Global Scientific JOURNALS

GSJ: Volume 12, Issue 2, February 2024, Online: ISSN 2320-9186
www.globalscientificjournal.com

IMPROVING BUSINESS INSIGHTS USING MACHINE LEARNING ALGORITHMS

Ismail, I.A., Chukwuka, E.E., Abiola, M.K

Abstract

In today's data-driven world, the utilization of business intelligence and machine learning has become imperative for companies seeking a competitive edge and informed decision-making. This article examines the employment of machine learning algorithms towards improving business actionable insights. An apple dataset and three classifiers namely, Logistic Regression, K-Nearest Neighbor, and Naïve Bayes were employed. The performance of developed models is evaluated using precision, recall, f1-score, and accuracy metrics. Results indicate that all three classifiers offer better performances, although K-Nearest Neighbor outperforms the others with a 90% accuracy, precision, recall and f1-score rates. This suggests that K-Nearest Neighbor is the best classifier among the three, with the highest fitness for the dataset used in this study. This study also highlights the significance of integrating business analytics and machine learning in driving business success and providing better business insights in business organizations.

Keywords: business intelligence**,** business insight, machine learning, logistic regression, k-nearest neighbor, naïve bayes

1. Introduction

Business decisions in the past were frequently made using intuition and foresight—a method referred to as the "traditional approach." This approach has drawn criticism for assuming rationality and full knowledge, which may not necessarily match actual business situations. According to Jones and George (2017), the conventional method has long been an essential component of organizational strategy since it provides a well-organized framework for weighing possibilities and deciding on the optimal course of action. But the contemporary business environment necessitates creative approaches to decision-making.

Making sound business decisions involves a variety of skills, including creativity, critical thinking, problem-solving skills, and technological involvement beyond human capabilities. This has brought about the data revolution, the emergence of digital technologies, and the introduction of business intelligence tools, predictive analysis, and machine learning.

Utilization of business intelligence and machine learning tools has become essential for companies looking to acquire a competitive edge and support well-informed decision-making in today's data-driven world. It appears that businesses use analytics to use previous data to make decisions more quickly and get insights. It is impossible to overstate the importance of business analytics in the modern world, since all ethical companies strive to glean insights from their data to inform decisions and gain a competitive edge. Because it enables data-driven decision-making, competitive advantage, enhanced performance, better customer understanding, and risk management, business analytics is therefore critical to businesses.

According to Eitle and Buxmann (2019), modern businesses must adapt their business processes to be competitive and adaptable due to the rapid pace of business changes and the continuous digital transformation of the global environment. Businesses have implemented business analytics using historical data in an effort to boost performance. Business analytics is the process of turning data into insights to enhance business choices. Analytics is defined as the application of mathematics, statistics, and machine learning to discover significant patterns in data.

A subfield of computer science and artificial intelligence known as "machine learning" deals with the technique of using past data to teach machines in order to predict the future. Bharadiya (2023) claims that machine learning algorithms are used to examine past data and spot patterns and trends because they help companies forecast future results with accuracy. By predicting demand, maximizing inventory levels, and enhancing supply chain management, predictive analytics assists companies in streamlining their operations (Mungoli, 2023).

The process of implementing a business vision begins with a traditional system that uses written papers as data. For the purpose of deriving insights, this technique is laborious and ineffective. Business intelligence tools such as Power BI, Tableau, Excel, and many more are used nowadays. These tools are effective and efficient when compared to traditional methods, and they are useful for dashboarding and data visualization. However, there are certain drawbacks to this strategy, including its inability to anticipate the future and promote products. On the other hand, as algorithms learn and change over time with fresh data inputs, the integration of business analytics and machine learning enables constant development and adaptation.

Organizations can fully utilize their data thanks to the partnership between business analytics and machine learning, which spurs innovation, streamlines processes, and improves consumer experiences. Businesses may find hidden patterns, forecast future trends, and automate decision-making processes by utilizing machine learning algorithms and advanced analytics methodologies. As a result, this article investigates the potential applications of machine learning algorithms to historical data in order to provide better business insights for commercial organizations.

2. Literature Review

2.1 Business Analytics

Business analytics gives firms the ability to turn structure and unstructured data into insightful knowledge by facilitating data-driven decisions that enhance overall performance. Business analytics, according to Wolniak and Grebski (2023), is the process of applying statistical techniques and data analysis to obtain insightful knowledge and make defensible business decisions. In order to find trends, patterns, and correlations that can inform strategic planning and operational enhancements, business analytics comprises the investigation, analysis, interpretation, and visualization of data from several sources (Drozd & Wolniak, 2021; Gajdzik & Wolniak, 2021, 2022).

Business analytics, an interdisciplinary field that integrates machine learning, statistics, information systems, operations research, and management science, is crucial in enabling this new approach to decision-making (Schmitt, 2023). It can be broadly classified into three categories: descriptive, predictive, and prescriptive analytics (Delen & Ram, 2018). Other category of business analytics is diagnostic analytics.

Using its three functions—descriptive, such as "who are the most profitable customers that we need to protect?"; predictive, such as "what will be the impact of a new pricing strategy on our most profitable customers?"; and prescriptive, such as "how can we reach the most profitable customers at the lowest cost?"; business analytics can offer useful insights in strategic decision making (Kunc & O'brien, 2019). Additionally, diagnostic analytics shed light to root of problems.

In order to comprehend prior performance and what has transpired in a business, descriptive analytics entails the analysis and interpretation of historical data. Data visualization tools like charts, graphs, and dashboards are frequently used in this process (Wolniak & Grebski, 2023).

Furthermore, the process of making predictions about future events or results is known as predictive analytics, and it involves using historical data and statistical algorithms. Predictive analytics helps businesses foresee possible outcomes and trends by finding patterns and relationships in the data. They can now plan ahead and make better decisions by doing so, such as forecasting consumer behavior, product demand, or financial performance (Wolniak & Grebski, 2023).

As expressed by Wolniak & Grebski, (2023), prescriptive analytics offers actionable recommendations that assist decision-makers in optimizing efficiency, minimizing risks, and accomplishing strategic objectives. It uses sophisticated algorithms and decision models to ascertain the optimal course of action under various conditions.

2.2 Machine Learning

Machine learning is a subfield of artificial intelligence that has the ability to learn from historical data and draw out patterns for the purpose of making insights with less human intervention. A key component in boosting and extending business intelligence's capabilities is machine learning. It helps businesses generate precise forecasts, automate procedures, and glean insightful information from massive volumes of data (Bharadiya, 2023). To help machines learn and manage data more effectively, machine learning (ML) is utilized (Mahesh, 2020). This implies that big data is interpreted and explained using machine learning in order to extract information from the data. Supervised and unsupervised learning are the two common types of machine learning.

Supervised learning is a machine learning technique that trains algorithms and generates predictions using labeled datasets. The input dataset is split into train and test datasets for supervised learning algorithms. The output variable in the train dataset has to be classified or forecasted using a new or test dataset. Figure 1 depicts the steps involved in the supervised learning method:

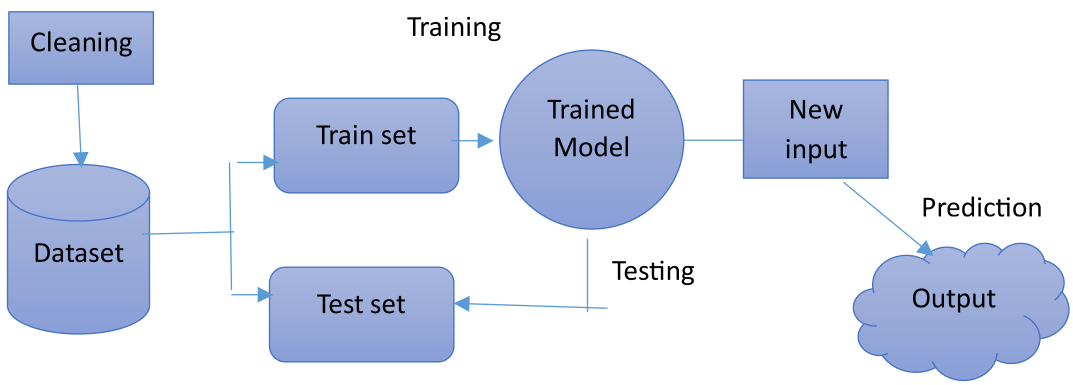

Figure 1: Supervised machine learning process

Typical supervised machine learning algorithms include k-nearest neighbor, naïve bayes, decision trees, and support vector machines. The primary uses of **support vector machines** are in regression analysis and classification. By implicitly translating their inputs into high-dimensional feature spaces, it is able to effectively carry out a non-linear classification utilizing a technique known as the kernel trick. In essence, it draws boundaries between the different classes (Mahesh, 2020).

A **decision tree** is an algorithm that resembles a tree and shows decisions and their outcomes as a tree. Every node in the tree represents a decision or an event, and the edges stand for the circumstances or rules governing decisions. Every tree is made up of nodes and branches. Depending on the problem, decision tree leaf nodes have classes, probabilities, or continuous values in case of regression (Badilo, Benfai, Birzele, Davydov, Hutchinson, Kam-Thong, Siebourg-Poister, Steiert, and Zhang, 2020).

The **Naïve Bayes** approach relies on the premise of predictor independence and is based on the Bayes Theorem. According to Mahesh (2020), the naive Bayes classifier makes the assumption that a feature's presence in a class is independent of the existence of any other feature. Another supervised learning technique, **K-Nearest Neighbor (KNN),** takes advantage of nearness to predict or classify how a single data point will be grouped.

However, unlike supervised learning, unsupervised learning algorithms do not require labeled input; instead, they identify the class of the data by using the few properties they have previously learned from historical data. Clustering and feature reduction are its primary applications. The common example of unsupervised learning is Principal Component Analysis (PCA). The primary function of PCA is feature reduction or dimensionality reduction.

In 2020, Mahesh defined PCA as a statistical technique that reduces the dimension of data as a result of reduced features in order to facilitate quicker and simpler computations. Dropping uninformative features could improve the model's performance and convergence time (Badilo et al., 2020). The K-means algorithm is another unsupervised learning algorithm employed for resolving the well-known clustering problem.

Ensemble learning, on the other hand, enhances another model's performance or lessens the possibility of choosing a bad one by accident. Ensemble learning involves multiple models merged in some fashion, like averaging and voting (Ganaie, Hu, Malik, Tanveer, and Suganthan, 2022), to tackle a specific computer intelligence problem (Mahesh, 2020). The algorithms used in ensemble learning are bagging and boosting. A class of algorithms known as "boosting" is used to turn weak learners into strong ones. Whenever a machine learning algorithm has to have its accuracy and stability improved, bagging is used. It works well with regression and classification.

2.3 Machine learning contribution towards improving business insights

Large-scale corporate data analysis, complex behavior pattern detection, and individual preference identification are all areas in which machine learning algorithms shine, enabling the creation of extremely detailed client groups. By doing this, businesses can enhance the consumer experience by customizing marketing efforts and making product recommendations.

Businesses may streamline procedures, reduce expenses, find new sources of income, and provide outstanding customer experiences by utilizing machine learning and artificial intelligence (AI) in business intelligence. Businesses that strategically use cutting-edge technologies gain a competitive advantage and are better positioned to prosper in the data-driven, quickly changing business environment.

2.4 Related works

While little research had used machine learning algorithms for business analytics, the majority of the reviewed articles theoretically addressed the application and contribution of machine learning in business analytics. Eitle and Buxmann (2019) put forth a proposal titled "Business Analytics for Sales Pipeline Management in the Software Industry: A Machine Learning Perspective”. The CatBoost, Random Forest, Support Vector Machine, XGBoost, and Decision Tree algorithms were compared in the study. The results demonstrated the superiority of the Random Forest method (63% AUC) and CatBoost algorithm (88%) over other supervised classifiers including Support Vector Machine, XGBoost, and Decision Tree.

Żbikowski and Antosiuk (2021) created a machine learning-based predictive model to predict a company's performance. Three techniques were compared in the study: gradient boosting classifiers, logistic regression, and support vector machines. According to the study, the best model had precision, recall, and F1 scores of 57%, 34%, and 43%, respectively. The gradient boosting classifier produced the best results.

Schmitt (2023) examined the potential of AutoML for business analytics applications. Three real-world datasets were used as benchmarks against a manually adjusted stacked ML model. The study found that the H2O AutoML framework is a useful instrument for facilitating quick prototyping, which could reduce the duration of development.

3. Methodology

This article uses an apple dataset that was kindly donated by an American agriculture company. The dataset was acquired from the Kaggle repository to illustrate how machine learning may help with business insight and competitive advantage. The dataset contains 4000 instances and 9 variables. The variables in consideration are weight, size, juiciness, maturity, acidity, crunchiness, sweetness, and quality of fruits. As seen in figure 2, the number of good fruit (represented by 0) is 2,006 and the number of bad fruit (represented by 1) is 1,996, affirming the balance of the dataset in use which in turn assure accuracy in prediction.

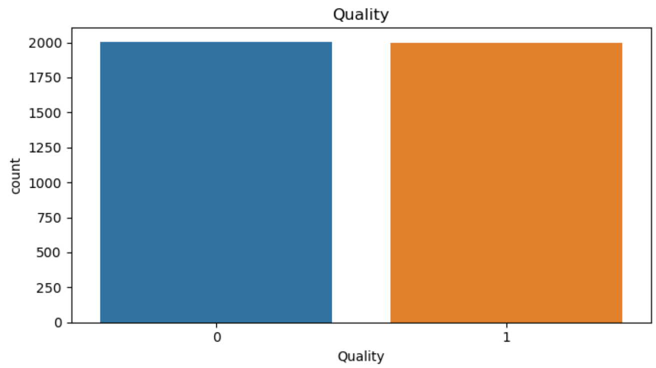

Figure 2: Balanced dataset

However, the amount of data in the training and validation sets is displayed in Figure 3. The figure shows that, correspondingly, x_train and x_test is 2800 and 1200. Likewise, y_train and y_test is, correspondingly, 2800 and 1200.

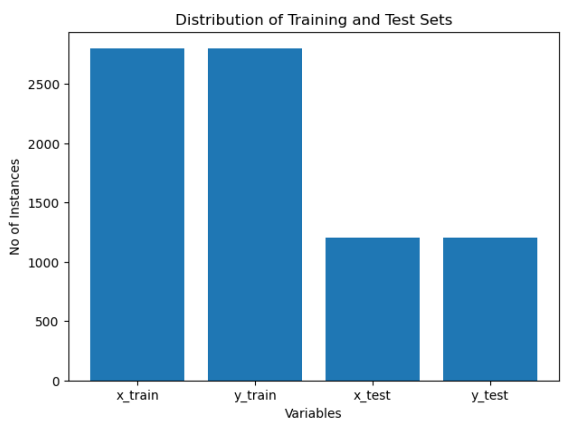

Figure 3: Train-Test set split

Ultimately, for this article, three classifiers—Logistic Regression, K-Nearest Neighbor, and Naïve Bayes—were selected, and metrics for accuracy, precision, recall, and f1-score were used to assess each one's performance.

4. Results and discussion

Figure 4.1: Model 1 – Logistic Regression

Logistic Regression Classification Report

	precision	recall	f1-score	support
0	0.75	0.76	0.76	607
1	0.75	0.74	0.75	593
accuracy			0.75	1200
macro avg	0.75	0.75	0.75	1200
weighted avg	0.75	0.75	0.75	1200

Figure 4.1 gives an overview of the performance of the first model developed with Logistic Regression in terms of precision, recall, f1-score and accuracy. Taking weighted average into consideration, Logistic Regression had 75% precision, 75% recall, 75% f1-score and 75% accuracy.

Figure 4.2: Model 2 – K-Nearest Neighbor

KNN Classification Report

	precision	recall	f1-score	support
0	0.90	0.89	0.90	607
1	0.89	0.90	0.89	593
accuracy			0.90	1200
macro avg	0.89	0.90	0.89	1200
weighted avg	0.90	0.90	0.90	1200

Figure 4.2 reflects an overview of the performance of the second model (K-Nearest Neighbor) in terms of precision, recall, f1-score and accuracy. However, considering the result of weighted average, KNN resulted to 90% precision, 90% recall, 90% f1-score and 90% accuracy.

Figure 4.3: Model 3 – Naïve Bayes

```
Naive Bayes Regression Classification Report
              precision    recall  f1-score   support

           0       0.76      0.75      0.75       607
           1       0.74      0.76      0.75       593

    accuracy                           0.75      1200
   macro avg       0.75      0.75      0.75      1200
weighted avg       0.75      0.75      0.75      1200
```

Figure 4.3 shows the overview of the performance of the third model, Naïve Bayes in terms of precision, recall, f1-score and accuracy. Hence, considering the result of weighted average, Naïve Bayes had 75% precision, 75% recall, 75% f1-score and 75% accuracy.

Table 4.1: Models comparison

	Precision	Recall	F1-score	Accuracy
Logistic Regression	75%	75%	75%	75%
K-Nearest Neighbor	90%	90%	90%	90%
Naïve Bayes	75%	75%	75%	75%

Table 4.1 provides the comparison of the performances of all three developed models in terms of precision, recall, f1-score and accuracy. When comparing classification techniques, all the three algorithms offer similar performances and exceed the baseline. Taking accuracy into account, KNN had 90% to outperform the other classifiers, meanwhile similar best result was also reached by KNN in the other performance metrics including precision, recall and f1-score. Consequently, KNN is the best classifier that with highest fitness to the data in use among other classifiers.

Based on the aforementioned findings, firms may quickly identify and reduce risks associated with selling defective items, as well as facilitate the prediction of possible supply chain disruptions, by having instantaneous knowledge of the frequency of good and bad goods via predictive models. Predicting the quality of items before they are offered for sale also helps firms optimize their operations by pointing out inefficiencies and potential improvement areas. Most essential, though, is that it makes it possible to customize commodities, services, or products to better satisfy customers.

5. Conclusion

This study underscores the contribution of machine learning algorithm in driving informed decision-making and gaining a competitive edge in today's data-driven business landscape. The integration of analytics and machine learning enables businesses to extract valuable insights from historical data, predict future trends with accuracy, and streamline operations for improved performance.

This article demonstrates the potential of machine learning algorithms in enhancing business insights and gaining a competitive advantage. By leveraging machine learning algorithms, organizations can uncover hidden patterns, forecast future trends, and automate decision-making processes. The results also highlight the importance of selecting the right classifier for predictive modeling, with K-Nearest Neighbor demonstrating the highest performance among the classifiers evaluated by 90%.

In conclusion, this article advances company insight through the application of machine learning since better meeting client wants requires tailoring products or services based on predicted insights. It is therefore recommended that future studies should use different large datasets and techniques to get even greater improvement.

References

Badillo, S., Banfai, B., Birzele, F., Davydov, I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. (2020). An Introduction to Machine Learning. *Clinical Pharmacology & Therapeutics*. https://doi.org/10.1002/cpt.1796

Bharadiya, J. (2023). Machine learning and AI in business intelligence: Trends and opportunities. *International Journal of Computer (IJC)*, *48*(1), 123-134.

Delen, D., & Ram, S. (2018). Research challenges and opportunities in business analytics. Journal of Business Analytics, 1(1), 2–12. https://doi.org/10.1080/2573234x.2018.1507324

Drozd, R., & Wolniak, R. (2021). Systematic assessment of product quality. Journal of Open 11 Innovation: Technology, Market, and Complexity, 7(4), 1-12.

Eitle, & and Buxmann, (2019). Business Analytics for Sales Pipeline Management in the Software Industry: A Machine Learning Perspective. *Proceedings of the 52nd Hawaii International Conference on System Sciences, 1013-1022,* 978-0-9981331-2-6. https://hdl.handle.net/10125/59541

Gajdzik, B., & Wolniak, R. (2021). Digitalization and innovation in the steel industry in 21 Poland - selected tools of ICT in an analysis of statistical data and a case study. Energies, 22 14(11), 1-25.

Ganaie, M., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151.

Jones, G., & George, J. (2019). *Essentials of contemporary management*. McGraw-hill.

Kunc, M., & O'brien, F. A. (2019). The role of business analytics in supporting strategy processes: Opportunities and limitations. *Journal of the Operational Research Society*, *70*(6), 974-985.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381-386.

Mungoli, N. (2023). Adaptive Ensemble Learning: Boosting Model Performance through Intelligent Feature Fusion in Deep Neural Networks. arXiv preprint arXiv:2304.02653.

Schmitt, M. (2023). Automated machine learning: AI-driven decision making in business analytics. *Intelligent Systems with Applications*, *18*, 200188.

Wolniak, R., & Grebski, W. (2023). The five stages of business analytics. *Silesian University of Technology Scientific Papers. Organization and Management Series*, *178*, 735-752.

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management*, *58*(4), 102555.