Journal of Future Artificial Intelligence and Technologies

E-ISSN: 3048-3719

Research Article

RetenNet: A Deployable Machine Learning Pipeline with Explainable AI and Prescriptive Optimization for Customer Churn Management

Amirthanathan Prashanthan [1,*], Rinzy Roshan [1], and MWP Maduranga [2]

[1] Faculty of Business Analytic Center, National Innovation Centre, National Institute of Business Management, Colombo 00600, Sri Lanka; e-mail : prashanthana@hotmail.com; rinzyroshan@gmail.com
[2] Department of Electrical and Electronic Engineering, University of Sri Jayewardenepura, Nugegoda 10250, Sri Lanka; e-mail : pasanm@sjp.ac.lk
* Corresponding Author : Amirthanathan Prashanthan

Abstract: This study presents RetenNet, a comprehensive framework for managing customer churn in telecommunications, integrating predictive modelling, prescriptive optimization, and explainable artificial intelligence (XAI) incorporated with Large Language Models (LLMs). The process commences with the IBM Telco dataset, divided in an 80:20 ratio into training and testing sets. Categorical variables are converted by one-hot and label encoding, whilst class imbalance is mitigated using SMOTEENN. Min-max scaling and mutual information-based feature selection guarantee data appropriateness for machine learning models. Five classification algorithms, i.e., Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boosting (XGB), Logistic Regression (LR), and Multi-Layer Perceptron (MLP) are assessed. The SVM model utilizing an RBF kernel exhibits optimal performance. In conjunction with nested cross-validation, Bayesian optimization guarantees excellent hyperparameter optimization and generalization. Performance is evaluated using the F1-score to highlight the implications of false negatives and false positives in churn situations. The methodology additionally incorporates fuzzy rule-based clustering, facilitating flexibility in customer segment identification for intervention priority. Prescriptive optimization uses linear integer programming to distribute retention budget according to model results and business constraints. SHAP waterfall plot employed to guarantee transparency and facilitate actionable insights. Furthermore, Gemini 1.5 Flash, a multimodal LLM, generates analysis and produces contextual recommendations derived from the SHAP waterfall plot. RetenNet offers a comprehensive and interpretable approach to the churn management pipeline, including classical machine learning, prescriptive optimization, and LLM-driven explainable artificial intelligence to enhance decision-making in customer retention efforts.

Keywords: Churn Management; Customer Churn Prediction; Customer Retention; Explainable Artificial Intelligence; Fuzzy Rule-Based Clustering; Large Language Models; Machine Learning; Prescriptive Optimization.

Received: May, 23rd 2025
Revised: June, 9th 2025
Accepted: June, 14th 2025
Published: June, 17th 2025
Curr. Ver.: June, 17th 2025

1. Introduction

Customer churn prediction has garnered significant interest across various sectors, including banking, telecommunications, insurance, gaming, and academia [1]. Customer churn represents a considerable challenge for businesses when customers discontinue services or transition to alternative providers [2]. In the telecommunications business, customer churn significantly diminishes profit and company revenue [3]. Telecommunications firms are currently seeing challenges in enhancing profit margins due to elevated license fees, costs associated with spectrum allocations, and increasing customer expenditures while simultaneously lowering pricing. Moreover, acquiring new consumers is far more expensive than retaining current ones [4]. Numerous research studies have examined The telecommunications churn issue, employing various machine learning and artificial neural network techniques for customer churn prediction [5]–[13].

Explainability, fundamentally a human-centric aspect associated with perception and cognition, positions the findings of an AI system. XAI has focused chiefly on interpretability, defined as the system's ability to clarify or communicate model predictions in understandable terms [14]. Telecommunication churn with XAI techniques often lacks practical application [15]–[18]. In the telecommunications sector, XAI techniques, such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP), are essential. These strategies provide a more profound understanding of customer churn prediction determinants [19].

Optimization concerns the best distribution of finite resources among competing activities, subject to limitations dictated by the specific situation under investigation. Prescriptive optimization is broadly characterized as a mathematical formulation designed to optimize the allocation of limited resources [20]. Prescriptive optimization, focusing on active decision-making from predictive results, is understudied in telecommunications customer churn management. The application of LLMs to various downstream tasks has gained significant traction in AI research communities and enterprises, with numerous novel uses being identified and investigated daily. LLMs proficient at comprehending and producing human-like writing have discovered significant applications in numerous domains [21]. The efficacy of LLMs is augmented by their capacity to conform to the particular style and tone of the text they analyze, rendering the outputs more accessible and contextually relevant [22]. However, the applicability of XAI with LLMs in telecommunication churn analysis is unexplored.

No past study has successfully combined predictive analytics, prescriptive optimization, and XAI with LLMs to interpret and suggest actionable strategies for retaining customers in telecommunications. This gap presents a unique opportunity to develop a deployable machine learning pipeline that combines predictive classification, prescriptive optimization, and explainable AI-augmented by LLMs to provide a thorough and effective solution for customer churn management. The primary objective of this study is to develop an integrated, interpretable, and deployable customer retention system (RetenNet) by integrating ML-based prediction, SHAP-based explanation, and budget-aware prescriptive optimization, utilizing LLMs to translate interpretability into actionable recommendations.

Our principal contributions are as follows:

- Developed RetenNet, a comprehensive and implementable customer retention framework that incorporates machine learning-driven churn prediction, dual-layer explainable AI, and prescriptive optimization designed for pragmatic use in the telecommunications industry.
- Performed a comparative assessment of various classification algorithms, utilizing comprehensive data preprocessing (e.g., balancing, scaling, feature selection), hyperparameter optimization, and cross-validation to determine the best efficient model for churn prediction.
- Utilized XAI to improve model transparency and user confidence by identifying and visualizing the principal features affecting churn prediction local scales. Utilized LLM to convert XAI outputs into clear, context-sensitive customer intervention messages, facilitating the transition from technical insights to practical business communication.
- Created a prescriptive optimization framework that integrates fuzzy rule-based clustering with linear integer programming to produce a cost-effective retention budget optimization strategy in accordance with anticipated churn risks and business constraints.

The structure of the paper is as follows. Section 2 presents the literature review, succeeded by the methodology described in Section 3. Section 4 delineates the result and discussion of RetenNet. Section 5 presents the strengths, limitations, and future works, while Section 6 provides the conclusion.

2. Literature Review

2.1. Predictive Analytics

Predictive analytics has emerged as a potent instrument for mitigating customer churn in the telecommunications sector [8]. Numerous studies in the telecoms business illustrate how predictive analytics employs machine learning and artificial neural network classification techniques to predict churn likelihood [3], [6], [16], [18], [23]–[34]. Machine learning refers to algorithms that rely on models and inferences from data processing, avoiding given instructions explicitly [35]. Machine learning methods develop a computational model from sample data, known as "training data," to enable predictions or assessments without direct programming for the specific task. It is considered a particular type of artificial intelligence. Machine Learning algorithms are primarily classified into four categories: Supervised learning, Unsupervised learning, Semi-supervised learning, and Reinforcement learning [36], [37].

Numerous machine learning algorithms are employed for classification prediction, such as Naive Bayes (NB) [38]–[44], Linear Discriminant Analysis (LDA) [42], [45], [46], Logistic Regression (LR) [39]–[41], [43], [47], [48], Decision Tree (DT) [40], [42], [49]–[53], K-Nearest Neighbours (KNN) [38], [39], [41], [42], [54]–[56], Support Vector Machines (SVM) [39], [41]–[43], [57], Random Forest (RF)[38]–[40], [43], [51], [58], [59], Adaptive Boosting (AdaBoost) [40]–[42], [60], and Extreme Gradient Boosting (XGBoost) [13], [40], [42], [61], are widely used in machine learning classification algorithms in predictive analytics.

Artificial neural networks (ANN) emulate biological brain networks via statistical models. The fundamental unit of an ANN is the artificial neuron or perceptron, which mathematically represents a neuron and comprises four basic components: inputs, weights, a net sum function, and an activation function to generate the output [62]. An artificial neural network consists of three main processing layers: the input, hidden, and output layers. Numerous ANN algorithms are employed for classification tasks, but Multi-layer Perceptron (MLP) [5], [41], [61], [63], Convolutional Neural Network [5], [31], [41], [64], and Long Short-Term Memory Recurrent Neural Network [31], [41], [43], [65]–[67] are most popular for classification prediction in literature.

2.2. Prescriptive Analytics

Prescriptive analytics is the most sophisticated category of business analytics, providing substantial insights and value to organizations. It seeks to recommend optimal decision-making alternatives to capitalize on anticipated future outcomes by leveraging extensive data sets [68]. Prescriptive analytics integrates predictive analytics results and employs artificial intelligence, optimization algorithms, and expert systems within a probabilistic framework to deliver adaptive, automated, constrained, time-sensitive, and optimal decisions [69], [70]. Predictive analytics can be utilized alongside prescriptive analytics for enhanced decision-making in advance [71], [72]. This study utilized logic-based modelling and mathematical programming to optimize the retention budget for customers with a high likelihood of churn.

Logic-based modelling is a theoretical depiction of the causal sequence leading to a certain conclusion. This includes rule-based systems, expert knowledge representation, and domain-specific information gathering information to enhance proactive decision-making in prescriptive analytics applications. Fuzzy rule-based systems, or fuzzy rules, are typically categorized as logic-based models for prescriptive analytics [71]. Fuzzy rule-based clustering employs a supervised classification approach to do unsupervised cluster analysis. It seeks to independently analyze the potential clusters within the data patterns and define them using comprehensible fuzzy rules. The simultaneous categorization of data patterns using these fuzzy rules clarifies the actual boundaries of the clusters [73]. Optimization methods have garnered research attention in prescriptive analytics to obtain optimum solutions across several objectives. The primary aim is often considered to be the reduction of operating expenses [74]. Linear programming and its expansions are arguably the most utilized optimization techniques in prescriptive analytics. It is a method for optimizing a linear objective function constrained by linear equalities and inequalities [75].

2.3. Explainable AI

Highlighting the notion of transparency, researchers utilized XAI to transform the “black box” into a "glass box" [76] by offering technical solutions that autonomously produce explanations for the rationale behind specific decisions or recommendations over others. Merely around 20% of XAI assessment studies used human participants [77], and social science theories related to human information processing and decision-making were infrequently integrated [14], [78]. This investigation aims to enhance the development of user research in XAI by focusing on non-expert users. The limited user studies in XAI have mostly concentrated on interpretability, defined as the system's ability to clarify or communicate model predictions in understandable terms [14].

Dodge et al. [79] suggest that XAI can function as a conduit for human engagement, enabling users to identify and address fairness issues in AI systems. An XAI method is considered model agnostic if it can elucidate the predictions of any machine learning model, such as SHAP and LIME [33], [80]. Alternatively, it is model-specific, clarifying the predictions of a certain machine learning model, such as feature significance incorporated inside the Random Forest framework [59]. An XAI methodology is classified as local when it elucidates a singular prediction (e.g., Shapley values, LIME) and as global when it explains the complete model (e.g., feature importance techniques, SHAP) [81], [82]. SHAP utilizes Shapley values from game theory to deliver local or global explanations regarding feature importance for any machine learning model [80], [83]. The Shapley value is characterized as the incremental contribution of a variable's value to prediction over all possible "coalitions" or subsets of features [80]. Benefits include SHAP facilitating contrastive explanations, possessing robust theoretical support, and providing comprehensive explanations that are equitably allocated among feature values. Drawbacks: SHAP necessitates substantial computational resources [84]. LIME uses a local surrogate model to interpret individual predictions (local interpretation) [85]. Local surrogate models are interpretable models, such as linear regression and decision trees, utilized to elucidate individual predictions of opaque machine learning algorithms [83]. Advantages as LIME exhibits local fidelity, enabling the interpretation of individual predictions, and provides user-friendly explanations. Drawbacks include the selection of a surrogate model being subjective, the delineation of an instance's neighborhood being ambiguous, and outcomes being contingent upon the selection of neighbors [84].

2.4 Large Language Model

Recently, substantial advancements have been observed in language models, mostly because of transformers [86], enhanced processing power, and the accessibility of extensive training data. These advancements have facilitated a transformative shift by allowing the development of LLMs to emulate human-level performance across several tasks [87]–[89]. The capabilities of LLMs to address various jobs with human-level proficiency are accompanied by prolonged training and inference times, substantial hardware demands, and elevated operational expenses. These requirements have constrained their adoption and created chances to develop superior architectures [90]–[92] and training methodologies [93], [94].

Multimodal LLM (MLLMs) [95]–[98] have garnered considerable interest due to their exceptional multimodal functionalities. LLMs are enhanced by aligning them with a pre-trained visual encoder through text-image datasets, enabling LLMs to engage in dialogues incorporating image inputs [99], [100]. MLLMs have experienced significant breakthroughs in the past year, enhancing off-the-shelf LLMs to accommodate multimodal inputs or outputs through cost-effective training methodologies [101]. The bulk of LLMs have mostly focused on text processing. There is a growing tendency towards a multimodal strategy that includes studying images, text, and video data. OpenAI and Google have introduced general-purpose multimodal models, specifically GPT-4-vision and Gemini-Pro-Vision. The large Language and Vision Assistant (LLaVA), an open-source project that combines vision encoding with a large language model, has demonstrated outstanding results in several visual-based tasks [102].

3. Methodology

This section delineates the methodological procedures of the study. Figure 1 concisely delineates the top-level steps of the methodology.

3.1. Data and preprocessing

3.1.1. Dataset

The research utilizes the publicly accessible IBM Telco Dataset [103]. The study dataset consists of 7,043 instances and 21 attributes, with 20 distinct features and one target variable. The dataset consists of 26.54% of samples that are classified as churn.

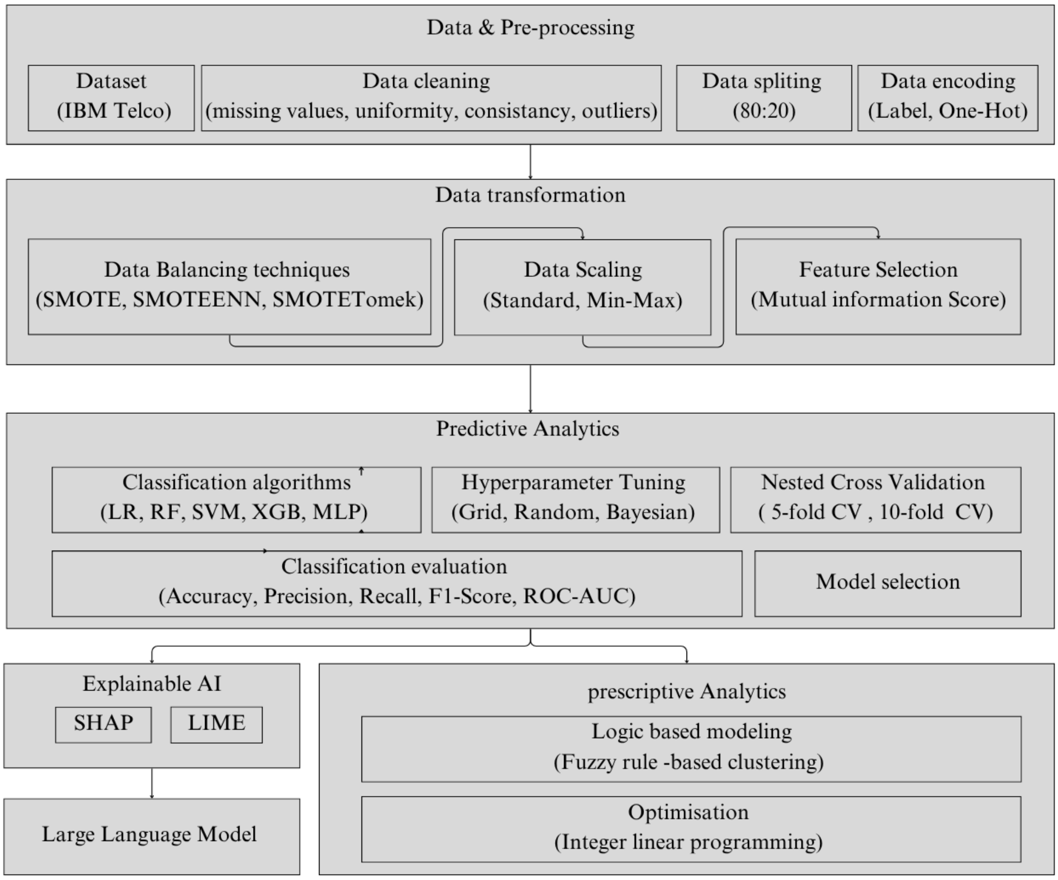


Figure 1. Top-level methodology of research

3.1.2. Data cleaning

Data preparation was conducted to refine the dataset, encompassing the removal of extraneous columns, standardization of column names for uniformity, validation of data types, and resolution of missing values and outliers. Consequently, box plots were employed to evaluate the existence of data outliers in the numerical variables of tenure, monthly charges, and total charges, indicating that none of the variables displayed outliers in the overall distribution.

3.1.3. Data encoding

The study employed label encoding and One-Hot encoding methods to transform identified nominal data into numerical data. The dataset was split into training and testing subsets at an 80:20 ratio before encoding.

3.2. Data transformation

3.2.1. Data balancing

Due to substantial data imbalance in the employed datasets, it was necessary to implement data balancing strategies to address the issue. Three data balancing techniques, namely SMOTE, SMOTE+ENN, and SMOTE+Tomek, were employed to identify the appropriate model for selecting the optimal classification model.

3.2.2. Data scaling

Standard Scalar and Mini-Max techniques were employed to convert "Monthly Charges", "Total Charges" columns to identify the appropriate model for classification tasks.

3.2.3. Feature selection

Mutual Information is the simplest concept for implementing feature selection. A substantial amount of literature has focused on using mutual information (MI) as a metric for selecting an appropriate subset of features [104]. This study employed mutual information with cross-validation to identify the most appropriate features for predicting the target class.

3.3 Predictive analytics

3.3.1. Selection of algorithms

This study employed a suite of sophisticated classification techniques to provide consistent and accurate models. This includes SVM [105], a function-based algorithm acknowledged for its effectiveness in high-dimensional spaces, and LR [106], a primary technique for binary classification problems. Furthermore, the Random Forest Classifier, a tree-based ensemble learning method proposed by Breiman [59], was implemented. This method is particularly adept at managing large datasets with a multitude of features. The XGBoost [107] was chosen for its exceptional predictive accuracy, which was achieved by combining multiple weak prediction models into a single and robust model. Ultimately, MLP, a Neural Network [65], has been implemented due to its exceptional ability to identify intricate data patterns by means of interconnected nodes, rendering our methodology both comprehensive and effective.

3.3.2. Hyperparameter optimization

To determine the best machine learning parameters for classification, this study uses three hyperparameter optimization techniques: grid search, random search, and Bayesian search, along with five-fold cross-validation.

3.3.3. Cross-validation

A nested cross-validation procedure is used in the standard methodology: the inner cross-validation selects the hyperparameters, and the outer cross-validation uses cross-validation to refine the hyperparameters to produce an objective estimate of the algorithm's expected accuracy [108]. This study, 5-fold cross-validation was used for hyperparameter tuning, while 10-fold cross-validation was used as the outer validation method to assess model performance. Strikes a balance between model selection and computational efficiency, making it easier to adjust hyperparameters effectively without significantly lowering the performance of the finished model.

3.3.4. Performance evaluation

The effectiveness of a classification model can be evaluated using various performance metrics. The performance metrics include accuracy (1), precision (2), recall (3), and F-measure (4). The metrics' values can be calculated from the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values obtained from the confusion matrix.

$$\mathrm{Accuracy} = \frac{\mathrm{Accurate\ predictions}}{\mathrm{Total\ number\ of\ predictions}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

$$\mathrm{Precision} = \frac{\mathrm{Correctly\ classified\ actual\ positives}}{\mathrm{Everything\ classified\ as\ positive}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$\mathrm{Recall} = \frac{\mathrm{Correctly\ classified\ actual\ positives}}{\mathrm{All\ actual\ positives}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$\mathrm{F1\text{-}score} = 2 * \frac{\mathrm{Precison}\ *\ \mathrm{Recall}}{\mathrm{Precisison}\ +\ \mathrm{Recall}} = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

3.4. XAI implementation with LLM integration

3.4.1. Selection of XAI technique

This study used XAI methods like SHAP (beeswarm, dependence, force, waterfall, bar, scatter, and dependence plots) and LIME to generate global and local explanations for specific predictions. The following hypothesis was developed to test and find a suitable plot for XAI.

- **H1**: Local explanations exert a greater influence than global explanations in decision support.

Typically, a distinction is made between global and local explainability. Global explanation enables the comprehensive explanation of AI model predictions, thereby 'opening' the black box of these algorithms. Local interpretation explains the model's rationale for each specific sample. The two forms of explainability represent different aspects of the same concept: global explainability identifies overarching patterns, whereas local explainability examines these patterns at the individual level [109]. End-users, including analysts, managers, or operational staff, frequently want actionable insights for specific projections (e.g., the reasons behind a particular customer's expected churn) rather than merely abstract global patterns.

- **H2**: SHAP-based methodologies provide more reliable and theoretically robust explanations than heuristic approaches such as LIME.

SHAP and LIME are the predominant local explainable artificial intelligence methods. LIME is entirely model-agnostic, signifying its independence from the prediction model allowing its application to any linear or nonlinear model. Conversely, the SHAP toolbox encompasses both model-agnostic XAI tools, like the SHAP Kernel Explainer, and mod-el-specific XAI tools, such as the TreeExplainer, which is optimized for tree-based mod-els[[110]. Nonetheless, LIME has been criticized for its inconsistency, as identical inputs do not consistently provide the same outputs [111], and its local approximation fails to maintain a stable correlation with the model's global level. Conversely, SHAP possesses four advantageous properties: efficiency, symmetry, dummy, and additivity [112], offering mathematical assurances to overcome the local-to-global constraint. The rationale for SHAP is based on Shapley values from cooperative game theory, which guarantees consistency and local correctness in feature attribution [111], [112] LIME evaluates how data variances affect machine learning model predictions. LIME explains the importance of feature for individual prediction, but there are no stronger theoretical assurances than SHAP [83], [113].

- **H3**: SHAP waterfall plots provide superior interpretability for non-technical stakeholders compared to alternative SHAP visualizations.

In contrast to SHAP force plots or summary plots, waterfall plots graphically deconstruct a prediction incrementally, emulating human logical reasoning from the fundamental value to the ultimate outcome. The following table 1 illustrates the application of various SHAP plots.

Table 1. Application of various SHAP plots .

Plot	Application
SHAP Beeswarm plot	This plot can present several features, prioritized by significance, and show the impact's direction and shape. Colour as a scale can make it hard to see the relationship's structure. Explaining global effects and feature importance [83], [113], [114]
SHAP Waterfall plot	Waterfall plots show all facets of a sample's expected SHAP values. This plot focuses on a single sample for local interpretation. This helps examine model outliers, SHAP value outliers, and specific interest [83], [113], [114]
SHAP Bar plot	The bar plot measure of feature relevance is easier to understand because SHAP values are expressed in the same units as model predictions. This representation lacks clarity about impact direction and relationship monotonicity [83], [113], [114]
SHAP Scatter plot	The scatter plot shows the correlation between a single feature value (x-axis) and SHAP values (y-axis). Detailing feature effects [83], [113], [114]
SHAP Decision plot	SHAP values are directly represented in decision graphs, simplifying interpretation. Vertical representation of Waterfall plot [83], [113], [114]
SHAP Force Plot-Single	Arrow Rearrangement from Waterfall plot [83], [113]
SHAP Force Plot-Interactive	The SHAP force plot lets you visualise all study populations interactively by feature value similarity and model output. Vertically stacked force graphs. Hard to understand [83], [113]
SHAP Dependence plot	Dependence plots are necessary to understand the relationship between feature values and model predictions. Details on feature effects [83], [113]

- **H4**: Interactive and narrative-driven SHAP waterfall plots enhance trust and usability in high-stakes contexts.

A waterfall plot visualizes the expected SHAP values of an individual sample, displaying all aspects. The model characteristics are positioned along the y-axis, with the corresponding value for each unique sample indicated in grey. The SHAP value for each feature pertaining to this specific sample is displayed in the main panel, indicated by an arrow for each row or feature. The row is red (blue) if the SHAP value elevates (diminishes) the prediction f(x) relative to the anticipated or mean prediction. This plot style offers a localized interpretation as it concentrates on a singular sample. Figure 2. presents the waterfall plot for a specific customer.

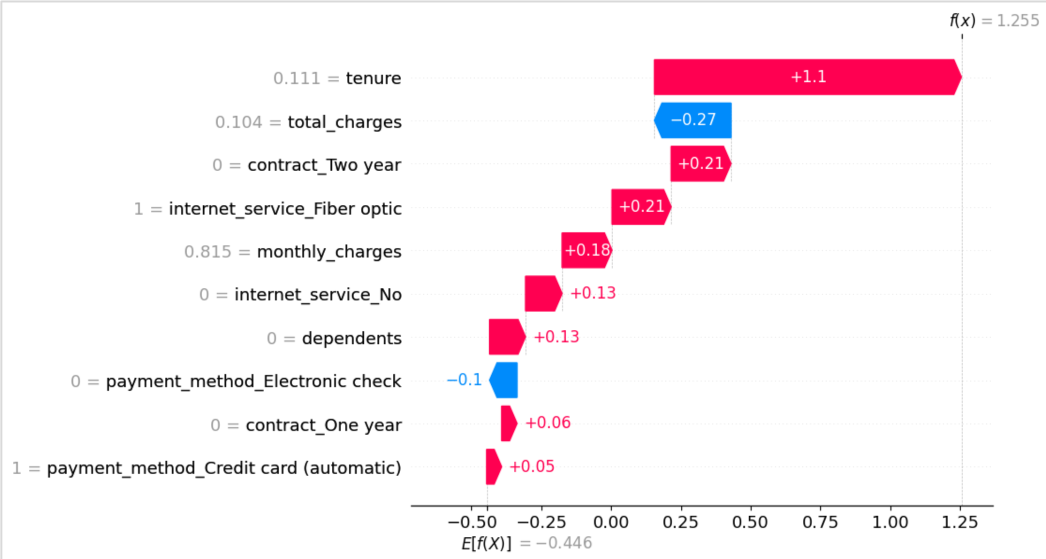


Figure 2. SHAP waterfall plot

3.4.2. Selection of LLM

The Gemini 1.5 Flash transformer decoder model is intended to optimize the utilization of tensor processing units (TPUs) with reduced latency for model serving. It is trained using higher-order preconditioned methods and performs parallel computation of attention and feedforward components. English queries generate over 650 characters per second, over 30% faster than the second-fastest model evaluated. Gemini 1.5 Flash achieves the fastest output generation for all languages tested [115]. The Gemini 1.5 Flash model offers users alternative payment structures based on their requirements, accommodating little chores in the Free Tier and extensive operations in the Paid Tier. The Free Tier delivers substantial value for customers with moderate demands, whereas the Paid Tier offers additional possibilities for those with greater volume and intricate requirements [116].

3.5. Prescriptive optimisation

3.5.1. Fuzzy rule-based clustering

Fuzzy rule-based systems, or fuzzy rules, are typically categorized as logic-based models for prescriptive analytics [71]. The study uses a fuzzy rule-based approach to prioritize customers with high Customer Lifetime Value who are at risk of leaving. The process involves fuzzy set generation, rule-based development, clustering, and cluster assignment. Fuzzy sets indicate the degree to which a value relates to a set, with three categories: poor, average, and good. A set of fuzzy rules determines the link between features and output, classifying customers as Low Engagement, Standard, or Premium. The fuzzy inference system establishes how each data point (customer) is related to each fuzzy category. After calculating membership values, consumers are placed in one of three clusters by the highest membership degree.

3.5.2. Linear integer programming

Linear programming is a widely employed subset of mathematical programming, especially as most of the publications that have been reviewed concentrate on optimization [71].

The study developed a linear programming-based optimization model to determine the optimum number of customers likely to churn who qualify for discounts within a limited budget. The model prioritizes high-value consumers identified in the preceding customer lifetime value-based clustering phase using fuzzy rule-based clustering while adhering to budgetary limitations. Linear integer programming denotes a category of combinatorial constrained optimization problems characterized by integer variables, whereby the objective function is linear, and the constraints are represented as linear inequalities [117]. This study aims to optimize the number of customers whose churn probability is decreased to below the threshold value by strategically selecting customers for intervention. The decision variables indicate the selection of customer intervention and the extent to which churn probability can be mitigated for every customer. The constraints in the optimization model ensure that the chosen consumers adhere to the budget and attain the targeted churn reduction. The optimization strategy involves customer classification, formulating the linear programming problem, solution selection, and an iterative solution process. Premium customers receive the utmost priority, followed by Standard and Low Engagement customers. The deterministic solution (PULP's CBC solver) is used to identify the ideal customer interventions while minimizing costs and maximizing churn reduction.

4. Results and Discussion

4.1. Deployment of RetenNet

We created a robust and deployable web application named RetenNet to implement retention strategies informed by predictive data. The system incorporates predictive analytics, explainable AI (XAI), prescriptive optimization, and LLMs into a cohesive architecture. The comprehensive configuration of the deployed RetenNet is outlined in Table 2 below.

Table 2. Deployed RetenNet configurations.

Step	Techniques / Methods	Selection / Tools
Data transformation	Data encoding	One-hot encoding, label encoding
	Data scaling	Min-max scaling
	Feature selection	Mutual information score
Predictive analysis	Classification model	Support vector machine (SVM) with RBF kernel
Prescriptive analytics	Clustering	Fuzzy Rule-Based Clustering
	Optimization	Linear Integer Programming
Explainable AI	Model explanation	Shap waterfall plots
Natural language layer	LLM integration	Gemini 1.5 flash for explanation translation
Deployment Framework	Frontend & backend integration	Streamlit
	Development tools	VS Code, GitHub (version control)

The architecture of RetenNet, seen in Figure 4, facilitates seamless interaction between the user interface and the foundational model logic. Streamlit is employed to develop a dynamic web interface that takes user inputs, transmits them to backend prediction and optimization models, and presents interpretable results. The development and deployment process uses Code for codebase management and GitHub for version control and collaborative development.

4.2. Predictive analysis

4.2.1. Classification analysis

Machine learning algorithms, including SVM, RF, LR, XGBoost, and MLP, were used to evaluate the most effective classification model for churn prediction. Various configurations were examined by systematically integrating data balancing techniques (SMOTE, SMOTEENN, and SMOTE Tomek), data scaling methods (Min-Max Scaler and Standard Scaler), and hyperparameter optimization strategies (Grid Search, Randomized Search, and Bayesian Optimization). These configurations sought to optimize model performance, improve generalizability, and rectify class imbalance present in churn datasets. Before model training, feature selection was conducted via the Mutual Information Score, which assesses the relationship between each feature and the target variable, thereby retaining only the most informative qualities and minimizing noise. The models were meticulously assessed using essential performance indicators, i.e., accuracy, precision, recall, and F1-score, to guarantee a thorough and equitable evaluation of predictive efficacy. The optimum classifier was discovered through lengthy experimentation because of its consistent and high performance, rendering it ideal for incorporation into the interpretable and deployable RetenNet system.

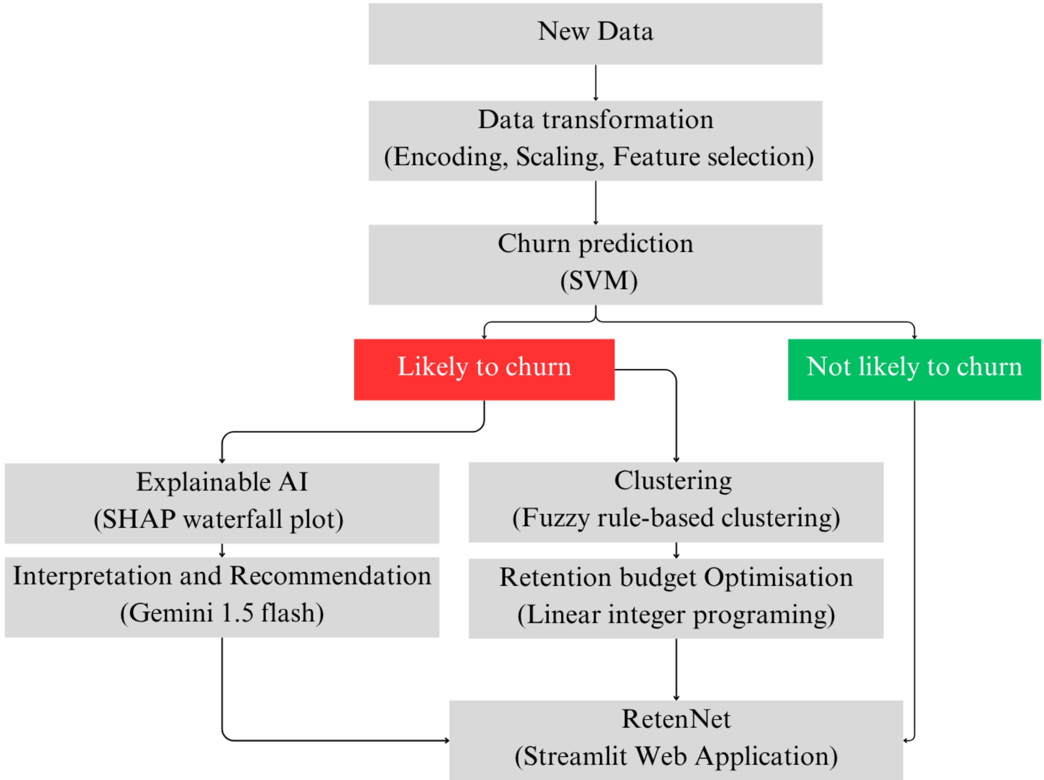


Figure 3. Deployed RetenNet Architecture

4.2.2. Performance metric trade-off

The business loses retention possibilities if a customer is likely to leave but misclassified as unlikely. False negative churn forecasts cost the organization a lot. Unique offers, customized services, and proactive communication require churn awareness. This may affect sales and loyalty. Even if it misclassifies non-churners, the company must detect as many churn projections as feasible. It's cheaper than ignoring churn but may cause overtreatment. Recall and precision matter. Overly optimistic churn models may waste resources on doubtful clients. Discounts or promotions for departing customers may raise costs. Precision increases the model's customer turnover prediction. Saving money on non-leaving client retention incentives is crucial. Because of false negatives (missing churn estimates) and false positives (misclassifying loyal customers as churners), model selection requires the F1-Score. Churn prediction needs perfect equilibrium. Reduce re-source-wasting churn prediction errors and false positives. The F1-Score penalizes exaggerated false negatives and positives equally to establish equilibrium and prevent the model from favoring one mistake type. F1-Score was used as a ranking metric to choose the best classifier model. The following Table 3 exhibits the top 10 model configurations ranked by F1-score.

4.2.3. Optimum classifier

SVM configured with SMOTEENN, Min-Max scaling, and Bayesian Optimisation demonstrates an optimal F1-Score and elevated Recall (Sensitivity), making this model appropriate for the classification task based on the utilized dataset. The optimal parameters for the SVM were C = 8.648674412457455 and the kernel type configured to "rbf," as established by the optimization procedure. Figure 4 below illustrates the confusion matrix of the optimum classifier.

Table 3. Top 10 model configurations ranked by F1-score

Rank	Model Configuration	Accuracy	Precision	Recall	F1-Score
1	SVM + SMOTEENN + Min-Max scaling + Bayesian Search	0.750177	0.519504	0.783422	0.624733
2	SVM + SMOTEENN + Min-Max scaling + Grid Search	0.750177	0.519573	0.780749	0.623932
3	SVM + SMOTEENN + Standard scaling + Grid Search	0.732434	0.497496	0.796791	0.612539
4	SVM + SMOTEENN + Standard scaling + Random Search	0.731015	0.495826	0.794118	0.610483
5	SVM + SMOTEENN + Standard scaling + No HPO	0.727466	0.491749	0.796791	0.608163
6	RF + SMOTE + Min-Max scaling + Grid Search	0.781405	0.585492	0.604278	0.594737
7	RF + SMOTE + Standard scaling + Grid Search	0.781405	0.585492	0.604278	0.594737
8	XGB + SMOTE + Standard scaling + Grid Search	0.779986	0.582474	0.604278	0.593176
9	XGB + SMOTE + Standard scaling + No HPO	0.779986	0.582474	0.604278	0.593176
10	RF + SMOTETomek + Standard scaling + Grid Search	0.781405	0.586387	0.598930	0.592593

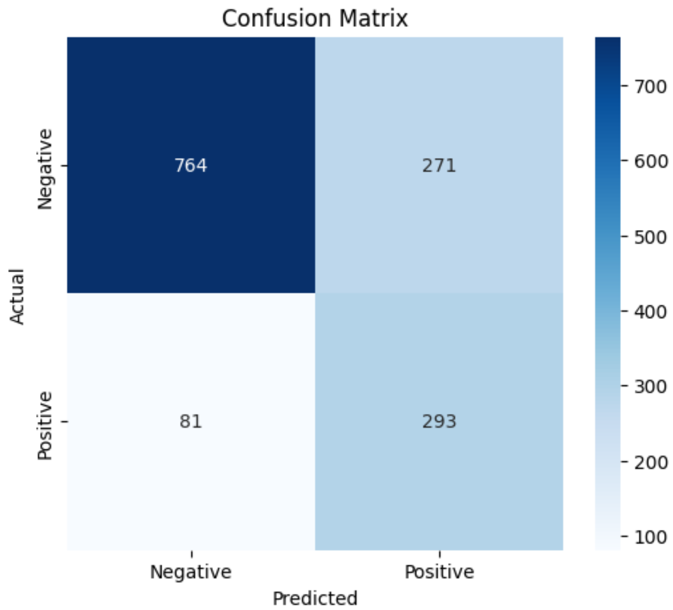


Figure 4. Confusion matrix of optimum classifier

4.3. XAI Analysis

According to predictive analytics, the customer identified by ID: 7130-VTEWQ has been classified as likely to churn. To facilitate transparent and elucidative decision-making, the RetenNet autonomously produced a SHAP waterfall plot, which delineates the contribution of each feature to the ultimate churn prediction. Figure 5 illustrates how several features affect the anticipated probability of churn, either increasing or decreasing it.

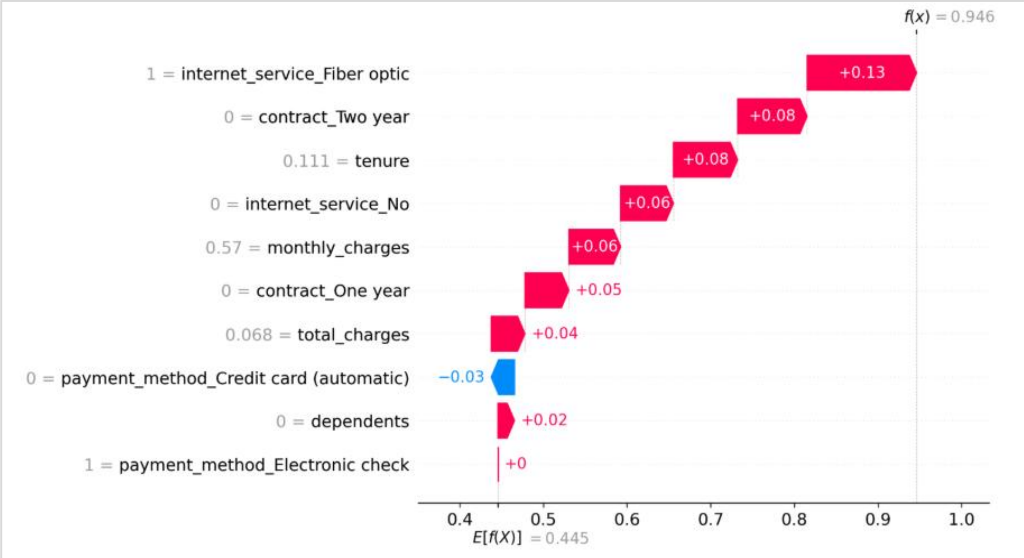


Figure 5. SHAP waterfall plot for customer ID: 7130-VTEWQ

Figure 6 demonstrates the application of the system prompt in Gemini 1.5 Flash to transform the SHAP waterfall plot into a natural language explanation. This Image-to-Text production approach facilitates an effortless and automated interpretation of intricate model outputs, enhancing accessibility for business stakeholders and non-technical users.

"""

You are a highly experienced professional in the field of telecommunications, with deep expertise in Explainable Artificial Intelligence (XAI) methodologies. Your primary analytical tool is the SHAP (SHapley Additive exPlanations) waterfall plot, which you use to investigate how individual features influence predictive outcomes at the customer level. Your role is critical in translating complex model behavior into clear, actionable insights that support strategic business decisions.

Your task is to analyze the SHAP waterfall plot for specific customer prediction, focusing on the following aspects:

1. **Feature Contributions**: Identify which features contribute positively or negatively to the prediction score.
2. **Magnitude of Impact**: Assess the magnitude of each feature's contribution, highlighting those that have the most significant influence on the prediction.
3. **Behavioral Insights**: Derive insights into customer behavior or operational patterns based on the feature contributions.
4. **Actionable Recommendations**: Provide recommendations based on the analysis that can inform business strategies or customer engagement approaches.
5. **Transparency and Clarity**: Ensure that the analysis is presented in a clear and understandable manner, suitable for stakeholders who may not have a technical background.

Your analysis should be objective, data-driven, and focused on the specific features and their contributions as depicted in the SHAP waterfall plot. Your insights will be used to guide business decisions, so accuracy and clarity are paramount.

Disclaimer

This explanation serves as a preliminary analytical review of the SHAP waterfall plot for specific customer prediction. It is not intended to be the sole basis for strategic or operational decisions. For a comprehensive evaluation, these insights must be corroborated with further analysis and reviewed in consultation with both data science and business operations teams. Additional contextual or cross-model investigation may be necessary to validate conclusions.

"""

Figure 6. System prompt used in Gemini 1.5 Flash for generating insights from SHAP waterfall plot

Figure 7 illustrates a specific SHAP waterfall plot analysis segment, emphasizing significant contributing characteristics. Figure 8 presents a collection of actionable insights and customized recommendations from the analysis. These insights indicate certain retention measures, such as the provision of discounts, alteration of contract types, or improvement of client interaction. These recommendations facilitate proactive measures to mitigate churn risk for this specific customer.

Key Drivers of High Churn Probability:

The plot clearly indicates that three features are the most significant drivers pushing this customer's churn probability higher:

1. **Internet Service (Fiber Optic):** This feature has the largest positive impact (+0.13). Having fiber optic internet service strongly increases the likelihood of churn for this customer. This suggests that customers with fiber optic service might be more prone to switching providers, possibly due to price sensitivity, competition, or dissatisfaction with the service.

2. **Contract Type (Two-Year):** This feature contributes positively (+0.08). The lack of a two-year contract indicates a higher churn risk. Customers without long-term contracts are more likely to switch providers.

3. **Tenure:** This feature also contributes positively (+0.08). While the exact tenure is not explicitly stated, the normalized value (0.111) suggests a relatively short tenure. This is consistent with the observation that customers with shorter tenures are more likely to churn.

Figure 7. Interpretation of SHAP waterfall plot (Figure 5) using Gemini 1.5 Flash

Actionable Recommendations to Improve Customer Retention:

Given the customer's profile, the key areas for intervention are their short tenure and payment method. The fiber optic internet service is a positive factor, suggesting satisfaction with the service itself.

1. **Proactive Engagement & Onboarding:** Since the customer has a very short tenure, a proactive onboarding program is crucial. This could include:
   - **Welcome call/email:** A personalized welcome call or email within the first week to address any initial concerns or questions.
   - **Tutorial/FAQ access:** Providing easy access to tutorials or frequently asked questions to help them understand their plan and services fully.
   - **Early Engagement Survey:** A short survey after a month to gauge their satisfaction and identify potential issues.
2. **Payment Method Incentives:** While Electronic Check is a common method, it may be worth considering offering incentives to switch to a more convenient or automated payment method (e.g., auto-pay). This could include:
   - **Discount on monthly bills:** A small discount for setting up auto-pay.
   - **Promotional offers:** Bundled offers or discounts for switching to auto-pay.
   - **Improved payment experience:** Ensure a seamless and user-friendly online payment system.
3. **Upselling/Cross-selling (Cautious Approach):** Given the short tenure, avoid aggressive upselling. However, consider:
   - **Targeted offers:** Based on their fiber optic plan, suggest add-on services that complement their current plan (e.g., premium streaming packages, enhanced security features) only after demonstrating value through the onboarding program.

Figure 8. Actionable recommendation generated from SHAP waterfall plot (Figure 4) using Gemini 1.5 Flash

4.4. Optimisation analysis

This interface (Figure 9) enables business users to evaluate retention budget eligibility for at-risk customers. The sample indicates customer ID 7130-VTEWQ, which was identified as potentially at risk of churning. RetenNet acquires pertinent churn insights and indicates whether the customer is eligible for a retention discount based on predictive risk and policy criteria.

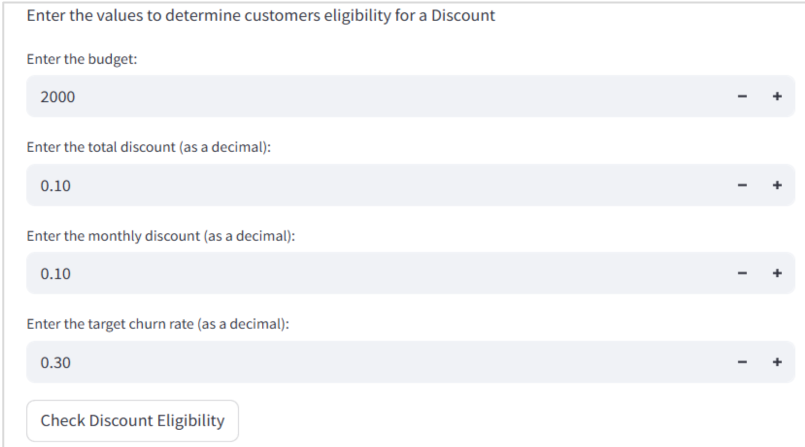
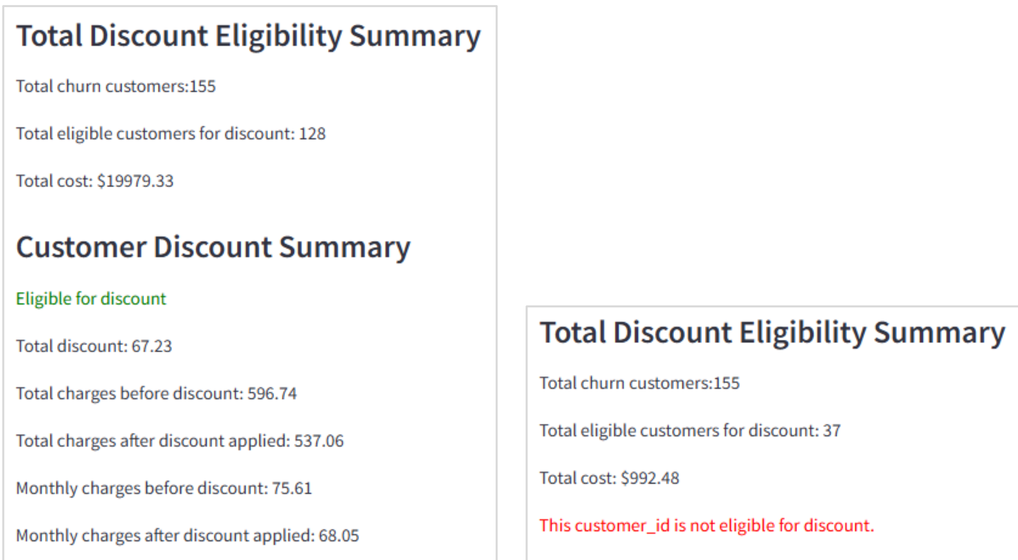


Figure 9. User interface for verifying customer eligibility for retention discount

The graphical interface, Figure 10(a), offers a detailed view of discount distribution for customer ID 7130-VTEWQ. The overall discount eligibility summary presents cumulative metrics, comprising the overall count of churn-prone customers (155), the number of customers who qualified for a discount (128), and the corresponding retention expenditure ($19979.33).

The Customer Discount Summary section outlines the discount granted to a designated customer ID: 7130-VTEWQ identified as qualified for the discount. The system delineates financial data, including the overall discount amount ($67.23), total charges before and after the discount from $596.74 to $537.06, and monthly billing adjustments—from $75.61 to $68.05 following the application of the discount. This capability assists business teams in making informed retention decisions based on predictive risk and cost-effectiveness. The graphical interface, Figure 10(b), offers a detailed view of discount distribution for customer ID 7130-VTEWQ, which is not eligible for a discount due to budget limitations.



(a) (b)

Figure 10. (**a**) User interface for displaying customer eligibility for retention discount; (**b**) Interface shows customer's non-eligibility for retention discount due to limited budget

To demonstrate the transformation of churn prediction into a real, quantifiable, and economically beneficial business strategy, we executed a series of simulations employing fuzzy clustering to discern actionable customer segments and integer programming to optimize discount distributions within diverse retention budget limitations. Table 4 offers compelling proof that churn prediction, when combined with prescriptive optimization methods, can yield concrete and strategically viable business outcomes:

Table 4. Simulation results for optimization

Available Retention Budget	Discount of "Total Charges"	Discount of "Monthly Charges"	Total Likely to Churn	Eligible for Discount	Utilized Retention Budget	Discount for customer ID: 7130-VTEWQ
$20,000.00	10%	10%	155	128	$19,979.33	$67.23
$10,000.00	20%	10%	155	78	$9,752.89	$126.91
$7,500.00	15%	15%	155	76	$7,283.94	$100.85
$5,000.00	10%	10%	155	72	$4,855.96	$76.15
$4,000.00	10%	5%	155	70	$3,956.11	$63.45
$3,000.00	5%	5%	155	82	$2,925.24	$33.62
$2,000.00	0%	20%	155	122	$1,999.26	$15.12
$1,000.00	10%	10%	155	37	$992.48	Not eligible

The results indicate that the following points provide substantial evidence reinforcing the significance of prescriptive optimization in improving business decision-making.

- Resource-efficient strategy via Linear Integer Programming:

Integer programming was employed to efficiently allocate available retention budgets, ensuring that the highest number of high-risk consumers received retention incentives without surpassing budget constraints. This underscores the operational viability of AI-generated insights.

- Scalability and adaptability:

The system optimized discount techniques and consumer targeting to enhance effectiveness through several budget tiers, ranging from $1,000 to $20,000. For example A budget of $20,000 was allocated to target 128 high-risk consumers with moderate discounts of 10%, resulting in extensive coverage and balanced cost-effectiveness. Conversely, with merely $2,000, the system emphasized consumers who might be retained by a substantial monthly discount (20%) without an overall reduction, successfully targeting 122 cases with low expenditure.

- Personalized recommendations:

The "Discount for Customer ID: 7130-VTEWQ" column indicates that the system generates individualized suggestions based on customers' propensity to churn and financial worth. This accuracy prevents generalized strategies and facilitates micro-targeted retention initiatives.

- Economic Justification:

Each configuration illustrates how a data-driven methodology guarantees optimal budget allocation, proving that customer retention strategies are financially justifiable and quantifiable in ROI.

5. Strengths, limitations, and future work

This section provides a comprehensive analysis and assessment of the research study's strengths, limitations, and future directions into the broader knowledge domain. This research addresses the gap in the literature on customer churn in the telecoms sector by combining predictive analytics, prescriptive optimization, and XAI to formulate proactive retention tactics. It enhances transparency and interpretability by using SHAP to make machine learning models understandable to non-technical stakeholders. The study optimizes retention budgets using rule-based clustering and linear programming, ensuring cost-effective budget allocation. The system is user-friendly, making recommendations accessible to non-technical business users. The research offers practical business impact, integrating theoretical insights with practical methods to reduce churn and improve retention rates.

Based on a public dataset, the study presents a framework for predicting customer behavior in telecom services. However, the model's generalizability and applicability are limited due to its use of public datasets and lack of real-world validation. The model's predictive performance is moderate, with a recall of 79% and an F1-score of 62%. The framework's generalization challenges include the need for retraining and revalidation in different environments. The system's computational complexity and scalability may also pose issues. The model's explainability and LLM strategy generation constraints may lead to oversimplification or miscommunication, and the model's effectiveness in formulating retention techniques may be limited. The static budget optimization model, which relies on rule-based clustering and linear programming, may not adapt to changing corporate environments or customer responses unless continuously revised.

This study offers a thorough framework for predicting customer churn and formulating retention strategies in the telecommunications sector; however, numerous opportunities for further research and development exist to improve its scope, robustness, and adaptability. The study provides a framework for predicting customer churn and formulating retention strategies in the telecommunications sector. Future research could focus on real-time churn monitoring systems, integrating unstructured data, and using reinforcement learning for retention strategy optimization. Personalized incentive recommendation systems could be developed to adjust customer preferences and behavioral patterns. The system's cross-domain applicability could be extended to other sectors with high customer churn rates. Enhanced explainability through domain-specialized and multi-agent LLMs could be achieved by utilizing domain-specific or agent-based frameworks to analyze churn determinants and support customer-facing personnel in real-time.

6. Conclusions

This study aimed to develop RetenNet, an integrated, interpretable, and deployable customer retention system that combines machine learning-based churn prediction, SHAP-driven explainability, and budget-aware prescriptive optimization. RetenNet utilizes LLM to convert XAI outputs into practical, non-technical recommendations, facilitating strategic decision-making in the telecommunications sector. The solution employs predictive analytics to identify customers susceptible to churn precisely. SHAP-based explanations improve transparency by identifying the primary factors influencing customer churn, ensuring interpretability for business users. These insights are subsequently implemented using prescriptive optimization methods, including fuzzy rule-based clustering and linear integer programming, to propose personalized retention strategies, such as discount allocation, all within specified budget limitations. RetenNet effectively bridges the divide between complex analytics and practical business requirements by integrating prediction, explanation, and optimization within a unified framework and turning analytical results into intuitive recommendations through LLMs. This research provides a useful, scalable, and interpretable method for customer retention, especially applicable to data-intensive sectors such as telecommunications.

Author Contributions: Conceptualization: A.P. and R.R.; Methodology: A.P.; Software: A.P.; Validation: R.R., M.M. and A.P.; Formal analysis: M.M.; Investigation: A.P.; Resources: R.R.; Data curation: A.P.; Writing—original draft preparation: A.P.; Writing—review and editing: M.M.; Visualization: R.R.; Supervision: M.M.; Project administration: A.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available https://www.kaggle.com/datasets/blastchar/telco-customer-churn; Code available https://github.com/DataInsighty/Customer-Churn-Retention-Project.git; RetenNet app available https://customer-churn-retention.streamlit.app/.

Acknowledgments: No AI tools were used to prepare this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Y. Liu, M. Shengdong, G. Jijian, and N. Nedjah, "Intelligent Prediction of Customer Churn with a Fused Attentional Deep Learning Model," *Mathematics*, vol. 10, no. 24, p. 4733, Dec. 2022, doi: 10.3390/math10244733.

[2] 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, doi: 10.1109/ACCESS.2023.3257352.

[3] C.-P. Wei and I.-T. Chiu, "Turning telecommunications call details to churn prediction: a data mining approach," *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, Aug. 2002, doi: 10.1016/S0957-4174(02)00030-1.

[4] S. Saleh and S. Saha, "Customer retention and churn prediction in the telecommunication industry: a case study on a Danish university," *SN Appl. Sci.*, vol. 5, no. 7, p. 173, Jul. 2023, doi: 10.1007/s42452-023-05389-6.

[5] Somya, S. Kaushik, S. Saini, S. Singhal, and S. Kumar, "Telecom Churn Prediction Using Data Science," *Educ. Adm. Theory Pract.*, vol. 30, no. 5, pp. 11026–11034, 2024, doi: 10.53555/kuey.v30i5.4880.

[6] S. Saha, C. Saha, M. M. Haque, M. G. R. Alam, and A. Talukder, "ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry," *IEEE Access*, vol. 12, pp. 4471–4484, 2024, doi: 10.1109/ACCESS.2024.3349950.

[7] B. Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *Int. J. Intell. Networks*, vol. 4, pp. 145–154, 2023, doi: 10.1016/j.ijin.2023.05.005.

[8] M. R. Mohaimin *et al.*, "Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market," *J. Comput. Sci. Technol. Stud.*, vol. 7, no. 1, pp. 30–45, Jan. 2025, doi: 10.32996/jcsts.2025.7.1.3.

[9] A. Ben, "Enhanced Churn Prediction in the Telecommunication Industry," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3577712.

[10] R. Srinivasan, D. Rajeswari, and G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches," in *2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, Jan. 2023, pp. 1–6. doi: 10.1109/ICECONF57129.2023.10083813.

[11] T. J. Shen, A. Samad, and B. Shibghatullah, "Customer Churn Prediction Model for Telecommunication Industry," *J. Adv. Artif. Life Robot.* , vol. 3, no. 2, pp. 85–91, 2022, [Online]. Available: https://alife-robotics.org/jallr.html

[12] S. Höppner, E. Stripling, B. Baesens, S. vanden Broucke, and T. Verdonck, "Profit driven decision trees for churn prediction," *Eur. J. Oper. Res.*, vol. 284, no. 3, pp. 920–933, Aug. 2020, doi: 10.1016/j.ejor.2018.11.072.

[13] G. Sam, P. Asuquo, and B. Stephen, "Customer Churn Prediction using Machine Learning Models," *J. Eng. Res. Reports*, vol. 26, no. 2, pp. 181–193, Feb. 2024, doi: 10.9734/jerr/2024/v26i21081.

[14] Y. Rong *et al.*, "Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 4, pp. 2104–2122, Apr. 2024, doi: 10.1109/TPAMI.2023.3331846.

[15] T. Xu, Y. Ma, C. Ao, M. Qu, and X. Meng, "A Novel Telecom Customer Churn Analysis System Based on RFM Model and Feature Importance Ranking," *Interdiscip. J. Information, Knowledge, Manag.*, vol. 18, pp. 719–737, 2023, doi: 10.28945/5192.

[16] V. Chang, K. Hall, Q. Xu, F. Amao, M. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," *Algorithms*, vol. 17, no. 6, p. 231, May 2024, doi: 10.3390/a17060231.

[17] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Mach. Learn. with Appl.*, vol. 17, p. 100567, Sep. 2024, doi: 10.1016/j.mlwa.2024.100567.

[18] D. M. Melian, A. Dumitrache, S. Stancu, and A. Nastu, "Customer Churn Prediction in Telecommunication Industry. A Data Analysis Techniques Approach," *Postmod. Openings*, vol. 13, no. 1 Sup1, pp. 78–104, Mar. 2022, doi: 10.18662/po/13.1Sup1/415.

[19] C. Özkurt, "Transparency in Decision-Making: The Role of Explainable AI (XAI) in Customer Churn Analysis," *Inf. Technol. Econ. Bus.*, Jan. 2025, doi: 10.69882/adba.iteb.2025011.

[20] E. K. P. Chong and S. H. Żak, *An Introduction to Optimization*. Wiley, 2008. doi: 10.1002/9781118033340.

[21] C. Qin, A. Zhang, Z. Zhang, J. Chen, M. Yasunaga, and D. Yang, "Is ChatGPT a General-Purpose Natural Language Processing Task Solver?," *ArXiv*. Feb. 08, 2023. [Online]. Available: http://arxiv.org/abs/2302.06476

[22] H. Naveed *et al.*, "A Comprehensive Overview of Large Language Models," *ArXiv*. Jul. 12, 2023. [Online]. Available: http://arxiv.org/abs/2307.06435

[23] N. N. Y., T. Van Ly, and D. V. T. Son, "Churn prediction in telecommunication industry using kernel Support Vector Machines," *PLoS One*, vol. 17, no. 5, p. e0267935, May 2022, doi: 10.1371/journal.pone.0267935.

[24] Y. Fareniuk, T. Zatonatska, O. Dluhopolskyi, and O. Kovalenko, "Customer churn prediction model: a case of the telecommunication market," *Economics*, vol. 10, no. 2, pp. 109–130, Dec. 2022, doi: 10.2478/eoik-2022-0021.

[25] N. Mustafa, L. Sook Ling, and S. F. Abdul Razak, "Customer churn prediction for telecommunication industry: A Malaysian Case Study," *F1000Research*, vol. 10, p. 1274, Dec. 2021, doi: 10.12688/f1000research.73597.1.

[26] N. I. Mohammad, S. A. Ismail, M. N. Kama, O. M. Yusop, and A. Azmi, "Customer Churn Prediction In Telecommunication Industry Using Machine Learning Classifiers," in *Proceedings of the 3rd International Conference on Vision, Image and Signal Processing*, Aug. 2019, pp. 1–7. doi: 10.1145/3387168.3387219.

[27] A. Amin *et al.*, "Just-in-time customer churn prediction in the telecommunication sector," *J. Supercomput.*, vol. 76, no. 6, pp. 3924–3948, Jun. 2020, doi: 10.1007/s11227-017-2149-9.

[28] A. Amin *et al.*, "Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods," *Int. J. Inf. Manage.*, vol. 46, pp. 304–319, Jun. 2019, doi: 10.1016/j.ijinfomgt.2018.08.015.

[29] S. W. Fujo, S. Subramanian, and M. A. Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," *Inf. Sci. Lett.*, vol. 11, no. 1, pp. 185–198, Jan. 2022, doi: 10.18576/isl/110120.

[30] R. E. Ako *et al.*, "Effects of Data Resampling on Predicting Customer Churn via a Comparative Tree-based Random Forest and XGBoost," *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 86–101, Jun. 2024, doi: 10.62411/jcta.10562.

[31] M. D. Gabhane, A. Suriya, and S. B. Kishor, "Churn Prediction in Telecommunication Business using CNN and ANN," *J. Posit. Sch. Psychol.*, vol. 2022, no. 4, pp. 4672–4680, 2021, [Online]. Available: http://journalppw.com

[32] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi, “Improved churn prediction in telecommunication industry using data mining techniques,” *Appl. Soft Comput.*, vol. 24, pp. 994–1012, Nov. 2014, doi: 10.1016/j.asoc.2014.08.041.

[33] H. Ribeiro, B. Barbosa, A. C. Moreira, and R. G. Rodrigues, “Determinants of churn in telecommunication services: a systematic literature review,” *Manag. Rev. Q.*, vol. 74, no. 3, pp. 1327–1364, Sep. 2024, doi: 10.1007/s11301-023-00335-7.

[34] N. Almufadi and A. Mustafa Qamar, “Deep Convolutional Neural Network Based Churn Prediction for Telecommunication Industry,” *Comput. Syst. Sci. Eng.*, vol. 43, no. 3, pp. 1255–1270, 2022, doi: 10.32604/csse.2022.025029.

[35] C. M. Bishop and N. M. Nasrabadi, “Pattern Recognition and Machine Learning,” *J. Electron. Imaging*, vol. 16, no. 4, p. 049901, Jan. 2007, doi: 10.1117/1.2819119.

[36] L. Deng, “A tutorial survey of architectures, algorithms, and applications for deep learning,” *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. 1, 2014, doi: 10.1017/atsip.2013.9.

[37] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine Learning*. CRC Press, 2016. doi: 10.1201/9781315371658.

[38] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[39] M. Badawy, N. Ramadan, and H. A. Hefny, “Healthcare predictive analytics using machine learning and deep learning techniques: a survey,” *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, p. 40, Aug. 2023, doi: 10.1186/s43067-023-00108-y.

[40] A. Manzoor, M. Atif Qureshi, E. Kidney, and L. Longo, “A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners,” *IEEE Access*, vol. 12, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.

[41] I. H. Sarker, “Deep Cybersecurity: A Comprehensive Overview from Neural Network and Deep Learning Perspective,” *SN Comput. Sci.*, vol. 2, no. 3, p. 154, May 2021, doi: 10.1007/s42979-021-00535-6.

[42] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, doi: 10.1289/EHP4713.

[43] I. H. Sarker, “A machine learning based robust prediction model for real-life mobile phone data,” *Internet of Things*, vol. 5, pp. 180–193, Mar. 2019, doi: 10.1016/j.iot.2019.01.007.

[44] Q. A. Mumuni, O. M. Olaniyan, O. Ipinnimo, L. A. Akinyemi, and A. I. Olayiwola-Mumuni, “Intelligent PID Gain Selection via Supervised Machine Learning Approach for Decoupled MIMO Control Systems,” *J. Futur. Artif. Intell. Technol.*, vol. 2, no. 1, pp. 163–181, Jun. 2025, doi: 10.62411/faith.3048-3719-92.

[45] I. H. Sarker, A. S. M. Kayes, and P. Watters, “Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage,” *J. Big Data*, vol. 6, no. 1, p. 57, Dec. 2019, doi: 10.1186/s40537-019-0219-y.

[46] S. Zhao, B. Zhang, J. Yang, J. Zhou, and Y. Xu, “Linear discriminant analysis,” *Nat. Rev. Methods Prim.*, vol. 4, no. 1, p. 70, Sep. 2024, doi: 10.1038/s43586-024-00346-y.

[47] S. Le Cessie and J. C. Van Houwelingen, “Ridge Estimators in Logistic Regression,” *Appl. Stat.*, vol. 41, no. 1, p. 191, 1992, doi: 10.2307/2347628.

[48] C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, “An up-to-date comparison of state-of-the-art classification algorithms,” *Expert Syst. Appl.*, vol. 82, pp. 128–150, Oct. 2017, doi: 10.1016/j.eswa.2017.04.003.

[49] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986, doi: 10.1007/BF00116251.

[50] S. L. Salzberg, “C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993,” *Mach. Learn.*, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/BF00993309.

[51] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification And Regression Trees*. Routledge, 2017. doi: 10.1201/9781315139470.

[52] I. H. Sarker, Y. B. Abushark, and A. I. Khan, “ContextPCA: Predicting Context-Aware Smartphone Apps Usage Based On Machine Learning Techniques,” *Symmetry (Basel).*, vol. 12, no. 4, p. 499, Apr. 2020, doi: 10.3390/sym12040499.

[53] I. H. Sarker, A. Colman, J. Han, A. I. Khan, Y. B. Abushark, and K. Salah, “BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model,” *Mob. Networks Appl.*, vol. 25, no. 3, pp. 1151–1161, Jun. 2020, doi: 10.1007/s11036-019-01443-z.

[54] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991, doi: 10.1007/BF00153759.

[55] F. Sanei, A. Harifi, and S. Golzari, “Improving the precision of KNN classifier using nonlinear weighting method based on the spline interpolation,” in *2017 7th International Conference on Computer and Knowledge Engineering (ICCKE)*, Oct. 2017, pp. 289–292. doi: 10.1109/ICCKE.2017.8167893.

[56] K. Muhamada *et al.*, “Exploring Machine Learning and Deep Learning Techniques for Occluded Face Recognition : A Comprehensive Survey and Comparative Analysis,” *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 2, pp. 160–173, Sep. 2024, doi: 10.62411/faith.2024-30.

[57] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to Platt’s SMO Algorithm for SVM Classifier Design,” *Neural Comput.*, vol. 13, no. 3, pp. 637–649, Mar. 2001, doi: 10.1162/089976601300014493.

[58] Y. Amit and D. Geman, “Shape Quantization and Recognition with Randomized Trees,” *Neural Comput.*, vol. 9, no. 7, pp. 1545–1588, Oct. 1997, doi: 10.1162/neco.1997.9.7.1545.

[59] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[60] Y. Freund and R. E. Schapire, “Experiments with a New Boosting Algorithm,” in *Proceedings of the 13th International Conference on Machine Learning*, 1996, pp. 148–156. doi: 10.1.1.133.1040.

[61] A. Sikri, R. Jameel, S. M. Idrees, and H. Kaur, “Enhancing customer retention in telecom industry with machine learning driven churn prediction,” *Sci. Rep.*, vol. 14, no. 1, p. 13097, Jun. 2024, doi: 10.1038/s41598-024-63750-0.

[62] K. Suzuki, *Artificial Neural Networks - Methodological Advances and Biomedical Applications*. InTech, 2011. doi: 10.5772/644.

[63] J. Singh and R. Banerjee, "A Study on Single and Multi-layer Perceptron Neural Network," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Mar. 2019, pp. 35–40. doi: 10.1109/ICCMC.2019.8819775.
[64] M. U. Tariq, M. Babar, M. Poulin, and A. S. Khattak, "Distributed model for customer churn prediction using convolutional neural network," *J. Model. Manag.*, vol. 17, no. 3, pp. 853–863, Aug. 2022, doi: 10.1108/JM2-01-2021-0032.
[65] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. London, England: MIT Press, 2016.
[66] S. H. Iranmanesh, M. Hamid, M. Bastan, G. Hamed Shakouri, and M. M. Nasiri, "Customer churn prediction using artificial neural network: An analytical CRM application," *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, no. July, pp. 2214–2226, 2019.
[67] S. Bakhvalov, E. Osadchy, I. Bogdanova, R. Shichiyakh, and E. L. Lydia, "Intelligent System for Customer Churn Prediction using Dipper Throat Optimization with Deep Learning on Telecom Industries," *Fusion Pract. Appl.*, vol. 14, no. 2, pp. 172–185, 2024, doi: 10.54216/FPA.140214.
[68] D. Frazzetto, T. D. Nielsen, T. B. Pedersen, and L. Šikšnys, "Prescriptive analytics: a survey of emerging trends and technologies," *VLDB J.*, vol. 28, no. 4, pp. 575–595, Aug. 2019, doi: 10.1007/s00778-019-00539-y.
[69] M. Ewert, D. P. Brodowicz, and P. Pospieszny, "Artificial Intelligence and machines : A curse or blessing for corporate real estate?," *Corp. Real Estate J.*, vol. 7, no. 4, p. 337, Jun. 2018, doi: 10.69554/KHWS4814.
[70] Y. Engel and O. Etzion, "Towards proactive event-driven computing," in *Proceedings of the 5th ACM international conference on Distributed event-based system*, Jul. 2011, pp. 125–136. doi: 10.1145/2002259.2002279.
[71] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, "Prescriptive analytics: Literature review and research challenges," *Int. J. Inf. Manage.*, vol. 50, pp. 57–70, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.04.003.
[72] J. Krumeich, D. Werth, and P. Loos, "Prescriptive Control of Business Processes," *Bus. Inf. Syst. Eng.*, vol. 58, no. 4, pp. 261–280, Aug. 2016, doi: 10.1007/s12599-015-0412-2.
[73] E. G. Mansoori, "FRBC: A Fuzzy Rule-Based Clustering Algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 960–971, Oct. 2011, doi: 10.1109/TFUZZ.2011.2158651.
[74] A. Achenbach and S. Spinler, "Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights," *Oper. Res. Perspect.*, vol. 5, pp. 265–279, 2018, doi: 10.1016/j.orp.2018.08.004.
[75] G. Sierksma, G. Sierksma, and Y. Zwols, *Linear and Integer Optimization*. Chapman and Hall/CRC, 2015. doi: 10.1201/b18378.
[76] A. Rai, "Explainable AI: from black box to glass box," *J. Acad. Mark. Sci.*, vol. 48, no. 1, pp. 137–141, Jan. 2020, doi: 10.1007/s11747-019-00710-5.
[77] M. Nauta *et al.*, "From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI," *ACM Comput. Surv.*, vol. 55, no. 13s, pp. 1–42, Dec. 2023, doi: 10.1145/3583558.
[78] J. Zhou, F. Chen, and A. Holzinger, "Towards Explainability for AI Fairness," in *xxAI - Beyond Explainable AI*, 2022, pp. 375–386. doi: 10.1007/978-3-031-04083-2_18.
[79] J. Dodge, Q. V. Liao, Y. Zhang, R. K. E. Bellamy, and C. Dugan, "Explaining models: an empirical study of how explanations impact fairness judgment," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Mar. 2019, pp. 275–285. doi: 10.1145/3301275.3302310.
[80] S. Lundberg, "Commentary," *Ann. Emerg. Med.*, vol. 70, no. 1, pp. 30–31, Jul. 2017, doi: 10.1016/j.annemergmed.2017.05.019.
[81] M. Fisher, M. Robb, D. M. Wolf, and J. Slade, "Implementing a Standardized Discussion Forum Rubric Across 3 Online Nursing Programs," *Nurse Educ.*, vol. 44, no. 6, pp. 291–292, Nov. 2019, doi: 10.1097/NNE.0000000000000652.
[82] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *Ann. Stat.*, vol. 29, no. 5, Oct. 2001, doi: 10.1214/aos/1013203451.
[83] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges," in *ECML PKDD 2020 Workshops*, Springer, Cham, 2020, pp. 417–431. doi: 10.1007/978-3-030-65965-3_28.
[84] D. Tchuente, J. Lonlac, and B. Kamsu-Foguem, "A methodological and theoretical framework for implementing explainable artificial intelligence (XAI) in business applications," *Comput. Ind.*, vol. 155, p. 104044, Feb. 2024, doi: 10.1016/j.compind.2023.104044.
[85] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?,'" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, vol. 13-17-Augu, pp. 1135–1144. doi: 10.1145/2939672.2939778.
[86] A. Chernyavskiy, D. Ilvovsky, and P. Nakov, "Transformers: 'The End of History' for Natural Language Processing?," in *Machine Learning and Knowledge Discovery in Databases. Research Track*, 2021, pp. 677–693. doi: 10.1007/978-3-030-86523-8_41.
[87] D. Adiwardana *et al.*, "Towards a Human-like Open-Domain Chatbot," *arXiv*. Jan. 27, 2020. doi: 10.48550/arXiv.2001.09977.
[88] A. Wang *et al.*, "SuperGLUE: a stickier benchmark for general-purpose language understanding systems," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.
[89] A. Iorliam and J. A. Ingio, "A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing," *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 311–325, Feb. 2024, doi: 10.62411/jcta.9447.
[90] S. Wang *et al.*, "ERNIE 3.0 Titan: Exploring Larger-scale Knowledge Enhanced Pre-training for Language Understanding and Generation," *ArXiv*. Dec. 23, 2021. [Online]. Available: http://arxiv.org/abs/2112.12731
[91] Y. Wang, H. Le, A. D. Gotmare, N. D. Q. Bui, J. Li, and S. C. H. Hoi, "CodeT5+: Open Code Large Language Models for Code Understanding and Generation," *ArXiv*. May 13, 2023. [Online]. Available: http://arxiv.org/abs/2305.07922
[92] A. Zeng *et al.*, "GLM-130B: An Open Bilingual Pre-trained Model," *ArXiv*. Oct. 05, 2022. [Online]. Available: http://arxiv.org/abs/2210.02414
[93] Z. Hu *et al.*, "LLM-Adapters: An Adapter Family for Parameter-Efficient Fine-Tuning of Large Language Models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 5254–5276. doi: 10.18653/v1/2023.emnlp-main.319.
[94] J. He, C. Zhou, X. Ma, T. Berg-Kirkpatrick, and G. Neubig, "Towards a Unified View of Parameter-Efficient Transfer Learning," *ArXiv*. Oct. 08, 2021. [Online]. Available: http://arxiv.org/abs/2110.04366
[95] Q. Ye *et al.*, "mPLUG-Owl2: Revolutionizing Multi-modal Large Language Model with Modality Collaboration," *ArXiv*. Nov. 07, 2023. [Online]. Available: http://arxiv.org/abs/2311.04257

[96] L. Chen *et al.*, "ShareGPT4V: Improving Large Multi-Modal Models with Better Captions," *ArXiv*. Nov. 21, 2023. [Online]. Available: http://arxiv.org/abs/2311.12793

[97] W. Dai *et al.*, "InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning," *ArXiv*. May 10, 2023. [Online]. Available: http://arxiv.org/abs/2305.06500

[98] J. Bai *et al.*, "Qwen-VL: A Versatile Vision-Language Model for Understanding, Localization, Text Reading, and Beyond," *ArXiv*. Aug. 24, 2023. [Online]. Available: http://arxiv.org/abs/2308.12966

[99] A. Q. Jiang *et al.*, "Mixtral of Experts," *ArXiv*. Jan. 08, 2024. [Online]. Available: http://arxiv.org/abs/2401.04088

[100] H. Touvron *et al.*, "LLaMA: Open and Efficient Foundation Language Models," *ArXiv*. Feb. 27, 2023. [Online]. Available: http://arxiv.org/abs/2302.13971

[101] D. Zhang *et al.*, "MM-LLMs: Recent Advances in MultiModal Large Language Models," *ArXiv*. Jan. 24, 2024. [Online]. Available: http://arxiv.org/abs/2401.13601

[102] H. Liu, C. Li, Q. Wu, and Y. J. Lee, "Visual Instruction Tuning," *ArXiv*, Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.08485

[103] Blastchar, "Telco Customer Churn," *Kaggle.com*, 2018. https://www.kaggle.com/datasets/blastchar/telco-customer-churn (accessed May 15, 2025).

[104] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, Jan. 2014, doi: 10.1007/s00521-013-1368-0.

[105] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

[106] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley, 2013. doi: 10.1002/9781118548387.

[107] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[108] J. Wainer and G. Cawley, "Nested cross-validation when selecting classifiers is overzealous for most practical applications," *Expert Syst. Appl.*, vol. 182, p. 115222, Nov. 2021, doi: 10.1016/j.eswa.2021.115222.

[109] A. Kalyakulina, I. Yusipov, A. Moskalev, C. Franceschi, and M. Ivanchenko, "eXplainable Artificial Intelligence (XAI) in aging clock models," *Ageing Res. Rev.*, vol. 93, p. 102144, Jan. 2024, doi: 10.1016/j.arr.2023.102144.

[110] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nat. Mach. Intell.*, vol. 2, no. 1, pp. 56–67, Jan. 2020, doi: 10.1038/s42256-019-0138-9.

[111] M. Saarela and L. Geogieva, "Robustness, Stability, and Fidelity of Explanations for a Deep Skin Cancer Classification Model," *Appl. Sci.*, vol. 12, no. 19, p. 9545, Sep. 2022, doi: 10.3390/app12199545.

[112] M. Saarela and V. Podgorelec, "Recent Applications of Explainable AI (XAI): A Systematic Literature Review," *Appl. Sci.*, vol. 14, no. 19, p. 8884, Oct. 2024, doi: 10.3390/app14198884.

[113] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. LeanPub, 2020. [Online]. Available: https://christophm.github.io/interpretable-ml-book/

[114] A. V. Ponce‐Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, "Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development," *Clin. Transl. Sci.*, vol. 17, no. 11, Nov. 2024, doi: 10.1111/cts.70056.

[115] Gemini Team *et al.*, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," *ArXiv*, Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.05530

[116] Google AI for Developers, "Gemini models." https://ai.google.dev/gemini-api/docs/models#gemini-2.5-pro-exp-03-25 (accessed Mar. 27, 2025).

[117] K. Genova and V. Guliashki, "Linear integer programming methods and approaches - A survey," *Cybern. Inf. Technol.*, vol. 11, no. 1, pp. 3–25, 2011, [Online]. Available: https://cit.iict.bas.bg/CIT_2011/v11-1/3-25.pdf