Seattle Weather Prediction Using Regression Analysis

- **Chitti babu Yelugubandi**

Title:

Seattle Weather Prediction Using Regression Analysis

Aim:

To develop a predictive model for forecasting daily maximum temperatures in Seattle using historical weather data and regression analysis implemented in R.

Objective:

The project aims to create a reliable model to predict the maximum daily temperature (temp_max) in Seattle. By analyzing historical weather data, features such as minimum temperature (temp_min), precipitation, wind speed (wind), and weather type (weather) are used to enhance prediction accuracy. The tasks involve data preprocessing, exploratory data analysis (EDA), model training, and evaluation using metrics like RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error). The ultimate goal is to support weather-related decisions through the predictive model.

Abstract:

Weather prediction plays a vital role in numerous fields. This project focuses on predicting Seattle's daily maximum temperature using regression analysis implemented in **R programming language**. Historical weather data is analyzed to uncover relationships between temp_max and features such as temp_min, precipitation, wind speed, and weather type.

The project involves data cleaning and preparation, including handling missing values and normalizing the data. Exploratory Data Analysis (EDA) is carried out to explore trends and correlations. A regression model is trained using R's statistical and machine learning packages, and its performance is evaluated using RMSE and MAPE. The final model is validated on unseen data to ensure its reliability for forecasting future maximum temperatures.

Features Analyzed:

1. **Minimum Temperature (temp_min)**: The lowest daily temperature.
2. **Precipitation**: The amount of rainfall or snowfall.
3. **Wind Speed (wind)**: Daily average wind speed.
4. **Weather Type (weather)**: Categorical classification of weather conditions.

Methodology:

1. **Data Collection:**

- Use historical weather datasets specific to Seattle.

2. **Data Preprocessing:**
    - Handle missing values, normalize data, and encode categorical variables (e.g., weather).
3. **Exploratory Data Analysis (EDA):**
    - Use R's visualization packages to explore data distributions and correlations.
4. **Model Training:**
    - Fit regression models (lm or advanced machine learning methods like randomForest or xgboost) for predicting temp_max.
5. **Model Evaluation:**
    - Evaluate the model using RMSE and MAPE to validate accuracy.
6. **Prediction:**
    - Forecast future maximum temperatures using the trained model.

Tools Used:

1. **Programming Language:** R
2. **Libraries:**
    - dplyr for data manipulation
    - ggplot2 for visualizations
    - caret for model training and evaluation
    - Metrics for calculating RMSE and MAPE
3. **Environment:** RStudio

Evaluation Metrics:

1. **RMSE (Root Mean Squared Error):** Measures prediction accuracy by computing the average squared difference between actual and predicted values.
2. **MAPE (Mean Absolute Percentage Error):** Provides an interpretable error measure in percentage form.

Output:

- A regression model trained to predict maximum daily temperatures.
- Visualizations showing relationships between features and target variables.
- RMSE and MAPE metrics to evaluate model accuracy.

- Predicted future temperatures based on new input data.

Conclusion:

Using regression analysis in R, this project successfully forecasts Seattle's maximum daily temperature. The resulting model provides accurate predictions and can be a valuable tool for weather-related decision-making. The approach demonstrated here can also be adapted for predicting other weather variables or applied to different geographic regions.

References:

- R Documentation: https://www.r-project.org
- CRAN Packages: dplyr, ggplot2, caret, Metrics
- Historical Weather Data Repository (e.g., NOAA)