INTRODUCTION

Examining and protecting air quality has become one of the most essential activities for the government in many industrial and urban areas today. The meteorological and traffic factors, burning of fossil fuels and industrial parameters play significant roles in air pollution. With this increasing air pollution, Were in need of implementing models which will record information about concentrations of air pollutants (so2,no2,etc) [1]. Air pollution is a global environmental problem that influences mostly the health of urban population, and repeated exposures to ambient air pollutants over a prolonged period of time increases the risk of being susceptible to airborne diseases such as cardiovascular and respiratory diseases and lung cancer [2].An air quality index (AQI) can be defined as a communication tool and a standardized summary measure of ambient air quality used to express the level of health risk related to particulate and gaseous air pollution [3].Air Quality Index is also used by the government for easier understanding of the air pollution to common people. Awareness of the daily levels of air pollution is important for the citizens, because of the diseases spreading in the air exposed by the air pollution [4].

METHODOLOGY

Artificial intelligence and machine learning are areas of biggest rise in the last year. The science of artificial intelligence where system makes decision on its own, instead of working only by orders given by programmer as traditional programming works, gradually started influencing all aspects of our life [5]. After texting the accuracy of different regression model, the decision tree model has been used to predict the air quality index of Guwahati city. Different fields used Decision Tree algorithms and used it in their respective application. These algorithms can be used as to find data in replacement statistical procedures, to extract text, medical certified fields and also in search engines [6]. The decision tree here is use to predict the value of air quality index for a specific date. A decision tree classifies data items by posing a series of questions about the features associated with the items. Each question is contained in a node, and every internal node points to one child node for each possible answer to its question. The questions thereby form a hierarchy, encoded as a tree. In the simplest form we ask yes-or-no questions, and each internal node has a ‘yes’ child and a ‘no’ child. An item is sorted into a class by following the path from the topmost node, the root, to a node without children, a leaf, according to the answers that apply to the item under consideration. An item is assigned to the class that has been associated with the leaf it reaches [7].

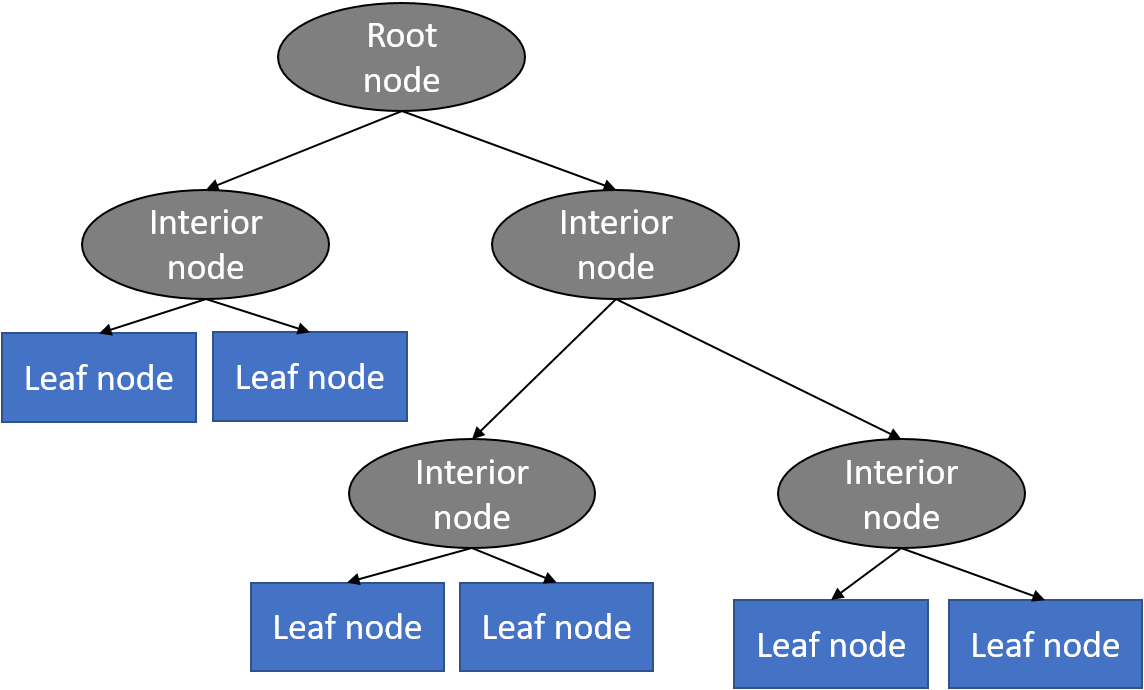


Figure 1

WORKING MODEL

date date

pollutant

Training set(80%)

Test set(20%)

date

pollutant

Figure 2

RESULTS ANALYSIS

Two types of graphs were plotted :

1. The AQI value of a single pollutant for all locations in one graph.

2. The AQI value of all the pollutants in one location.

Comparison has also been done between actual and predicted values.

Variation of the AQI parameters

The matplotlib library of python was used to plot the pollutants against the days the pollutants was measured.Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.Matplotlib was originally written by John D. Hunter. Python 3 support started with Matplotlib 1.2. Matplotlib 1.4 is the last version to support Python 2.6[8]

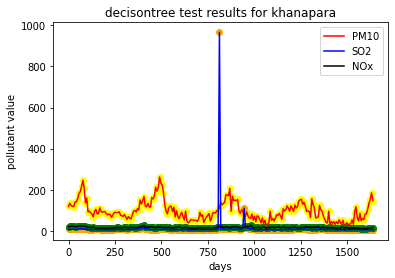


Figure 3.

In the above graph we can see that all the pollutants have been plotted for the station Khanapara. The pollutants have been plotted in the y-axis against the days the pollutants have been measured , plotted in the x-axis. The pollutants have been labelled as follows: PM10 has been represented with orange colour, SO2 has been represented by blue and Nox has been represented with black colour.

Dealing with Anomalies

In the above graph ,we can also see an abnormal spike in the SO2 value for a day. It has occurred due to an error in the sensor at that location. We chose to take that anomaly because we wanted to keep the collected data as much unaltered as possible. In the prediction from our model , the value was seen to be balanced. The results for predicted values for the khanapara are given below

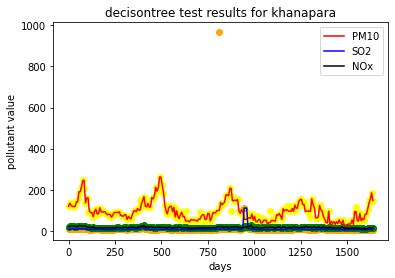
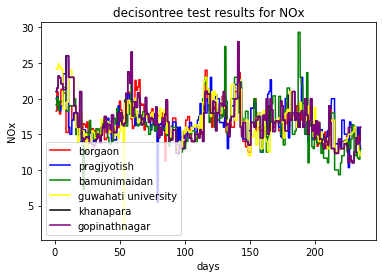


Figure 4

COMPARISON OF VALUES ACTUAL VS PREDICTED IN ALL THE STATIONS OF GUWAHATI

Total six number of graphs were plotted. The graphs were plotted for the three pollutants separately and included all the stations’ data.

1.For pollutant Nox



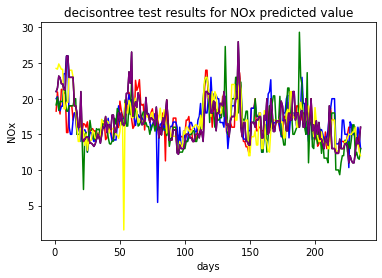
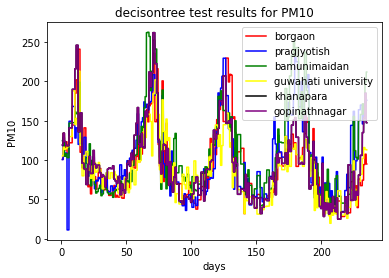


Figure 5

2.For Pollutant PM10



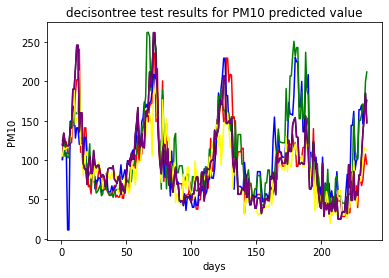
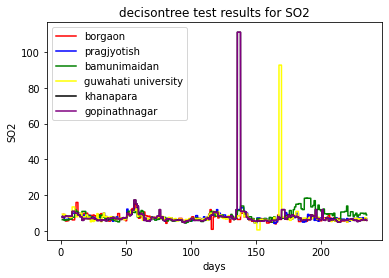


Figure 6

3.For pollutant SO2



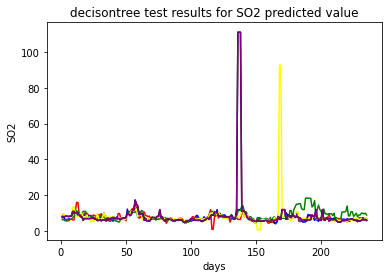
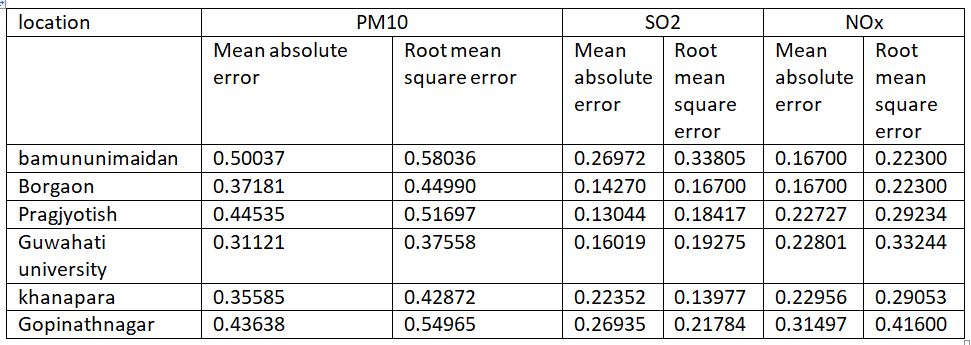


Figure 7

ACCURACY ANALYSIS

The following mean absolute error and root mean square errors were found



CONCLUSION

1. From the results obtained we can see that out of all the considered algorithms, the Decision Tree algorithm gives the highest accuracy, hence is most precise in predicting the AQI values of different locations of Guwahati.
2. The PM10 Values were higher than that of SO2 and NOx in all the locations.
3. The actual and predicted values of pollutants are almost linearly comparable.

REFERENCES

1. Title: Air Quality Prediction using Machine Learning Algorithms

International Journal of Computer Applications Technology and Research Volume 8–Issue 09, 367-370, 2019, ISSN:-2319–8656

Pooja Bhalgat , Sejal Pitale , Sachin Bhoite

1. WHO (2009) Global health risks: mortality and burden of diseases attributable to selected major risks. World Health Organization, Geneva, Available online at <http://www.who.int/healthinfo/global_burden_ disease/GlobalHealthRisks_report_full.pdf>
2. Kowalska M, Osrodka L, Klejnowski K, Zejda JE, Krajny E, Wojtylak M (2009) Air quality index and its significance in environmental health risk communication. Arc Environ Prot 35(1):13–21, ISSN: 0324-8461
3. Title : analysis of air quality index

Coimbatore Institute of Technology

Nikila Varshini.E1 , Sreeha.MR2 , Lhavanya Roobini. VN3 ,Vijayarangam.J4 , Sujithra.M5

1. Air Quality Index Prediction Using Simple Machine Learning Algorithms

International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)

Volume 7, Issue 1, January - February 2018 ISSN 2278-6856

Kostandina Veljanovska1 , Angel Dimoski2

1. Study and Analysis of Decision Tree Based Classification Algorithms

International Journal of Computer Sciences and Engineering

Vol.-6, Issue-10, Oct. 2018 E-ISSN: 2347-2693

Harsh H. Patel , Purvi Prajapati

1. NIH Public Access Author Manuscript

Nat Biotechnol. Author manuscript; available in PMC 2009 June 24

Published in final edited form as:

Nat Biotechnol. 2008 September ; 26(9): 1011–1013. doi:10.1038/nbt0908-1011.

8. <https://en.wikipedia.org/wiki/Matplotlib>