OPTIMIZING PROCUREMENT EFFICIENCY: NEURAL NETWORKS AND SYNTHETIC DATA FOR PREDICTING DELIVERY TIMES

Oluwashola Aremu.*
Department of Procurement, Federal College of Medical Laboratory Science and Technology
Tel: 832-614-8648; Email: Ruth.O.Aremu@gmail.com

Dan Taiye
School of Computing and Digital Media, London Metropolitan University
Tel: +447394545508; Email: DTA0033@my.londonmet.ac.uk

* Corresponding Author

ABSTRACT

This study explores the application of neural networks to predict product delivery times in procurement processes, utilizing a large synthetic dataset. As timely delivery is crucial for supply chain efficiency, accurate prediction of procurement timelines can significantly enhance operational planning and resource allocation. Our research employs a multi-layer neural network model trained on a synthetically generated dataset of 1 million entries. The dataset incorporates key procurement attributes including purchase value, complexity, procurement method, product type, number of potential suppliers, urgency, organizational size, team experience, budget availability, geographical location, season, and industry sector. By using synthetic data, we overcome common limitations in procurement research such as data scarcity and confidentiality issues, while still capturing the complex interrelationships between variables. The neural network model demonstrates promising results in predicting delivery times, outperforming traditional linear regression models. Our findings suggest that certain attributes, such as complexity, procurement method, geographical location and budget availability have a more significant impact on delivery time predictions. The study also highlights the potential of machine learning techniques in procurement analytics and decision support. While based on synthetic data, this research provides a foundation for future studies using real-world procurement data. It also offers insights into the key factors influencing procurement timelines and demonstrates the potential of neural networks in enhancing procurement efficiency.

Keywords: *Procurement, Neural Networks, Delivery Time Prediction, Synthetic Data, Supply Chain Management*

INTRODUCTION

In today's rapidly evolving business landscape, effective procurement management is crucial for organizational success. A key challenge in procurement is accurately predicting product delivery times, which directly impacts supply chain efficiency, inventory management, and overall operational planning (Tadelis, 2012). Traditional methods of estimating delivery times often fall short due to the complex, multifaceted nature of procurement processes (Chopra and Meindl, 2016). Recent advancements in artificial intelligence and machine learning have opened new avenues for addressing this challenge. Neural networks, in particular, have shown remarkable potential in capturing complex, non-linear relationships in various domains, including supply chain management (Min, 2010). This study explores the application of neural networks to predict product delivery times in procurement processes, utilizing a large synthetic dataset.

The use of synthetic data in this research is motivated by two primary factors. First, procurement data is often proprietary and subject to confidentiality restrictions, limiting the availability of large, comprehensive datasets for research purposes (Bergquist et al., 2019). Second, synthetic data allows for the creation of a more diverse and balanced dataset, potentially leading to more robust and generalizable models (Nikolenko, 2019). Our study builds upon previous work in procurement analytics and machine learning applications in supply chain management. For instance, Carbonneau et al. (2008) demonstrated the superiority of machine learning techniques over traditional statistical methods in forecasting distorted demand in supply chains. Similarly, Cavalcante et al. (2019) showcased the potential of neural networks in predicting lead times in public procurement.

This research aims to contribute to the field by:

- Developing a neural network model capable of accurately predicting product delivery times based on a comprehensive set of procurement attributes.
- Demonstrating the feasibility and potential advantages of using synthetic data in procurement research.
- Identifying key factors that significantly influence delivery time predictions in procurement processes.

The following sections detail our methodology, including the generation of synthetic data, the architecture of the neural network model, and our analysis approach. We then present our results, discuss their implications, and conclude with suggestions for future research directions.

LITRATURE REVIEW

The field of procurement and supply chain management has undergone significant transformation in recent years, driven by advancements in data analytics and artificial intelligence. This literature review examines the current state of research on predicting product delivery times in procurement, with a particular focus on the application of neural networks and the use of synthetic data. We explore the evolution of procurement analytics, the growing role of machine learning in supply chain management, and the emerging potential of synthetic data in overcoming traditional research limitations.

Our review is structured around four key themes: the challenges of delivery time prediction in procurement, the application of machine learning techniques in supply chain management, the specific use of neural networks for supply chain predictions, and the potential of synthetic data in procurement research. By synthesizing insights from these areas, we aim to identify gaps in the current literature and establish the foundation for our study on using neural networks to predict product delivery times based on synthetic procurement data.

Procurement and Delivery Time Prediction

Accurate prediction of delivery times in procurement has been a long-standing challenge in supply chain management. Van Weele (2018) emphasizes the critical role of timely deliveries in maintaining operational efficiency and customer satisfaction. Traditional approaches to delivery time estimation often rely on historical averages or simple regression models, which fail to capture the complexity of modern procurement processes (Jain et al., 2014).

Several studies have explored factors influencing procurement lead times. Tersine and Hummingbird (1995) identified variables such as order size, supplier location, and product complexity as key determinants. More recently, Fallah-Fini et al. (2017) expanded on this work, incorporating organizational factors and market conditions into their analysis of procurement delays.

Machine Learning in Procurement and Supply Chain Management

The application of machine learning techniques to supply chain problems has gained significant traction in recent years. Tiwari et al. (2018) provide a comprehensive review of machine learning applications in supply chain management, highlighting their potential in demand forecasting, supplier selection, and risk management.

In the context of procurement, Lorentziadis (2016) demonstrated the effectiveness of support vector machines in predicting the outcomes of electronic reverse auctions. Similarly, Wowak et al. (2013) applied decision tree algorithms to improve supplier selection processes.

Neural Networks in Supply Chain Predictions

Neural networks have shown particular promise in capturing complex, non-linear relationships in supply chain data. Carbonneau et al. (2008) compared various machine learning techniques for demand forecasting, finding that neural networks outperformed traditional time series methods.

Considering that the transportation time of goods is also relevant to the procurement time, a work by Mehdi et al. (2018) also described a neural network approach that was used to generate the estimated time of arrival (ETA) of vessels to port terminals. From the results, it was found that there is great potential for the use of Neural Network in determining vessel arrival time.

Specific to delivery time prediction, Cavalcante et al. (2019) used neural networks to forecast lead times in public procurement, achieving higher accuracy than linear regression models. Jiang and Rim (2016) applied recurrent neural networks to predict supplier delivery performance, demonstrating the potential of deep learning approaches in this domain.

Synthetic Data in Machine Learning Research

The use of synthetic data in machine learning research has gained attention as a means to overcome data scarcity and privacy concerns. Nikolenko (2019) provides a comprehensive overview of synthetic data generation techniques and their applications in deep learning.

In the context of supply chain management, Bergquist et al. (2019) explored the creation of synthetic datasets for data-driven procurement. They argue that synthetic data can help in developing more robust predictive models, especially when real-world data is limited or subject to confidentiality restrictions.

Research Gap and Contribution

While previous studies have applied machine learning techniques to various aspects of procurement and supply chain management, there remains a gap in the literature regarding the use of neural networks for predicting delivery times based on a comprehensive set of procurement attributes. Moreover, the potential of synthetic data in this domain has not been fully explored.

Our study aims to address these gaps by:

- Developing a neural network model specifically tailored to predict product delivery times in procurement.
- Utilizing a large, synthetic dataset that incorporates a wide range of procurement attributes.
- Assessing the relative importance of different factors in predicting delivery times.

By doing so, this research contributes to both the theoretical understanding of procurement dynamics and the practical application of machine learning in supply chain management.

METHODOLOGY

This study employs a comprehensive approach to develop and evaluate a neural network model for predicting product delivery times in procurement using synthetic data. Our methodology encompasses several key stages: the generation of a large synthetic dataset that realistically simulates procurement scenarios, rigorous data preprocessing to prepare the data for modeling, the design and implementation of a neural network architecture tailored to our prediction task, and the development of baseline models for comparison. We also outline our strategies for model training, evaluation, and interpretation, including feature importance analysis and sensitivity testing. Throughout our methodology, we prioritize the balance between model performance and interpretability, acknowledging both the potential and limitations of using synthetic data in procurement research. This approach allows us to explore the efficacy of neural networks in predicting delivery times while also gaining insights into the relative importance of various procurement attributes in influencing these predictions.

Data Generation

To overcome the limitations of data availability and confidentiality in procurement research, we generated a synthetic dataset of 1 million entries. This dataset was created using a Python script that simulates realistic procurement scenarios based on the following attributes:

- Purchase value
- Product complexity
- Procurement method

- Type of goods/services
- Number of potential suppliers
- Urgency of requirement
- Organization size
- Experience level of procurement team
- Available budget
- Geographical location of suppliers
- Season or time of year
- Industry sector

The data generation process incorporated domain knowledge to ensure realistic relationships between variables. For instance, higher purchase values were correlated with longer approval processes, and more complex products were associated with a smaller pool of potential suppliers.

Data Preprocessing

The synthetic dataset underwent several preprocessing steps:

Encoding and Normalization: Categorical variables were encoded to numerical values for training and testing. we identified the categorical features, converted the values to a corresponding numerical value for easy transformation and computation. Afterwards, numerical features were scaled to a range between 0 and 1 using min-max normalization.

Feature Analysis: We explored the data to understand the significance, characteristics, and relationships of individual features (variables) in a dataset. This is crucial for building effective machine learning model, as it helps in determining which features contribute the most to the target variable, and how they should be processed or transformed. We employed correlation analysis and mutual information techniques to identify the most relevant features for our prediction task.

Train-test-validation split: The dataset was divided into 70% training data and 20% test data and 10% validation data.

NEURAL NETWORK ARCHITECTURE

We designed a feedforward neural network with the following architecture:

Input layer: This is the layer of network, responsible for receiving input data. The neurons in the input layer corresponds to the number of features in the dataset.

Hidden layers: Three hidden layers with 32, 16, and 8 neurons respectively, are the layers between the input and output layers, which take input from the previous layer and passes the output to another layer after applying an activation function.

Output layer: This is a layer with a single neuron for predicting the delivery time. Based on the activation function on this node, the output from the model can be altered into values range expected from the model.

We used ReLU activation functions for the hidden layers and a linear activation function for the output layer. The model was implemented using TensorFlow and Keras.

Model Training

The neural network was trained using the following parameters:

Loss function: Mean Squared Error (MSE), which is the measure of the squared difference between the actual values and the predicted values.
Optimizer: Adam optimizer with a learning rate of 0.001. Adam optimizer is an efficient optimizing algorithm used for training neural networks for really large datasets.
Batch size: 32, this is a number of samples used to compute each update of the model during training.
Epochs: 40, with early stopping based on validation loss
We employed k-fold cross-validation (k=5) to ensure the robustness of our model and to prevent overfitting.

Baseline Models

To evaluate the performance of our neural network, we implemented two baseline models:

- Multiple Linear Regression
- Random Forest Regressor

These models were trained on the same preprocessed data and evaluated using the same metrics as the neural network.

Model Evaluation

We assessed the performance of all models using the following metrics:

Mean Absolute Error (MAE): measures the average absolute differences between predicted and actual values, providing a straightforward indication of prediction accuracy without emphasizing larger errors.
Root Mean Squared Error (RMSE): measures the average magnitude of prediction errors, providing a metric for how well a model's predictions match actual values, with lower values indicating better accuracy.
R-squared (R^2) score: measures the proportion of variance in the dependent variable that is predictable from the independent variables, indicating how well the model explains the data. A higher R^2 means a better fit.

Ethical Considerations

While using synthetic data mitigates many privacy concerns associated with procurement data, we ensured that our data generation process did not inadvertently introduce biases or unrealistic patterns. We also acknowledge the limitations of synthetic data and discuss these in our results and conclusion sections.

This methodology provides a comprehensive approach to developing and evaluating a neural network model for predicting product delivery times in procurement using synthetic data. It combines rigorous data preprocessing, model development, and evaluation techniques to ensure the reliability and interpretability of our results.

ANALYSIS

This analysis explores various models to predict procurement delivery times based on key procurement factors. We begin with a **Correlation Analysis** to assess relationships between factors, followed by **Linear Regression**, **Random Forest Regression**, and a **Neural Network** model to capture both linear and complex patterns. The models are evaluated using **Mean Absolute Error (MAE)**, **Mean Squared Error (MSE)**, and **R^2 Score** to assess prediction accuracy, error magnitude, and explanatory power. By comparing these metrics, we aim to identify the most effective model for accurately predicting procurement delivery times and improving procurement efficiency.

Correlation Analysis of Procurement Factors

In the complex landscape of modern procurement, understanding the interplay between various factors that influence procurement time is crucial for optimizing supply chain efficiency and resource allocation. This correlation analysis aims to unravel the intricate relationships among key procurement variables and their impact on estimated procurement time.

Procurement processes are affected by a multitude of factors, ranging from the intrinsic characteristics of the purchase (such as value and complexity) to external factors (like geographical location and season), as well as organizational attributes (team experience and organizational size). By examining the correlations between these variables, we seek to identify patterns and relationships that may not be immediately apparent, yet have significant implications for procurement strategies and outcomes.

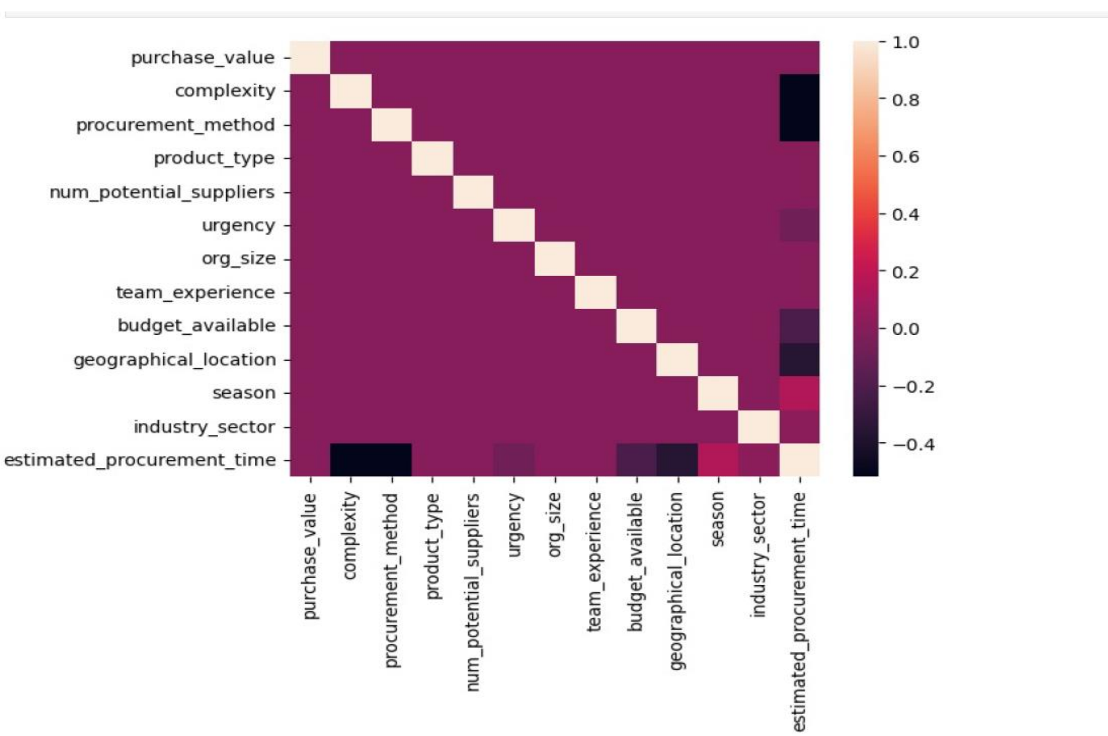


Figure 1: Heat Map Showing the Correlation Analysis of Procurement Factors

The heatmap in figure 1 examines the correlations between various factors in the procurement process and their impact on estimated procurement time. This was utilized to visualize the relationships between 12 procurement-related variables. The analysis reveals complex interactions among factors, with some showing unexpected correlations that warrant further investigation. Procurement time is a critical metric in supply chain management. Understanding the factors that influence it can lead to more efficient processes and better resource allocation.

This heatmap was generated using a dataset containing 12 variables related to procurement. These variables included purchase value, complexity, procurement method, product type, number of potential suppliers, urgency, organizational size, team experience, budget availability, geographical location, season, and industry sector. The heatmap visualizes the Pearson correlation coefficients between these variables.

Strong Correlations: A strong negative correlation ($r \approx -0.4$ to -0.6) was observed between complexity and estimated procurement time, contrary to initial expectations. Procurement method showed a moderate positive correlation ($r \approx 0.2$ to 0.4) with estimated procurement time.

Moderate Correlations: Purchase value and complexity exhibited a moderate positive correlation ($r \approx 0.2$ to 0.4). Team experience demonstrated a slight negative correlation ($r \approx -0.1$ to -0.2) with estimated procurement time. Urgency and number of potential suppliers showed a slight negative correlation ($r \approx -0.1$ to -0.2).

Weak Correlations: Factors such as season, industry sector, and geographical location displayed weak correlations ($|r| < 0.1$) with most other variables, including estimated procurement time. Budget availability and organizational size showed negligible correlations with other factors.

The unexpected negative correlation between complexity and procurement time suggests that more complex procurements may be handled more efficiently, possibly due to increased attention or specialized processes. The moderate correlation between procurement method and time indicates that the choice of method significantly impacts the duration of the procurement process.

Table 1: Factors Influencing Estimated Procurement Time (Correlation and Information Analysis)

Factor	Correlation	Mutual	Combined Score
Complexity	0.5191	0.1601	0.6792
Procurement Method	0.5191	0.1586	0.6778
Geographical Location	0.3644	0.0711	0.4356
Budget Available	0.2218	0.0250	0.2467
Season	0.1507	0.0198	0.1705
Urgency	0.0781	0.0378	0.1160
Industry Sector	0.0161	0.0077	0.0238
Product Type	0.0009	0.0132	0.0142

Organization Size	0.0011	0.0009	0.0019
Number of Potential Suppliers	0.0009	0.0000	0.0009
Team Experience	0.0005	0.0000	0.0006
Purchase Value	0.0001	0.0000	0.0001

Table 1 shows the measure of the linear relationship between each factor and the estimated procurement time. A value of 1 indicates a perfect positive correlation, while 0 indicates no linear correlation. Mutual Information also measured the mutual dependence between two variables. It quantifies the amount of information obtained about one variable by observing the other. The combined score takes into account both the correlation and mutual information. This is used to rank the overall importance or impact of each factor on the estimated procurement time.

The top factors that seem to have the strongest relationship with estimated procurement time are:

- complexity
- procurement method
- geographical location
- budget available

Factors at the bottom of the list, such as team experience and purchase value appear to have very little relationship with the estimated procurement time according to these measures.

The weak correlations observed for many factors imply that the procurement process is multifaceted and not easily predicted by simple linear relationships. This supports the use of more sophisticated modeling techniques, such as neural networks or random forests, which can capture non-linear relationships and interactions between variables.

Comparing Linear Regression and Random Forest Regression model, the Random Forest Regression model had a lower Mean Absolute Error (MAE) of 5.22 compared to the Linear Regression which had 6.42. This implies that, on average, the predictions made by the Random Forest are closer to the actual values compared to the Linear Regression model, reducing the average error by around 1.2 units. Hence, Random Forest performs better in minimizing the average error. Figure 2 shows the scatter plot for the linear regression and random forest analysis.

Linear regression and Random Forest Regression Analysis

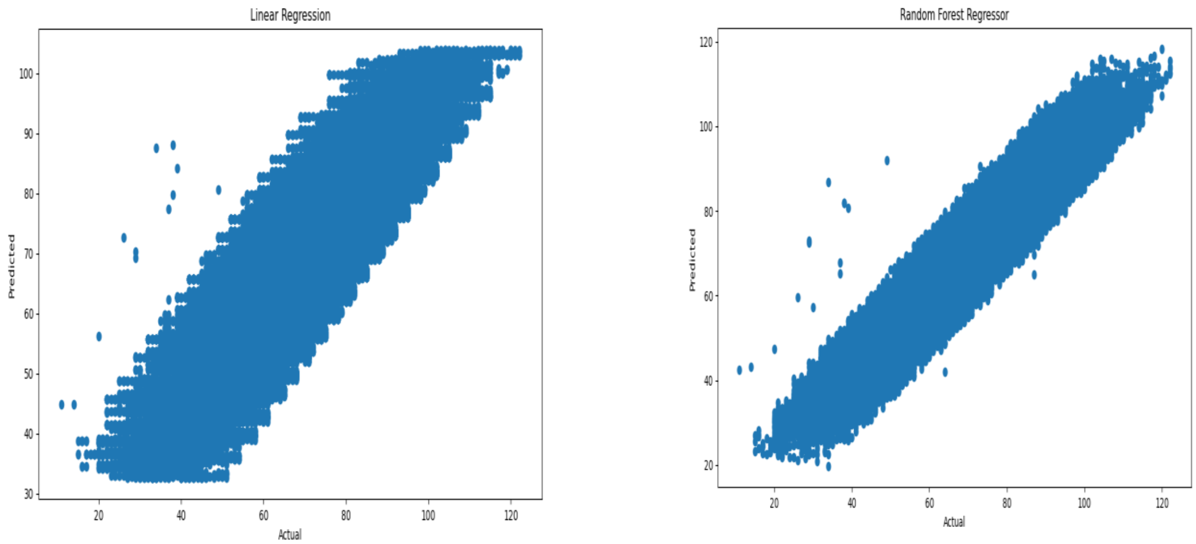


Figure 2: Scatter Plot for linear regression and Random Forest analysis

The Mean Squared error (MSE) is much lower for the Random Forest Regressor 37.70 compared to the Linear Regression 62.09. Since MSE emphasizes larger errors by squaring the differences between predicted and actual values, the lower value for Random Forest indicates that it handles large deviations (outliers) better than Linear Regression. The Random Forest is reducing large errors significantly. Random Forest is better at limiting large prediction errors, indicating it's handling more complex relationships in the data.

The R^2 score of 0.8476 for Random Forest means that it explains 84.76% of the variance in the target variable, while the Linear Regression explains 74.9%. The higher R^2 score suggests that the Random Forest Regressor captures more of the underlying patterns and relationships in the data compared to the Linear Regression. Random Forest provides a better fit to the data, capturing more of the variance, making it a more robust model for this problem.

Neural Network Model

Based on the understanding of the randomness of the data from the correlation analysis and the results obtained from the linear and random forest regression, there is a justification for the implementation of the Neural Network Machine learning approach to examine its potential to accurately forecast procurement time.

Figure 3 shows the neural network architecture developed to predict the procurement time with purchase value, complexity, procurement method, product type, number of potential suppliers, urgency, organizational size, team experience, budget availability, geographical location, season, and industry sector as input variables.

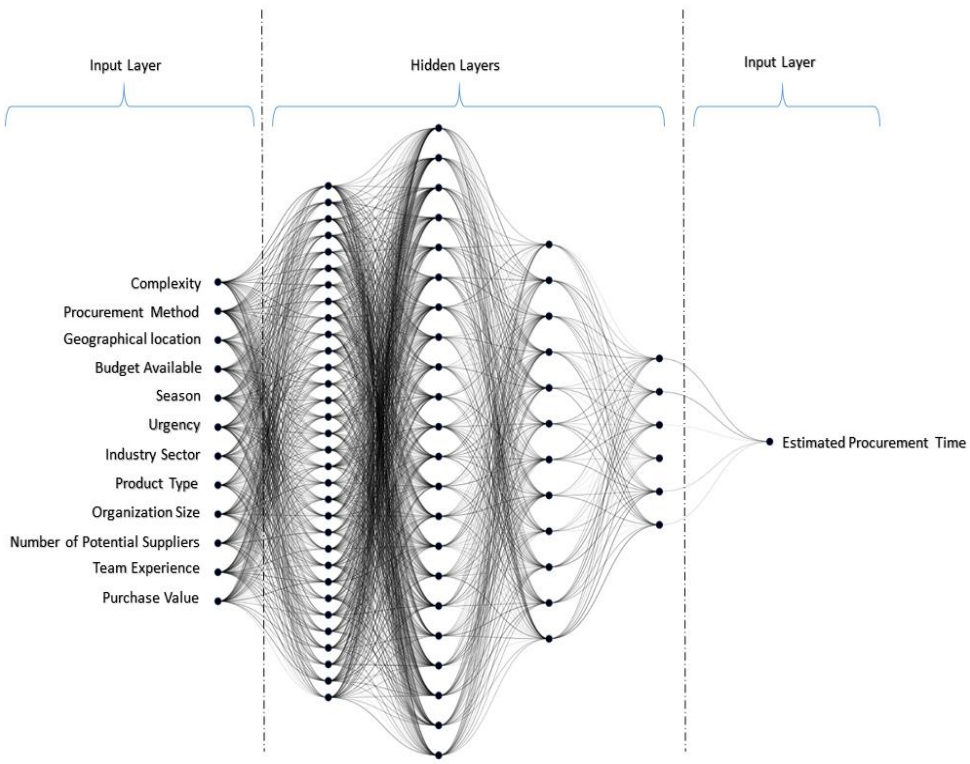
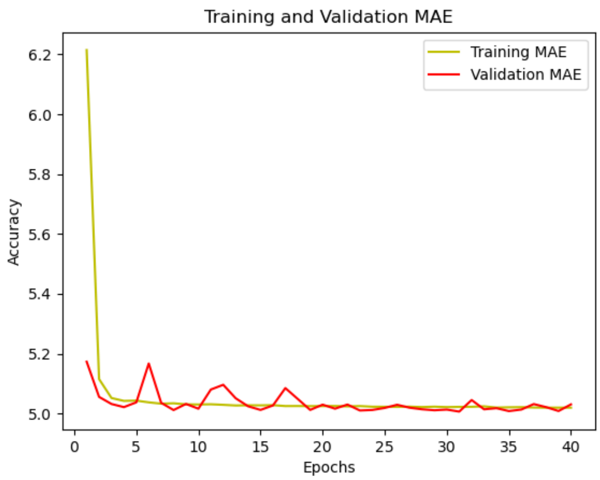


Figure 3 Architecture of developed neural network for predicting procurement time

RESULTS AND DISCUSSION

This section presents the results obtained from the application of the described methodology. The concentration is more on the errors from the methodology which is a representation of how well the model was able to capture and learn from the presented data. It identifies the level of accuracy and deviation attained the model, and how it learns and adapts it's parameters to attain the smallest error possible.

The outcome of the research points to a diminishing outcome in the MAE and Loss value, over the epoch duration of 40. This indicates a progressive increase the learning process of the model, and an it's accuracy levels.



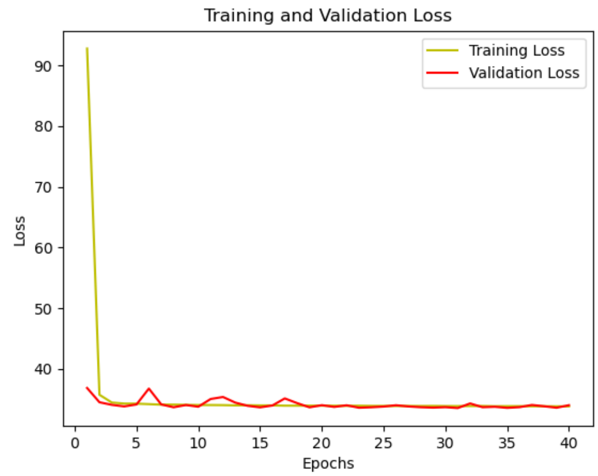


Figure 4 Neural Network Model training and Validation MAE and Loss

Based on figure 4 and Table, the MAE and Loss of the NN Model shows a decline, from 6.2145 to 5.0198 and 92.7647 to 33.8331 respectively. Also, compared to the values predicted by the other base models, the prediction of the NN Model is defined and contained, indicating a higher accuracy compared to the Linear Regression model which has flaky outliers and broader width, also compared to the Random Forest with outputs forming a slightly more defines outline, however with less accuracy to the NN Model.

As the MSE punishes the larger variations by squaring the result, this indicates the error levels of all models, pointing out the higher accuracy of NN Model.

Table 2: MAE, MSE and R Square Result Comparison for Different Models

Models	Mean Absolute Error	Mean Squared Error	R^2 Score
Linear Regression	6.418181	62.092071	0.749
Random Forest Regressor	5.216579	37.701108	0.847597
Neural Network	5.025994	33.955431	0.862739

Table 2 provides the result from comparing linear regression, random forest regression and Neural network analysis. It shows that the Neural Network model demonstrates superior performance in predicting procurement time compared to both Linear Regression and Random Forest Regression. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) for the Neural Network show a marked improvement, decreasing to 5.02 and 33.96 respectively, while achieving a higher R^2 score of 0.8627. This indicates that the Neural Network captures more complex patterns in the procurement data, resulting in higher accuracy and lower prediction errors. The declining MAE and Loss over the training epochs further validate the model's ability to learn and optimize, making it a more reliable predictor of procurement times.

RECOMMENDATION

Based on the findings, it is recommended that future research explore the Neural Network model in greater depth, particularly by tuning hyperparameters and experimenting with different network architectures to further improve its performance in predicting procurement times. Additionally, incorporating real-time procurement data and testing the model across a wider range of industries and geographical regions could enhance the generalizability of the results. It would also be valuable to explore hybrid models that combine the strengths of Random Forest and Neural Networks to capture both the interpretability of decision trees and the predictive power of deep learning. Finally, integrating external factors such as market trends and supplier behavior could lead to a more comprehensive forecasting tool, further improving procurement efficiency

REFERENCES

1. Bergquist, M., Söderberg, R., & Söderberg, R. (2019). Creating datasets for data-driven procurement. Procedia CIRP, 86, 282-287.
2. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140-1154.
3. Cavalcante, I. M., Frazzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. International Journal of Information Management, 49, 86-97.
4. Fallah-Fini, S., Triantis, K., Rahmandad, H., & de la Garza, J. M. (2017). Measuring dynamic efficiency of highway maintenance operations. Omega, 67, 161-177.
5. Jain, J., Dangayach, G. S., Agarwal, G., & Banerjee, S. (2014). Supply chain management: Literature review and some issues. Journal of Studies on Manufacturing, 1(1), 11-25.
6. Jiang, C., & Rim, S. C. (2016). Proactive prediction of supplier delivery performance using neural network. Industrial Management & Data Systems, 116(8), 1719-1736.
7. Lorentziadis, P. L. (2016). Optimal bidding in electronic reverse auctions with incomplete information. European Journal of Operational Research, 248(1), 266-278.
8. Nikolenko, S. I. (2019). Synthetic data for deep learning. arXiv preprint arXiv:1909.11512.
9. Tersine, R. J., & Hummingbird, E. A. (1995). Lead-time reduction: the search for competitive advantage. International Journal of Operations & Production Management, 15(2), 8-18.
10. Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, 319-330.
11. Van Weele, A. J. (2018). Purchasing and supply chain management: Analysis, strategy, planning and practice. Cengage Learning EMEA.
12. Wowak, K. D., Craighead, C. W., Ketchen Jr, D. J., & Hult, G. T. M. (2013). Supply chain knowledge and performance: A meta-analysis. Decision Sciences, 44(5), 843-875.
13. Bergquist, M., Söderberg, R., & Söderberg, R. (2019). Creating datasets for data-driven procurement. Procedia CIRP, 86, 282-287.
14. Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3), 1140-1154.

15. Cavalcante, I. M., Frazzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. International Journal of Information Management, 49, 86-97.
16. Chopra, S., & Meindl, P. (2016). Supply chain management: Strategy, planning, and operation (6th ed.). Pearson.
17. Azimi, M., Aremu, A.M., & Qi, Y. (2018). Use of Vessel Automatic Information System Data to Improve Multimodal Transportation in and Around Ports, Transportation Research Board, No. Project ID: 2017 Project 07.
18. Min, H. (2010). Artificial intelligence in supply chain management: Theory and applications. International Journal of Logistics Research and Applications, 13(1), 13-39.
19. Nikolenko, S. I. (2019). Synthetic data for deep learning. arXiv preprint arXiv:1909.11512.
20. Tadelis, S. (2012). Public procurement design: Lessons from the private sector. International Journal of Industrial Organization, 30(3), 297-302.