# A Fine Tuned-based Framework to Predict Salesforce Data using Machine Learning in Business Analytics

**Naveen Kumar**
Salesforce Inc., Dallas, Texas, USA
nkumar5@salesforce.com (corresponding author)

*Received: 8 September 2024 | Revised: 12 October 2024 | Accepted: 21 October 2024*

 *DOI: https://doi.org/10.48084/etasr.8948*

## ABSTRACT

**Sales forecasting is one of the critical areas in business analytics where business organizations aim to enhance efficiency and, therefore, revenues. An excellent example of a CRM program is Salesforce, which produces massive amounts of sales data that are essential for forecasting and decision-making. Data analysis involves the use of complex and effective tools for its processing. This study proposes a framework based on the following classification algorithms: Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF). The proposed framework follows a fine-tuned approach to improve the prediction of sales data. Regarding the fine-tuning of these algorithms, it was observed that specific changes were required within the hyperparameters to better relate to the inherent patterns and other factors that exist in the sales data. The optimization process was very crucial in improving the performance of the model. The proposed framework was used on a sales dataset and evaluated in terms of accuracy, precision, data loss, and F1 score. Fine-tuned algorithms had higher accuracy and lower data loss.**

***Keywords-SVM; RF; DT; fine-tuning; sales data; CRM; salesforce***

## I. INTRODUCTION

Data in the modern corporate world, where development takes place at a high pace and organizations moving towards the digital world, have become an essential factor in strategic management. Companies collect huge volumes of data from tasks, clients, markets, and social networks [1]. As with all things, there are advantages and disadvantages to the large amount of data being dealt with today; the former is the potential for creative solutions and more insightful knowledge, whereas the latter falls in the form of more difficulties in handling them. This has necessitated business analytics as a discipline to help organizations comprehend large amounts of data and make the right decisions to enhance performance and gain competitive advantage [2]. Business analytics includes tools such as descriptive, predictive, and prescriptive analytics, based on statistical models and Artificial Intelligence (AI) to predict outcomes and suggest strategies. The utilization of big data technologies and Machine Learning (ML) is an essential tool to make this field more efficient by automating large dataset analysis [3]. Interconnected systems, such as Hadoop and Apache Spark, allow companies to capture and integrate many forms of transactional, social media, and sensor data, thus improving their marketing knowledge [2]. Compared to statistical approaches, ML can adjust parameters within and between tasks, such as clustering, classification, and regression of consumer audit trails, to find behavior patterns [4]. AI also increases business analytics' capabilities using technologies such as NLP and computer vision to optimize features. As a result, intelligent business analytics systems have emerged, which require information supplemented with minimal human touch [5].

This study adds to this field by presenting an ML approach that enhances the accuracy of sales forecasts by fine-tuning models such as Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF). The performance of the proposed framework was evaluated using CRM sales data with accuracy, precision, and F1 score. This work discusses the prospects of using AI business analytics for the development of new strategies and the identification of further research avenues for optimizing strategic and operational business activities.

## II. RELATED WORKS

Current ML studies in business analytics have presented various methodologies that demonstrate its influence on decision-making, business productivity, and strategies for different industries. In [6], SVM along with DT helped increase customer satisfaction and revenues, as well as better positioning on the vehicle market. In [7], real-time sales forecasting was discussed using neural networks and regression algorithms. In [8], the K-Nearest Neighbors (KNN) algorithm was used in sales data analysis, helping customer segmentation and improving strategic marketing. Similarly, the centrality of ML in enhancing the speed of predictive analytics was highlighted in [9], especially for large datasets, noting how it could be applicable across different fields. In [10], a meta-ML framework was presented for business networks, which focuses on data privacy and reduces data transmission during the big data prediction processes. In [11], a conceptual framework was proposed, which focused on the utilization of ML to improve organization-level knowledge and inform decision-making using meta-modeling. In [12], ML was used for pricing strategies, describing how businesses can enjoy competitive pricing strategies and, therefore, enhanced profitability using supervised learning models. In [13], a conceptual modeling approach was proposed to enhance organizational performance by connecting data development and computational capabilities.

The study in [14] examined how ML can help customer relations, especially through NLP and sentiment analysis, to improve services. In [11, 15], the importance of ML in real-time business data analysis was highlighted. In [16], the functions of AI and ML on innovative and productive outputs in business analytics were discussed. In [17], the role of AI and ML in managing business issues was discussed, especially during the COVID-19 pandemic, making them more scalable and efficient. The studies in [18, 19] built on previous successful implementations by discussing additional ML applications in the context of optimizing business activities. In [20], an innovative framework with ML elements was proposed to improve employee performance through predictive analysis. In [21], the contribution of ML to marketing was discussed, along with how it helps business organizations make decisions. In [22, 23], it was stated that ML brings consistency to e-commerce, especially in predicting market trends and enhancing customer satisfaction.

Analyzing the literature on the application of ML in business analytics revealed some research gaps. Existing works show how ML affects business activities positively, in terms of improving decision-making and increasing forecast accuracy and customer satisfaction. Furthermore, some studies elaborated that there is a need to develop broad frameworks for real-time operational ML data processing using large computer networks with appropriate data privacy mechanisms. Most studies encompass special uses or strategies, with few attempts to examine how the integration of two or more ML strategies can be applied in a dynamic organizational setting. In addition, current knowledge does not permit a comprehensive prediction of the long-term effects of ML projects within an organization regarding the performance or employees' use. This study proposes an ML framework comprising several ML algorithms, providing an opportunity for corporations' strategic and tactical advantage.

## III. RESEARCH METHODOLOGY

To evaluate the effects of ML on business analytics, the method was divided into specific steps, as shown in Figure 1.

- Data collection and preprocessing: A dataset was obtained from [24]. Data cleaning was performed to address missing values and outliers, followed by normalization using a standard scaling pipeline to standardize features such as TV and radio promotion budgets.
- Feature engineering: Additional features were created by analyzing historical data, including combining TV and radio promotion budgets to assess their joint impact on sales.
- Model selection and training: ML models, including RF, DT, and SVM, were applied to classify sales into three categories: low, medium, and high. These models were chosen for their scalability and interpretability in business contexts.
- Hyperparameter tuning: Fine-tuning was performed using grid search and cross-validation to optimize hyperparameters, such as the number of trees in RF, depth of DT, and regularization parameters in SVM.
- Model evaluation: Evaluation was performed using accuracy, precision, recall, and F1-score. Further optimization improved classification accuracy and reduced the number of false positive and false negative results.
- Deployment and monitoring: The trained models were integrated into business intelligence environments for both online analytical processing and batch intervention.
- Ethical considerations: Accountability, efficacy, privacy, and legal compliance were protected when collecting and processing data.

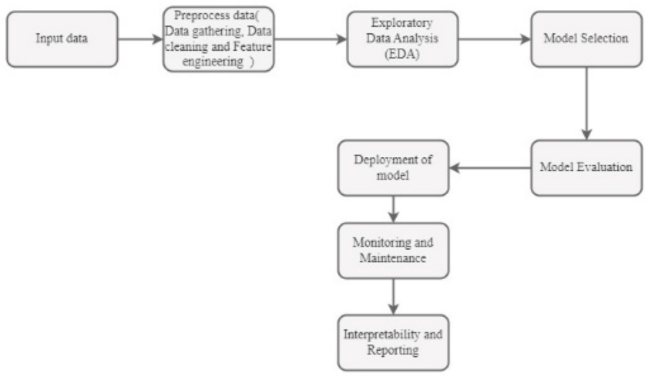


Fig. 1. Research method.

## IV. PROPOSED WORK

AI has become prevalent in the current world, where organizations rely on data to operate efficiently, maintain viable competitiveness, and control expenses. Recent years have shown a higher popularity for ML algorithms such as SVM, DT, and RF, as they are more suitable for storing and analyzing large sets of data and provide high forecast accuracy. Due to its ability to handle features with high dimensionality, ML makes it easier to classify consumer trends and sales data. RF combines DT and is well suited for large-scale data, preventing overfitting, and providing detailed information on customer behavior such as churn rates and effective marketing solutions. These algorithms improve decision-making processes and provide massive business value.

### A. Fine Tuning

The hyperparameters were adjusted to increase efficiency. Tuning involves making adjustments to improve the model generalization and improve results by decreasing the error rates.

- Fine-tuning SVM: To fine-tune an SVM with an RBF kernel, the parameters $C, \gamma$ along with other discriminating parameters should be optimized. Applying hybrid models, decision-making and business results within analytics can be optimally enhanced. The SVM model can be represented as:

  $$f(x) = \sum_{i=0}^{n} \alpha_i y_i . K(x, x_i) + b$$

  where $f(x)$ is the output to be predicted with the help of the input feature vector $x$, $\alpha_i$ is the Lagrange multiplier, $b$ is the bias, and $K(x, x_i)$ is the RBF function. The RBF kernel function can be represented as:

  $$K(x, x_i) = e^{(-\gamma||x-x_i||^2)}$$

  where $\gamma$ is known as the kernel parameter. The purpose of hyperparameter tuning is to obtain the optimal $\gamma$ and $C$ values that will enhance the performance of the SVM model. This can be done using grid search or Bayesian optimization.

- Fine-tuning DT: Suppose a DT model with Gini impurity as the criterion, which is broadly employed in business analysis. The DT model can be represented as:

  $$f(x) = \sum_{m=1}^{M} \theta_m . I(x \in R_m)$$

  where $f(x)$ represents the output, $x$ is the input feature vector and $\theta_m$ represents the predicted value for the region $R_m$. The function $I(x \in R_m)$ gives 1 if $x$ belongs to the region $R_m$ or 0 otherwise. The objective of this method is to fine-tune some of the compile time parameters which include the maximum depth of the tree, the minimum number of samples required to split an internal node, and the minimum number of samples required to be at a leaf node.

- Fine-tuning RF: Suppose an RF model consisting of $N$ DTs, which is often used in business analytics. The RF model can be represented as:

  $$f(x) = \frac{1}{N} \sum_{n=1}^{N} f_n(x)$$

  where $f(x)$ is the overall predicted output, $x$ is the input feature vector, and $f_n(x)$ is the overall predicted output of the $n^{th}$ DT. Fine-tuning aims to determine the best benchmark principal values, such as the number of DTs, maximum tree depth, number of features to consider at each split etc.

- Hybrid Algorithm: A hybrid algorithm can combine the advantages of using SVMs, DTs, and RF. The formula of the hybrid algorithm can be written as :

  $$f(x) = \alpha \cdot f_{SVM}(x) + \beta \cdot f_{DT}(x) + \gamma \cdot f_{RF}(x)$$

where $f(x)$ is the predicted output, $x$ is the input feature vector, $f_{SVM}(x)$ is the estimated output reconstructed by the SVM model, $f_{DT}(x)$is the estimated output reconstructed by the DT model, and $f_{RF}(x)$ is the estimated output reconstructed by the RF model. The values $\alpha$, $\beta$, and $\gamma$ are used to sum up the values of the three models developed.

```
Algorithm 1: Fine-tuning SVM, DT, and RF
# Start creating a dictionary that would
# store the best parameters to use with
# each algorithm
best_parameters = {}
# Tunable Algorithm List
algorithms = [SVM, DT, RF]
For every algorithm,
  # Start the parameters of the current
  # algorithm
   InitializeParameters(algorithm)
  # Variables to store the best metrics
  # and parameters
  best_metrics = InitializeBestMetrics()
  best_params = InitializeBestParameters()
  # For each combination of the parameters
  # in the grid
  For each combination in the
  parameter_grid
    # Define the parameters to control in
    # the current combination
    SetParameters(algorithm, combination)
    # Score it with the current
    # hyperparameters
    metrics = EvaluateModel(algorithm)
    # Save the metrics of the evaluation
    StoreEvaluationMetrics(algorithm,
    combination, metrics)
    # If the current metrics are better
    # than the best metrics found so far
    If IsBetterMetrics(metrics,
    best_metrics)
      # Improve the best metrics and
      # parameters
      best_metrics = metrics
      best_params = combination
    End If
  End For
  # Save the best parameters for the
  # current algorithm
  best_parameters[algorithm] = best_params
End For
# Return the best parameters for all
# algorithms
Return best_parameters
End
```

In ML, hyperparameter optimization is crucial to enhance model efficiency. The process involves creating a 'best_parameters' dictionary to store the optimal hyperparameters for each algorithm, such as SVM, DT, and

RF. The initial values are assigned and a grid of potential combinations is built. Each combination is tested, and performance metrics are logged. If a combination outperforms the previous settings, it replaces the old ones in the dictionary. This systematic approach ensures that optimal hyperparameters are identified and maintained through regular fine-tuning for consistent model performance.

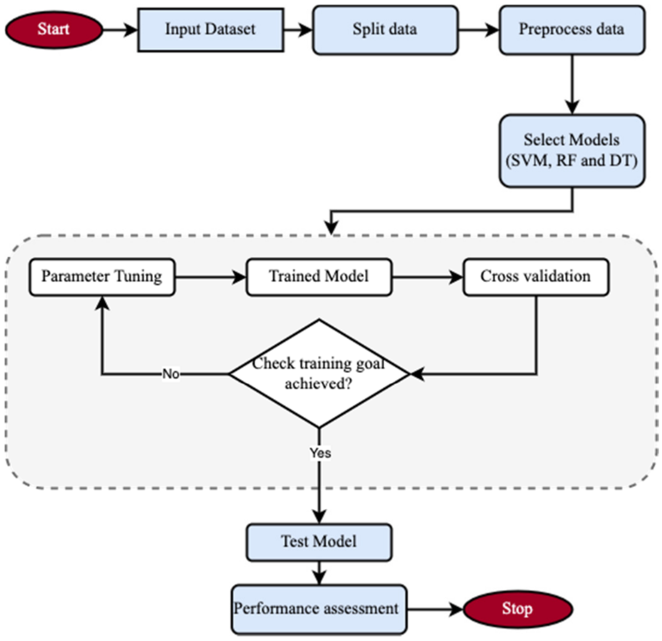


Fig. 2. Proposed architecture.

Figure 2 presents the proposed architecture for applying ML algorithms to sales data prediction. The process begins with obtaining raw data from Salesforce platforms, which are then divided into training and testing datasets. Preprocessing is crucial for handling missing values and inconsistent qualitative data, allowing for feature extraction essential for forecasting. The models are then selected and trained on the processed data to identify patterns related to sales outcomes. Fine-tuning is performed systematically to optimize hyperparameters, such as the regularization parameter, using techniques such as cross-validation to ensure that improvements are consistent across data subsets. The goal is to determine whether the model meets its sales prediction objectives. If the goals are not met, further re-tuning is necessary. If successful, the model proceeds to testing. Finally, the effectiveness of the proposed framework is evaluated using the testing dataset.

## V. RESULTS AND DISCUSSION

This study focuses on predicting sales performance based on advertising budgets for TV and radio promotions. The goal is to classify sales into three categories: low, medium, and high. A Kaggle dataset was used, which includes 4,572 rows of data covering advertising budgets for TV, radio, social media, and influencers [24]. This study specifically selected TV and radio budgets as input features and categorized the sales data into three groups. After preprocessing, 4,546 rows were used for analysis. ML classification algorithms, including RF, DT, and SVM, were employed to build models that predict the sales category based on advertising spending. The models were fine-tuned using a standard scaling pipeline to optimize performance. The process was carried out in Python using Google Colab.

### *A. Performance Metrics*

In the following metrics, TP denotes True Positives, FP denotes False Positives, TN denotes True Negatives, and FN denotes False Negatives.

Accuracy: Used while testing a given model, it is a measure that works together with the assessment of the positive results by also considering the negative results in a given model.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN+TN}$$

Precision: It is a measure that defines the rate of accurately predicted positive cases out of all the cases that were anticipated to be positive.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall (Sensitivity): It is the rate of positive results for the total amount of expected positive samples.

$$\text{Recall} = \frac{TP}{TP+FN}$$

F1 Score: This is a preferred metric for circumstances where the positive and negative classes are not balanced in a dataset. It measures the harmonic balance of precision and recall, so it gives a single value for both of them.

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

### *B. Results*

As shown in Figure 3, the performance of the basic SVM was compared to that of the optimized SVM. Fine-tuning improves the SVM model from 81% accuracy to 97%, and its FP rate from 17.8% to 2.1%, meaning that there were fewer FP. The log loss decreased from 48.1% to 9.3%. Accuracy increased from 80.2% to 99.2% and the F1-score, which integrates precision and recall rates, increased from 81.9% to 97%.

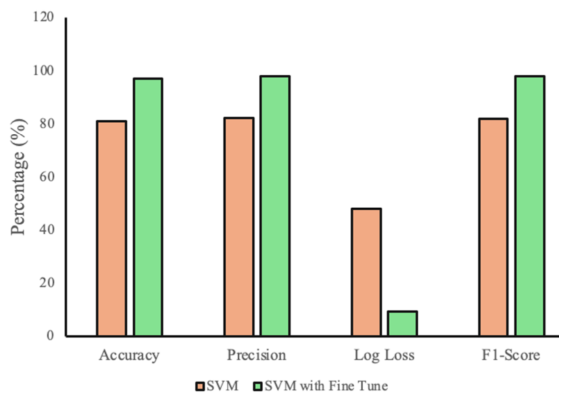


Fig. 3. SVM comparison results.

Figure 4 compares an RF model with its fine-tuned version across four metrics, where the latter showed significant improvements. Accuracy increased from 81.5% to 98%. The log loss decreased from 43.4% to 9.3 % and the F1-score increased from 81.7% to 98.1%, showing that the fine-tuned model had better accuracy for both positive and negative instances.

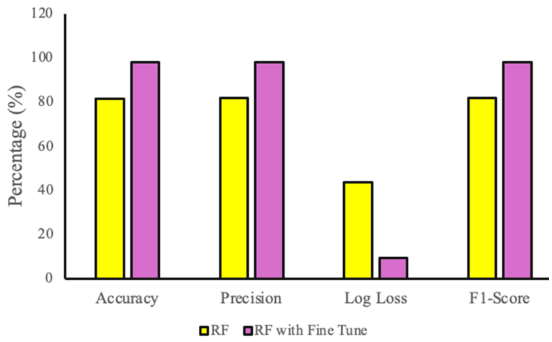


Fig. 4. RF comparison results.

Figure 5 shows the results of a DT along with its fine-tuned version. Fine-tuned DT presents a marked improvement in accuracy, from 79.7% to 97.9%. The F1 score increased from 79.8% to 97.9%, which also shows the improvement in the system's accuracy and its chance to recall the correct object. The log loss decreased from 13.2 to 5.2, indicating that the model had more confidence than before. The fine-tuned model reduced the general number of FP cases, as the precision increased from 82.2% to 97.9%.

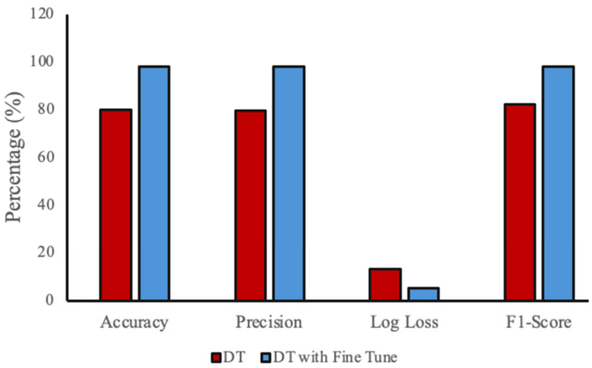


Fig. 5. DT comparative results.

These results show that fine-tuning helps to improve the performance of these ML models. However, this process may cause overfitting or compatibility issues. Nevertheless, fine-tuning enhances the prediction ability of the models, thus it is an effective means of enhancing accuracy and generalization.

## VI. CONCLUSION AND FUTURE WORK

Today, organizations use analytical methods to enhance the impact of strategies and sales in the context of the current environment. Data features from CRMs, such as Salesforce, entail methods of feature analysis on aspects of sales data. This study recommends an approach to making a sales forecast using feature selection, data preprocessing, model selection, and finally hyperparameter tuning. The evaluation revealed that the proposed framework offered higher values of estimated test accuracy than the default methods. The results show that fine-tuning enhances the models' performance and validity and reduces the chances of false positives. Future work should investigate further directions proposed in the framework, such as initiating a comfortable list of the kind of engagement, expanding its scope to the big data phenomenon, and considering ethical issues in ML.

## REFERENCES

[1] S. San, "Optimizing Sales Performance in Creative as a Service (CaaS) Companies: A Machine Learning Approach to Opportunity Time-Series Forecasting," M.S. Thesis, Universidade NOVA de Lisboa, Portugal, 2023.

[2] A. A. K. Saja and R. M. Sawsan, "A Systematic Literature Review on Construction Management Productivity Enhancement by utilizing Business Information Modeling," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13702–13705, Apr. 2024, https://doi.org/10.48084/etasr.7048.

[3] R. Agnihotri, K. Afshar Bakeshloo, and S. Mani, "Social media analytics for business-to-business marketing," *Industrial Marketing Management*, vol. 115, pp. 110–126, Nov. 2023, https://doi.org/10.1016/j.indmarman.2023.09.012.

[4] M. N. Khan and Z. Shao, "Impact of Big Data and Knowledge Management on Customer Interactions and Consumption Patterns: Applied Science Research Perspective," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14125–14133, Jun. 2024, https://doi.org/10.48084/etasr.7203.

[5] A. G. Al-Ali, R. Phaal, and D. Sull, "Deep Learning Framework for Measuring the Digital Strategy of Companies from Earnings Calls." arXiv, Dec. 03, 2020, https://doi.org/10.48550/arXiv.2010.12418.

[6] R. Adebiaye, M. Alshami, and T. Owusu, "Machine Learning Models for Extrapolative Analytics as a Panacea for Business Intelligence Decisions," *International Journal of Engineering Technologies and Management Research*, vol. 10, no. 6, pp. 13–32, Jun. 2023, https://doi.org/10.29121/ijetmr.v10.i6.2023.1333.

[7] F. Wang and J. Aviles, "Enhancing Operational Efficiency: Integrating Machine Learning Predictive Capabilities in Business Intellgence for Informed Decision-Making," *Frontiers in Business, Economics and Management*, vol. 9, no. 1, pp. 282–286, May 2023, https://doi.org/10.54097/fbem.v9i1.8694.

[8] R. Bogue, "The role of machine learning in robotics," *Industrial Robot: the international journal of robotics research and application*, vol. 50, no. 2, pp. 197–202, Feb. 2023, https://doi.org/10.1108/IR-11-2022-0279.

[9] S. Kavitha, "Machine Learning Enabled E-Commerce Analytics," *International Scientific Journal of Engineering and Management*, vol. 2, no. 3, Mar. 2023, https://doi.org/10.55041/ISJEM00214.

[10] M. Gottam, "How Machine Learning Can Be Used To Improve Predictive Analytics," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 12, pp. 2199–2203, Dec. 2022, https://doi.org/10.22214/ijraset.2022.48432.

[11] R. Hirt, N. Kühl, D. Martin, and G. Satzger, "Enabling inter-organizational analytics in business networks through meta machine learning," *Information Technology and Management*, Jun. 2023, https://doi.org/10.1007/s10799-023-00399-7.

[12] B. P. Gadekar and T. Hiwarkar, "From Business Objectives to Analytics and Machine Learning Solutions: A Framework for Conceptual Modeling," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 277–285, Jan. 2023, https://doi.org/10.48175/IJARSCT-7876.

[13] B. P. Gadekar and T. Hiwarkar, "A Conceptual Modeling Framework to Measure the Effectiveness using ML in Business Analytics," *International Journal of Advanced Research in Science, Communication*

*and Technology*, pp. 399–406, Dec. 2022, https://doi.org/10.48175/IJARSCT-7703.

[14] K. Hemachandran, S. Khanra, R. V. Rodriguez, and J. R. Jaramillo, *Machine Learning for Business Analytics: Real-Time Data Analysis for Decision-Making*, 1st ed. New York, NY, USA: Productivity Press, 2022.

[15] U. K. Thakur, "The Role of Machine Learning in Customer Experience:," in *Advances in Business Information Systems and Analytics*, M. S. Hossain, R. C. Ho, and G. Trajkovski, Eds. IGI Global, 2023, pp. 80–89.

[16] P. Kaushik, "Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation," *International Journal for Global Academic & Scientific Research*, vol. 1, no. 3, pp. 1–11, Oct. 2022, https://doi.org/10.55938/ijgasr.v1i3.15.

[17] H. S. Sanil *et al.*, "Role of machine learning in changing social and business eco-system – a qualitative study to explore the factors contributing to competitive advantage during COVID pandemic," *World Journal of Engineering*, vol. 19, no. 2, pp. 238–243, Jan. 2022, https://doi.org/10.1108/WJE-06-2021-0357.

[18] M. Rahmaty, "Machine learning with big data to solve real-world problems," *Journal of Data Analytics*, vol. 2, no. 1, pp. 9–16, Apr. 2023, https://doi.org/10.59615/jda.2.1.9.

[19] A. Verma, J. Kumar, H. M. Gaur, V. Singh, and V. E. Balas, *Advances in Cyber Security and Intelligent Analytics*, 1st ed. Boca Raton, FL, USA: CRC Press, 2022.

[20] B. Hurry and W. Jay, "Predictive Performance Analytics: a Strategic Framework for Enhancing Employee Productivity through an Integrated Business Analytics and Machine Learning Approach." EasyChair Preprint 12189, 2024.

[21] Y. Duan, G. Cao, and J. S. Edwards, "Understanding the impact of business analytics on innovation," *European Journal of Operational Research*, vol. 281, no. 3, pp. 673–686, Mar. 2020, https://doi.org/10.1016/j.ejor.2018.06.021.

[22] C. O. O. Tsahat, N. A. -Ndzeli, and C. G. L. N. Moukengue, "Opportunities for Using Machine Learning and Artificial Intelligence in Business Analytics," *Computer and Information Science*, vol. 17, no. 2, Jun. 2024, https://doi.org/10.5539/cis.v17n2p1.

[23] R. Zhang, "Research on the application of machine learning in business analytics: Cases of Amazon and eBay," *Applied and Computational Engineering*, vol. 51, pp. 59–65, Mar. 2024, https://doi.org/10.54254/2755-2721/51/20241165.

[24] H. Samuel, "Dummy Marketing and Sales Data." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/harrimansaragih/dummy-advertising-and-sales-data.