Looking into Raw Material Costs Through Machine Learning to Improve Efficiency

Adruce Bin Khairudin , Lee Jia Fong , Wadthanak Khongsuwan , Teh Yu Xiang , Bryan Thong Khai Junn , Noor Ul Amin *

Posted Date: 24 February 2025

doi: 10.20944/preprints202502.1861.v1

Keywords: Machine Learning; SME; Linear Regression; Decision Tree Regression; Random Forest Regression; Data Preprocessing; Mean Squared Error; R2 Score; Mean Absolute Error

 doi:10.20944/preprints202502.1861.v1

Article

Looking into Raw Material Costs Through Machine Learning to Improve Efficiency

Adruce Bin Khairudin, Lee Jia Fong, Wadthanak Khongsuwan, Teh Yu Xiang, Bryan Thong Khai Junn and Noor Ul Amin *

School of Computer Science, Taylors university, Malaysia

* Correspondence: nooraminnawab@gmail.com

Abstract: This document is meant to show how machine learning can analyze the raw material costs that are affecting the business operations of small-to-medium enterprises. Currently, SMEs face the challenge of rising raw material costs, cost inefficiencies and lack of predictive insights into cost management, which create problems for trying to operate a small business smoothly in Malaysia. Machine learning can help find solutions to those operational problems through using regression to analyze and forecast the costs of raw materials in general. To do that, three regression models are used which are Linear Regression (LR), Decision Tree Regression (DTR) and Random Forest Regression (RFR). By analyzing the dataset by raw material prices, it is shown that the DTR outperforms the other models in forecasting the prices of raw materials in general with an R2-score of 0.91, followed by the RFR at 0.87, and LR at 0.79. With these findings in mind, it could serve as how machine learning can help in enhancing in how SMEs can adapt to factors with the implementations of cost management strategies, helping decision-making to ensure competitiveness and profit sustainability, with the framework serving as a solution that is both practical and scalable to offer SMEs opportunities to respond to market shifts and mitigate risk.

Keywords: Machine Learning; SME; Linear Regression; Decision Tree Regression; Random Forest Regression; Data Preprocessing; Mean Squared Error; R2 Score; Mean Absolute Error

---

I. Introduction

1. Background to the Proposed Solution

a). Problem to be Solved and Proposed Relevant Solution Problem Identification

The operation challenges of SMEs under JPW Melaka involve a lack of data analysis due to the following reasons. These challenges include:

- **Rising Raw Material Costs**

One of the biggest challenges for SMEs is that the cost of materials varies over time, and has a direct impact on the profitability of the venture and the capacity to set affordable prices.

- **Cost Inefficiencies**

Since cost control entails the identification of costs and the forecast of their behavior, SMEs often encounter problems with productivity, purchasing, and inventory that result in the wastage of resources or unnecessary spending.

- **Lack of Predictive Insights into Cost Management**

The major problem of many SMEs is that they cannot predict how cost drivers (including material and labor costs) evolve or how certain factors like seasonality or demand impact the costs.

Proposed Solution

In order to address these challenges, this research presents the Cost Optimization and Forecasting Framework which can assist SMEs in cost prediction. The framework leverages three regression

 doi:10.20944/preprints202502.1861.v1

models to analyze the relationship between raw material costs, production, and overall business expenses:

Linear Regression (LR)

A basic method that was used to estimate the future changes in cost based on past records. It makes it easier for SMEs to identify the direct link between the cost of raw materials, seasonality, and other factors.

Decision Tree Regression (DTR)

A model that provides more detailed information about the behavior of different variables—different categories of products or high demand periods—on costs. The tree structure makes it easier for SMEs to understand the model's decision-making process as compared to the original model.

Random Forest Regression (RFR)

A set of decision trees that integrate several tree-based models to give a better and more stable prediction by minimizing the chances of overfitting. This model will assist in the determination of interactions between variables that may affect costs in a given organization.

Objective

The first is to assist SMEs in improving the efficiency of their cost control through the timely prediction of the costs of raw materials and other operational costs. This will enable them to forecast better, minimize wastage and therefore improve on their revenues.

The framework will also train and test all three models and then compare the results to determine the best model to use for cost forecasting and optimization.

b). Compare and Discuss the Proposed Solution with Recent Exciting Solutions

Solution 1: Walmart's Predictive Analytics for Supply Chain Management

Walmart employs sophisticated predictive models to drive its supply chain, especially on issues to do with costs of raw materials, inventory, and demand. The company uses machine learning models to forecast the changes in raw material costs and product demand so that they can avoid overstocking or understocking. Through studying of past sales records, seasonality, and other outside factors, Walmart can cut down its operating expenses and increase its gross profit.

Key Features:

- It employs sophisticated algorithms in cost prediction, demand forecasting, and inventory management.
- Provides real-time data updates for effectiveness in decision-making.

Comparison:

Similarity: The rationale of the proposed solution and the strategy of Walmart are based on cost estimation and optimization of expenditures.

Difference: Walmart's solution is on a different level of scale, which has access to enormous amounts of data and high technologies. On the other hand, the proposed solution is for the SMEs with less data and resources and mainly targeting the cost-efficient regression model like Linear Regression, Decision Trees, and Random Forests.

Solution 2: Starbucks' Ingredient Procurement Optimization

Decision trees and random forests are used at Starbucks to predict and manage the procurement and cost of ingredients. Starbucks can easily predict the cost of raw ingredients by analyzing the customer demand, seasons, and preferences of products. It then adapts its sourcing techniques to avoid wastage and cut on costs of production but at the same time ensure the availability of the ingredients to satisfy the consumers.

Key Features:

- Decision trees and random forest models are used to predict the costs of the ingredients using past data and demand estimates.
- Concentrates on the proper control of the procurement process so that the cost of procurement is kept low.

 doi:10.20944/preprints202502.1861.v1

Comparison:
Similarity: Both solutions rely on machine learning algorithms such as Decision Trees and Random Forests to enhance cost control and thus increase the business's profitability.
Difference: While Starbucks concentrates on the costs of ingredients of its products, the proposed framework will give a more general approach, which will look for the optimal operational costs of the SMEs in various sectors based on the various cost drivers such as raw materials, manpower, and seasonality [24–26].

II. Literature Review

1. Linear Regression Model

Linear regression Model is an algorithm that use to predict a variable's value based on another variable's value. Dependent variable is the variable's value we want to predict while independent variable is the variable's value that we use to predict (IBM, n.d) perform well at both regression and classification tasks (IBM,n.d.).

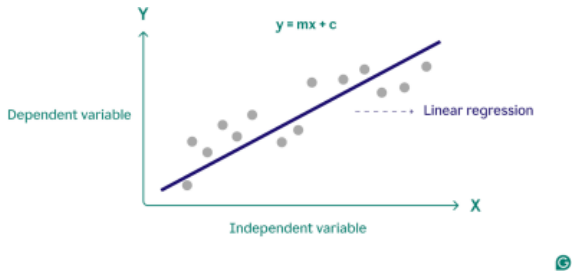

Figure 1. Linear Regression Model (Grammarly, 2024).

2.3.1. Challenges of Random Forest Regression

Although it can help to prevent overfitting, it still has a chance. It particularly will occur when using noisy datasets. .Random forest regression will computationally expensive when working with large datasets.It need a huge storage to store those data (AIML, 2024).Other than that, when working with huge data sets, it can take a long time due to the need to process each decision tree (IBM, n.d.).

2.1.1. Benefits of Using Linear Regression Model:

Linear regression is easy to implement as it does not require significant overhead during deployment and maintenance. Then, if compared to other deep learning models[21–23] such as neural networks, it is highly interpretable, allowing a clear understanding of how input variables affect the output. Furthermore, its lightweight computing properties make it ideal for scalability and efficient processing of large data sets in big data scenarios (Vijay, 2023)

2.1.2. Disadvantages of Linear Regression Model:

When relationship of dependent and independent value is not linear or occur high correlation, the model will has change to give unstable predictions. Next, linear regression model is very easy to have overfitting and underfitting. Overfitting occur when training data too well but fail togeneralize unseen data while underfitting happen when it is too simple (GeeksforGeeks, 2024).

2.2. Decision Tree Regression

Decision tree regression is use to predict continuous numerical values. It is very easy to understand its rules and patterns. Besides,it handles both numerical and categorical data, making it adaptable to variable use cases (Viswa, 2023).Rather than that, it will automatically solve missing values and not need to use feature scaling. However, decision trees are easy to overfit training data then deep with many nodes.Other than that, decision trees may be unstable due to small

 doi:10.20944/preprints202502.1861.v1

variations of data. It will cause different tree structures, affecting consistency and reliability (GeeksforGeeks, 2024).

2.3. Random Forest Regression

Random Forest Regression is a technique that is used to predict numerical values. It combines the predictions those gain from decision trees, by using multiple predictions to prevent overfitting and also help to improve accuracy. Since it uses multiple predictions to create an accurate and stable prediction so it is a machine learning technique that calls ensemble learning method(GeeksforGeek, 2024). Besides, it is very popular among data scientists because it is able to

III. Analysis

Before conducting an analysis for the machine learning methods that are being used for the machine learning models, it is important that first and foremost, the data must first be preprocessed.

3.1. Overview of Data Preprocessing

Data preprocessing is a critical step in performing data analysis. For our dataset that was given for the purpose of research, it was given in the Excel file format, .xlsx, and this caused a lot of problems. While formatting in Excel provides advantages such as formatting to make it more visually appealing, such as the use of colors in cells, however in machine learning, while it is possible to work on the xlsx file, the fact that it came with multiple sheets complicated the process, which is why it is better to split the sheets into several csv files, which are more preferred for machine learning algorithms.

García et al. (2014) gave several key negative factors that influenced the dataset that is given that may affect the quality of the data that will be used in the machine learning algorithm, from noise to high dimensionality, and if the data is not preprocessed, the machine learning algorithm will produce low-quality results from low-quality data.

In detail about the three main factors about the factors affecting data quality:

- Noise: Errors, outliers, and inconsistencies are examples of noisy data that can skew learning and impair model performance. (Arinze, 2024)
- Dimensionality: Sometimes, the dataset given has too many features and attributes that it has become impossible to work on without compromising data performance[18–20].

To ensure high model performance within the dataset, it is essential that the noise and dimensionality must be dealt with first as part of the data preprocessing before being analyzed for the machine learning models.

3.2. Data Preprocessing into Action

3.2.1. Data Acquisition

For the first step in data preprocessing, the data was gathered through direct contact with a food and beverage (F&B) small- to-medium enterprise (SME) owned by Mdm. Rohana Hananiaz in the state of Malacca in Malaysia. The data given details the SME's business performance, from the sales made between June and August, the cost of raw materials that were recorded within those months, and the profit and losses made in each month. For the purpose of optimizing cost efficiency for raw materials, only the records regarding the raw materials were considered.

At first, the dataset was organized into separate datasets due to how the dataset was initially given in an xlsx file, with its ability to store multiple sheets in one file and visually format them to be easier to understand to others who are not computer scientists. For the purpose of the research, only sheets that are necessary, like the records showing the purchases of raw materials from June to August. Those sheets are later concatenated in a Google Colab file to be processed as one dataset.

 doi:10.20944/preprints202502.1861.v1

3.2.2. Exploratory Data Analysis

After concatenation, the dataset then underwent an exploratory data analysis (EDA) to identify the columns within the dataset. The identified columns include:

- BULAN: representing the months
- TARIKH: representing the day of the month
- BAHAN: representing the item that was purchased
- CATATAN: representing the location of where the item was purchased
- JUMLAH: representing the cost of the item that was purchased
- KATEGORI: classifying what kind of item that was bought, either as "BAHAN MENTAH" (raw materials) or "LAIN-LAIN" (others).

3.2.3. Handling Missing Values

For the noise, the dataset observed has many missing values and incompatible file formats through an exploratory data analysis (EDA), mainly in the "BULAN", "TARIKH", "CATATAN" and "KATEGORI" columns, representing the months and dates respectively. The columns representing the months have attributes written as strings and have missing values, totaling at 79 values, while the columns representing the day of each month have missing values, at around 66 of them, representing their day of purchase. Meanwhile, "CATATAN" and "KATEGORI", representing where the items in "BAHAN" were bought, are in a string format and "KATEGORI" classifies on what these items are, from "BAHAN MENTAH" (raw materials) to "LAIN-LAIN" (others).

To deal with missing values in both "BULAN" and "TARIKH", the first is to impute the values in the columns to fill up the missing values, using the ".fillna" method from the Pandas library in Python. To make sure that it follows exactly like the months that were defined, the .fillna() method uses "ffill" to copy each value into the rows that are missing their values with the previous, defined row.

3.2.4. Data Encoding

The "JUMLAH" column in the dataset, where it reveals the price of each item in "BAHAN", was formatted as currency. For machine learning purposes, the currency format was switched to a numeric format. This is done as machine learning models only work in numeric data, and currency formats are non-numeric, making it complicated for machine learning models to work on. (Raut, 2022)

The "BAHAN", "CATATAN" and "KATEGORI" columns within the dataset are then encoded with LabelEncoder, that turns the non-numeric formatting of the data in those columns into numeric columns, as again, machine learning models could only work well with numeric values. As a result, this can ensure that errors are not made when the dataset is run through machine learning models, ensuring integrity of the model performance.

Additionally, we could also take a look at the rising costs of specific raw materials, such as chicken. In the dataset that has been given, it is possible to create a new dataframe to filter the dataframe to only include the items in "BAHAN", labeled "AYAM CHOP" and "AYAM KING", and the same data encoding can be applied as well.

3.3. Machine Learning

After preprocessing the dataset that is going to be used later on the machine learning models, the next step is to split the dataset into training and validation (also known as test) sets. This step is critical to ensure that the machine learning models that are going to be utilized will have their performance evaluated based on their performance metrics that are used for regression analysis[27,28].

 doi:10.20944/preprints202502.1861.v1

3.3.1. Dataset Splitting

At first, the independent and dependent variables are defined, with “BULAN”, “TARIKH”, “BAHAN”, “CATATAN” and “KATEGORI” being defined as the independent variables, labeled “X”, and the “JUMLAH” column being labeled the dependent variable, “Y”.

Afterwards, by utilizing the train_test_split library, the dataset is then split into training and test sets by an 80:20 ratio. The training dataset, comprising 80% of the dataset, will be used for fitting the machine learning models that are going to be used, while the test dataset, comprising the remaining 20%, that was created will be used for evaluation of the machine learning model.

3.3.2. Model Fitting

A function is then created to fit the machine learning models into the dataset. For the purpose of research, the models used and why they are justified are:

- Linear Regression:
- Decision Tree Regression:
- Random Forest Regression:

3.3.3. Model Evaluation

Then, after fitting the models, the evaluation metrics were then given through a library within Python to evaluate the machine learning model by three main matrices, commonly used for regression models:

- Mean Squared Error (MSE): measures the closeness between a regression line to its data points. A lower score indicates a better performance. (Gupta, 2024)
- R^2 score: How well the machine learning model interprets the data that is observed. A high r-squared scored indicates the best performance, but it is not necessarily good as it may indicate problems in the dataset, such as overfitting or underfitting. (Taylor, 2024)
- Mean Absolute Error (MAE): measures the overall accuracy of the machine learning model used. The lower the score, the better it can align with predictions. (Ahmed, 2023)

3.3.3. Performance Visualization

After the evaluation metrics were given for each model, the results for the actual and predicted values that were given by the machine learning models were later visualized through tools in the ‘matplotlib’ library in Python, which allows graphs to be created within the Google Colab file itself.

IV. Result and Discussion

4.1.1. Reporting and Visualization of the Proposed Solution

The purpose of this report as well as the visual aids present in the report is so to help SMEs understand the process of what we have achieved thus far with their provided dataset albeit missing some crucial information in the report is discussed on how we as a group tackle those issues. Since there were some missing values within the dataset that were needed for us to complete our machine-learning models we used data preprocessing to tackle that issue. For us to ensure that the model performed properly as well as accurately we also evaluated it using model evaluation methods like mean squared error score, R2 score, and mean absolute error score; these scores help us evaluate the machine learning model to see whether they have performed to our expected standards. Lastly, we included performance visualization of how each of the machine learning models we trained performed. With all these steps it should in theory help the SMEs understand and make actionable insights in their business model.

 doi:10.20944/preprints202502.1861.v1

4.1.2. In Detailed Steps Taken in the Report (Name WIP)

The ways in which reporting and visualization are utilized in our proposed solution are mentioned below:

Data Preprocessing Visualization

- Utilizes tabular summaries to show encoded and cleaned data typically used to handle missing values in the dataset if there were any.

Model Evaluation Report

- As seen in Table 1 this is typically used to evaluate a model in our case we utilize it to obtain the mean squared error score, R2 score, and mean absolute error score which all together should allow us to evaluate how each of the models performed.

Performance Visualization

- Shown in Figure 2 these scatter plots are utilized to compare the actual and predicted values obtained from the trained machine-learning models. These graphs should help SMEs interpret complex data which in hand helps them in their decision-making for their business model/plan.

Table 1. Results of the machine learning models that were utilized for raw materials in general.

Model	Mean Squared Error	R2 score	Mean Absolute Error
Linear Regression	7859.67	0.79	59.43
Decision Tree Regression	3263.07	0.91	27.69
Random Forest Regression	4906.82	0.87	41.58

 doi:10.20944/preprints202502.1861.v1

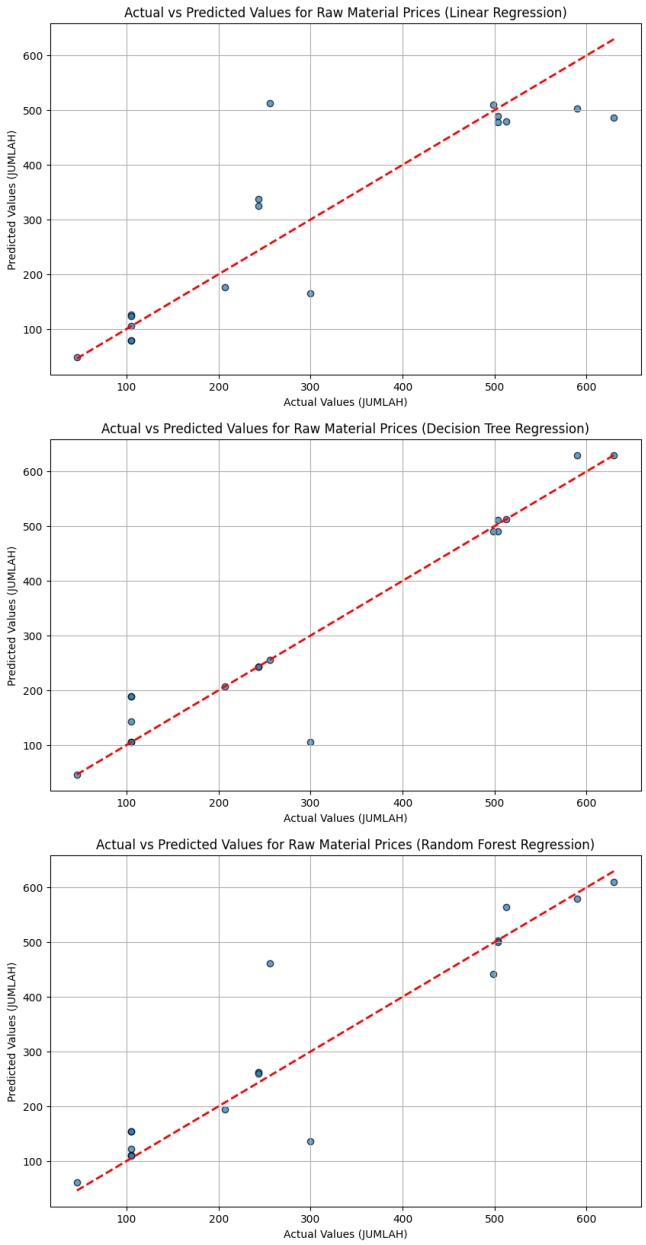

Figure 2. Graphs showing results of each machine learning method by raw materials in general.

4.2.1. Tools and Used for Reporting and Visualization

To achieve the results we are expecting we require tools to accomplish that, like certain libraries needed to make the graph from the trained model which will then generate scatter plot graphs like the ones shown in Figure 3. In this section we discuss the tools used in our machine-learning models.

1. Matplotlib ("import matplotlib.pyplot as plt")
    a. This library is mainly used to provide a comprehensive visualization aid that are either in a static form, animated form, or an interactive form. An example of matplotlib usage is the graphs in Figure 3 where it is used to make scatter plots and line plots and can be utilized to show the predicted against the actual price of the raw materials.
2. Seaborn ("import seaborn as sns")
    a. This is a library used to create statistical graphs. It in a way functions similarly

 doi:10.20944/preprints202502.1861.v1

to matplotlib except it is way more focused on the statistical graph side.

3. Pandas DataFrame ("import pandas as pd")
    a. Is a type of library used in importing the dataset which will be used as the training info for the machine-learning model as well as used in helping display evaluation metrics of the model's performance.
4. Google Colab Compiler
    a. Used in sharing the progress of the code is a cloud-based platform mainly where all the codes are viewed and tracked for progress in the group. Used by everyone in the group to get an understanding of the code as well as discuss and provide solutions in the case of any error occurring within the codes.

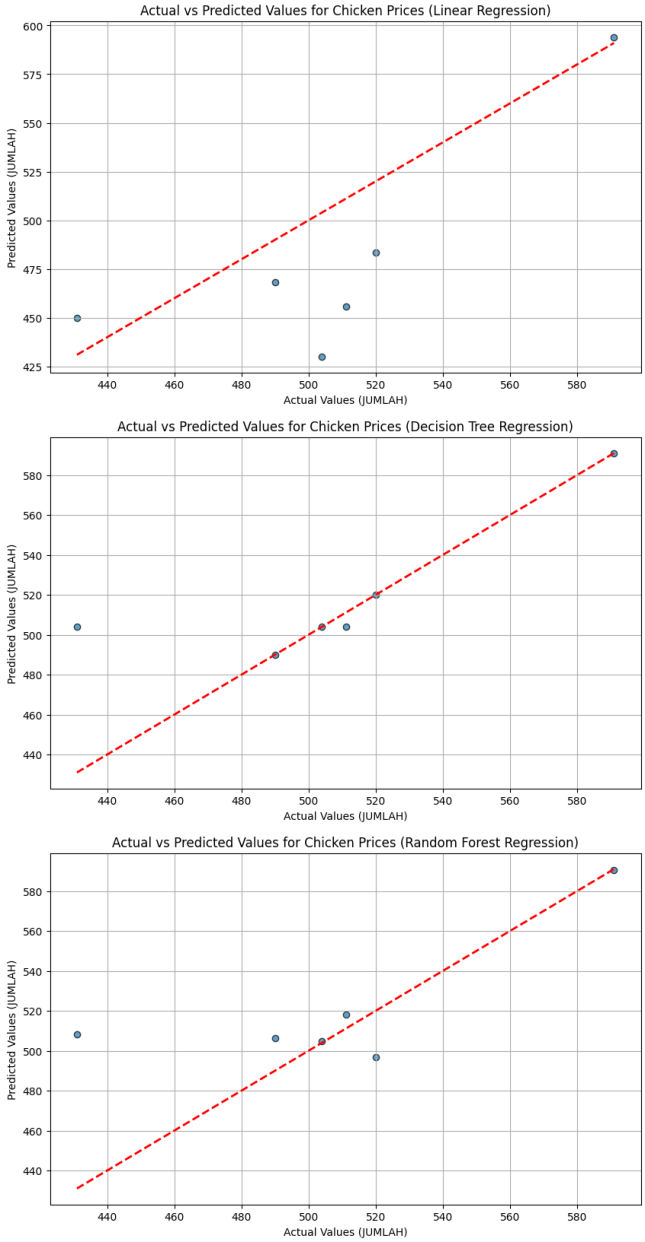

Figure 3. Graphs showing results of each machine learning method by chicken prices.

 doi:10.20944/preprints202502.1861.v1

Table 2. Results of the machine learning models that were utilized for chicken prices.

Model	Mean Squared Error	R2 score	Mean Absolute Error
Linear Regression	1780.37	0.20	34.88
Decision Tree Regression	894.86	0.60	13.38
Random Forest Regression	1135.72	0.49	20.79

As per the results given by the machine learning models, it is predicted that there would be a slight increase in rising material costs, with the Decision Tree Regression (DTR) showing how the values predicted for overall material costs in general returns with an either slightly higher or the same actual value. The DTR model is the most accurate out of all of the three machine learnings models used, with an R2-score of 0.91, or about 91%, followed by the RFR score of about 0.87 (87%), and the LR model at around 0.79 (79%). This similar trend can be said for the data frame that looks into chicken prices in specific, where the evaluation metrics shows that the DTR model is the most accurate with an R2- score of 0.60 (60%), followed by the RFR's score at 0.40 (49%), and the LR's score at 0.20 (20%). However, it is to note that these metrics are lower than that of raw materials in general, as the filtered data frame for chicken may be suffering from underfitting or overfitting.

Cost efficiency is one of the many concerns businesses, large and small, sought to seek and retain to maintain competitiveness and sustainability in the long term. To ensure lower raw material costs for better efficiency and better long- term sustainability, several strategies can be taken, such as diversifying supplies to depend not only from just any single supplier, adjust prices in response to fluctuating raw material prices to maintain customer loyalty and meet expectations, and empowerment of business owners through technology to find and learn the best uses of raw materials to reduce waste while also ensuring high profitability, such as using the internet to find higher-quality raw materials to provide high quality products and services. (Wilson, 2024)

V. Conclusions

Technology can help with aiding small business owners to succeed in where they can. Small businesses should not be stuck in a cycle of complacency despite its perceived comfort, but by taking the risk with great caution and full understanding, it can yield the best results.

Machine learning has its benefits to help uplift people of different backgrounds, regardless of race, religion, and socioeconomic class. Machine learning aids in the assistance of small business owners to understand the performance of their business in the terms of how underlying factors affect their profitability and their sustainability to how they operate their businesses. With machine learning, it can help predict outcomes that can aid in decision-making to help small business owners innovate new ideas on how to run businesses.

With that in mind, the possibilities are endless for many small to medium enterprises across the country, uplifting owners and the community and enable them social mobility to ensure better standards and practices through the SME world in Malaysia.

Acknowledgment: This paper is for a project under Impact Lab Digital Innovation & Smart Society in Taylor's University, Malaysia.

 doi:10.20944/preprints202502.1861.v1

References

1. Ahmed, M. W. (2023, August 24). Understanding mean absolute error (MAE) in regression: A practical guide. *Medium.* https://medium.com/@m.waqar.ahmed/understanding-mean-absolute-error-mae-in-regression-a-practical-guide-26e80ebb97df
2. AIML. (2024). What are the advantages and disadvantages of random forest? *AIML.* https://aiml.com/what-are-the-advantages-and-disadvantages-of-random-forest/
3. Arinze, C. P. (2024, November 25). Effective strategies for handling noisy data in machine learning. *Medium.* https://medium.com/@InsightCoder/effective-strategies-for-handling-noisy-data-in-machine-learning-79f02f216b63
4. Brownlee, J. (2020). Train-test split for evaluating machine learning algorithms. *Machine Learning Mastery.* https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/
5. García, S., Luengo, J., & Herrera, F. (2014). *Data preprocessing in data mining.* Springer International Publishing.
6. GeeksforGeeks. (2024). Linear regression in machine learning. *GeeksforGeeks.* https://www.geeksforgeeks.org/ml-linear-regression/#cost-function-for-linear-regression
7. GeeksforGeeks. (2024). Decision tree. *GeeksforGeeks.* https://www.geeksforgeeks.org/decision-tree/
8. GeeksforGeeks. (2024). Random forest regression in Python. *GeeksforGeeks.* https://www.geeksforgeeks.org/random-forest-regression-in-python/
9. Grammarly. (2024). What is linear regression in machine learning? *Grammarly.* https://www.grammarly.com/blog/ai/what-is-linear-regression/
10. Gupta, A. (2024). Mean squared error: Overview, examples, concepts, and more. *Simplilearn.* https://www.simplilearn.com/tutorials/statistics-tutorial/mean-squared-error
11. IBM. (n.d.). What is linear regression? *IBM.* https://www.ibm.com/topics/linear-regression
12. IBM. (n.d.). What is random forest? *IBM.* https://www.ibm.com/think/topics/random-forest
13. Kanade, V. (2023). What is linear regression? Types, equation, examples, and best practices for 2022. *Spiceworks.* https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/
14. Raut, O. (2022, January 6). Data pre-processing - Machine learning concepts. *Medium.* https://medium.com/machine-learning-concepts/data-pre-processing-edfc1eecd0c2
15. Taylor, S. (2024). R-Squared. *Corporate Finance Institute.* https://corporatefinanceinstitute.com/resources/data-science/r-squared/
16. Viswa. (2024). Unveiling decision tree regression: Exploring its principles, implementation. *Medium.* https://medium.com/@vk.viswa/unveiling-decision-tree-regression-exploring-its-principles-implementation-beb882d756c6
17. Wilson, R. (2024, April 17). Strategies to mitigate high raw material cost impact. *AuditComply.* https://www.auditcomply.com/2024/04/17/strategies-to-mitigate-high-raw-material-cost-impact/
18. Gill, S. H., Razzaq, M. A., Ahmad, M., Almansour, F. M., Haq, I. U., Jhanjhi, N. Z., ... & Masud, M. (2022). Security and privacy aspects of cloud computing: a smart campus case study. Intelligent Automation & Soft Computing, 31(1), 117-128.
19. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In Employing Recent Technologies for Improved Digital Governance (pp. 151-163). IGI Global.
20. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications (pp. 49-64). IGI Global.
21. Lee, S., Abdullah, A., & Jhanjhi, N. Z. (2020). A review on honeypot-based botnet detection models for smart factory. International Journal of Advanced Computer Science and Applications, 11(6).
22. Attaullah, M., Ali, M., Almufareh, M. F., Ahmad, M., Hussain, L., Jhanjhi, N., & Humayun, M. (2022). Initial stage COVID-19 detection system based on patients' symptoms and chest X-ray images. Applied Artificial Intelligence, 36(1), 2055398.
23. Aldughayfiq, B., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2023). Explainable AI for retinoblastoma diagnosis: interpreting deep learning models with LIME and SHAP. Diagnostics, 13(11), 1932.

 doi:10.20944/preprints202502.1861.v1

24. Kumar, M. S., Vimal, S., Jhanjhi, N. Z., Dhanabalan, S. S., & Alhumyani, H. A. (2021). Blockchain based peer to peer communication in autonomous drone operation. Energy Reports, 7, 7925-7939.
25. Aherwadi, N., Mittal, U., Singla, J., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2022). Prediction of fruit maturity, quality, and its life using deep learning algorithms. Electronics, 11(24), 4100.
26. Jena, K. K., Bhoi, S. K., Malik, T. K., Sahoo, K. S., Jhanjhi, N. Z., Bhatia, S., & Amsaad, F. (2022). E-learning course recommender system using collaborative filtering models. Electronics, 12(1), 157.
27. Jhanjhi, N. Z., Humayun, M., & Almuayqil, S. N. (2021). Cyber security and privacy issues in industrial internet of things. Computer Systems Science & Engineering, 37(3).
28. Babbar, H., Rani, S., Masud, M., Verma, S., Anand, D., & Jhanjhi, N. (2021). Load balancing algorithm for migrating switches in software-defined vehicular networks. Comput. Mater. Contin, 67(1), 1301-1316.
29. Alferidah, D. K., & Jhanjhi, N. Z. (2020, October). Cybersecurity impact over bigdata and iot growth. In 2020 International Conference on Computational Intelligence (ICCI) (pp. 103-108). IEEE.
30. Alkinani, M. H., Almazroi, A. A., Jhanjhi, N. Z., & Khan, N. A. (2021). 5G and IoT based reporting and accident detection (RAD) system to deliver first aid box using unmanned aerial vehicle. Sensors, 21(20), 6905.
31. Alex, S. A., Jhanjhi, N. Z., Humayun, M., Ibrahim, A. O., & Abulfaraj, A. W. (2022). Deep LSTM model for diabetes prediction with class balancing by SMOTE. Electronics, 11(17), 2737.
32. Fatima-tuz-Zahra, N. Jhanjhi, S. N. Brohi, N. A. Malik and M. Humayun, "Proposing a Hybrid RPL Protocol for Rank and Wormhole Attack Mitigation using Machine Learning," 2020 2nd International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2020, pp. 1-6, doi: 10.1109/ICCIS49240.2020.9257607.