**SHRI VAISHNAV VIDHYAPEETH ISHWAVIDHYALAYA**

**INDORE**



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**SESSION: 2022-2023**

**Enrolment No. :**20100BTCSDSI07273

**Year:** III

**Semester:** 5th

**Subject –** **Data Science**

BTIT505

**Submitted to Submitted by**

Prof. Om Kant Sharma Sir Harsh Chourasiya

**Guidelines for major project synopses project**

1. **Introduction :**

Air pollution is contamination of the indoor or outdoor environment by any chemical, physical or biological agent that modifies the natural characteristics of the atmosphere.

Household combustion devices, motor vehicles, industrial facilities and forest fires are common sources of air pollution. Pollutants of major public health concern include particulate matter, carbon monoxide, ozone, nitrogen dioxide and sulfur dioxide. Outdoor and indoor air pollution cause respiratory and other diseases and are important sources of morbidity and mortality.

WHO data show that almost all of the global population (99%) breathe air that exceeds [WHO guideline limits](https://www.who.int/publications-detail-redirect/9789240034228) and contains high levels of [pollutants](https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants), with low- and middle-income countries suffering from the highest exposures.

Air quality is closely linked to the earth’s climate and ecosystems globally. Many of the drivers of air pollution (i.e. combustion of fossil fuels) are also sources of greenhouse gas emissions. Policies to reduce air pollution, therefore, offer a win-win strategy for both climate and health, lowering the burden of disease attributable to air pollution, as well as contributing to the near- and long-term mitigation of climate change.

1. **Problem Domain :**

From smog hanging over cities to smoke inside the home, air pollution poses a major [threat to health](https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts) and climate.

Ambient (outdoor) air pollution in both cities and rural areas is causing fine particulate matter which result in strokes, heart diseases, lung cancer, acute and chronic respiratory diseases.

Additionally, around 2.4 billion people are exposed to dangerous levels of household air pollution, while using polluting open fires or simple stoves for cooking fuelled by kerosene, biomass (wood, animal dung and crop waste) and coal.

The combined effects of ambient air pollution and household air pollution is associated with 7 million premature deaths annually.

Sources of air pollution are multiple and context specific. The major outdoor pollution sources include residential energy for cooking and heating, vehicles, power generation, agriculture/waste incineration, and industry. Policies and investments that support sustainable land use, cleaner household energy and transport, energy-efficient housing, power generation, industry, and better municipal waste management can effectively reduce key sources of ambient air pollution.

1. **Solution Domain :**

**Logistic regression :**

* Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

**Decision tree classifier :**

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.

**Linear regression :**

**Linear Regression** is a machine learning algorithm based on **supervised learning**. It performs a **regression task**. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

**Decision tree regressor :**

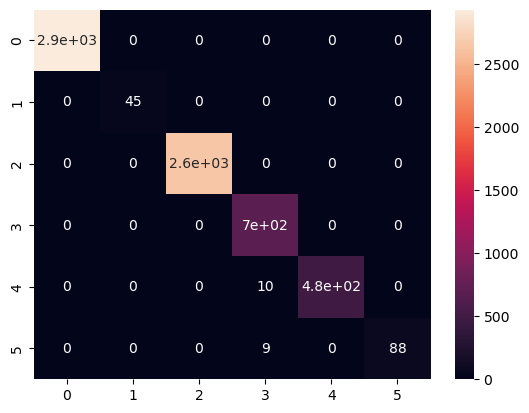
Decision trees is a type of supervised machine learning algorithm that is used by the Train Using Auto ML tool and classifies or regresses the data using true or false answers to certain questions. The resulting structure, when visualized, is in the form of a tree with different types of nodes—root, internal, and leaf

1. **System Domain :**

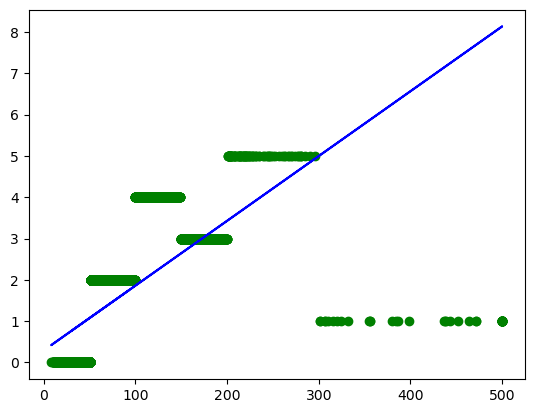
* **Python:** Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.
* **Jupyter Notebook :** JupyterNotebook (or “notebooks”, all lower case) are documents produced by the [Jupyter Notebook App](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html" \l "notebook-app), which contain both computer code (e.g. python) and rich text elements (paragraph, equations, figures, links, etc…). Notebook documents are both human-readable documents containing the analysis description and the results (figures, tables, etc..) as well as executable documents which can be run to perform data analysis.
* **Kaggle (Data set) : https://www.kaggle.com/**

1. **Application Domain :** This project is gives a prediction of air quality in all over the world by predicting the data set. So that we can control the pollution level of Air.
2. **Expected Outcome :**

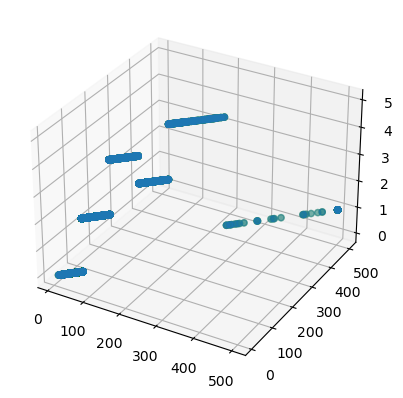
**Logistic Regression:**

****

**Linear Regression:**

****

**Decision tree regressor:**

****

**7. Reference :**

**Kaggle:** [**https://www.kaggle.com/**](https://www.kaggle.com/)