Can a simple customer review outperform a feature set for predicting churn?

William Jones Beckhauser[1], **Renato Fileto**[1]

[1]Department of Informatics and Statistics Federal University of Santa Catarina, Florianópolis, SC, Brazil.

beckhauserwilliam@gmail.com, r.fileto@ufsc.br

Abstract. *Churn prediction traditionally employs customer profile and transaction data, leaving textual features like customer comments unexploited. This work compares machine learning models for churn prediction that use conventional data with those using reviews posted by customers about their purchases. Our experiments with the most used models for churn prediction in the literature reveal that using conventional data the models perform the best with RFM segmentation, achieving up to 93% F1-Score. It drops to less than 75% without RFM segmentation. In contrast, by using BERT embeddings of review texts, an F1-Score of 96% is achieved.*

Resumo. *A previsão de perda de clientes (churn) tradicionalmente usa dados de perfis e transações, deixando inexploradas características textuais como comentários dos clientes. Este trabalho compara modelos de aprendizado de máquina para previsão de churn que usam dados convencionais com aqueles que usam revisões postadas pelos clientes sobre suas compras. Nossos experimentos com os modelos mais utilizados para previsão de churn na literatura revelam que, usando dados convencionais, os modelos apresentam o melhor desempenho com a segmentação RFM, alcançando até 93% de F1-Score. Esse valor cai para menos de 75% sem a segmentação RFM. Em contraste, usando embeddings BERT dos textos das avaliações um F1-Score de 96% é alcançado.*

1. Introduction

Effective churn prediction, i.e., identifying the customers with a likelihood of leaving, allows preventive measures to strengthen relationships with them [27]. It is essential for the survival and growth of businesses amidst intense corporate competition [9]. A modest 5% customer retention rate allows companies to increase their profits from 25% to 95% [8]. Therefore, churn prediction is a critical aspect of Customer Relationship Management (CRM). Improving customer retention helps balance the consumer scale, reducing the imperative for acquiring new customers and consequently cutting marketing costs. This is achieved by realizing financial gains through effective retention strategies [13].

The growing demand for churn prediction motivates the use of Machine Learning (ML) technologies [26] to automate this task. A variety of ML models have been investigated in recent years for churn prediction, especially in sectors like telecommunication [25, 15]. However, despite the frequent application in specific sectors, most ML models for churn prediction rely just on conventional data like customer profile and transaction data[1, 10], leaving aside Natural Languagem Processing (NLP) techniques that could take advantage of textual data like customer comments or reviews of their purchases. In a recent bibliographical review, we found only 12 studies directly related to churn prediction using NLP, and they are restricted to microblog data [4, 3].

This work addresses the scarcity of research on churn prediction using NLP techniques by conducting experiments comparing ML models such as Random Forest (RF), Multilayer Perceptron (MLP), Support Vector Machine (SVM) and Decision Tree (DT), for churn prediction using conventional data about the customer profiles and transactions and using textual reviews posted by customers to comment about their purchases. We apply NLP techniques and models, Word2Vec (both Skip-gram and CBOW) and BERT, to generate embeddings of the textual reviews to feed ML models.

Our experiments used data from a food delivery company comprising 75,000 orders from 7,249 customers, with over 40% providing textual reviews. Transactional data underwent balancing, OneHotEncoding, and RFM segmentation, while textual data underwent preprocessing and language translation. The models were trained with conventional profile and transaction data and alternatively with two subsets of review data, one multilingual and the other one translated into English. In the experiments, RF achieved over 90% accuracy with RFM segmentation. However, it felt be 15% without RFM. On the other hand, in the experiments with textual data of customer reviews, MLP trained with BERT embbeding achieved the best performance, with over 96% accuracy.

To the best of our knowledge, this is the first work in the literature to apply Word2Vec and BERT embeddings of customer reviews for churn prediction, and the first to apply NLP techniques for churn prediction in the Food Delivery sector. Its main contributions are: (i) systematic training and evaluation of ML models, with conventional data and textual data handled with NLP techniques; (ii) experimental results showing the superior results obtained by using review text embeddings as features for the ML predictors; (iii) analysis of the effects of using RFM segmentation in churn prediction, including possibly biased results widespread in proposals of the literature; (iv) provision of a rich sub-dataset with customer reviews from a food delivery app for the research community.

The remainder of this paper is organized into 5 other sections. Section 2 lays the groundwork for understanding the current landscape in churn prediction and comparing our proposal with related work. Section 3 describes the analysis and preparation of conventional data and textual data. Section 4 reports the experiments conducted and their outcomes. Finally, Section 5 presents the conclusions drawn from the research and suggestions for future research directions.

2. Related work

The term "churn" is commonly used to refer to any type of customer attrition, voluntary or involuntary [6]. Churn occurs when a customer discontinues using a product or service, becoming inactive. Dissatisfaction with the quality of service, high costs, unattractive plans, lack of understanding of the service plan, poor support, and other factors may contribute to customer churn [20]. The following subsections 2.1 and 2.2 discuss scholarly works about churn prediction that use conventional data of customer profiles and transactions, and those that apply NLP techniques on textual features, respectively.

2.1. Churn prediction with customer profile and transaction data

In the last decade, studies on churn prediction focused mainly on sectors like telecommunications [15, 1, 10]. More than 45% of all research on churn prediction used case studies from this sector, followed by the financial sector [25]. This concentration in telecom is probably due to the abundancy of data available in this sector. It caused a need for techniques to segment data for model training. On the other hand, emerging sectors with high churn rates, such as gamming [22], e-commerce [21] and non-contractual companies [24], are still under-explored in the churn prediction research [5].

Some models stand out in our bibliographical search about churn prediction during the last 5 years, namely: Random Forests (RF) used in 30 articles, Scalable Vector Machine (SVM) in 29 articles, and Decision Trees (DT) in 52 articles. On the other hand, the Multilayer Perceptron (MLP) was used in just 4 articles published until 2020. Publications using other techniques are taking short steps, according to [25]. However, many proposals employed RFM segmentation [7], which classifies customers based on:

Recency (R) the time since the last purchase (the shorter this time the higher the R);
Frequency (F) the number of transactions within a specific period (e.g., year, month);
Monetary value (M) the amount spent on purchases within a specific time period.

RFM segmentation enables around 15% accuracy gains in several works [5, 19, 23, 2]. However, recency depends on the same criteria used to define churn (elapsed time since last purchase). It may inadvertently compromised experimental results, as we discuss in the end of section 4.

2.2. Churn prediction by applying NLP on textual features

We also conducted a bibliographical search for churn prediction models exploiting textual features and NLP techniques. Table 1 lists the selected articles, with the kind of data employed in their respective experiments, ML techiques used and the highest performance obtained. Notice that four of these studies use data from social media, specifically tweets, and one from phone calls. Only our work exploit reviews posted by customers about their purchases. According to Zhong and Li, two factors make social media less relevant for predicting churn: (1) customers prefer to contact the company directly instead of posting on social networks and (2) there is a shortage of training data [29]. In addition, while previous work use SVM, Recurrent Neural Network (RNN) e Convolutional Neural Network (CNN) models, our work is the first to perform more extensive experiments with 3 distinct textual datasets and BERT, MLP, RF e SVM models.

Table 1. NLP in Churn Prediction: Comparison of selected articles.

Work	Data	Techniques	Performance
[4]	Microblog (tweets)	SVM	F1 Score: 75%
[3]	Microblog (tweets)	RNN	F1 Score: 78%
[11]	Microblog (tweets)	CNN	F1 Score: 84%
[14]	Microblog (tweets)	SVM	Accucary: 94%
[29]	Phone Calls (telecom)	CNN	F1 Score: 91%
This Work (2024)	Customer Reviews (Food Delivery)	BERT, MLP, RF, SVM	F1 Score: 96%

After reviewing the related articles, we identified the opportunity to implement these approaches with greater robustness and evaluate them on real datasets from companies that record customer reviews or interactions arose. Furthermore, we exploit contemporary model like BERT, which, to the best our knowledge, have not being applied for churn prediction yet [16], neither for just generating embeddings for further processing nor for being fine-tuned to solve the churn classification task. For comparative purposes, we complement the experiments by using Word2vec with CBOW and Skip-gram for generating token embeddings to be feed as features to MLP [5], SVM [14] and RF [25] models. In our experiments the MLP model fed with BERT embeddings yielded the highest performance, as detailed in section 4.

3. Data understanding and preparation

The methodology adopted in this work draws from the CRISP-DM [?] general process, incorporating five of its six core phases: business understanding, data understanding, data preparation, modeling, and validation. It guarantees a structured process for training and classifying models for predicting customers at risk of churn. Section 3.1 describes the Food Delivery dataset employed in our experiments. Analysis and preparation were performed on conventional data and textual data separately, as described in the following sections 3.2 and 3.3, respectively.

3.1. The Food Delivery dataset

This dataset especially obtained for our experiments, came from a food delivery. It has data about 76,446 deliveries to 7,249 customers, mainly for lunch and dinner on weekdays (Monday to Friday), during the 12-month period between April 20, 2022, and April 21, 2023. Table 2 lists the features available in this dataset grouped by their data types. About 40% of customers reviewed their last purchase. For this dataset, we consider 30 days without purchasing as churn, because it is the period of time commonly waited in loyalty programs to intensify promotions [28, 5].

Table 2. Food Delivery dataset summary type of featuring.

Data Type	Quantity	Variables
Numeric	12	ItemsOrderedCount; DessertsCount; AppliedDiscountsCount; DrinksCount; MainDishesCount; NonMainItemsCount; TransactionTotalValue; AverageTransactionValue; TotalDiscountValue; AverageDiscountValue; ReviewScores; DaysSinceLastOrder
String	4	OrderAreaName; TransactionLanguage; CustomerGender; RFMSegment
Boolean	8	IncludesDinnerItems; IncludesLunchItems; EmailMarketingConsent; SMSMarketingConsent; LoyaltyPlanStatus; PromoCodeUsed; IsFirstOrder; Churn (target)
Text	1	CustomerReviewsText

3.2. Conventional data analysis and preparation

The analysis and preparation of the conventional data (about customer profiles and transactions) began with data cleaning. The data quality was assessed by identifying missing values, outliers, and duplicated records. The percentage of missing values per column was calculated and statistical measures like the Z-score was used to detect discrepancies. Then, we analyzed the dispersion of feature values and their correlation with the target variable, churn. Adjustments made included renaming columns and converting data types, to facilitate analysis. Registers with null values were excluded, and categorical data managed through one-hot encoding.

Following data cleaning, RFM analysis was conducted for customer segmentation. Relevant columns were selected and renamed to align with the RFM framework. Quartiles for each RFM metric were calculated, assigning scores to customers. These scores enabled segmentation into the labels “Not_Fan”, “Switchers”, “Loyal”, and “Champions”. Figure 1 shows the number of customers associated each RFM label for the Food Delivery dataset used in the experiments reported in this paper.

Figure 1. RFM segmentation of the customers in the Food delivery

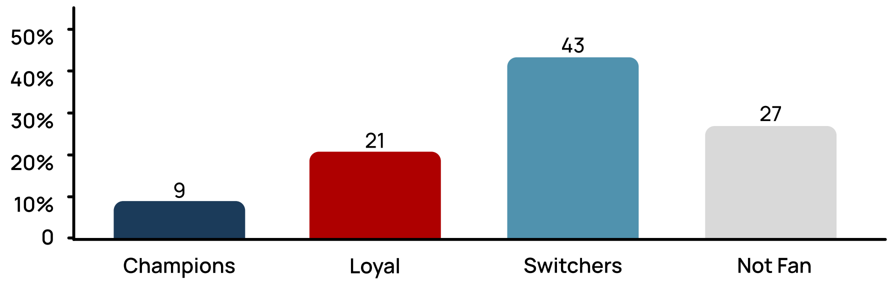


Table 3 provides a snapshot of descriptive statistics for some selected features of the Food Delivery dataset. A more comprehensive descriptive and exploratory analysis of its features and values is publicly available at GitHub[1] . Data analysis and preparation used Python libraries: pandas for data frame manipulation, numpy for numerical computations, matplotlib and seaborn for visualization, and Google Colab’s files module for file operations. Steps include normalizing data with MinMaxScaler and encoding categorical variables with OneHotEncoder from sklearn.preprocessing.

Table 3. Distribution of selected features from the Food Delivery dataset

index	orders	plates	totalValue	discountTotal	DaysLastOrder
count	7249	7249	7249	7249	7249
mean	10.55	13.75	88.78	27.51	51.48
std	16.68	44.96	132.57	97.08	53.16
min	1.0	0.0	0.0	0.0	1.0
max	237.0	2672.0	985.0	978.0	182.0

[1] github.com/WilliamBeckhauser/Can-a-simple-customer-review-outperform-a-feature-set-for-predicting-churn-

3.3. Textual data analysis and preparation

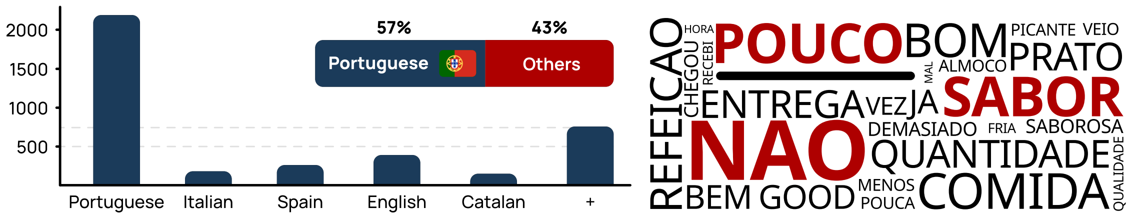
Customer reviews about their purchases are short texts describing their experiences. They encompass a range of topics, ranging from complaints about undelivered items to discussions about meal prices, portion sizes and compliments. Figure 2 shows three examples of customer reviews, one per line, with the original version of the reviews on the left column, and their respective English version, obtained by using Googletrans, on the right column. Most reviews are short as the ones in these examples.

Figure 2. Examples of reviews from the Food Delivery dataset.

Original	Translated
Excelente, e o estafeta muito educado	Excellent, and the delivery guy was very polite
Gostei, mas pouca quantidade	I liked it, but not much
O pedido chegou muito tarde	The order arrived very late

The Food Delivery dataset contains reviews written in more than ten languages. Our analysis identified Portuguese (specifically from Portugal) as the most common language, appearing in 57% of the reviews. English, Spanish, Italian, and Catalan also make up a significant portion, accounting for the remaining 43%. Figure 3 shows the most prominent languages (left) and words (right) found in these reviews.

Figure 3. Most used languages and words in the Food Delivery reviews.



The translation of the multi-language reviews into English aimed to asses possible challenges and deviations of experimental results due to linguistic diversity. Despite the overall precision of the translations generate by the Googletrans library, we manually uncovered about 30 reviews whose translations deviated from their original meanings. It happened mainly with reviews one or two words long. This manual verification and fix of translation fidelity was limited to Portuguese, Italian, Spanish, and Catalan. This process resulted in two datasets: one with the original reviews in multiple languages and the other with the reviews translated into English.

Sentiment analysis of the reviews was done using SiEBERT [12], which has demonstrated higher performance than generative models like GPT-4 [17]. On the original dataset, 82% of the non-churn reviews are positive, while 40% of the churn reviews are negative. Among reviews translated into English, 84% of the non-churn are positive and 86% of the churn reviews are negative.

Prior to model training, the textual data underwent preprocessing. We employed a variety of tokenization and embedding techniques, including BERT and Word2vec with both skip-gram and CBOW models. The tokenization and embedding processes were greatly facilitated by the transformer libraries from Hugging Face, such as BertTokenizer and the BertModel for embeddings. Furthermore, the PyTorch library, with its DataLoader and Dataset classes, played a crucial role in managing the data efficiently, enabling data processing in segments during both training and evaluation. Then, we divided the dataset into training and test sets, with the test set comprising 20% of the data, using the train_test_split function from sklearn.model_selection.

4. Experiments

4.1. Churn prediction using profile and transaction features

The experiments with conventional data used Python on a Google Colaboration Pro plan environment with 12.7 GB of RAM and 78.2 GB of disk space. Model training used scikit-learn libraries: MLPClassifier, SVC, DecisionTreeClassifier, and RandomForestClassifier. The train_test_split library was used for data partitioning. The libraries GridSearchCV and cross_val_score were used for hyperparameter tuning and model evaluation, respectively. Various metrics from sklearn.metrics were used for performance assessment, including: accuracy_score, f1_score, precision_score, recall_score, roc_auc_score, roc_curve, classification_report, confusion_matrix. A test size of 20% was adopted. Table 4 lists the optimal hyperparameter configurations found for each model.

Table 4. Best combinations of hyperparameter values

RF	SVM	DT	MLP
criterion: 'entropy' max depth: 10 n estimators: 50	kernel: poly C: 10 gamma: 'scale'	criterion: 'entropy' max_depth: 10 min_samples_split: 10 min_samples_leaf: 4 splitter: 'best'	hidden layer sizes: 5 alpha: 0.001 solver: 'adam' activation: 'relu'

The optimal settings for RF were determined to be a maximum depth of 10 and 50 estimators. The SVM model was fine-tuned with an 'rbf' kernel, C=1, and 'scale' gamma for the Gambling and E-commerce datasets. A minimum sample split of 10 was found optimal for Food Delivery, with the MLP model sharing the same hyperparameters: hidden layer sizes of 5, alpha=0.001, solver='adam', and activation='relu'.

Table 5 summarizes the outcomes. The RF model achieved the highest accuracy (90.69%), followed by the Decision Tree (90.21%) and MLP models (88.41%).

Table 5. Final results after parameter optimization.

		Accuracy	F1-score	Precision
Food Delivery	Multilayer Perceptron	88.4%	88.4%	89.7%
	Support Vector Machine	87.9%	88.0%	78.6%
	Decision Tree	90.2%	89.9%	90.2%
	Random Forest	**90.7%**	**90.9%**	**91.1%**

Further experimentation, excluding RFM segmentation indicated a consistent decline in accuracy by at least 15%, with RF achieving the highest accuracy of 74%.

4.2. Churn prediction using textual reviews

The experiments with customer reviews were conducted on a MacBook Pro with an Apple M2 Pro chip, a 12-core CPU, a 19-core GPU, 16GB of RAM and a 512GB SSD. Jupyter Notebook served as the development and execution environment. Python libraries were selected for ML and NLP tasks. The transformers library was used for loading and fine-tuning BERT. The gensim library facilitated the training and the application of Word2vec models for text vectorization. For ML models such as SVM and RF, and for data preprocessing tasks like dataset splitting, the scikit-learn library was employed.

Only review texts of the most recent purchase of each customer from the Food Delivery dataset were employed. These reviews are available for around 40% of the customers. They went through a preprocessing procedure involving the removal of emojis and punctuation marks, as well as being converted to lowercase. The dataset was divided into training and test sets, using a holdout method, with each subset representing 80% and 20% of the data, respectively. Text tokenization and encoding into embeddings were accomplished using BERT, as well as Word2vec with both Skip-gram and CBOW. The BERT model was loaded and fine-tuned for the task of binary sequence classification. The embeddings produced by Word2vec were utilized as inputs for the SVM, MLP, and RF.

Churn prediction models using Word2vec embeddings

Table 6 shows the accuracy, F1-score and hyperparameters of the best churn prediction models trained with Word2vec embeddings, both CBOW and Skip-gram. All models performed better with reviews translated into English, with MLP giving the best results.

Table 6. Models' performance and hyperparameters with Word2vec embeddings

Model	Language	Accuracy	F1-Score	Hyperparameters
				Word2vec CBOW
RF	Multilingual	72.95%	72.26%	Max depth: 10, Estimators: 100
RF	Translated	75.74%	75.64%	Max depth: 20, Estimators: 200
SVM	Multilingual	62.05%	52.07%	C: 10, Kernel: linear
SVM	Translated	72.52%	71.41%	C: 10, Kernel: linear
MLP	Multilingual	67.56%	67.60%	Activation: relu, Layers: (100,), Iter: 300
MLP	Translated	71.85%	71.69%	Activation: tanh, Layers: (100,), Iter: 200
				Word2vec Skip-gram
RF	Multilingual	74.02%	73.61%	Max depth: 10, Estimators: 200
RF	Translated	78.41%	78.37%	Max depth: 10, Estimators: 200
SVM	Multilingual	73.76%	73.09%	C: 10, Kernel: linear
SVM	Translated	78.55%	78.65%	C: 10, Kernel: rbf
MLP	Multilingual	75.23%	74.65%	Activation: tanh, Layers: (50,), Iter: 300
MLP	Translated	**79.76%**	**79.95%**	Activation: relu, Layers: (100,), Iter: 300

Churn prediction using BERT embeddings

Table 7 shows the accuracy, F1-score and hyperparameters of the best models using BERT embeddings in the first 6 lines, and of the fine-tuned BERT models in the last 2 lines. The performance gains using BERT embeddings instead of Word2Vec embeddings are considerable for RF, SVM, and MLP. On the other hand, fine-tuned BERT models were not so competitive. As occurred using Word2Vec embeddings, all these models performed better on reviews translated into English, and MLP delivered the best results. The optimal hyperparameters varied between the original and English datasets, reflecting the distinct characteristics of the multilingual and translated texts.

Table 7. Models' performance and hyperparameters using BERT

Model	Language	Accuracy	F1-Score	Hyperparameters
RF	Multilingual	93.40%	92.20%	bootstrap: False, max_depth: None, min_samples_leaf: 4, min_samples_split: 2, n_estimators: 100
RF	Translated	96.11%	95.15%	bootstrap: True, max_depth: 30, min_samples_leaf: 4, min_samples_split: 2, n_estimators: 200
SVM	Multilingual	93.46%	93.64%	C: 10, degree: 2, gamma: 'scale', kernel: 'rbf'
SVM	Translated	96.19%	96.34%	C: 0.1, degree: 2, gamma: 'scale', kernel: 'linear'
MLP	Multilingual	93.03%	93.23%	activation: 'tanh', alpha: 0.05, hidden_layer_sizes: (50, 50), learning_rate: 'adaptive', solver: 'sgd'
MLP	Translated	**96.61%**	**96.65%**	activation: 'tanh', alpha: 0.001, hidden_layer_sizes: (50, 100, 50), learning_rate: 'constant', solver: 'sgd'
Fine-tune	Original	86,94%	84,48%	l̀r': 2e-05, 'batch_size': 32
Fine-tune	Translated	89,14%	86,58%	l̀r': 2e-05, 'batch_size': 16

4.3. Discussion

The experimental results are clearly superior when using BERT embeddings of review texts as features for RF, SVM and MLP classifies (first 2 lines of Table 7), when compared to using conventional data (Table 5) or the textual features vectorized as Word2Vec embeddings (Table 6). It shows the potential of textual features for churn prediction, and the superiority of contextualized embeddings for codifying their contents.

Translating the multilingual reviews into English before model training consistently improves performance. This improvement can be attributed to the extensive pre-training of the embedding models on diverse English corpora, which likely enhances their representation capacity for English. Furthermore, the variation in optimal hyperparameters between the original and translated texts suggests that the characteristics of the language significantly influence the classification model. Specifically, models trained on English texts tended to perform better with different configurations of hyperparameters.

Another relevant issue regards RFM segmentation in the experiments using conventional data. RFM segmentation of customers led to an average increase of 15% in prediction accuracy. However, RFM segmentation could have introduced a bias in the training data, since RFM considers the date of the last purchase and the churn in each dataset was defined as at least a certain number of days without new purchases after the last one. We carried out further experiments to investigate this possible bias. Table 8 shows model accuracy in experiments using the same methodology used to generated the results presented in Table 5, but also using alternative segmentations, based on pairs of RFM measures. Excluding F or M led to an accuracy increase of up to 5% over RFM. However, the omission of the R measure resulted in an accuracy decline between 13% and 20%. It suggests that, in fact, the use of the R measure for customer segmentation, as we initially did, allows dubious gains in models performance. We found out that it is a common practice in the literature [5, 23, 18], without mentioning this bias.

Table 8. Models' accuracy with alternative segmentations

Model	RFM	RF	RM	FM
MLP	88.40%	90.21%	90.28%	**75.59%**
SVM	87.90%	90.62%	90.83%	74.21%
Decision Trees	90.20%	**90.83%**	90.62%	74.69%
Random Forest	**90.70%**	90.55%	**91.17%**	75.31%

5. Conclusions and future work

The prediction of customer churn is crucial for many enterprises because it enables taking measures in advance to restore loyalty. This research undertook a comprehensive comparison of various ML and language models to forecast customer churn within the food delivery sector. In our initial analysis, we focused on conventional data features about customer profiles and transactions. The RF surpassed other algorithms, including SVM and MLP, using these features. Particularly notable was the performance improvement using RFM segmentation. Without it, we observed a significant 15% drop in performance. Further investigation into RFM segmentation's components (RF, RM, FM) identified a critical dependence on the "Recency" factor. Our subsequent experiment revealed that feeding just BERT embeddings of customer reviews into the MLP and SVM models allowed superior predictive accuracy than using conventional data features. Remarkably, this approach yielded superior results even when the analysis was constrained to a single variable, far outpacing the outcomes obtained with transactional data alone, regardless of RFM segmentation application.

As future work we envision: (i) experiments with other datasets that include texts like customer comments; (ii) investigating the impact of polarity and emotions found in these texts on churn and (iii) exploiting new NLP technologies like current generative models to identify causes of churn on available text data. We are currently performing experiments using LLMs, such as GPT, Llama, Gemini, and Claude for multi-label classification of potential churn due to specific reasons, like problems in food quality or food delivery, which can be expressed in the reviews of the Food Delivery dataset.

Acknowledgements

This work has been supported by a 2022 CNPq Universal grant, FAPESC grant 2021TR1510, the Print CAPES-UFSC Automation 4.0 Project, and indirectly by the Céos project, financed by the Public Ministry of Santa Catarina State (MPSC), which has contributed significantly to the improvement of our Laboratory (LISA) and the high-performance processing infrastructure at UFSC.

References

[1] Ahn, J., Hwang, J., Kim, D., Choi, H., Kang, S.: A survey on churn analysis in various business domains. IEEE Access **8**, 220816–220839 (2020). https://doi.org/10.1109/ACCESS.2020.3042657

[2] Aleksandrova, Y.: Application of machine learning for churn prediction based on transactional data (rfm analysis). In: 18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings. vol. 18, pp. 125–132 (2018)

[3] Amiri, H., Daume III, H.: Target-dependent churn classification in microblogs. Proceedings of the AAAI Conference on Artificial Intelligence **29**(1) (Feb 2015). https://doi.org/10.1609/aaai.v29i1.9532

[4] Amiri, H., Daume III, H.: Short text representation for detecting churn in microblogs. Proceedings of the AAAI Conference on Artificial Intelligence **30**(1) (Mar 2016). https://doi.org/10.1609/aaai.v30i1.10333

[5] Beckhauser, W.J., Fileto, R.: Churn prediction in enterprises with high customer turnover. In: International Conference on Information Integration and Web Intelligence. pp. 176–191. Springer (2023)

[6] Berry, M.J., Linoff, G.S.: Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons (2004)

[7] Cheng, C.H., Chen, Y.S.: Classifying the segmentation of customer value via rfm model and rs theory. Expert systems with applications **36**(3), 4176–4184 (2009)

[8] Gallo, A.: The value of keeping the right customers. Harvard Business Review (Nov 2014), `https://hbr.org/2014/10/the-value-of-keeping-the-right-customers`

[9] Gao, L., de Haan, E., Melero-Polo, I., Sese, F.J.: Winning your customers' minds and hearts: Disentangling the effects of lock-in and affective customer experience on retention. Journal of the Academy of Marketing Science **51**(2), 334–371 (2023)

[10] Geiler, L., Affeldt, S., Nadif, M.: A survey on machine learning methods for churn prediction. International Journal of Data Science and Analytics **14**(3), 217–242 (2022)

[11] Gridach, M., Haddad, H., Mulki, H.: Churn identification in microblogs using convolutional neural networks with structured logical knowledge (11 2017). https://doi.org/10.18653/v1/W17-4403

[12] Hartmann, J., Heitmann, M., Siebert, C., Schamp, C.: More than a feeling: Accuracy and application of sentiment analysis. International Journal of Research in Marketing **40**(1), 75–87 (2023). https://doi.org/https://doi.org/10.1016/j.ijresmar.2022.05.005

[13] Hochstein, B., Voorhees, C.M., Pratt, A.B., Rangarajan, D., Nagel, D.M., Mehrotra, V.: Customer success management, customer health, and retention in b2b industries. International Journal of Research in Marketing **40**(4), 912–932 (2023)

[14] Ibitoye, A., Onifade, O.: Sequentially clustered social opinion for improved customer management in churn prediction (06 2022). https://doi.org/10.21203/rs.3.rs-1797516/v1

[15] Jain, H., Khunteta, A., Srivastava, S.: Telecom churn prediction and used techniques, datasets and performance measures: a review. Telecommunication systems **76**(4), 613–630 (2021)

[16] Koroteev, M.V.: Bert: A review of applications in natural language processing and understanding. ArXiv **abs/2103.11943** (2021), `https://api.semanticscholar.org/CorpusID:232307525`

[17] Krugmann, J.O., Hartmann, J.: Sentiment analysis in the age of generative ai. Customer Needs and Solutions **11**(1), 1–19 (2024)

[18] Mena, G., Coussement, K., De Bock, K.W., De Caigny, A., Lessmann, S.: Exploiting time-varying rfm measures for customer churn prediction with deep neural networks. Annals of Operations Research pp. 1–23 (2023)

[19] Mitrovic, S., Singh, G., Baesens, B., Lemahieu, W., de Weerdt, J.: Scalable rfm-enriched representation learning for churn prediction. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA). pp. 79–88 (2017). https://doi.org/10.1109/DSAA.2017.42

[20] Mustafa, N., Sook Ling, L., Abdul Razak, S.F.: Customer churn prediction for telecommunication industry: A malaysian case study [version 1; peer review: 2 approved]. F1000 research **10**, 1274–1274 (2021)

[21] Nagaraj, P., Muneeswaran, V., Dharanidharan, A., Aakash, M., Balananthanan, K., Rajkumar, C.: E-commerce customer churn prediction scheme based on customer behaviour using machine learning. In: Intl. Conf. on Computer Communication and Informatics (ICCCI). pp. 1–6. IEEE (2023)

[22] Perišić, A., Pahor, M.: Clustering mixed-type player behavior data for churn prediction in mobile games. Central European journal of operations research **31**(1), 165–190 (2023)

[23] Perišić, A., Pahor, M.: Rfm-lir feature framework for churn prediction in the mobile games market. IEEE Transactions on Games **14**(2), 126–137 (2022). https://doi.org/10.1109/TG.2021.3067114

[24] Shobana, J., Gangadhar, C., Arora, R.K., Renjith, P., Bamini, J., devidas Chincholkar, Y.: E-commerce customer churn prevention using machine learning-based business intelligence strategy. Measurement: Sensors **27**, 100728 (2023)

[25] Sobreiro, P., Martinho, D.D.S., Alonso, J.G., Berrocal, J.: A slr on customer dropout prediction. IEEE access **10**, 14529–14547 (2022)

[26] Suh, Y.: Machine learning based customer churn prediction in home appliance rental business. Journal of big Data **10**(1), 41 (2023)

[27] Tran, H., Le, N., Nguyen, V.H.: Customer churn prediction in the banking sector using machine learning-based classification models. Interdisciplinary Journal of Information, Knowledge & Management **18** (2023)

[28] Wenger, M.: Strategic Business Models in the Online Food Delivery Industry-Detailed Analysis of the-" Order and Delivery" Business Model. Master's thesis, Universidade NOVA de Lisboa (Portugal) (2021)

[29] Zhong, J., Li, W.: Predicting customer churn in the telecommunication industry by analyzing phone call transcripts with convolutional neural networks. In: 3rd Intl. Conf. on Innovation in Artificial Intelligence. p. 55–59. ICIAI 2019, ACM, New York, NY, USA (2019). https://doi.org/10.1145/3319921.3319937