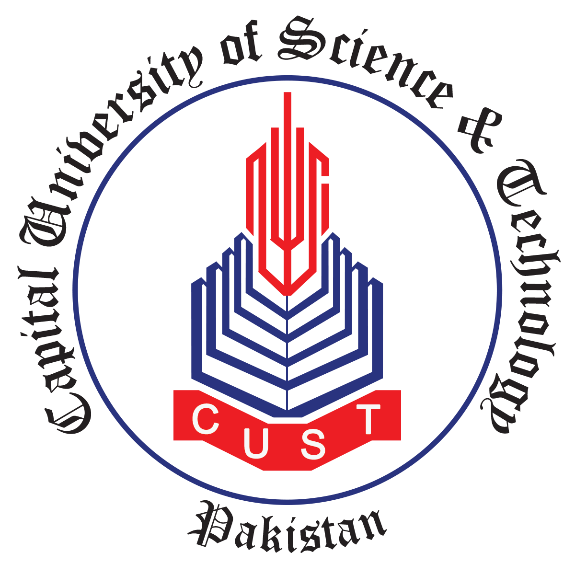
**PROJECT REPORT**

**THEORY OF PROGRAMMING LANGUAGES**



SUBMITTED BY:

Omer Tanveer

BCS213004

Muhammad Adil

BCS213002

**SUBMITTED TO:**

**Maam. Memoona Billal**

**DEPARTMENT OF COMPUTER SCIENCE**

**CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY**

# Rainfall Prediction with Support Vector Machines

## Abstract

Rainfall forecasting is crucial in island regions, especially those vulnerable to extreme climate events. We used Support Vector Machine (SVM) for predicting rainfall. Variables which are crucial were used for prediction. The SVM model exhibited a precision value of 82% for rainfall, with a ROC curve score of 0.74, indicating its ability to distinguish between positive and negative rainfall events.

## Introduction

Kepulauan Riau Province in Indonesia, made up of 96% ocean and over 2,400 islands, is prone to tropical storms. These storms can lead to heavy rain, strong winds, and high ocean waves, causing significant damage. Rainfall forecasting is crucial in island regions due to limited water resources and the reliance on rainfall for freshwater supply. Accurate forecasts are essential for water management, agriculture, and preparing for extreme weather events. Support Vector Machine (SVM), a superior model in predicting extreme rainfall, is used in this study. Unique geographical characteristics, such as ocean currents and mountain ranges, make forecasting challenging. Developing an SVM-based rainfall forecasting model tailored for islands improves accuracy. This aids local authorities in informed decision-making for water management, agriculture, disaster preparedness, and infrastructure development. SVM's use in island rainfall forecasting shows promise, enhancing accuracy and reliability for effective resource management and decision-making in vulnerable environments.

## Problem statement

This project addresses the challenge of accurately forecasting rainfall in archipelago. The unique geography and susceptibility to extreme weather events make current forecasting methods less reliable, impacting water resource management, agriculture, and infrastructure planning. The project aims to enhance accuracy by developing and evaluating a specialized Support Vector Machine (SVM) model tailored for island conditions.

## Objective

Develop and evaluate a specialized Support Vector Machine (SVM) model tailored for island conditions.

## Methodology

### Preparation

Utilized tools such as Google Colab, Panda, Matplotlib for data processing and machine learning model development. Google Colab is a cloud-based platform, facilitated the creation and training of machine learning models in a Python-based environment. Panda, a data analysis library for structured data analysis. Matplotlib was utilized for data visualization through graphs and plots.

### Data Collection

Collected data from the Raja Haji Fi Sabilillah meteorological and geophysical station in Tanjungpinang City, Riau Islands, Indonesia. Data spanned from January 2021 to January 2023. Attributes included crucial factors in weather prediction. Data divided into 80% training, 20% testing.

### Data Preprocessing

A crucial first step is data preprocessing. Data preprocessing helps to clean the data from errors, missing values, and outliers. It improves the quality of the data, which in turn can lead to better models. Data preprocessing makes it possible to transform data into a form that is more suitable for machine learning algorithms.

### Training Data

In using the SVM technique, it tries to find the optimal classifier to separate two different classes. It tries to find the best dividing function (hyperplane) among an unconstrained number of functions to divide two types of objects. A good hyperplane is one that lies in the middle between two sets of objects from two classes. The definition of the equation on the separating hyperplane can be written as

𝑊 ∙ 𝑋 + 𝑏 = 0

W is the weight of a vector, i.e. 𝑊 = {𝑤1, 𝑤2, 𝑤3, … , 𝑤𝑛}; where 𝑛 is the number of attributes and 𝑏 is a scalar value or often called bias.

## Results

### Data Loading and Cleaning

The dataset was successfully loaded from a CSV file, and a systematic check for missing values was conducted. Rows containing missing values were removed to ensure the integrity of the dataset.

### ****Data Splitting****

The dataset was divided into training and testing sets, with an 80-20 split ratio.

### Support Vector Machine (SVM) Model Training

The SVM classifier, employing a linear kernel, was trained on the prepared training data. This kernel choice is suitable for linearly separable datasets.

### Model Prediction

The trained SVM model demonstrated its predictive capabilities by forecasting rainfall or no rainfall on the test set.

### Confusion Matrix

Confusion Matrix is used to analyze the performance of the Support Vector Machine (SVM) classifier on the test set. The confusion matrix is a table that summarizes the model's predictions, allowing for a detailed evaluation of classification results.

## Conclusion

In conclusion, the developed SVM model effectively predicted rainfall based on collected data. The model demonstrated reliable performance, showcasing its potential for practical weather forecasting applications. Indicates its ability to distinguish between positive and negative rainfall events.