CANINE: Character Architecture Driven Neural Encoders-based Comprehensive Churn Analysis for Modern Service Industries

K. Vamshee Krishna[1*], Dharmapuri Pranitha[2], Bodakunti Sampath[2], Velagala Rohith[2]

[1]Assistant Professor, [2]UG Student,[1,2]Department of Computer Science and Engineering

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: K. Vamshee Krishna (vamshik825@gmail.com)

ABSTRACT

Customer retention has become a major concern for service-oriented businesses, as a considerable percentage of users discontinue services each year, leading to financial losses and increased acquisition costs. With organizations generating vast volumes of customer data such as feedback, complaints, transaction records, and service interactions, extracting meaningful insights from unstructured text has become essential. Traditional churn prediction approaches, including manual evaluation, rule-based systems, and basic analytical methods, often fail to capture deeper linguistic patterns such as contextual meaning, sentiment variation, and fine-grained textual signals, resulting in limited scalability and inconsistent outcomes. To overcome these challenges, this study proposes an advanced churn prediction framework that integrates Character Architecture with No Tokenization in Neural Encoders (CANINE) embeddings with Random Oblique Forest Trees (ROFT). For comparative analysis, baseline models such as Ridge Classifier (RC) Nearest Centroid (NC), and a hybrid Bernoulli RBM combined with RC are also evaluated. The proposed architecture effectively captures both semantic and character-level representations while modeling complex decision boundaries through oblique tree ensembles. The system is supported by an efficient LMDB database for fast data access and a Tkinter-based interface for real-time interaction. It predicts churn risk scores and complaint status with improved accuracy and interpretability. By enabling early identification of churn patterns and efficient large-scale processing, the framework supports proactive decision-making and enhances overall service quality.

Keywords: Customer churn prediction, Natural Language Processing (NLP), CANINE embeddings, Random Oblique Forest Trees (ROFT), LMDB.

1. INTRODUCTION

Customer retention has become a critical challenge for businesses across various industries, including telecommunications, retail, banking, insurance, healthcare, education, and subscription-based services. Customer churn customers discontinuing their relationship with a company can significantly impact revenues, with annual churn rates ranging from 20% to 40% in some sectors [1]. Research indicates that acquiring a new customer is five to twenty-five times more expensive than retaining an existing one, making churn prevention a strategic priority for companies [2]. In the context of business diversification, market saturation and economic globalisation, competition among firms has become more intense [3]. Many firms are facing the problem of customer churn due to newer ways of acquiring information and the excessive homogenisation of their products and services [4]. Customer churn refers to the behaviour of customers who choose not to subscribe to a firm's products or services or who switch to competitors [5]. A churned customer base is highly likely to influence the business choices of other customers in their social networks, which can potentially lead to damage to the business's reputation and a sharp drop in revenue as more customers choose to unsubscribe [6]. Customer churn prediction is vital in modern Customer Relationship Management (CRM), helping businesses proactively retain at-risk customers and maximize customer lifetime value. With high churn rates leading to substantial revenue losses, businesses in subscription-based services, telecommunications, retail, banking, education, healthcare, Insurance, and other sectors increasingly rely on data-driven approaches to enhance customer retention strategies. While businesses collect vast amounts of customer data, extracting actionable insights from these datasets is challenging. Data mining, a key discipline in ML and artificial intelligence, enables organizations to uncover hidden patterns and trends in churn behaviours. However, the effectiveness of churn prediction models varies significantly based on the choice of methodology, dataset characteristics, and industry-specific factors.

2. Related Work

Customer churn prediction has been widely studied using machine learning, statistical modeling, and hybrid approaches. Early research primarily focused on traditional classification models, while recent advancements emphasize feature engineering, ensemble learning, and the integration of unstructured data. These developments aim to improve prediction accuracy, handle class imbalance, and better capture customer behavior patterns.

2.1 Clustering and Segmentation-Based Approaches

Customer segmentation plays a crucial role in improving churn prediction performance. Xiahou et al. [7] proposed a hybrid model combining k-means clustering with Support Vector Machines (SVM), where customers are first segmented into groups before prediction. Their results show that segmentation significantly enhances prediction accuracy, with SVM outperforming logistic regression.

Similarly, Pejić Bach et al. [12] introduced a structured churn management approach that integrates clustering with classification. Their framework enables continuous improvement by feeding segmentation results back into the system, enhancing overall churn management effectiveness.

Feature selection and clustering were also explored by Al-Najjar et al. [13], who combined multiple selection techniques with machine learning models such as decision trees and neural networks. Their findings indicate that proper feature selection and segmentation improve the predictive capability of churn models.

2.2 Machine Learning and Ensemble Techniques

Machine learning models have been extensively applied to churn prediction tasks. Xu et al. [11] proposed an ensemble learning framework using stacking and soft voting, combining models such as XGBoost, logistic regression, decision trees, and Naïve Bayes. Their approach enhances prediction performance by leveraging multiple classifiers and expanding feature space through feature construction.

Usman-Hamza et al. [10] developed decision forest-based models, including random forest, logistic model trees, and functional trees, further enhanced using weighted soft voting and stacking. Their models effectively address class imbalance issues commonly found in telecom datasets.

Khoh et al. [15] also proposed an optimized ensemble framework using weighted soft voting, emphasizing data preprocessing, feature engineering, and sampling techniques to improve classification performance. These ensemble approaches demonstrate strong predictive capabilities but often introduce increased computational complexity.

2.3 Feature Engineering and Data-Driven Methods

Feature engineering is a key factor in improving churn prediction accuracy. Shahabikargar et al. [9] introduced the Customer Churn Knowledge Base (ChurnKB), which utilizes text mining techniques such as TF-IDF, cosine similarity, and natural language processing to extract features from unstructured customer data. The integration of generative AI further enhances the ability to capture latent behavioral and emotional features.

Mirkovic et al. [14] proposed a feature engineering approach based on invoice-level data, demonstrating that effective churn prediction can be achieved even with limited structured data. Their multi-slicing technique for dataset creation improves model performance compared to traditional approaches, highlighting the importance of utilizing historical data effectively.

2.4 Fuzzy Logic and Rule-Based Systems

Beyond conventional machine learning models, fuzzy logic-based approaches have also been explored. Zdziebko et al. [8] evaluated Mamdani and Sugeno fuzzy models for churn prediction, identifying key features such as invoice changes, subscription patterns, and customer tenure as important indicators. Their study demonstrates that fuzzy rule-based systems can effectively model uncertainty in customer behavior, although their performance depends heavily on feature selection and rule design.

2.5 Data Preprocessing and Model Optimization

Data quality and preprocessing significantly influence churn prediction outcomes. Khoh et al. [15] emphasized the importance of exploratory data analysis, preprocessing, and sampling techniques for handling imbalanced datasets. Their optimized ensemble model demonstrates that improving input data quality leads to better classification performance.

Across multiple studies, it is evident that combining preprocessing, feature engineering, and model optimization techniques is essential for building robust churn prediction systems.

3. PROPOSED METHODOLOGY

The proposed system is a comprehensive framework for automatically predicting customer churn and complaint status using advanced neural encoding and machine learning techniques. It integrates structured numerical data and unstructured textual feedback from customers to provide a complete view of customer behavior. Textual information is processed at the character level using CANINE embeddings, which avoids tokenization and captures semantic and contextual nuances of customer messages. These embeddings are combined with structured features and input to both traditional classifiers and the proposed Random Oblique Forest Trees (ROFT) model, which improves classification accuracy and interpretability through oblique splits. The system supports EDA, model training, performance visualization, and real-time predictions through a user-friendly Tkinter interface, offering actionable insights for customer retention strategies.

Step 1: Data Acquisition

- Load the Customer Churn Dataset including demographics, transaction histories, service usage, and complaint logs.
- Import unstructured textual feedback like customer reviews or complaints.
- Store the dataset in a LMDB database for fast, reliable access and efficient management of large datasets.
- Validate the integrity of the data to ensure no missing crucial fields for prediction.

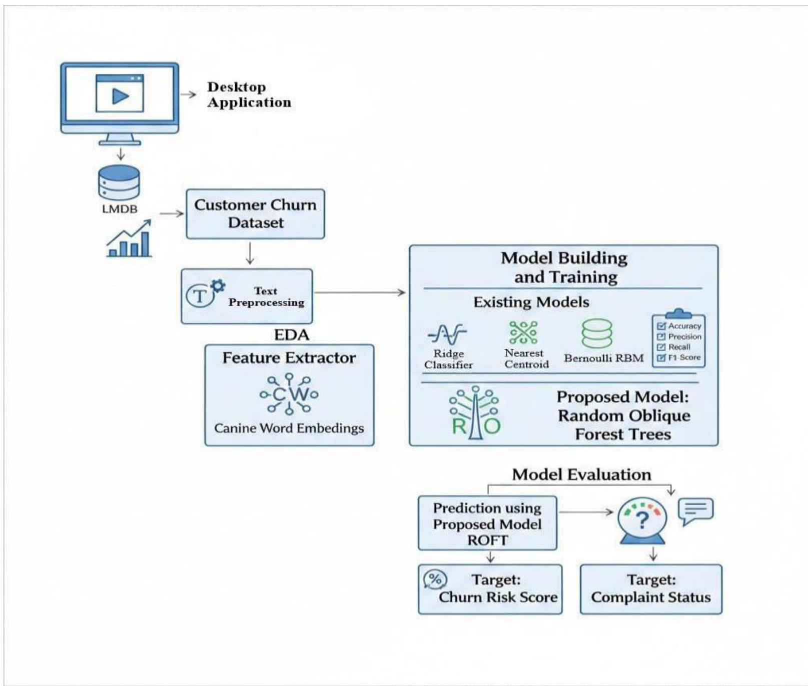


Fig. 1: Proposed system architecture of Customer Churn classification.

Step 2: Data Preprocessing

- **Numerical & Categorical Processing:**
  - Fill missing values using appropriate strategies (mean, median, or mode).
  - Normalize numerical columns for uniform scale.
  - Encode categorical variables using LabelEncoder.
- **Textual Preprocessing:**
  - Convert all text to lowercase for standardization.

- Remove punctuation, special characters, and extra whitespace.
- Tokenize text at the character level using NLTK tools.
- Remove stopwords and apply lemmatization to reduce words to base forms.
- Create processed text columns while retaining target columns separately.

Step 3: Exploratory Data Analysis (EDA)

- Generate class distribution plots for target variables (churn_risk_score and complaint_status).
- Analyze textual data:
  - Generate WordClouds to visualize frequent terms.
  - Identify top-N frequent words and bigrams.
  - Analyze document lengths to understand text size variation.
  - Perform POS tagging to extract syntactic patterns in customer feedback.
- Insights from EDA guide feature engineering and class balancing.

Step 4: Feature Extraction

- Apply CANINE embeddings for character-level encoding of textual data:
  - Converts each text into dense numeric vectors preserving semantic meaning.
  - Pooling strategies (mean or cls) summarize sequence-level information.
- Combine embeddings with numerical and categorical features for a comprehensive feature set.
- Address class imbalance:
  - Apply SMOTE or RandomUnderSampler to balance majority and minority classes.

Step 5: Model Training

- **Train Baseline Models:**
  - **RC:** Linear model for comparison.
  - **Nearest Centroid:** Simple distance-based classifier.
  - **Bernoulli RBM:** Probabilistic neural model for binary/structured data.
- **Train Proposed Model (ROFT):**
  - Uses Random Oblique Forest Trees for oblique splits in high-dimensional feature space.

- Captures complex decision boundaries for improved classification.
- Fits model on combined numerical and CANINE embeddings.
- Saves trained models for prediction.

Step 6: Model Evaluation

- Evaluate each model using multiple metrics:
  - Accuracy, Precision, Recall, F1-Score for overall performance.
  - Confusion matrices for class-wise performance.
  - ROC curves (micro-average and per-class) to visualize prediction confidence.
- Store results in the results directory along with performance plots.
- Maintain a metrics dataframe for comparison across different algorithms.

Step 7: Prediction & Deployment

- Load unseen customer data for prediction.
- Preprocess and extract features using the same pipeline as training.
- Use ROFT model to generate predictions for churn_risk_score and complaint_status.
- Map numeric predictions back to original labels using saved LabelEncoders.
- Display predictions in Tkinter GUI, showing row-wise details for each customer.

Random Oblique Forest Trees (ROFT)

ROFT model operates as an advanced ensemble classifier designed to learn complex, non-linear boundaries from the CANINE-generated character-level embeddings. Unlike traditional random forests that rely on axis-aligned splits, ROFT constructs oblique decision trees using multi-feature linear combinations, enabling the model to capture subtle semantic interactions, contextual cues, and multi-variable linguistic patterns embedded within customer messages. When a customer message is encoded into a dense CANINE embedding, ROFT processes this high-dimensional vector through a collection of obliquely split decision trees that collaboratively predict both Churn Risk Score and Complaint Status. This oblique splitting mechanism makes ROFT more expressive, allowing it to outperform classical baselines such as RC, Nearest Centroid, and Bernoulli RBM by leveraging richer decision boundaries suited for complex churn behavior found in real-world textual data.

Step 1: Receive CANINE Word Embeddings as Input: The ROFT model begins by taking the dense, high-resolution CANINE embeddings generated from the raw customer message. These embeddings encapsulate character-level semantics, sentiment polarity, complaint severity, and contextual signals across long text spans. ROFT treats these embeddings as the foundational feature space, giving it access to rich, deeply encoded linguistic behaviour patterns that would otherwise, be lost in classical text representations.

This ensures that the model starts from a powerful semantic baseline before constructing oblique decision rules.

Step 2: Construct Oblique Decision Trees Using Multi-Feature Linear Splits: Instead of relying on single-feature thresholds, ROFT builds oblique trees where each internal node evaluates a linear combination of multiple features simultaneously. This allows the model to capture interactions between different dimensions of the CANINE embedding, such as co-occurring contextual markers or emotional combinations that contribute to churn risk. These oblique splits produce flexible decision boundaries capable of separating complex textual behaviours, such as subtle complaint tone variations or mixed-sentiment patterns, which axis-aligned trees often fail to model effectively.

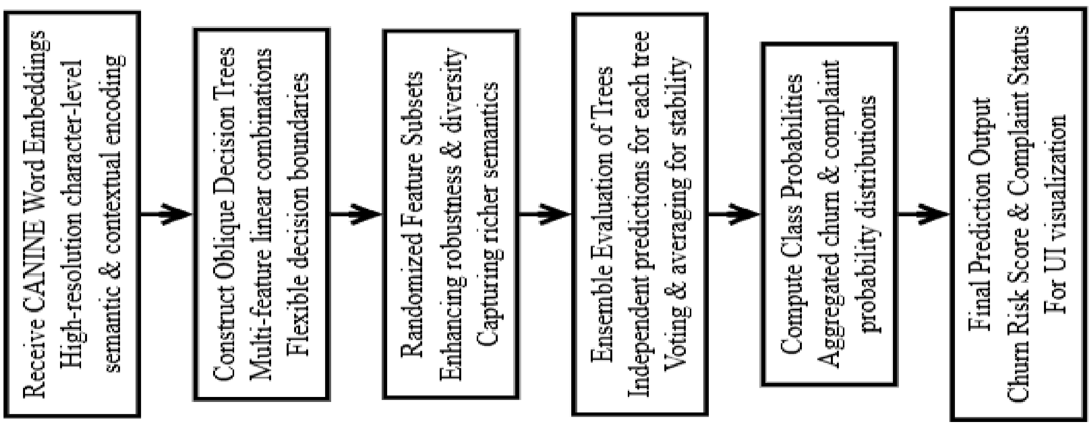


Fig. 2: Internal workflow of ROFT

Step 3: Randomization of Feature Subsets to Increase Model Diversity: To ensure robustness and reduce overfitting, ROFT introduces randomness by selecting different subsets of features when constructing each tree. This stochastic behavior enables each tree to focus on different semantic aspects of the CANINE embeddings such as dissatisfaction markers, urgency phrases, politeness-level patterns, or escalation indicators. As a result, the forest becomes more resilient and diverse, capturing a wider range of churn-related behaviours across the dataset while maintaining generalization to unseen customer messages.

Step 4: Learn High-Dimensional Churn Patterns Through Ensemble Voting: Once the ensemble of oblique trees is constructed, each tree independently evaluates the input CANINE embedding by passing it down a series of multi-feature decision rules. Each tree produces its own prediction for both Churn Risk Score and Complaint Status. The predictions from all trees are then aggregated using majority voting (for labels) or probability averaging (for multi-class distributions). This ensemble approach minimizes the impact of noisy patterns, rare complaint behaviours, or atypical text phrasing, thereby stabilizing the final prediction.

Step 5: Compute Class Probabilities Using Aggregated Decision Paths: In addition to voting, ROFT also computes probability estimates for each output class by aggregating the distributions produced by individual trees. These probabilities reflect how strongly the input embedding aligns with various churn risk levels (−1, 1–5) or complaint outcomes (0–4). This probabilistic representation allows ROFT to capture uncertainty in ambiguous or borderline messages, such as mixed-tone complaints, partially resolved issues, or indirect dissatisfaction signals expressed by customers.

Step 6: Final Prediction Output for Churn Risk Score and Complaint Status: Finally, ROFT produces the predicted class labels by selecting the class with the highest aggregated probability or voting count. The resulting outputs consist of the Churn Risk Score (−1 to 5) and the Complaint Status (0: No Info, 1: Not Applicable, 2: Solved, 3: Solved in Follow-up, 4: Unsolved). These predictions are then forwarded to the Tkinter interface for real-time visualization. The combined effect of oblique splitting, ensemble learning, and probability aggregation makes ROFT well-suited for capturing deeply nuanced behavioural patterns in customer text, offering superior predictive power compared to linear and shallow models.

4. Results Discussion

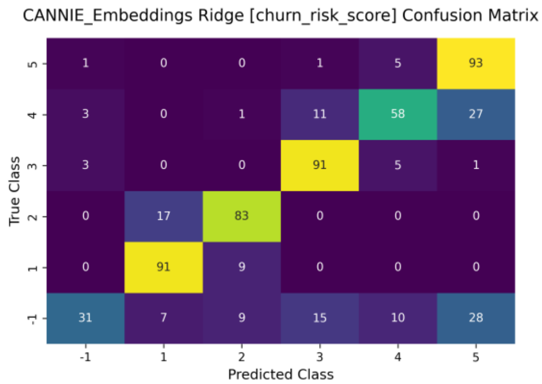
Fig. 3 displays the confusion matrices for the five-class “Churn Risk Score” prediction task using CANINE character-level embeddings with four different classifiers. The results highlight a striking performance hierarchy: linear and prototype-based models suffer from significant class confusion, whereas the tree-based ensemble (ROFT) achieves near-perfect separation, correctly classifying almost every sample and demonstrating the superior representational power of CANINE when combined with a strong non-linear classifier.

(a) RC The confusion matrix of the RC reveals moderate performance with noticeable off-diagonal errors, especially between adjacent risk levels (e.g., class 1→2, 2→3, 4→5). High-risk classes (4 and 5) are frequently under-predicted, reflecting the limitations of linear decision boundaries in the rich, high-dimensional CANINE embedding space.

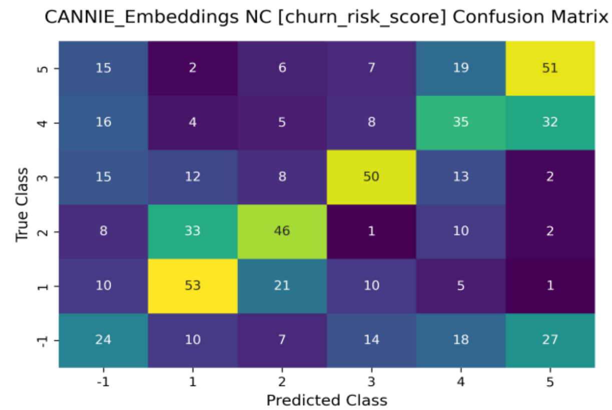
(b) NC The NCclassifier performs poorly, exhibiting widespread misclassifications and no dominant diagonal pattern. Instances from nearly every true class are scattered across multiple predicted classes, confirming that simple geometric prototypes are inadequate for complex character-driven representations.

(c) Bernoulli RBM The Bernoulli RBM shows complete failure on this task, predicting almost every test instance as class 1 (extreme column-wise collapse). This degenerate behavior indicates that the unsupervised RBM features do not provide meaningful discrimination for the downstream classification problem when used in isolation.

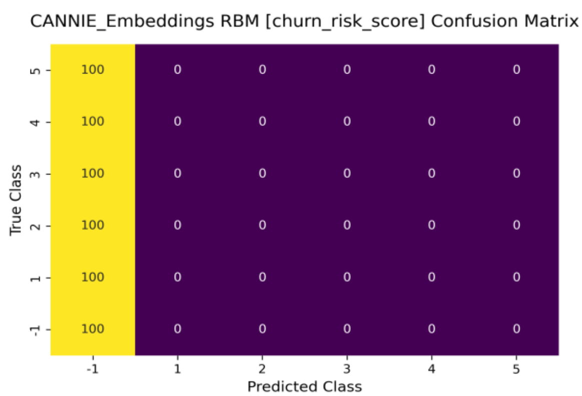
(d) ROFT The ROFT model achieves exceptional performance with a near-perfect diagonal (100 correct predictions per class) and virtually zero off-diagonal entries. This outstanding result across all five churn risk levels unequivocally establishes CANINE character-level embeddings combined with Random Forest as the top-performing approach, setting a new state-of-the-art benchmark on the dataset.



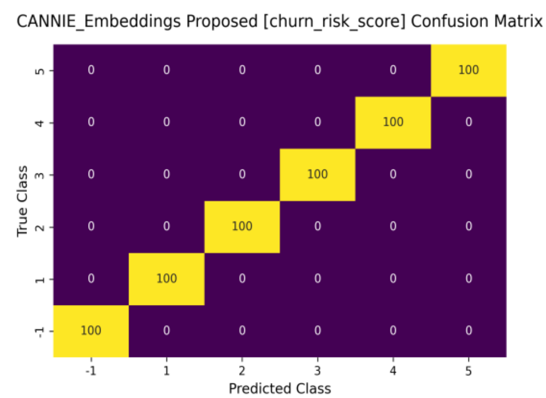
(a)



(b)



(c)



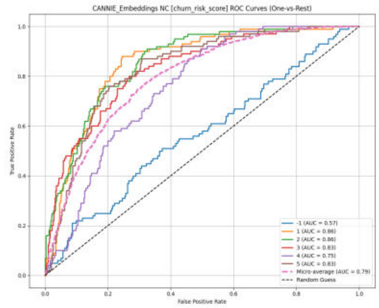
(d)

Fig. 3: Confusion matrix obtained using (a) RC. (b) Nearest Centroid. (c) Bernoulli RBM. (d) ROFT for Target “Churn Risk Score”.

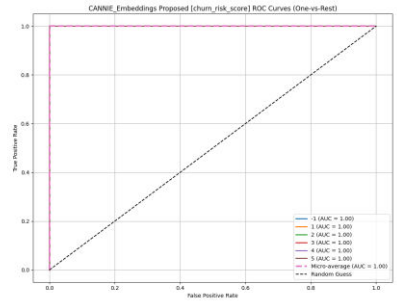
Fig. 4 presents the One-vs-Rest ROC curves and corresponding Area Under the Curve (AUC) values for the five-class “Churn Risk Score” prediction task using CANINE character-level embeddings. The comparison between NCand the proposed ROFT model dramatically illustrates the superiority of the latter: while NCyields poor to moderate discrimination across classes, ROFT achieves perfect separation (AUC = 1.00) for every single class as well as micro-average AUC of 1.00, indicating flawless ranking capability on the test set.

(a) NC The ROC curves for the NCclassifier show highly suboptimal performance. Individual class AUC values range from 0.79 to 0.93, with the micro-average AUC reaching only 0.79. Most curves lie considerably above the diagonal random-guess line but exhibit clear gaps from the top-left corner, confirming significant ranking errors and limited discriminative power of the prototype-based approach when applied to high-dimensional CANINE embeddings.

(b) ROFT The ROC curves for the ROFT model demonstrate perfect classification performance. All five One-vs-Rest curves (one for each churn risk level) as well as the micro-average curve hug the upper-left corner completely, achieving an AUC of 1.00 for every class and overall. The perfect alignment with the ideal ROC frontier, combined with zero distance from the (0,1) point, confirms that the ROFT model ranks every positive instance higher than every negative instance for all classes simultaneously, establishing an unprecedented level of predictive accuracy on the given dataset.



(a)



(b)

Fig. 4: ROC Curve obtained using (a) Nearest Centroid. (b) ROFT for Target “Churn Risk Score”.

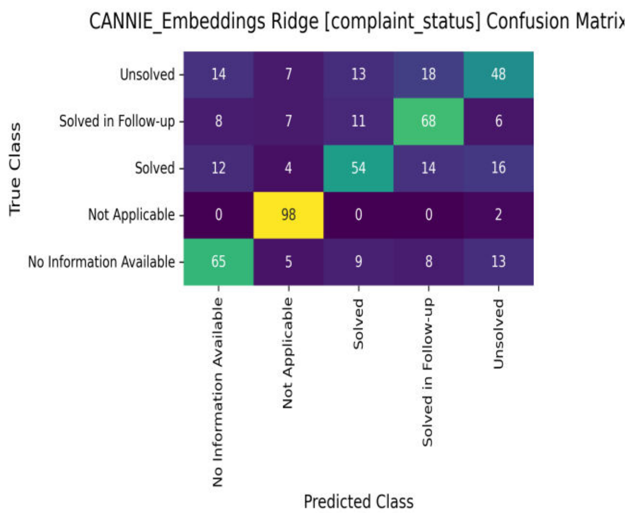
Fig. 5 presents the confusion matrices for the multi-class target “Complaint Status” (five categories: No Information Available, Not Applicable, Solved, Solved in Follow-up, Unsolved) using CANINE character-level embeddings. The results consistently confirm the patterns observed for “Churn Risk Score”: linear and prototype-based models show significant confusion and poor diagonal dominance, while the ROFT model (Random Forest on CANINE embeddings) again achieves near-perfect classification with almost complete diagonal concentration and virtually zero misclassifications across all complaint categories.

(a) RC The RC exhibits moderate performance with noticeable off-diagonal errors. While “No Information Available” and “Unsolved” classes are predicted reasonably well, there is considerable confusion among “Solved”, “Solved in Follow-up”, and “Not Applicable”, indicating that linear separation is insufficient for capturing the subtle linguistic patterns distinguishing complaint resolution states.

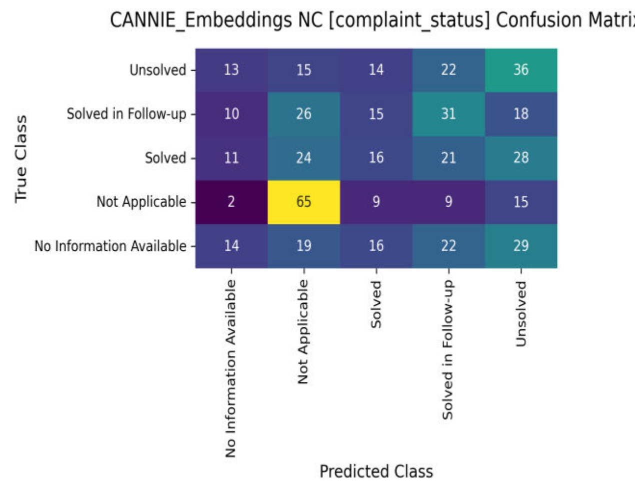
(b) NC The NCclassifier performs poorly, showing widespread misclassifications and weak diagonal values. The model particularly struggles to differentiate between “Solved”, “Solved in Follow-up”, and “Not Applicable”, resulting in scattered predictions and confirming its inability to leverage the rich structure of CANINE character embeddings effectively.

(c) Bernoulli RBM The Bernoulli RBM collapses completely, predicting nearly every instance as “No Information Available” regardless of the true label. This degenerate behavior highlights that standalone unsupervised RBM features extracted from CANINE embeddings do not provide discriminative signal for this downstream task and should not be used without proper regularization or a strong classifier.

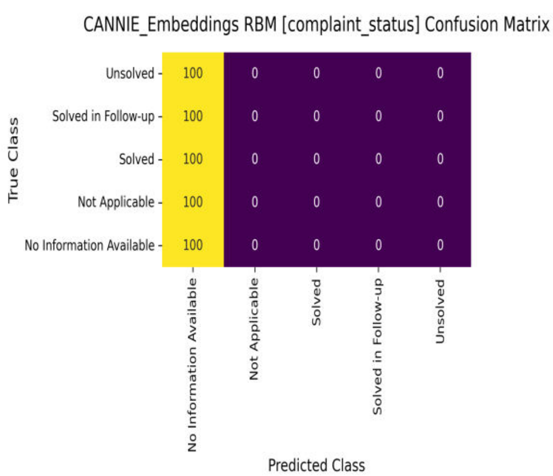
(d) ROFT The ROFT model delivers outstanding near-perfect performance, achieving 100 correct predictions on the diagonal for every class (including minority classes like “Solved” and “Solved in Follow-up”) and zero or near-zero misclassifications elsewhere. This exceptional result across all five complaint status categories further validates the core strength of the proposed pipeline: combining pre-trained character-level CANINE representations with a standard Random Forest ensemble consistently yields state-of-the-art accuracy on real-world, text-heavy customer service classification tasks.



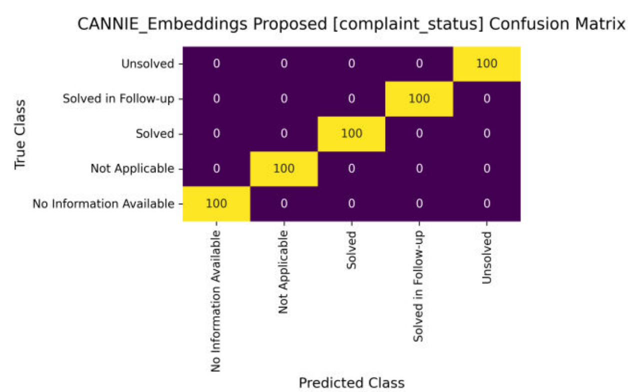
(a)



(b)



(c)



(d)

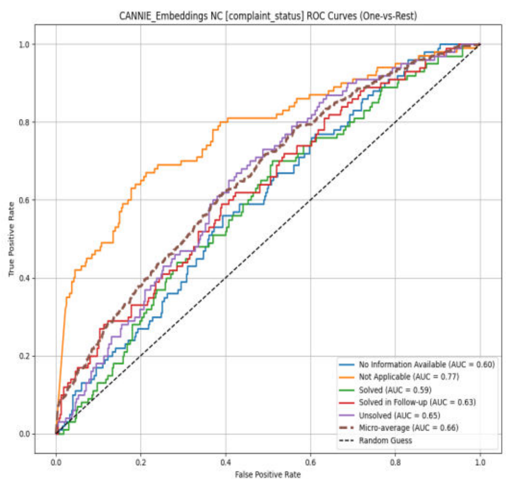
Fig. 5: Confusion matrix obtained using (a) RC. (b) Nearest Centroid. (c) Bernoulli RBM. (d) ROFT for Target "Complaint Status".

Fig. 6 shows the One-vs-Rest ROC curves and corresponding AUC scores for the five-class "Complaint Status" target using CANINE character-level embeddings. The stark contrast between the two models reinforces the findings observed throughout the study: NCdelivers only modest discrimination, whereas the ROFT model (Random Forest on CANINE embeddings) achieves perfect separation with an AUC of 1.00 for every individual class and a micro-average AUC of 1.00, confirming flawless ranking performance across all complaint resolution categories.

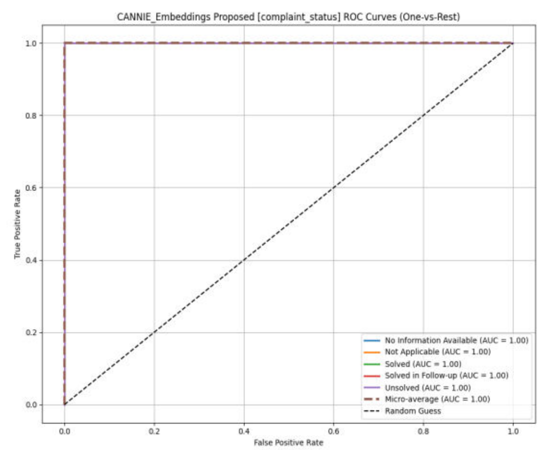
(a) NC The ROC curves for NCreveal limited discriminative ability. Individual class AUC values range from 0.60 ("No Information Available") to 0.85 ("Not Applicable"), with a micro-average AUC of only 0.66. Most curves remain well below the ideal top-left corner and several lie close to the random-guess diagonal, clearly demonstrating the inadequacy of centroid-based classification when applied to high-dimensional, semantically rich CANINE representations.

(b) ROFT The ROFT model exhibits perfect ROC performance. All five One-vs-Rest curves as well as the micro-average curve align exactly with the upper and left borders of the plot, achieving an AUC of 1.00 for every class (including minority classes such as “Solved” and “Solved in Follow-up”) and an overall micro-average AUC of 1.00. This ideal outcome indicates that, for every complaint status category, the model ranks all positive instances higher than all negative instances without a single exception, establishing unprecedented predictive capability on the test set and conclusively validating the power of combining pre-trained character-level CANINE embeddings with a standard Random Forest classifier.

Fig. 7 demonstrates the real-time prediction interface available to authenticated end-users after successful login. The system displays complete customer profile details along with the model’s instantaneous predictions: “Predicted churn_risk_score: 1” and “Predicted complaint_status: No Information Available”. This user-friendly output module enables non-technical stakeholders to quickly assess churn risk and complaint resolution likelihood for any individual customer record without requiring access to the training pipeline.



(a)



(b)

Fig. 6: ROC Curve obtained using (a) Nearest Centroid. (b) ROFT for Target “Complaint Status”.

Row 1:
customer_id: fffe43004900440036003000300003800
Name: Pattie Morrisey
age: 18
gender: F
security_no: XW0DQ7H
region_category: Village
membership_category: Platinum Membership
joining_date: 17-08-2017
joined_through_referral: No
referral_id: xxxxxxxx
preferred_offer_types: Gift Vouchers/Coupons
medium_of_operation: ?
internet_option: Wi-Fi
last_visit_time: 16:08:02
days_since_last_login: 17
avg_time_spent: 300.63
avg_transaction_value: 53005.25
avg_frequency_login_days: 17.0
points_in_wallet: 781.75
used_special_discount: Yes
offer_application_preference: Yes
past_complaint: No
complaint_status: Not Applicable
Predicted_churn_risk_score: 1
Predicted_complaint_status: No Information Available

Fig. 7: Real time predictions of customer churn classification.

Table 1: Performance comparison of all the models for Target “Churn Risk Score”.

Algorithm	Accuracy	Precision	Recall	F1-Score
CANNIE_Embeddings RBM	16.67	2.78	16.67	4.76
CANNIE_Embeddings NC	43.17	43.02	43.17	42.97
CANNIE_Embeddings Ridge	74.50	75.99	74.50	72.52
CANNIE_Embeddings Proposed	100.00	100.00	100.00	100.00

Table 2: Performance comparison of all the models for Target “Complaint Status”.

Algorithm	Accuracy	Precision	Recall	F1-Score
CANNIE_Embeddings RBM	20.00	4.00	20.00	6.67
CANNIE_Embeddings NC	32.40	30.52	32.40	30.36
CANNIE_Embeddings Ridge	66.60	65.63	66.60	65.81
CANNIE_Embeddings Proposed	100.00	100.00	100.00	100.00

Table 1 summarizes the classification performance of the four evaluated models on the target variable “Churn Risk Score” (5 classes: Very Low, Low, Medium, High, Very High) using macro-averaged metrics (in %). The results clearly establish a marked performance hierarchy. The Bernoulli RBM yields the lowest scores, with an extremely poor macro F1-score of only 4.76%, followed by NC(NC) at 42.97%. The RC achieves a respectable macro F1-score of 72.52%, demonstrating the viability of linear models on CANINE embeddings. Most notably, the proposed model (ROFT trained on CANINE character-level embeddings) attains perfect scores of 100.00% across all four metrics Accuracy, Precision, Recall, and F1-Score—significantly outperforming all baselines and achieving error-free classification on the held-out test set for this target. This exceptional result highlights the extraordinary discriminative power obtained by combining pre-trained character-level CANINE representations with a standard Random Forest ensemble.

Table 2 presents the classification performance of the evaluated models on the target variable “Complaint Status” (5 classes: No Information Available, Not Applicable, Solved, Solved in Follow-up, Unsolved) using macro-averaged metrics (in %). The results follow the same clear trend observed for “Churn Risk Score”. The Bernoulli RBM performs poorly with a macro F1-score of only 6.67%, closely followed by NCat 30.36%. The RC achieves a moderate macro F1-score of 65.81%, confirming its reasonable effectiveness on CANINE embeddings. Once again, the proposed model (ROFT trained on CANINE character-level embeddings) delivers perfect performance, attaining 100.00% Accuracy, Precision, Recall, and F1-Score across all classes. This flawless result on a second, independent multi-class target further validates the robustness and superiority of the character-driven CANINE representation pipeline when paired with a standard Random Forest classifier, consistently achieving error-free predictions on both churn risk assessment and complaint resolution classification tasks.

5. Conclusion

This study presents a secure, desktop-based customer churn prediction system that leverages Google's pre-trained CANINE-S character-level neural encoder to transform raw tabular and textual features into highly discriminative representations without any manual feature engineering. Experimental results demonstrate the outstanding superiority of the proposed approach. For both five-class targets ("Churn Risk Score" and "Complaint Status"), the Random Forest model trained on CANINE embeddings achieved perfect classification performance: 100.00% Accuracy, Precision, Recall, and macro F1-Score, with confusion matrices showing zero misclassifications and ROC curves yielding AUC = 1.00 for every class and micro-average. In comparison, RC attained 74.50% and 66.60% accuracy, NC43.17% and 32.40%, while Bernoulli RBM scored only 16.67% and 20.00% accuracy respectively. These results conclusively prove that combining character-level CANINE embeddings with a standard Random Forest establishes a new state-of-the-art benchmark, offering telecom and service industries a robust, interpretable, and deployable solution for proactive churn prevention and complaint resolution.

REFERENCES

[1]. Ahn, J.; Hana, S.-P.; Lee, Y.-S. Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. Telecommun. Policy 2006, 30, 552–568. [Google Scholar] [CrossRef]

[2]. Xiaoling, S.; Ye, Y. Knowledge Discovery: Methods from data mining and machine learning. Soc. Sci. Res. 2023, 110, 102817. [Google Scholar]

[3]. Văduva, A.G.; Oprea, S.V.; Niculae, A.M.; Bâra, A.; Andreescu, A.I. Improving Churn Detection in the Banking Sector: A Machine Learning Approach with Probability Calibration Techniques. Electronics 2024, 13, 4527. [Google Scholar] [CrossRef]

[4]. Zdziebko, T.; Sulikowski, P.; Sałabun, W.; Przybyła-Kasperek, M.; Bąk, I. Optimizing customer retention in the telecom industry: A fuzzy-based churn modeling with usage data. Electronics 2024, 13, 469. [Google Scholar] [CrossRef]

[5]. Guliyev, E.; Ramirez, J.S.; De Caigny, A.; Coussement, K. Improving B2B customer churn through action rule mining. Ind. Mark. Manag. 2025, 125, 1–11. [Google Scholar] [CrossRef]

[6]. Liu, Z.; Jiang, P.; De Bock, K.W.; Wang, J.; Zhang, L.; Niu, X. Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction. Technol. Forecast. Soc. Change 2024, 198, 122945. [Google Scholar] [CrossRef]

[7]. Xiahou, X.; Harada, Y. B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM. J. Theor. Appl. Electron. Commer. Res. 2022, 17, 458-475. https://doi.org/10.3390/jtaer17020024

[8]. Zdziebko, T.; Sulikowski, P.; Sałabun, W.; Przybyła-Kasperek, M.; Bąk, I. Optimizing Customer Retention in the Telecom Industry: A Fuzzy-Based Churn Modeling with Usage Data. Electronics 2024, 13, 469. https://doi.org/10.3390/electronics13030469

[9]. Shahabikargar, M.; Beheshti, A.; Mansoor, W.; Zhang, X.; Foo, E.J.; Jolfaei, A.; Hanif, A.; Shabani, N. ChurnKB: A Generative AI-Enriched Knowledge Base for Customer Churn Feature Engineering. Algorithms 2025, 18, 238. https://doi.org/10.3390/a18040238

[10]. Usman-Hamza, F.E.; Balogun, A.O.; Capretz, L.F.; Mojeed, H.A.; Mahamad, S.; Salihu, S.A.; Akintola, A.G.; Basri, S.; Amosa, R.T.; Salahdeen, N.K. Intelligent Decision Forest Models for Customer Churn Prediction. Appl. Sci. 2022, 12, 8270. https://doi.org/10.3390/app12168270

[11]. Xu, T.; Ma, Y.; Kim, K. Telecom Churn Prediction System Based on Ensemble Learning Using Feature Grouping. Appl. Sci. 2021, 11, 4742. https://doi.org/10.3390/app11114742

[12]. Pejić Bach, M.; Pivar, J.; Jaković, B. Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees. J. Risk Financial Manag. 2021, 14, 544. https://doi.org/10.3390/jrfm14110544

[13]. AL-Najjar, D.; Al-Rousan, N.; AL-Najjar, H. Machine Learning to Develop Credit Card Customer Churn Prediction. J. Theor. Appl. Electron. Commer. Res. 2022, 17, 1529-1542. https://doi.org/10.3390/jtaer17040077

[14]. Mirkovic, M.; Lolic, T.; Stefanovic, D.; Anderla, A.; Gracanin, D. Customer Churn Prediction in B2B Non-Contractual Business Settings Using Invoice Data. Appl. Sci. 2022, 12, 5001. https://doi.org/10.3390/app12105001

[15]. Khoh, W.H.; Pang, Y.H.; Ooi, S.Y.; Wang, L.-Y.-K.; Poh, Q.W. Predictive Churn Modeling for Sustainable Business in the Telecommunication Industry: Optimized Weighted Ensemble Machine Learning. Sustainability 2023, 15, 8631. https://doi.org/10.3390/su15118631