Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Program: Bachelor of Science in Computer Science and Engineering

Course No: CSE 4108

Course Title: Artificial Intelligence Lab

Project Name: **Cooling Load Prediction of a Building.**

Date of Submission: 4 September, 2022

Submitted to:

Mr. Md. Siam Ansary
Lecturer, Department of CSE, AUST.

Ms. Tamanna Tabassum
Lecturer, Department of CSE, AUST.

Submitted by:

Yumna Islam Student ID: **18.02.04.046**

Farhana Azad Student ID: **18.02.04.068**

Introduction

We will predict cooling load requirement of a building based on different features of building shapes. It will help us to perform energy analysis of a certain building if the features are given and decide if the building structure is energy efficient or not. Feature-wise comparison of cooling load requirement can play significant role in energy analysis.

Dataset

We have chosen a dataset consisting of 450 samples. The dataset contains eight attributes (or features, denoted by X1...X8) and one response (or outcome, denoted by Y1). The aim is to use the eight features to predict the response which is cooling load requirement of a building.

Specifically:
X1 Relative Compactness
X2 Surface Area
X3 Wall Area
X4 Roof Area
X5 Overall Height
X6 Orientation
X7 Glazing Area
X8 Glazing Area Distribution
Y1 Cooling Load

The main dataset was consisted of 768 samples, 8 features and two outcomes from where we have chosen 450 samples, 8 features and one outcome.
Dataset link: https://archive.ics.uci.edu/ml/datasets/Energy+efficiency

ML Models

1. **Multiple Linear Regression**: Multiple linear regression is the most common form of linear regression analysis .In multiple linear regression instead of having a single independent variable, the model has multiple independent variables to predict the dependent variable.
2. **Random Forest Regression**: Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model. It builds decision trees on different samples and takes average in case of regression.
3. **Support Vector Regression**: Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. Support Vector Regression uses the same principle as the SVMs. The basic idea behind SVR is to find the best fit line. In SVR, the best fit line is the hyperplane that has the maximum number of points.

4. **K Nearest Neighbors Regression:** KNN regression is a non-parametric method that, in an intuitive manner, approximates the association between independent variables and the continuous outcome by averaging the observations in the same neighborhood. KNN algorithm uses '**feature similarity**' to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set.

Comparison of Performance Scores

Name of Algorithm	Mean Absolute Error	Mean Squared Error	R Squared Error	Mean Absolute Percentage Error
Multiple Linear Regression	2.399021950169777	12.46317279463418	0.8664264901866876	9.371461304432534%
Random Forest Regression	1.0149111111111142	3.117236088592597	0.9666712086948541	3.4415446425327954%
Support Vector Machines Regression	4.646701157789325	40.236705802669704	0.5697981377120531	18.736106316512203%
K Nearest Neighbors Regression	1.0629444444444442	2.3331886111111113	0.9750540690261371	4.684061825001497%

Discussion

Random Forest Regression model performs better as it give less erroneous result among the models we have used. We can get better result from our models by using large dataset to train the models. We are keeping it as the future improvement scope of our project.