To Analyze Fog weather conditions of Delhi and Forecasting rain in Australia

Harsh Manoj Chudasama(x18187340)
Masters in Data Analytics(Group B)
School Of Computing
National College Of Ireland
Dublin, Ireland
x18187340@student.ncirl.ie

***Abstract*—Weather forecasting has become a new trend in today's generation. Technology has advanced so much that it can predict whether it will Rain Tomorrow or not. My key objective for this research topic is to predict whether it will rain tomorrow in Australia or not and also predict whether there be fog in Delhi or not. Air quality forecasting is a hot topic for many scientists owing to its impact on health. Scientists from all over the world are trying to create a model that can predict the quality of the air (i.e. Fog). Similar is the case of Australia where it rains throughout the year. In the case of, air pollution in Delhi, inadequate research had been done in the past. Fog is not harmful in nature unless it's mixed with polluted gases i.e. Smog which results in lung cancer and respiratory problems. Hence, I would like to explore further by analyzing factors like Temperature, Humidity, etc that affect the Fog mixed with air pollutants and its density.**

I. Introduction

Air pollution has become a growing problem in the world in recent years. As per the report published by the World Health Organization (WHO), the majority (80%) of the population is suffering from air pollution especially people who are residing in the urban area where the air is more toxic and exceeding the standards set by WHO[1]. During 2018, Indian cities including New Delhi, Varanasi or Patna, which identified Particulate Matter(PM) less than 2.5 micrograms in each cubic meter of the surrounding air, that was detected in samples collected from 4300 cities worldwide and concluded them to be most polluted cities in the world.

In several developing economies, notably, India, air pollution has become a major issue and may have negative impacts on people. From the Past 10 Years, Delhi, the capital of India, has been facing this issue. Hence the key focus is to determine the root cause of harmful fog that is affected the person's health and to find the parameters that increase the fog level in Delhi.

Australia is one of the countries that have multiple weather zone depending on the area.[1] The northerly segment has extra tropical influenced weather, warm and humid inside the summer, and quite heat and dry within the winter, even as the southern components are cooler with moderate summers and cool, sometimes rainy winters. Although seasons are like North and South pole which are completely different from each other.[1] Rainfall is very rare in Australia that being said it's a fact it Sydney even in monsoon season it only rains 13 days per year. Hence, I had the eagerness t find out the prediction of rain in Australia as it will be a challenge to study weather patterns and then predict based on factors like humidity, temperature, etc.

This research paper is charted in the following ways. First describes the Introduction of the topic which includes the motivation for choosing the topic, its current scenario and research question for the topic. Section 2 outlines Related Work which is a brief on literature survey which highlights Key results achieved by researchers and evaluation positive and negative aspects of the research. Limitations of the models and datasets used for research are described in detail. Section 3 describes in detail about the Data mining methodology used for project implementation. Section 4 illustrates the model used for implementation with its interpretation and Section 5 concludes the research paper with future work and a list of references.

II. Related Work

Delhi weather dataset :

Authors Kumar and Ankindar compared 5 data mining models on the Delhi weather dataset based in their recent paper[2]. They discovered that when the training dataset is comparatively small i.e. around 10,000 records, K Nearest Neighbor (KNN) performs the best followed by Support Vector Machine algorithm. The least error rate was demonstrated by the Random Forest algorithm. From this research paper, we will be interested in KNN and Random Forest models to compare the performances.

In recent research stated author that the Support vector machine(SVM) is the best model when it comes to supervised learning models and outperforms in predicting accurate results when compared with other models[3]. Although there is a drawback of SVM that it sometimes overfit

[1] https://www.weatheronline.co.uk/reports/climate/Australia.htm

the models and identifies incorrect points that lie on the other side of the hyperplane.SVM is time-consuming as it models the data around the hyperplane which is a three-dimensional linear function of x and y axes. From this research paper, we will try to tune the SVM model and rework on the time factor of model implementation.[4]

In agreement with the above-mentioned investigation authors of American Meteorological Society publication initiated similar results asserting random forest achieve better results than SVM and KNN[5]. Their research also found out that as temperature decreases the dew precipitation increases which results in a fall of humidity of the surrounding environment[4], [5]. Also, Fog is directly proportional to the dew factor.

In one of the papers, the author followed a different approach where he subsetted the dataset as he found out that not all variables contribute to the end result. For this task of elimination of variables, he found out the correlation between variables and selected the best ones[6]. Then the author performed Principal Component Analysis for further reduction of variables and in the end, he was able to predict the accurate result with fewer variables and thus increasing the efficiency as well as the accuracy of the model[6]. Here, he discovered the decision tree outnumbered other models in terms of accuracy and specificity.

Rain in Australia dataset:

One of the authors discovered that Temperature is a key factor in predicting Rain[3], [7]. Here author selected the predictor variables Temperature, Humidity, Viscosity and imputed them with the missing values using mice library of random forest. After preprocessing, the model was trained using a logistic regression model[8]. Post prediction it was found out that Temperature plays a key role in forecasting rain. The model was highly efficient with an accuracy of 98% and a kappa value of 0.9 with the highest sensitivity of 0.977. This will provide a base for my research[9].

The author discovered that Step Index Regression was most effective with a dimension of predictors being reduced when backward regression was implemented[10]. In his research, the author stated that applying step regression not only saves time but predicts best model accuracy without human intervention of the elimination of variables based on the p-value[11]. Some authors have stated that the Neural network performs best when then there is a more complex relationship between predictors although the Neural network is a time-consuming process but guarantees accurate predictions of 98% with ROC values of 0.96 which is so far best predicted by any researcher in terms of neural network implementation[12].

Researcher Wiley performed a unique experiment where he combined Linear Discriminant Analysis and KNN and applied it to Random forest as input[13]. The outcome was surprising not only it improved its accuracy but increased the Negative Prediction rate and thus he was able to get the corner cases and get rid of them.his Model predicted a balanced accuracy of 0.89 with No information rate 0.85. This insight could help us to understand how multiple model combination works in the machine learning algorithm[13].

One such researcher stated that "Neural Network doesn't perform well when data is highly imbalanced[14]. The author proved it by passing biased data and it took so long to build the model and the outcome was very poor with an accuracy of 30% which is a sign of bad model that one wants to consider while predicting[15]. The author also stated that imagine if this model was used in predicting life and death situation which is very critical where people life is on the line. He gave an example of predicting the time it takes to open airbags in case of a car crash experiment. So model selection plays a key role which we will take into consideration while implementing[16].

Based on the research proposed by Springer Nature publication, key information is always looking at how well the model predicts but to understand its underlying principle[17]. It may lack performance but conceptually it satisfies the objective[18]. For instance, a decision tree in some cases performs worst but when it comes to evaluating and making branches and pruning the tree it's the best model[19].

So as suggested its better mix match the models and try to find optimal solutions[20]. As said by Donald "There are thousands of ways to solve a problem but we need the perception to look at it from the right point of view "[21].

III. Data Mining Methodology

For analyzing the datasets, I have surveyed Knowledge Discovery in Databases i.e. KDD Methodology. The key objective of KDD is to extract knowledge from the dataset and provide key insights. It's a process-oriented approach (refer fig 1) and each step will be explained in further analysis with respect to dataset selected. Two relatively large datasets of more than 10,000 rows with each consisting of 24 columns in Rain in Australia dataset and 20 columns in Delhi Weather dataset.

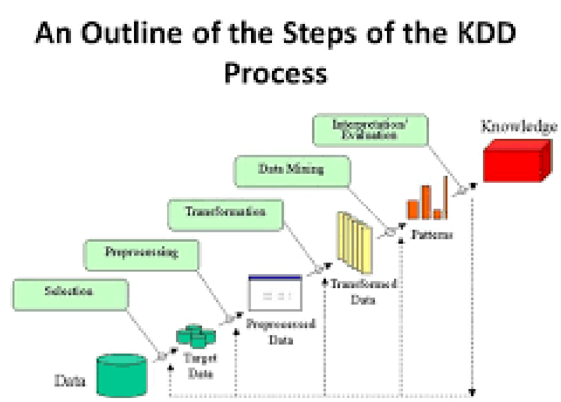

Fig. 1. KDD Methodology[2]

The second steps involve cleaning the dataset i.e. pre-processing data. This is a very important step as data needs to be normalized or cleaned without consisting of any missing values or outliers. These imbalanced values can affect the results of predictions. For both the dataset, Missing values were imputed with median values and some values were predicted

[2] Image Source :https://images.app.goo.gl/iKUekDKMYa4yzftr6

using mice library of random forest. The distribution of data was plotted after imputing values to check its structure. (Refer fig 2 for Missing Values interpretation)

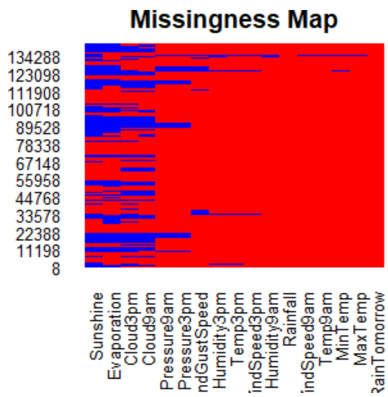

Fig. 2. Count of Missing values in Dataset

The third step involves the transformation of the datasets. Here categories were recorded in the form of Yes and No to numerical values to (1,0) in order to smear the machine learning algorithm. Variables were converted to factors so that they can be used on classification models.

Post transformation, data was analyzed considering the numerical and categorical values a report was generated. In order to understand the data report checked the spreads of data based on parameters like mean, mean, mode,1st Quartile, standard deviations, etc. It was made sure that were no missing values at this stage.

For the dataset, rain in Australia, the objective was to predict whether it will rain in Australia tomorrow or not and the objective of Delhi Weather data was to determine whether the climate will be Foggy or not based on various factors like Wind, Humidity, Pressure, Precipitation, etc.. Based on the objective stated i.e. the response variable is categorical (Yes/No) in nature based on that data mining techniques were used.

In rain in Australia dataset, it consisted of 24 attributes like Min Temperature, Max Temperature, Humidity, Rain Tomorrow, etc and 142,193 records. Dataset was taken from the Kaggle website - https://www.kaggle.com/jsphyg/weather-datasetrattle-package.

Following columns RISKMM, WindGustDir, WindDir9am, WindDir3pm, RainToday, Date, Location were dropped from the dataset because more than 90% of data was missing. So overall 17 variables were considered for analysis out of 24. Attribute RainTomorrow is categorical is nature which is encoded in numerical while transforming (Yes means 1 and No means 0). Other attributes like Max Temp, Min Temp, etc are continuous in nature and respective attributes were imputed with missing values. The good part about Dataset was it self-explanatory in nature and hence it was easy to interpret values.

The second dataset, Delhi Weather data, involved 20 attributes like Temperature, Pressure, Humidity, Precipitation, Fog, etc and a total of 100,990 rows. Dataset was taken from the Kaggle website - https://www.kaggle.com/mahirkukreja/delhi-weather-data.

Following columns X_heatindexm, X_precipm, X_wgustm, X_windchillm were dropped from the dataset reason being more than 90% values were NA. So overall 15 variables out of 20 were considered for further analysis. Attribute "Fog" is categorical is nature which is encoded in numerical while transforming (0 as Yes and 1 as No). Other attributes like Humidity, Pressure, Dew, etc are continuous in nature and respective attributes were imputed with missing values. The weakest link about the dataset was that no data dictionary was provided and hence values were assumed.

In order to apply a machine learning algorithm, random samples of data were divided in the ratio of 75% training set 25% test set. The model was trained on 75% data and Evaluated the performance of the model on the remaining 25% test data.

IV. Evaluation

As per guidelines by the mentor, a minimum of 5 machine learning algorithms needed to be implemented in total. The following algorithms were used for analysis K- nearest neighbor, Logistic Regression, Random Forest, Support Vector Machine(SVM), Decision trees, Linear Discriminant Analysis, and Multiple Linear Regression. All the methods take predictor value as a binary categorical value.

1. Logistic Regression :

This method is known as the base class for all classification models. It's the easiest model to implement. Although there are chances that predicted values will overfit the model. It requires large datasets to predict accurate results. This model is said to predict a higher number of negative test cases i.e. High Specificity Value. If we plot the model it will be S curve-shaped. It takes Binary value as the dependent variable and independent variables can be continuous or binary in nature. The below image is the equation used for logistic regression where Beta 0 is the Y-intercept when X is not true and Beta 1 is a slope that intercepts the logistic line and e is the exponential function.

$$\frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$$

Fig. 3. Logistic Regression Equation[3]

2. Random Forest: It's derived from decision trees but the only difference is that it classifies the trees based on entropy value. It's a supervised learning model that makes rules at every branch and further branches as yes or no and so on. Instead of creating a single regression model it creates multiple linear models and avoids the problem of overfitting of data. The plan was to build more trees in order to reduce the interaction between the plants. Here several branches are made with different outputs at the end, in order to compute the final result and average is taken of all the generated output as final

[3] Image Source : https://images.app.goo.gl/bzorFT282Ssj3F429

combined output. The below diagrams depicts the working of decision trees.

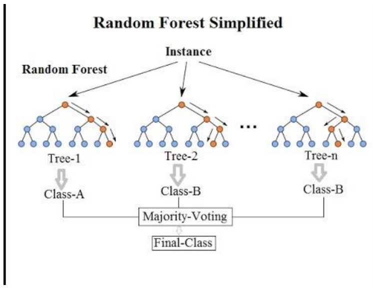

Fig. 4. working of Decision Trees[4]

3. K – Nearest Neighbour (KNN): KNN is an unsupervised learning algorithm that is best used when you don't have any idea about data sets. Its the easiest and simplest algorithm to implement. Basically, it forms clusters of a group based on Euclidean distances on the training dataset and applies the learning on test dataset to predict the proper category.

The following mentioned methods were applied to the two datasets. As all models belong to the classification category, the confusion matrix along with the ROC curve was used for evaluating the models. The confusion matrix is a table that is divided into 4 quadrants used to tabulate Predicted vs actual values(refer Fig 5) The terminologies used in ***confusion***

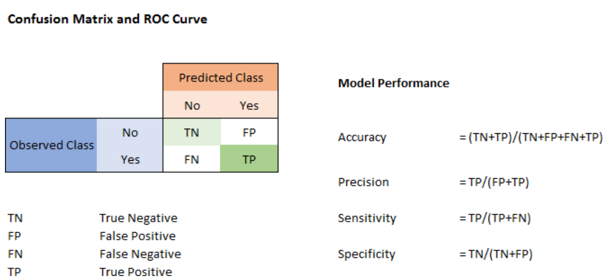

Fig. 5. Confusion Matrix[5]

matrix:

- True Postive(TP) is the Actual values that are correctly predicted.
- False Positive(FP) usually referred to as Type-1 error.
- TP is the Actual values that are incorrectly predicted.
- False Negative(FN) they are usually referred to as Type-2 error.
- FN are actual values that are incorrect but predicted as True.
- lastly True Negative(TN) as it goes by the name they are negative values that are correctly predicted.
- It is a normal convention that more values should be Predicted along the diagonal as TP and TN and fewer values on the other diagonals(FP and FN).

Other factors of confusion matrix include Accuracy i.e. chances of being correctly classified.No Information rate that is probably of choosing the classifier without prior knowledge. Ideally, it should be less than accuracy. Cohen kappa is a derived version of accuracy only difference is that it takes the likelihood of an event in the picture. Ideal kappa value should be at least 0.6 and above and 0.8 and more is outstanding. The ***detection rate or Sensitivity*** is the number of samples of TP correctly identified. Its a ratio of TP/(TP+FN) which gives better model understanding. ***Specificity*** is the ratio of TN/(TP+FN). This value states that with how much accuracy can you predict the TNs.Following parameters will be assessed to identify the ideal model :

1) High Accuracy.
2) No information rate should be less than accuracy.
3) Cohen kappa value should tend nearest to 1.
4) High Detection rate.
5) P-value to be statistically significant i.e. less than 0.05.

Apart from the Confusion matrix and its parameter for accessing the model, we are using the Receiver Operating Curve(ROC). It checks for the goodness of fit of the model by plotting Sensitivity or True Postive rate vs Specificity or False Postive rate. The ideal value is located at the top left corner of the curve in order to predict accurate TP's.ROC value ranges from 0 to 1. Values that are closest to 1 are considered the best model.

Also, we are using the Regression model which we are evaluated on the following factor :

RMSE: It's the Root mean square error of the residuals obtained from predictions. It should be as low as possible.

RSS: It's the Residual Sum of squares. It should as low as possible.

MAE: It is computed by calculating the Mean of Absolute errors of the residues. It should as low as possible.

R Square: It is the percentage of the adjustment in the dependent variable that is predictable from the independent variable. It is used for testing the goodness of fit of the model. The best model has the value closest to 1.

A. Delhi Weather Dataset[22] **:**

Looking at the table, we can say that the Regression model doesn't perform well with R square of 0.652, RMSE of 4.15 and MAE of 2.86. Hence the model is rejected.

Now, coming to classification models, random forest yields the best result with an accuracy of 98.6%. Also, its value of No information rate is less than the accuracy value. This model has the best Kappa value of 0.893 which makes it the ideal model along with the detection rate of 0.924.

[4] Image Source: https://images.app.goo.gl/YUVgC6qw3ebrFaPd8
[5] Image Source : https://images.app.goo.gl/Y6Pu4pTkmyLLWbfs8

Legend	
<NA>	Not Applicable for the model

Evaluation Method	Delhi Weather dataset						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Multiple Regression	Logistic Regression	Decision Trees	KNN	SVM	Random Forest	LDA
RMSE	4.15	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
MAE	2.86	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
Adj. R Sq	0.652	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
RSS	0.264	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
Accuracy	< NA>	0.826	0.982	0.982	0.964	0.986	0.943
95% CI	< NA>	(0.822,0.83)	(0.98,0.983)	(0.822, 0.83)	(0.961, 0.966)	(0.985,0.988)	(0.94, 0.946)
No Information rate	< NA>	0.778	0.93	0.778	0.930	0.93	0.93
P-value [ACC > NIR]	< NA>	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
Kappa	< NA>	0.33	0.851	0.33	0.689	0.893	0.413
Mcnemar's Test P-Value	< NA>	<2e-16	<2e-16	<2e-16	<2e-16	0.0115	<2e-16
Sensitivity	< NA>	0.987	0.994	0.987	0.989	0.994	0.99
Specificity	< NA>	0.262	0.815	0.262	0.629	0.888	0.316
Pos Pred value	< NA>	0.824	0.986	0.824	0.973	0.992	0.951
Neg Pred value	< NA>	0.848	0.912	0.848	0.810	0.913	0.713
Prevalence	< NA>	0.778	0.93	0.778	0.930	0.93	0.93
Detection Rate	< NA>	0.768	0.925	0.768	0.92	0.924	0.921
Detection Prevalence	< NA>	0.932	0.938	0.932	0.946	0.932	0.969
Balanced Accuracy	< NA>	0.624	0.904	0.624	0.809	0.941	0.653
AUC	< NA>	0.836	0.949	0.949	0.891	0.952	0.832

Fig. 6. Results of Delhi Dataset

Followed by random forest, decision trees are the second-best model with a bit higher detection rate of 0.01 and also it has No information rate less than accuracy and rest all other factors are falling short of few numbers.
Whereas KNN is the worst amongst all the models. Its kappa value is 0.33 which is very low but it does have high sensitivity and accuracy is 82.6% but the overall performance of the model is not good and hence it is discarded.
Observing the ROC curves it is quite evident that random

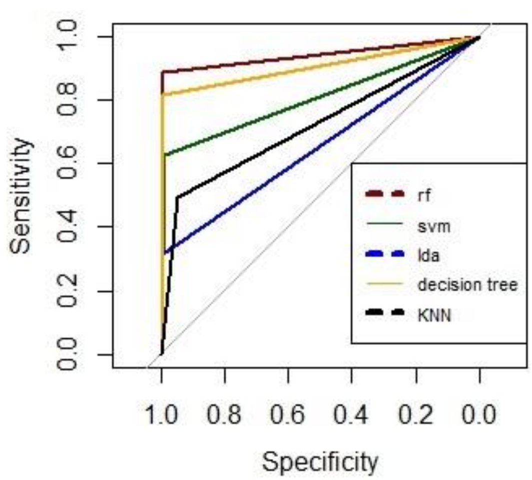

Fig. 7. ROC curve for Delhi Dataset

the forest is performed really well as compared to others. The random forest has the highest AUC value of 0.952 followed by a clash between KNN and decision trees with the same value of 0.949. Whereas LDA is the worst with an AUC value of 0.832.

B. Rain in Australia Dataset [23]:

Looking at the table, we can see that Regression doesn't perform well with R square as 0.5, RMSE as 4.19 and MAE as 2.98. Hence the model is discarded.//
Now, coming to classification models, Support Vector Machine(SVM) yields the best result with an accuracy of

Legend	
<NA>	Not Applicable for the model

Evaluation Method	Rain in Australia dataset						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Multiple Regression	Logistic Regression	KNN	Decision Trees	SVM	Random Forest	LDA
RMSE	4.19	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
MAE	2.98	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
Adj. R Sq	0.5	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
RSS	0.264	< NA>	< NA>	< NA>	< NA>	< NA>	< NA>
Accuracy	< NA>	0.826	0.981	0.982	0.987	0.964	0.943
95% CI	< NA>	(0.822,0.83)	(0.98,0.983)	(0.822, 0.83)	(0.985,0.988)	(0.961, 0.966)	(0.94, 0.946)
No Information rate	< NA>	0.778	0.93	0.778	0.93	0.930	0.93
P-value [ACC > NIR]	< NA>	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
Kappa	< NA>	0.33	0.851	0.33	0.893	0.689	0.413
Mcnemar's Test P-Value	< NA>	<2e-16	<2e-16	<2e-16	0.0115	<2e-16	<2e-16
Sensitivity	< NA>	0.987	0.994	0.987	0.994	0.989	0.99
Specificity	< NA>	0.262	0.815	0.262	0.888	0.629	0.316
Pos Pred value	< NA>	0.824	0.986	0.824	0.992	0.973	0.951
Neg Pred value	< NA>	0.848	0.912	0.848	0.913	0.810	0.713
Prevalence	< NA>	0.778	0.93	0.778	0.93	0.930	0.93
Detection Rate	< NA>	0.768	0.925	0.768	0.924	0.92	0.921
Detection Prevalence	< NA>	0.932	0.938	0.932	0.932	0.946	0.969
Balanced Accuracy	< NA>	0.624	0.904	0.624	0.941	0.809	0.653
AUC	< NA>	0.836	0.948	0.949	0.952	0.891	0.832

Fig. 8. Results of Rain in Australia Dataset

98.7%. Also, its value of No information rate is less than the accuracy value. This model has the best Kappa value of 0.893 which makes it the ideal model along with the detection rate of 0.924. Followed by SVM, decision trees are the second-best model with a bit higher detection rate of 0.01 and also it has No information rate less than accuracy and rests all other factors are falling short of few numbers Whereas

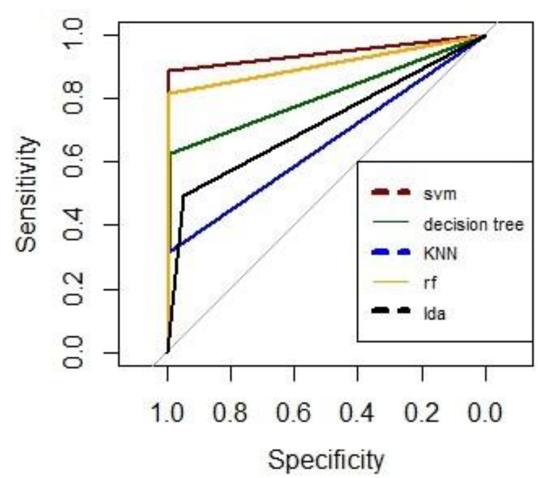

Fig. 9. ROC curve for Delhi Dataset

logistic regression is the worst amongst all the models. Its kappa value is 0.33 which is very low but it does have high sensitivity and accuracy is 82.6% but the overall performance of the model is not good and hence it is discarded.
Observing the ROC curves it is quite evident that SVM has performed really well as compared to others. The SVM has the highest AUC value of 0.952 followed by KNN and decision trees with a difference of 0.01. Whereas LDA is the worst with an AUC value of 0.832.

Some insights were drawn while were extracted while performing experiments :

1. The model's accuracy is directly dependent on the number of samples. If More records are chosen, the probability is high that the model will yield the highest accuracy.

2. Some models like decision trees performed extremely well when conditioned with fewer records.
3. But in achieving the end goal of obtaining the highest accuracy, other performance parameters were compromised which led to a decrease in its rank.
4. KNN models' accuracy is increased as k is gradually increased.
5. In Random Forest, OOB error rates were stabilized after 20 mtry(refer below graph).

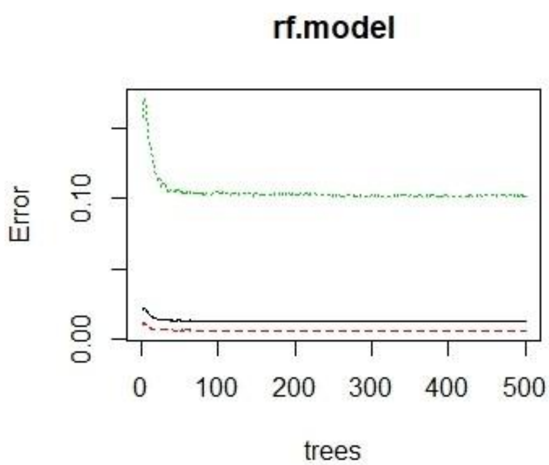

Fig. 10. Random Forest Performance

V. Conclusions and Future Work

1. To Conclude SVM has outperformed in both the datasets by maintaining its consistency. But one drawback was that as the payload was increased performance started to degrade and which decreased the Detection rate.
2. In forecasting rain, dew plays a keys role provided humidity predictors are normally distributed throughout the model.

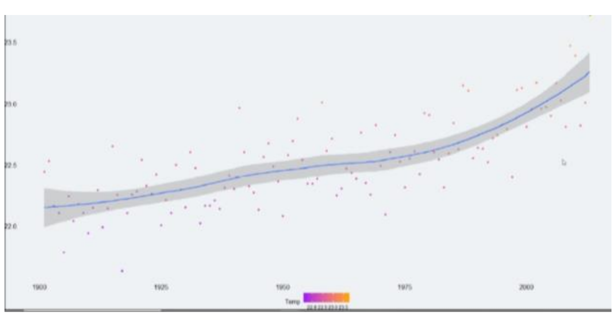

Fig. 11. Scatter Plot of Temperature based on Year

3. In the case of fog prediction, viscosity was the deciding factor in the decision tree for branching. (refer below fig)

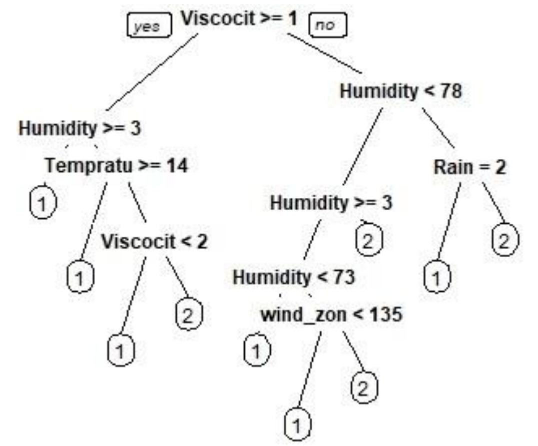

Fig. 12. Random Forest Performance

4. In future work, I would like to apply Principal Component Analysis to the existing models in order to reduce dimensions of the dataset and then calculate its efficiency.

5. Although no one can predict exactly whether of a particular instance as there is a probability of failure cases as well. But, Analysis can be carried to learn from the patterns and trying to predict accurate results.

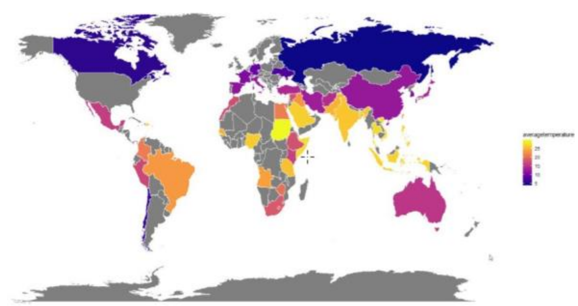

Fig. 13. Trend Analysis based on location

References

[1] H. Wang and L. Zhao, "A joint prevention and control mechanism for air pollution in the Beijing-Tianjin-Hebei region in china based on long-term and massive data mining of pollutant concentration," Atmos. Environ., vol. 174, pp. 25–42, Feb. 2018.
[2] A. Kumar and P. Goyal, "Forecasting of air quality in Delhi using principal component regression technique," Atmos. Pollut. Res., 2011.
[3] "Scopus preview - Scopus - Weather and Climate Extremes." [Online]. Available: https://www.scopus.com/sourceid/21100265617. [Accessed: 18-Dec-2019].
[4] "Scopus preview - Scopus - Monthly Weather Review." [Online]. Available: https://www.scopus.com/sourceid/13482. [Accessed: 18-Dec-2019].
[5] "Scopus preview - Scopus - Weather and Forecasting." [Online]. Available: https://www.scopus.com/sourceid/14764. [Accessed: 18-Dec-2019].
[6] "Scopus preview - Scopus - Journal of Space Weather and Space Climate." [Online]. Available: https://www.scopus.com/sourceid/21100286958. [Accessed: 18-Dec-2019].
[7] "Scopus preview - Scopus - Weather, Climate, and Society." [Online]. Available: https://www.scopus.com/sourceid/21100215718. [Accessed: 18-Dec-2019].

[8] "Scopus preview - Scopus - Weather." [Online]. Available: https://www.scopus.com/sourceid/14762. [Accessed: 18-Dec-2019].
[9] "Scopus preview - Scopus - World Scientific Series on Asia-Pacific Weather and Climate." [Online]. Available: https://www.scopus.com/sourceid/21100817633. [Accessed: 18-Dec-2019].
[10] "Scopus preview - Scopus - Climate Research." [Online]. Available: https://www.scopus.com/sourceid/12175. [Accessed: 18-Dec-2019].
[11] "Scopus preview - Scopus - Nature Climate Change." [Online]. Available: https://www.scopus.com/sourceid/21100198409. [Accessed: 18-Dec-2019].
[12] "Scopus preview - Scopus - Current Climate Change Reports." [Online]. Available: https://www.scopus.com/sourceid/21100873474. [Accessed: 18-Dec-2019].
[13] "Scopus preview - Scopus - Wiley Interdisciplinary Reviews: Climate Change." [Online]. Available: https://www.scopus.com/sourceid/19900191605. [Accessed: 18-Dec-2019].
[14] "Scopus preview - Scopus - Journal of Climate." [Online]. Available: https://www.scopus.com/sourceid/13550. [Accessed: 18-Dec-2019].
[15] "Scopus preview - Scopus - Climatic Change." [Online]. Available: https://www.scopus.com/sourceid/12177. [Accessed: 18-Dec-2019].
[16] "Scopus preview - Scopus - Urban Climate." [Online]. Available: https://www.scopus.com/sourceid/21100220478. [Accessed: 18-Dec-2019].
[17] "Scopus preview - Scopus - Climate Dynamics." [Online]. Available: https://www.scopus.com/sourceid/12172. [Accessed: 18-Dec-2019].
[18] "Scopus preview - Scopus - Climate Policy." [Online]. Available: https://www.scopus.com/sourceid/22524. [Accessed: 18-Dec-2019].
[19] "Scopus preview - Scopus - Climate Services." [Online]. Available: https://www.scopus.com/sourceid/21100463842. [Accessed: 18-Dec-2019].
[20] "Scopus preview - Scopus - Advances in Climate Change Research." [Online]. Available: https://www.scopus.com/sourceid/21100310030. [Accessed: 18-Dec-2019].
[21] "Scopus preview - Scopus - Climate and Development." [Online]. Available: https://www.scopus.com/sourceid/21100198541. [Accessed: 18-Dec-2019].
[22] "Delhi Weather Data | Kaggle." [Online]. Available: https://www.kaggle.com/mahirkukreja/delhi-weather-data. [Accessed: 18-Dec-2019].
[23] "Rain in Australia | Kaggle." [Online]. Available: https://www.kaggle.com/jsphyg/weather-dataset-rattle-package/metadata. [Accessed: 18-Dec-2019].