WEATHER PREDICTION USING MACHINE LEARNING

Problem Description:

The goal is to predict weather conditions based on historical data and various meteorological features.

This involves forecasting weather parameters such as temperature, humidity, precipitation, wind speed, and other relevant factors like atmospheric pressure, cloud cover, etc. The aim is to build a model that can accurately predict these weather conditions for a given location and time.

The prediction can help in providing accurate weather forecasts, which is crucial for various applications such as agriculture, transportation, water management and disaster management.

Type of Problem:

Weather prediction can be framed as a **Regression** problem as we aim to predict numerical values such as min. temperature, max. temperature, humidity, pressure....

For a regression problem like weather prediction, labelled data with actual weather measurements is crucial for training the model.

Supervised learning is a technique where a model learns from labelled data to make predictions or classifications. Supervised learning can be used for regression tasks where the output is a continuous variable, or classification tasks, where the output is a discrete variable.

In the context of weather and climate forecasting, supervised learning can be used for tasks such as predicting the temperature or precipitation for a given location and time.

Machine Learning Algorithm can be broadly classified into three types:

Supervised Learning Algorithms

Unsupervised Learning Algorithms

Reinforcement Learning algorithm.

There are various machine learning algorithms that can be used for the prediction such as like Linear regression, Decision Tree, Random Forest, Support Vector Machines, K-Nearest Neighbour, K-Means Clustering, Logistic regression.

Linear regression is used for predicting continuous variables like temperature or humidity.

Linear regression is divided into two types.

Simple Linear Regression, Multiple Linear Regression

Simple Linear Regression - a single independent variable is used to predict the value of the dependent variable.

Multiple Linear Regression - more than one independent variables are used to predict the value of the dependent variable.

Data Sources:

For this, various sources of data can be considered:

Historical weather data from meteorological agencies or weather stations.

Satellite imagery data for cloud cover, temperature distribution, etc.

IoT devices providing real-time weather updates.

Weather APIs that offer historical and current weather data.

The Data used here is obtained from an open source Kaggle – Historical Weather Data for Indian Cities.

The Dataset can be found [here](https://www.kaggle.com/datasets/hiteshsoneji/historical-weather-data-for-indian-cities).

Data covers a significant period from 2009 – 2020 which contains hourly weather data to ensure completeness. 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.

Displaying top 5 rows of the dataset

A screenshot of a computer

Description automatically generated

The describe() method returns description of the data in the DataFrame.

A screenshot of a computer screen

Description automatically generated

Checking if there are any null values in the dataset.

A black and white screen with white text

Description automatically generated

A screen shot of a computer

Description automatically generated

Checking the datatypes of the data

A screen shot of a computer

Description automatically generated

Number of rows and columns in the data.

A screen shot of a computer program

Description automatically generated

Obtaining the required fields

A screenshot of a computer

Description automatically generated

New Columns

A screen shot of a computer

Description automatically generated

Plotting the columns

A screenshot of a computer screen

Description automatically generated

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python.

From scikit-learn(sklearn) library, library linear regression, decision tree regression, random forest regression models can be used for the prediction. L Linear Regression method can be used to create linear regression object. Using fit() method can be used to train the model and predict() method can be used to make new predictions on the new data.  train\_test\_split() module of scikit-learn can be used for splitting the data into a train set and a test set.