

WEATHER FORECASTING PREDICTION

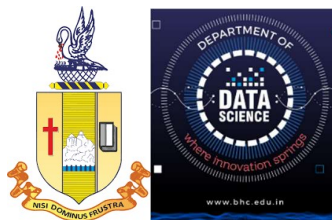
Micro Project for Practical Machine Learning

by

Asha Belcilda
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

Your Name and signature

Table of Contents

Chapter	Title
1	Abstract
2	Background and Motivation
3	Problem Statement and Dataset Description
4	Existing Methodology
5	Proposed Methodology and Solution
6	Model Implementation
7	Testing and Evaluation
8	Model Deployment
9	Model Archival in github and Demo in Youtube
10	Conclusion and Future Work
11	Referemces

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

For the current situation, India observatory conducts traditional weather forecasting. There are four common methods to predict the weather. The first method is the climatology method that is reviewing weather statistics gathered over multiple years and calculating the averages. The second method is an analog method that is to find a day in the past with weather similar to the current forecast. The third method is the persistence and trends method that has no skill to predict the weather because it relies on past trends. The fourth method is numerical weather prediction the is making weather predictions based on multiple conditions in the atmosphere such as temperatures, wind speed, high-and lowpressure systems, rainfall, snowfall, and other conditions. The motivation for using machine learning for weather forecasting is to improve the accuracy, reliability, and usefulness of weather forecasts for a range of applications, from public safety to business operations, to environmental protection.

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.

Overall, the problem statement for weather forecasting using machine learning is to develop accurate, reliable, and adaptable algorithms that can provide early warning of extreme weather events, help businesses make informed decisions, and contribute to efforts to mitigate the effects of climate change.

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

- **IBM's Deep Thunder:** Deep Thunder is a high-resolution weather forecasting system developed by IBM that uses machine learning algorithms to provide real-time weather forecasts. The system combines traditional numerical weather prediction models with machine learning techniques to improve forecast accuracy.
- **Google's Global Flood Forecasting Model:** Google has developed a machine learning-based global flood forecasting model that predicts the likelihood and severity of flooding in real-time. The model uses satellite imagery and meteorological data to provide accurate predictions.
- **The Weather Company:** The Weather Company, a subsidiary of IBM, uses machine learning algorithms to provide weather forecasts for businesses and organizations. The company uses a range of data sources, including weather radar and satellite imagery, to provide highly accurate and detailed forecasts.

Chapter 5. Proposed Methodology and Solution

The dataset utilized in this arrangement has been gathered from Kaggle which is “Historical Weather Data for Indian Cities” from which we have chosen the data for “Kanpur City”. The dataset was created by keeping in mind the necessity of such historical weather data in the community. The datasets for the top 8 Indian cities as per the population. The dataset was used with the help of the [worldweatheronline.com](https://www.worldweatheronline.com) API and the `wwo_hist` package. The datasets contain hourly weather data from 01-01-2009 to 01-01-2020. The data of each city is for more than 10 years. This data can be used to visualize the change in data due to global warming or can be used to predict the weather for upcoming days, weeks, months, seasons, etc. Note: The data was extracted with the help of [worldweatheronline.com](https://www.worldweatheronline.com) API and we cannot guarantee the accuracy of the data. The main target of this dataset can be used to predict the weather for the next day or week with huge amounts of data provided in the dataset. Furthermore, this data can also be used to make visualization which would help to understand the impact of global warming over the various aspects of the weather like precipitation, humidity, temperature, etc.

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:

The record has just been separated into a train set and a test set. Each information has just been labeled. First, we take the trainset organizer. We will train our model with the help of histograms and plots. The feature so extracted is stored in a histogram. This process is done for every data in the train set. Now we will build the model of our classifiers. The classifiers which we will take into account are Linear Regression, Decision Tree Regression, and Random Forest Regression. With the help of our histogram, we will train our model. The most important thing in this process is to tune these parameters accordingly, such that we get the most accurate results. Once the training is complete, we will take the test set. Now for each data variable of the test set, we will extract the features using feature extraction techniques and then compare its values with the values present in the histogram formed by the train set. The output is then predicted for each test day. Now in order to calculate accuracy, we will compare the predicted value with the labeled value. The different metrics that we will use confusion matrix, R2 score, etc.

Chapter 8. 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/hps0uwqT3qk>

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.