A Machine Learning-based decision support system for Fast-Moving Consumer Goods vendor selection

Authors:

1. **XXX**
   XXX
2. **XXX**
   XXX

Abstract

In a competitive business environment, procurement is a profit-contributing activity. This is particularly true for fast-moving consumer goods, where delays and quality issues can result in missed sales. Classically, the main vendor selection factor is price.

This study conducts vendor classification based on performance data to reduce supply chain risk by enabling informed procurement decisions. Employed performance characteristics are product quality, delivery time, communication, reliability, and geographical distance.

Three different prototype classifiers are employed based Random Forest, K-Nearest Neighbour, and Naïve Bayes as classification algorithms. Performance metrics are accuracy, precision, recall, specificity, and f-1 score.

Using an artificial dataset, Random Forest shows the best performance results, followed by Naïve Bayes and K-Nearest Neighbours. Random Forest also exceeds at detecting features with the most impact in the data. The results provide a first step towards the implementation of ML-based vendor classification by indicating a suitable algorithm and evaluating its performance.

Keywords: Supply-chain management, machine learning, vendor selection, artificial intelligence

1 Introduction

In a competitive business environment, making the right choice when it comes to selecting vendors can indicate the difference between success and failure. Businesses are beginning to understand that procurement is more than just placing orders; it is a profit-contributing activity (Moynihan et al.,2006). Furthermore, according to Tan et al. (1998), purchasing practices that consider supplier capabilities are positively and significantly correlated with most performance indicators as well as increases in market share, sales, and return on assets.

Suppliers can be seen as business partners, and the efficiency of many of their operations, such as production or quality control, is heavily dependent on supplier performance. In 2022, the cost of supply chain disruptions in the United Kingdom was estimated to result in revenue losses of $ 142 million (Resilience 2022: Interos Annual Global Supply Chain Report, 2022). This, of course, would be a mix of external factors that could not be avoided, but also internal factors such as poor planning and other performance issues on the supplier’s side. The procurement departments and the entire business would benefit from a tool that utilises supplier behaviour data to improve vendor evaluation.

This is particularly relevant for fast-moving consumer goods (FMCG) companies. FMGC are defined as products that are sold quickly and at a relatively low cost. Examples include non-durable household goods such as packaged foods, beverages, toiletries, candies, cosmetics, over-the-counter drugs, dry goods, and other consumables. For instance, issues such as poor-quality testing media or lack of raw materials for production can cause a delay in the product's release to the market. This delay can lead to missed sales opportunities for the manufacturer as consumers may opt for alternative products from other brands. The longer the delay in product release, the more significant the potential loss of sales for a company.

The aim of this project is to build a machine learning (ML) decision support system (DSD) for vendor classification based on vendor performance. DSD can be defined as a system that is expandable and has built-in flexibility to assist in decision modelling and ad hoc data analysis (Moore, 1980). This type of system uses various data collected for analysis and report production to assist organisational leaders and managers in decision-making and problem-solving. By analysing data on vendor behaviour and product release rates, this system will allow procurement departments to make more informed decisions and reduce the risk of supply chain disruptions.

Large organisations’ procurement departments may lack sufficient information regarding their vendors' actual performance, which often leads them to prioritise securing better discounts. Although this may appear to be a logical business decision aimed at increasing profitability, it is a short-term approach that overlooks other critical factors. To make better-informed decisions, procurement should consider the vendor's impact on the product release (PR) rate and end-user feedback. In the context of this project, the term PR is used to describe the process in the FMCG environment, starting from manufacturing a batch of established product through quality control testing and approval to the point of being ready to be dispatched to retailers.

The aim of this project is to develop a machine learning based decision support system to classify suppliers into categories in terms of their impact on the product release rate in FMCG companies.

2 Literature Review

This literature review covers the current procurement processes and criteria in supplier selection to ensure that DSD fits into purchasing standards and that the most relevant features are selected for the model. It also examines various previously studied ML solutions and their performance results both within and outside of the supplier selection problem to identify the most promising solutions as well as the literature gap in the current research.

2.1 Supplier selection process

The standard supplier selection process comprises four steps (Taherdoost and Brard, 2019):

1. Choice of method of subcontracting (decide if whole or partial contract to be awarded)
2. Pre-qualification of potential suppliers
3. Request for quotation and bid analysis
4. Final supplier selection

De Boer *et al*. (2001) analysed the different traditional methods used for supplier selection. No method was found for problem definition, and very few methods were found for criteria formulation. The next stage, pre-selection, is the initial sorting process without ranking potential vendors. There were several techniques identified as used at this stage: Categorical methods (assigning vendors to 'positive', 'neutral' and 'negative' categories), Data envelopment analysis (analysing suppliers based on efficiency to cost ratio), Cluster analysis and Case-based reasoning systems (software advising users by analysing similar cases from the past). Final-choice phase methods are used to rank vendors. Linear weighting models (focused on assigning weights to criteria), Total cost of ownership (analysis of all quantifiable life-cycle cost of purchased item), various mathematical and statistical models and Artificial Intelligence are discussed in this research. In conclusion, this study points out that the same method is often used without consideration for the context and suggests further research on the suitability of the methods in specific purchasing situations and environments.

2.2 Supplier selection criteria impact on business performance

Tracey and Leng Tan (2001) analysed the correlation between supplier selection criteria and firm performance. The research was based on questionnaire responses received from 180 subscribers of the 'Industry Week' publication. Participants who were characterised as high-level executives in manufacturing companies were asked to rate which criteria (Unit Price, Quality, Reliability and Performance) were the most significant in their vendor selection decisions. These responses were later linked to companies' own performance self-assessment data, which consisted of growth in sales, return on assets, market share gain, and overall competitive position. The hypothesis that a supplier's pricing will have a significant impact on the dimensions of firm performance was not supported by the evidence in the analysis, and surprisingly, no strong correlation was found between supplier pricing and companies' pricing competitiveness. The second hypothesis, stating that vendor selection based on the three remaining criteria (Quality, Reliability and Performance) is positively correlated with stronger firm performance, was found to be supported.

Kanan and Choon Tan (2002) conducted a similar correlation-based study looking at the link between supplier assessment criteria and company performance measures, such as Market

Share, Return on Assets, Product Quality and Competitive Position. There was a strong positive correlation between supplier responsiveness and return on assets and between suppliers' delivery and service and companies' product quality. These results suggest that these criteria are worth including in decision support systems for vendor selection.

2.3 Machine learning classifiers and their performance analysis

The use of machine learning to support supplier selection decisions has been attempted more often in recent years, as machine learning popularity has grown. However, the literature findings may not provide a full picture of current developments in this area because it can be assumed that businesses develop their own internal systems that are not shared with the public.

Omurca (2013) proposed a hybrid system of fuzzy c-means (FCM) and rough set theory (RST)to solve the selection problem. The first step is to cluster vendors using the FCM algorithm and label clusters. The RST is then used to determine the importance of different evaluation criteria and extract decision rules. The following criteria were included in the test dataset: quality management practices and systems, self-audit, process/manufacturing capability, management of the firm, design and development capabilities, cost reduction capability, quality performance, price performance, delivery performance, cost reduction performance and other. The model grouped the suppliers into three clusters. Cluster 1 members were the best performers who could be chosen for long-term relationships. Vendors from Cluster 2 could be selected if some improvements to their performance have been made and Cluster 3 contained candidates for pruning. One of the main benefits of this system is the visibility of features to improve, which can then be passed on as feedback to vendors in Clusters 2 and 3.

Harikrishnakumar *et al.* (2019) used and evaluated number of classification algorithms such as support vector machines (SVM), logistic regression, k-nearest neighbours (KNN), and Naïve Bayes (NB) methods in the context of supplier assessment. The following attributes have been used for vendor evaluation: quality audits (a systematic review to assess whether quality process comply with existing procedure of the organisation), price (the total cost of goods and logistics), service (ensures supply of goods to meet the customer demands), reliability (ability to provide supplies to crucial activities in supply chain), sustainability (management of social, economic, and environmental aspects of supply chain), geographic location (the supplier and buyer are within close proximity to ensure on-time delivery) and delivery (capability to maintain accurate delivery schedules). In the test, SVM and logistic regression provided the most accurate models, with polynomial SVM and NB methods ranking the lowest. The importance of the attributes was ranked in the following order: geographic location, delivery, price, quality audits, service, sustainability, and reliability.

Wilson *et al*. (2020) proposed Random Forest (RF) model for vendor evaluation in situations where there are many criteria to be considered. In this model, 20 different features are used with weighting criteria assigned according to the Analytical Hierarchical Process (AHP). The proposed system would help to save a substantial amount of time and costs involved in otherwise complex calculations. AHP is a method of measurement in which priority scales are determined based on the judgments of experts (de FSM Russo and Camanho, 2015). Usually, it consists of the following steps:

- Definition of alternatives
- Definition of criteria
- Criteria assessment
- Calculation of criteria weight and priorities
- Consistency analysis

Study of the performance of RF, KNN and NB in the prediction of chronic kidney disease (CKD)by Devika *et al.* (2019) shows that RF results in the highest accuracy and overall was best at the prediction of CKD. Furthermore, Fernandez-Delgado et al (2014) in their study 'Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?' evaluated 179 classifiers from 17 families and concluded that the classifiers most likely to be the bests are the RF versions. However, Wainberg *et al*. (2016) argued that, in the mentioned study, the outcomes do not demonstrate that the best RF outperform the best SVMs and Artificial Neural Networks (ANN) as the accuracy differences between the top-ranked RF and the other top eight models, such as SVM, neural networks, and other RFs, were not statistically significant, according to the authors' paired t-tests. The conclusion that RF are the best classifiers is called into question by this.

The advantage of RF models over the single decision tree (DT) classifier was shown by Esmaily *et al.*(2019) in their large sample size study, where the RF model resulted in a far better level of accuracy shown in the Area Under the Curve (AUC), which is an index of model performance and accuracy based on the proportion of True Positives and False Positives.

The optimal number of trees in RF was researched by Probst et al (2019), who concluded that for most of the examined datasets, the biggest performance gain is achieved when training the first 100 trees. Based on this, it can be assumed that any additional gain in the RF performance achieved by the modification of this number would be insignificant.

Wahyono and Kang-Hyun (2014) compared ANN, KNN, SVM and RF as classification methods for traffic sign recognition. In several tests, RF and KNN showed the highest accuracy; however, KNN was the slowest algorithm. The ANN provided the lowest accuracy and highest error rate. Suksomboon and Ritthipakdee (2022) investigated the differences in performance between RF and KNN and concluded that KNN performs better on a smaller number of instances of data (e.g. Iris and Computer datasets, 105 and 209 instances, respectively), while RF deals better with large datasets (such as Dry Bean and Yeast, 13611 and 1484 instances, respectively).

In a document classification study by Ting *et al.* (2011) NB classifier showed the highest precision, recall, and f1-score compared to SVM, ANN, and DT. This study recommends NB classifiers because of their simplicity, highlighting their computational efficiency, as NB allows each characteristic to influence the final decision equally and independently from other attributes.

2.4 Literature gap

As De Boer *et al.* (2001) suggest, more research is needed to improve supplier selection decisions and processes with consideration for specific contexts. This project is dedicated to the FMCG sector, in which timely and efficient product release is crucial.

Price and discount are included as standard features in most machine learning models for supplier evaluation. This project will attempt a new approach, focusing only on supplier performance in areas related to product release rate, such as delivery, communication, product quality, and other similar factors. This is to ensure that a vendor with a clearly negative impact on company profits will not be suggested as the preferred vendor because they offer good prices. In most cases, the loss caused by supply chain disruption is not worth the discount offered by vendors with unsatisfactory performance. Then, only the best-performing vendors can be selected, and at this stage, the best offers can be compared.

Another point of improvement is the reliability of the collected supplier data. In the ML models above, they originated from purchasing managers. Purchasing managers may not provide a good overview of real supplier performance. Data would be more accurate if collected from each specific department that uses vendors, preferably from end users, people responsible for ordering goods, communicating with vendors for quotes, and receiving deliveries.

This project attempts to answer the following questions:

1. *How would this specific machine learning decision support system for vendor selection be designed?*
2. *What nonmonetary features should be selected as the most appropriate?*
3. *Which algorithm performs best in this context?*

3 Research Methodology

The required DSD needed to be built and evaluated within a relatively short timeframe of a few months, but with good foundations to be reused and developed further. Real-world data collection would be very challenging in these circumstances because of the required time and internal regulations within businesses which protect their internal data and limit external access to them. Therefore, the simulated dataset was used with a careful choice of features which can be reused and partially random numerical values which can be replaced with actual data. These numerical data can be easily measured using quantitative methods and common machine learning performance metrics, giving this project a quantitative nature. The choice of algorithms was aimed at balancing the complexity of the problem and the simplicity of the solution. As a result, simpler suitable methods were chosen over more sophisticated ones.

3.1 Feature selection

Based on the literature review, the variables showing the strongest correlation with companies' good performance outcomes were chosen: 'Quality' and 'Reliability' (Leng Tan, 2001), 'Communication' and 'Delivery' (Kanan and Choon Tan, 2002). In addition, another variable, 'Location', was proposed because sourcing locally should typically result in more efficient and simpler processes. In this case, deliveries are usually cheaper and faster, communication is easier, and closer quality control is possible. Harikrishnakumar *et al.* (2019) identified geographic location as the most important criterion in supplier selection.

3.2 Data

The dataset was generated with Makaroo (https://www.mockaroo.com/), an online data generator. The final dataset contains 500 rows of observations with the following variables:

‘vendor id’, ‘vendor name’, ‘quality’, ‘delivery’, ‘communication’, ‘reliability’ and ‘location’, which were rated on the scale of 1-10 and ‘PR impact’, which classifies supplier into ‘positive’, ‘neutral’ and ‘negative’ categories. Data distribution rules were applied to the dataset to prevent complete randomness so that the model could be tested and evaluated. This has been done using the ‘Create a Custom Distribution’ option in Makaroo. It evaluates each given rule as true or false and then applies a distribution according to the assigned number. For example, if there is ‘reliability >= 5’ as the first rule with value 3 assigned to class ‘positive’, and 0 to other classes, this will result in class ‘positive’ being assigned 3 times more often relative to other classes when variable ‘reliability’ is rated equal or higher than 5. Rules are evaluated in order of definition; therefore, the first rule has higher importance than the second one.

The following rules were applied in this order:

1. If ‘reliability’ is equal or higher than 5, then supplier PR impact is classed as ‘positive’.
2. If ‘delivery’ is equal or higher than 5, then supplier PR impact is classed as ‘negative’.
3. If ‘quality’ is equal or higher than 5, then supplier impact is classed as ‘positive’.

These rules do not necessarily represent real-world vendor behaviour impact on PR, but model performance evaluation would not be possible without them.

75% of the dataset was used for training, and the remaining 25% for testing.

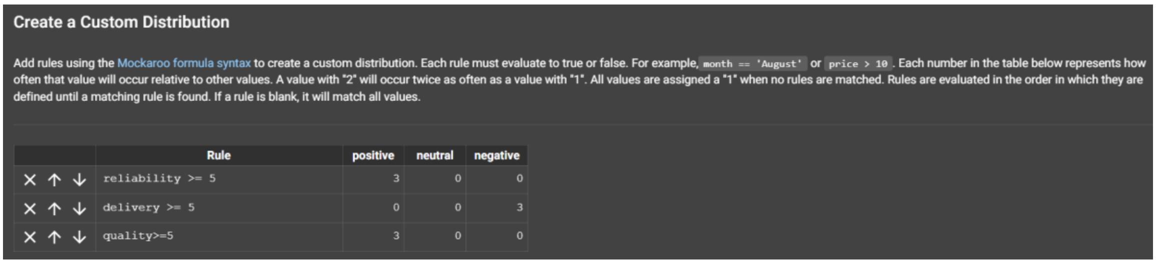

Figure 1: Custom Distribution in Makaroo

3.3 Classification methods

RF, KNN, and NB were implemented and evaluated in this study. These classifiers were chosen on the basis of their overall simplicity and accuracy in previous studies, such as that of Fernandez-Delgado *et al.* (2014), whose evaluation of 179 classifiers proved the highest accuracy of RF; Wahyono and Kang-Hyun (2014), whose test showed the best accuracy for RF and KNN from four classifiers tested; and Ting *et al*. (2011), who suggested NB as the most accurate method in their classification problem.

3.3.1 Random Forest

RF combines predictions from multiple individual decision trees and averages those results, offering improved accuracy compared with a single decision tree. RF was the first choice of algorithm because of its numerous advantages in the context of this project. It. It also provides accurate insights into feature importance which can help with further decision making in the supplier selection area, is easy to interpret, and can handle large complex datasets, which are common in supply chain data.

3.3.2 K-nearest neighbours

KNN is another supervised learning algorithm which is commonly used for both classification and regression problems. It classifies new data points by looking at the class of the closest neighbours in the training data. The distance between two data points in KNN is calculated using Euclidean distance, which is a measure of the straight-line distance between two points in a multidimensional space.

3.3.3 Naive Bayes

NB is a probabilistic algorithm commonly used for classification tasks. It is built on Bayes' theorem, which is a core concept in the field of probability. There is a simplified assumption that there is no dependence between features in terms of class label. It determines the likelihood of each potential class label for a given input, and subsequently selects the label with the highest probability as the predicted label.

3.4 Tools

Most of the project was completed using Python version 3.10.11 and Google Collaboratory, which is a cloud Jupyter Notebook environment. Visual Studio Code version 1.78.2 was used to build the user interface. The following Python libraries were used.

- Pandas: A data manipulation library which provides data structure and functions for data manipulation and analysis. It was used to load data from the CSV files, create data frames, perform operations on them, and display them.
- NumPy: Fundamental library for performing numerical operations in Python.
- tkinter: Library to create graphical user interfaces (GUI). It was used to create a GUI window and input fields to make the predictions.
- scikit-learn: a machine learning library that provides various algorithms and tools for tasks such as classification, regression, clustering, and model evaluation. The classification module in the code is used to train the data and make predictions, and the metrics module is used to calculate the accuracy of the predictions.
- Matplotlib: Plotting library in Python. It allows the creation of various types of visualisations, such as line plots, scatter plots, bar plots or histograms. In this project, it was used to create a bar plot to visualise the feature importance score.
- Seaborn: Statistical data visualisation library built on Matplotlib, providing a higher-level interface. This was used to enhance the styling of the bar plots.

3.5 Model performance measures

To evaluate the classification model, first, true positives, true negatives, false positives, and false negatives need to be identified:

- **True positives (TP)**: number of suppliers correctly classified into their categories.
- **True negatives (TN)**: number suppliers correctly predicted as not belonging to given categories.
- **False positives (FP)**: number of suppliers predicted to belong to given categories but not belonging there.
- **False negatives (FN)**: number of suppliers predicted to not belong to given categories but actually belonging there.
- **All Positives (P):** total of TP and FN.

- **All Negatives(N)**: total of TN and FP.

From there, a confusion matrix can be created, and more meaningful evaluation metrics can be calculated (Tharwat, 2020):

- **Accuracy**: the ratio of correctly predicted data to all predicted data.
- **Sensitivity (Recall)**: TP / (TP + FN) = TP / P
  This will show how good the test is at predictive positives (true positive rate).
- **Specificity:** TN / (TN + FP) = TN / N
  This shows the proportion of true negatives correctly identified by the model.
- **Precision**: TP / (TP + FP)
  This yields a fraction of the predicted positives that are positive.
- **F1 score**: 2 * (precision * recall) / (precision + recall)
  It is computed using the average of precision and recall, providing a more accurate measure, particularly in the case of unbalanced data. The closer the f1-score is to 1, the better it is.

For this project, three target values (labels) in ML models were used: Positive, Neutral and Negative (according to their impact on the Product Release Rate; this naming convention is unrelated to TP, TN, FP, and FN and should not be confused). As classification is non-binary in this case, sensitivity, precision, specificity and f1-score need to be calculated separately for each category, and the mean of the three categories is used for each metric.

4 Result and Analysis

4.1 Random forest model evaluation

The RF algorithm was found to be the most accurate with three features used per tree and was run each time using a default value of 100 trees.

The accuracy in this case was 0.97, which is close to perfect.

Among the three methods used, RF was the best at detecting the features that had the most impact on the given data pattern. Its feature importance visualisation provided an accurate representation of the actual trends in the dataset.

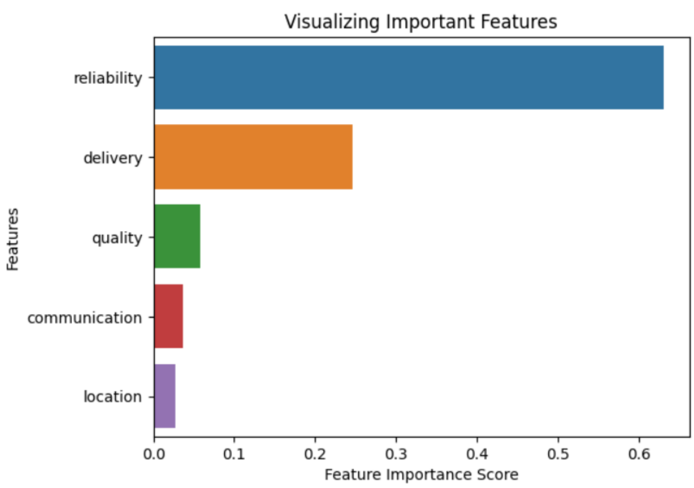

Figure 2: Feature Importance, Random Forest

Precision	**0.99**
Recall	0.80
Specificity	0.95
F1 Score	0.86

Table 1: Random Forest, other metrics

4.2 KNN model evaluation

The KNN model achieved the highest accuracy with n=3.

The accuracy in this case was 0.88, which was significantly lower than that of the RF model. The possible reason for this might be that KNN struggled with the unbalanced dataset, which had a much higher proportion of instances classed as 'positive' than two other classes:

Negative	**129**
Neutral	10
Positive	361

Table 2: Number of instances for each class in dataset

KNN is not recommended for feature selection or feature importance evaluation because the algorithm relies on all features equally and does not assume any underlying relationship between features and the target variable. Therefore, KNN would not be the best choice in this case, as the visibility of feature importance would be very helpful in the procurement decision-making process.

Precision	**0.89**
Recall	0.72
Specificity	0.89
F1 Score	0.78

Table 3: KNN, other metrics

4.3 Naive bayes model evaluation

The NB model achieved an accuracy of 0.91, slightly higher than KNN. This could probably be improved by parameter tuning. However, feature importance visualisation is not a truthful representation of the dataset used, as it only represents the mean value of the feature in each class. Naive Bayes is not considered the best algorithm for feature importance evaluation because it is based on probabilities and does not rank features in terms of importance.

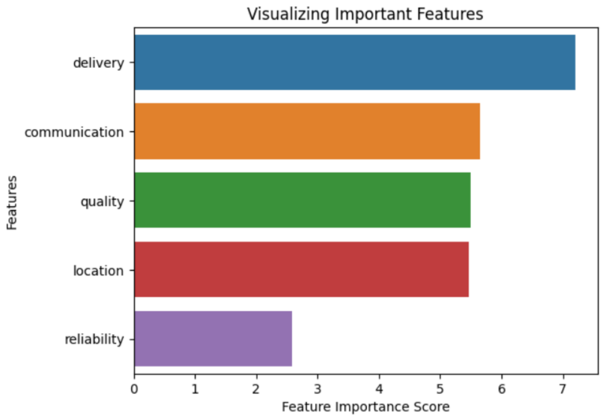

Figure 3: Feature Importance, Naïve Bayes

Precision	**0.94**
Recall	0.73
Specificity	0.89
F1 Score	0.80

Table 4: Naïve Bayes, other metrics

4.4 Model evaluation summary

As shown in the summary table below, RF outperformed the two other algorithms in all metrics used.

Algorithm	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1 score*	*Feature importance*
RF	0.97	0.99	0.80	0.95	0.86	Accurate
KNN	0.88	0.89	0.72	0.89	0.78	Not recommended
NB	0.91	0.94	0.73	0.89	0.80	Not meaningful in terms of impact

Table 5: Models evaluation summary table

5 Discussion and Conclusions

This study investigates which of the three common machine learning methods is most suitable for the vendor selection process and the justification for this. It also offers a fresh perspective on features and algorithm selection in supplier selection machine learning models, emphasising the importance of choosing nonmonetary features, what should ultimately result in higher profitability.

Models have been trained on an artificially generated dataset that does not reflect real-world trends and does not allow the investigation of the real impact on companies' profits.

As the data were not real, the project did not bring new insights into the rank of importance of the criteria in the supplier selection process. However, it provides a framework which can be used to determine this rank when trained and used with business data.

The results suggest that RF is the best-performing algorithm for this relatively small (500 instances), unbalanced, multidimensional, feature-scaled (ranges were normalised to all variables) and non-binary (multiclass) dataset used. This finding is consistent with the study of the performance of RF, KNN and NB in the prediction of chronic kidney disease (CKD)by Devika et al. (2019), where RF showed the highest accuracy and overall was best at the prediction of CKD. RF was also one of the top-performing classifiers in other reviewed studies, such as Wahyono and Kang-Hyun (2014) and Suksomboon and Ritthipakdee (2022). In addition, RF appears to be the best tool for feature importance detection. Unfortunately, there were no real data in this project; therefore, the most important features identified likely do not reflect reality. However, in the business environment, if this model is applied to procurement data, this advantage of RF would prove very valuable for procurement decision-making.

The actual impact on profit could be subject to larger research, in which relevant data can be collected for a period of time and then analysed.

Although the RF performance was the best of the three classifiers used, Wainberg et al. (2016) suggested that there are other classifiers that may work equally well or even better, such as SVM and ANN. These are beyond the scope of this project, but can be evaluated in further research.

This DSD is suitable for the initial stage of supplier selection, classification. In the future, it can be further enhanced to be usable in a later selection stage, where suppliers classified as having a positive impact on PR would be ranked according to their monetary offers.

6 References

interos (2022) Resilience 2022: Interos Annual Global Supply Chain Report. Available at: https://www.interos.ai/resources/global-supply-chain-report/ (Accessed: 01 March 2024).

De Boer, L., Labro, E. and Morlacchi, P. (2001) A review of methods supporting supplier selection, *European journal of purchasing & supply management,* 7(2), pp. 75-89.

de FSM Russo, R. and Camanho, R. (2015) Criteria in AHP: a systematic review of literature, *Procedia Computer Science,* 55, pp. 1123-1132.

Devika, R., Avilala, S.V. and Subramaniyaswamy, V. (2019) Comparative study of classifier for chronic kidney disease prediction using naive bayes, KNN and random forest. In *3rd International conference on computing methodologies and communication ICCMC* (2019), pp. 679-684.

Esmaily, H., Tayefi, M., Doosti, H., Ghayour-Mobarhan, M., Nezami, H. and Amirabadizadeh, A. (2018) A comparison between decision tree and random forest in determining the risk factors associated with type 2 diabetes, *Journal of research in health sciences,* 18(2), pp. 412.

Fernández-Delgado, M., Cernadas, E., Barro, S. and Amorim, D. (2014) Do we need hundreds of classifiers to solve real world classification problems?, *The journal of machine learning research,* 15(1), pp. 3133-3181.

Harikrishnakumar, R., Dand, A., Nannapaneni, S. and Krishnan, K. (2019) Supervised machine learning approach for effective supplier classification. In: *18th IEEE International Conference On Machine Learning And Applications ICMLA* (2019), pp. 240-245.

Kannan, V.R. and Tan, K.C. (2002) Supplier Selection and Assessment: Their Impact on Business Performance, *Journal of Supply Chain Management,* 38(3), pp. 11-21 Available at: 10.1111/j.1745-493X.2002.tb00139.x.

Moore, J.H. and Chang, M.G. (1980) Design of decision support systems, *ACM SIGOA Newsletter,* 1(4-5), pp. 8-14.

Moynihan, G.P., Saxena, P. and Fonseca, D.J. (2006) Development of a decision support system for procurement operations, *International Journal of Logistics Systems and Management,* 2(1), pp. 1-18.

Omurca, S.I. (2013) An intelligent supplier evaluation, selection and development system, *Applied Soft Computing,* 13(1), pp. 690-697.

P. Suksomboon and A. Ritthipakdee (2022) Performance Comparison Classification using k-Nearest Neighbors and Random Forest Classification Techniques. In: *3rd International Conference on Big Data Analytics and Practices IBDAP* (2022), pp. 43-46.

Probst, P., Wright, M.N. and Boulesteix, A. (2019) Hyperparameters and tuning strategies for random forest, *Wiley Interdisciplinary Reviews: data mining and knowledge discovery,* 9(3), pp. e1301.

Taherdoost, H. and Brard, A. (2019) Analyzing the process of supplier selection criteria and methods, *Procedia Manufacturing,* 32, pp. 1024-1034.

Tan, K., Kannan, V.R. and Handfield, R.B. (1998) Supply chain management: supplier performance and firm performance, *International Journal of Purchasing & Materials Management,* 34(3).

Thanh Noi, P. and Kappas, M. (2017) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery, *Sensors,* 18(1), pp. 18.

Tharwat, A. (2020) Classification assessment methods, *Applied computing and informatics,* 17(1), pp. 168-192.

Ting, S.L., Ip, W.H. and Tsang, A.H. (2011) Is Naive Bayes a good classifier for document classification, *International Journal of Software Engineering and Its Applications,* 5(3), pp. 37-46.

Tracey, M. and Leng Tan, C. (2001) Empirical analysis of supplier selection and involvement, customer satisfaction, and firm performance, *Supply chain management: An international journal,* 6(4), pp. 174-188.

Wahyono and Kang-Hyun Jo (2014) *A comparative study of classification methods for traffic signs recognition* IEEE.

Wainberg, M., Alipanahi, B. and Frey, B.J. (2016) Are random forests truly the best classifiers?, *The Journal of Machine Learning Research,* 17(1), pp. 3837-3841.

Wilson, V.H., NS, A.P., Shankharan, A., Kapoor, S. and Rajan A, J. (2020) Ranking of supplier performance using machine learning algorithm of random forest, *International Journal of Advanced Research in Engineering and Technology (IJARET),* 11(5).