WEATHER FORECASTING PREDICTION

Micro Project for Practical Machine Learning

by

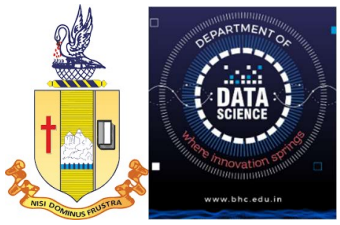
Asha Belcilda P

225229104

Submitted To

Dr. K. RAJKUMAR

Course Instructor



DEPARTMENT OF DATA SCIENCE

BISHOP HEBER COLLEGE (AUTONOMOUS)

TIRUCHIRAPPALLI 620017

MARCH 2023

CERTIFICATE

I hereby acknowledge that this project is the original work done by me for the requirements for Micro Project in Practical Machine Learning Course. This micro project is not copied from internet or whatsoever.

Tiruchirappalli
20 March 2023

Asha Belcilda P
Your Name and signature

Table of Contents

Chapter 1. ABSTRACT

Weather forecasting is the prediction of the state of the atmosphere for a given location using the Application of science and technology. This includes temperature, rain, cloudiness, wind speed, and Humidity. Weather forecasts are often made by collecting scientific quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve in future. Forecasting could be applied in air traffic, severe weather alerts, marine, agriculture, utility companies, private sector and military application. Weather forecasting is a complex and challenging science that depends on the efficient interplay of weather observation, data analysis by meteorologist and computers, and rapid and rapid communication system.

Chapter 2. Background and Motivation

Weather forecasting prediction is the process of using scientific methods and mathematical models to predict atmospheric conditions and weather patterns. This process is essential for providing accurate and timely information to people, businesses, and governments, allowing them to make informed decisions and take appropriate action in response to changing weather conditions.

The motivation for weather forecasting prediction is multifaceted. One of the primary reasons is to protect public safety and prevent damage to property and infrastructure. Severe weather events such as hurricanes, tornadoes, and floods can cause significant harm and devastation, and accurate forecasting can help people prepare and evacuate in advance, reducing the risk of injury and loss of life.

Chapter3. Problem Statement and Dataset Description

The problem statement for weather forecasting using machine learning is to accurately predict weather conditions based on current and historical meteorological data. The goal is to improve the accuracy and reliability of weather forecasts, which can have significant implications for public safety, business operations, and environmental protection.

In this project, we are concentrating on the temperature prediction of Kanpur city with the help of various machine learning algorithms and various regressions. By applying various regressions on the historical weather dataset of Kanpur city we are predicting the temperature like first we are applying Multiple Linear regression, then Decision Tree regression, and after that, we are applying Random Forest Regression.

Chapter 4. Existing Methodology

XGBoost (eXtreme Gradient Boosting) is a popular machine learning algorithm that can be used for regression and classification tasks. XGBoost is an ensemble learning algorithm that combines multiple weak models to create a stronger model.
Here is a high-level overview of how XGBoost can be used for weather forecasting prediction:

Data Preprocessing: The first step is to gather and preprocess the weather data. This may involve collecting data from various sources such as weather stations, satellites, and radar data. The data may also need to be cleaned, transformed, and normalized to prepare it for training.

Feature Selection: Once the data is preprocessed, the next step is to select the most relevant features for the model. This can be done using techniques such as feature importance ranking or correlation analysis.

Training: After the data is preprocessed and the features are selected, the XGBoost model can be trained on the data. During training, XGBoost iteratively builds a set of decision trees to predict the target variable (e.g., temperature). The algorithm uses gradient boosting to minimize the errors of each decision tree and improve the overall prediction accuracy.

Evaluation: Once the model is trained, it needs to be evaluated to ensure it is accurate and reliable. This can be done by measuring metrics such as mean absolute error (MAE) or mean squared error (MSE) on a test dataset.

Prediction: Finally, the trained model can be used to make weather forecasting predictions for new data. The model takes in the relevant weather data as input and generates a predicted value for the target variable (e.g., temperature).

XGBoost has been shown to be an effective method for weather forecasting prediction, particularly for short-term forecasts. However, as with any machine learning model, XGBoost has its limitations and may not always be the best choice for every weather forecasting problem. Therefore, it is important to carefully consider the specific requirements of the problem and explore other methods and algorithms as needed.

Chapter 5. Proposed Methodology and Solution

Random Forest Regression is another popular machine learning algorithm that can be used for weather forecasting prediction. Random forest regression is an ensemble learning algorithm that combines multiple decision trees to make predictions.

Here is a proposed methodology for using Random Forest Regression for weather forecasting:

Data Preprocessing: The first step is to gather and preprocess the weather data. This may involve collecting data from various sources such as weather stations, satellites, and radar data. The data may also need to be cleaned, transformed, and normalized to prepare it for training.

Feature Selection: Once the data is preprocessed, the next step is to select the most relevant features for the model. This can be done using techniques such as feature importance ranking or correlation analysis.

Training: After the data is preprocessed and the features are selected, the Random Forest Regression model can be trained on the data. During training, the algorithm builds a set of decision trees to predict the target variable (e.g., temperature). The algorithm uses bagging to create multiple random subsets of the training data and trains a decision tree

on each subset. The final prediction is then generated by averaging the predictions of all the decision trees.

Evaluation: Once the model is trained, it needs to be evaluated to ensure it is accurate and reliable. This can be done by measuring metrics such as mean absolute error (MAE) or mean squared error (MSE) on a test dataset.

Prediction: Finally, the trained model can be used to make weather forecasting predictions for new data. The model takes in the relevant weather data as input and generates a predicted value for the target variable (e.g., temperature).

Chapter 6. Model Implementation

import pandas as pd

Pandas is a Python library used for working with data sets. It has functions for analysing, cleaning, exploring, and manipulating**.**

from sklearn.model_selection import train_test_split

Splitting your dataset is essential for an unbiased evaluation of prediction performance. In most cases, it's enough to split your dataset randomly into three subsets.(1.Training, 2.Validation, 3.Test).

From sklearn.model_selection import cross_val_score

Cross_val_score is a function in the scikit-learn package which trains and tests a model over multiple folds of your dataset. This cross validation method gives you a better understanding of model performance over the whole dataset instead of just a single train/test split**.**

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix

These performance metrics include accuracy, precision, recall, and F1-score. Because it helps us understand the strengths and limitations of these models when making predictions in new situations, model performance is essential for machine learning**.**

from sklearn.preprocessing import StandardScaler

StandardScaler removes the mean and scales the data to the unit variance. However, outliers have an influence when calculating the empirical mean and standard deviation, which narrows the range of characteristic values.

Chapter 7. Testing and Evaluation

Random Forest Regression has been shown to be an effective method for weather forecasting prediction, particularly for short to medium-term forecasts. However, as with any machine learning model, Random Forest Regression has its limitations and may not always be the best choice for every weather forecasting problem. Therefore, it is important to carefully consider the specific requirements of the problem and explore other methods and algorithms as needed.

Mean absolute error: 0.47
Residual sum of squares (MSE): 0.63
R2-score: 0.99

Chapter8. Model Archival in github and Demo in Youtube

Details of code repository in Github:

https://github.com/ashabelcilda/Weather-Forecasting.git

Youtube Video Link:

https://youtu.be/eJTM3lImems

Chapter 9. Conclusion and Future Work

All the machine learning models: linear regression, various linear regression, decision tree regression, random forest regression were beaten by expert climate determining apparatuses, even though the error in their execution reduced significantly for later days, demonstrating that over longer timeframes, our models may beat genius professional ones. Linear regression demonstrated to be a low predisposition, high fluctuation model though polynomial regression demonstrated to be a high predisposition, low difference model. Linear regression is naturally a high difference model as it is unsteady to outliers, so one approach to improve the linear regression model is by gathering more information. Talking about Random Forest Regression, it proves to be the most accurate regression model. Likely so, it is the most popular regression model used, since it is highly accurate and versatile. Below is a snapshot of the implementation of Random Forest in the project. Weather Forecasting has a major test of foreseeing the precise outcomes which are utilized in numerous ongoing frameworks like power offices, air terminals, the travel industry focuses, and so forth. The trouble of this determining is the mind-boggling nature of parameters. Every parameter has an alternate arrangement of scopes of qualities.

Chapter 10.References

https://www.kaggle.com/code/kalsangsherpa100/predicting-climate-using-xgbregressor