VISIBILITY PREDICTION
FOR AIR TRAFFIC CONTROL

PREPARED BY – YATRIK SHAH

DEPARTMENT OF ARTIFICIAL INTELLIGRNCE AND MACHINE LEARNING

SCHOOL OF EMERGING SCIENCE AND TECHNOLOGY

GUJARAT UNIVERSITY

1. INDEX

2. ABSTRACT

In today's world, the problem of human safety is increasing day by day due to absorption of light to reduce the clarity and colours of what you see. This affects the visibility in the environment, and it goes down. Many factors can cause visibility in the environment like haze, air pollution, climate change etc. Considering this, I have decided to make the machine to predict the level of visibility in the environment taking various features like temperature, wind speed, humidity etc. It becomes very important in cities like Delhi, where the air pollution remains at peak level, or the cities where there is high amount of haze which can cause the air traffic problem. My model will predict the level of visibility level in the environment so we can avoid these problems before it arises.

3. INTRODUCTION

Visibility is a measure of the distance at which an object or light can be clearly perceived. We predict the visibility of an object from a particular distance in kilometers based on: Dry Bulb Temp, Wet Bulb Temp, Relative humidity and many more. Visibility is affected by certain environmental factors and may

vary according to the direction and angle of view and the height of the observer. In some certain situations Visibility as a limitation factor affects a lot. Football match, f1 racing, takeoff and landing in Air Traffic Control are some of the examples of these situations where the factor visibility affects a lot. There visibility plays a key role in the public safety factor.

4. REAL LIFE APPLICATIONS

- Key value in deciding if the flight would take off or land or not.
- Flight's Visibility is an important factor in all phases of flight, but especially when an aircraft is maneuvering on or close to ground, at that time weather must be clear in order to avoid accidents.
- In addition, it has very important value in knowing distance from the land while docking a ship at a dockyard.
- Car Racing is not allowed if there isn't proper visibility.
- Useful in day-to-day life as well while driving vehicles.

5. Data description

1. VISIBILITY - Distance from which an object can be seen.

2. DRYBULBTEMPF-Dry bulb temperature (degrees Fahrenheit). Most reported standard temperature.

3. WETBULBTEMPF-Wet bulb temperature (degrees Fahrenheit).

4. DewPointTempF-Dew point temperature (degrees Fahrenheit).

5. Relative Humidity-Relative humidity (percent).

6. Wind-Speed-Wind speed (miles per hour).

7. WindDirection-Wind direction from true north using compass directions.

8. StationPressure-Atmospheric pressure (inches of Mercury; or 'in Hg').

9. SeaLevelPressure- Sea level pressure (in Hg).

10. PreciP Total-precipitation in the past hour (in inches).

6. WORKFLOW STEPS

- Raw Data Collection
- Data Pre-processing
- Eda and Data Visualization
- Feature Engineering
- Logging
- Maintain Training Data in Database
- Cross Validation and Hyper Parameter Tuning
- Model Selection
- Maintaining Prediction Database
- Creating API
- Integrating with WEB FRAMEWORK - flask
- Model Deployment

7. Important Tools used in the Project

8. METHODOLOGY

8.1 DATA PRE-PROCESSING AND EDA

8.1.1 Observation:

As a first step, the data was observed, and all the features were taken into consideration.

8.1.2 Null Values:

In the dataset, it was checked whether there are the null values or not and if it is, how many records it contains.

8.1.3 Trends and Patterns:

Visualized some important graphs and charts to observe the considerable   trends and some patterns to know the relationship between the features.

8.2 FEATURE ENGINEERING

8.2.1   Feature Selection:

Basically, Feature selection is considered as the first step in Feature Engineering. Here all the features are selected which are truly helpful in achieving Machine Learning end-goal i.e.(here) Regression.

There are different methods in feature selection to select important features. Some of these are: Chi^2 test, feature importance (by the model), Correlation Matrix Heatmap

(using Karl Pearson's correlation and Spearman's Rank correlation), Gini impurity, Entropy and Information Gain, etc. Here the Correlation Matrix Heatmap is used to select the best features for the machine learning model furthermore, Gini impurity is used in the Random Forest model as attribute selection method.

8.2.2 Scaling:

In Machine Learning, scaling is a very useful step of Feature Engineering. The goal of feature scaling in machine learning is to scale all the important features into same range or scale. Some Machine Learning models expect all the features to be scaled properly [Generally in Standard Normal distribution].

Standardization was used as a method of scaling to bring down all the features to a common scale without distorting the differences in the range of the values.

8.2.3Multicollinearity:

In Machine Learning Multi-collinearity is the problem where two or More independent columns or features are correlated internally. It may Cause. In Regression Analysis it is assumed that there is no exact relationship among the exploratory variables and in regression analysis this assumption is violated then the problem of multi-collinearity occurs. Removed multi-collinearity by creating heatmap of correlation metrics.

8.2.4Imputation [Handling null values]:

In Feature Engineering one of the important step is considered as Imputation or handling the Null values which means, observing Null values or empty values(thresholds) and replacing them with a best-fit meaningful substitute.

Null values were critically observed and it was removed from all the features in the dataset.

8.3 LOGGING

Logging is a very fundamental part of every end-to-end project or software. Logging let us track events when the program gets executed, so that when the code crashes, we can check the logs and identify that what caused it. Very robust software and programs can be created using Logging.

Here, logs are maintained and get informed about each information, warnings, errors in any program, which helps us to track the program. Logging is very useful specially in case of program crashing. It is an essential part of troubleshooting application and infrastructure performance.

8.4 MAINTAINED TRAINING AND PREDICTION DATABASE

A database is defined as a structured set of data which is stored in computer's memory or some cloud platform. In large and advanced projects storing each data is a challenging task. Which can further be used in some other tasks like maintaining

training data database which can further be divided into good data and bad data by data validation.

After data cleaning, exploratory data analysis and feature engineering, dumping the final data for model training in the database.

This data can directly be taken for model retraining purpose.

Here the **prediction database** is the database where the data is stored which were tested on the deployed API after deployment.

In more advanced data science projects Prediction Database plays more important role mostly in retraining approach.

Here SQLite database is maintained.

8.5 CROSS VALIDATION AND HYPER PARAMETER TUNING

- Cross validation and Hyperparameter tuning is one of the most important steps in Machine Learning, since the performance of the machine learning model is affected upon the hyperparameters those are selected for model training.

- Performed Cross validation and Hyperparameter tuning on all the models and selected the best models those performs best on the validation sets according to the given parameter grid.

- Found and selected the best hyperparameters of all the models for model training   on Full data.

8.6 CREATING API

- A real-world Data Science based does not end just after successfully building a machine learning model that performs well on the test dataset.
- In real-world industrial end-to-end Artificial Intelligence or Data Science based projects creating Front-End API is also having equal importance as back-end tasks like- Model Training or Model Evaluation
- The Front-End is created with the help of HTML (Hypertext Markup Language) and CSS (Cascading Style Sheets), which are very powerful tools.
- Flask is used as WEB – API in python that integrates back-end and front-end.
- Flask depends on the Jinja template engine and the Werkzeug WSGI toolkit.

8.7 MODEL DEPLOYMENT

- Generally Model Deployment is considered as the last step in any Machine Learning project.
- After successfully creating front-end API with HTML and CSS and using Flask as web Framework only Model Deployment is left.
- Here HEROKU platform is chosen as Model Deployment.

- Heroku is a container-based cloud Platform as a Service (PaaS). Developers use Heroku to deploy, manage, and scale modern apps.