Uluslararası İleri Doğa Bilimleri ve Mühendislik Araştırmaları Dergisi
Sayı 8, S. 69-74, 6, 2024
© Telif hakkı IJANSER'e aittir
Araştırma Makalesi

International Journal of Advanced Natural Sciences and Engineering Researches
Volume 8, pp. 69-74, 6, 2024
Copyright © 2024 IJANSER
Research Article

https://as-proceeding.com/index.php/ijanser
ISSN: 2980-0811

Regression-Based Temperature Prediction Using Humidity and Pressure Data for Smart Factories

Merve Sayar*, Merih Palandöken[2]

[1]*Department of Electrical and Electronics Engineering/Faculty of Engineering and Architecture, İzmir Katip Celebi University, Turkey*
[2]*Department of Electrical and Electronics Engineering/Faculty of Engineering and Architecture, İzmir Katip Celebi University, Turkey*

*(*merve.sayar.smarttech@gmail.com*)*

(Received: 09 July 2024, Accepted: 24 July 2024)

(4th International Conference on Scientific and Academic Research ICSAR 2024, July 19 - 20, 2024)

ATIF/REFERENCE: Sayar, M. & Palandöken, M. (2024). Regression-Based Temperature Prediction Using Humidity and Pressure Data for Smart Factories. *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(6), 69-74.

Abstract – With the rapid growth of technology, numerous new innovations appear in our life. One of these advancements is Industry 4.0, often known as the Fourth Industrial Revolution, which brings together the Internet of Things (IoT) and cyber-physical systems. IoT enables the transport of sensor data via the internet, allowing for data exchange and central control systems that do not require human participation. Smart Factories, which are equipped with automation technologies that allow for real-time monitoring and remote control, improve production efficiency by guaranteeing that development in each department is handled from a single location. The Smart Factory Management Information System is a computer-based information system that generates management reports by aggregating and summarizing transaction records obtained via sensors. This method improves energy efficiency in response to rising energy demand. It also gives responsible workers with quick access to information, cost savings, and system security. The primary goal of this study is to estimate temperature using humidity and pressure data. The BME680 sensor was used to make pressure and humidity-based temperature forecasts. Temperature predictions were made using Gaussian Process Regression, a machine learning methodology, based on sensor data, and it was discovered that this method produces more accurate predictions of actual values. The study demonstrates that a Smart Factory Management Information System loaded with Industry 4.0 innovations can make substantial contributions to enhancing energy efficiency, lowering costs, and maximizing efficiency.

Keywords – Industry 4.0, Smart Factory, Sensor, Regression.

I. INTRODUCTION

Machine learning, a vital aspect of artificial intelligence, involves creating algorithms that enable computers to learn from data and make decisions based on that data. Unlike traditional programming that relies on explicit instructions, machine learning uses data-driven approaches that allow computers to autonomously improve their performance over time. The main goal is to develop models that learn from historical data, identify complex patterns, and solve new problems effectively. This capability is required in a variety of fields, including predictive analytics, natural language processing, and image recognition.

A key advantage of machine learning is the ability to quickly and accurately process and analyze large amounts of data, where traditional methods often fall short due to the complexity and volume of data [1].
Machine learning models, especially advanced models such as deep learning and reinforcement learning, are excellent at uncovering hidden patterns and insights by processing multiple layers of data. These models are trained on historical data to predict future outcomes or classify data points into different categories. Machine learning equips computers with the ability to learn from data and experience, similar to the human brain [2]. The key task is to design a reasonable model that learns from historical data to complete specific target tasks. A model that fails to fully learn from historical data is underfitting, while a model that overlearns noise information is overfitting [3].
Machine learning models are classified as traditional and advanced models. Traditional models include supervised, semi-supervised, and unsupervised learning, with supervised learning involving labeled data and unsupervised learning focusing on unlabeled data. Typical supervised models include neural networks, support vector machines, and regression algorithms. Unsupervised models include K-means clustering, principal component analysis, and self-encoding algorithms [4-5]. Advanced models such as deep learning and reinforcement learning include deep neural networks, convolutional neural networks, and Q-Learning models.
This study investigates the application of machine learning to predict temperature based on humidity and pressure values. It is aimed to provide accurate temperature predictions from sensor data by using a BME680 sensor. This approach demonstrates the potential of machine learning in improving data analysis and prediction capabilities in a variety of industrial applications.

II. MATERIALS AND METHOD

This work introduces a regression model that predicts temperature from humidity and pressure variables, providing valuable scientific and practical information. The process for model development includes five key stages:

A. *Data Collection*

In this first stage, raw sensor data is systematically collected from the BME680 sensor over a period of time. The BME680 sensor is capable of measuring a variety of environmental parameters including temperature, humidity and pressure. Consistent and accurate data collection forms the basis for subsequent modeling stages. Data points are often recorded at regular intervals to capture variability in environmental conditions. This step ensures that sufficient data is available to create a reliable regression model.

B. *Data Pre-processing*

Once the raw data is collected, it goes through an extensive pre-processing phase to prepare it for modeling. This stage includes several important steps. First, anomalies or outliers that may reduce the performance of the model are cleared; This includes removing missing values, duplicates, and data inconsistencies. The data is then normalized to ensure that the different properties (humidity and pressure) are on a similar scale; because many machine learning algorithms perform better when input features are standardized. Finally, new features are created or existing features are modified to better represent underlying patterns in the dataset. This stage ensures that the data set is of high quality and suitable for training an accurate and reliable regression model. Raw data is cleaned, normalized and feature extracted before modelling. This phase ensures that the data is of high quality and consistent enough to train an appropriate model [6].

C. *Seperation of Data into Training and Test Sets*

The pre-processed data is then divided into two subgroups: training and testing. This distinction is critical for evaluating model performance on previously unseen data [7]. The training set is used to train the regression model and contains enough examples for the model to learn. The test set is completely separated from the training process and is used for evaluation purposes only. This subset allows us to evaluate how well the model generalizes to new, unseen data. This phase helps prevent overlearning, where the model performs well on training data but poorly on new data, and ensures that the model's performance metrics accurately reflect its true predictive capabilities. The dataset is divided into training and testing subsets to facilitate training and evaluation of models.

D. Training and Optimization of the Model

Many techniques have been used to create regression models. Gaussian Process Regression (GPR) technique was chosen as the most appropriate. GPR is a powerful nonparametric Bayesian approach well adapted for regression tasks. The model is trained using the training data set and learns the relationship between humidity, pressure and temperature. The hyperparameters of the model are fine-tuned to improve prediction accuracy. This involves systematically evaluating different sets of hyperparameters using techniques such as cross-validation and selecting the best-performing configuration. The goal of this phase is to develop a model that can accurately predict temperature based on humidity and pressure readings. [8].

E. Evaluation of the Model

The final stage involves evaluating the trained model using the test data set. This step is critical to understanding the model's performance in real-world scenarios. The evaluation is done using several important metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). While MAE measures the average magnitude of errors between predicted and actual temperature values, RMSE gives greater weight to larger errors by taking the square root of the mean of the squares of the differences between predicted and actual values. These measurements allow us to comprehensively evaluate the accuracy and robustness of the model. The low MAE and RMSE indicate that the model is effective in predicting temperature based on humidity and pressure variables [9].

III. RESULTS

Sensor data was collected at various points and in different time periods at Katip Çelebi University Smart Factory Systems Application and Research Center and in the garden. 1450 data collected were displayed on the interface and saved to Google Firebase Real-Time Database. Data in JSON format was converted to Excel format for machine learning applications. Details of the data are shown in Table 1.

Table 1. Some of the data collected from temperature, humidity and pressure values

Humidity	Pressure	Temperature
23.744	1014.37	35.52
23.918	1014.38	35.5
24.013	1014.36	35.57
22.696	1014.59	34.50
22.48	1014.54	34.88
21.92	1014.64	34.61
21.758	1014.66	34.59
23.845	1014.36	35.52
23.872	1014.37	35.52
23.877	1014.37	35.52
21.75	1014,73	34.31
23.885	1014.36	35.51
25.902	1014.56	31.48
21.685	1014.85	34.93

Abnormal deviations in temperature and humidity values in the data set, which may be caused by sensor errors or environmental factors, are seen as extreme values and NaN values. During the data preprocessing stage, these deviations are detected and replaced with the average value of the relevant column. This ensures the homogeneity of the data set. Additionally, by applying the normalization method, data values are reduced to the [0,1] range, thus ensuring that the model is trained in a more balanced and effective manner. This process reduces noise in the data set, ensures stable training of the model, and helps make more accurate predictions. In this project, the dataset is divided into 80% training, 15% testing, and 5% validation. This project uses various machine learning methods to predict temperature from humidity and pressure values. In the data analysis and modeling stages, simple linear regression was used initially, followed by multiple linear regression methods. Models were then improved using more complex algorithms such as support vector machine (SVM), k-nearest neighbors (KNN), decision trees, and random forest. As a result, the predictions obtained by combining different algorithms and ensemble methods provide average results for temperature prediction from humidity and pressure values. Random forest is an ensemble learning method created by combining many decision trees. Each decision tree is trained on different subsets of the data set, and these trees are then combined to make predictions. According to the results obtained by the project with the random forest method, the MSE value was determined as 0.57917 and the R-square value was determined as 0.8037.

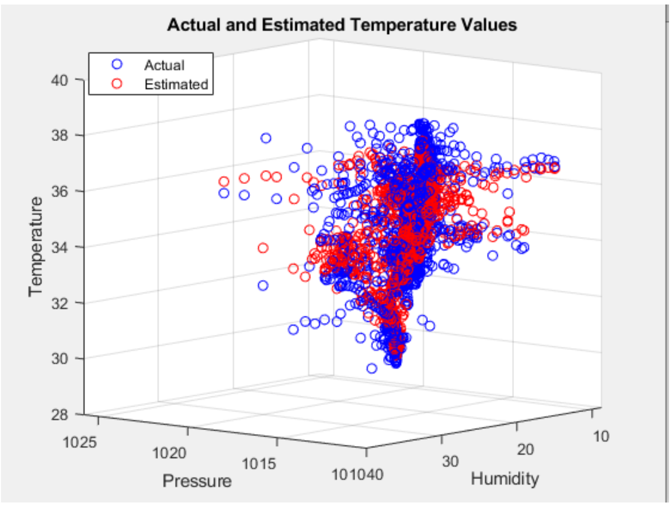

Fig. 1 Result of Random Forest

Figure 2 illustrates the R-square and MSE values for each regression type. These graphs show a visual comparison of the accuracy and error rates of each regression model. R-square values indicate the model's ability to explain data, whereas MSE values reveal the average inaccuracy in predictions. As a result, it will be easier to determine which regression model performs best, making the model selection process more efficient.

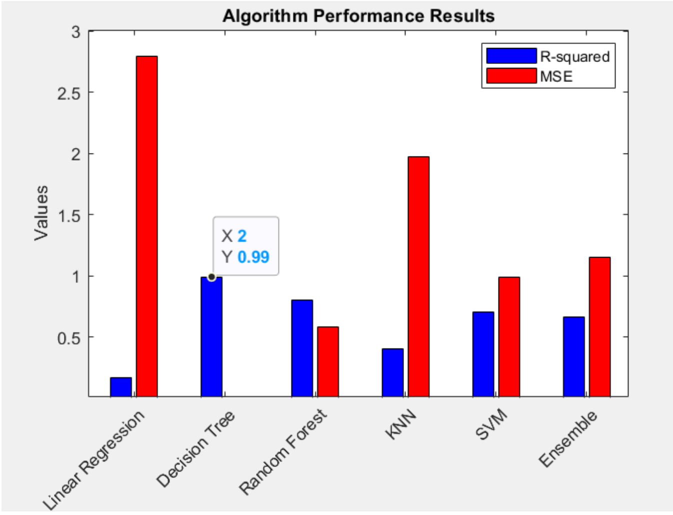

Fig. 2 Result of Algorithm Performance

IV. DISCUSSION

The emergence of Industry 4.0 integrates IoT, internet services and cyber-physical systems, enabling efficient data transfer and centralized control without human intervention. This work leverages sensor networks to develop intelligent systems that aim to increase convenience, reduce costs, and increase efficiency in both personal and industrial applications. Smart Factories, which exemplify this evolution, use advanced automation for real-time monitoring and remote control, optimizing production processes and resource allocation in terms of central management. This research focused on predicting temperature using pressure and humidity data from the BME680 sensor, and it was decided that the best application for accurate temperature prediction was the practical application of Gaussian Process Regression (GPR). GPR has proven effective in capturing complex data relationships and provides more accurate temperature estimates than traditional methods. This approach not only improves operational efficiency and energy management, but also highlights the potential of advanced machine learning in improving industrial processes compliant with Industry 4.0 standards. In summary, this study shows how smart systems can revolutionize industrial operations by leveraging IoT and predictive analytics. These systems improve data accuracy and predictive capabilities, paving the way for future innovations in smart factory technologies and promising continued advances in efficiency and adaptability across industrial sectors.

V. CONCLUSION

The rapid advancement of technology has ushered in a new era known as Industry 4.0 or the Fourth Industrial Revolution, which seamlessly integrates the Internet of Things (IoT), internet services and cyber-physical systems. IoT facilitates the transfer of sensor data over the internet, enabling data sharing and central control mechanisms without human intervention. This study aimed to benefit from these developments by developing an intelligent system using sensor networks. The main objectives were to provide significant convenience for both personal and business applications, reduce operational costs and increase overall efficiency. Smart Factories represent a significant evolution, especially in industrial production. These factories are equipped with advanced automation technologies that facilitate real-time monitoring and remote control, thereby increasing production efficiency. Smart Factories can optimize workflows and resource allocation by managing the development in each department from a single central point. The research attempts to predict temperature values based on pressure and humidity readings, specifically leveraging sensor data from the BME680 sensor. The application of Gaussian Process Regression has proven to be particularly effective, providing more accurate estimates of actual temperature values than other methods. As a result, this study has successfully developed a smart system compatible with Industry 4.0 standards. The system not only improves operational efficiency and energy management, but also integrates advanced machine learning techniques to increase data accuracy and predictive capabilities. The results of this research demonstrate the significant potential of smart systems to transform industrial operations, providing a roadmap for future innovations in smart factory technologies.

ACKNOWLEDGMENT

This study is supported by TUBITAK 2209-A University Students Research Projects Support Program within the scope of project numbered 1919B012108131.

REFERENCES

[1] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
[2] Gör, İ. (2014). Vektör nicemleme için geometrik bir öğrenme algoritmasının tasarımı ve uygulaması (Master's thesis, Adnan Menderes Üniversitesi).
[3] Liao, Y., Weng, Y., Liu, G., & Rajagopal, R. (2018). Urban MV and LV distribution grid topology estimation via group lasso. IEEE Transactions on Power Systems, 34(1), 12-27.
[4] Tang, L., Yang, F., & Ma, J. (2014, May). A survey on distribution system feeder reconfiguration: Objectives and solutions. In 2014 IEEE Innovative Smart Grid Technologies-Asia (ISGT ASIA) (pp. 62-67). IEEE.
[5] Amanulla, B., Chakrabarti, S., & Singh, S. N. (2012). Reconfiguration of power distribution systems considering reliability and power loss. IEEE transactions on power delivery, 27(2), 918-926.
[6] Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Elsevier.
[7] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
[8] Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press.
[9] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.