International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Data Analytics for Decision-Making in Evaluating the Top-Performing Product and Developing Sales Forecasting Model in an Oil Service Company

Ronggo Saputro[1], Santi Novani[2]
[1,2] Master of Business Administration, School of Business Management, Institut Teknologi Bandung, Indonesia

ABSTRACT: This study addresses the strategic challenges faced by a company specialising in the manufacture of oil and gas equipment. Following organisational restructuring, which involved the dissolution of one business unit and the creation of another, the company is navigating complexities in product focus and manpower allocation within the Asia-Pacific region. The research problem centres on identifying the top-performing product, determining potential countries for establishing a support base facility based on sales performance, and developing a method for forecasting future sales.

The research involved retrieving and pre-processing historical sales data, then performing a thorough descriptive and predictive analysis. The data was partitioned into training and testing sets to facilitate predictive analytics. Several predictive models were developed and tested, including neural networks, linear regression, gradient-boosted trees, random forests, and ARIMA methods. Tableau Public was utilised for descriptive analytics, whereas RapidMiner Studio was employed for predictive analytics.

The study's results, derived through both descriptive and predictive analytic methods, reveal critical insights. The Blowout Preventer (BOP) emerged as the top-performing product in the Asia-Pacific region. In terms of establishing support base facilities, Malaysia was identified as the ideal location for the BOP, while Indonesia was found suitable for the Manifold product group. Furthermore, the Random Forest model was determined to be the most effective for forecasting future sales. These findings provide strategic guidance for the company in product focus, regional expansion, and resource allocation, contributing significantly to the company's decision-making process in a competitive industry.

KEYWORDS: Analytics, Descriptive, Predictive, Product performance, Random forest, Sales forecasting,

INTRODUCTION

In recent years, the company underwent significant organisational changes, including the dismissal of its welding department a couple of years ago. The welding department was responsible for supporting the fabrication of skids for manifold product. Then the company's former general manager redirected the business focus towards surface wellhead and christmas tree equipment, which resulted in forming a new department and becoming the engineering centre for surface wellhead products. However, the revenue generated from this product line did not exhibit a significant increase. This restructuring is further complicated by challenges in strategic focus, particularly in identifying the most suitable product group for focus within the Asia-Pacific region. The other issue is that the company is faced with a pressing requirement from customers to extend its operating facility outside of Singapore. The main focus of this expansion is the development of support bases facility in these countries. The support bases facility serves multiple purposes, including storing spare parts, conducting maintenance, and servicing company's equipment. This facility enables faster delivery and reaction times to support customer ongoing operations. Furthermore, the company currently lacks a scientific method for sales forecasting, leading to potential issues such as missed sales opportunities, increased carrying costs, and the accumulation of unsold stock. Thus, achieving more accurate sales forecasts is crucial for the company to efficiently allocate resources, plan production schedules, and manage supply chain operations.

This study aims to achieve three key objectives. First, it seeks to identify the top-performing product lines in the region, crucial for strategic decision-making and resource allocation. Second, the research aims to pinpoint the most suitable country for establishing a support base facility, essential for operational efficiency and market penetration. Lastly, the study focuses on identifying the optimal predictive analytics model to forecast future sales. These objectives are designed to enhance the company's strategic and operational performance in the Asia-Pacific market.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

LITERATURE REVIEW

Data analytics refers to the scientific process of transforming data into insights to facilitate better decision-making (Mariani & Wirtz, 2023). It encompasses various advanced analytics methods that enhance the intelligence and capabilities of applications through smart decision-making in different scenarios (Sarker, 2021). Analytics refers to the use of intelligent data, learner-produced data, and analysis models to discover information and connections and to predict and advise on learning (Ferguson, 2012). This definition emphasises the use of data and analysis models to gain insights and make predictions in making better decision. This also highlights the role of analytics in uncovering patterns within data, which is essential for both descriptive and predictive analytics. Data analytics consists of both descriptive and predictive analytics.

A. Descriptive Analytics

Descriptive analytics involves the exploration and interpretation of historical data to understand patterns and trends. It provides insights into what has happened in the past and helps in summarising and interpreting data to make it understandable.

B. Predictive Analytics

Predictive analysis, known as predictive modeling or data mining, is a method for analysing historical data and predicting future outcomes. It entails the application of statistical algorithms and machine learning techniques to identify patterns, trends, and relationships in the data (Wisesa et al., 2020). Predictive analysis can be applied to sales data to forecast future sales based on factors such as customer behavior, market conditions, and promotional strategies (Chong et al., 2016).

C. Data-Driven Decision Making

The combined use of descriptive analysis, predictive analysis, sales data, and sales forecasting enables companies to make data-driven decisions and improve sales performance. By leveraging historical sales data and applying predictive analysis techniques, businesses can generate accurate sales forecasts, which inform decision-making and drive business strategies (Wang & Aviles, 2023). This progressive process enables businesses to adapt to varying market conditions, optimize resource allocation, and increase overall sales performance (Blanco-Mesa et al., 2018).

D. Executive Dashboard

An executive dashboard is a data-driven tool that executive management uses to track, examine, and visualise key performance indicators and important business metrics. There are different types of dashboards, including strategic, tactical, and operational dashboards, each serving different functions and targeted towards various users (Buttiġieġ et al., 2017).

E. Sales Forecasting by Predictive Model

Sales forecasting is the process of predicting future sales based on historical data, market trends, customer behaviour, and other relevant factors (Trento et al., 2021). It offers invaluable insights for decision-making, including production planning, inventory management, pricing strategies, and resource allocation (Silva et al., 2021). Accurate sales forecasting is essential for effective decision-making and can help businesses optimize their operations and maximize profitability (Rodrigues, 2021).

Pavlyshenko (2019) investigates the application of machine learning models for predictive sales analytics. The study examines various approaches and case studies. The findings indicate that the use of advanced predictive analytics models has shown promising results in improving the accuracy of sales forecasting, especially in time series analysis (Pavlyshenko, 2019; Li & Zhang, 2022). The integration of predictive analytic methods in time series forecasting has significantly enhanced the accuracy and reliability of sales predictions, offering valuable insights for strategic decision-making in various business domains.

F. Previous Study on Predictive Analytics for Forecasting

In the business and analytics, sales forecasting stands as a critical area of study, especially in the context of predictive analytics. The following Table I tabulates the research studies that have been conducted in this field.

These studies encompass a range of methodologies, models, and tools used in predictive analytics, each contributing to the evolving understanding of sales forecasting. The study highlights the model they used, its implementation, and the implications of their work in the practical domain of forecasting.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table I. Previous Study on Forecasting by Predictive Analytics

Study	Method	Subject
Abdel-Khalik (1983)	ARIMA	Sales forecasting
Abdullahi et al. (2021)	Random Forest, Extreme Gradient Boosting, Support Vector Machine, Ensemble Model	Sales in supermarket
Afrianto et al. (2020).	Logistic Regression, Decision Tree, K-NN, Random Forest	Booking prediction for accommodation
Andariesta et al. (2022)	Artificial Neural Network, Support Vector Regression, Random Forest	Predicting international tourists in Indonesia during the COVID-19 pandemic
Ashraf et al. (2022)	Linear Regression, Random Forest, Gradient Boosting, ARIMA	Sales in IT retail chain stores
Chen et al. (2022)	Linear Regression, Gradient Boosting, Support Vector Machine, Artificial Neural Network	Sales in fashion retail company
Feng et al. (2022)	Support Vector Machine, Random Forest, Gradient Boosting, Extreme Gradient Boosting	Sales in short-term e-commerce
M et al (2021)	Linear Regression, Sentiment Based Approach, Econometric Model, Bass Model	Sales in e-commerce
Majhi et al. (2009)	Differential Evolution	Sales in retail
Pavlyshenko (2019)	Holt-Winters, ARIMA, SARIMA, SARIMAX, GARCH	Sales for new product or store
Raizada et al. (2021)	Multiple Linear Regression, Random Forest, K-NN, Support Vector Machine, Extra Tree Regression	Sales in retail stores
Rodrigues (2021)	ARMA, ARIMA	Sales in food-related business
Schmidt et al. (2022)	Recurrent Neural Network (RNN)	Sales in restaurant
Vineeth et al. (2020)	Support Vector Machine, Ridge Regression, Gradient Boosting, Random Forest	Sales for truck components (Volvo)
Wasesa et al. (2020)	SARIMA, Artificial Neural Network	Electricity consumption
Wasesa et al. (2022)	Extreme Gradient Boosting, Support Vector Regression, ARIMAX	Electricity consumption

A brief comparison of the predictive models utilised in this research is presented in Table II This table outlines the characteristics and differences of each model, providing a clear overview of their respective functionalities and applications within the study.

Table i. Brief Comparison of Predictive Model Characteristics

Model	Characteristics
Neural Network	It is known for its ability to capture complex patterns in data, making it suitable for nonlinear relationships in sales data (Abdou et al., 2008).
Linear Regression	It is a fundamental technique that is commonly used in various fields, including sales forecasting, due to its simplicity and interpretability (Dombrowsky, 2023).
Gradient Boosted Trees	It is known for its high predictive accuracy and ability to handle large datasets efficiently (Koduru, 2020; Li et al., 2022).
Random Forest	It is known for its robustness to overfitting and capability to work well with uninformative features (Munsarif et al., 2022).
ARIMA	A time series model that is specifically designed for analysing and forecasting time-dependent data makes it suitable for sales forecasting when dealing with seasonality and trend components (Juba, 2016).

G. Conceptual Framework

The conceptual framework for this research, as shown in Figure I, establishes the foundation for the study by defining key concepts, theories, and relationships between variables. A diagram showing all concepts or models that will be employed to solve the business issue.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

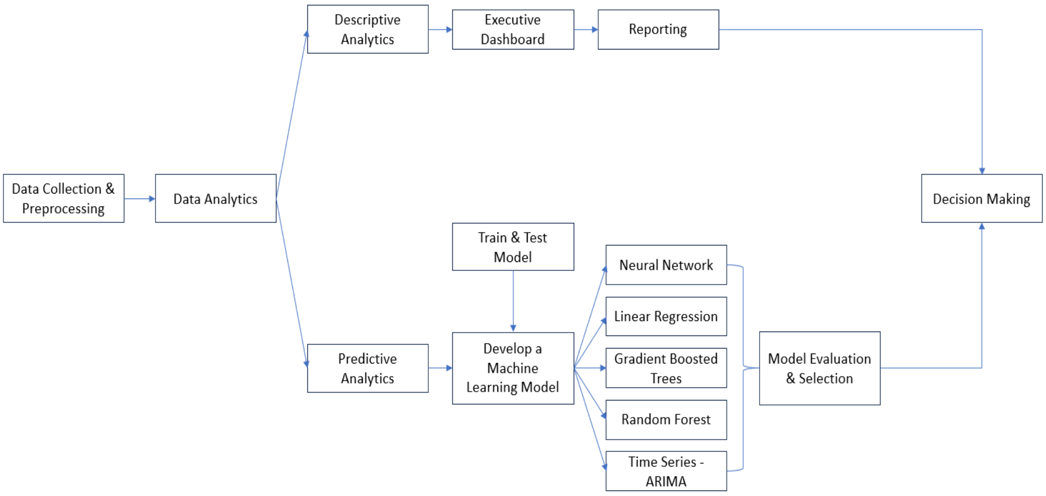

Figure I. Conceptual Framework

RESEARCH METHODOLOGY

In this research, the company employs an ERP system named Infor ERP SyteLine Version 8.02 for recording its business transactions. The essential data for this study is the historical sales data, extracted from two specific modules within SyteLine: “Customer Order” and “Customer Order Lines.” The Customer Order module captures the aggregate value of orders placed by customers per purchase order, while the Customer Order Line module details the individual items and pricing derived from the Customer Order module. The dataset selected for this research encompasses sales data spanning from 2017 to 2023. The "Customer Order" module contributed a total of 1,321 data entries, each record comprising 30 distinct fields. In parallel, the "Customer Order Lines" module provided a more extensive dataset with 12,391 individual records, each encompassing 64 fields.

The data collected from these modules are subsequently consolidated into a singular dataset. This initial dataset then undergoes a series of preprocessing steps. These steps encompass cleaning the data, addressing missing values, and selecting relevant fields, thereby ensuring the data's readiness for further analysis. After the merger, a thorough process of data cleansing was carried out, specifically targeting the filtration and removal of entries that contained missing values. The extensive process of data cleaning led to a cleaned dataset consisting of 9,627 records and 22 fields, hence improving the integrity and dependability of a subsequent study.

The chosen methodology for analysing historical sales records in this study is data analytics. This scientific methodology effectively transforms data into substantial insights, which eventually improves decision-making processes. Data analytics comprises two fundamental components: descriptive analytics and predictive analytics.

Descriptive analytics looks into historical data to identify patterns and trends. It enhances comprehension of historical events by summarising and analysing data, making it understandable. By employing data aggregation and data mining approaches, descriptive analytics offers valuable insights into past trends, with emphasis on answering the question "What has occurred?".

Predictive analytics, conversely, employs machine learning algorithms to analyse both present and historical data for predicting future events or behaviours. It is aimed at forecasting future trends and behaviours by recognizing patterns in data. Employing methods such as neural network, linear regression, gradient boosted trees and random forest, predictive analytics seek to answer, "What is likely to happen?".

RESULTS AND DISCUSSION

In this study, data visualisation was performed using an online interactive data visualisation tool specifically designed with a focus on business data analytics. The capability of this tool is crucial in helping individuals and organisations become more data-driven, allowing for a more detailed and efficient analysis of complex information. This software is essential for visually representing analytical findings, which improves the clarity and impact of the research.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

The primary purpose of utilising this visualisation tool in the study was to generate a comprehensive Executive Sales Overview Dashboard using the datasets. This interactive dashboard was specifically created to offer the company's management a clearer picture and extensive insights into sales performance. Figure II demonstrates that the Executive Sales Overview Dashboard consists of seven separate graphs or visualisations, each contributing to a comprehensive understanding of the sales dynamics. This interactive visualisation tool is essential for management to analyse extensive data patterns and make well-informed decisions by examining graphical representations of sales trends and metrics. The Executive Sales Overview Dashboard also features interactive tooltips that display detailed information and visualisations upon cursor hover.

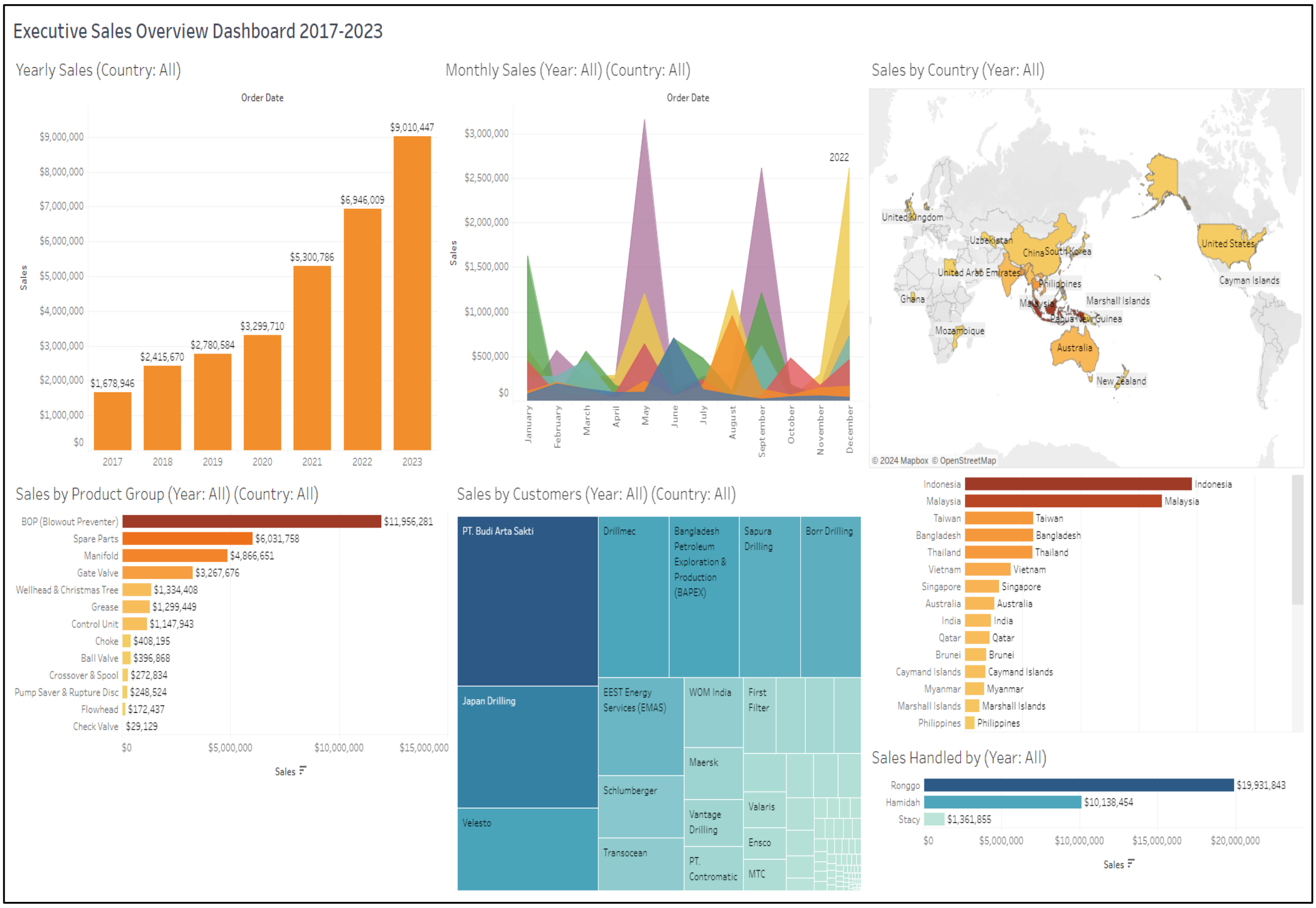

Figure II. Executive Sales Overview Dashboard

The initial visualisation presented in the Executive Sales Overview Dashboard is the Yearly Sales chart. This graphic is important for demonstrating the yearly sales trends. The data, covering the period from 2017 to 2023, clearly illustrates a consistent trend of increasing sales throughout these years. The bar chart accurately captures and represents the increasing trend in the sales figures. The second chart in the dashboard is the Monthly Sales chart. This chart provides an insightful view of sales booking patterns over the years. It effectively highlights consistent spikes in sales data, particularly noticeable in certain months. The analysis reveals that these sales peaks typically occur in May, June, August, September, and December. This pattern of sales fluctuations is crucial for understanding the seasonal or periodic trends in the company's sales activities, thereby offering valuable insights for strategic planning and decision-making. The next visualisation in the sequence focuses on analysing sales by country. This analysis utilises two different types of visual representation: a geographical map and a bar chart. The visualisations clearly show the geographic distribution of sales, accurately identifying the main countries where the company's products are mostly sold. Based on this data, it is clear that

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Indonesia is the primary recipient of these products, with Malaysia closely following. The fifth visualisation in the series is the 'Sales by Product Group' chart, which is displayed as a bar chart. This chart provides a concise overview of the product categories that have recorded the highest levels of sales from 2017 to 2023. The data unmistakably shows that the BOP, Spare Parts, and Manifold have been the top-selling products over this time period. This bar chart is essential for illustrating the product preferences within the sales portfolio, offering a distinct indication of customer demand trends and product popularity over the observed period. Following in the sequence of visualisations is the Sales by Customer analysis. This visualisation employs a tree map layout, where the size of each box correlates directly with the sales volume associated with each customer. Larger boxes within the chart indicate a higher volume of purchases made by those customers. This form of representation effectively demonstrates the varying levels of sales contributions across different customers, highlighting key clients who significantly contribute to the company's sales revenues. Such a visualisation is instrumental in identifying major clients and understanding customer-based sales distribution. The last visualisation focuses on the sales professionals who are in charge of handling particular sales orders. This chart serves as an important tool for management, as it simplifies the process of identifying and evaluating the workload and performance of every sales representative. The chart illustrates the distribution of sales orders among the individuals, offering clear insights into individual contributions and efficiency within the sales team.

The BOP, Spare Parts, and Manifold were the three best-selling products from 2017 to 2023, according to observations. Further analysis focuses on the last three years: 2021, 2022, and 2023. This refined analysis also indicates that the BOP remains the top-selling product. Manifold, however, has moved up to the second spot in this more recent period, ahead of Spare Parts. The analysis also enables identification of the primary countries where the BOP was most utilised during the period from 2021 to 2023. The findings indicate that Malaysia was the leading country in terms of BOP usage, followed closely by Indonesia. This geographical distribution of product usage provides valuable insights into market penetration and regional preferences for the BOP product, highlighting key areas of focus for future business strategies and market development initiatives. When the filter is set to Indonesia, the targeted analysis reveals that in Indonesia, Manifold emerges as the top-performing product, followed by the BOP. Conversely, a different outcome is observed when the filter is applied to Malaysia. In the Malaysian market, the BOP is identified as the top-selling product, with Spare Parts following in terms of sales volume. These variations in product performance across different countries underscore the importance of regional market dynamics and customer preferences in shaping sales trends.

This study integrates predictive analysis to perform sales forecasting using well-acknowledged open-source software designed for data mining and comprehensive data analysis. The extensive features of this software enable detailed predictive modelling, which is crucial for predicting future sales trends using previous data patterns and variables. Partitioning the dataset into training and testing sets is a fundamental aspect of data preprocessing, especially relevant in the context of this research, where the dataset comprises historical sales data and thus is categorised as a time-series dataset. The division of the dataset was conducted in a chronological manner: the training set encompasses all data records from the years 2017 to 2022, while the testing set includes the data records from the year 2023, as illustrated in Figure III.

Year	2017	2018	2019	2020	2021	2022	2023
Data	Training Set						Testing Set

Figure III. Data Training and Data Test

This study involves the processing of data using five different predictive analytics techniques: Neural Networks, Linear Regression, Gradient-Boosted Trees, and Random Forest. For each of these methodologies, five separate models have been developed. These models, tailored to their respective analytical methods, are detailed in Table III below. This approach of employing multiple models per method allows for a comprehensive exploration of the predictive capabilities of each technique, thereby enhancing the depth and breadth of the analysis.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table ii. Model Parameter

Method	Parameter	Model 1	Model 2	Model 3	Model 4	Model 5
Neural Net	Hidden Layer	2	4	6	4	4
	Training Cycle	200	200	200	300	300
	Learning Rate	0.01	0.01	0.01	0.01	0.1
Linear Regression	Feature Selection	M5 Prime	Greedy	T-Test	T-Test	Iterative T-Test
	Alpha	-	-	0.05	0.1	0.05
	Min Tolerance	0.05	0.05	0.05	0.05	0.1
	max iteration	-	-	-	-	10
Gradient Boosted Tree	Number of Trees	50	100	100	100	200
	Maximal Depth	5	5	10	10	20
	Learning Rate	0.01	0.01	0.01	0.1	0.5
Random Forest (least square)	Number of Trees	100	200	200	100	200
	Criterion	Least Square	Least Square	Least Square	Least Square	Least Square
	Maximal Depth	10	10	20	20	50
ARIMA	p,d,q	1,0,0	1,1,0	1,0,1	1,1,1	0,1,1

Upon the completion of processes for all five variations of the Neural Network models, it was established that Neural Net Model 4 achieved the most favourable results in terms of prediction accuracy. This particular model, distinguished by its configuration with four hidden layers, 300 training cycles, and a learning rate of 0.01, demonstrated the lowest values in Root Mean Square Error (RMSE) and Mean Absolute Error (AE), recorded at 30,871.04 and 6,203.809, respectively. In the case of the Linear Regression models, despite variations in parameters, all models yielded identical results. The RMSE for each model stood at 30,847.781, and the MAE was recorded at 5,852.623. These consistent outcomes across different parameter settings highlight the stability of the Linear Regression approach in this context. In the Gradient-Boosted Tress methodology, the lowest values of RMSE and MAE were achieved by Model 4, registering an RMSE of 30,297.012 and a MAE of 4,485.585. The parameters defining Model 4 include a configuration of 100 trees, a maximal depth of 10, and a learning rate of 0.1. Within this forecasting methodology, Model 5 emerged as the most accurate, achieving the lowest RMSE and MAE among its counterparts. The RMSE recorded was 30,140.55, and the AE stood at 4,306.051. The defining parameters of Model 5 comprised a configuration of 200 trees, the least square criterion as the decision-making criterion, and a maximal depth of 10. The RMSE and MAE of the Autoregressive Integrated Moving Average (ARIMA) model are higher compared to other methods, with Model 1 achieving the most favourable results, an RMSE of 81,097.958 and a MAE of 54,752.641, indicating a variance in predictive accuracy relative to other modelling techniques in this study.

Upon completing the data processing for all models with various variables, a comparative evaluation was conducted based on the RMSE and MAE scores. Table IV comprehensively presents the RMSE and MAE scores for each model. It showed similar RMSE and MAE scores for Neural Networks, Linear Regression, Gradient-Boosted Trees, and Random Forest, while the ARIMA method exhibited notably higher scores compared to these methods. It was also observed that the best RMSE and AE scores were achieved by the Random Forest method, Model 5, with an RMSE of 30,140.55 and a MAE of 4,306.51. The result of all the predictive models used in this research for sales forecasting is graphically depicted in Figure IV.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

Table iii. Comprehensive Result of RMSE and MAE

Method	Model	RMSE	MAE
Neural Net	Model 1	30934.447	7778.202
	Model 2	30871.515	6207.602
	Model 3	30921.346	7583.417
	Model 4	**30871.04**	**6203.809**
	Model 5	30916.951	7634.508
Linear Regression	Model 1	30847.781	5852.623
	Model 2	30847.781	5852.623
	Model 3	30847.781	5852.623
	Model 4	30847.781	5852.623
	Model 5	30847.781	5852.623
Gradient Boosted Tree	Model 1	30638.991	5202.092
	Model 2	30586.752	4884.272
	Model 3	30437.595	4836.924
	Model 4	**30297.012**	**4485.585**
	Model 5	30242.047	4518.896
Random Forest (least square)	Model 1	30240.567	4424.82
	Model 2	30241.655	4434.847
	Model 3	30140.691	4306.842
	Model 4	30140.959	4294.907
	Model 5	**30140.55**	**4306.051**
ARIMA	Model 1	**81097.958**	**54752.641**
	Model 2	118417.569	81964.598
	Model 3	102650.47	71952.262
	Model 4	215718.738	160540.663
	Model 5	118571.273	91379.177

The Executive Sales Overview Dashboard provides insights into the business issue of identifying the top-performing product of the company in the Asia-Pacific region. Analysis from the last three years indicates that the BOP leads in performance, followed by Manifold and Spare Parts. This outcome not only highlights BOP as one of the company's strongest products but also suggests strategic directions for the company. Focusing more on BOP, such as establishing a dedicated service facility for this product—a plan previously considered but not executed—or forming a specialised support team in Asia, could enhance operational efficiency. This approach would reduce reliance on the Houston team and enable swifter responses in the Asia-Pacific market.

In addressing the strategic decision of where to establish a support base, insights can be obtained from the Executive Sales Overview Dashboard as well. This involves identifying the countries where the company's products, particularly the BOP, are most in demand. The dashboard's visualisations indicate that Malaysia is the leading market for BOP, suggesting it as a potential location for a support base dedicated to this product. Conversely, when considering overall sales, Indonesia emerges as the top destination for the company's products, with Manifold being the most popular. Therefore, Indonesia also presents itself as a viable option for establishing a support base, with a focus on Manifold products. This analysis aids in strategic planning for geographical expansion and resource allocation.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

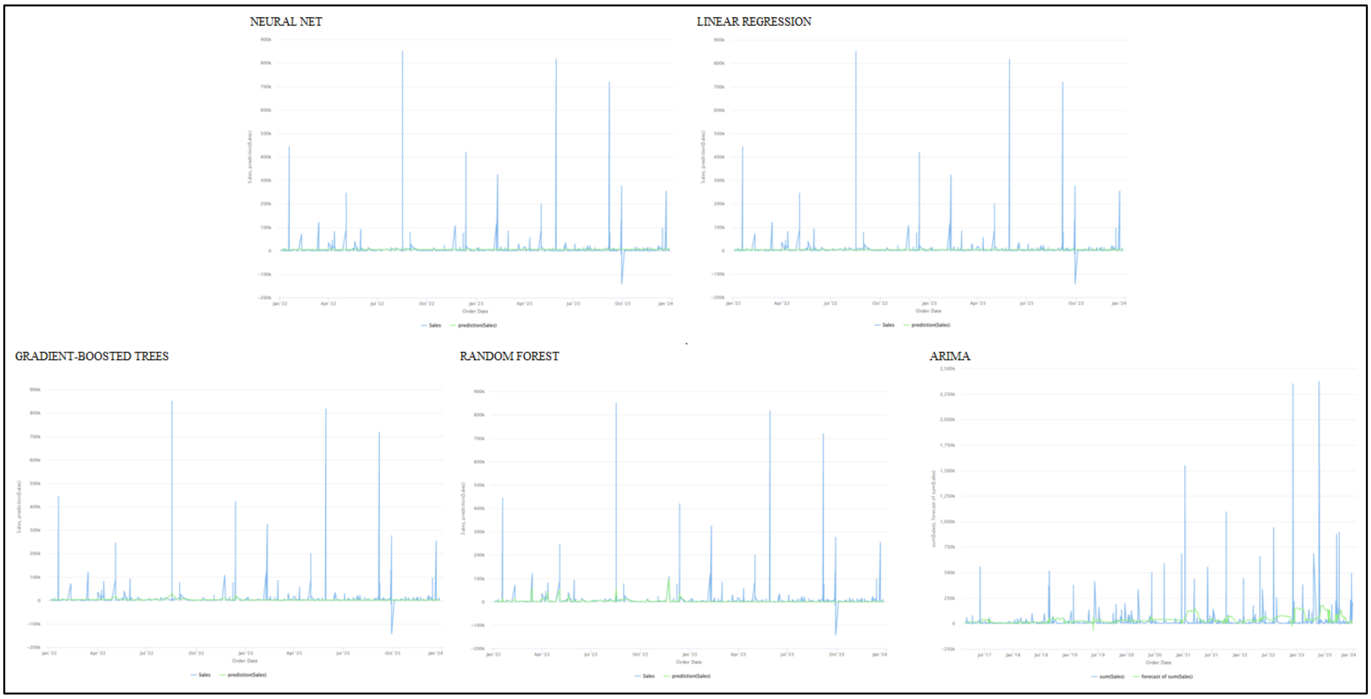

Figure IV. Predictive Model Results

The presence of outliers has the potential to impact the accuracy of forecasting models. Nevertheless, in this research, the focus was restricted to exploring approaches for handling these outliers, specifically the discussion of whether to remove them or apply alternative treatments without excluding them from the dataset. The outliers within this research are identified in Figure V with green dots. The presence of outliers in this study is attributed to instances of exceptionally high sales volumes on specific dates, often resulting from securing substantial tender values or due to significant discounts offered, as indicated by negative sales values.

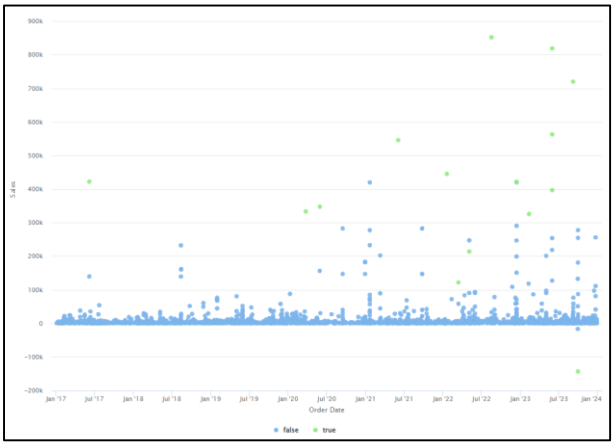

Figure V. Outliers

To resolve the concerns raised by the sales team regarding forecasting methods, the application of predictive analytics can be utilised to project future sales. Based on the analysis conducted in the preceding section, the Random Forest model emerges as the most effective predictive approach.

International Journal of Current Science Research and Review
ISSN: 2581-8341
Volume 07 Issue 02 February 2024
DOI: 10.47191/ijcsrr/V7-i2-14, Impact Factor: 7.943
IJCSRR @ 2024
www.ijcsrr.org

CONCLUSION

The top-performing product in the Asia-Pacific region is Blowout Preventer (BOP), which has the highest sales volume among the other products. Following the BOP in sales performance are the Spare Parts and Manifold product groups. Overall, the country that utilises the company's products most is Indonesia, followed by Malaysia. The BOP, as the top-performing product, is used the most in Malaysia. While in Indonesia, the top-selling product is the Manifold. Consequently, this suggests that Malaysia presents a strategic opportunity for establishing a support-based facility focused on the BOP, while Indonesia offers potential for a facility focused on the Manifold product group. The best predictive analytics model for forecasting future sales at MAX Singapore Pte Ltd is identified as the Random Forest model. This model is distinguished by achieving the lowest scores among the others in Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

REFERENCES

1. Abdel-Khalik, A. and El-Sheshai, K. (1983). sales revenues: time-series properties and predictions. Journal of Forecasting, 2(4), 351-362. https://doi.org/10.1002/for.3980020405
2. Abdou, H. A., Pointon, J., & El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in egyptian banking. Expert Systems With Applications, 35(3), 1275-1292. https://doi.org/10.1016/j.eswa.2007.08.030
3. Abdullahi, M., Aimufua, G., & Muhammad, U. (2021). Application of sales forecasting model based on machine learning algorithms. Proceedings of the 28th iSTEAMS Multidisciplinary &Amp; Inter-Tertiary Research Conference. https://doi.org/10.22624/aims/isteams-2021/v28p17
4. Afrianto, M. and Wasesa, M. (2020). booking prediction models for peer-to-peer accommodation listings using logistics regression, decision tree, k-nearest neighbor, and random forest classifiers. Journal of Information Systems Engineering and Business Intelligence, 6(2), 123. https://doi.org/10.20473/jisebi.6.2.123-132
5. Andariesta, D. and Wasesa, M. (2022). machine learning models for predicting international tourist arrivals in indonesia during the covid-19 pandemic: a multisource internet data approach. Journal of Tourism Futures. https://doi.org/10.1108/jtf-10-2021-0239
6. Ashraf, D. (2022). A Predictive Analysis Of Retail Sales Forecasting Using Machine Learning Techniques. Research Journal Of Computer Science And Information Technology, 04(6), 23–33. https://doi.org/10.54692/lgurjcsit.2022.0604399
7. Blanco-Mesa, F., León-Castro, E., Merigó, J., Xu, Z. (2018). Bonferroni Means With Induced Ordered Weighted Average Operators. Int J Intell Syst, 1(34), 3-23. https://doi.org/10.1002/int.22033
8. Buttiġieġ, S. C., Pace, A., & Rathert, C. (2017). Hospital performance dashboards: a literature review. Journal of Health Organization and Management, 31(3), 385-406. https://doi.org/10.1108/jhom-04-2017-0088
9. Chong, A. Y., Li, B., Ngai, E. W., Ch'ng, E., Lee, F. (2016). Predicting Online Product Sales Via Online Reviews, Sentiments, and Promotion Strategies. International Journal of Operations &Amp; Production Management, 4(36), 358-383. https://doi.org/10.1108/ijopm-03-2015-0151
10. Chen, C., Liu, Z., Zhou, J., Li, X., Yuan, Q., Jiao, Y., … & Zhong, X. (2019). How much can a retailer sell? sales forecasting on tmall. Advances in Knowledge Discovery and Data Mining, 204-216. https://doi.org/10.1007/978-3-030-16145-3_16
11. Dombrowsky, T. (2023). Linear regression. Nursing, 53(9), 56-60. https://doi.org/10.1097/01.nurse.0000946844.96157.68
12. Feng, T., Niu, C., & Song, Y. (2022). short term e-commerce sales forecast method based on machine learning models., 1020-1030. https://doi.org/10.2991/978-2-494069-31-2_119
13. Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5/6), 304. https://doi.org/10.1504/ijtel.2012.051816
14. Juba, B. (2016). Conditional sparse linear regression. https://doi.org/10.48550/arxiv.1608.05152
15. Koduru, M. (2020). Rf-xgboost model for loan application scoring in non-banking financial institutions. International Journal of Engineering Research And, V9(07). https://doi.org/10.17577/ijertv9is070117
16. Li, Z. and Zhang, N. (2022). Short-term demand forecast of e-commerce platform based on convlstm network. Computational Intelligence and Neuroscience, 2022, 1-10. https://doi.org/10.1155/2022/5227829
17. M., P. A. D. and S., A. V. (2021). Comparative study of product sales forecasting methods. International Research Journal on Advanced Science Hub, 3(Special Issue 7S), 117-124. https://doi.org/10.47392/irjash.2021.220

18. M. Wasesa, A. R. Tiara, M. A. Afrianto, F. I. Ramadhan, I. N. Haq and J. Pradipta, "SARIMA and Artificial Neural Network Models for Forecasting Electricity Consumption of a Microgrid Based Educational Building," 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, Singapore, 2020, pp. 210-214, doi: 10.1109/IEEM45057.2020.9309943
19. Majhi, R., Panda, G., Majhi, B., Panigrahi, S., & Mishra, M. (2009). forecasting of retail sales data using differential evolution. https://doi.org/10.1109/nabic.2009.5393740
20. Mariani, M. and Wirtz, J. (2023). A critical reflection on analytics and artificial intelligence-based analytics in hospitality and tourism management research. International Journal of Contemporary Hospitality Management, 35(8), 2929-2943. https://doi.org/10.1108/ijchm-08-2022-1006
21. Munsarif, M., Sam'an, M., & Safuan, S. (2022). Peer to peer lending risk analysis based on embedded technique and stacking ensemble learning. Bulletin of Electrical Engineering and Informatics, 11(6), 3483-3489. https://doi.org/10.11591/eei.v11i6.3927
22. Pavlyshenko, B. (2019). Machine-learning Models For Sales Time Series Forecasting. Data, 1(4), 15. https://doi.org/10.3390/data4010015
23. Raizada, S. and Saini, J. R. (2021). Comparative Analysis Of Supervised Machine Learning Techniques For Sales Forecasting. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/10.14569/ijacsa.2021.0121112
24. Rodrigues, A. (2021). Food Sales Forecasting Using Machine Learning Techniques: A Survey. IJRASET, 9(9), 869-872. https://doi.org/10.22214/ijraset.2021.38069
25. Sarker, I. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. Sn Computer Science, 2(5). https://doi.org/10.1007/s42979-021-00765-8
26. Schmidt, A., Kabir, M., & Hoque, T. (2022). machine learning based restaurant sales forecasting. machine learning and Knowledge Extraction, 4(1), 105-130. https://doi.org/10.3390/make4010006
27. Silva, R., Ribeiro, M., Larcher, J., Mariani, V., Coelho, L. (2021). Artificial Intelligence and Signal Decomposition Approach Applied To Retail Sales Forecasting. https://doi.org/10.21528/cbic2021-25
28. Trento, L., Pereira, G., Jabbour, C., Ndubisi, N., Mani, V., Hingley, M., … & Souza, M. (2021). Industry-retail symbiosis: what we should know to reduce perishable processed food disposal for a wider circular economy. Journal of Cleaner Production, 318, 128622. https://doi.org/10.1016/j.jclepro.2021.128622
29. Vineeth, V., Kusetogullari, H., & Boone, A. (2020). forecasting sales of truck components: a machine learning approach. https://doi.org/10.1109/is48319.2020.9200128
30. Wang, F., Aviles, J. (2023). Contrasting Univariate and Multivariate Time Series Forecasting Methods For Sales: A Comparative Analysis. ASIR, 2(7), p127. https://doi.org/10.22158/asir.v7n2p127
31. Wasesa, M., Andariesta, D., Afrianto, M., Haq, I., Pradipta, J., Siallagan, M., … & Putro, U. (2022). predicting electricity consumption in microgrid-based educational building using google trends, google mobility, and covid-19 data in the context of covid-19 pandemic. Ieee Access, 10, 32255-32270. https://doi.org/10.1109/access.2022.3161654