

Using Machine Learning Techniques to Enhance the Efficiency of Management Information Systems

Mohanad Ali Hussein

Al-Furat Al-Awsat Technical University, Iraq. E-mail: muhannad.hussain.cku@atu.edu.iq

Abstract - Organizations have become increasingly large and business operations more complex, resulting in a number of challenges for traditional Management Information Systems (MIS). For effective managerial decision making, Intelligent Systems that can handle huge amounts of data and see patterns, trends and insights in the data are crucial for modern organizations. In this regard the Machine Learning techniques have been found to be very useful in improving the efficiency and analysis capability of Management Information Systems.

The aim of this research is to see how Machine Learning can be used to enhance the performance of MIS by predicting customer churn. The study seeks to build predictive models to detect customers at risk of attrition from an organization. An accurate churn prediction model enables businesses to act proactively in order to prevent customers from churning, to prevent wasted monetary resources, and to optimize business operations.

A number of Machine Learning algorithms implemented in the study were Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). The experimental steps involved in the process included data preprocessing, a feature selection technique, feature scaling and dataset balancing with the Synthetic Minority Oversampling Technique (SMOTE). Standard classification metrics (accuracy, precision, recall, and F1-score) were used to test the models.

After optimizing and pre-processing the best model with respect to overall predictive performance according to the results found was Logistic Regression. The optimized model achieved a precision of 98%, and was extremely accurate in both recall and f1-score, indicating very good performance on identifying the churn customers.

The results also emphasized the need to solve the class imbalance issue and use the right preprocessing methods to enhance the performance of Machine Learning in business datasets.

Keywords: Machine Learning, Management Information Systems, Customer Churn Prediction, Predictive Analytics, Logistic Regression, Artificial Intelligence, SMOTE, Data Preprocessing, Decision Support Systems.

I. INTRODUCTION

Over the past few years, organisations have seen the amount of data they generate on a day to day basis with their business, customers and financial transactions, as well as digital systems of communication, shooting up [7]. With the constant increase in data, organizations are facing new challenges in handling information effectively and converting raw data into useful information that can be utilized by the management for decision making. In the light of this, Management Information Systems (MIS) are vital tools that are indispensable in the contemporary organizations which helps managers to gather, process, store and analyze information needed for planning, controlling and strategic decision making processes [8].

Management Information Systems are important to enhance the efficiency of an organization by delivering relevant, accurate and timely information to decision makers [9]. The main objective of traditional MIS is to store and organize data, produce reports and facilitate routine management activities. In complex business environments and large amounts of data in the organization, it is difficult to perform advanced predictive analysis and detect patterns in a large data set within a traditional system [10]. Organizations need more intelligent and adaptive systems that can help support proactive decisions instead of reactive decisions.

Using ML techniques in Management Info Systems offers a whole lot of opportunities for the organization to lower down their operations, boost the quality of their choices and automate their analytical workflows optimally. Smart MIS systems can be used to study customer actions, anticipate future business trends, flag off abnormalities, detect risks, and provide strategic planning support activities. This integration enables them to shift from descriptive reporting to predictive and prescriptive decision support systems.

Customer churn prediction is one of the most useful applications of Machine Learning used in Management

Information Systems. Customer churn is a phenomenon in which customers stop using a company's products and services and/or end their relationship with the company. As retaining customers becomes more cost effective than winning them over, churn prediction has emerged as a key concern of organisations. Not identifying potential churn customers can result in the loss of revenue, a drop in customer satisfaction and a drop in competitiveness of the organization.

Although the technologies of Machine Learning are increasingly being implemented, there are still many organizations that struggle to use intelligent predictive systems in their Management Information Systems. The difficulties faced are data quality, imbalances in the data, algorithm selection, and the requirement for appropriate assessment methods. Hence, devising effective predictive models that can accurately classify churn customers is still an important research field in the domain of Management Information Systems and Machine Learning.

II. PROBLEM STATEMENT

The use of Management Information Systems (MIS) is now a common feature of modern organizations, providing them with assistance in their operations, managerial tasks, and strategic decisions. These systems are used to create and manipulate terabytes of organizational and customer data on a daily basis. However, traditional information management systems tend to be siloed and may be limited to reporting basic data and transactions, without providing the predictive and analytical features needed in today's fast-paced business environment.

The inability to optimally utilize large amounts of data for proactive decision-making is one of the biggest challenges facing businesses today. Traditional data analysis methods fail to identify patterns, predict future trends, or pinpoint risks to businesses, such as customer loss. As a result, organizations may experience decreased operational efficiency, poor customer retention, management delays, and financial losses.

The potential for customer loss is a critical issue in the business world, as retaining existing customers is far less expensive than acquiring new ones. Poor identification of customers who are at risk of leaving can have a negative effect on the profitability of the organization, its customer relationship management (CRM), and its competitive advantage. There are many organizations, however, that are still facing challenges in implementing intelligent predictive systems that will enable them to make effective moves to prevent customer churn, despite having access to vast amounts of customer information.

Also, there are various technical and analytical challenges when combining ML with MIS, such as data quality problems, class imbalance, feature selection complexity, and selecting the appropriate predictive algorithms. Organizations often don't have effective machine learning integration strategies that align with MIS features to enhance decision-making and operational efficiency.

Thus, this research focuses on the issue of improving the efficiency of the Management Information Systems by the integration of Machine Learning techniques in the prediction of customer churn. The aim of the study is to analyze the predictive accuracy and compare different Machine Learning algorithms and study the influence of pre-processing and optimization techniques and develop an intelligent and efficient decision support system for modern organizations.

III. RESEARCH OBJECTIVES

This research primarily aims to determine the effectiveness of machine learning techniques in improving the efficiency and analytical capabilities of corporate management information systems (MIS) by predicting customer churn. The study seeks to explore the potential of using intelligent predictive models to support organizational decision-making and, consequently, enhance customer retention strategies in the current business environment.

The research targets of this research will be as follows:

1. To explore how implementation of Machine Learning techniques can help improve the performance and efficiency of Management Information Systems.
2. To build predictive models that are effective in profiling customer who are likely to churn, from customer related business information.
3. To implement and evaluate several machine learning algorithms for customer churn prediction problems such as Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbour (KNN) and Support Vector Machine (SVM).
4. To compare different data preprocessing methods (feature scaling, class balancing using Synthetic Minority Oversampling Technique (SMOTE)) on their influences on predictive performance.
5. Find the most significant business related factors that impact on customer churn.
6. To validate the predictive analytics model for its effectiveness in managerial decision-making and customer market strategy.
7. To highlight the application of the Machine Learning technologies in the Management Information Systems field in order to make them more efficient and useful in the context of Business Intelligence.

IV. RELATED WORK

The business community is in a constant struggle to protect their customers, as the issue of customer churn is a widely studied topic in the field of machine learning because loss of customers can have a great impact on profit in livelihood-related sectors, primarily telecommunications. A number of researchers have devised their own predictive frameworks with various classification and ensemble learning approaches.

Study [1] had suggested a multi-stage churn prediction model consisting of data preprocessing, feature analysis and feature selection via gravitational search algorithms. Different classification algorithms were employed in the study including Logistic Regression, Naive Bayes, Decision Trees, Random Forest and Support Vector Machines. Furthermore, ensemble Learning models like: AdaBoost and XGBoost were employed to boost the accuracy of prediction. K-fold cross validation method was used for the study and models were analyzed using the AUC and confusion matrix. AdaBoost and XGBoost were the best performing models with accuracy of 81.71% and 80.8% respectively.

The study in [2] has been concentrated on among Customer churn analysis in Telecom industry using classification approach like Random Forest, K-Nearest Neighbors, and Decision Trees. One aspect of the study was to highlight the importance of retaining a customer as opposed to acquiring one. The performance of the proposed system was found to be very high where Random Forest obtained result of around 99% accuracy, precision, and recall. The study emphasized the necessity to carry out the study of

consumption pattern to help develop retaining strategies and limit revenue losses.

In Study [3] the authors examined the use of multiple machine learning techniques with customer data obtained over a nine month period to make an early churn prediction. The evaluated models consisted of Random Forest, Logistic Regression, K-Nearest Neighbor and the Stochastic Gradient Boosting. The outcomes revealed that Logistic Regression made excellent accuracy of 82.9% and Random Forest and Gradient Boosting performed at similar levels. It was determined that depending on the characteristics of the data set multiple algorithms can be competitive.

Research [4] suggested a mixed method of the classification and clustering techniques in predicting churn in the telecom segment. Several different algorithms such as Random Forest, XGBoost, KNN, SVM, and Decision Trees were followed and compared on the study. Results showed that the accuracy, precision, recall and F1 score of the ensemble techniques as XGBoost and random forest were better than the conventional classifiers.

Study [5] looked at application of machine learning models for churn prediction with several sophisticated models such as CatBoost and Light Gradient Boosting Machines. For this work, 16 different models were experimented with, both with a train-test split and cross-validation. LightGBM had the best performance when validated using cross validation with accuracy of 96.39%, while CatBoost gained the best performance with the accuracy of 97.54%. Also, the feasibility of applications in real-time deployment in cloud-based environments, along with the importance of large datasets was highlighted by the study.

Table 1: Comparative Analysis of Related Studies and Proposed Work

Study	Methods Used	Best Model	Accuracy	Key Contribution	Comparison with Our Work
[1]	Logistic Regression, SVM, RF, AdaBoost, XGBoost	AdaBoost / XGBoost	~81%	Feature selection + ensemble learning + AUC analysis	Higher complexity models than our work; our model focuses on interpretability
[2]	RF, KNN, Decision Tree	Random Forest	~99%	High-performance churn classification	Very high accuracy but possible overfitting; our work is more balanced and realistic
[3]	LR, RF, KNN, Gradient Boosting	Logistic Regression	82.9%	Comparison of classical ML models	Similar approach to our methodology but without SMOTE balancing
[4]	RF, XGBoost, SVM, KNN	XGBoost / RF	High comparative performance	Hybrid classification + clustering	Focus on ensemble learning; our study emphasizes MIS integration and preprocessing
[5]	16 ML models (CatBoost, LightGBM, etc.)	CatBoost	97.54%	Deep benchmarking + cross-validation	Much more complex models; our work prioritizes simplicity and MIS applicability

Our Study	LR, DT, RF, KNN, SVM + SMOTE	Logistic Regression	98%	MIS-focused churn prediction with balanced dataset	Focus on interpretability, preprocessing, and business applicability
------------------	------------------------------	---------------------	-----	--	--

Discussion of Literature Gap

The research papers reviewed herein illustrate the extent to which machine learning methods have been used to predict customer churn in a variety of contexts, some successful and some failing, throughout the years. Most previous work emphasized themselves on enhancing prediction accuracy by adopting complex ensemble techniques like XGBoost, CatBoost or LightGBM. These models can be very accurate, although generally they focus on computational complexity, predictive performance, rather than being interpretable and integrated into Management Information Systems.

One factor that cannot be neglected is that some studies mention very high accuracy values, often over 95%, which could suggest over-fitting or not being generalizable to other data sets. In addition, some of the studies used to solve the class imbalance issue are inadequate to solve class imbalance problem, and it is a critical issue in churn prediction.

The current research rather concentrates on introducing machine learning techniques in the context of MsIS, with preserving balance between model performance and interpretability. It focuses on the application of Logistic Regression model as a basic model because its application is transparent and suitable for decision support applications, as opposed to very complicated ensemble methods. Furthermore, the study incorporates data processing methods and SMOTE to overcome class imbalance and boost the reliability of the model.

The gap between existing and needed research is therefore filled by this research, which can provide a balanced but still practical approach towards predicting, which can be directly applied in the environment of MIS in real business applications.

V. METHODOLOGY

The study is of a quantitative nature and experiment type, with multiple Machine Learning models trained and tested on structured customer data to classify the most suitable predictive model to classify the customer churn.

Dataset Description

This study utilized data that has been collected from customers for the sole purpose of studying customer churn. It's a "normal" structured business data, typical of a MSIS data environment.

The dataset includes the following attributes [6]:

- **Age:** Represents the age of the customer.
- **Total_Purchase:** Total monetary value of purchases made by the customer.
- **Account_Manager:** Indicates whether the customer has an assigned account manager.
- **Years:** Number of years the customer has been associated with the company.
- **Num_Sites:** Number of services or sites used by the customer.
- **Churn:** Target variable indicating whether the customer has left the company (1) or remained active (0).

The data at first was also including irrelevant information like the names of the customers, the names of the companies located in their respective places, and dates they were on boarded. These are attributes that have been eliminated during prep-processing, as they will not help to build predictions, and can lead to noise within the data.

Data Preprocessing

Creating a new dataset is critical and may have a major impact on the result of the models and their perceived accuracy and trustworthiness. For this study, the following are some of the preprocessing techniques used:

1. Data Cleaning

Any missing data and inconsistencies were identified in the data set. The analysis showed that there were no missing values in the data, which ensured its completeness and reliability.

2. Feature Selection

Irrelevant features such as Names, Company and Location were neglected and Onboard_Date was dropped to improve the efficiency of the model and computational complexity.

3. Feature Scaling

In the dataset there are numerical features which have different scales; therefore, feature scaling was applied to normalize the data. This step is especially crucial for algorithms based on distance and related algorithms based on gradients.

4. Handling Class Imbalance

There was a high level of class imbalance in the dataset with respect to churn and non-churn. In order to overcome this problem, a synthetic minority oversampling technique (SMOTE) was used to create synthetic data points for minority class. This supported better ability of the model to correctly identify churn cases.

Machine Learning Models

To this end, several ML algorithms have been implemented and tested in this study. The models chosen are:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)

These algorithms were chosen due to their widespread use in classification problems and their applicability in business analytics and Management Information Systems.

Experimental Setup

The data set was split in 80:20 ratio, meaning that 0.8 or 80% of data were placed in the training set while 0.2 or 20% were put in the test set. The model were trained using the data in the training set, and the data in the testing set were used to test the models for their performance on unseen data.

All data presented have been the same to allow for proper comparison among the models. All the experiments were carried out under the same experimental conditions to ensure consistency and reliability of results.

Evaluation Metrics

Model performance was evaluated by employing different classification metrics:

- **Accuracy:** How accurate overall the predictions are.
- **Precision:** How many correctly predicted positive cases to how many cases predicted as positive.
- **Recall:** Percentage of actual cases (churn customers) the model successfully catches.
- **F1-Score:** Offers a balance between the precision and recall.

This method is designed to leverage the technologies of the Machine Learning (ML) area of AI as it relates to managing management information systems (MIS) and augment its predictive analytics. The data preprocessing, along with multiple classification approaches and thorough

evaluation metrics, guarantees a robust and dependable experimental design to evaluate customer churn conduct. The data preprocessing, along with several classification methods and a thorough evaluation metrics, ensure a powerful and reliable experimental design which can be used for studying customer churn behavior.

VI. RESULTS AND DISCUSSION

In this section, the results of various experiments conducted using different Machine Learning (ML) algorithms to predict customer churn are discussed in the context of a Management Information System (MIS). This experiment aims is to test the effectiveness of predictive analytics in the loyalty program and on how successful it is in improving the decision making process of managers.

Experimental Setup

For determining the most effective predictive model for integration use within Management Information Systems framework, several Machine Learning algorithms were implemented, and compared. Evaluated models comprise of:

- Logistic Regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)

The data set was split into training and testing subsets, 80:20 can be evaluated with some common classification metrics: Accuracy, Precision, Recall, F1-Score. The above metrics have been chosen for the sound reasons that these metrics offer a complete evaluation of predictive performance, notably in the case of imbalanced classification problems.

Initial Model Comparison

The initial experimental results demonstrated varying levels of predictive performance among the tested algorithms.

Table 2 summarizes the performance metrics before applying advanced optimization techniques.

Table 2

Model	Accuracy	F1-Score
Logistic Regression	90%	0.90
Decision Tree	85%	0.81
Random Forest	86.67%	0.80
KNN	80.55%	0.80
SVM	80%	0.80

Figure 1 shows the performance metrics before applying advanced optimization techniques.

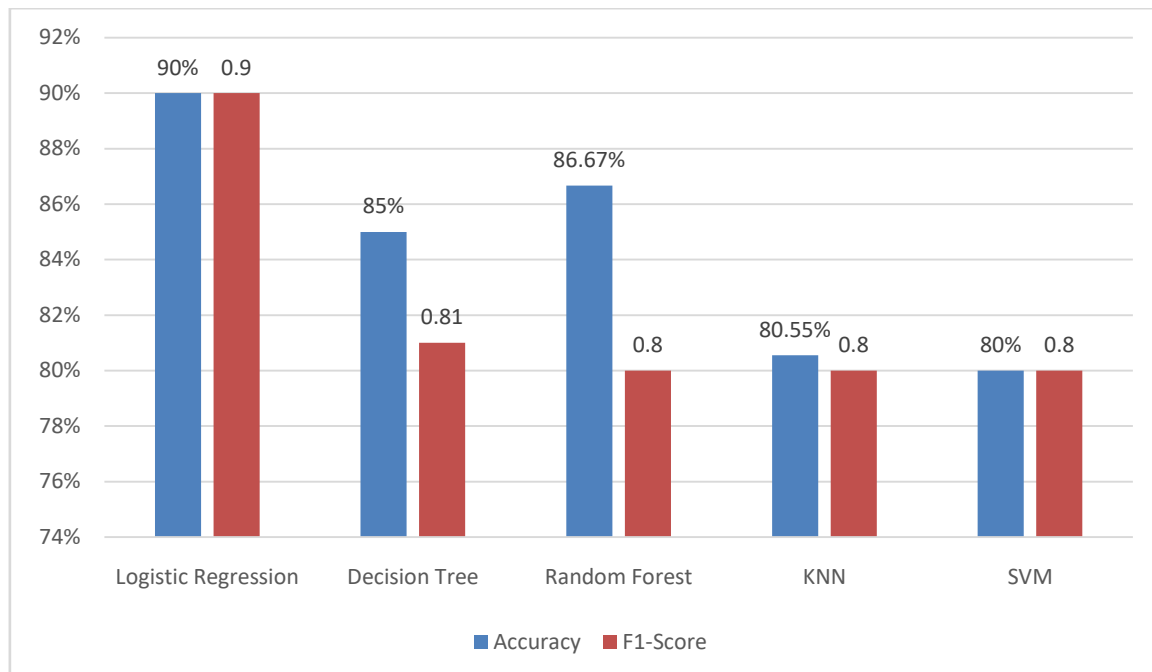


Figure 1: performance metrics before applying advanced optimization techniques

From the results, Logistic Regression was able to achieve the highest overall performance in terms of classification accuracy and churn detection capability. Support Vector Machine has good recall but its accuracy was very low and its classification behavior was unstable, so they were not suitable for practice. Likewise, KNN couldn't detect any cases with good accuracy for class Churn, this indicates its performance was not satisfactory in minority class prediction.

The recall value of Random Forest was still lower than that of Logistic Regression and was able to provide stable performance in acceptable accuracy. Although Decision Tree achieved moderate performance, it did not have high generalization ability and was less consistent in predictions.

Model Optimization and Performance Improvement

To enhance the prediction effect, a few optimisation techniques have been applied such as feature scaling, class balancing by SMOTE and parameters fine-tuning for LR. These enhancements caused the model to have a much better ability to identify churn customers. Optimized, the Logistic Regression model obtained the following performance results (Table 3).

Table 3: Performance Improvement

Metric	Value
Accuracy	98%
F1-Score	98%

The optimized model had significant gain in accuracy and f1score performance than the initial experiments. Which means that the model turned out to be more effective in differentiating customers who are more likely to defect from the organization.

The enhancement of this metric is of significant value in real world Management Information System applications – it is directly related with financial sustenance and Customer Relationship Management. In real world business scenarios, the identification of potential ‘churn’ customers can make a real impact in the saving of revenue, customer satisfaction, and resource allocation.

Moreover, model optimization achieved a balance in the classification performance of each class, whether it's churn or non-churn. This balance helps to reduce the predictive bias and run well automated decision support.

Feature Importance Analysis

The number of sites used by customers was considered the most important feature in predicting customer churn in feature importance analysis. Importantly, customer tenure and the amount of purchases positively affected customer action, too.

Table 4 presents the relative importance of each feature.

Table 4

Feature	Importance Score
Num_Sites	46.56%
Years	21.84%
Total_Purchase	16.13%
Age	12.54%
Account_Manager	2.92%

These results are useful stewardship decisions for a customer attrition behavioral issue. More users having more services or sites could have more varied usage patterns which can impact exit probability of customers from the organization. Similarly, in terms of customer outcomes, customer tenure is seen to be an important factor with purchase activity also being important.

Discussion

The results of the experiment validate that using Machine Learning techniques throughout Management Information Systems applications can significantly enhance predictive analytics and decision-making within organizations. The designed system showed excellent potential to provide comprehensive business intelligence based on the operational data of customers that enables effective customer retention strategies.

Among the most critical findings of this study, it has been found that simpler algorithms like Logistic Regression can yield better performance with the use of optimal Pre-processing and optimization machineries when provided to the algorithms. The implication from this finding is that data quality, feature engineering and class balancing are key to getting a robust prediction.

Moreover, this study reveals the significance of recall in the critical evaluation measures of churn prediction systems. While the accuracy is important, the right to accurately classify risk associated customers is more strategically relevant for companies aiming to enhance their long-term sustainability and profitability through decreasing customers' attrition.

Finally, the proposed intelligent MIS framework was presented to enhance the performance of the Customer churn prediction and evaluated its efficiency and usefulness in real-world business scenario, presenting opportunities for the use of Machine Learning. These findings align with the growing importance of adopting intelligent decision-support systems, which can help improve both operational and strategic management, and the efficiency of organizations.

VII. CONCLUSION

The purpose of this study was to compare the existing Machine Learning techniques to make the Management Information System more efficient with Customer Churn prediction. The study demonstrated the value of predictive analytics in the managerial decision-making process and how it can add value to the delivery of an organization's performance by turning raw customer data into valuable business insights.

Finally, machine learning algorithms were applied and tested including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours and a Support Vector Machine. Experiments indicated that what type of data is used, how the data is pre-processed, and which model is used leads to different degrees of accuracy in this prediction. Logistic Regression was found to be leading algorithm among evaluated algorithms as it gives best reliability and best balance by introducing the improvements in data pre-processing and balancing the classes.

As might be expected this study highlighted the importance of data pre-processing for the applications to be used in a machine learning. Feature scaling and SMOTE were among the methods that greatly contributed to better classification results and minimized the model's class skewness. So, Recall and F1 value for the optimized model were high, which indicated the efficiency of recognising those who were going to quit the organization.

Results indicate that machine learning implementation in MIS is likely to be very beneficial for the predictive ability, strategic planning as well as managing customer retention. Intelligent predictive systems can help organizations predict potential business risks, optimize resource usage and improve operational efficiency.

Also the research proves that the successful implementation of Machine Learning in MIS environment doesn't always need advanced and complex algorithms. Much simpler and more seemingly intelligent models can perform better in many instances when the right and appropriate "pre-processing" and "optimizing" algorithms are used and given appropriate acceptance.

Finally, this study underscores the potential of Machine Learning to improve business decision-making processes, its relevance for modern business firms as well as its effectiveness in managing company performance.

VIII. FUTURE WORK

The forecast results from the suggested system were very good but a few suggestions are presented for further research and improvements.

Synthetic Datasets implementation of more advanced ensemble methodologies and application of Deep Learning models is recommended as future work to further enhance classification stability and enhance the prediction accuracy. Furthermore, larger and more diverse datasets might be able to be used to enhance the model's generalization and robustness for different business sectors.

The embedding of real-time predictive analytics in the cloud-based Management Information Systems environment for adaptive decision-making within organizations is another potential research direction. Moreover, other measures of customer behaviour, such as customer satisfaction measures, extent of service usage or service interactions might further improve predictive power.

One of the potential solutions is the hybrid intelligence systems development – a mixture of Machine Learning algorithms, Business Intelligence and the visualization of data, transforming into fully automated decision supporting systems.

Finally, future research could implement techniques from XAI to make the model more interpretable and give more managers confidence in the automation predictive systems.

REFERENCES

[1] Lalwani, P., Mishra, M.K., Chadha, J.S. et al. Customer churn prediction system: a machine learning approach. *Computing* 104, 271–294 (2022). <https://doi.org/10.1007/s00607-021-00908-y>

[2] Shahimoon, M., et al. "Evolving Predictive Models for Customer Churn in Telecommunications." 2025 2nd

International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS). IEEE, 2025.

[3] B. Prabadevi, R. Shalini, B.R. Kavitha, Customer churning analysis using machine learning algorithms, *International Journal of Intelligent Networks*, Volume 4, 2023, Pages 145-154, ISSN 2666-6030, <https://doi.org/10.1016/j.ijin.2023.05.005>.

[4] Glory Sam a, Philip Asuquo a, and Bliss Stephen Customer Churn Prediction using Machine Learning Models, *Journal of Engineering Research and Reports*, 2024

[5] Mohammad Raquibul Hossain, Predicting Customer Churn in Telecommunications with Machine Learning Models, *Asian Journal of Research in Computer Science*, 2025.

[6] <https://www.kaggle.com/datasets/hassanamin/customer-churn>

[7] Ashrafuzzaman Hera, Abdullah Al Rian, et al. "Leveraging information systems for strategic management: Enhancing decision-making and organizational performance." *American Journal of Industrial and Business Management* 14 (2024): 1045-1061.

[8] Pearlson, Keri E., Carol S. Saunders, and Dennis F. Galletta. Managing and using information systems: A strategic approach. *John Wiley & Sons*, 2024.

[9] Rusilowati, Umi, et al. "Optimizing human resource planning through advanced management information systems: A technological approach." *Aptisi Transactions on Technopreneurship (ATT)* 6.1 (2024): 72-83.

[10] Dominic, M. Lishmah, et al. "The synergy of management information systems and predictive analytics for marketing." *Data-driven decision making for long-term business success. IGI Global Scientific Publishing*, 2024. 49-63.

Citation of this Article:

Mohanad Ali Hussein. (2026). Using Machine Learning Techniques to Enhance the Efficiency of Management Information Systems. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(6), 121-128. Article DOI <https://doi.org/10.47001/IRJIET/2026.106013>
