PREDICTA 1.0
PROBLEMS AND GUIDELINES

~Organized by~

IEEE Student Branch
University of Peradeniya

Table of Contents

1. General Information

Competition Name: Predicta 1.0

Competition Dates: June 21, 2024, to June 23, 2024

Venue: The competition is held completely Online

Platform: Kaggle (for submission and evaluation)

Objective: To develop a machine learning model to predict the evolution of a given real-world dataset.

2. Problem 1: Predict the Unpredictable ([Click Here](#))

Problem Description

Accurate weather prediction is essential for planning and decision-making in various fields such as agriculture, disaster management, energy supply, and transportation. In this challenge, participants are tasked with predicting the average temperatures for the first week of 2019 across 100 cities worldwide, using historical weather data from January 1, 2014, to December 31, 2018. This challenge emphasizes the importance of precise weather forecasting in improving safety, productivity, and overall quality of life.

Data Description

historical_weather.csv

- **Columns:**
    - `city_id`: Unique identifier for each city.
    - `date`: Date of the weather observation.
    - `avg_temp_c`: Average temperature in Celsius.
    - `min_temp_c`: Minimum temperature in Celsius.
    - `max_temp_c`: Maximum temperature in Celsius.
    - `precipitation_mm`: Precipitation in millimeters.
    - `snow_depth_mm`: Snow depth in millimeters.
    - `avg_wind_dir_deg`: Average wind direction in degrees.
    - `avg_wind_speed_kmh`: Average wind speed in kilometers per hour.

sample_submission.csv

- **Columns:**
    - `submission_ID`: Identifier for each prediction.
    - `avg_temp_c`: Column for participants to enter their predicted average temperatures.

submission_key.csv

- **Columns:**
    - `submission_ID`: Identifier for each prediction.
    - `city_id`: Identifier for each city.
    - `date`: Date for which the prediction is made.

Evaluation

Participants' predictions will be evaluated based on the Root Mean Squared Error (RMSE) metric, which measures the difference between the predicted and actual average temperatures.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

Where:

- y_i is the actual average temperature.
- $\hat{y}_i$ is the predicted average temperature.
- N is the total number of predictions.

3. Problem 2: Classify the Weather (Click Here)

Problem Description

Classifying daily weather conditions accurately is critical for public safety, agriculture, transportation, event planning, and health management. This challenge involves predicting the weather condition for specific days and cities using historical weather data. Participants will classify daily weather conditions into one of nine categories, such as "Clear and Sunny," "Partly Cloudy," "Light Precipitation," and others. By improving weather classification models, we can better understand and predict weather patterns, leading to enhanced preparedness and response strategies.

Data Description

daily_data.csv

- **Columns:**
    - `day_id`: Unique identifier for each day.
    - `city_id`: Identifier for each city.
    - `temperature_celsius`: Temperature in Celsius.
    - `condition_text`: Text description of the weather condition (target variable).
    - `wind_kph`: Wind speed in kilometers per hour.
    - `wind_degree`: Wind direction in degrees.
    - `pressure_mb`: Atmospheric pressure in millibars.
    - `precip_mm`: Precipitation in millimeters.
    - `humidity`: Humidity percentage.
    - `cloud`: Cloud cover percentage.
    - `feels_like_celsius`: Feels-like temperature in Celsius.
    - `visibility_km`: Visibility in kilometers.
    - `uv_index`: UV index.
    - `gust_kph`: Wind gust speed in kilometers per hour.
    - `air_quality_us-epa-index`: Air quality index based on US EPA standards.
    - `sunrise`: Time of sunrise.
    - `sunset`: Time of sunset.

submission.csv

- **Columns:**
    - `day_id`: Identifier for each day.
    - `condition_text`: Column where participants will enter their predicted weather conditions.

Weather Condition Types

Participants have to classify the weather conditions into one of the following nine types:

1. Clear and Sunny
2. Partly Cloudy
3. Light Precipitation
4. Cloudy and Overcast
5. Mist or Fog

6. Rain Showers
7. Light Rain with Thunder
8. Thunderstorms
9. Moderate to Heavy Rain

Evaluation

Participants' submissions will be evaluated based on accuracy, which measures the proportion of correctly predicted weather conditions out of the total predictions made.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

4. Report Submission Guidelines

Report Format (Report Format)

- **Length**: Maximum 12 pages. (10 excluding the cover page and the contents page)
- **Font**: Times New Roman, 11pt.
- **Spacing**: 1.15 line spacing.
- **Margins**: 1-inch margins on all sides.
- **Format**: PDF file.
- **File Name**: `Predicta_Report_PXXX.pdf`, where XXX is your unique registration number.
- **Submission Method**: Submit through the provided Google Form.
- **Deadline**: June 23, 2024, at 23:59 UTC+05:30.

5. Scoring and Final Submission

Scoring

- **Problem 1: Predict the Unpredictable**: 25 marks
- **Problem 2: Classify the Weather**: 25 marks
- **Report**: 50 marks

Final Submission

- **Predictions for Problem 1**: Submit `sample_submission.csv` with predicted average temperatures.
- **Predictions for Problem 2**: Submit `submission.csv` with predicted weather conditions.
- **Report**: Submit the report in PDF format as specified in the Report Submission Guidelines.

Ensure that all files are correctly named and submitted before the deadline.

.

6. Rules and Conduct

General Rules

1. **Original Work**: All submissions must be the original work of the participants. Plagiarism or use of unauthorized data will result in disqualification.
2. **Code Sharing**: Teams must share their code via a public GitHub repository. The repository link should be included in the report.
3. **Fair Play**: Participants must not engage in any activity that compromises the fairness of the competition, including but not limited to collusion, sabotage, or tampering with other participants' work.
4. **Communication**: All official communication will be conducted via the competition's official channels (Kaggle forum, official email, and official WhatsApp announcements group). Participants are expected to stay updated with any announcements or changes.

Participation

- **Eligibility**: Only registered teams are eligible to compete and should use the emails used in the registration.
- **Registration**: Participants must register on the competition's official webpage by June 16, 2024.
- **Team Composition**: Teams must be formed and registered before the competition begins. No changes to team composition are allowed after the competition starts.

Dataset

- **Dataset Access**: The dataset will be provided via the Kaggle competition page on June 21, 2024, at 08:00 UTC+5.30.

- **Usage**: Participants may only use the provided dataset for model training and prediction. External data is not allowed unless explicitly permitted by the competition organizers.

Model Development

- **Techniques**: Participants can use any machine learning or data analytic techniques to develop their models.
- **Software**: Any programming language or tool can be used, provided the final submission is compatible with the Kaggle platform.

Submission

Kaggle Submission

- Participants must submit their predictions on the Kaggle platform.
- Each team is allowed up to 3 submissions per day.
- The final submission deadline is June 23, 2024, at 23:59 UTC+05:30.

Important Submission Instructions:

- **Team Name Format:** Each team or individual participant must submit their answers after updating their team name to match the registration number provided by the Organizing Committee. This should be in the format 'PXXX', where XXX is your unique registration number. **Important:** Submitting without the correct team name will result in disqualification.
- **Team Creation and Joining:**
  - Team Leaders: **You must create the team on the Kaggle platform.**
  - Team Members: **Ensure you join the team created by your leader.**
- **Further Assistance:**
  - For detailed guidance on using Kaggle for team creation and submission, please refer to the Kaggle Help Page or contact us.

Report Submission

- A comprehensive report detailing the methodology, data preprocessing, feature selection, model training, and evaluation must be submitted.
- The report should be in PDF format and submitted through a Google Form provided by June 23, 2024, at 23:59 UTC.
- The report should follow the provided template and include sections: Summary, GitHub Repository, Data Cleaning and Preprocessing, Feature Selection and Engineering, Feature Scaling and Standardization, Model Selection and Training, Results and Conclusion, and References.

7. Prizes and Recognition

- Details of prizes will be announced on the competition's official webpage and digital media channels during the competition.
- Winners and notable participants will be featured on the competition's official webpage and other promotional materials.
- E-certificates will be awarded to all the participants.

8. Disqualification

Non-compliance

- Failure to comply with any of the rules and regulations will result in immediate disqualification.

Misconduct

- Any form of misconduct, including but not limited to cheating, will result in disqualification.