RainSenseAI

Harsh Mithaiwala
harsh.mithaiwala.hm@gmail.com

Raveena Choudhary
raveena.choudhary@gmail.com

Darshakkumar Kachchhi
darshak.kachchhi@gmail.com

Sumit Monapara
sumitmonapara@gmail.com

I. Problem Statement and Application

Problem Statement: Despite the evident importance of rainfall prediction, achieving a high level of accuracy remains a complex challenge. Australia's vast geographical expanse, diverse climatic zones, and unique weather patterns pose significant hurdles in accurately forecasting rainfall events. This poses limitations on the ability of various sectors to plan effectively, allocate resources efficiently, and mitigate potential risks associated with extreme weather conditions. [1]

Application: The application of rainfall prediction spans a wide range of sectors and industries, playing a crucial role in decision-making and resource management.

1) Agriculture
2) Water Resource Management
3) Urban Planning and Infrastructure
4) Disaster Management
5) Energy Sector

Challenges: Below are the challenges associated with our selected problem application:

- **Spatial and Temporal Coverage:** Ensuring an adequate spatial and temporal coverage of data is crucial to capture the diversity of rainfall patterns. Lack of data from certain regions or time periods may lead to biased or incomplete models, affecting the accuracy of predictions.
- **Preprocessing of data**: identifying the missing values, converting strings to relevant numbers, and also, Identifying the most relevant features and transforming them into suitable representations is critical for accurate predictions.
- **Model selection**: Choosing the right classification model, and appropriate evaluation metrics, such as precision, recall, or F1-score, is necessary to assess the model's performance, and validation methods for rain prediction are crucial.

Goal: Our goal is to develop an accurate rainfall forecast model using key features like temperature, humidity, wind speed, and atmospheric pressure. We aim to uncover patterns and correlations that contribute to precise rain prediction through advanced data exploration, feature engineering, and machine learning algorithms.

II. Dataset Selection

We have selected the "Rain in Australia" dataset, sourced from Kaggle, for our analysis. [2]

Parameters	Statistics
Total Records	145460
Total Features	23
Date Field	1
Text Fields	10
Numeric Fields	12
Missing values	Yes

III. Possible Methodology

Pre-requisite step for all the models:

Data Pre-processing: To accurately predict the possibility of rain, first, we will determine the correlation of features. Understanding the correlations will enable us to identify the most influential features and effectively utilize them for rain prediction models.
We are planning to preprocess the data by replacing missing values using skewness-based imputation and converting text and date to numeric constants.

Feature Extraction: We are planning to transform raw data into a more meaningful representation capturing the relevant information.

k-cross validation:: We plan to use this technique to evaluate and compare the performance of the models, aiding in the selection of the best model configuration or hyper-parameters, and providing a more robust assessment of the model's generalization performance. [3]

Visualization: To visualize the model, we are planning to use t-SNE (t-Distributed Stochastic Neighbor Embedding).

Accuracy score: To evaluate the accuracy of the model, we are thinking to use any one of these accuracy scores: **F1, RMSC(Root Mean Square Error) scores**, which will help us to predict the accuracy of the application using models such as Decision Tree supervised, Decision Tree semi-supervised, and DNN supervised. []

IV. Expected Results:

The following are the anticipated results of this study:

- Determining the best categorization models for correctly predicting rainfall.
- An in-depth understanding of the compromises among computing complexity, forecasting precision, and model complexity in the prediction of rainfall.
- This research has the potential to support data-driven decision-making across a variety of sectors by producing precise and accurate rainfall predictions.

REFERENCES

[1] A. S. Cabezuelo, "Prediction of rainfall in australia using machine learning," 2022.

[2] "Rain in australia (dataset)." https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package, 2018.

[3] "Rain in australia (dataset)." http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml, 2018.