Learning to Predict Flight Delay

DataSets

We have the following sources of datasets to be used in our project

1. **Statistical Computing & Statistical Graphics** - http://stat-computing.org/dataexpo/2009/the-data.html. This data is taken from the Research and Technology Administration (RITA) database and structured for our use. But the above dataset doesn't have any information related to weather conditions at the origin and destination airports.

2. For weather data we have:
   Hourly land-based weather observations from NOAA - http://cdo.ncdc.noaa.gov/qclcd_ascii/ . This source contains hourly and daily data of weather at various airports

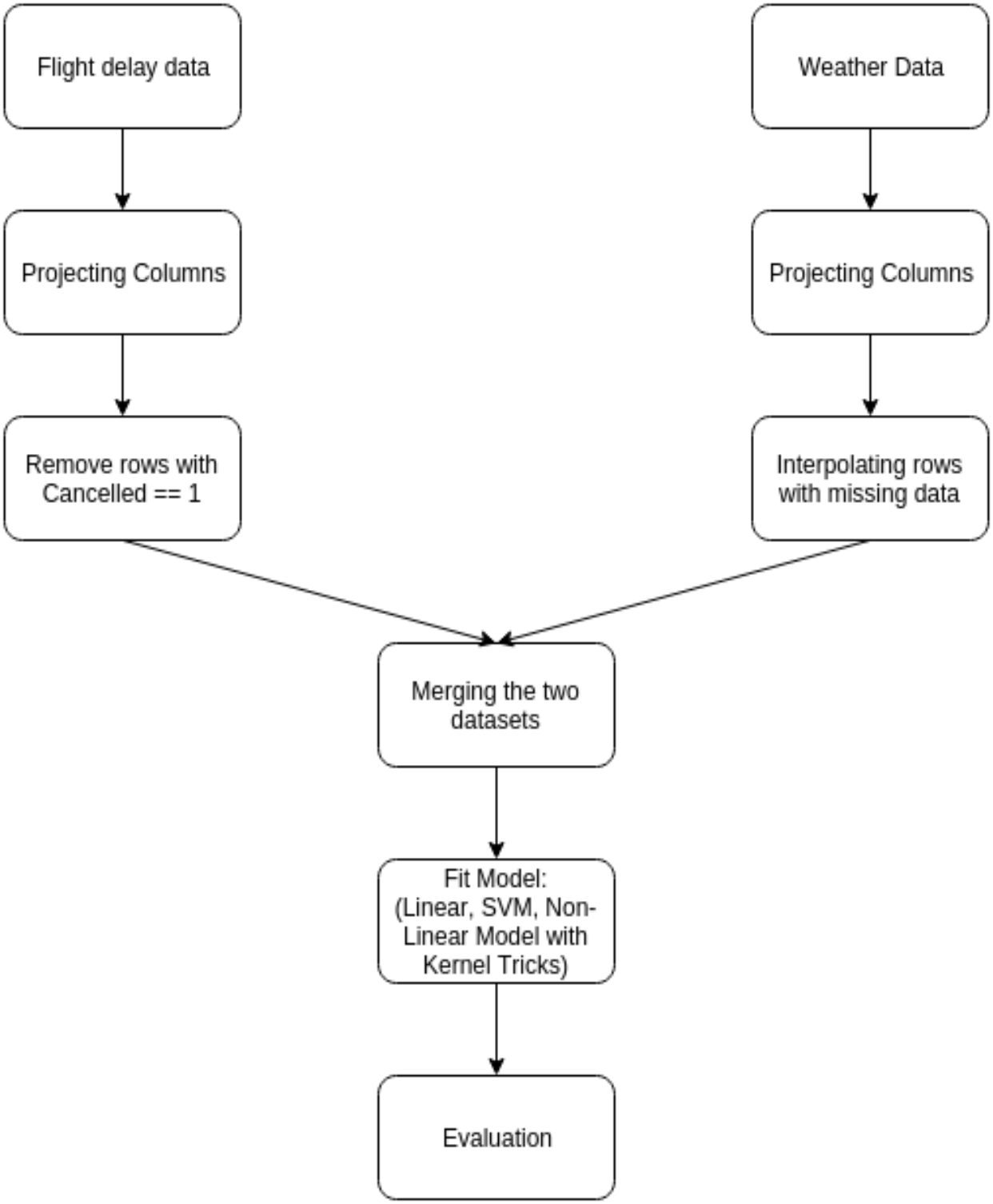
Data Preprocessing

The columns which are relevant in the RITA dataset are as follows:
DayOfWeek, UniqueCarrier, Origin, Dest, Distance, ArrDelay, DepDelay

From the weather data the following columns will be extracted:
Sky Conditions, Visibility, Weather Type, Dry Bulb Temp, Dew Point Temp, Wet Bulb Temp, % Relative Humidity, Wind Speed (kt), Wind Direction, Wind Char. Gusts (kt), Val for Wind Char., Station Pressure, Pressure Tendency, Sea Level Pressure, Record Type, Precip. Total

The final step of data preprocessing will require the concatenation of both the datasets by mapping airport id with weather station id.

Flow-chart


Flight delay data
Weather Data
Projecting Columns
Projecting Columns
Remove rows with Cancelled == 1
Interpolating rows with missing data
Merging the two datasets
Fit Model: (Linear, SVM, Non-Linear Model with Kernel Tricks)
Evaluation

Algorithms

Currently we haven't merged the weather data with original flight dataset. We have tried to fit linear model to the available data after basic data preprocessing.
The models which we plan to use will be linear regression , ridge regression , svr and non linear model with kernel tricks. We will use Root mean square method to calculate the error between the predicted data and the test data.

Team Members:

1. Prateek Chandan - 120050042
2. Maninder Singh Saluja - 120040029
3. Nishant Kumar Singh - 120050043
4. Shubham Jangir - 12D070030

Mentors - Shubham and Dheeraj

References

1. http://en.wikipedia.org/wiki/Flight_delay
2. http://www.transtats.bts.gov/homedrillchart.asp
3. http://stat-computing.org/dataexpo/2009/the-data.html
4. https://github.com/caesar0301/awesome-public-datasets#transportation