

London Weather Prediction with Apache Spark



COURSE:

CSYE 7200: BIG-DATA ENGINEERING USING SCALA

PROFESSOR:

ROBIN HILLYARD

TEAM 9

RENTENG HUANG, SHUANGSHUANG XU, BALAJI MUDALIYAR

Goals of the project

- ▶ To predict the future weather condition of London city(1 week).
- ▶ To develop Apache - Spark Scala code to clean, train, model the data.
- ▶ To use Apache - Spark Scala MLib(Machine Learning Library) to predict the weather.
- ▶ We will be implementing the UI using play framework.

Use cases

Model input

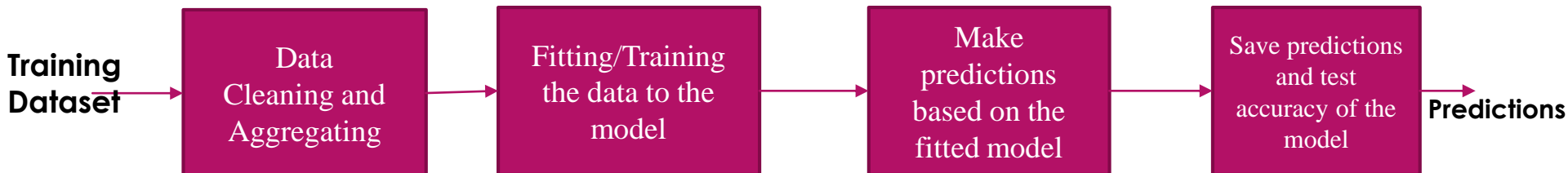
- Date, Temperature, Humidity, Dewpoint, Pressure, Wind
- Date

Model Output

- Weather prediction (Sunny, Cloudy, Rainy, Foggy, Clear)
- Temperature, Pressure

Methodology

- ▶ The cleaning of the training dataset
- ▶ Fitting of the data to the model
- ▶ Making predictions based on fitted model
- ▶ Calculating accuracy of the model
- ▶ Save predictions



Data source

- ▶ **Kaggle :**
(<https://www.kaggle.com/jeanmidev/smart-meters-in-london/data>)
- ▶ **21000 record dataset containing hourly weather information of London city**

PROJECT DETAILS

DATE	MILESTONE
15-Mar	Project Start Planning & Data work
23-Mar	Data finding & Cleaning(complete)
24-Mar	Spark Self-Learning, Mlib study
1-Apr	Coding
5-Apr	Implementation
9-Apr	Testing
13-Apr	Final Presentation & Documentation
15-Apr	Project End

Milestones/sprints

Programming in Scala and code repository

- ▶ **Most part of the project will be programmed in Scala including**

- ▶ Cleaning
- ▶ Splitting(Training& Testing)
- ▶ Fitting/Training Data to Model
- ▶ Predictions
- ▶ Accuracy Calculation

- ▶ **Code repository : GitHub**

<https://github.com/001239511ShuangShuangXu/csye7200-spring2018-group9>

Acceptance criteria

- ▶ **The accuracy of the model predicting weather will be more than 90 percent.**
- ▶ **Target a Root Mean Square Percentage Error (RMSPE) of 0.10**



THANK YOU

FOR LISTENING

End