International Scientific Journal of Engineering and Management ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024 DOI: 10.55041/ISJEM01801
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A STUDY ON COMPETITOR ANALYTICS AS A DECISION MAKING CATALYST IN A LEATHER MANUFACTURING COMPANY

*SHARLI M **Mrs. J. SOFIA VINCENT*

**MBA Student, Panimalar Engineering College, Chennai.*
***Assistant Professor, Panimalar Engineering College, Chennai.*

ABSTRACT

This study investigates the impact of competitor analytics on decision-making at a leather manufacturing company, using machine learning algorithms such as Random Forest, Support Vector Regression (SVR), and XGBoost. Trained on a dataset of 5,871 samples, the models were evaluated with Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). SVR emerged as the most efficient model. Python libraries NumPy and Pandas in Jupyter Notebook facilitated the data analysis. The SVR model provided valuable insights for strategic decision-making and competitor performance predictions. This study underscores the importance of competitor analytics and machine learning in enhancing manufacturing strategies.

Keywords: competitor analytics, machine learning algorithms, strategic decision making, predictive modeling, python, data science

INTRODUCTION

In today's competitive landscape, manufacturing companies must make informed strategic decisions to stay ahead. This study explores the role of competitor analytics, using machine learning algorithms such as Random Forest, Support Vector Regression (SVR), and XGBoost, in enhancing decision-making processes. With a dataset of 5,871 samples, the study employed Python libraries NumPy and Pandas within Jupyter Notebook for data analysis, using metrics like RMSE and MAE for evaluation.

The findings highlight the efficacy of SVR in generating accurate predictions and valuable insights for strategic decisions in manufacturing. By leveraging machine learning, companies can better understand competitor behavior, market trends, and consumer preferences, enabling them to anticipate market shifts and mitigate risks. This proactive approach enhances resource allocation, reduces costs, and improves competitive positioning.

Moreover, integrating competitor analytics fosters continuous improvement and innovation by identifying market gaps and opportunities. The study emphasizes the need for manufacturing companies to invest in data infrastructure and develop a data-driven culture to fully realize these benefits. As technological advancements and data availability grow, the potential for actionable insights will increase, driving strategic decision-making and sustainable growth in the manufacturing sector.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

REVIEW OF LITERATURE

David Rey- Blanco(2024), By addressing the complexities of developing comprehensive location indices for house price modeling, which stem from issues like achieving consensus, selecting variables, and data granularity. The study propose a novel method using computer algorithms to create car and walk accessibility indices. These indices enhance prediction accuracy by 13% in regression models and 21.6% in random forests, while also reducing spatial autocorrelation by 35%. The indices are interpretable, scalable, and applicable beyond residential real estate, offering valuable insights for urban analysis and different property types across various geographic areas. **Jean- Laurent Viviani(2024),** This paper introduces an innovative approach to precisely forecast gold price movements and interpret predictions. Initially, it evaluates six machine learning models, including the recent eXtreme Gradient Boosting (XGBoost) and CatBoost methods. Empirical results demonstrate the superiority of XGBoost over other advanced models. Additionally, it suggests employing Shapley additive explanations (SHAP) to aid policymakers in interpreting complex machine learning model predictions and assessing the importance of various features influencing gold prices. **Guna Sekhar Reddy Thummala(2023),** This research project focuses on utilizing random forest algorithms to predict cardiac illness and evaluates their performance compared to logistic regression. Both methodologies are examined, with two groups statistically evaluated, each consisting of 20 samples. The logistic regression model achieved an 80% accuracy in predicting heart disease, while the random forest classifier achieved a mean accuracy of 87.64% for the same prediction. . Initially, it evaluates six machine learning models, including the recent eXtreme Gradient Boosting (XGBoost) and CatBoost methods. Empirical results demonstrate the superiority of XGBoost over other advanced models. **Dragana Rajkovic(2023),** This study analyzed the quality of rapeseed oil from 40 genotypes over four years using two machine learning techniques: artificial neural network (ANN) and random forest regression (RFR). These models predicted fatty acid content (C16:0, C18:0, C18:1,C18:2, C18:3, and C22:1), α-tocopherol, γ-tocopherol, and total tocopherols based on production year and rapeseed genotype. The developed models demonstrated high predictive performance, with strong r2 values during training. machine learning algorithms were employed to thoroughly analyze datasets across training, testing, modeling, and cross-validation stages. Four performance metrics— MAE, MSE, MAPE, and R2—were used to assess and compare the algorithms' accuracy.Based on the coefficient of determination results, the MDT algorithm achieved the highest prediction accuracy of 0.9284, followed by LightGBM with an accuracy of 0.8765, and XGBoost with an accuracy of 0.8493. The MDT can serve as a decision support tool for equipment sellers, buyers, and owners, aiding in equipment life cycle analysis and decision-making related to selling, purchasing, overhauling, repairing, disposing, and replacing equipment.

OBJECTIVES OF THE STUDY

- To evaluate the performance of the most used predictive analytics models like Random Forest, SVR(Support Vector Regression) and XGboost algorithm.
- To conduct a comparative analysis of these models by evaluating performance metrics like RMSE(Root Mean Squared Error) and MAE(Mean Absolute Error).
- To give suggestions for the company on the best model.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

NEED OF THE STUDY

Harnessing machine learning (ML) enables businesses to gain a competitive edge by meticulously analyzing rivals' strategies, strengths, weaknesses, and market positioning. ML algorithms efficiently process vast datasets, unveiling market trends, client preferences, and demand fluctuations. Predictive analytics forecasts competitors' future moves and market paths. Insights gleaned empower product or service adjustments to meet consumer demands. Continuous monitoring and improvement through ML algorithms ensure agility in responding to dynamic competitive landscapes. By scrutinizing competitors' moves and strategies, organizations unearth market gaps oropportunities for surpassing rivals. ML-driven evaluation of brand positioning amplifies uniqueness and engagement with target audiences, enhancing market presence. Moreover, ML algorithms dissect global market patterns and competitor activities, furnishing invaluable intelligence for organizations operating in diverse foreignmarket places.

SCOPE OF THE STUDY

This study explores using Python libraries and machine learning models for competitor analysis in various business settings. It covers data preprocessing, feature engineering, and model development techniques tailored for this purpose. Evaluation includes using confusion matrices and accuracy rates to gauge predictive accuracy. The aim is to offer valuable insights and recommendations for organizations adopting ML-driven competitive intelligence strategies. Although focused on the manufacturing sector, itdoesn't encompass all machine learning techniques. Leveraging Python libraries like scikit-learn, it delves into strategic implications and actionable insights derived from competitor analytics.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

RESEARCH METHODOLOGY

Research Design:

Analytical research is a methodical, scientific approach to investigation that entails evaluating data, facts, or information in order to make significant inferences, spot trends, and obtain new perspectives. It uses methods like statistical analysis, data modeling, and experimentation to try to make sense of complicated occurrences by breaking them down into their constituent parts and linking them together.

Data Collection:

This study employed secondary data collection methodology for our data collection. In the context of research, secondary data is information that has been gathered for a different reason by somebody else but is used by researchers for their own studies. Since this kind of data has already been obtained, cleaned, and saved, it is valuable since it saves time and resources. We collected the data for this research from a manufacturing company. Data regarding FN(FINANCIAL YEAR), DATE, CODE NUMBER,COMPANY NAME, COMPANYGROUP, CUSTOMERS GROUP, COUNTRY, CONTINENT, PRODUCT DESCRIPTION, PRODUCT TYPE, QTY MNPRS, VALUE(USD) UNIT PRICECATEGORY, NAN, NAN, GENDER, are all included in the framed data set.

Tools Applied:

This study uses Random forest techniques, Support vector regression(SVR) and XGboost algorithm machine learning (ML) models to analyze the data set. This study uses numpy, pandas, seaborn, matplotlib and scikit-learn, among other Python packages, for data preprocessing, feature engineering, and model development which is illustrated below:

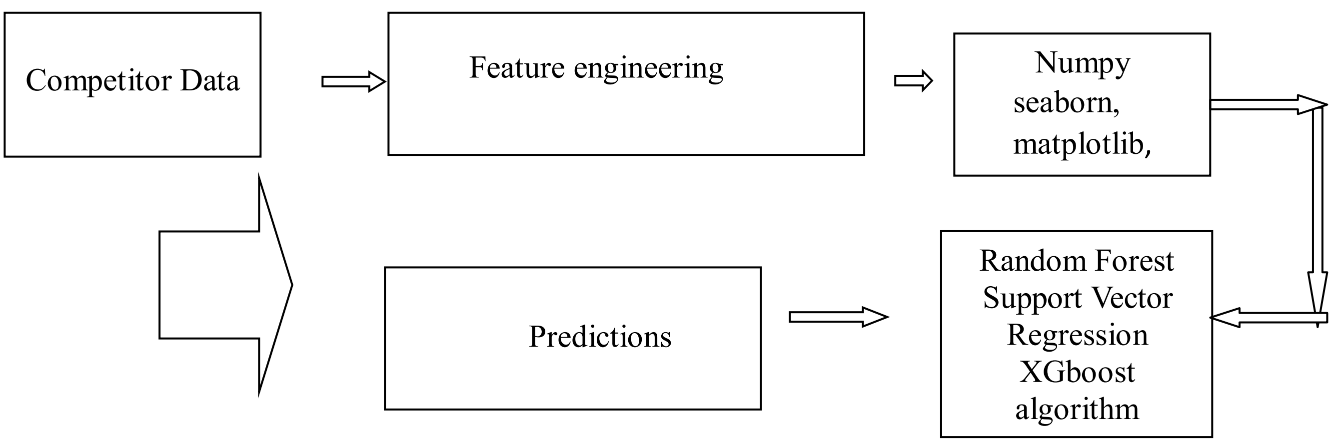

Fig 1. Tools applied

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

DATA ANALYSIS AND INTERPRETATION

RANDOM FOREST

Random Forest is a versatile and powerful machine learning algorithm commonly used for both classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode (classification) or average (regression) prediction of the individual trees. Each decision tree in the Random Forest is trained on a random subset of the training data and a random subset of features, which introduces diversity and reduces the risk of overfitting.

TABLE 1 RANDOM FOREST VALUES

Metrics	Values(training)	Values(testing)	Values(validation)
Root Mean Squared Error(RMSE)	1203.5269547700136	951.5929886120165	951.592988612016
Mean Absolute Error(MAE)	455.14363231673927	482.9316302825129	482.9316302825129

Interpretation:

The RMSE values indicate moderate prediction errors for the Random Forest model, with the training set at 1203.53 and both test and validation sets at 951.59. Similarly, the MAE values show consistent accuracy, with 455.14 for the training set and 482.93 for both test and validation sets. These metrics suggest reasonable model performance but highlight the need for further refinement to enhance real-world applicability.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

Support Vector Regression(SVR)

A Support Vector Machine (SVM) is a supervised machine learning technique employed for classification and regression assignments. SVM operates by identifying a hyperplane within a multi-dimensional space that effectively divides data into distinct classes. Its objective is to optimize the margin (the gap between the hyperplane and the nearest data points of each class) while reducing classification errors. SVM is capable of addressing both linear and non-linear classification challenges through the utilization of diverse kernel functions. SVM regression is classified as a nonparametric method due to its dependency on kernel functions.

TABLE 2 SUPPORT VECTOR REGRESSION(SVR) VALUES

Metrics	Values(Training)	Values(Testing)	Values(Validation)
Root Mean Squared Error(RMSE)	1420.93319163255	936.7201994936187	1029.5907898049362
Mean Absolute Error(MAE)	433.466226637449	414.0127989057772	440.07515913512395

Interpretation:

The SVR model shows moderate prediction errors, with RMSE values ranging from 936.72 to 1420.93 and MAE values from 414.01 to 440.08. The lowest errors are observed in the testing set, indicating better performance on unseen data. These metrics suggest that while the model is reasonably accurate, there is room for improvement.

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 05 | May – 2024
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

XGBoost algorithm

XGBoost is a meticulously crafted distributed gradient boosting library engineered for the effective and scalable training of machine learning models. This ensemble learning technique amalgamates the predictions of numerous weak models to generate a more robust prediction. Renowned as "Extreme Gradient Boosting," XGBoost has risen to prominence as one of the most favored and extensively applied machine learning algorithms.

TABLE 3 XGBOOST ALGORITHM VALUES

Metrics	Values(Training)	Values(Testing)	Values(Validation)
Root Mean Squared Error(RMSE)	1133.678949342474	951.5609182855438	1394.533669897692
Mean Absolute Error(MAE)	444.9907559544394	480.661730541976	478.817133073567

Interpretation:

The XGBoost model shows moderate prediction errors across datasets. Training RMSE is approximately 1133.68, with MAE around 444.99. The validation set has higher errors, with RMSE at 1394.53 and MAE at 478.82. The testing set, however, shows relatively lower errors, with RMSE around 951.56 and MAE about 480.66. These metrics highlight variability in performance across different data sets, suggesting potential for further model optimization.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

FINDINGS

TABLE 4 COMPARISON OF ALGORITHMS WITH RMSE AND MAE

Metrics	Random Forest	Support Vector Regression(SVR)	XGBoost algorithm
Training RMSE	1203.526954	1420.93319163	1133.678949
Testing RMSE	951.592988	936.72019949	951.5609182
Validation RMSE	951.595988	1029.5907868	1394.53366989
Training MAE	455.1439323	433.466226637	444.9907559544
Testing MAE	482.931630	414.01298905777	480.661730541
Validation MAE	482.931630	440.075159135	478.8171330735

The Support Vector Regression (SVR) model is the top performer, exhibiting the lowest RMSE and MAE values across all datasets compared to Random Forest and XGBoost. This indicates the smallest average deviation and absolute difference between predicted and actual values. These metrics highlight SVR's superior predictive accuracy and generalization capabilities, making it the most suitable choice for the given task. Its effectiveness in capturing underlying data patterns underscores its status as the best-performing algorithm in this scenario.

SUGGESTIONS

The company should implement the Support Vector Regression (SVR) algorithm for decision support in manufacturing, as it efficiently predicts performance based on competitor analytics. Investing in robust infrastructure for collecting and analyzing competitor data using advanced tools is crucial for extracting actionable insights. Continuous monitoring and updating of machine learning models are necessary to maintain accuracy and relevance amidst evolving business needs. Collaborating with manufacturing experts can enrich the interpretation of model results, translating findings into practical strategies. Exploring ensemble techniques like model stacking or blending can enhance predictive performance. Prioritizing model interpretability ensures that insights are understandable and actionable for decision-makers. Additionally, investing in employee training programs to build proficiency in data analytics and machine learning will empower the company to leverage advanced tools effectively and drive strategic decision-making.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

CONCLUSION

The study highlights the significant potential of machine learning algorithms, especially Support Vector Regression (SVR), in driving strategic decision-making in manufacturing through competitor analytics. By analyzing competitor data, companies can gain valuable insights into performance and operations. The study emphasizes investing in robust analytics infrastructure, continuous model updates, and collaboration with domain experts to maximize machine learning utility. Enhancing model interpretability, investing in employee training, and validating models in real-world scenarios are crucial. The findings underscore the importance of data-driven decision-making for gaining a competitive edge, uncovering hidden patterns, and making informed decisions that drive efficiency and innovation. Embracing a culture of continuous learning and adaptation is essential. The study demonstrates the transformative impact of competitor analytics and machine learning in optimizing processes and fostering growth, urging companies to invest in analytics capabilities and stay agile to thrive in a data-driven landscape.

REFERENCES

- Davenport, T. H., & Harris, J. G. (2007). Competing on Analytics: The New Science of Winning. Harvard Business Press.
- Schroeder, R. G., Rungtusanatham, M. J., & Goldstein, S. M. (2019). Operations Management in the Supply Chain: Decisions and Cases. McGraw-Hill Education.
- Porter, M. E. (1998). Competitive Advantage: Creating and Sustaining Superior Performance. Free Press.
- Liaw, A., & Wiener, M. (2002). Classification and regression by RandomForest. R news, 2(3), 18-22.
- Smola, A. J., &Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3), 199-222.
- Chen, T., &Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
- "Rey-Blanco, David, José L. Zofío, and Julio González-Arias. "Improving hedonic housing price models by integrating optimal accessibility indices into regression and random forest analyses." *Expert Systems with Applications* 235 (2024): 121059.
- "Jabeur, Sami Ben, Salma Mefteh-Wali, and Jean-Laurent Viviani. "Forecasting gold price with the XGBoost algorithm and SHAP interaction values." *Annals of Operations Research* 334.1 (2024): 679-699.
- "Thummala, Guna Sekhar Reddy, Radhika Baskar, and S. RimlonShibi. "Predictionof Heart Disease using Random Forest in Comparison with Logistic Regression to Measure Accuracy." *2023 International Conference on Advances in Computing,Communication and Applied Informatics (ACCAI)*. IEEE, 2023.
- "Rajković, Dragana, et al. "Artificial neural network and random forest regression

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

models for modelling fatty acid and tocopherol content in oil of winter rapeseed." *Journal of Food Composition and Analysis* 115 (2023): 105020.

ISSN: 2583-6129
DOI: 10.55041/ISJEM01801

-