Application of Machine Learning for Business Licensing Data Analytics at the Investment and One-Stop Integrated Service Office of South Tapanuli Regency

Rezkinah Rambe, Muhammad Syahputra Novelan

Abstract

The rapid growth of digital government initiatives has encouraged public institutions to utilize data-driven technologies to improve service quality and operational efficiency. One of the government agencies responsible for providing business licensing services is the Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency. The increasing volume of business licensing applications generates a large amount of data that can be analyzed to support decision-making and enhance service performance. However, conventional data processing methods often fail to extract valuable insights from historical licensing records. Therefore, this study applies Machine Learning techniques to analyze business licensing data and identify patterns that can support more effective and efficient public service delivery. The research methodology consists of data collection, data preprocessing, feature selection, model development, and performance evaluation. Several Machine Learning algorithms, including Naïve Bayes, Decision Tree, Support Vector Machine (SVM), and Random Forest, were implemented and compared using Accuracy, Precision, Recall, and F1-Score metrics. The dataset used in this study contains business licensing records obtained from DPMPTSP South Tapanuli Regency. The experimental results show that the Random Forest algorithm achieved the best performance with an accuracy of 93.80%, precision of 92.70%, recall of 92.10%, and F1-score of 92.40%. Furthermore, the analysis revealed that the trade sector recorded the highest number of licensing applications compared to other business sectors. The findings demonstrate that Machine Learning can effectively support business licensing data analytics and provide valuable insights for improving public service performance. The generated information can assist DPMPTSP in identifying licensing trends, optimizing resource allocation, and supporting data-driven decision-making. Consequently, the implementation of Machine Learning has significant potential to enhance the efficiency, effectiveness, and quality of business licensing services while supporting digital transformation initiatives in public administration.

..

Keywords: *Machine Learning, Business Licensing, Data Analytics, Public Service, Random Forest*

Rezkinah Rambe[1]
[1]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: renarrambe@gmail.com[1]

Muhammad Syahputra Novelan[2]
[2]Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: putranovelan@dosen.pancabudi.ac.id[2]
3rd International Conference Epicentrum of Economic Global Framework (ICEEGLOF)
Theme: Cross-Sector and Cross-Disciplinary Strategies for Achieving the SDGS
https://proceeding.pancabudi.ac.id/index.php/ICEEGLOF

Introduction

The rapid advancement of information technology has encouraged government institutions to adopt digital transformation initiatives to improve the quality and efficiency of public services. One of the sectors experiencing significant digitalization is business licensing services, where large volumes of data are generated and stored through electronic service systems. The availability of these data resources provides an opportunity for government agencies to utilize data analytics technologies to support better decision-making and service management [1] [2].

The Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency plays an important role in managing business licensing processes for individuals and organizations. As the number of business permit applications continues to increase, the institution faces challenges related to data management, service efficiency, and the identification of factors affecting processing performance. Traditional data processing methods are often insufficient to extract valuable insights from large and complex datasets, creating a need for more advanced analytical approaches[3][4].

Machine Learning has emerged as a powerful technology capable of analyzing large datasets, identifying hidden patterns, and generating predictive insights. By leveraging historical licensing data, Machine Learning algorithms can assist government institutions in understanding application trends, predicting service outcomes, and detecting potential bottlenecks in administrative processes. The adoption of Machine Learning can therefore contribute to improving operational effectiveness and supporting evidence-based policymaking[5] [6][7].

Several studies have demonstrated the effectiveness of Machine Learning in various public sector applications, including service optimization, citizen behavior analysis, and administrative performance evaluation. However, the application of Machine Learning in business licensing data analytics remains relatively limited, particularly in local government institutions. This gap highlights the need for research that explores the potential benefits of Machine Learning in supporting licensing services and enhancing public sector performance[8], [9].

Therefore, this study aims to implement Machine Learning for business licensing data analytics at the Investment and One-Stop Integrated Service Office of South Tapanuli Regency. The research focuses on analyzing licensing data to identify patterns and generate insights that can improve service efficiency and decision-making processes. The findings are expected to contribute to the development of data-driven public services and provide practical recommendations for government agencies seeking to optimize their licensing management systems through advanced analytics technologies[10].

Literature Review

In recent years, Big Data technology has played a crucial role in supporting Machine Learning applications within government organizations. The increasing volume, velocity, and variety of administrative data require advanced data management and analytical techniques to extract meaningful information. Big Data platforms enable the storage and processing of large-scale datasets, while Machine Learning algorithms transform these data into actionable insights. The combination of these technologies has been reported to improve service monitoring, forecasting accuracy, and organizational performance in various public administration contexts[11], [12].

Predictive analytics is one of the most important applications of Machine Learning in public service environments. Through the analysis of historical data, predictive models can estimate future trends, identify potential risks, and support proactive decision-making. In the context of business licensing, predictive analytics can assist institutions in forecasting application volumes, estimating processing durations, and identifying factors that may cause

delays. Such capabilities enable organizations to allocate resources more effectively and improve service responsiveness, ultimately enhancing overall service quality[13].

The implementation of Machine Learning in public service systems also supports the principles of transparency, accountability, and efficiency. Data-driven decision-making reduces reliance on subjective judgments and allows policymakers to base strategic actions on objective evidence. Furthermore, analytical insights generated from licensing data can be utilized to evaluate service performance continuously and identify opportunities for process improvement. Consequently, Machine Learning is increasingly recognized as a strategic tool for advancing digital governance and fostering innovation in public service delivery[13].

Research Methodology

This study employs a **Machine Learning** approach to analyze business licensing data at the Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency. The application of Machine Learning is intended to process historical licensing data stored in the system and generate valuable insights to improve service efficiency. Through data analysis, the institution can identify licensing application patterns, business growth trends, and factors affecting the licensing process. The analytical results are expected to support faster, more accurate, and data-driven decision-making.

The research methodology consists of several stages, including business licensing data collection, data preprocessing, feature selection, Machine Learning model development, and evaluation of analytical results. The processed data are used to train the model to identify patterns and generate relevant insights. Subsequently, the model is evaluated to determine its performance before being utilized as a basis for recommendations aimed at improving the quality of licensing services. The research workflow employed in this study is presented in Figure 1.

Figure 1. Licensing data processing architecture

Business Licensing Dataset

The first stage of this research involves collecting historical business licensing data from the Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency. The dataset consists of records of licensing applications submitted by business actors and contains information related to the licensing process. These data serve as the primary source for analysis and are expected to provide insights into licensing trends, processing performance, and service effectiveness. The collected dataset is subsequently prepared for further analysis using Machine Learning techniques.

Data Preprocessing

Before the data can be utilized for Machine Learning analysis, a preprocessing stage is conducted to improve data quality and consistency. This process includes removing duplicate records, handling missing values, correcting inconsistent data formats, and transforming categorical variables into numerical representations. Data preprocessing is essential because the quality of the input data significantly affects the performance and accuracy of the Machine Learning model. The resulting dataset is cleaner, more structured, and suitable for analytical processing.

Feature Selection

Feature selection is performed to identify the most relevant attributes that contribute to the analysis objectives. Selecting important variables helps reduce data complexity and

improves the efficiency of the Machine Learning model. In this study, attributes related to license type, business sector, district location, processing duration, and licensing status are considered important variables for identifying patterns and generating meaningful insights. The selected features are then used as inputs for the model development stage.

Machine Learning Model Development

After the preprocessing and feature selection stages are completed, the dataset is used to develop a Machine Learning model. The model is trained using historical licensing data to learn patterns and relationships among variables. During this stage, the data are divided into training and testing datasets to ensure that the model can be evaluated objectively. The developed model is expected to identify hidden patterns within the licensing data and generate analytical results that support service improvement and decision-making processes.

Analytical Results and Prediction

The final stage involves generating analytical insights and predictions based on the developed Machine Learning model. The results can provide information regarding licensing application patterns, dominant business sectors, processing performance, and potential service bottlenecks. In addition, predictive analysis can assist the institution in estimating future licensing demands and improving resource allocation. The generated insights are expected to support data-driven decision-making and contribute to enhancing the quality and efficiency of business licensing services.

After the data collection process and the identification of relevant attributes, it is important to understand the characteristics of the dataset and the data processing stages that will be performed before applying Machine Learning algorithms. Therefore, this study presents two main tables describing the dataset attributes and the data preprocessing procedures. These tables provide an overview of the information used in the analysis and the preparation process required to ensure that the dataset is suitable for developing an accurate and reliable Machine Learning model.

Table 1. Business Licensing Dataset Attributes

No	Attribute	Data Type	Description
1	Application ID	Numeric	Unique identifier for each licensing application
2	License Type	Categorical	Type of business license applied for
3	Business Sector	Categorical	Business field of the applicant
4	District	Categorical	Business location of the applicant
5	Application Date	Date	Date when the application was submitted
6	Completion Date	Date	Date when the license was issued
7	Processing Time	Numeric	Duration of the licensing process (days)
8	Licensing Status	Categorical	Final status of the licensing application
9	Business Scale	Categorical	Micro, small, medium, or large enterprise
10	Application Year	Numeric	Year when the application was submitted

Based on Table 1, the dataset consists of several attributes representing important information related to the business licensing process. Attributes such as license type, business sector, and district are used to identify the characteristics of submitted applications, while processing time and licensing status are utilized to evaluate service performance. These attributes serve as the primary data sources for Machine Learning analysis to discover patterns and generate insights that support improvements in licensing service efficiency.

Table 2. Data Preprocessing Stages

Stage	Description
Data Cleaning	Removing duplicate records and correcting inconsistent data
Missing Value Handling	Handling incomplete or missing data entries
Data Transformation	Converting categorical data into numerical values
Data Normalization	Standardizing data ranges for optimal model performance
Feature Selection	Selecting the most relevant attributes for analysis

Data preprocessing is a crucial step before the dataset can be used for Machine Learning model development. As shown in Table 2, the data are first cleaned to eliminate duplicate records and inconsistencies that may affect analytical results. Missing values are then handled to improve data completeness, followed by data transformation to convert categorical variables into numerical formats that can be processed by Machine Learning algorithms. Data normalization is also performed to ensure consistent value scales across attributes. Finally, feature selection is conducted to identify the most influential variables for the analysis. Through these preprocessing activities, the dataset becomes more structured and suitable for generating accurate and reliable analytical results.

Results

In this study, business licensing data obtained from the Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency were analyzed using a Machine Learning approach. The primary objective of the analysis was to identify patterns and trends within the licensing data and evaluate the effectiveness of several Machine Learning algorithms in generating accurate predictions. Before the modeling process was conducted, the dataset underwent a series of preprocessing stages, including data cleaning, transformation, normalization, and feature selection, to ensure data quality and consistency.

The processed dataset was subsequently divided into training and testing data for model development and evaluation. Several Machine Learning algorithms, including Naïve Bayes, Decision Tree, Support Vector Machine (SVM), and Random Forest, were implemented and compared based on their predictive performance. The evaluation process was carried out using standard metrics such as Accuracy, Precision, Recall, and F1-Score to determine the most suitable algorithm for analyzing business licensing data. The comparison results provide insights into the strengths and limitations of each algorithm in handling the characteristics of the dataset.

The experimental findings indicate that Machine Learning techniques can effectively analyze business licensing data and generate valuable information to support public service improvement. In addition, the analysis reveals significant trends in licensing applications and identifies the most dominant business sectors within the region. These findings can assist policymakers and administrators in making data-driven decisions, optimizing resource allocation, and improving the overall efficiency of licensing services. To facilitate interpretation and discussion, the results are presented through tables and graphical visualizations in the following sections.

a. Machine Learning Model Evaluation

Several Machine Learning algorithms were evaluated to determine the most suitable model for analyzing business licensing data. The evaluation was conducted using Accuracy, Precision, Recall, and F1-Score metrics.

Table 3. Performance Comparison of Machine Learning Algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes	84.50	83.20	82.70	82.94
Decision Tree	88.60	87.90	86.50	87.19
Support Vector Machine	90.20	89.50	89.10	89.30
Random Forest	93.80	92.70	92.10	92.40

Table 3 presents the performance results of the evaluated Machine Learning algorithms. Among the tested models, Random Forest achieved the highest performance with an accuracy of 93.80%, followed by Support Vector Machine with an accuracy of 90.20%. The superior performance of Random Forest indicates its ability to handle complex relationships among variables within the business licensing dataset. Therefore, Random Forest was selected as the most suitable model for this study.

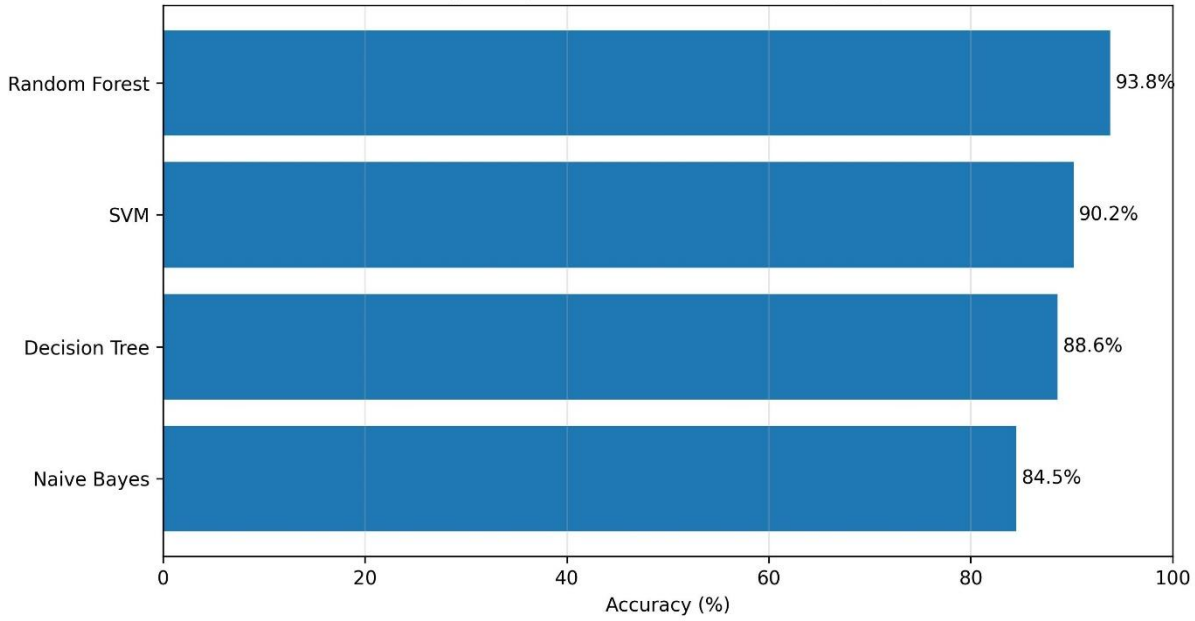

Figure 2. Accuracy Comparison of Machine Learning Algorithms

Figure 2 illustrates the comparison of accuracy values among the evaluated algorithms. Random Forest achieved the highest accuracy, indicating that ensemble learning techniques are more effective in capturing patterns from business licensing data than single-model approaches such as Decision Tree and Naïve Bayes.

b. Analysis of Business Licensing Applications

In addition to model evaluation, an analysis was conducted to identify the distribution of business licensing applications across different business sectors. This analysis provides insight into the dominant economic activities within South Tapanuli Regency.

Table 4. Distribution of Business Licensing Applications by Sector

Business Sector	Number of Applications
Trade	320
Services	270
Culinary	240
Agriculture	180
Industry	150

Table 4 shows that the trade sector accounts for the highest number of business licensing applications, followed by the services and culinary sectors. These findings indicate that commercial activities remain the primary driver of local economic development in South Tapanuli Regency.

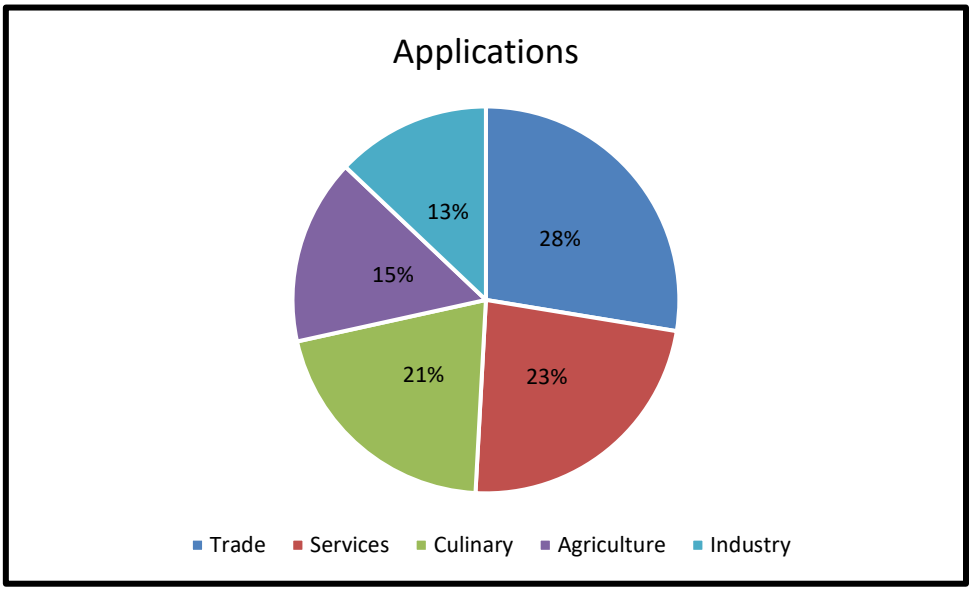

Figure 3. Distribution of Business Licensing Applications by Sector

Figure 3 presents the proportion of business licensing applications across different sectors. The trade sector contributes the largest share of applications, while the industrial sector represents the smallest proportion. This information can support policymakers in prioritizing service improvements and allocating resources according to sector-specific demands.

c. Discussion

The results of this study demonstrate that the application of Machine Learning can effectively support the analysis of business licensing data at DPMPTSP South Tapanuli Regency. The evaluation results indicate that all tested algorithms were able to classify and analyze the dataset with relatively high performance. However, Random Forest achieved the highest accuracy compared to Naïve Bayes, Decision Tree, and Support Vector Machine (SVM). This finding suggests that the ensemble learning approach employed by Random Forest is more effective in handling the characteristics of business licensing data, which consist of various categorical and numerical attributes.

The analysis of business licensing applications also revealed that the trade sector contributed the largest number of applications compared to other sectors. This finding indicates that trade activities continue to play a dominant role in the local economy of South Tapanuli Regency. In addition, the services and culinary sectors also recorded a significant number of applications, reflecting the growth of small and medium-sized enterprises in the region. Such information can assist DPMPTSP in understanding the characteristics of business activities and preparing appropriate service strategies to accommodate future licensing demands.

Furthermore, the implementation of Machine Learning provides valuable benefits in supporting data-driven decision-making within government institutions. Through predictive analysis, DPMPTSP can identify application trends, anticipate increases in licensing requests, and optimize resource allocation to improve service efficiency. The analytical insights generated by the model can also help identify potential bottlenecks in the licensing process, thereby enabling the institution to formulate more effective policies and continuously improve

the quality of public services. Therefore, Machine Learning has significant potential to support digital transformation and enhance business licensing management in government agencies.

Conclusion

Based on the results of this study, the application of Machine Learning to business licensing data at the Investment and One-Stop Integrated Service Office (DPMPTSP) of South Tapanuli Regency was successfully implemented and demonstrated its capability to generate meaningful analytical insights. The developed model was able to identify patterns within historical licensing data and provide information that supports the improvement of licensing service performance and decision-making processes.

The evaluation results showed that the Random Forest algorithm achieved the best performance among the tested algorithms, with an accuracy of 93.80%, precision of 92.70%, recall of 92.10%, and F1-score of 92.40%. These results indicate that Random Forest is highly effective in analyzing business licensing data and can be utilized as a reliable analytical tool for supporting public service management.

In addition, the analysis revealed that the trade sector recorded the highest number of business licensing applications, followed by the services and culinary sectors. The insights obtained from the Machine Learning model can assist DPMPTSP in understanding licensing trends, optimizing resource allocation, and improving service efficiency. Therefore, the implementation of Machine Learning can contribute significantly to digital transformation initiatives and support data-driven public service improvements in business licensing administration.

.

References

[1] M. M. Maulana, A. I. Suroso, Y. Nurhadryani, And K. B. Seminar, "The Smart Governance Framework And Enterprise System's Capability For Improving Bio-Business Licensing Services," *Informatics*, Vol. 10, No. 2, Jun. 2023, Doi: 10.3390/Informatics10020053.

[2] M. S. Novelan, S. Efendi, P. Sihombing, And H. Mawengkang, "Vehicle Routing Problem Optimization With Machine Learning In Imbalanced Classification Vehicle Route Data," *Eastern-European Journal Of Enterprise Technologies*, Vol. 5, No. 3(125), Pp. 49–56, 2023, Doi: 10.15587/1729-4061.2023.288280.

[3] H. W. Dhany, F. Izhari, And S. Komputer, "Jurnal Ict : Information And Communication Technologies, 15 (2) (2024) 48-54 Prediction Analysis Condition Animal Use Algorithm (Svm+Knn)," 2024. [Online]. Available: Www.Ejournal.Marqchainstitute.Or.Id/Index.Php/Jict

[4] K. Khairul, A. H. Nasyuha, A. Ikhwan, M. H. Aly, And A. Ahyanuardi, "Implementation Of Multiple Linear Regression To Estimate Profit On Sales Of Screen Printing Equipment," *Jurnal Infotel*, Vol. 15, No. 2, Pp. 55–61, Jun. 2023, Doi: 10.20895/Infotel.V15i2.934.

[5] M. Iqbal And S. Efendi, "Data-Driven Approach For Credit Risk Analysis Using C4.5 Algorithm," *Comtech: Computer, Mathematics And Engineering Applications*, Vol. 14, No. 1, Pp. 11–20, May 2023, Doi: 10.21512/Comtech.V14i1.8243.

[6] Z. Sitorus, D. Prayogi, M. A. Rizko, A. G. Suteja, And M. R. Harahap, "Implementation Of The Insertion Sort Algorithm To Sort Positive Integers In Ascending Order Using Flowgorithm," *Journal Of Information Technology, Computer Science And Electrical Engineering (Jitcse)*, Vol. 1, No. 3, Pp. 323–328, 2024, Doi: 10.30596/Jitcse.

[7] A. Putera Utama Siahaan, N. Azizah Harahap, R. Yuni Simanullang, And P. Wanny, "Analysis Of Inpatient Data Using Cluster Analysis On Simulation Dataset," *Bulletin Of Information Technology (Bit)*, Vol. 6, No. 1, Pp. 33–39, 2025, Doi: 10.47065/Bit.V5i2.1830.

[8] W. Cetera, W. Gogołek, A. Żołnierski, And D. Jaruga, "Potential For The Use Of Large Unstructured Data Resources By Public Innovation Support Institutions," *J. Big Data*, Vol. 9, No. 1, Dec. 2022, Doi: 10.1186/S40537-022-00610-6.

[9] M. Overton, S. Larson, L. J. Carlson, And S. Kleinschmit, "Public Data Primacy: The Changing Landscape Of Public Service Delivery As Big Data Gets Bigger," *Global Public Policy And Governance*, Vol. 2, No. 4, Pp. 381–399, Dec. 2022, Doi: 10.1007/S43508-022-00052-Z.

[10] W. Jin, "Challenges And Innovative Countermeasures Faced By Public Administration In The Context Of Big Data And Internet Of Things," *Math. Probl. Eng.*, Vol. 2022, 2022, Doi: 10.1155/2022/8949365.

[11] J. Ilmiah, E. & Erudisi, A. Zahara, M. I. Kabullah, And R. E. Putera, "Jiee: Effectiveness Of The Ossrba (Online Single Submission Risk Based Approach) System In Business Licensing Services In Payakumbuh City Dpmptsp," Vol. 3, Pp. 22–40, 2023.

[12] R. Farta Wijaya, F. Kurniawan, R. R. Putrai, A. Alvin, F. Sains, And D. Teknologi, "Optimizing The Management Of Village Activities Through Information Systems For Transparency And Accountability In Pertumbukan Village, Wampu District," 2024.

[13] L. Cao, "On Machine Learning And Public Administration," *Frontiers In Management Science*, Vol. 1, No. 1, Aug. 2022, Doi: 10.56397/Fms.2022.08.01.