Customer Churn Prediction and Recommendation using an Intelligent System

M V K Om Aditya[1]
[1]M. Tech Student
Department of
Computer Science and Engineering
School of Engineering
Anurag University
Hyderabad, Telangana, India
mvkomaditya@gmail.com

Dr. Vadipina Amarnadh[2]
[2]Assistant Professor
Department of
Computer Science and Engineering,
School of Engineering,
Anurag University
Hyderabad, Telangana, India
amarnadhcse@anurag.edu.in

Dr. G. Vishnu Murthy[3]
[3]Professor and Dean
Department of
Computer Science and Engineering,
School of Engineering,
Anurag University
Hyderabad, Telangana, India
deancse@anurag.edu.in

***Abstract*—Customer churn remains a critical challenge in the retail sector, often resulting in significant revenue loss and increased acquisition costs. This paper presents an integrated deep learning framework that combines unsupervised clustering, predictive modeling, and rule-based recommendation to identify and retain at-risk customers. The approach begins with customer segmentation using KMeans clustering enhanced by Principal Component Analysis (PCA) for reduction of dimensional complexity of the dataset, followed by segment-specific churn prediction using a hybrid BiLSTM–CNN architecture. This model leverages temporal patterns and local feature extraction to improve prediction accuracy. Based on the predicted churn probability, a rule-based recommendation engine issues targeted digital coupons or retention strategies, enhancing business decision-making. The entire workflow is deployed through a user-friendly Streamlit web application, enabling real-time data analysis, visualization, and strategic action. Experimental results demonstrate that segment-specific models outperform global models, and the rule-based system offers interpretable, customizable recommendations. The proposed system is scalable, domain-adaptable, and effective in reducing churn while supporting informed, cost-efficient marketing interventions.**

Keywords— Churn Prediction, Deep Learning, Customer Segmentation, Recommendation System

I. INTRODUCTION

Customer retention is a vital aspect in business continuality, as holding on to current customers is substantially more lucrative than bringing in new ones [1]. Studies suggest that the cost incurred in gaining a new customer is about five times higher than the cost of holding the current customer [2]. In addition, businesses that successfully minimize the experience of customer churn not only have higher profitability but also enhanced brand recognition, as satisfied customers contribute to positive word of mouth and long-term loyalty. As a result, solutions for predicting and retaining customers have been developed [3].

Customer churn refers to a business experiencing significant losses due to frequent customer departures. This phenomenon, which is also termed customer attrition, happens when current customers cease utilizing a business's offerings. To retain current customers, a business must evaluate its customer information base to uncover the underlying reasons for their exit [4]. The prediction of customer attrition is to recognize customers that are most likely to exit the organization, which allows businesses to forestall possible whys and wherefores for churning and take precautionary strategies to resolve them [5].

Churn indicators vary across industries, making it difficult to develop a one-size-fits-all model for churn prediction. For example, if a customer's average annual account balance and annual transactions drop by 30 percent, banking industry professionals classify them as churners [6]. In a similar vein, if there is no action from a consumer for 90 days in a row, Edwine et al. [7] labelled them a churner. In addition, traditional customer churn management often assumes that all customers behave similarly, overlooking the diverse transaction histories and behavioural patterns that influence retention. This oversimplification leads to ineffective churn prediction models that fail to capture the complexities of customer behaviour. The primary objective of churn prediction is to minimize customer attrition, particularly among high-value clients, by directing retention efforts toward those who significantly impact profitability. Businesses must balance the expected lifetime value of a customer with the cost of retention strategies to ensure cost effectiveness. Retaining unprofitable customers may not justify the expense of a retention campaign. Therefore, instead of striving to improve overall retention rates indiscriminately, companies should focus their resources on retaining a strategically chosen subset of customers where the return on investment is maximized. Therefore, an effective churn management strategy must incorporate customer segmentation to tailor retention efforts to specific customer groups. By identifying distinct segments based on behavioural patterns and transaction history, businesses can optimize their retention strategies and allocate resources efficiently.

The simplest and most natural method to execute a retention effort is to immediately provide customers with customized coupons, ensuring timely and personalized engagement to encourage continued patronage. Moreover, sales can be augmented by enhancing the purchase conversion rate without incurring additional expenses, such as promotional events, by identifying customers at risk of churn in real-time and offering them tailored discount coupons, necessitating an AI-driven strategy for execution.

The three-step plan includes clustering of customers, followed by predicting at-risk customers, and recommending incentives in the form of discount coupons.

Customer segmentation is the technique of categorizing customers according to their common traits, forming the foundation for marketing strategies suited to each group [8].

Most of the machine learning Algorithms used to segment customers are unsupervised learning methods, such as K-means models or self-organizing maps (SOMs), or supervised learning methods like decision trees [9]. In order to predict customer attrition, there have been significant studies in the past of models that were learned using individual algorithms such as ANNs, Logistic Regression. And Decision Trees. However, lately, ensemble models or hybrid models that connect various models are also being attempted [10]. The recommendations are mainly based on association mining or the probability that a particular product was bought by consumers. Now, techniques like collaborative filtering, content-based, hybrid methods, and deep learning techniques, combining numerous supporting processing techniques, have gained momentum and have been applied by companies, like Amazon and Netflix, for their recommendation services [11], [12].

In this research, we employ DL techniques on customer transaction data to predict churn and provide personalized retention strategies. Our approach has the following key contributions: At first, we segment customers using K-Means and PCA, ensuring that each group receives a customized churn prediction model.

Then, we develop a BiLSTM–CNN-based churn prediction model for each segment, improving accuracy and robustness. We also integrate a rule-based recommendation system that issues targeted coupons to at-risk customers, enhancing retention efforts. Finally, we implement a Streamlit-based web application, allowing seamless customer segmentation, churn prediction, and recommendation deployment in a real-world retail setting.

II. Related Work

A. Customer Segmentation

Consumer segmentation is an important aspect of the modelling technique, but it is not incorporated in most of the proposed models. The first step in marketing research is to segment the consumer. When consumers are categorized according to comparable traits, marketing strategies can be developed that are tailored to each target segment. However, rather than being a standalone strategy, consumer segmentation should be used in conjunction with other marketing techniques. It is found that by adopting segmentation tactics, Businesses can outperform their competitors. This is achieved through unique and effective marketing strategies for each set of customers. Furthermore, segmentation allows businesses to better understand the wants and needs of their customers [15].

Traditionally, the most sought-after method for customer segmentation has been the RFM model. This model analyses customer purchasing behaviour based on three key attributes-recency, frequency, and monetary. The customers are then segmented into distinct groups according to their RFM scores [13], [14], [15], [16]. Christy et al. [17] discuss the RFM ranking technique to assess customer value. Their study utilizes RFM values computed after data preprocessing and employs the following clustering algorithms: repetitive median K-means, fuzzy C-means.

The K-means assigns each data point to the nearest of the k centroids, assuming that each point belongs to a unique singular cluster. The algorithm calculates the Euclidean distance between each data point and the centroids, iteratively refining cluster assignments until the total intra-cluster distance is minimized. Fuzzy C-means, unlike K-means, allows data points to belong to multiple clusters by calculating the probability that a data point belongs to each cluster. Repetitive median K-means initialize centroids using the medians of the variables [17].

Beyond traditional methods like RFM, modern customer segmentation increasingly uses machine learning. unsupervised learning algorithms such as K-means, DBSCAN, and Hierarchical Clustering are commonly used to identify hidden patterns in customer data. When predefined labels are available, supervised learning algorithms such as Decision Trees, Random Forests, or XGBoost can be employed. Dimensionality reduction techniques, such as Principal Component Analysis or deep learning methods such as autoencoders, are often applied before clustering to improve performance by focusing on the most relevant features [15]. Alkhayrat et al. [18] used PCA and deep learning, specifically autoencoders, to reduce dimensionality. Autoencoders, composed of an encoder, code, and decoder, compress input data into a lower-level code and later rebuild the input from that code. By training on a dataset, the autoencoder learns to reduce dimensionality while preserving essential information. Their study compared the performance of K-means clustering on the original dataset against the transformed PCA and auto-encoder datasets using the silhouette coefficient as an evaluation metric [18]. They found that reducing dimensionality with these techniques improved clustering performance and facilitated the identification of more meaningful customer segments.

Businesses are advantaged by this as it allows them to target and personalize their marketing efforts for a better customer experience.

B. Customer Attrition Prediction

The Customer churn, the loss of existing customers, is a significant concern for businesses for two primary reasons: it damages a company's reputation, and getting new customers is considerably expensive than holding existing ones [1]. Subsequently, research on churn prediction has shifted from empirical studies based on hypotheses to data-driven machine learning approaches [1]. Churn prediction, classically framed as a two-class classification problem, aims to categorize customers as either "churn" or "non-churn".

Several researchers have explored various machine learning algorithms for predicting churn. [19] compared the performance of the RF regressor, Decision Tree, Logistic Regression, Adaboost, and Gradient Boosting Classifier on the IBM Attrition dataset, finding that linear models performed best. Other studies have focused on segmenting customers before applying prediction models. For example, [20] used clustering to create different customer segments based on behaviour and then trained custom Random Forest models for each segment. This personalized approach aimed to improve the accuracy of churn predictions and enable more targeted retention strategies.

Tree-based algorithms have also been widely used for churn prediction. [21] Focused on predicting customer churn for Syriatel, comparing the performance of XGBOOST, Random Forest, GBM tree, and Decision Tree. XGBOOST consistently outperformed other algorithms, highlighting the importance of feature engineering and selection for the success of this model. Deep learning techniques have also gained traction in churn prediction. [22] compared the performance of Convolution Neural Networks to Artificial Neural Networks for predicting customer churn in the retail industry, finding that CNN achieved approximately 99 percent accuracy. The authors also suggested that AI techniques could be used to identify hidden customer behavioural patterns to further enhance model performance. Despite the promising results achieved by existing models, accurately predicting churn on unseen data (i.e., generalizing to new data distributions) remains a challenge.

Many models struggle with distribution shifts in real-world scenarios. To address this issue, more research is needed to investigate optimal parameter combinations and network architectures. A comparative analysis with existing models is also crucial to establish the effectiveness of any proposed approach. The model proposed in this study aims to estimate the likelihood of churn in the retail industry by focusing on these challenges.

C. Recommendation Engine

In the past decade, personalized recommendations, such as coupons, discounts, and perks, have become a central focus of marketing. Early personalization efforts relied primarily on association mining or simple probabilistic models of purchase behaviour. However, more sophisticated techniques have emerged, including collaborative filtering, content-based filtering, hybrid methods, and deep learning approaches, often incorporating various auxiliary processing techniques. Companies such as Amazon and Netflix have successfully implemented these advanced methods for their recommendation services.

The design of a recommendation system is heavily influenced by its intended objective. Several key techniques underpin modern recommendation systems. Collaborative filtering, explored by [23], analyses user preferences based on web data and is particularly well suited for platforms with large user bases. This approach constructs a user-item matrix, identifies user preferences for items, and predicts future purchases by suggesting items favoured by similar users. As (Rafsanjani et al., 2013) explain, while both the ranking and the recommendation systems present items in a sorted manner, recommenders aim to help users discover items they might not have found otherwise and strive for diversity to avoid overspecialization.

Other techniques used in recommendation systems include content-based filtering, knowledge-based systems, classifier-based models, and clustering-based models [24],[25]. Based on the user's history, previously liked items are identified and then used to recommend items similar to those in Content-based methods, while knowledge-based systems use explicit user preferences and product features to make recommendations. Classifier-based models frame recommendation as a classification task, predicting the likelihood of a user liking an item. Clustering-based models group users or items based on similarities and make recommendations within those groups. More recently, deep learning models have become increasingly popular due to their ability to handle time-series data, non-linear relationships, and diverse input formats. For example, [26] used a time series-based model for recommendations in the context of social IoT (sIOT).

III. CUSTOM COUPON ISSUANCE

The BiLSTM-CNN network is employed, and a recommendation system approach is used to issue digital discounts. These digital coupons are provided to clients with a high risk of churn. Our method(see fig. 1) includes the following steps:

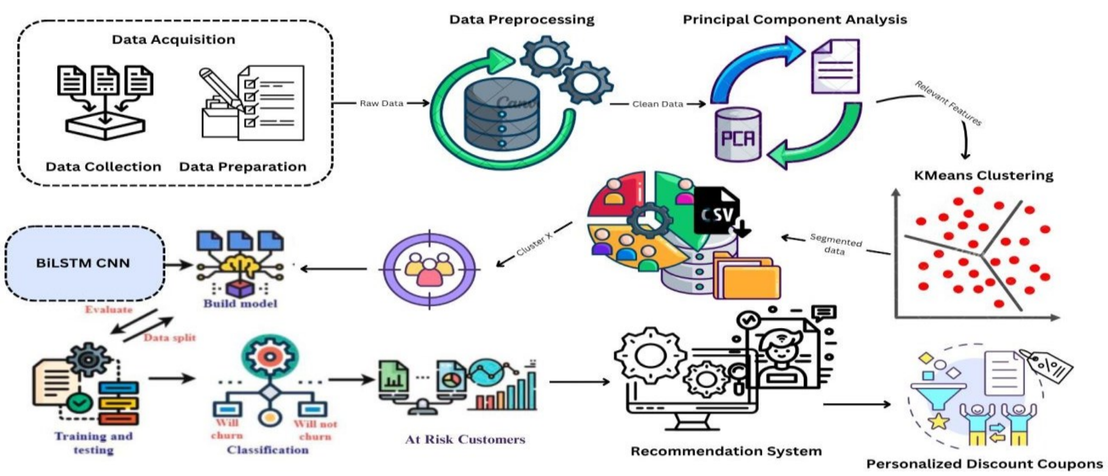
A. Customer Base Classification

In this study, K-Means clustering and Principal Component Analysis (PCA) are used to effectively segment customers. The data set includes key attributes such as age, usage frequency, support calls, payment delay, total spend, contract length, and subscription type. Before clustering, data undergo preprocessing, including handling missing values, feature selection, and standardization using RobustScaler to ensure consistency and mitigate the influence of outliers.

The Elbow Method is applied by analysing the Within-Cluster Sum of Squares (WCSS) to get the number of clusters that can be formed. The data set is then partitioned using K- Means, assigning each customer to the cluster with the nearest centroid. Since high-dimensional data can be challenging, PCA is utilized to reduce dimensionality while preserving maximum variance.

B. Churn Rate Estimation

Following the clustering procedure, the BiLSTM-CNN hybrid deep learning model is utilized to estimate the churn rate based on the separate segment's homogeneous behavior. Several layers make up our BiLSTM + CNN model, including the output layer, maxpool layer, convolutional layer, bidirectional LSTM layer, and flatten layer. Figure 2 depicts the overall design and functioning of the proposed hybrid model for predicting customer turnover.



The BiLSTM layer proves to be really helpful when used with data that has several features. This layer enables access to both future and past contexts by gathering long-term dependencies. The BiLSTM effectively retains the extra data required for precise predictions, in contrast to unidirectional LSTMs, which only look at past data and disregard future input. By combining forward and backward LSTM layers, the BiLSTM gathers data from the past and the future, enabling us to use the dataset's characteristics to provide precise predictions.

The convolutional layer extracts local n-gram features of the output of the previous layer. The CNN layer is composed of a number of distinct parts, including the flatten, maxpooling, and convolutional layers, which combine to provide an extensive feature vector. To create the feature map, the convolution filter matrix is applied to every possible window over the matrix. The size of the feature map is reduced by the maxpool layer. This strategy improves calculation speed by deleting non-maximal (less significant) features. The Flatten layer creates a feature vector of the previous step's pooled feature map that can be fed into the final classification layer. For prediction, a dense layer with only one neuron is employed. This dense layer computes the likelihood of "churn" and "non-churn" classes using the sigmoid activation function.

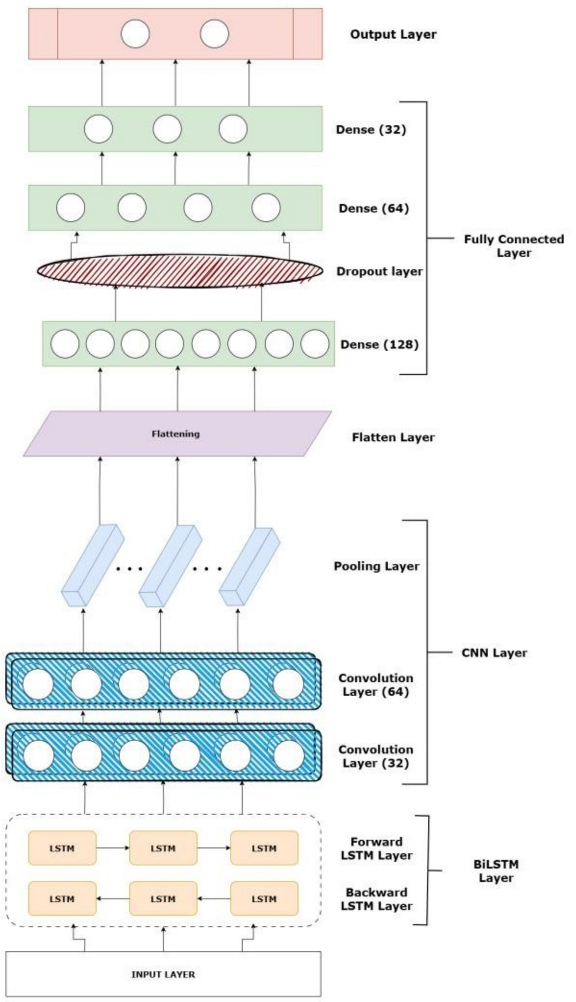


Fig. 2. Architecture of Proposed Hybrid model for Prediction of Churn

C. *Recommender System*

We aim to issue custom coupons to customers who are identified as high-risk churners using a deep learning-based churn prediction model. Specifically, our system recommends tailored retention offers—such as discount coupons or personalized emails—based on churn probability thresholds. These thresholds are defined through a rule-based system that maps the predicted churn probability of each customer to an appropriate retention strategy.

Unlike hybrid recommendation systems that combine multiple scoring algorithms, our rule-based engine emphasizes interpretability and business control by allowing domain experts to define discrete rules in a configurable JSON format. Each rule consists of a probability threshold and an associated recommendation. For instance, if a customer's churn probability exceeds 0.95, the system might issue a high-value coupon, whereas a medium-risk customer may only receive a retention email. This method ensures that retention efforts are both cost-effective and targeted.

In addition to the recommendation, the system provides a confidence score, which helps prioritize which customers should be targeted first. The simplicity of rule-based mapping allows easy testing and adjustment of recommendation thresholds using historical churn data. This makes the system ideal for real-time deployment in business environments where explainability, configurability, and rapid decision-making are essential.

D. *Streamlit Web Architecture*

To provide an interactive and user-friendly interface for the proposed customer retention framework, a web-based application was developed using Streamlit, a lightweight Python library for building data-driven web apps. The application serves as a unified platform that integrates customer segmentation, churn prediction, and personalized recommendation functionalities.

IV. Methodology

The objective of this study was to develop a predictive framework for customer churn using real-world retail data. The data set used consisted of 64,374 observations in 12 characteristics, carefully chosen to reflect relevant customer attributes in a retail setting. This data set was acquired in machine- readable CSV format, making it compatible with a wide range of machine learning tools and techniques. Upon acquisition, the data was loaded into a pandas data frame, providing a flexible and efficient structure for further processing.

Before applying any analytical techniques, the data set underwent a thorough pre-processing phase to ensure consistency, quality, and suitability for machine learning. The initial inspection revealed that several columns contained string- formatted data, which could not be directly used for numerical calculations. To address this, a comprehensive data cleaning step was performed. Categorical fields, particularly the target variable 'Churn', were encoded in numeric form: specifically, 'Yes' was converted to 1 and 'No' to 0. Furthermore, rows and columns containing only missing or null values were removed to maintain the completeness and integrity of the data set. Any inconsistencies in the formatting or irrelevant textual data were also corrected during this phase.

A critical observation during preprocessing was the imbalance in feature scales that posed a significant issue for algorithms such as KMeans, which rely on distance metrics that are sensitive to scale. Without proper standardization, features with wider numeric ranges could dominate the clustering process, leading to biased and less interpretable groupings. To mitigate this, data standardization was performed using the RobustScaler method. This approach was chosen specifically because the dataset exhibited right-skewed distributions and contained outliers as shown in figure 3 — conditions under which standard scaling methods like Min-Max or Z-score normalization tend to perform poorly. RobustScaler operates based on the interquartile range, making it more resistant to the influence of outliers.

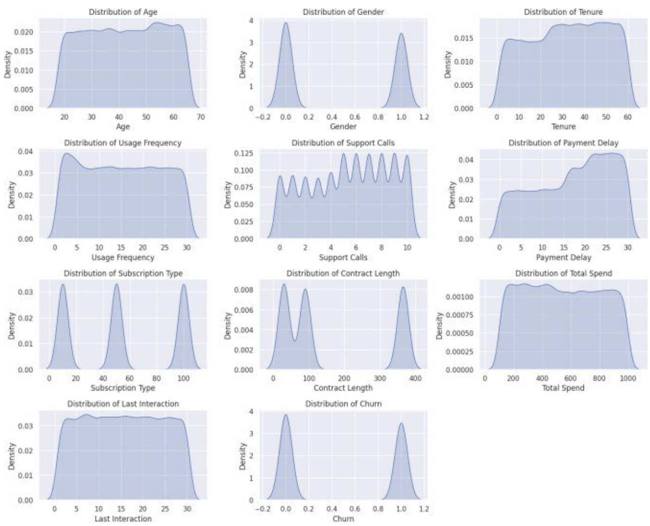


Fig. 3. Frequency Distribution Plot of all the features in the dataset

The numerical features were scaled using this technique, while categorical features such as 'Gender' and 'Churn' were excluded from scaling to preserve their qualitative characteristics.

This refined data set served as the foundation for downstream tasks, including customer segmentation using KMeans and PCA, as well as churn prediction through deep learning models such as BiLSTM-CNN.

Before developing our predictive model for churn, it was essential to segment the customer base into meaningful groups. Clustering allows us to identify underlying patterns in the data that are not immediately visible. The Elbow Method, which is based on evaluating the Within-Cluster Sum of Squares (WCSS), was used to determine the ideal number of segments for segmentation.

We computed the WCSS for a range of cluster values and plotted them to identify the 'elbow point' — the point beyond which additional clusters no longer significantly reduce the WCSS. In our case, the graph showed a sharp decline in WCSS values up to three clusters, after which the curve plateaued as seen in Figure 4. This distinct "elbow" at k = 3 suggested that a three-cluster solution would strike a good balance between model simplicity and interpretability.

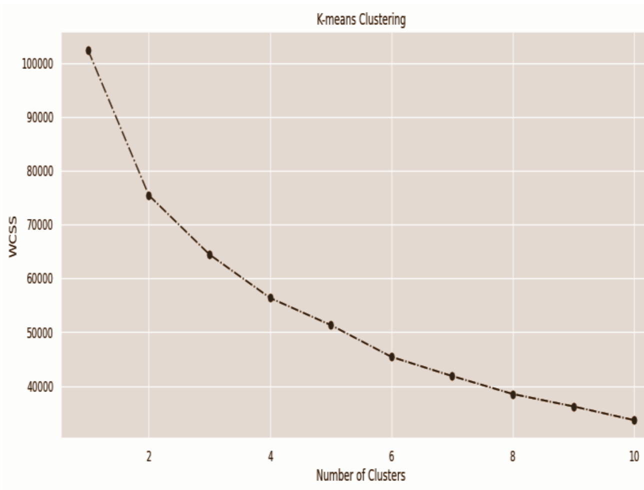


Fig. 4. WCSS plot with number of clusters equal to 10

Initially, K-Means clustering was performed directly on the feature set. A scatter plot of the clusters revealed some degree of separation. The green cluster, which had the highest values of Payment Delay, appeared well separated, while the red cluster showed significant overlap with the green one, especially along the vertical axis. Although the clusters were identifiable, the plot shown in figure 5 indicated that the segmentation relied heavily on just a few features, failing to capture more nuanced relationships present in the broader dataset. This limitation motivated the need for a more holistic dimensionality reduction approach.

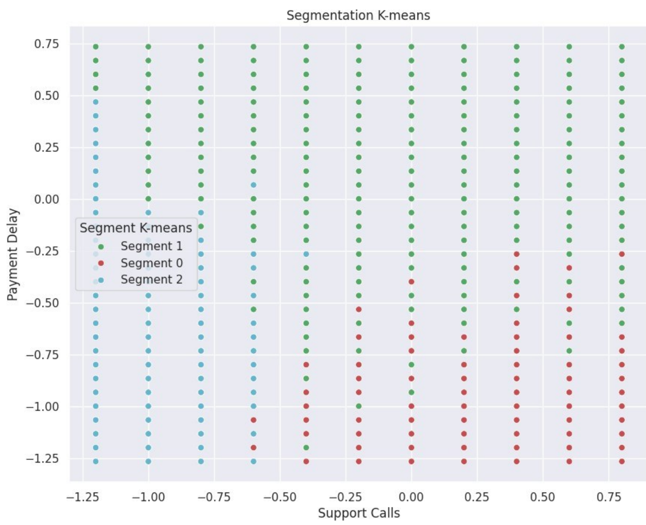


Fig. 5. Scatter plot of distinct segments formed by K-Means Clustering without PCA

To improve segmentation and better capture underlying structures, we employed Principal Component Analysis (PCA). PCA transforms the original feature space into a set of uncorrelated components that preserve most of the data's variance. We selected the top four principal components, based on their explained variance values: Component 1 (53.95), Component 2 (36.88), Component 3 (34.66), and Component 4 (33.53). This choice, supported by a scree plot, described in Figure 6, aligns with the elbow rule and reflects a trade-off between dimensionality reduction and information retention.

Table 1 Customer Segmentation Summary

Segment	Age	Gender	Tenure	Usage Freq.	Support Calls	Payment Delay	Subscription Type	Contract Length	Total Spend	Last Interaction	Churn
High-Value Churners	42.41	0.40	37.49	14.67	6.98	23.33	52.81	155.70	522.70	15.45	0.98
Low-Maintenance Customers	41.52	0.47	30.88	16.04	1.85	15.66	53.68	161.06	513.70	15.89	0.14
Loyalists	41.83	0.56	25.03	14.58	7.15	9.53	53.64	168.30	600.31	15.12	0.10

These four components collectively captured a substantial proportion of the data's variability and served as the basis for refined clustering. Having extracted the four principal components, we transformed the original dataset into PCA scores, resulting in a new 4-feature matrix. This transformation not only reduced dimensionality but also removed multicollinearity, enabling more effective clustering. We then re-applied the K-Means algorithm using the transformed PCA scores, with the number of clusters fixed at k = 3, consistent with our earlier analysis.

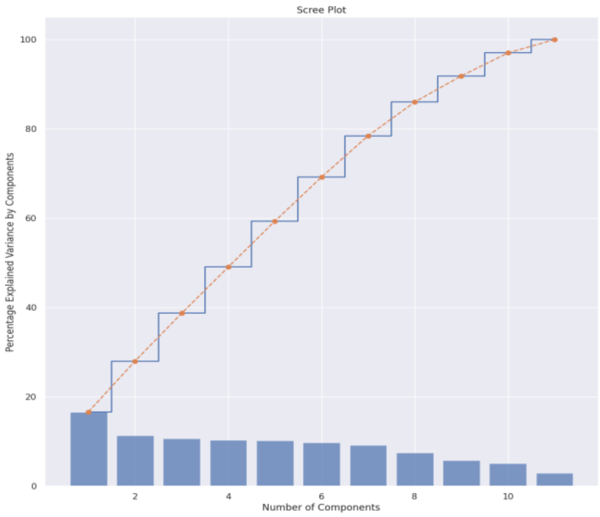


Fig. 6. Explained variance ratio of the features with principal components.

The results from K-Means clustering on PCA-transformed data showed significantly improved cluster separation and more meaningful groupings. Unlike the original feature-based clusters, the PCA-based clusters captured higher-order relationships and offered clearer visual delineation when plotted in a 2D or 3D space using the first few components. The segments formed were more coherent, both spatially and in terms of business interpretation, suggesting that PCA-enhanced clustering provides a superior segmentation strategy.

This segmentation was instrumental in downstream tasks, such as churn prediction and targeted recommendation strategies, providing a strong foundation for personalized interventions and marketing actions.

In order to estimate client turnover, we then used a dataset to run a number of simulations, which was sourced from Kaggle. It contains 12 features across 64374 rows. The data set was divided into two subsets to make it easier to create and evaluate the models. This was done in order to assist with the development and evaluation of the suggested churn prediction model. 70% of the samples were included in the training dataset, and the remaining 30% were included in the testing dataset, to achieve a well-balanced and representative data distribution for successful model assessment.

We identified four specific features from this dataset: tenure, usage frequency, support calls, and payment delay, which are given as input to the model.

These features were sent back and forth the LSTM layers, whose outputs were then combined and given to the next layer for processing.

Filtering is then executed, followed by the alignment of the filter matrix, after which element-wise multiplication is conducted between the filter matrix and the selected segment of the sentence matrix. The feature v2 is derived by aggregating all values from the preceding phase. Pooling is performed next on the CNN process. The appropriate window size, followed by a stride is selected. The window's movement around the feature map is determined by the stride. The maximum value, which is the output of the pooling process, is chosen in the following phase using the window from the previous stage. Next, the pooled feature matrix is turned into a feature vector by the flattening layer. The last classification then classifies in the customer into one of the two classes based on the features that were acquired.

Once the customers are classified as "No-Churn" or "Churn" respectively, the rule-based recommendation system is used to provide personalized retention strategies for customers identified as being at risk of churn. It operates by analysing the churn probability scores—predicted by a BiLSTM-CNN model—and matching them against a predefined set of business rules specified in a JSON configuration file. Each rule consists of a churn probability threshold and a corresponding retention action, such as offering a discount or sending a personalized email.

The recommendation system evaluates each customer's score against the defined thresholds in descending order of severity. Once a rule is matched, the associated recommendation is assigned. If no rules are matched, a default "No action" label is returned.

V. RESULTS

A. *Segmentation of Customers*

The customer segmentation analysis used K-Means clustering, Principal Component Analysis (PCA) for dimensionality reduction, while preserving most of the variance.

The K-Means algorithm applied on to the data set with the first four principal components, grouped customers into three segments. The three segments exhibit distinct behavioral and financial characteristics as described in Table 1. The High-Value Churners have the longest tenure and substantial spending but suffer from the highest churn rate (0.98), likely due to frequent payment delays and moderate support calls.

Table 2: Retention strategies generated based on confidence score by the recommendation system

Index	Tenure	Usage Frequency	Support Calls	Payment Delay	Predicted Churn	Confidence Score	Recommendation
0	25	14	4	27	1	85	Offer 15% discount
1	37	15	9	28	1	97	Offer 25% discount
2	42	27	9	21	1	95	Offer 25% discount
3	44	7	8	16	1	71	Send personalized retention email
4	31	6	2	29	1	92	Offer 15% discount

In contrast, Low-Maintenance Customers are highly stable, with a churn rate of just 0.14, minimal support interactions, and low payment delays, indicating satisfaction and self-sufficiency. Meanwhile, The Loyalists — who exhibit the highest total spending (600.31) and the longest contract lengths—demonstrate an even lower churn rate (0.10) despite making more support calls, suggesting that their concerns are effectively addressed. Among these, churn estimation is most critical for High-Value Churners, as their departure represents a significant revenue loss. Understanding their reasons for churn and implementing proactive retention strategies are essential to sustaining business profitability.

Figure 7 represents a scatter plot of PC1 vs. PC2, that was generated to visualize how the clusters were distributed in the transformed space.

Fig. 7. Scater Plot of the customer segments formed by the PCA transformed dataset

B. *Customer Churn Prediction*

To enhance the utility of the proposed churn prediction model, customers were first segmented using KMeans clustering on PCA-transformed features. This preprocessing step enables the model to focus on more homogeneous subgroups, allowing for more accurate churn detection.

To assess the efficacy of this, a comparative evaluation was conducted using both the unsegmented dataset and the individual customer segments derived via KMeans clustering on PCA-reduced feature space. The BiLSTM-CNN model was independently trained and evaluated on each of the identified segments, as well as on the full, unsegmented dataset.

To measure the performance standard classification metrics, including F1 Score, Precision, Accuracy and Recall were used. The empirical results described in figure 8, indicate that models trained on segment-specific data consistently outperformed the global model. Notably, the model corresponding to Segment 0 achieved an accuracy of 0.906 and an F1 score of 0.91, compared to 0.837 and 0.84, respectively, for the model trained on unsegmented data. These findings substantiate the hypothesis that prior segmentation enhances predictive performance by enabling the model to capture intra-group behavioral consistency and reducing variance introduced by heterogeneous customer profiles.

Accordingly, this supports the design rationale of the proposed workflow, which integrates customer segmentation as a prerequisite to churn prediction and recommendation generation.

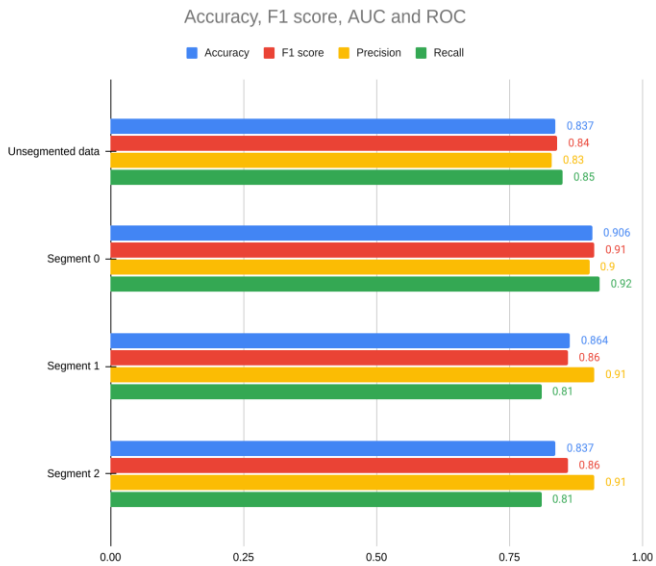


Fig. 8. Comparison of various performance metrices of the BiLSTM CNN model across various segmented and unsegmented data

C. *Comparison of Churn Prediction Models*

To weigh the predictive power of the proposed hybrid architecture, we benchmarked its performance against several classical machine learning models—Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and XGBoost—as well as baseline deep learning models, namely Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), and BiLSTM. The evaluation metrics included Accuracy, F1 Score, and ROC-AUC.

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	0.65	0.64	0.66	0.65	0.63
Random Forest	0.85	0.78	0.97	0.86	0.97
XGBoost	0.89	0.85	0.96	0.90	0.98
SVM	0.91	0.95	0.86	0.90	0.95
MLP	0.82	0.82	0.82	0.82	0.90
CNN	0.85	0.85	0.85	0.85	0.94
BiLSTM	0.71	0.74	0.71	0.71	0.69
BiLSTM-CNN	0.90	0.90	0.90	0.91	0.91

As shown in Table 3, the BiLSTM-CNN model achieved an F1 Score of 0.90 and ROC-AUC of 0.90, demonstrating a strong overall performance. Although XGBoost and SVM slightly outperformed BiLSTM-CNN in ROC-AUC (0.98 and 0.95 respectively), the hybrid model exhibited superior balance across metrics. Notably, BiLSTM-CNN outperformed the standalone BiLSTM model (F1 Score: 0.76, ROC-AUC: 0.70), indicating that the integration of convolutional layers significantly enhanced its ability to extract spatial features from sequential customer data.

The ROC-AUC/F1 Score trends depicted in Fig. 6 reinforce this observation. Classical models like Random Forest and XGBoost achieved high AUC values but at the cost of greater feature engineering complexity. In contrast, BiLSTM-CNN leveraged raw sequential patterns with minimal preprocessing and maintained competitive performance, suggesting its adaptability across diverse customer behavior datasets.

From a practical standpoint, the high F1 Score achieved by BiLSTM-CNN indicates its robustness in handling class imbalance—a common characteristic of churn datasets. This is particularly important in real-world scenarios where both false positives and false negatives incur business costs. Furthermore, its deep learning architecture enables the model to generalize better to subtle variations in customer behavior over time, thus offering greater utility in dynamic environments.

The BiLSTM-CNN model not only compares favorably with state-of-the-art classical models but also brings distinct advantages in terms of temporal pattern recognition, scalability, and minimal feature engineering. These attributes position it as a compelling solution for operationalizing churn prediction in real-world customer retention systems.

D. Rule Based Recommendation

The rule-based recommendation system was applied to a test dataset consisting of 19,396 at-risk customers, each with a predicted churn probability generated by the deep learning churn prediction model. Table 2 describes the retention strategy suggested by the system based on the confidence score. Customers with a confidence score above 0.95 were consistently recommended high-value retention actions such as a 25% discount coupon. Customers with moderately high churn probabilities (e.g., between 0.85 and 0.95) received mid-tier incentives such as a 15% discount coupon. Those with churn probabilities around 0.70 to 0.85 were advised to receive non-monetary retention strategies, such as personalized retention emails.

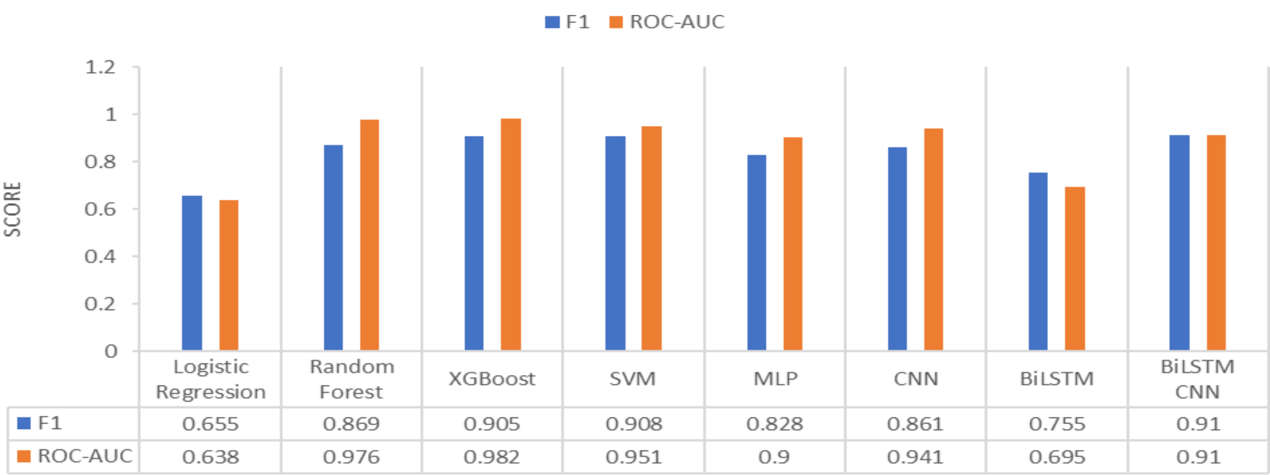
CONCLUSION

Our research presents an integrated framework for customer retention that leverages unsupervised and deep learning techniques to address the persistent challenge of customer churn in the retail sector. By combining KMeans clustering with PCA, BiLSTM-CNN-based churn prediction, and a rule-based personalized coupon recommendation system, the proposed system demonstrates a comprehensive approach to segmentation, prediction, and actionable decision-making. The deployment of the model through a user-friendly Streamlit web application enables marketers and decision-makers to seamlessly upload customer data, visualize segmentation, predict churn risks, and generate tailored retention strategies.

Experimental results show that segment-specific churn models provide more accurate predictions compared to a generic model, and rule-based recommendations offer interpretable and easily modifiable logic to drive real-world marketing campaigns. The modularity and scalability of the framework ensure that it can be adapted to diverse datasets and evolving business needs. Overall, the system offers a practical solution to reduce churn rates and improve customer engagement through intelligent data-driven insights.

However, certain limitations that open up opportunities for future research do exist. First, instead of static rules, reinforcement learning can be employed to dynamically adapt recommendation strategies based on campaign performance and user feedback.

Second, incorporating interpretability techniques such as SHAP or LIME into the churn prediction module can help



business users understand why a customer is predicted to churn, thus increasing trust in the system.

Third, future versions could support streaming data and real-time churn prediction to enable proactive customer interventions.

Fourth, with appropriate adjustments, the system architecture can be generalized for use in other sectors such as telecom, banking, or subscription services.

REFERENCES

[1] G. Nie, ''Finding the hidden pattern of credit card holder's churn: A case of China,'' in Proc. Int. Conf. Comput. Sci. Cham, Switzerland: Springer, 2009, pp. 561–569.

[2] A. D. Athanassopoulos, ''Customer satisfaction cues to support market segmentation and explain switching behavior,'' J. Bus. Res., vol. 47, no. 3, pp. 191–207, Mar. 2000.

[3] J. Ahn, J. Hwang, D. Kim, H. Choi, and S. Kang, ''A survey on churn analysis in various Bus. Domains,'' IEEE Access, vol. 8, pp. 220816–220839, 2020.

[4] Kumar, S. & Chandarkala, D. A survey on customer churn prediction using machine learning techniques. *Int. J. Comput. Appl.* 154(10), 13–16 (2016)

[5] Rodan, A., Faris, H., Alsakran, J. & Al-Kadi, O. A support vector machine approach for churn prediction in telecom industry. *Int. J. Inf.* 17(8), 3961–3970 (2014).

[6] M. Alizadeh, D. S. Zadeh, B. Moshiri, and A. Montazeri, ''Development of a customer churn model for banking industry based on hard and soft data fusion,'' IEEE Access, vol. 11, pp. 29759–29768, 2023.

[7] N. Edwine, W. Wang, W. Song, and D. Ssebuggwawo, ''Detecting the risk of customer churn in telecom sector: A comparative study,'' Math. Problems Eng., vol. 2022, pp. 1–16, Jul. 2022.

[8] M.-S. Chang, H. Kim, and Joong, ''A customer segmentation scheme base on big data in a bank,'' J. Digit. Contents Soc., vol. 19, no. 1, pp. 85–91, 2018.

[9] N. Chang, ''Improving the effectiveness of customer classification models: A pre-segmentation approach,'' Inf. Syst. Rev., vol. 7, no. 2, pp. 23–40, 2005.

[10] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, ''Applying data mining to telecom churn management,'' Expert Syst. Appl., vol. 31, no. 3, pp. 515–524, Oct. 2006.

[11] J. Wen and W. Zhou, ''An improved item-based collaborative filtering algorithm based on clustering method,'' J. Comput. Inf. Syst., vol. 8, no. 2, pp. 571–578, 2012.

[12] M. Pham and Cuong, ''A clustering approach for collaborative filtering recommendation using social network analysis,'' J. Univers. Comput. Sci., vol. 17, pp. 583–604, Feb. 2011.

[13] W. Jo-Ting, L. Shih-Yen, and W. Hsin-Hung, ''A review of the application of RFM model,'' African J. Bus. Manage., vol. 4, no. 19, pp. 4199–4206, 2010.

[14] J. T. Wei, S.-Y. Lin, Y.-Z. Yang, and H.-H. Wu, ''The application of data mining and RFM model in market segmentation of a veterinary hospital,'' J. Statist. Manage. Syst., vol. 22, no. 6, pp. 1049–1065, Aug. 2019.

[15] M. Pakyurek, M. S. Sezgin, S. Kestepe, B. Bora, R. Duzagac, and O. T. Yildiz, ''Customer clustering using RFM analysis,'' in Proc. 26th Signal Process. Commun. Appl. Conf. (SIU), May 2018, p. 2.

[16] P. A. Sarvari, A. Ustundag, and H. Takci, ''Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis,'' Kybernetes, vol. 45, no. 7,pp. 1129–1157, Aug. 2016.

[17] A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa, "RFM ranking – An effective approach to customer segmentation", Journal of King Saud University - Computer and Information Sciences, Volume 33, Issue 10, 2021, Pages 1251-1257, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2018.09.004

[18] Alkhayrat, M., Aljnidi, M. & Aljoumaa, K. A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA. J Big Data 7, 9 (2020). https://doi.org/10.1186/s40537-020-0286-0

[19] Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R. & Aghamdi, H. S. Prediction of employee attrition using machine learning and ensemble methods. Int. J. Mach. Learn. Comput. 11(2), 110–114 (2021).

[20] Li, W. & Zhou, C. Customer churn prediction in telecom using big data analytics. IOP Conf.Ser. Mater. Sci. Eng. 768(5), 1–6 (2020).

[21] Ahmad, A. K., Jafar, A. & Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. J. Big Data 6(1), 1–24 (2019).

[22] Seymen, O. F., Olmez, E., Dogan, O., Orhan, E. R. & Hiziroglu, K. Customer churn prediction using ordinary artificial neural network and convolutional neural network algorithms: A comparative performance assessment. J. Sci. 36(2), 720–733 (2023).

[23] G. Chandrasekaran and D. J. Hemanth, ''An intelligent framework for online product recommendation using collaborative filtering,'' in Proc. Int. Conf. Sustain. Expert Syst. (ICSES). Cham, Switzerland: Springer, 2021, p. 249.

[24] A. Felfernig and R. Burke, ''Constraint-based recommender systems: Technologies and research issues,'' in Proc. 10th Int. Conf. Electron. Commerce, Aug. 2008, pp. 1–10.

[25] T. Zhang and V. S. Iyengar, ''Recommender systems using linear classifiers,'' J. Mach. Learn. Res., vol. 2, pp. 313–334, Mar. 2002.

[26] Y. Chen, M. Zhou, Z. Zheng, and D. Chen, ''Time-aware smart object recommendation in social Internet of Things,'' IEEE Internet Things J., vol. 7, no. 3, pp. 2014–2027, Mar. 2020.