Bixi Weather

CEBD1260 Big Data Analytics

By

Donna Chan

Problem Definition

Explore and analyze if the following weather attributes, humidity, pressure, temperature, wind speed and wind direction, have an impact on a specific area in our lives

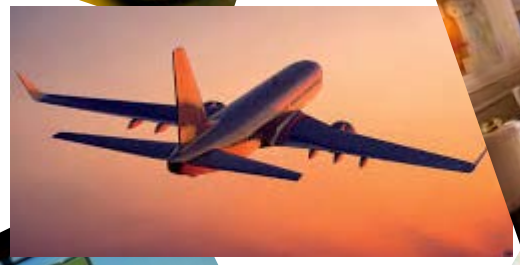
Problem Definition

- Economy
- Transportation
- Agriculture
- Human aspects
- Society
- etc













Weather Dataset Description

The weather dataset was obtained from Kaggle which originally acquired from OpenWeatherMap website.

Data on 36 cities and their weather attributes for 2012 to 2017, were contained in 7 csv files:

city_attributes, humidity, pressure, temperature, weather_description, wind_speed, wind_direction

Each weather attribute csv file contain the datetime and the value of attribute for each city.

Problem Definition - redefined

Explore and analyze if the following weather attributes,

humidity, pressure, temperature, wind speed and wind direction,

have an impact **on bixi ridership in the city Montreal, Quebec for 2017**

Bixi Dataset Description

The bixi ride dataset was also obtained from Kaggle

Data comprised of 2014 to 2017 bixi rides and its associated station locations.

Each csv of each year contains the start date, start station code, end date, end station code, duration of the ride, is_member.

Only 2017 dataset was used due to time limitation.

Main Use Case

Can we predict the number of bixi bike rides for a given day and hour based on weather condition?

Results

Six regression model were applied to the merge data.

The Random Forest Regressor models appear to have higher predictive accuracy than the other models. They show lower mean absolute errreors (mae) and root mean square errors (rmse)

	model	mae	rmse
3	RandomForestRegressor100	154.183227	260.638500
1	RandomForestRegressor	160.633867	271.663691
2	RandomForestRegressor10	166.716872	277.785557
5	DecisionTreeRegressor	214.565271	387.548419
4	KNeighborsRegressor	330.619212	488.276417
0	LinearRegression	521.976636	673.659273

Results

Haroku application was created to predict the number of bixi bikes rides for a given day and hour based on weather condition.

Here are a few results: