Exploring A Machine Learning Model in Seasonal Prediction of Temperature over Dhaka

Mantasha Rahman Poushi and Shuchi Chaki

Department of Applied Mathematics, University of Dhaka, Dhaka, 1000, Bangladesh
mantasha-2018725564@amath.du.ac.bd, shuchi@du.ac.bd

Abstract

Accurate temperature predictions are crucial for various applications, including agriculture, energy management, and urban planning. Over the decades, sophisticated numerical prediction software has been developed and employed with the goal of better forecasting. In this era of machine learning, with machine learning models being used for different predictive jobs, exploration of such models in weather forecasting is becoming popular across the world. This research article investigates the use of Support Vector Machine (SVM) model for seasonal temperature forecasting over Dhaka City. We explore distinct scaling techniques, namely Min-Max Scaling, Standard Scaling, and Robust Scaling, to determine which method improves the SVM model's accuracy in predicting seasonal temperature variations using daily data for February to July collected from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data ($0.25° \times 0.25°$ global grid) for the years 2011-2020.

Keywords: Machine Learning (ML), Support Vector Machine (SVM), Seasonal forecasting.

Introduction

Weather prediction has proven to help provide early warning of the effects of weather on several facets of human life [1]. Weather forecasting has evolved significantly over time, from traditional methods based on barometric pressure and historical data to modern approaches incorporating advanced technologies, data analytics, and machine learning [2]. This evolution has improved the accuracy and efficiency of predicting atmospheric conditions. Early forecasting relied on direct observations and barometers, but meteorologists developed instruments like thermometers and hygrometers to collect more data. The introduction of Numerical Weather Prediction (NWP) models in the mid-20th century allowed for more accurate forecasting. At the moment, a large number of computer-based simulation models are employed in weather prediction. The atmosphere's present state is sampled in these simulations, and the equations of fluid dynamics and thermodynamics are numerically solved to determine the atmosphere's future state [3] Satellites provided real-time imagery and data on cloud cover and precipitation, while radar and Doppler technology improved warning systems. Ensemble forecasting acknowledges uncertainty in atmospheric processes, improving forecast reliability. In this development process, machine learning is a comparatively new addition. Machine learning is a field of AI that seeks to automate the processes of data analysis, prediction, and decision-making without requiring human intervention or training [4]. It involves the development of algorithms that enable computers to learn and improve from experience, enabling them to handle complex tasks and make accurate predictions or decisions. Machine learning has numerous applications, ranging from image and speech recognition to natural language processing, recommendation systems, and autonomous vehicles, thereby revolutionizing numerous industries and improving the efficiency and precision of decision-making processes [5]. Over the last few years, researchers have been investigating the application of different machine learning models in various sectors of weather prediction, starting from daily forecast to seasonal forecast, including regular monitoring to extreme event exploration [6], [7], [8]. However, in the context of Bangladesh, study of machine learning model in weather prediction is limited. For this reason, in this paper, we investigate a machine learning model, Support Vector Machine (SVM) in forecasting seasonal temperature over Dhaka city. Support Vector Machine algorithm is sometimes referred to as support vector networks or support vector machines, they examine data for relapse and league using associated learning algorithms [4]. It is a supervised machine learning model. This type of work is novel in the context of Bangladesh, and the study may open new direction in research related to forecasting.

Methodology

In order to predict summer season temperature over Dhaka city, we used six meteorological parameters of reanalysis data set, namely, temperature, relative humidity, geopotential height, surface potential temperature, Uwind and Vwind. These data were collected at atmospheric pressure level of 1000 hPa. Daily data for February to July were collected from the National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data ($0.25° \times 0.25°$ global grid) for the years 2011-2020. Support Vector Machine (SVM) model was built to predict the seasonal temperature over Dhaka. Monthly data of February, March and April was selected as predictors. The aim of the study was to predict the temperature of the month of May, June and July.
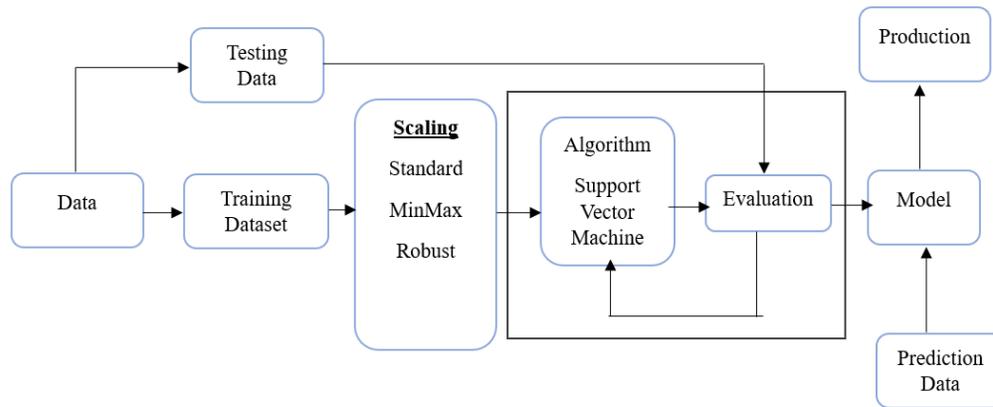
Figure 1: Flow diagram of methodology

Our experimental design involves training and evaluating SVM models with each scaling process to compare their predictive performance. After the initial processing of data, it was divided into train and test samples. The train set was used to model the data and the obtained model was validated using test data. For the verification, three performance metrics were employed, namely, mean squared error, mean absolute error and R^2 score. The SVM model is employed as a powerful machine learning tool due to its capability to handle non-linear relationships and high-dimensional datasets effectively.

Support Vector Machine Algorithm

SVM determines an optimal hyperplane as a solution to the problem of learning. The linear SVM formulation, in which the hyperplane is located on the input data x, is the most straightforward. The hypothesis space in this instance is a subset of all hyperplanes with the following form [4]:

$f(x) = wx + b$

In its most generic version, SVM discovers a hyperplane in a location other than the stretch of the captured data [4, 9].

Scaling in SVM

Scaling is a crucial aspect of Support Vector Machines (SVMs) for several reasons. It ensures that all features contribute proportionally to the model's decision-making process, preventing the dominance of certain features due to their larger scale. It also facilitates quicker convergence during the optimization process, enhancing the efficiency of the SVM training process. Additionally, scaling techniques inherently handle outliers, resulting in a more resilient and reliable SVM model. These advantages contribute to the overall effectiveness of SVM models in various applications. Standard Scaling, also known as Z-score normalization or Z-score standardization, is a data scaling technique that transforms numerical features to have a mean of 0 and a standard deviation of 1 [10]. This scaling method is particularly useful when the features in the dataset have different units or different scales. Min-Max Scaling, also known as Min-Max normalization, is a data scaling technique used to transform numerical features to a specific range. The purpose of Min-Max scaling is to bring all the values within a certain range, typically between 0 and 1.

Robust Scaling is a data scaling technique that is particularly useful when dealing with numerical features that may contain outliers. Unlike Min-Max Scaling and Standard Scaling, which are sensitive to outliers, Robust Scaling uses the median and interquartile range (IQR) to scale the data [10].

Performance Metric

1. Mean Absolute Error: It represents the difference between the original and predicted values extracted by averaging the absolute difference over the data set [11]. It can be expressed as:

$$\frac{1}{N}\sum_{i=1}^{N} |yi - \hat{y}|$$

2. Mean Squared Error: It represents the difference between the original and predicted values extracted by averaging the squared difference over the data set [11]. It can be expressed as:

$$\frac{1}{N}\sum_{i=1}^{N} (yi - \hat{y})^2$$

3. R^2 Score: It represents the coefficient of how well the values fit compared to the original values. The value from 0 to 1 is interpreted as percentages. In the context of machine learning, it is commonly used to evaluate the goodness of fit of a regression model [11]. The formula for R^2 score is given by:

$$R^2 = 1 - \frac{\text{Sum of Squared Residuals}}{\text{Total Sum of Squares}}$$

Results & Discussion

In this section, the results obtained are discussed. The results values for model application to test data set are plotted against the actual values are shown in the figure. From the figure we observe that, plot points obtained from the model trained with SVM without scalar are scattered, therefore it shows poor performance. Among the plots, the robust scalar
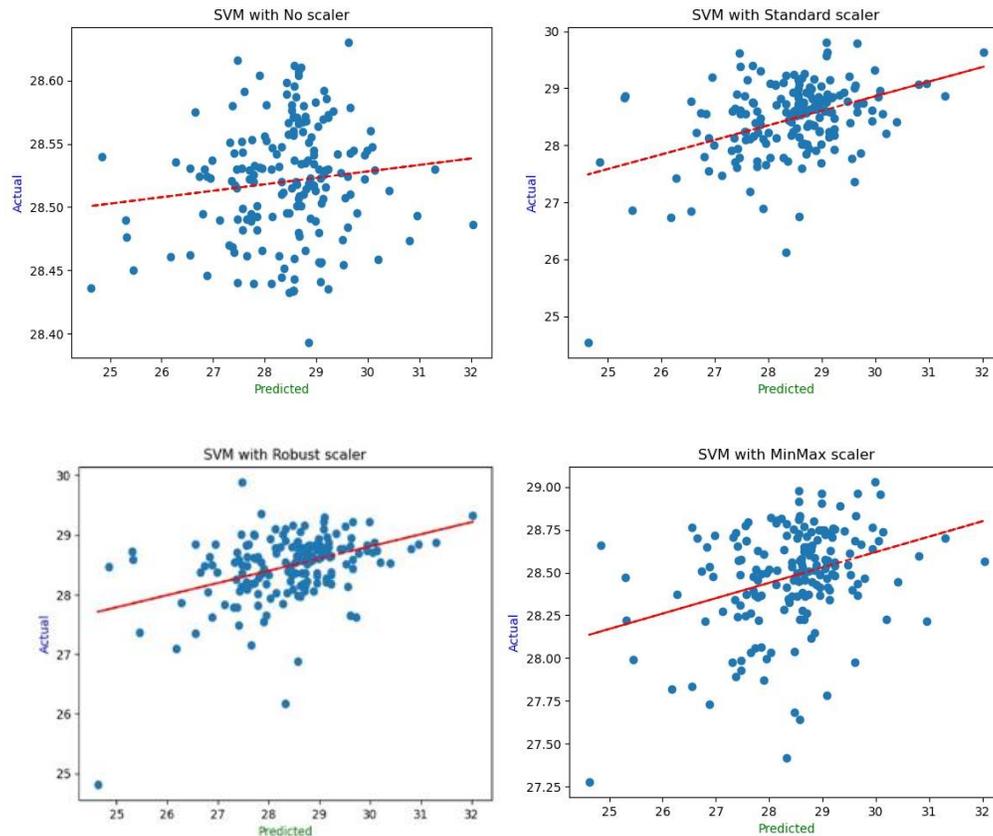
Figure 2: SVM implementation results for different scaling

graph resembles better fit. The performance matrices for these scaling processes are shown in the table. In Table 1, various procedures for Support Vector Machine (SVM) regression are compared based on their Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared, R^2 values. The procedures involve different feature scaling techniques applied to the SVM model. The SVM model without any feature scaling exhibits relatively higher errors and a low R^2 value, indicating suboptimal predictive performance. Standardizing the features using the Standard Scaler results in improved performance, with reduced errors and a higher R^2 value compared to the unscaled SVM model. Applying MinMax scaling leads to intermediary model performance, with errors and R^2 falling between the unscaled and Standard Scaler-scaled SVM models. Utilizing the Robust Scaler, which is resilient to outliers, demonstrates favorable results, with decreased errors and a higher R^2 value compared to both unscaled and MinMax Scaler-scaled SVM models. This table facilitates a comparative analysis of the impact of different feature scaling techniques on the performance of SVM regression, providing insights into the effectiveness of each scaling approach in improving predictive accuracy.

From the above discussion it is evident that, the value of R^2 is poor overall. Using support vector machine this regard does not better results. Seasonal forecasting is still very hard. Considering this fact, we can conclude that there is scope of development in the process.

Table 1: Comparative analysis of SVM with different scaling

Procedures	MSE	MAE	R^2
SVM without feature scaling	1.10988	0.7691	0.0595
SVM with Standard Scaler	0.9886	0.7211	0.1538
SVM with MinMax Scaler	1.0495	0.7424	0.1017
SVM with Robust Scaler	1.00222	0.7266	0.1422

Conclusion

In this era of machine learning, this study investigates the application of machine learning, specifically the Support Vector Machine (SVM) model, for seasonal temperature forecasting in Dhaka City. The research explores different scaling techniques aiming to check SVM model's performance. The findings derived from the study may guide future endeavors in predicting temperature in the context of Bangladesh.

Reference

[1]   M. Abdulraheem, J. B. Awotunde, E. A. Abidemi, D. O. Idowu and S. O., Weather prediction performance evaluation on selected machine learning algorithms, IAES International Journal of Artificial Intelligence, vol. 11, no. 4, p. 1535, 2022.

[2]   M. a. L. D. a. V. C. Holmstrom, Machine learning applied to weather forecasting, Meteorol. Appl, vol. 10, pp. 1-5, 2016.

[3]   M. Holmstrom, D. Liu and C. Vo, Machine learning applied to weather, Meteorol. Appl, vol. 10, pp. 1-5, 2016.

[4]   S. Suthaharan, Machine Learning Models and Algorithms for Big Data Classification, Integrated Series in Information Systems, vol. vol. 36, pp. pp. 1-12, 2016.

[5]   I. a. M. M. J. El Naqa, What is machine learning?, Springer, 2015.

[6]   D. Watson-Parris, Machine learning for weather and climate are worlds apart, Philosophical Transactions of the Royal Society A, vol. 379, p. 20200098, 2021.

[7]   Salcedo-Sanz, S., Pérez-Aracil, J., Ascenso, G. et al. Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review. Theor Appl Climatol 155, 1–44 (2024). https://doi.org/10.1007/s00704-023-04571-5.

[8]   Khan, N., Shahid, S., Ismail, T. B., & Behlil, F., Prediction of heat waves over Pakistan using support vector machine algorithm in the context of climate change, Stochastic Environmental Research and Risk Assessment, doi:10.1007/s00477-020-01963-1, 2021.

[9]   M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, Support vector machines, in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428.

[10]   wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Feature_scaling.

[11]   javapoint, [Online]. Available: https://www.javatpoint.com/performance-metrics-in-machine-learning.