

JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

Department of CSE & IT



Bachelor of Technology, 5th Semester

ANALYSIS OF AGRICULTURAL BURNING IN NORTHERN INDIA AND ITS EFFECT ON AIR QUALITY OVER DELHI

MINOR PROJECT REPORT

(2019-2020)

Group details:

Shivam Bisht-17103199; B5
Meetanshi Mittal-17103343; B9
Soumya Agarwal-17103347; B9

Submitted to:

Dr. Anuja Arora
Dr. Vikas Hassija

Supervised by:

Dr. Parmeet Kaur

TABLE OF CONTENTS

<u>TOPICS</u>	<u>PAGE</u>
1. CERTIFICATE	02
2. ACKNOWLEDGEMENT	03
3. PROBLEM STATEMENT AND IMPORTANCE	03
4. RESEARCH PAPER SUMMARY	04
5. WORK PLAN	05
6. FLOW CHART	06
7. BRIEF DESCRIPTION OF DATASETS	07
8. BACK TRAJECTORY ANALYSIS	08
9. FAMILIARITY WITH TOOLS AND EQUIPMENTS	12
10.CONSLUSION	14
11.REFERENCES	14

CERTIFICATE

This is to certify that the project work for minor evaluation on the topic **ANALYSIS OF AGRICULTURAL BURNING AND ITS EFFECT ON AIR QUALITY OVER DELHI** submitted by

1. Shivam Bisht
2. Soumya Agarwal
3. Meetanshi Mittal

bearing enrolment number 17103199, 17103347, 17103343 and batch B5, B9, B9 respectively has been successfully completed under the guidance of Dr Parmeet Kaur.

Signature

Dr. Parmeet Kaur

Date: - 25 November 2019

ACKNOWLEDGEMENT

We would like to express our gratitude to Dr. Parmeet Kaur for her constant guidance which helped us complete the project. It was an interesting and an educational experience which helped us gain more knowledge in the field of machine learning.

PROBLEM STATEMENT

The purpose of our project is to study, analyze and observe the agricultural burning in the farming dominant states Punjab and Haryana and its adverse effect on the Air Quality (PM_{2.5}) over the Indian capital of Delhi.

We have used NASA's A-train satellite data (MODIS), ground-level PM_{2.5} measurements of New Delhi, and back-trajectory analysis, to show that the PM_{2.5} over New Delhi is strongly affected by the agricultural fires in the northwestern Indian states of Punjab and Haryana.

IMPORTANCE/RELEVANCE

In a recent study, scientists at Harvard University suggest that most of the air pollution that is responsible for public health emergencies in Delhi every winter is caused by crop burning in neighboring states. While Delhi is seeing an increase in PM_{2.5} particulate matter, the atmospheric PM_{2.5} levels show peaks in October-November, which is also the peak season for paddy harvesting, when farmers burn abundant plant residues to prepare for the next crop. Because the harvest season coincides with the post-monsoon conditions that favor wind stagnation in northern India, these conditions allow smoke to permeate gradually throughout the Indo-Gangetic region including Delhi. This smoke blends with current car and factory emissions which produces a dense, deadly haze. Notwithstanding the 2010 National Green Tribunal Act's national ban and regulations, the practice continues as farmers and government remain at loggerheads for a cost-effective alternate.

RESEARCH PAPER SUMMARY

Agricultural Burning and Air Quality over Northern India: A Synergistic Analysis using NASA's A-train Satellite Data and Ground Measurements

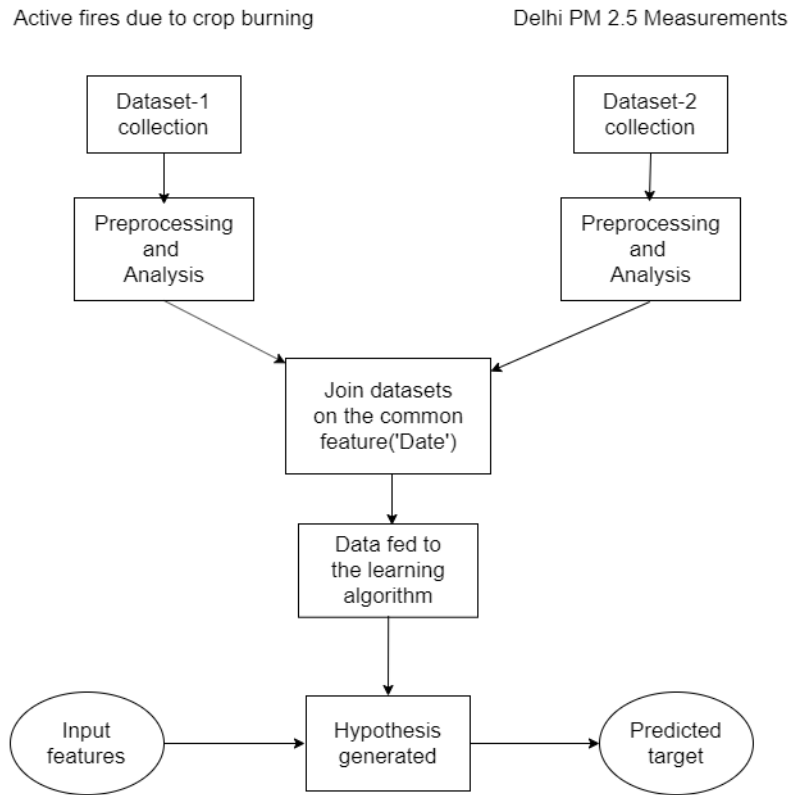
New Delhi, India's capital city, has ranked among the world's most polluted cities in recent years in terms of its submicron particulate matter (PM_{2.5}) air quality. Using NASA's A-train satellite data (MODIS), ground-level PM_{2.5} calculated in New Delhi, and back-trajectory measurements, we show that the PM_{2.5} over New Delhi is heavily affected by agricultural fires in the Punjab and Haryana northwestern Indian states during the post-monsoon season (October and November). During the peak burning period in early November, the mass concentration of PM_{2.5} increases from ~50 µg m⁻³ measured prior to residue burning in early October to as high as 300 µg m⁻³ (average 24-hour, 7-day running mean). A linear regression analysis shows that the variations in PM_{2.5} over New Delhi can be attributed to the simultaneous changes in the satellite recovery of fire counts and aerosols over the burning area. The back-trajectory analysis shows that most northwestern flow clusters (> 80 percent) near the ground intercepted the plant burning area before arriving at the receptor position in New Delhi; this further corroborates the transport trends inferred from the satellite data. A 15-year satellite record (2002–2016) shows a growing trend in November's agricultural fires (~617 per year) and aerosol charging (0.031 and 0.04 per year in optical aerosol depth and UV aerosol index). Increasing levels of crop residue burning and subsequent particulate matter emissions at an alarming rate over northern India is a pressing concern calling for corrective measures to substantially reduce or decrease crop burning by means of an active residue management system.

WORK PLAN

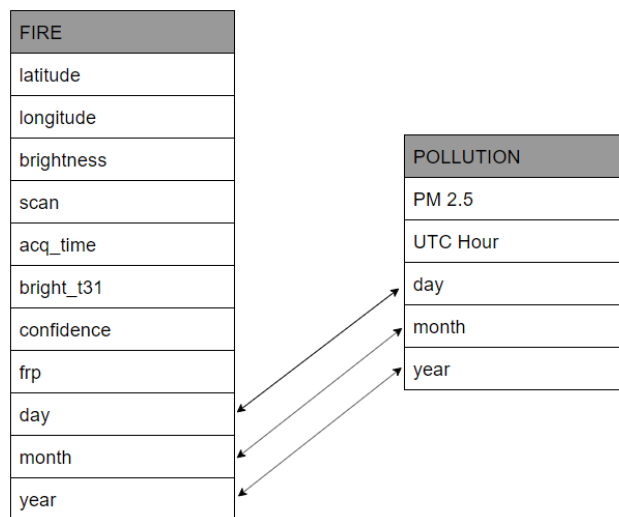
- Collection of datasets
 - i) NASA's A-train satellite data (MODIS)
 - ii) Ground-level PM2.5 measured in New Delhi
- Preprocessing and analysis
 - i) Visualization
 - ii) Correlation Matrix
 - iii) Feature Extraction
- Establishing relationships between datasets
- Mapping and joining of two datasets using python libraries
- Dimension reduction by PCA
- Training the combined datasets using various Supervised Machine Learning algorithms
- Classifying and analyzing the effect of crop burning in neighboring states of Delhi over its air quality.

Project Link- <https://github.com/Meetanshi18/crop-burning-analysis>

FLOW CHART



DATASET JOIN



BRIEF DESCRIPTION OF DATASETS

1. Active fires due to crop burning:

On board the Terra and Aqua platforms, the MODIS sensors detect fire spots globally on a daily basis at 1 / 1 km² of spatial resolution. Fire detection is carried out using a contextual algorithm (Giglio et al., 2003, 2016) which exploits the high mid-infrared radiation emission from fires. The algorithm consists of a series of spectral tests and each pixel of the MODIS image is classified. In the end, it assigns to each pixel one of the following classes: missing information, clouds, water, non-fire, flame, or unknown.

2. Delhi PM 2.5 measurements

Hourly observations of concentration for Delhi are given. All times are in UTC time. This data is based on the ground-level monitoring stations ' regional interpolation of real-time observations. Since the aim is to capture regional air quality differences, be aware that individual air quality monitoring that record values for PM_{2.5} concentrations that are somewhat higher or lower than those indicated by the local averages recorded here. These time series should be considered as preliminary and subject to change as real-time data is being used.

Back-trajectory Analysis

The figure below shows the 3-day (72-hour) back trajectory clusters calculated over northern India using the HYSPLIT model. Back trajectories are measured at three altitudes, i.e. 100 m (a), 500 m (b) and 1500 m (c) for each day of October and November 2013–2016, beginning at 1:30 PM local time in New Delhi. The transportation will take place over the first few hundred meters of the atmosphere in order for the ground sensor in New Delhi to test the smoke particles released over the plant burning areas. Therefore, for the back-trajectory analysis, we choose three altitudes, i.e. two closer to the ground (100 m and 500 m), which relate to the PM_{2.5} and one near the free boundary of the troposphere at 1500m. Both October and November trajectories (a total of 244) are grouped according to the concurrent 24-hour average PM_{2.5} measured at the US Embassy site in New Delhi (shown as the legends at the top). This helps us to connect PM_{2.5} variations at the receptor location via transport mechanism to the origin regions. Figure shows that most trajectories at the near surface level (100 m), quantitatively, 52%, 81%, 89%, and 84%, corresponding to the respective range of PM_{2.5} ($\mu\text{g m}^{-3}$), i.e., $0 < \text{PM}_{2.5} < 100$, $100 < \text{PM}_{2.5} < 200$, $200 < \text{PM}_{2.5} < 300$, and $\text{PM}_{2.5} > 300$, intercepted the crop burning region of Punjab and Haryana (depicted as a box in the figure) partially or fully before arriving at the receptor location in New Delhi. For the 500 m elevation, the corresponding percentages of all back trajectories are 50%, 89%, 86% and 84%. Noticeably, more than 80 percent of all trajectories are associated with higher PM_{2.5} rates at both altitudes that overlap plant burning areas. In other words, days with extreme levels of PM_{2.5} are correlated with the burning areas ' northwestern wind, indicating a profound impact of smoke transport on air quality regulation over New Delhi. Figure (bottom, d) shows as a function of time the average altitude dependence of the trajectory path. An average of all trajectories was performed on the hourly data. The average trajectories for the altitudes of 100 m and 500 nm showed significant subsidence of smoke-laden air mass along the transportation pathways suggesting downward movement of particulate pollution to the lower atmosphere levels while increasing the concentration of particles on the ground above the receptor position. An analysis of back-trajectories calculated for several individual days reveals that the transport of smoke produced from agricultural fires typically takes less than a day (~14– 22 hours) to be mobilized from the central areas of burning in Punjab to the receptor location in New Delhi. Since the burning of the paddy residue is widespread throughout the region, the actual time taken to mobilize smoke particles is attributable to the distance between the positions of the flame and the velocity of the receptor and wind. Assuming the wind speed of 5 m s⁻¹ of the northwest stream at the lower atmospheric levels (> 850 hPa), it would take between 8, 15 and 22 hours to transport smoke.

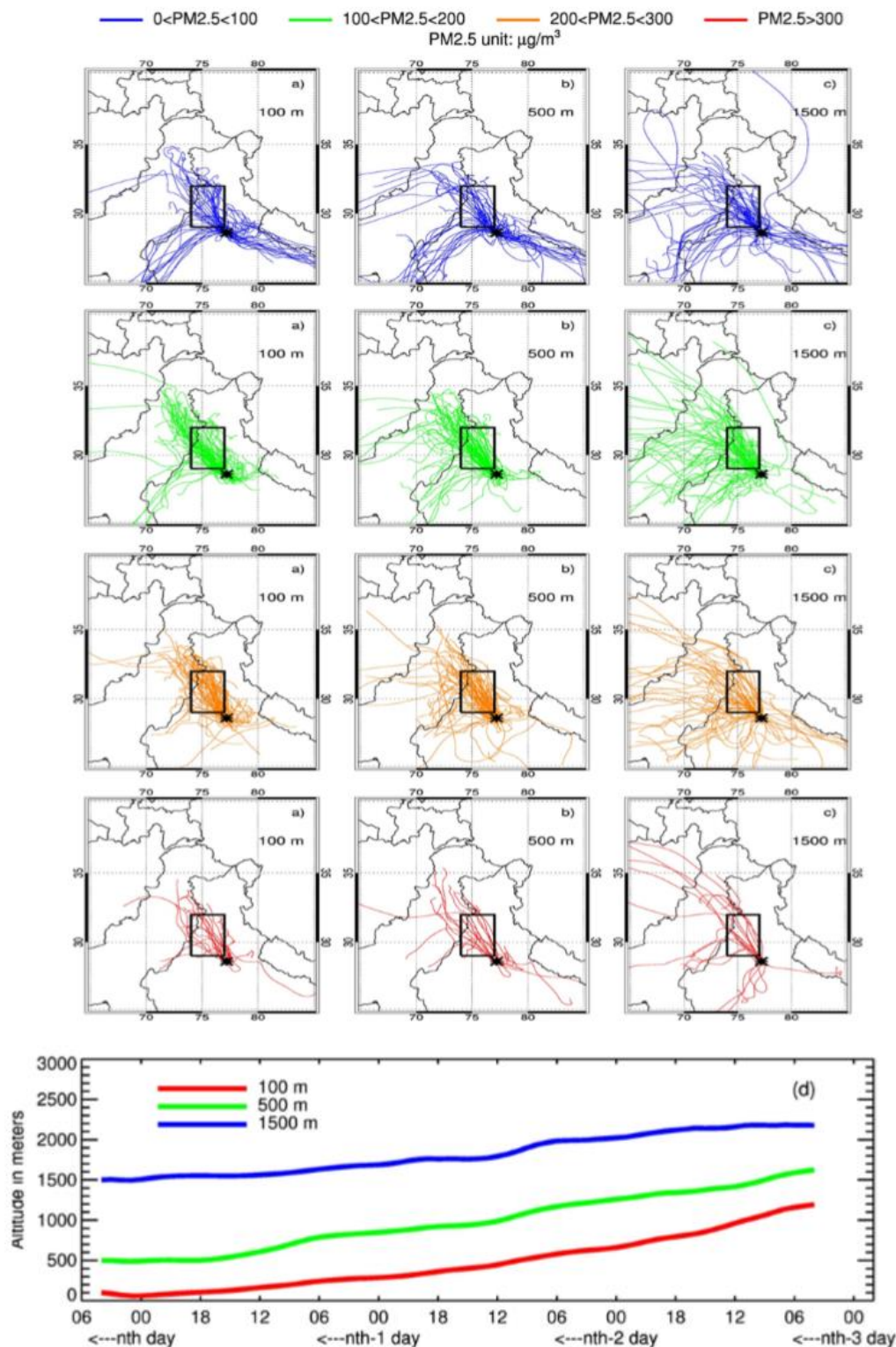
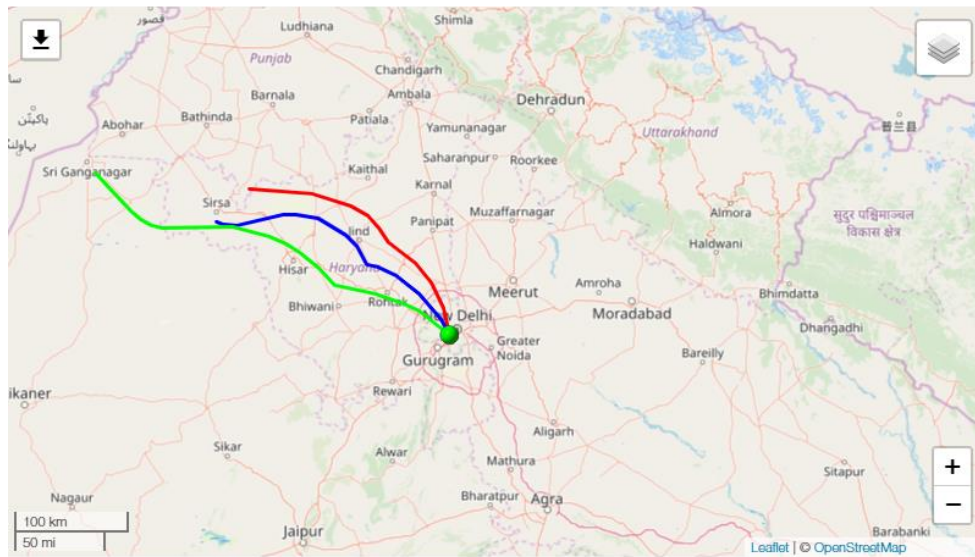
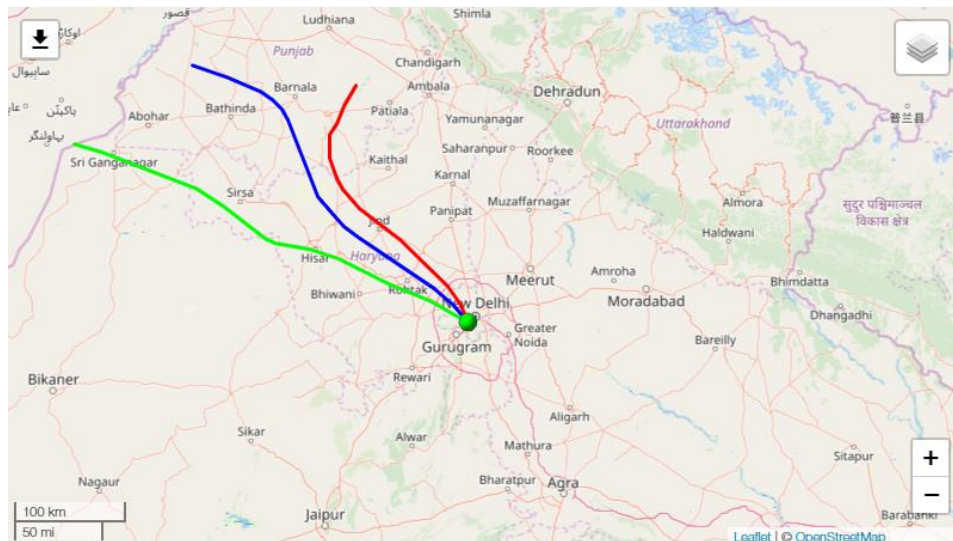


Fig. Clusters of HYSPLIT 3-day back-trajectories originating in New Delhi, India (marked with an asterisk) in October and November. Back-trajectories were calculated once per day with the starting local time of 1:30 PM and for the three altitudes, i.e., (a) 100 m (near-surface), (b) 500 m, and (c) 1500 m. Discrete color codes assigned to each back-trajectory (shown as legends on the top) represent a corresponding range of 24-hour averaged PM2.5 measured at the US Embassy site in New Delhi. The crop burning area of the Indian states of Punjab and Haryana is depicted as a bounding box. The bottom panel (d) shows the averaged altitudinal variations of the back-trajectories as a function of time for the three reference altitudes

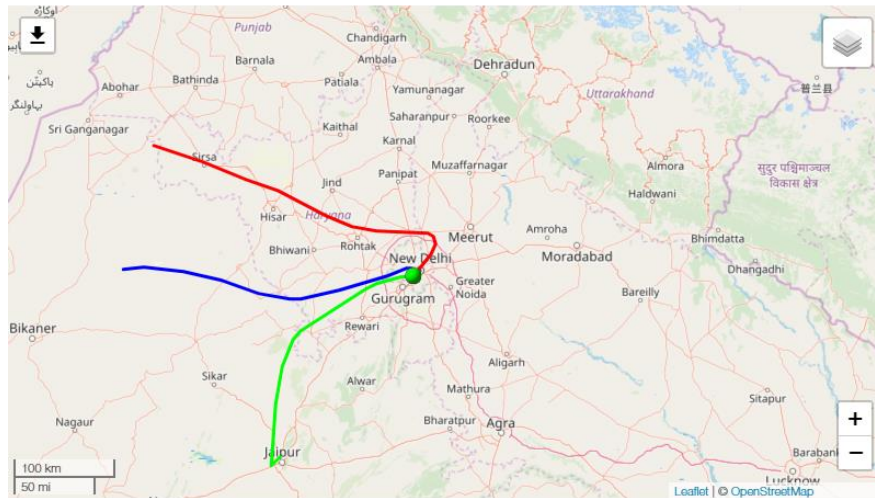
WIND FLOW DIRECTION SHOWN BY HYSPLIT MODEL



Date: 20 October 2017



Date: 10 November 2017



Date: 15 October 2018



Date: 27 October 2019 [Diwali]

FAMILIARITY WITH TOOLS AND EQUIPMENTS USED

Preprocessing Algorithm: Principal Component Analysis (PCA)

The most common technique of dimensionality reduction to date is the Principal Component Analysis (PCA). It allows us to take a n-dimensional feature-space and reduce it to a k-dimensional feature-space while in the reduced dataset keeping as much information as possible from the original dataset. In addition, PCA must create a new feature-space to capture as much variation in the original dataset as possible. We found for our dataset that the number of Principal Components was 4, hence reducing the dimension of our dataset from 11 to 4.

Training Algorithms:

1. **Linear Regression**-It is an algorithm for machine learning based on supervised learning. Regression models a target prediction value based on independent variables. The task of linear regression is to predict a dependent variable value (y) based on a given independent variable (x). This technique of regression considers a linear relationship between x (input) and y(output). The title is therefore Linear Regression.
The parameter's value used to train the model are:
0.01622218, 0.00257253, 0.03722611, -0.27733538
The accuracy score of fitting the model is 0.0874.
2. **Support Vector Machine**-This algorithm creates a large or infinite dimensional space hyperplane or set of hyperplanes that can be used for classification, regression or other tasks. Intuitively, the hyperplane that has the maximum distance to the closest training data points of any class (so-called functional margin) achieves a strong separation, since the greater the margin is usually the lower the classifier's generalization error.
The accuracy score of fitting the model is 0.8362637362637363
The F1 Score (average='micro') of testing model on test set is
0.793859649122807
The F1 Score (average='weighted') of testing model on test set is
0.7501773840995183
The accuracy score of testing model on test set is 0.793859649122807
3. **Logistic Regression** - is a statistical model that uses a logistic function to model a binary dependent variable in its basic form, even though there are many more complex extensions. Logistic regression (or logit regression)

estimates the parameters of a logistic model (a form of binary regression) in regression analysis.

The accuracy score of fitting the model is 0.7901098901098901

The F1 Score (average='micro') of testing model on test set is 0.75

The F1 Score (average='weighted') of testing model on test set is 0.6507644238874076

The accuracy score of testing model on test set is 0.75

4. **Decision Tree** -Through mathematics, learning from the Decision tree uses a decision tree to go from assumptions on an item to conclusions on the target value of the item. It is one of the methods used in analytics, data mining, and machine learning predictive modeling.

The accuracy score of fitting the model is 0.8626373626373627

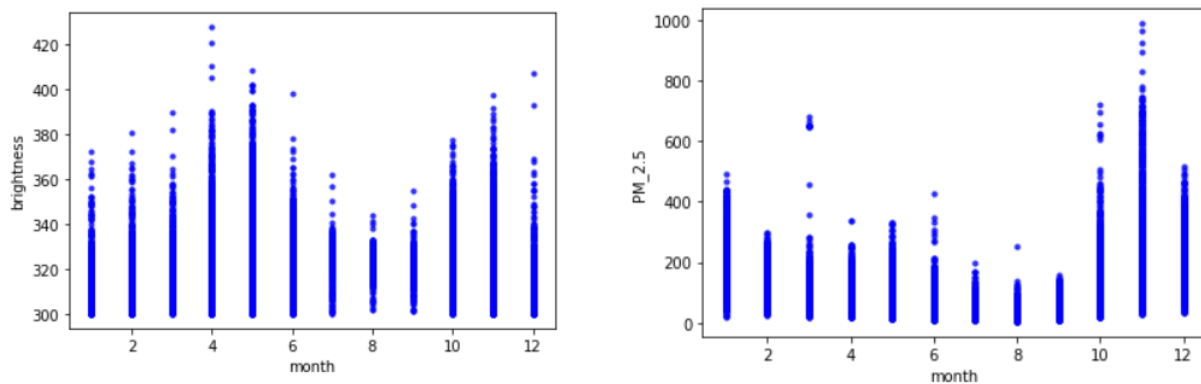
The F1 Score (average='micro') of testing model on test set is 0.8464912280701754

The F1 Score (average='weighted') of testing model on test set is 0.8815574322162372

The accuracy score of testing model on test set is 0.8464912280701754

CONCLUSION

The correlation analysis suggests a strong influence of crop-burning in Punjab and Haryana as observed by MODIS on the concentration of PM_{2.5} measured on the ground in New Delhi. We have developed a basis of the cause-and-effect relationship between the satellite observations over the burning region and ground measurements of PM_{2.5} downwind in New Delhi, as shown in figure below. These relationships are useful for predicting (classifying) the PM_{2.5}, albeit with some uncertainties, given the satellite measurements in the near-real time.



Different factors like wind speed, wind direction, boundary layer dynamics, and altitude dependence of particles can introduce uncertainties in the desired relationship. Due to the temperature drop in the months of October and November the dispersion effect of particles is reduced therefore settling over Delhi area and contributing towards increased pollution as opposed to high temperature and lighter air in months like April and May.

REFERENCES

<https://firms.modaps.eosdis.nasa.gov/download/create.php>

<http://berkeleyearth.lbl.gov/air-quality/maps/cities/India/NCT/Delhi.txt>

http://www.aaqr.org/files/article/6727/19_AAQR-17-12-2017AAC-0583_1756-1773.pdf

<https://www.ready.noaa.gov/hypub-bin/trajtype.pl>

medium.com

[google.com](https://www.google.com)