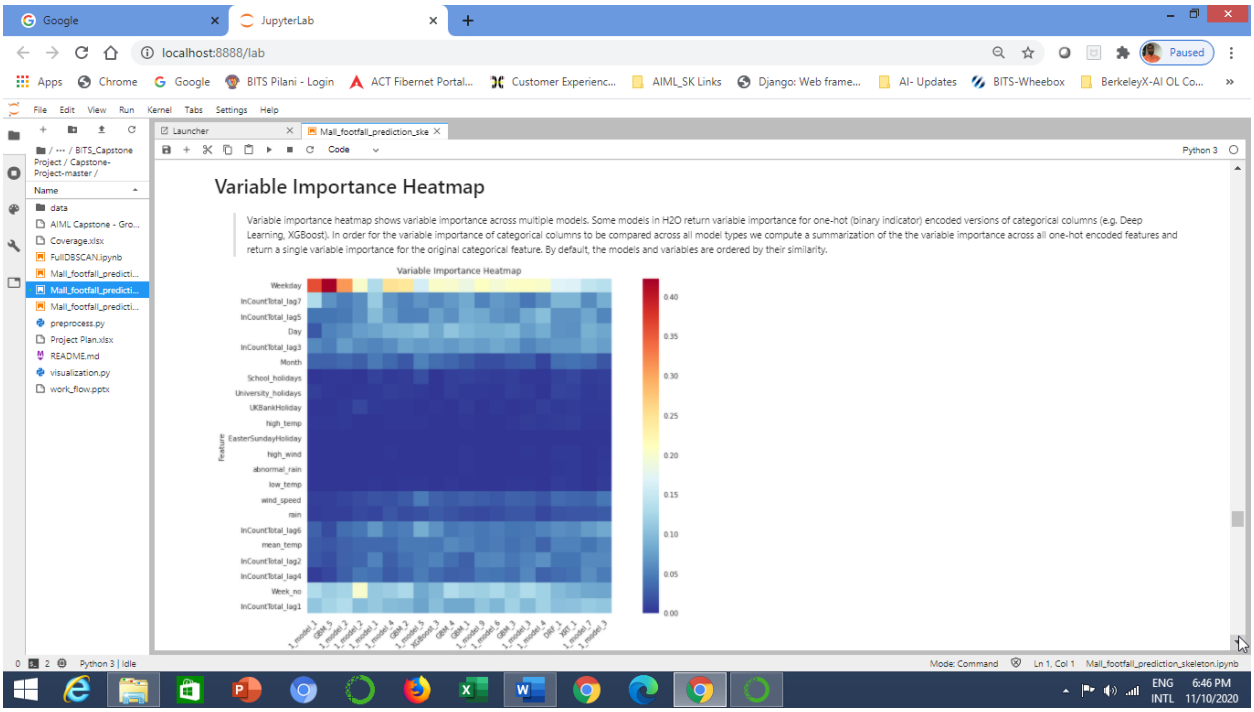

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
DeepLearning_grid_1_AutoML_20201106_162054_model_2 0.00816128 0.0903398 0.00816128 0.0635224 0.0646316
DeepLearning_1_AutoML_20201106_162054 0.00823825 0.0907648 0.00823825 0.0657861 0.0651732
GLM_1_AutoML_20201106_162054 0.00890638 0.0943736 0.00890638 0.0686886 0.0673304
DeepLearning_grid_1_AutoML_20201106_162054_model_1 0.0089483 0.0945955 0.0089483 0.0649044 0.0681545
Model Details
==========
H2OStackedEnsembleEstimator : Stacked Ensemble
Model Key: StackedEnsemble_BestOfFamily_AutoML_20201106_162054
No model summary for this model
ModelMetricsRegressionGLM: stackedensemble
** Reported on train data. **
MSE: 0.001047032747048587
RMSE: 0.032357885392047155
MAE: 0.023709803235846 7
RMSLE: 0.023796988845544 63
R^2: 0.8865476491103113
Mean Residual Deviance: 0.001047032747048587
Null degrees of freedom: 1413
Residual degrees of freedom: 1408
Null deviance: 13.049569284266353
Residual deviance: 1.480504384321430 3
AIC: -5675.820057866416
ModelMetricsRegressionGLM: stackedensemble
** Reported on validation data. **
MSE: 0.004074901690404259 5
RMSE: 0.06383495664919 23
MAE: 0.043788365173195135
RMSLE: 0.04622223892928287
R^2: 0.5157196776417972
Mean Residual Deviance: 0.004074901690404259 5
Null degrees of freedom: 347
Residual degrees of freedom: 342
Null deviance: 2.938925997379474
Residual deviance: 1.418065788260682 2
AIC: -913.4309937463518
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:42 PM 11/10/2020

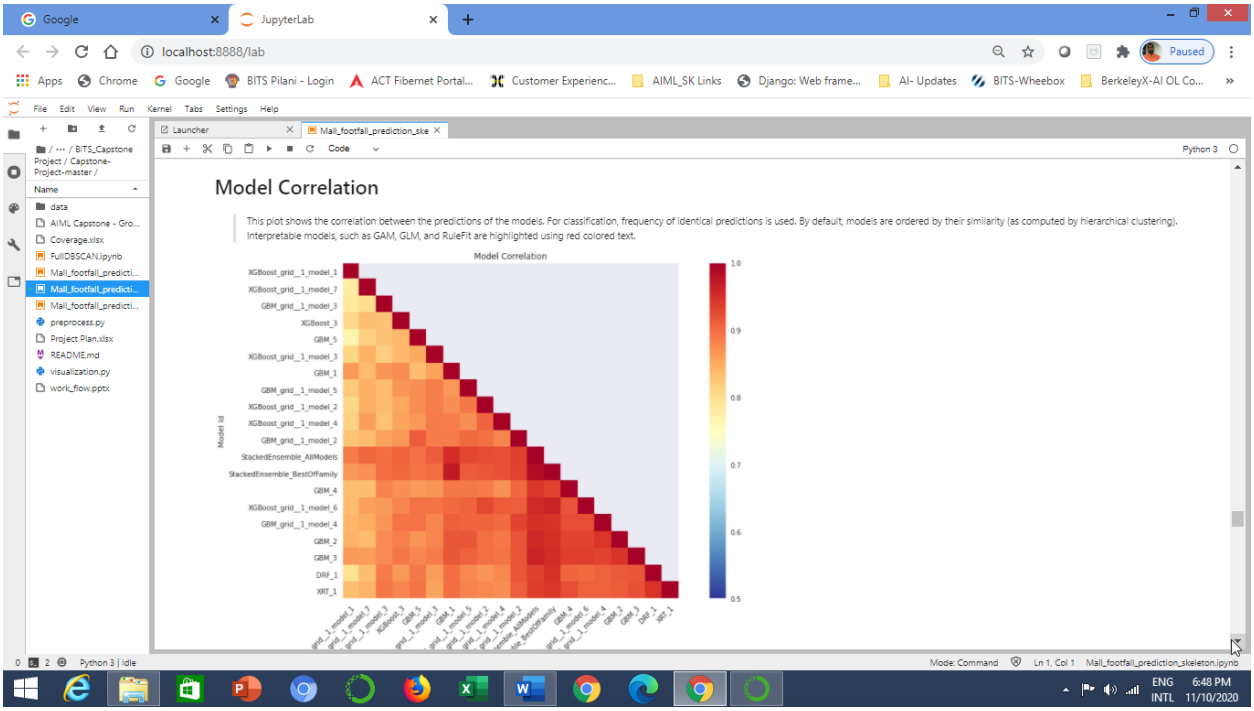

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
/ ... / BITS_Capstone Project / Capstone-Project-master /
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
ModelMetricsRegressionGLM: stackedensemble
** Reported on cross-validation data. **
MSE: 0.005201266349935215
RMSE: 0.07211980553173458
MAE: 0.049578563493098 31
RMSLE: 0.05237756732695406
R^2: 0.4364112329160963
Mean Residual Deviance: 0.005201266349935215
Null degrees of freedom: 1413
Residual degrees of freedom: 1408
Null deviance: 13.062289264475485
Residual deviance: 7.354590618808394
AIC: -3409.260188149267 4
Leaderboard
Leaderboard shows models with their metrics. When provided with H2OAutoML object, the leaderboard shows 5-fold cross-validated metrics by default (depending on the H2OAutoML settings), otherwise it shows metrics computed on the frame. At most 20 models are shown by default.
model_id mean_residual_deviance rmse mse mae rmsle training_time_ms predict_time_per_row_ms
StackedEnsemble_BestOfFamily_AutoML_20201106_162054 0.00604146 0.0777268 0.00604146 0.0512674 0.0553327 1082 0.138781
XRT_1_AutoML_20201106_162054 0.00605452 0.0778108 0.00605452 0.0518174 0.0555723 2102 0.02037
GBM_1_AutoML_20201106_162054 0.00614575 0.0783948 0.00614575 0.0518768 0.0557668 596 0.013982
StackedEnsemble_AllModels_AutoML_20201106_162054 0.00617715 0.0785948 0.00617715 0.0520592 0.0560273 1952 0.553043
GBM_3_AutoML_20201106_162054 0.006196 0.0787147 0.006196 0.0519991 0.056086 383 0.014493
GBM_2_AutoML_20201106_162054 0.00619931 0.0787357 0.00619931 0.0525321 0.056339 666 0.017252
DRF_1_AutoML_20201106_162054 0.00623003 0.0789306 0.00623003 0.0518157 0.0562195 1792 0.023751
XGBoost_grid_1_AutoML_20201106_162054_model_6 0.00632878 0.0795536 0.00632878 0.0536171 0.0566931 446 0.006724
GBM_grid_1_AutoML_20201106_162054_model_4 0.00649848 0.0806132 0.00649848 0.0550953 0.0575134 252 0.013486
GBM_4_AutoML_20201106_162054 0.00650937 0.0806807 0.00650937 0.0542919 0.057712 577 0.014093
GBM_grid_1_AutoML_20201106_162054_model_5 0.00661965 0.0813613 0.00661965 0.0553744 0.0580231 575 0.020366
GBM_grid_1_AutoML_20201106_162054_model_3 0.00665101 0.0815537 0.00665101 0.0545145 0.0581641 1124 0.020709
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
Python 3 | Idle
ENG INTL
6:44 PM 11/10/2020

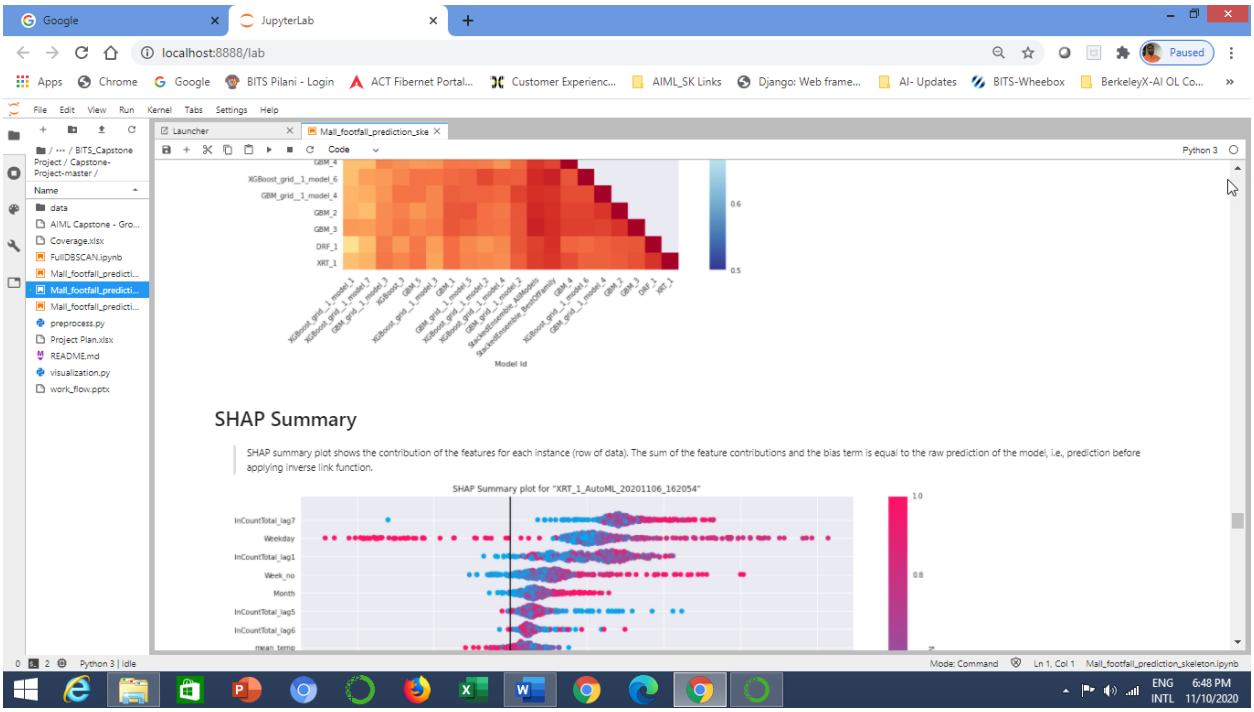

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
GBM_5_AutoML_20201106_162054 0.00666668 0.0816498 0.00666668 0.0555518 0.0583741 309 0.01082
GBM_grid__1_AutoML_20201106_162054_model_2 0.0067132 0.0819341 0.0067132 0.0561555 0.0584218 187 0.010788
XGBoost_grid__1_AutoML_20201106_162054_model_4 0.00673721 0.0820805 0.00673721 0.0568375 0.0585725 259 0.004185
XGBoost_3_AutoML_20201106_162054 0.00674128 0.0821053 0.00674128 0.0569568 0.0586813 188 0.00395
XGBoost_grid__1_AutoML_20201106_162054_model_2 0.00687664 0.0829255 0.00687664 0.0579128 0.0592073 343 0.003751
XGBoost_grid__1_AutoML_20201106_162054_model_1 0.00700662 0.0837056 0.00700662 0.0576526 0.0599603 212 0.003325
XGBoost_grid__1_AutoML_20201106_162054_model_3 0.00706011 0.0840245 0.00706011 0.0571254 0.0602028 371 0.005291
XGBoost_grid__1_AutoML_20201106_162054_model_7 0.00707039 0.0840856 0.00707039 0.0576915 0.0598428 572 0.006815
Residual Analysis
Residual Analysis plots the fitted values vs residuals on a test dataset. Ideally, residuals should be randomly distributed. Patterns in this plot can indicate potential problems with the model selection, e.g., using simpler model than necessary, not accounting for heteroscedasticity, autocorrelation, etc. Note that if you see "striped" lines of residuals, that is an artifact of having an integer valued (vs a real valued) response variable.
Residual Analysis for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
Residuals
0.4
0.2
0.0
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:45 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Residual Analysis
Residual Analysis plots the fitted values vs residuals on a test dataset. Ideally, residuals should be randomly distributed. Patterns in this plot can indicate potential problems with the model selection, e.g., using simpler model than necessary, not accounting for heteroscedasticity, autocorrelation, etc. Note that if you see "striped" lines of residuals, that is an artifact of having an integer valued (vs a real valued) response variable.
Residual Analysis for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
Residuals
0.4
0.2
0.0
-0.2
-0.4
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
fitted
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:45 PM
11/10/2020

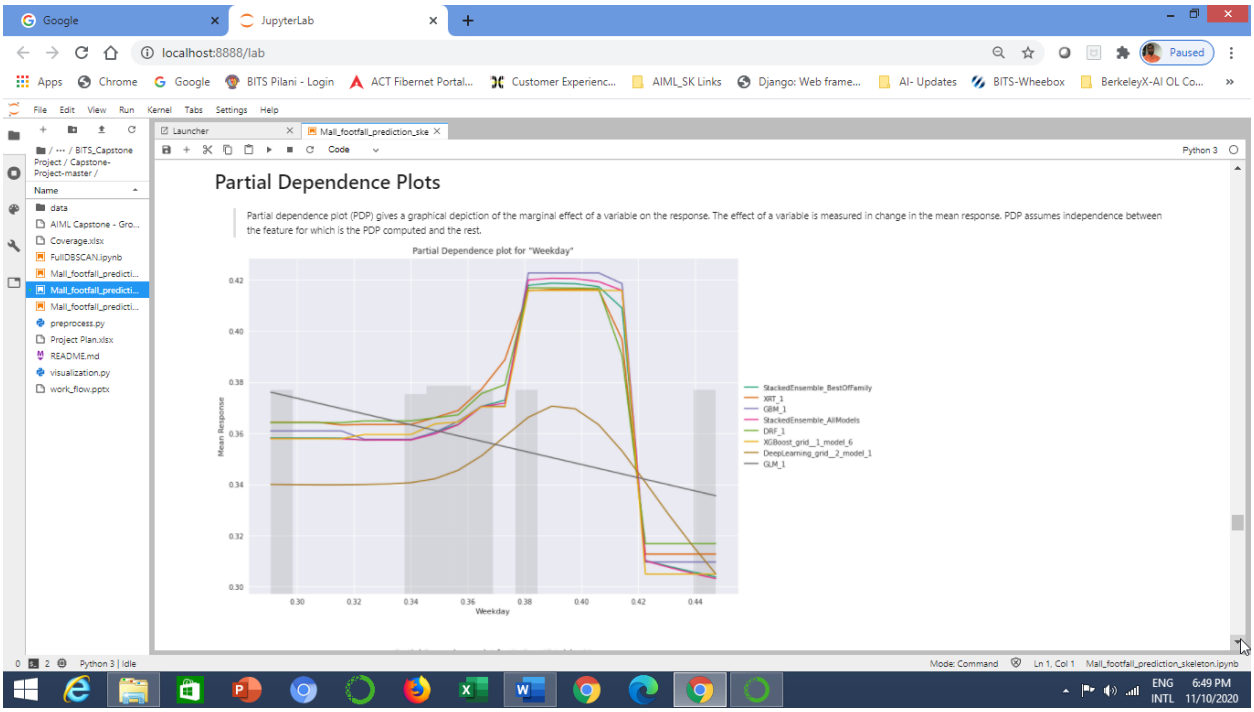

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Variable Importance
The variable importance plot shows the relative importance of the most important variables in the model.
Variable Importance for "XRT_1_AutoML_20201106_162054"
Weekday
lnCountTotal_lag1
lnCountTotal_lag7
Week_no
lnCountTotal_lag3
lnCountTotal_lag5
Day
lnCountTotal_lag6
mean_temp
lnCountTotal_lag2
0.0
0.2
0.4
0.6
0.8
1.0
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:46 PM
11/10/2020

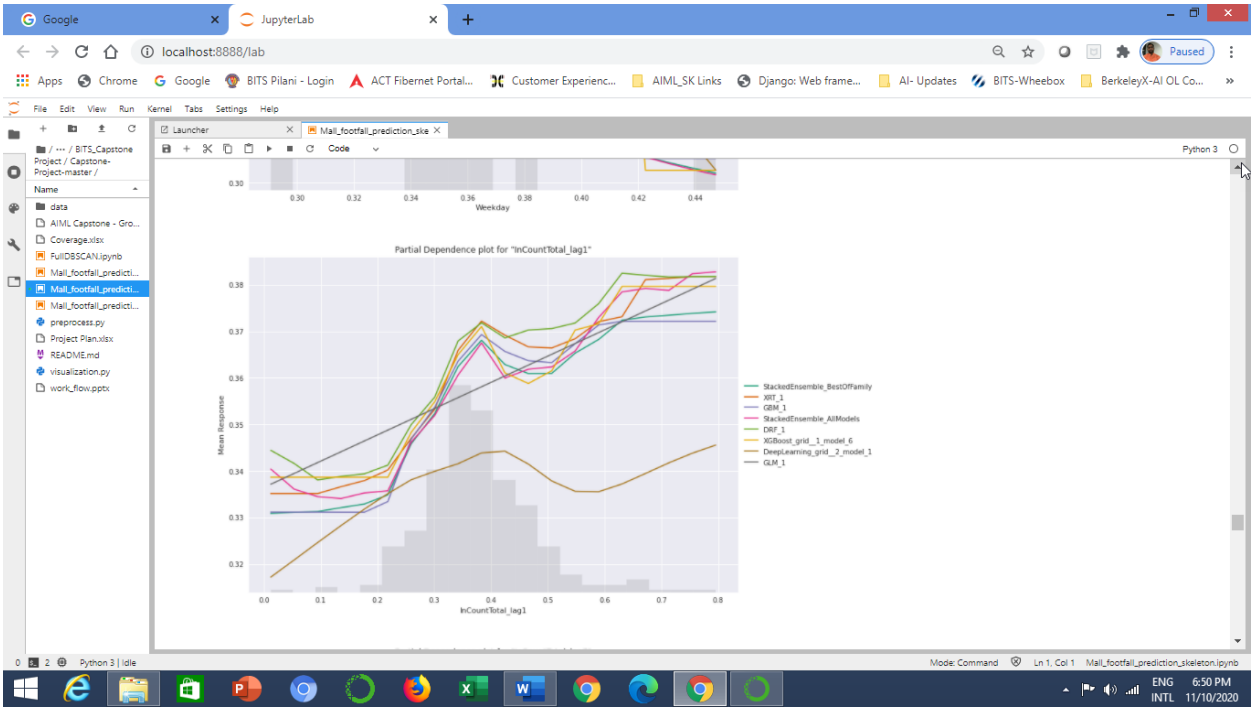

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Variable Importance Heatmap
Variable importance heatmap shows variable importance across multiple models. Some models in H2O return variable importance for one-hot (binary indicator) encoded versions of categorical columns (e.g. Deep Learning, XGBoost). In order for the variable importance of categorical columns to be compared across all model types we compute a summarization of the the variable importance across all one-hot encoded features and return a single variable importance for the original categorical feature. By default, the models and variables are ordered by their similarity.
Variable Importance Heatmap
Weekday
lnCountTotal_lag7
lnCountTotal_lag5
Day
lnCountTotal_lag3
Month
School_holidays
University_holidays
UKBankHoliday
high_temp
EasterSundayHoliday
high_wind
abnormal_rain
low_temp
wind_speed
rain
lnCountTotal_lag6
mean_temp
lnCountTotal_lag2
lnCountTotal_lag4
Week_no
lnCountTotal_lag1
Feature
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:46 PM
11/10/2020

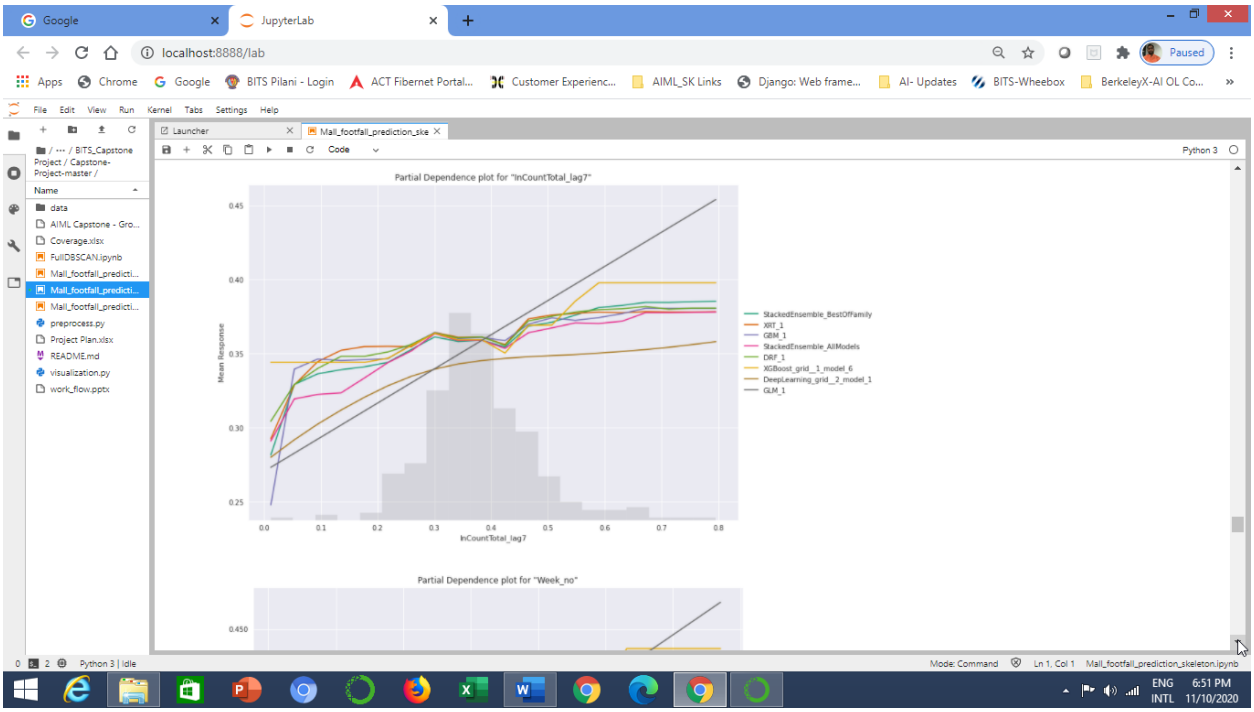

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Model Correlation
This plot shows the correlation between the predictions of the models. For classification, frequency of identical predictions is used. By default, models are ordered by their similarity (as computed by hierarchical clustering). Interpretable models, such as GAM, GLM, and RuleFit are highlighted using red colored text.
Model Correlation
Model Id
XGBoost_grid__1_model_1
XGBoost_grid__1_model_7
GBM_grid__1_model_3
XGBoost_3
GBM_5
XGBoost_grid__1_model_3
GBM_1
GBM_grid__1_model_5
XGBoost_grid__1_model_2
XGBoost_grid__1_model_4
GBM_grid__1_model_2
StackedEnsemble_AllModels
StackedEnsemble_BestOfFamily
GBM_4
XGBoost_grid__1_model_6
GBM_grid__1_model_4
GBM_2
GBM_3
DRF_1
XRT_1
1.0
0.9
0.8
0.7
0.6
0.5
Data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:48 PM
11/10/2020

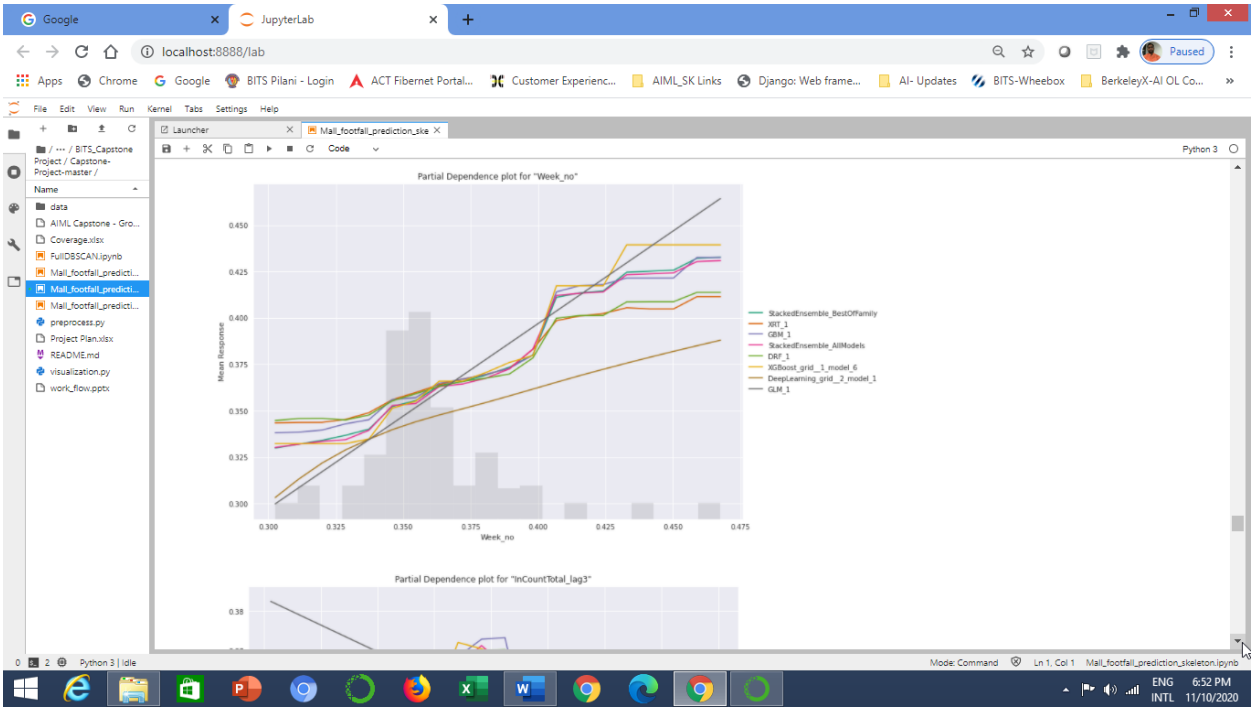

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Model Id
SHAP Summary
SHAP summary plot shows the contribution of the features for each instance (row of data). The sum of the feature contributions and the bias term is equal to the raw prediction of the model, i.e., prediction before applying inverse link function.
SHAP Summary plot for "XRT_1_AutoML_20201106_162054"
lnCountTotal_lag7
Weekday
lnCountTotal_lag1
Week_no
Month
lnCountTotal_lag5
lnCountTotal_lag6
mean_temp
1.0
0.8
Data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:48 PM
11/10/2020

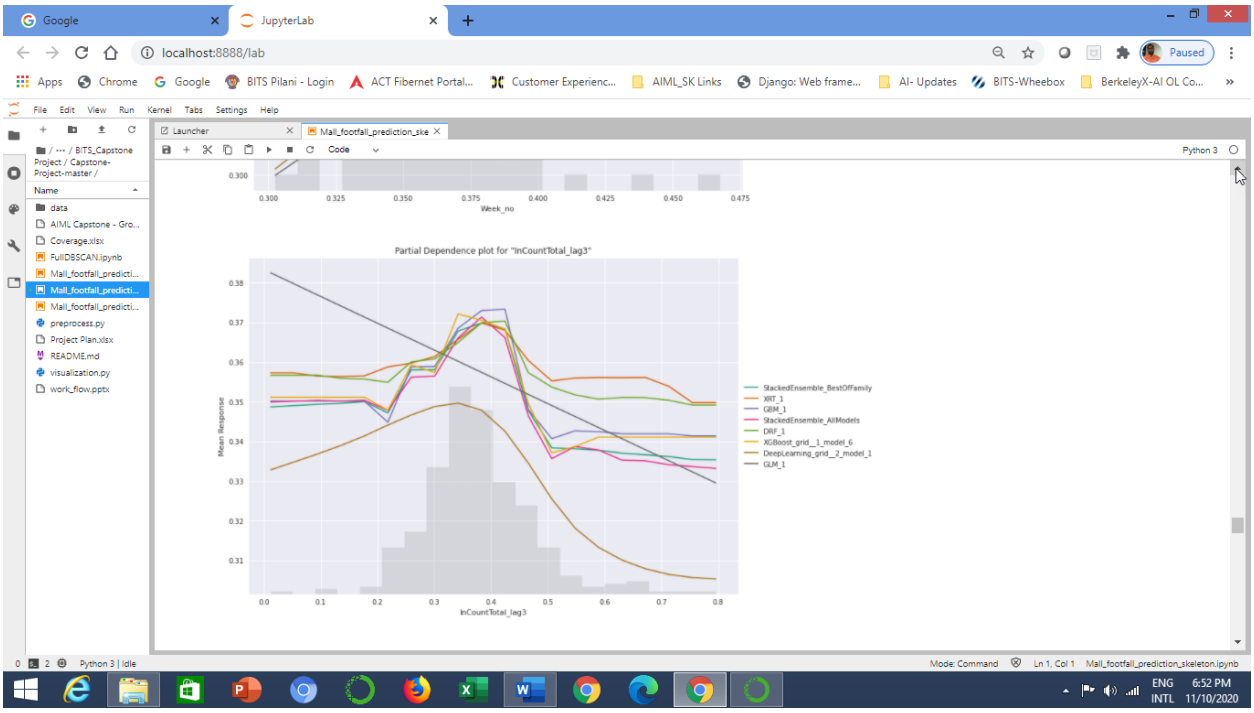

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
SHAP Summary plot for "XRT_1_AutoML_20201106_162054"
InCountTotal_lag7
Weekday
InCountTotal_lag1
Week_no
Month
InCountTotal_lag5
InCountTotal_lag6
mean_temp
Day
InCountTotal_lag3
InCountTotal_lag2
rain
InCountTotal_lag4
high_temp
wind_speed
University_holidays
School_holidays
UKBankHoliday
low_temp
high_wind
Feature
SHAP value
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:49 PM
11/10/2020

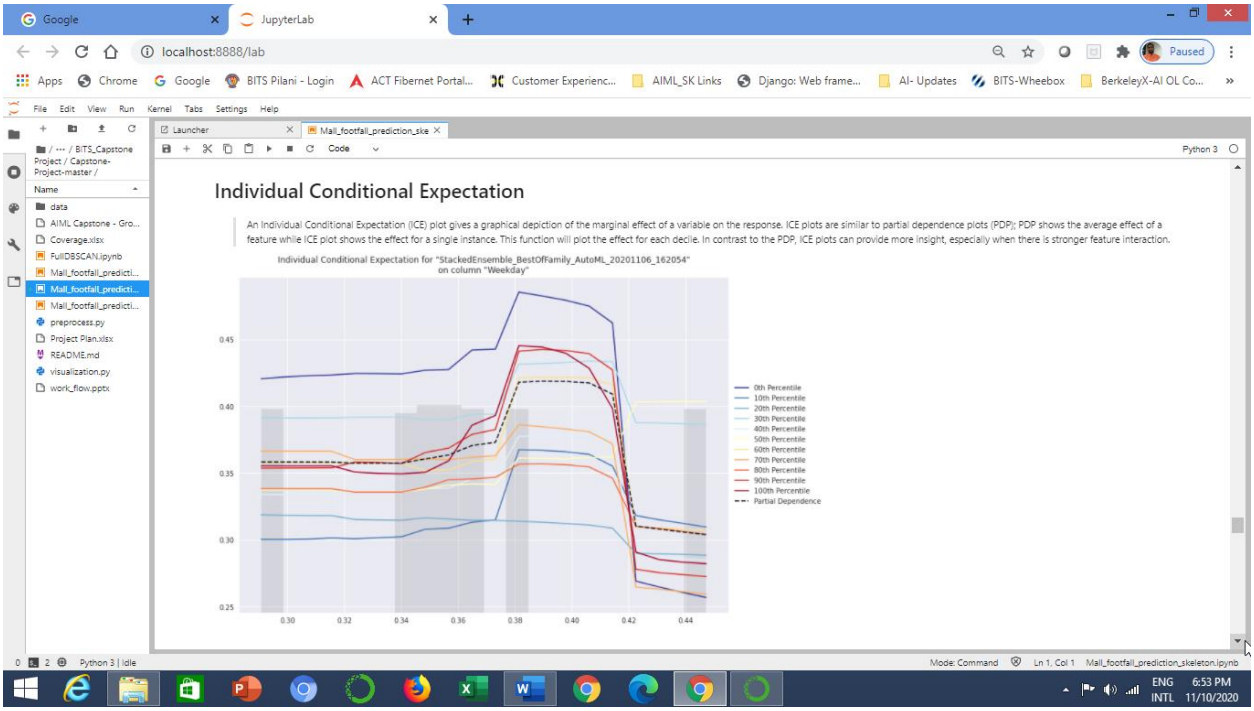

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Partial Dependence Plots
Partial dependence plot (PDP) gives a graphical depiction of the marginal effect of a variable on the response. The effect of a variable is measured in change in the mean response. PDP assumes independence between the feature for which is the PDP computed and the rest.
Partial Dependence plot for "Weekday"
Mean Response
Weekday
StackedEnsemble_BestOfFamily
XRT_1
GBM_1
StackedEnsemble_AllModels
DRF_1
XGBoost_grid__1_model_6
DeepLearning_grid__2_model_1
GLM_1
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:49 PM
11/10/2020

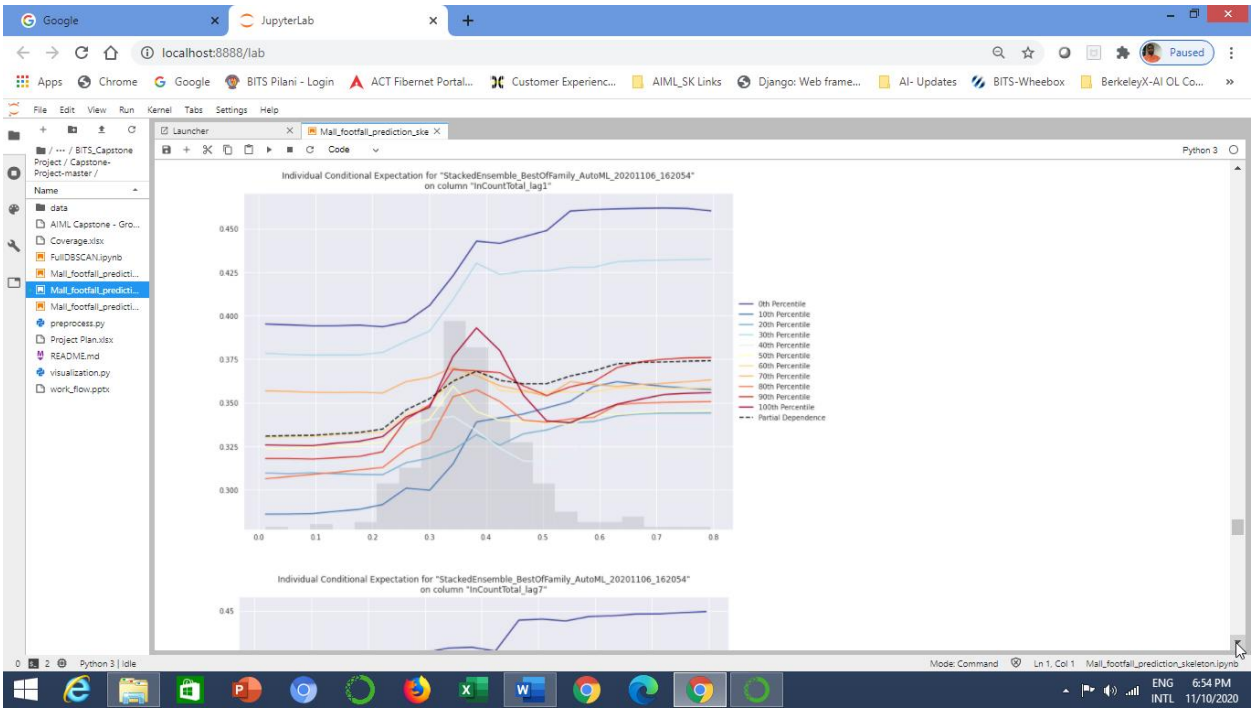

Partial Dependence plot for "lnCountTotal_lag1"
Mean Response
lnCountTotal_lag1
StackedEnsemble_BestOfFamily
XRT_1
GBM_1
StackedEnsemble_AllModels
DRF_1
XGBoost_grid__1_model_6
DeepLearning_grid__2_model_1
GLM_1

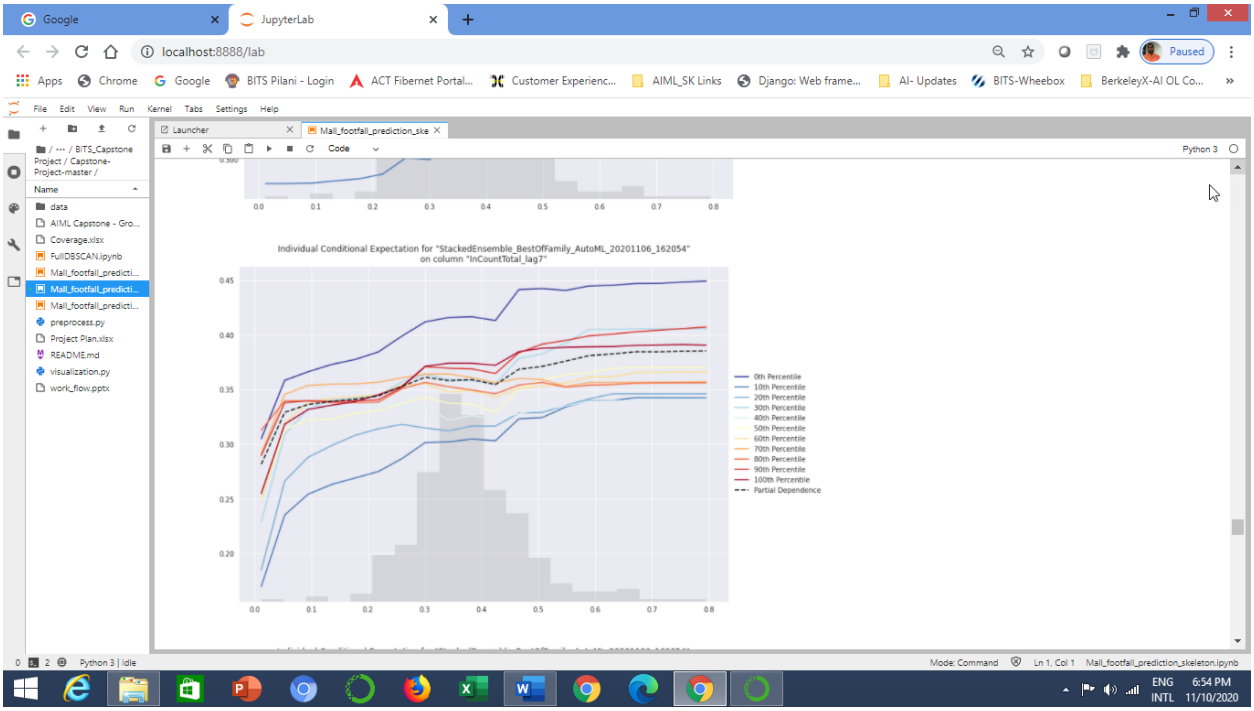

Partial Dependence plot for "lnCountTotal_lag7"
Mean Response
lnCountTotal_lag7
StackedEnsemble_BestOfFamily
XRT_1
GBM_1
StackedEnsemble_AllModels
DRF_1
XGBoost_grid__1_model_6
DeepLearning_grid__2_model_1
GLM_1
Partial Dependence plot for "Week_no"


Partial Dependence plot for "Week_no"
Mean Response
Week_no
Partial Dependence plot for "lnCountTotal_lag3"
StackedEnsemble_BestOfFamily
XRT_1
GBM_1
StackedEnsemble_AllModels
DRF_1
XGBoost_grid__1_model_6
DeepLearning_grid__2_model_1
GLM_1

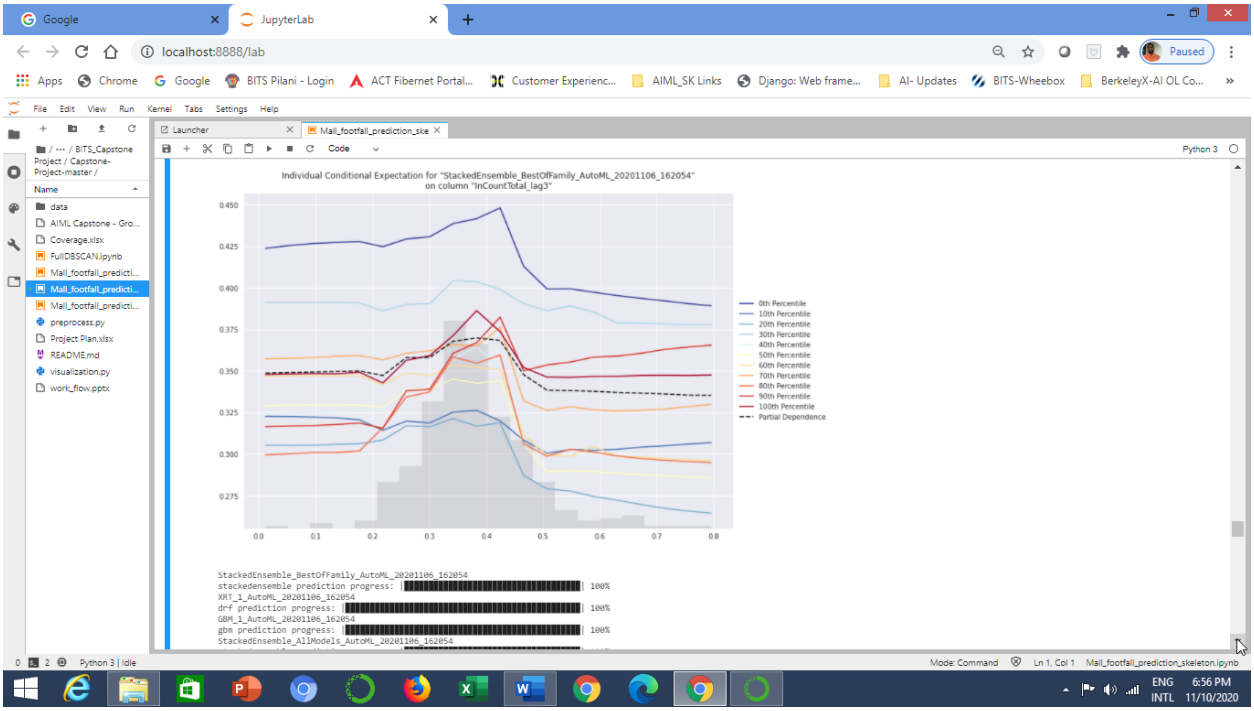

Partial Dependence plot for "lnCountTotal_lag3"
Mean Response
lnCountTotal_lag3
StackedEnsemble_BestOfFamily
XRT_1
GBM_1
StackedEnsemble_AllModels
DRF_1
XGBoost_grid__1_model_6
DeepLearning_grid__2_model_1
GLM_1

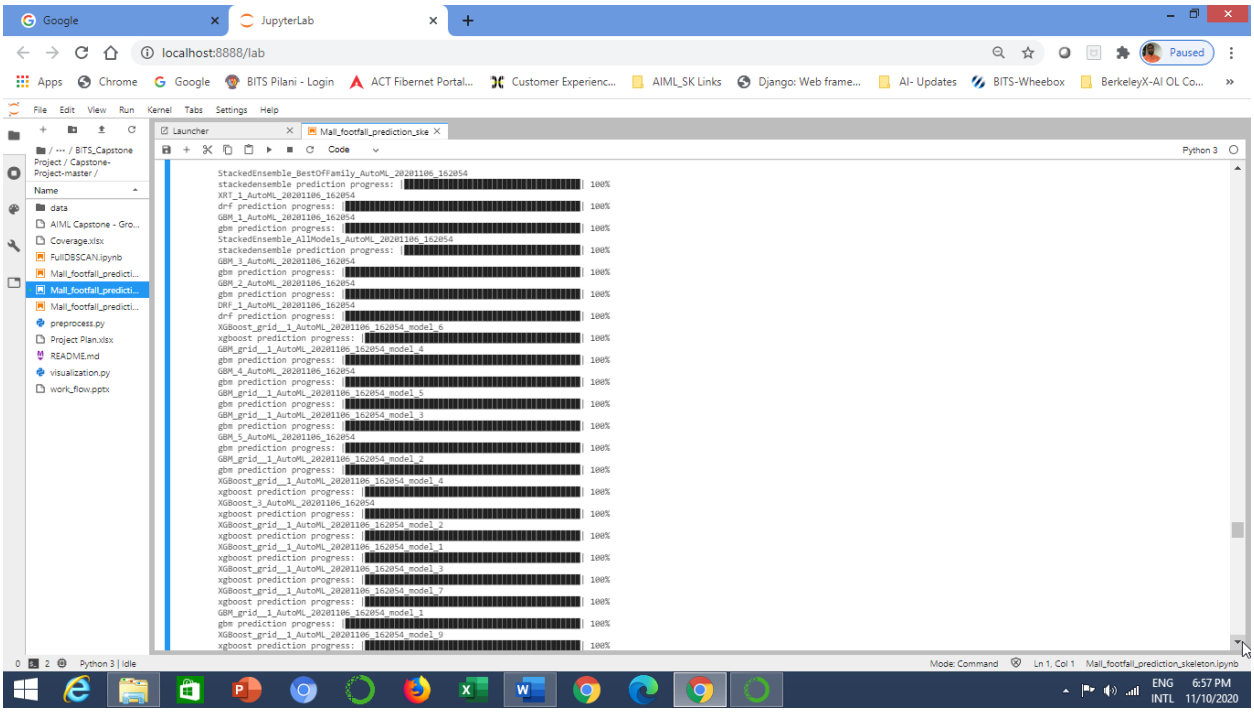

Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Individual Conditional Expectation
An Individual Conditional Expectation (ICE) plot gives a graphical depiction of the marginal effect of a variable on the response. ICE plots are similar to partial dependence plots (PDP); PDP shows the average effect of a feature while ICE plot shows the effect for a single instance. This function will plot the effect for each decile. In contrast to the PDP, ICE plots can provide more insight, especially when there is stronger feature interaction.
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054" on column "Weekday"
0th Percentile
10th Percentile
20th Percentile
30th Percentile
40th Percentile
50th Percentile
60th Percentile
70th Percentile
80th Percentile
90th Percentile
100th Percentile
Partial Dependence
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:53 PM 11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054" on column "InCountTotal_lag1"
0th Percentile
10th Percentile
20th Percentile
30th Percentile
40th Percentile
50th Percentile
60th Percentile
70th Percentile
80th Percentile
90th Percentile
100th Percentile
Partial Dependence
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054" on column "InCountTotal_lag7"
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG INTL 6:54 PM 11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
on column "lnCountTotal_lag7"
0th Percentile
10th Percentile
20th Percentile
30th Percentile
40th Percentile
50th Percentile
60th Percentile
70th Percentile
80th Percentile
90th Percentile
100th Percentile
Partial Dependence
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:54 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Name
data
AIML Capstone - Gro...
Coverage.xlsx
FullDBSCAN.ipynb
Mall_footfall_predicti...
Mall_footfall_predicti...
Mall_footfall_predicti...
preprocess.py
Project Plan.xlsx
README.md
visualization.py
work_flow.pptx
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
on column "Week_no"
0th Percentile
10th Percentile
20th Percentile
30th Percentile
40th Percentile
50th Percentile
60th Percentile
70th Percentile
80th Percentile
90th Percentile
100th Percentile
Partial Dependence
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
on column "lnCountTotal_lag3"
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:55 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
Individual Conditional Expectation for "StackedEnsemble_BestOfFamily_AutoML_20201106_162054"
on column "InCountTotal_lag3"
0.450
0.425
0.400
0.375
0.350
0.325
0.300
0.275
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0th Percentile
10th Percentile
20th Percentile
30th Percentile
40th Percentile
50th Percentile
60th Percentile
70th Percentile
80th Percentile
90th Percentile
100th Percentile
Partial Dependence
StackedEnsemble_BestOfFamily_AutoML_20201106_162054
stackedensemble prediction progress: 100%
XRT_1_AutoML_20201106_162054
drf prediction progress: 100%
GBM_1_AutoML_20201106_162054
gbm prediction progress: 100%
StackedEnsemble_AllModels_AutoML_20201106_162054
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:56 PM
11/10/2020


Google
JupyterLab
localhost:8888/lab
Paused
Apps
Chrome
Google
BITS Pilani - Login
ACT Fibernet Portal...
Customer Experienc...
AIML_SK Links
Django: Web frame...
AI- Updates
BITS-Wheebox
BerkeleyX-AI OL Co...
File Edit View Run Kernel Tabs Settings Help
Launcher
Mall_footfall_prediction_ske
Code
Python 3
StackedEnsemble_BestOfFamily_AutoML_20201106_162054
stackedensemble prediction progress: 100%
XRT_1_AutoML_20201106_162054
drf prediction progress: 100%
GBM_1_AutoML_20201106_162054
gbm prediction progress: 100%
StackedEnsemble_AllModels_AutoML_20201106_162054
stackedensemble prediction progress: 100%
GBM_3_AutoML_20201106_162054
gbm prediction progress: 100%
GBM_2_AutoML_20201106_162054
gbm prediction progress: 100%
DRF_1_AutoML_20201106_162054
drf prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_6
xgboost prediction progress: 100%
GBM_grid__1_AutoML_20201106_162054_model_4
gbm prediction progress: 100%
GBM_4_AutoML_20201106_162054
gbm prediction progress: 100%
GBM_grid__1_AutoML_20201106_162054_model_5
gbm prediction progress: 100%
GBM_grid__1_AutoML_20201106_162054_model_3
gbm prediction progress: 100%
GBM_5_AutoML_20201106_162054
gbm prediction progress: 100%
GBM_grid__1_AutoML_20201106_162054_model_2
gbm prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_4
xgboost prediction progress: 100%
XGBoost_3_AutoML_20201106_162054
xgboost prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_2
xgboost prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_1
xgboost prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_3
xgboost prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_7
xgboost prediction progress: 100%
GBM_grid__1_AutoML_20201106_162054_model_1
gbm prediction progress: 100%
XGBoost_grid__1_AutoML_20201106_162054_model_9
xgboost prediction progress: 100%
Python 3 | Idle
Mode: Command
Ln 1, Col 1
Mall_footfall_prediction_skeleton.ipynb
ENG
INTL
6:57 PM
11/10/2020