## Artificial Intelligence and

## Machine Learning

Project Report

Semester-IV (Batch-2022)

Title of the Project

**WEATHER PREDICTION**

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Faculty Name : Ms. Shagun Sharma Name: Harsh Bansal

Roll no.:2210990376

Group : G-9

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Abstract**

The project aims to develop a weather prediction system using machine learning techniques. The dataset used for this project consists of historical weather data from Seattle. The project involves exploratory data analysis, data preprocessing, model training, and evaluation. Various machine learning algorithms such as K-Nearest Neighbors, Support Vector Machine, Gradient Boosting Classifier, and XGBoost Classifier are implemented and evaluated for their performance in weather prediction. The trained model is saved for future use. The report provides a detailed overview of the project methodology, results, and conclusions.

**Table of Contents**

1. **Introduction**

1.1 Background

1.2 Objectives

1.3 Significance

1. **Problem Definition and Requirements**
2. **Proposed Design/Methodology**

3.1 Data Collection

3.2 Data Preprocessing

3.3 Model Training

3.4 Model Evaluation

1. **Results**

4.1 Exploratory Data Analysis

4.2 Model Performance

4.3 Prediction Results

1. **Conclusion**
2. **References**

**1. Introduction**

**1.1 Background**

Weather prediction plays a crucial role in various sectors such as agriculture, transportation, and disaster management. Accurate weather forecasting can help in making informed decisions and mitigating risks. Machine learning algorithms have shown promising results in weather prediction tasks by analyzing historical weather data and identifying patterns.

**1.2 Objectives**

The primary objective of this project is to develop a weather prediction system using machine learning techniques. Specific objectives include:

* Exploratory data analysis of historical weather data.
* Preprocessing the data to handle missing values and encode categorical variables.
* Training and evaluating machine learning models for weather prediction.
* Saving the trained model for future use.

**1.3 Significance**

The significance of this project lies in its potential to provide accurate weather forecasts, which can assist various industries and individuals in planning and decision-making. By leveraging machine learning techniques, the project aims to improve the accuracy and reliability of weather predictions.

**2. Problem Definition and Requirements**

The problem statement for this project involves predicting weather conditions (rain, sun, drizzle, snow, fog) based on features such as precipitation, temperature, and wind speed. The software requirements include Python programming language and libraries such as pandas, matplotlib, seaborn, scikit-learn, and xgboost. The hardware requirements are standard computer hardware with sufficient memory and processing power to handle the dataset and run machine learning algorithms.

**3. Proposed Design/Methodology**

**3.1 Data Collection**

The dataset used for this project is "seattle-weather.csv," which contains historical weather data from Seattle. The data includes features such as date, precipitation, temperature (max and min), wind speed, and weather condition.

**3.2 Data Preprocessing**

Data preprocessing steps include handling missing values, encoding categorical variables (weather condition), and splitting the dataset into training and testing sets.

**3.3 Model Training**

Four machine learning algorithms are trained for weather prediction: K-Nearest Neighbors, Support Vector Machine, Gradient Boosting Classifier, and XGBoost Classifier.

**3.4 Model Evaluation**

The trained models are evaluated on the testing set using accuracy as the evaluation metric. The performance of each model is compared, and the best-performing model is selected for further analysis.

**4. Results**

**4.1 Exploratory Data Analysis**

Exploratory data analysis includes visualizations such as count plots, histograms, violin plots, and correlation matrices to understand the relationships between weather features and weather conditions.

**4.2 Model Performance**

The performance of each machine learning model is evaluated using accuracy scores on the testing set. The results are presented in tabular format, showing the accuracy of each model.

**4.3 Prediction Results**

Sample predictions are made using the trained XGBoost model, and the predicted weather conditions are compared with the actual conditions.

**5. Conclusion**

The project successfully develops a weather prediction system using machine learning techniques. The trained model demonstrates good accuracy in predicting weather conditions based on historical data. The project highlights the importance of data preprocessing, feature engineering, and model selection in improving prediction performance.

**6. References**

Youtube

Kaggle for dataset