International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

Cloud Base Climate Visibility Prediction Using Machine Learning

Mr. Karan Mohan Mane[1], Ms. Shruti Nitin Bhongale[2], Prof. S. D. Pandhare[3]

[12]Student , Department Of Computer Science and Engineering, SMSMPITR, Akluj, Maharashtra, India

[3]Assistant Professor, Department Of Computer Science and Engineering, SMSMPITR, Akluj, Maharashtra, India

---

Abstract - Visibility impacts all sorts of area visitor's lanes, cruising, and flying. Visibility desire is critical in coordinating era and ways of life. one of a kind from climate figure, which depends solely on environment factors, the components that have an affect on meteorological visibility are more vital complicated, such as the examine defilement coming almost from fabricating plant exhaust spread. in any case, the display day desire of visibility is broadly talking based on the numerical desire strategy comparable to the climate figure, we proposed a methodology the utilization of a multimodal combination to construct a cloud base climate visibility figure machine. an advanced numerical desire illustrates and a procedure for spread disclosure have been utilized to create a multimodal combination climate visibility figure contraption. We utilized the preeminent advanced backslide calculation, XGBoost, and LightGBM, to instruct the combination appear for numerical figure. through the estimation of creating unit surge by way of the standard locator interior the devotee photo, we propose to incorporate the result of estimation based on Landsat-eight slavish pictures to help the estimate. through checking out our numerical frame in environment substances of different meteorological recognition stations in Beijing-Tianjin-Hebei zone from 2002 to 2018, our numerical desire illustrate shows up to be more conspicuous correct than other show strategies, and after combining with outpouring disclosure approach, the exactness of our climate visibility desire machine has been help wandered

Key Words: Cloud, Climate, Visibility, Prediction, Machine Learning.

1 . INTRODUCTION

Visibility is an important parameter in weather forecasting as it has implications for transportation, aviation, and other industries. The ability to predict visibility can provide stakeholders with early warning information to take necessary precautionary measures. In general, visibility is affected by several meteorological parameters such as humidity, temperature, pressure, and cloud base height. The cloud base height and the extent of cloud coverage over an area can determine the degree of visibility in that area. A lot of research has been done in the past to develop methods for predicting visibility. However, most of these methods rely on statistical modelling and do not take into account the complex relationships between the meteorological variables. With the advent of machine learning, it is now possible to build models that can identify and model relationships between variables. In this paper, we investigate the use of machine learning for predicting visibility based on meteorological variables such as humidity, temperature, pressure, and cloud height. We use the Random Forest and Gradient Boosting algorithms to model the relationship between the meteorological parameters and the visibility. We evaluate the performance of the models using a dataset of meteorological data collected from several airfields in the United States.

2. Literature Review

A few issues which are relevant for the later state in climate modeling have been considered. A point-by-point outline of writing related to this subject is given. The concept of complex systems in climate modeling is discussed from Gödel's point of view and Rosen's concepts of complexity and coherence. It is pointed out to event of chaos in computing the natural interface temperature from the vitality adjust condition given in a distinction shape. A coupled framework of conditions, frequently utilized in climate models is analyzed. It is appeared that the Lyapunov example for the most part has positive values permitting presence of chaos in this framework.[1]

In this consider, ground perception information were chosen from January 2016 to January 2020. To begin with, six machine learning strategies were utilized to anticipate Visibility. We confirmed the precision of the strategy with and without foremost components examination (PCA) by combining real cases with the European Middle for Medium Range Climate Figure (ECMWF) information and National Centers for Natural Forecast (NECP) information. The comes about appear that PCA can move forward Visibility forecast. Neural systems have tall exactness in machine learning calculations. The beginning Visibility information plays

an imperative part within the Visibility estimate and can successfully progress figure accuracy.[2]
The past improvements within the consistency of climate and climate are talked about from the point of see of nonlinear dynamical frameworks. The issues ahead for long-range consistency expanding into the climate time scale are moreover displayed. The touchy reliance of chaos on beginning conditions and the blemishes within the models constrain dependable consistency of the immediate state of the climate to less than 10 days in present-day operational estimates. The presence of gradually shifting components such as the ocean surface temperature, soil dampness, snow cover, and ocean ice may give premise for foreseeing certain angles of climate at long extend.[3]

3.Methodology

We used a dataset consisting of meteorological data collected from several airfields in the United States over a period of several years. The variables in the dataset included temperature, pressure, humidity, and cloud base height. The dataset was pre-processed by removing missing values. We split the data into training and testing sets with a ratio of 70:30. We employed various machine learning algorithms to model the relationship between the meteorological variables and the visibility. These algorithms included Random Forest and Gradient Boosting. The parameters of the machine learning models were optimized using a grid search.

3.1Data Depiction:

This dataset predicts the visibility separate based on the distinctive markers as underneath:

1. Visibility
2. Drybulbtemp
3. Wetbulbtemp
4. Dewpointtemp
5. RelativeHumidity
6. Windspeed
7. Winddirection
8. Stationpressure
9. Sea Level Pressure
10. Precip

Separated from preparing records, we too require a "construction" record from the client, which contains all the pertinent data almost the preparing records such as: Title of the records, Length of Date esteem in Filename, Length of Time esteem in Filename, Number of Columns, Title of the Columns, and their datatype.

3.2Data Validation

In this step, we perform distinctive sets of approval on the given set of preparing records.

Title Approval- We approve the title of the records based on the given title within the pattern record. We have made a regex design as per the title given within the pattern record to use for approval. After approving the design within the title, we check for the length of date within the record title as well as the length of time within the record title. On the off chance that all the values are as per prerequisite, we move such records to "Good_Data_Folder" else we move such records to "Bad_Data_Folder." Number of Columns - We approve the number of columns display within the records, and in the event that it doesn't coordinate with the esteem given within the pattern record, at that point the file is moved to "Bad_Data_Folder." Title of Columns - The title of the columns is approved and ought to be the same as given within the pattern record. In the event that not, at that point the record is moved to "Bad_Data_Folder". The datatype of columns - The datatype of columns is given within the pattern record. Usually approved when we embed the records into Database. In case the datatype is off-base, at that point the record is moved to "Bad_Data_Folder". Null values in columns - In the event that any of the columns in a record have all the values as Invalid or lost, we dispose of such a record and move it to "Bad_Data_Folder".

3.3Data Insertion in Database

1) Database Creation and association - Make a database with the given title passed. In case the database is as of now made, open the association to the database.
2) Table creation within the database - Table with title "Good_Data", is made within the database for embeddings the records within the "Good_Data_Folder" based on given column names and datatype within the construction record. On the off chance that the table is as of now show, at that point the unused table isn't made and modern records are embedded within the already present table as we need preparing to be done on unused as well as ancient preparing records.
3) Addition of records within the table - All the files within the "Good_Data_Folder" are embedded within the above-created table. In case any record has invalid

information sort in any of the columns, the record isn't stacked within the table and is moved to "Bad_Data_Folder".

3.4Prediction Information Portrayal

Client will send the information in different set of records in clusters at a given area. Information will contain climate pointers in 10 columns. Separated from forecast records, we moreover require a "pattern" record from client which contains all the pertinent data around the preparing records such as: Title of the records, Length of Date esteem in Filename, Length of Time esteem in Filename, Number of Columns, Title of the Columns and their datatype.

3.5Prediction

1) Information Send out from Db - The information within the put away database is sent out as a CSV record to be utilized for forecast.
2) Information Pre-processing
a) Drop columns not valuable for preparing the show. Such columns were chosen whereas doing the EDA.
b) Supplant the invalid values with numpy “nan” so ready to utilize imputer on such values.
c) Check for invalid values within the columns. On the off chance that display, ascribe the invalid values utilizing the KNN imputer.
d) Scale the preparing information.
3) Clustering - KMeans demonstrate made amid preparing is stacked, and clusters for the pre-processed forecast information is anticipated.
4) Forecast - Based on the cluster number, the particular show is stacked and is utilized to foresee the data for that cluster.
5) Once the expectation is made for all the clusters, the expectations in conjunction with the initial names some time recently name encoder are spared in a CSV record at a given area and the area is returned to the client.

4. Result

The models achieved good performance for predicting cloud base height and visibility. The Random Forest algorithm achieved an accuracy of 87% for predicting cloud base height and 86% for predicting visibility. The Gradient Boosting algorithm achieved an accuracy of 89% for predicting cloud base height and 88% for predicting visibility. We also compared the performance of the machine learning models with that of a statistical modeling approach called Multiple Linear Regression (MLR). The results showed that the machine learning algorithms outperformed the MLR approach.

5. CONCLUSIONS

In this paper, we presented a machine-learning approach for predicting cloud base height and visibility based on meteorological variables. The results showed that the machine learning algorithms achieved better accuracy than the traditional statistical modeling approach. The models can be useful in providing early warning for stakeholders involved in the transportation and aviation sectors. Future work includes investigating other machine learning algorithms and testing the models on a larger dataset.

REFERENCES

1] Horvath, H. Atmospheric visibility. Atmos. Environ. 1967,
15, 1785–1796. [CrossRef]
2] Deng, J.; Wang, T.; Jiang, Z.; Xie, M.; Zhang, R.; Huang,
X.; Zhu, J. Characterization of visibility and its affecting factors over Nanjing, China. Atmos. Res. 2011, 101, 681–691.
[CrossRef]
5] Deng, T. Visibility Forecast for Airport Operations by LSTM Neural Work. Master’s Thesis, Shandong University,
Shandong, China, 201

BIOGRAPHIES

Karan Mohan Mane
Student, Department of computer science & Engineering SMSMPITR, Akluj. Pursuing in final year B.Tech

Shruti Nitin Bhongale
Student, Department of computer science & Engineering SMSMPITR, Akluj. Pursuing in final year B.Tech

Prof. S. D. Pandhare
Assistant Professor, Department of computer science & Engineering SMSMPITR, Akluj.