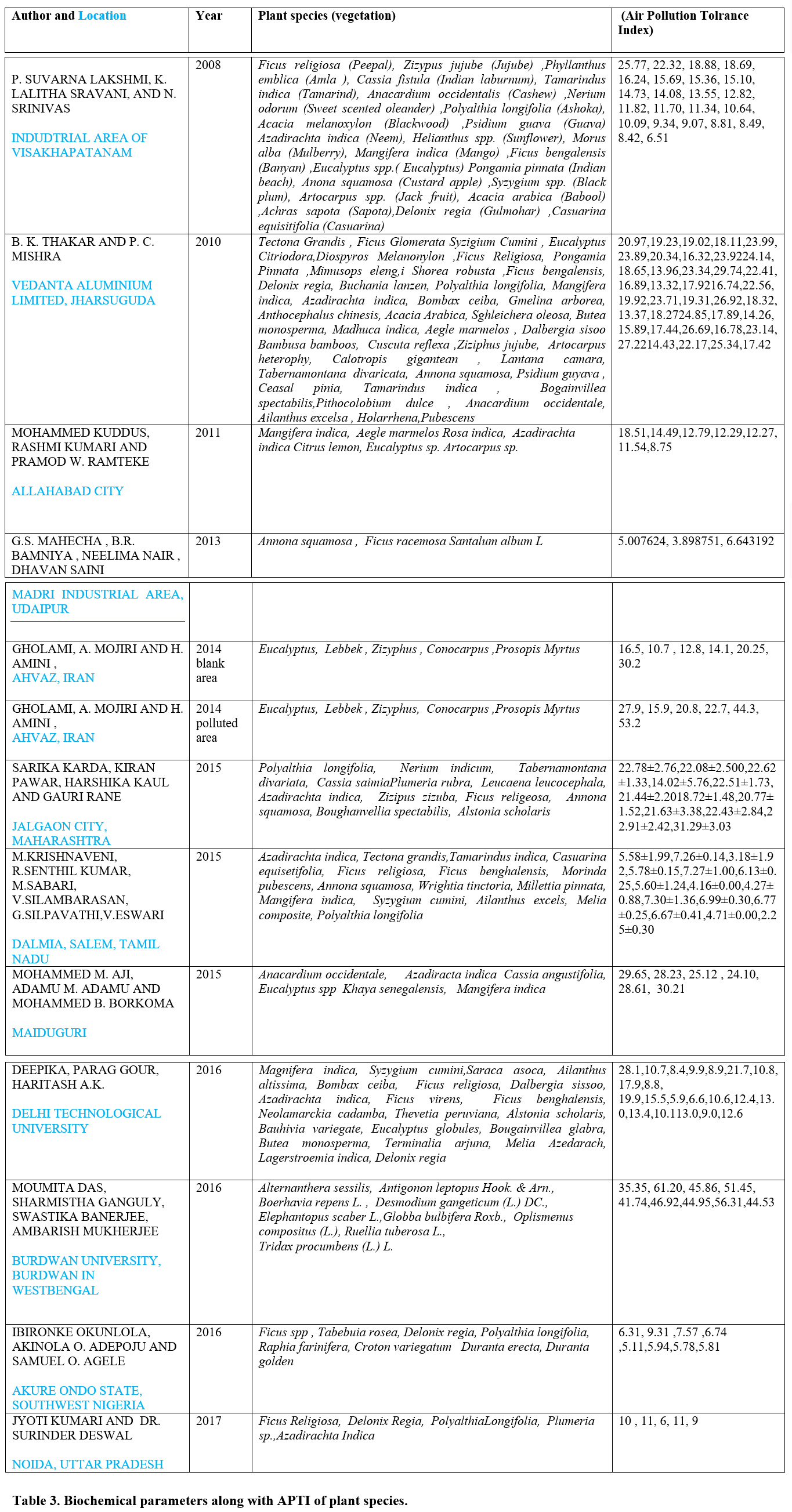
**BI-Project-Final from Team 13**

**DATA**

**Data Extraction**

* <https://www.kaggle.com/india-air-quality-data> ---> Air Quality from 2000-2015
* <https://www.aqicn.info/historical> ---> Air Quality from 2016-2019 Sate-Wise
* <http://www.frienvis.nic.in> ---> Tree and Forest Cover in alternate years
* <https://www.biodiversityofindia.org/> ---> Plant Species available in a State
* <http://www.cwejournal.org/vol13no1/a-review-on-air-pollution-tolerance-index–apti–and-anticipated-performance-index–api-/>
* <https://api.ocr.space/parse/image> ---> APTI Values (Vol13\_No\_Are\_Lak\_T1.jpg)



**Code**

**Data Pre - Processing:**

==> APTI Values (Vol13\_No\_Are\_Lak\_T1.jpg) ---> APTI Values.csv

==> Raw Air Dataset ---> india-air-quality-data2.zip; Extracted to ---> airdatasetnew.csv

==> Plant Species Extraction ---> plant\_scraping.ipynb; Species\_details.csv

==> Air Dataset ---> aircsv.ipynb; aircsv.csv (has state-wise csv files too in Folder ‘’air datasets”)

==> Green Cover ---> Tree Cover and Forest Cover datasets are merged into one; all the datasets are in the Folder “tree & forest cover extraction” with merged csv file as ---> Total\_Green\_Cover.csv; green\_cover.ipynb (the Tree Cover for the Year 2019 was taken from pdf file “ indian State forest Cover - 2019” manually)

**Data Visualization:**

==> For AQI vs APTI ---> AQI vs APTI.ipynb (the Air Quality Index is plotted against Air Pollution Tolerance Index to find out their relation on one another)

==> For AQI vs Green Cover ---> AQI vs Green Cover.ipynb (the Air Quality Index is plotted against Green Cover to find out their relation on one another)

**Data Modelling:**

==> Supervised Learning Algorithm, Linear regression is used to find out that how much AQI was needed according to that Green Cover in that state at that time and how much was already there available ---> “Linear\_Regression.ipynb”

Inputs: x-axis --> AQI; y-axis --> Green Cover

==>Unsupervised Learning Algorithm, K-Means Clustering is used to find out the Tolerant Plant Species region-wise for controlling air pollution ---> “k-means,ipynb”

**Help - Mapped Report**

In Section 3,

==> Section 3.1 and 3.2 are made from the following

>> Table 1: AQI Dataset is made from ---> aircsv.ipynb ---> aircsv.csv

>> Table 2: Tree Cover dataset is made from ---> tree\_cover.ipynb ---> tree\_cover.csv

>> Table 3: Forest cover is made from by combining all Forest Cover Datasets into one Dataset ---> merge.ipynb ---> forest\_cover.csv

Then both of these Tree and forest Cover datasets are merged to form one Green Cover Dataset ---> green\_cover.ipynb ---> Total\_Green\_Cover.csv

>> Table 4: Species Available Dataset is made from ---> plant\_scraping.ipynb ---> Species\_details.csv

>> Table 5: APTI Values Dataset is made from ---> Vol13\_No\_Are\_Lak\_T1.jpg ---> APTI Values.csv

In Section 4,

==> Section 4.1 is made from ---> AQI vs APTI.ipynb

==> Section 4.1 is made from ---> AQI vs Green Cover.ipynb

In Section 5,

==> Section 5.1 is made from ---> k-means.ipynb

==> Section 5.2 is made from ---> Linear\_Regression.ipynb

References are taken from,

==> Begum, A., Harikrishna, S.: Evaluation of some tree species to absorb air pollutantsin three industrial locations of south bengaluru, india. Journal of Chemistry7(122010). <https://doi.org/10.1155/2010/398382>

==> Panda, L.L., Aggarwal, R., Bhardwaj, D.: A review on air pollution tolerance index(apti) and anticipated performance index (api). Current World Environment13(1),55 (2018)

==> Tripathi, A., Gautam, M.: Biochemical parameters of plants as indicators of airpollution. Journal of Environmental Biology28(1), 127 (2007)

**Report**

Overleaf Link: <https://www.overleaf.com/project/5e7c94608eb4030001ac2fda>

Pdf File: BI\_Project\_Final\_Team\_13.pdf

All Codes Available: <https://github.com/Team13-ml/Air-Pollution-Tolerance>