**Assignment-10**

**Results (Draft Manuscript)**

**Sravan Kumar Borra**

We had the correlation results based on the experiments done in class using zeppelin. I performed precision correlation method to find the correlation between the variables. It shows that Temperature and density highest percentage of correlation. Our aim in this project is to provide an efficient model which helps to build a Smart city. Smart city mostly runs on connectivity. This connectivity is established by using big data. We are using big data techniques, machine learning techniques, data mining techniques to build a model that will be used for the prediction of weather which helps in building a smart city. There are various machine learning techniques that are used for predicting the weather. We can use Support vector machines method; Decision trees techniques, neural networks.

A data set with weather observations has been selected for this project. Data set observations are collected from the city of Aarhus in Denmark. This data set has been fetched form City pulse website, which is trying to provide models to build smart city. Data set has different types of fields such as Dew point, humidity, pressure, temperature, wind direction, wind speed. All these parameters are used for predicting weather conditions in different times of the year. We have two sets of data. One is from February-June 2014 and other set is from August-September 2014. These data sets were provided in JSON file format. We use Zeppelin platform to study the data. We initially load the data and study about the different fields in the data.

Our aim in this project is to provide an efficient model which helps to build a Smart city. Smart city mostly runs on connectivity. This connectivity is established by using big data. We are using big data techniques, machine learning techniques, data mining techniques to build a model that will be used for the prediction of weather which helps in building a smart city. There are various machine learning techniques that are used for predicting the weather. We can use Support vector machines method; Decision trees techniques, neural networks. All these methods are used in machine learning for predicting using the given data set. For this particular paper, we selected weather data of Aarhus city located in Denmark to predict the climatic changes. This prediction model helps in building a Smart city which runs on connectivity.

There are several ideas which are useful to predict the weather conditions, climatic changes, forecasting the weather which can be used for daily predictions, half yearly predictions. We can predict the temperatures changes such as maximum temperature of a particular day and minimum temperature of a particular day. We can also predict the weather condition for a particular day. We can predict whether the day is going to be sunny or it is going to rain. We can find out or predict the wind speed for a specific day so that people can take safety measures if there are high speed winds which can even destroy the buildings.

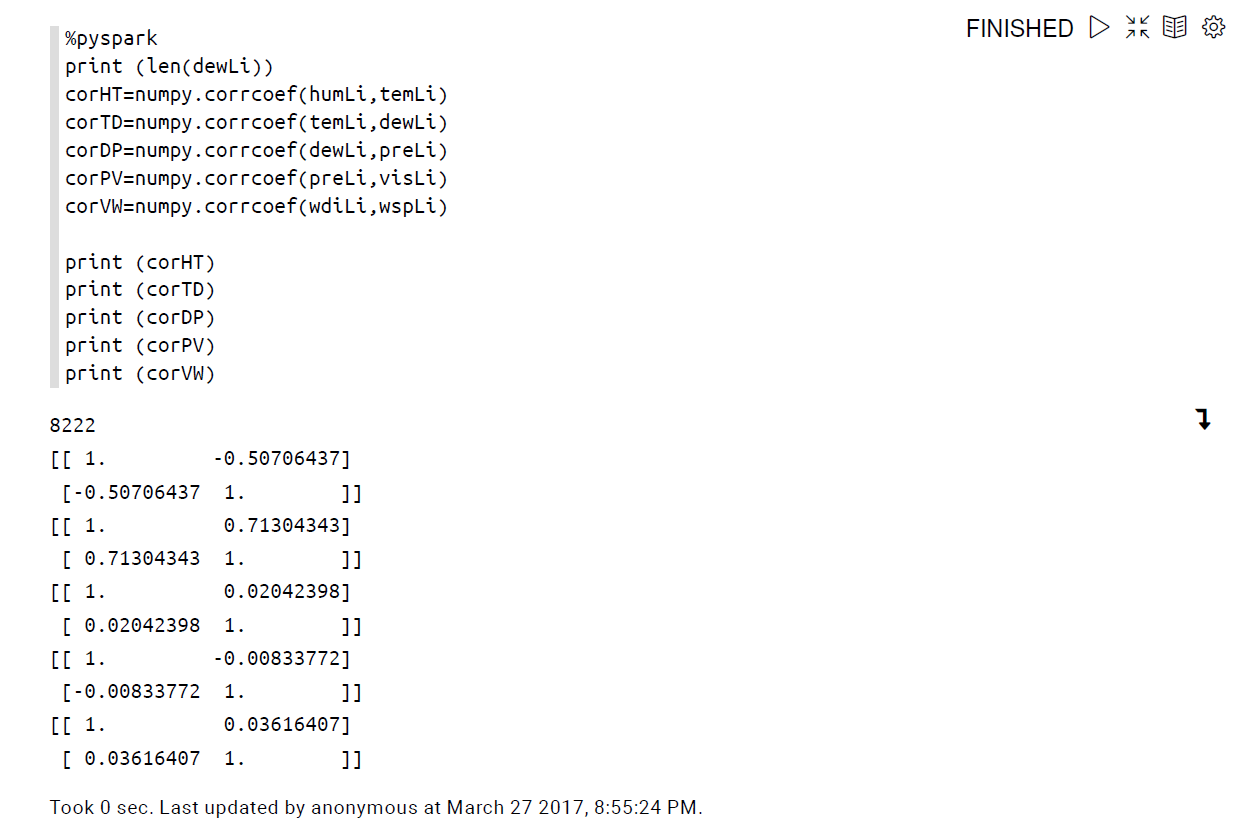
We can predict the climate changes seasonally. We can predict whether the summer season is going to be with high temperatures. We predict winter season whether it is going to have heavy snow fall or low temperatures so that people can be alerted if there is any problem. Natural calamities can be predicted if we can build a model using our machine learning technique methods. Now days it has become very important for detecting the blizzards and storms so that people can take shelters.

We predicted the maximum and minimum temperature of a particular using linear regression model. Therefore the main idea behind this project is to predict the temperatures to their maxima and minima for a particular day. We are going to predict wind speed and humidity for a particular day. Predicting climatic changes seasonally based on the data that we already have. We can also predict some of the areas in a city or a place where the solar energy production is suitable. We can also find the places which are helpful for producing wind energy. These predictions are done by using wind speed data. These are natural energies which are helpful for building a smart city as the city is going to be pollution free city. We can also predict the amount of rainfall that is going to come for a particular season which helps the farmers.

All the data sets provided are in JSON format. So I cleaned the data and saved to CSV file. There are sevendifferent data sets in weather data. So cleaning the data for each data set will consume time and makes the process slow. So write a loop for loading the data set where the code is optimized and took less than half of the time for loading the data separately. This makes my code optimized and efficient.

While plotting time series, the values presented in the data are not sorted. So the results are not accurate and it will not give sorted values. So I separately sorted the data set in CSV file separately and converted the strings into Date Time format. All this data manipulation will be useful for time series plotting and performing regression, SVM in the future.

We had the correlation results based on the experiments done in class using zeppelin. I performed precision correlation method to find the correlation between the variables. It shows that Temperature and density highest percentage of correlation.



These are the results obtained after performing the correlation for each time series variable. First set of results gives us the information about the correlation between humidity and temperature. It shows that we have fifty percent of correlation between them. After that we have results about the correlation between temperature and dew-point. It shows that temperature and dew point has more than seventy one percent of correlation. It is the highest percent of correlation happening in the weather data set. We also performed correlation analysis for dew point and pressure, wind direction and wind speed. All these operations took less than a second to run in zeppelin.