**Draft Manuscript**

**Sravan Kumar Borra**

**Discussion:**

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. 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 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.

In competitor article they performed ARIMA and ANFIS statistical methods. We are using SVM and ANN. We out performed them as these statistical models gives better results. In the competitor article, they predicted only Maximum and minimum temperature. We are predicting Max and Min temperature, wind speed on a particular day, dew point at a particular time. The tool used in competitor article is SPSS software. We used Zeppelin in our project. We again outperformed in the selection of tools as Zeppelin if far more efficient than SPSS. For excluding null values their code is lengthy and not accurate. It consumes more time. In our article, it took almost less than 10 lines of code to exclude null values as we are using spark. Pearson Correlation for temperature and dew point- more than 70%. Our data has less outlier as we excluded the null values before performing the statistical analysis. The regression value of temperature and humidity grouped by pressure is -7.500000 and the Intercept is 123.500. We still do not have the R^2 value for our data as we did not performed SVM on our data. The work is still going on. In the competitor article, people have not performed correlation on their data. So they don’t know the relation between their data variables. We performed correlation on our data. We have a 70% relation between temperature and dew point. With this result we can predict the temperature based on dew point variable.

**Conclusion:**

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.

Forecasting accurate weather conditions can also help in transportation. For flights and airplanes which are flying mostly in air perfect weather conditions are must and important. If we forecast the weather in advance it helps the aviation department to plan according and it also helps in damage control.

We can also build a model which is useful after building a smart city. We can predict the climatic changes, prediction of blizzards and storms. Due to that prediction one can take safety measures which helps in decreasing the losses which can be caused by a storm or blizzard. Now days due to pollution, the climate is changing without any warning. In those conditions these types of predictive models will help for damage control.