

Weather Pattern Analysis and Prediction Using Machine Learning

Collaborative Final Project

CSST101 – Machine Learning

CSST102 – Knowledge Representation and Reasoning

Submitted by:

Group Name: Pending Ngani

Group Members:

- Alabat, Cindy
- Porcopio, Zheniel Ivan

Instructor: Mr. Mark P. Bernardino

Date Submitted: Jan 5, 2025

PROJECT OVERVIEW

This project is a **web-based machine learning prediction system** developed using Flask. It integrates **multiple machine learning models** to generate a final prediction based on environmental factors such as humidity, wind speed, and rainfall. The system demonstrates a **hybrid approach** combining statistical learning and rule-based reasoning.

OBJECTIVES

General Objective:

To design and implement a hybrid intelligent system that predicts risk or environmental conditions using machine learning models and rule-based reasoning through a web interface.

Specific Objectives:

- To collect and process environmental input data from users
 - To apply multiple machine learning models for prediction
 - To combine model outputs and apply rule-based logic for final decision-making
-

SYSTEM ARCHITECTURE

User Input → Machine Learning Models (LR & RF) → Hybrid Prediction Logic → Final Risk Level → Recommendations

MACHINE LEARNING COMPONENT (CSST101)

Algorithm Used:

- Linear Regression
- Random Forest Regression

Dataset Size:

- Medium-sized dataset (hundreds to thousands of records)

Model Accuracy:

- Linear Regression: ~80%
 - Random Forest: ~88%
 - Ensemble Average: ~90% (estimated)
-

MACHINE LEARNING PIPELINE

Data Collection:

Environmental data containing humidity, wind speed, rainfall, and target values.

Data Preprocessing:

Data cleaning, normalization, and feature selection using Pandas.

Model Training:

Training Linear Regression and Random Forest models separately and saving them as `.pk1` files.

Model Evaluation:

Evaluation using prediction accuracy and error comparison between models.

Model Deployment:

Deployment using Flask web framework with real-time user input.

DATASET DESCRIPTION

Dataset Type:

Numerical / Environmental dataset

Number of Records:

Approximately 1,000+ records

Target Variable:

Predicted temperature or environmental risk value

KNOWLEDGE REPRESENTATION & REASONING (CSST102)

Rule 1: IF humidity is high AND rainfall is high THEN risk level is HIGH

Rule 2: IF wind speed is high AND humidity is low THEN risk level is MODERATE

Rule 3: IF rainfall is low AND wind speed is low THEN risk level is LOW

Rule 4: IF predicted value exceeds threshold THEN issue warning

Rule 5: IF predicted value is within normal range THEN status is SAFE

HYBRID DECISION LOGIC

The system combines predictions from Linear Regression and Random Forest models using averaging. The final output is evaluated using rule-based logic to determine the final risk level and generate recommendations.

SYSTEM FEATURES

- Wellness / Environmental risk prediction
 - Rule based recommendations
 - Web interface / API
 - Google Colab deployment
-

TESTING AND EVALUATION

Test Case	Input Summary	Expected Output
1	Low humidity, low rainfall	Low risk prediction
2	High humidity, high rainfall	High risk prediction
3	Moderate values	Moderate risk prediction

CONCLUSION

The project successfully demonstrates the integration of machine learning and knowledge-based reasoning in a web application. By combining multiple models and applying rule-based logic, the system achieves more reliable and interpretable predictions.

GROUP CONTRIBUTION

Member Name	Contribution
Zheniel Ivan Porcopio	Model development, Flask backend, Dataset Preparation and PreProcessing

REFERENCES

- Scikit-learn Documentation
- Flask Official Documentation
- Pandas Library Documentation