Time Series to observe DAILY temperature variations

Daily temperature prediction using Decision Tree

Following ARIMA, we use in the notebook below our first classifier:

Decision Tree (DT) : Similarly to ARIMA, We begin by loading all necessary libraries and paths to read the "pickles" as well as store image for the graph towards the end of our code. The pickles are read and the data is fed into an DT model.

Finally, we have two graphs showing the DT results vs. the fitted model as well as predicted results vs. actuals and test data

In [430]:
```
import warnings
warnings.filterwarnings('ignore')

%run helper_functions.py
%matplotlib inline
```

Create a folder for every run of the Decision tree to store our experiments

In [431]:
```
city='Houston' # New_York Atlanta Boston Dallas Houston Miami
analysis_type = 'Enhanced_Signals' # Basic, Inc_Signals, Enhanced_Signals
```

In [432]:
```
EXPERIMENT_DIR, EXPERIMENT_ID  = create_results_perrun()
print(f"Experiment ID: {EXPERIMENT_ID}")
print(f"Path of the results directory:{EXPERIMENT_DIR}")
```

```
Experiment ID: 18
Path of the results directory:../experiment_results/RUN-18
```

Here we are importing the train and test Data from pickle files created through the EDA file

In [433]:
```
X_train = pd.read_pickle(f'{PICKLE_PATH}/X_train_{city}_{analysis_type}.pkl')
Y_train = pd.read_pickle(f'{PICKLE_PATH}/Y_train_{city}_{analysis_type}.pkl')

X_test  = pd.read_pickle(f'{PICKLE_PATH}/X_test_{city}_{analysis_type}.pkl')
Y_test  = pd.read_pickle(f'{PICKLE_PATH}/Y_test_{city}_{analysis_type}.pkl')

print("Shape of Training Dataset " , X_train.shape)
print("Shape of Testing Dataset " , X_test.shape)
```

```
Shape of Training Dataset  (1432, 17)
Shape of Testing Dataset  (90, 17)
```

In [434]:
```
# Fitting a decision tree regressor with max depth
max_depth = 8
fitted_model = tree.DecisionTreeRegressor(max_depth=max_depth)
fitted_model.fit(X_train,Y_train)

# Dataframe to show features and their importances
top_features = len(fitted_model.feature_importances_)
features_importances_df= show_feature_importances(X_train.columns.values.tolist(),
                                                  fitted_model.feature_importances_,top_features)
features_importances_df.head(10)

# Store results
features_importances_df.to_csv(f'{EXPERIMENT_DIR}/feature_importances.csv')
```

In [435]:
```
# Run the model on the training dataset
Y_train_pred = fitted_model.predict(X_train)

# Calculate mean squared error for the predicted values
mse_train = mean_squared_error(Y_train, Y_train_pred)
print('Mean Squared Error for the training dataset: %.3f' % mse_train)
```

```
Mean Squared Error for the training dataset: 2.785
```

In [436]:
```
# Run the model on the testing dataset
Y_test_pred = fitted_model.predict(X_test)

# Calculate mean squared error for the test vs predicted values
mse_test = mean_squared_error(Y_test, Y_test_pred)
print('Mean Squared Error for the testing dataset: %.3f' % mse_test)
```

```
Mean Squared Error for the testing dataset: 10.205
```

In [437]:

```
# Creating a dataframe for predicted/fitted values
future_forecast = pd.DataFrame(Y_test_pred,index = Y_test.index,columns=['Fitted'])

# Concatenate the predicted/fitted values with actual values to display graphs
predictions = pd.concat([Y_test,future_forecast],axis=1)
predictions.columns = ["Actual","Fitted"]

# Displaying few of the predicted values
predictions.head(10)
```

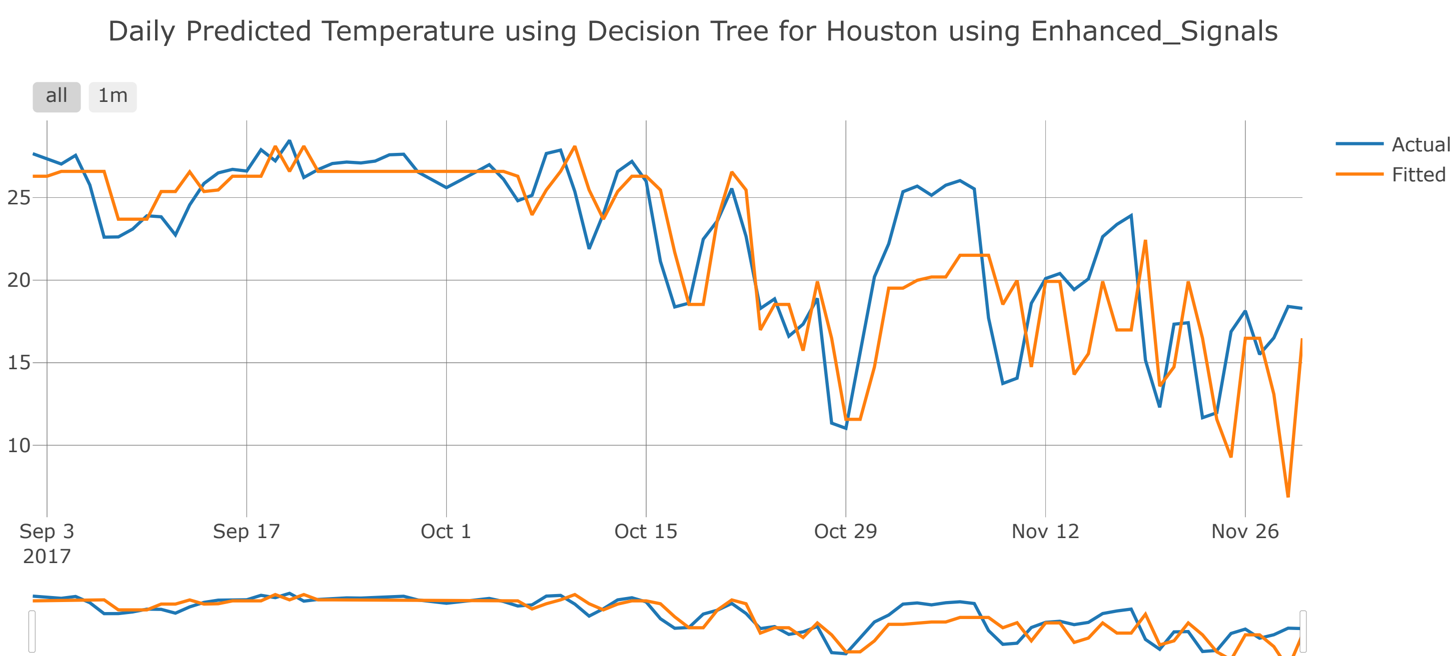
Out[437]:

datetime	Actual	Fitted
2017-09-02	27.658333	26.292523
2017-09-03	27.312917	26.292523
2017-09-04	27.032500	26.585889
2017-09-05	27.558333	26.585889
2017-09-06	25.762500	26.585889
2017-09-07	22.602917	26.585889
2017-09-08	22.617917	23.691244
2017-09-09	23.091667	23.691244
2017-09-10	23.897917	23.691244
2017-09-11	23.837500	25.365880

Mean Squared error (MAE), would be easier to interpret as they use the same scale as the data itself.

In [438]:
```python
city = city.replace('_',' ')
# Plotting the daily predicted temperature vs Actual Temperature - Decision Tree
fig = charter_helper_fitted(f"Daily Predicted Temperature using Decision Tree for {city} using {analysis_type}", predictions)
iplot(fig)

py.image.save_as(fig, f'{EXPERIMENT_DIR}/Daily_DT_actual_vs_predict.png')
```

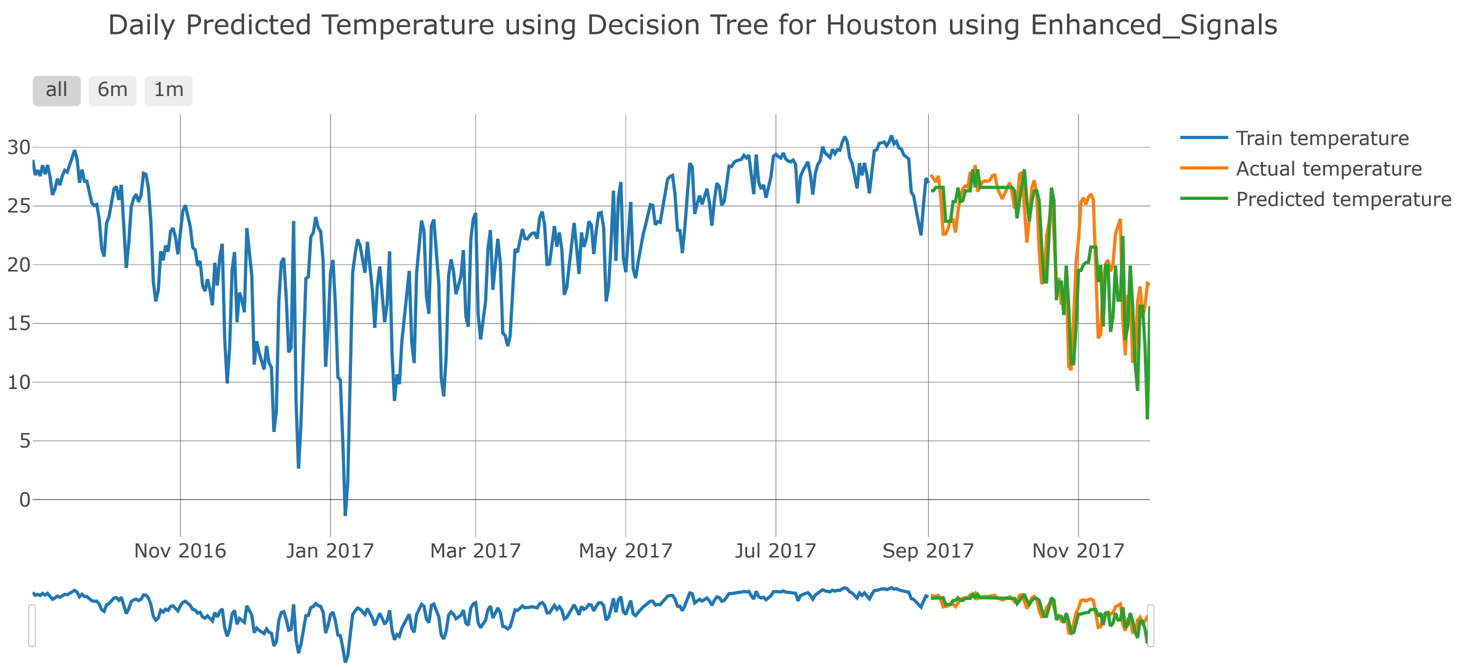


Export to plot.ly »

In [439]:
```python
# Plotting the training data for past year, Actual/test data and predicted temperature - Decision Tree
fig = charter_helper_prediction(f"Daily Predicted Temperature using Decision Tree for {city} using {analysis_type}",
                     X_train,Y_train,X_test,Y_test,future_forecast)

iplot(fig)

py.image.save_as(fig, f'{EXPERIMENT_DIR}/Daily_DT_predict.png')
```



Export to plot.ly »

In [440]:
```python
results = update_results_function(EXPERIMENT_ID, 'DECISION TREE', city, analysis_type,
                                  {'max_depth': max_depth, 'Info': X_train._metadata },
                                  {'features' : X_train.columns.values.tolist(),
                                   'importances':fitted_model.feature_importances_,
                                   'mse_train' : mse_train},
                                    mse_test)
```

In [441]:
```
results.tail(1)
```

Out[441]:

	RUN_ID	DATETIME	MODEL_NAME	CITY	FEATURE_TYPE	HOST_MACHINE	MODEL_PARAMETERS	MODEL_RESULTS	MEAN_SQUARED_ERROR
17	18	2018-08-14 19:48:11.736850	DECISION TREE	Houston	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	10.205057

```
In [442]: results = pd.read_pickle('../pickles/results.pkl')
results
```

Out[442]:

	RUN_ID	DATETIME	MODEL_NAME	CITY	FEATURE_TYPE	HOST_MACHINE	MODEL_PARAMETERS	MODEL_RESULTS	MEAN_SQUARED_ERROR
0	1	2018-08-14 19:41:45.275297	DECISION TREE	New York	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	5.799872
1	2	2018-08-14 19:42:13.654060	DECISION TREE	Atlanta	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	8.436149
2	3	2018-08-14 19:42:23.328525	DECISION TREE	Boston	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	16.504857
3	4	2018-08-14 19:42:30.715058	DECISION TREE	Dallas	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	12.523335
4	5	2018-08-14 19:42:41.506483	DECISION TREE	Houston	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	7.705077
5	6	2018-08-14 19:42:49.124681	DECISION TREE	Miami	Basic	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	1.833555
6	7	2018-08-14 19:43:04.972617	DECISION TREE	New York	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	6.518000
7	8	2018-08-14 19:43:12.930159	DECISION TREE	Atlanta	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	7.992946
8	9	2018-08-14 19:43:20.383683	DECISION TREE	Boston	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	22.967065
9	10	2018-08-14 19:43:32.825549	DECISION TREE	Dallas	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	14.223161
10	11	2018-08-14 19:43:39.374706	DECISION TREE	Houston	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	10.519872
11	12	2018-08-14 19:43:48.176029	DECISION TREE	Miami	Inc_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	1.577366
12	13	2018-08-14 19:44:00.243495	DECISION TREE	New York	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	6.767755
13	14	2018-08-14 19:44:07.609795	DECISION TREE	Atlanta	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	7.783687
14	15	2018-08-14 19:44:14.731752	DECISION TREE	Boston	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	25.109055

	RUN_ID	DATETIME	MODEL_NAME	CITY	FEATURE_TYPE	HOST_MACHINE	MODEL_PARAMETERS	MODEL_RESULTS	MEAN_SQUARED_ERROR
15	16	2018-08-14 19:44:20.754761	DECISION TREE	Dallas	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	15.432839
16	17	2018-08-14 19:44:46.805353	DECISION TREE	Miami	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	1.556905
17	18	2018-08-14 19:48:11.736850	DECISION TREE	Houston	Enhanced_Signals	DESKTOP-KN40C32	{'max_depth': 8, 'Info': {'feature_set_type': ...	{'features': ['temperature_lag1', 'temperature...	10.205057

In [ ]: