CUSTOMER CHURN PREDICTION FOR THE PAY-TV SECTOR

TUĞÇE AYDIN HATAŞ

MEF UNIVERSITY

OCTOBER 2023

MEF UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MASTER'S IN INFORMATION TECHNOLOGIES

M. Sc. THESIS

CUSTOMER CHURN PREDICTION FOR THE PAY-TV SECTOR

Tuğçe AYDIN HATAŞ

ORCID NO: 0000-0003-4466-9956

Thesis Advisor: Asst. Prof. Dr. Tuna ÇAKAR

OCTOBER 2023

ACADEMIC HONESTY PLEDGE

I declare that all the information in this study is collected and presented in accordance with academic rules and ethical principles, and that all information and documents that are not original in the study are referenced in accordance with the citation standards, within the framework required by the rules and principles as a graduation project Master's Degree in Information Technologies.

Name and Surname: Tuğçe AYDIN HATAŞ

Signature:

ABSTRACT

CUSTOMER CHURN PREDICTION FOR THE PAY-TV SECTOR

Tuğçe AYDIN HATAŞ

M.Sc. in Information Technologies

Thesis Advisor: Asst. Prof. Dr. Tuna ÇAKAR

October 2023, 37 Pages

Understanding the reasons for customer churn provides added value in terms of retaining existing customers, as customer attrition leads to revenue loss for companies and incurs marketing costs for acquiring new customers. In this study, the 6-month historical data of a Pay-TV company operating in Turkey was used, and due to the imbalanced nature of the dataset on a label basis, the oversampling method was applied. During the model development phase, various artificial learning algorithms (Random Forest, Logistic Regression, K-Nearest Neighbors, Decision Tree, AdaBoost, XGBoost, Extra Tree Classifier) were utilized, and their performances were compared. Based on the evaluation of success criteria for each model, it was observed that the tree-based Random Forest, Extra Tree Classifier and XGBoost achieved the highest performance for this dataset.

Keywords: Customer churn, churn prediction, machine learning, pay-tv industry, customer retention.

Numeric Code of the Field: 92404

ÖZET

PAY-TV SEKTÖRÜNDE MÜŞTERİ KAYIP TAHMİNİ

Tuğçe AYDIN HATAŞ

Bilişim Teknolojileri Tezli Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Tuna ÇAKAR

Ekim 2023, 37 Sayfa

Müşteri kaybı, şirketler için gelir kaybı ve yeni müşteri kazanımı için pazarlama maliyetleri yarattığından müşterilerin aboneliklerini neden sonlandırdıklarını anlamak, mevcut müşterileri elde tutmak açısından katma değer sağlamaktadır. Bu çalışma kapsamında Türkiye'de hizmet veren Pay-TV firmasının müşterilerinin 6 aylık geçmiş verileri kullanılmış ve veri setinin etiket bazında dengesiz olması sebebiyle aşırı örnekleme yöntemi de uygulanmıştır. Model geliştirme aşamasında farklı yapay öğrenme (Rassal Orman, Lojistik Regresyon, K-En Yakın Komşu, Karar Ağacı, AdaBoost, XGBoost, Ekstra Ağaç Sınıflandırıcı) algoritmaları kullanılmış ve model performansları karşılaştırılmıştır. Her bir model için başarı kriterleri incelenerek bu veri seti için en yüksek performans gösteren modellerin ağaç-bazlı Rassal Orman, Ekstra Ağaç Sınıflandırıcı ve XGBoost olduğu görülmüştür.

Anahtar Kelimeler: Müşteri kaybı, makine öğrenmesi, kayıp tahmini, ödemeli tv, müşteri tutma.

Bilim Dalı Sayısal Kodu: 92404

Tüm süreç boyunca desteğini esirgemeyen

eşime ve aileme ithaf ediyorum.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my thesis professor, Asst. Prof. Tuna Çakar, who contributed greatly to the completion of this study with his valuable guidance and knowledge.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

NN	: Neural Networks
LLM	: Logit Leaf Model
KNCn	: K-neighbors with the variants of centroids
KNCa	: K-neighbors with principal component analysis
SVC	: Support Vector Classification
SVCL	: SVC with its linear variant
NuSVC	: Nu-support vector classification
DT	: Decision Trees
OCP	: One Class Predictor
GBC	: DT with boosting based in gradients
HGBC	: DT with boosting based in histograms
LR	: Logistic Regression
LRVC	: Logistic Regression with crossvalidation
SVM	: Support Vector Machine
SVM rbf	: SVM with a radial basis function kernel
RF	: Random Forest
KNN	: K Nearest Neighbors
CRM	: Customer Relationship Management
CHAID	: Chi-square Automatic Interaction Detector
ML	: Machine Learning
EDA	: Exploratory Data Analysis
FCM	: Fuzzy c – Means
MAE	: Mean Absolute Error
MSE	: Mean Squared Error
RMSE	: Root Mean Squared Error
R^2	: R-Squared
AU-ROC	: Area Under the Receiver Operating Characteristic curve
TP	: True Positive
FP	: False Positive
FN	: False Negative

TN	: True Negative
FPR	: False Positive Rate
TPR	: True Positive Rate

INTRODUCTION

Problem to be Solved

The proliferation of new digital services and alternatives to traditional Pay-TV services is increasing competition in the Pay-TV industry. In today's world where competition is high, keeping customers, avoiding the cost of customer churn, focusing on the right customer group is an issue that many companies are looking for solutions. It is an issue that needs to be examined and researched for the Pay-TV industry as well.

The Aim of the Thesis

This thesis aims to reveal customer value and potential customer churn using data science and data analytics methods.

The Importance of the Thesis

By identifying the customers most likely to terminate their subscriptions beforehand, it will enable companies to take measures to retain these customers. They will be able to direct their marketing strategies according to these estimates.

Limitations of the Study

In this study, Pay-TV company serving in Turkey 6-month data of customers will be used for inactive customers, the 6-month data before the inactive date will be used.

Thesis Outline

In this section, the content of the thesis will be summarized.

This thesis starts by talking about the problem we want to solve, the purpose and importance of the thesis. Other parts are shared below.

Literature Review touches on the definitions of customer churn in the literature and the advantages it offers to companies. It also examines the machine learning methods used in previous research to predict customer loss.

Methodology theoretically explains the methods used in customer churn prediction. The stages of churn prediction modelling and machine learning methods are included.

Model Development and Validation includes examining the data used in the research, applying the selected machine learning methods to the data model, and evaluating their performance and accuracy.

Discussion and Conclusion shares our findings are discussed and our final conclusion of thesis.

LITERATURE REVİEW

What is Churn and Churn Prediction?

Churn definition is “When a customer quits using a service or cancels their subscription” [2]. In the telecom sector, Umayaparvathi and Iyakutti described churn as customers leaving a company due to dissatisfaction with services or better offerings from competitors within an affordable price range [27].

“Concretely, customer churn prediction is the practice of assigning a churn probability to each customer in the company database according to a predicted relationship between that customer’s historical information and its future churning behavior. Practically, the probability to end the relationship with the company is then used to rank the customers from most to least likely to churn, and customers with the highest propensity to churn receive marketing retention campaigns” [5].

Types of Churn

Types of Customer Churn can be grouped under two different headings as “voluntary churn” and “involuntary churn". Voluntary Churn is when customers willingly leave the existing business and prefer to benefit from the services of other businesses. Involuntary Churn, on the other hand, occurs due to reasons beyond the control of the customer and usually due to environmental conditions [6].

Based on customer purchase behavior, four types of business settings are mentioned. These are:

- Contractual and discrete business
- Contractual and continuous business
- Non-contractual and discrete business
- Non-contractual and continuous business

Those business type and business models have a significant impact on customers' behavior and lifecycles, which in turn affects the likelihood of customer churn [26].

The Cost of Churn

Customer retention is one of the foundational element for a subscription-oriented business model [1]. Previous academic research shows that acquiring a new customer is 6 times more expensive than retaining an existing customer [3]. Telephone service providers, internet service providers, Pay-TV companies, insurance companies, and alarm monitoring services utilize customer churn prediction to forecast profitability, recognizing the lower cost associated with retaining existing customers compared to acquiring new ones [4].

Data Mining Approach

While conventional research methods such as surveys and focus groups have been used in the past to investigate customer churn, data analytics and data science approaches have proven to be more efficient and effective solutions. The development of customer churn prediction models enables the identification of factors leading to customer loss and the anticipation of churn events. Data-driven models can be optimized to achieve higher levels of prediction accuracy. Customer segmentation is often combined with churn prediction to enhance management effectiveness. Decision trees and logistic regression are widely employed data science methods known for their prediction accuracy and interpretability. However, these methods have certain limitations. Decision trees struggle to manage linear relationships among variables, while logistic regression faces difficulties in handling interaction effects between variables. An alternative approach is the logit leaf model (LLM), which has shown higher levels of performance and interpretability compared to decision trees and logistic regression [11].

Related Works

Lopez et al. examine existing customer churn prediction models used by telecommunication companies and adapt them to Pay-TV in their research. Unlike previous studies, the data set they used in their study does not contain any personal metrics. Customer churn prediction was carried out by comparing different algorithms. Specifically models were applied that; neural networks (NN), K-neighbors with the variants of centroids (KNCn) and with principal component analysis (KNCa), support vector classification (SVC) and also SVC with its linear variant (SVCL) and Nu-support vector classification (NuSVC), one class predictor (OCP), decision trees (DT) and DT with boosting based in gradients (GBC) and histograms (HGBC), and finally logistic regression (LR), and LR with crossvalidation (LRCV). To address class imbalance, oversampling and undersampling methods were employed [7].

In a research conducted by Ulku and co-authors in 2021, the most suitable algorithm has been decided for predicting customer churn. Real data from a Pay-TV company in Turkey was used, and two datasets were prepared —one including all attributes and the other with selected attributes. Weka software, employing various data classification algorithms. The best model was proposed to the company, enabling targeted campaigns for the appropriate customer group [8].

Hadden et al. proposed a comprehensive five-step model for developing a customer churn management framework [9]. The steps involve identifying appropriate data, understanding data semantics, feature selection, developing a predictive model, and validating results. The selection of relevant data serves as the first step in establishing a churn management framework, followed by comprehending the context of the data. Feature selection plays a crucial role in identifying the best variables for developing a predictive model. The predictive model utilizes patterns discovered in the database to forecast future values. Finally, the results are often validated through methods such as cross-fold validation [9].

Simon and Martin in their study have presented churn prediction using machine learning in mobile telco. They have assessed a number of algorithms based on churn prediction performance experimentally. Using linear SVM (SVM linear), SVM with a radial basis function kernel (SVM rbf), random forest (RF), K-nearest neighbors (KNN) and logistic regression (LR) algorithms were done experiments. According to results after feature selection and parameter optimization, Random forest and K-NN learners improve results. Their results show that CRM features like contract info more effective in churn detection than traffic features. By the way, they report that it is necessary to be more careful in the period when the contract expiration date is approaching [10].

Bach et al have preferred a hybrid methodology when conducting churn analysis in the telecommunications industry. They have developed analysis of combined cluster and decision tree. An analysis of k-means clustering was utilized to identify the clusters with the highest level of loss, while a chi-square automatic interaction detector (CHAID) decision tree analysis was used to identify the premier characteristics of customers in the clusters with the highest loss level. They state that this approach will be beneficial for the downward migration prediction [13].

1. METHODOLOGY

1.1 Machine Learning Methods

ScienceSoft's Alex Bekker highlights the importance of machine learning for proactive churn management. He states that machine learning algorithms can do a good job of identifying potential customers to churn. Machine learning algorithms can predict potential churners based on some common behavioral patterns of customers who will leave the company [1].

1.1.1 Logistic Regression

Regression is among the statistical methods used to predict how is the relationship between variables and each other. Logistic regression is among the types of regression analysis. It is used to estimate the probability of something happening. While linear regression models are valuable for estimating continuous data, Logistic Regression is a probabilistic statistical classification model. It is also used for binary classification or binary estimation of a categorical value that depends on one or more parameters. The algorithm can be used in the case of customer churn [14].

The predicted output of the logistic regression is in the expression below. While Logistic function return a value between 0 and 1, input variables takes a value between -∞ and +∞ [23].

$$L_i = In\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta X_i + u_i$$

1.1.2 Decision Tree

Decision Tree is a supervised learning technique. Although it is generally used in classification problems, it is used in both classification and regression problems. It begins with the root node, which spreads further branches and forms a tree-like structure [12].

1.1.3 KNN (K-Nearest Neighbors)

KNN is one of the classification techniques. It only needs a k parameter, a labeled training set examples, and a metric measure to determine distances in an n-dimensional space. Mehmed Kantardzic states that the kNN classification process generally relies on the following steps [15]:

- k — number of nearest neighbors parameter is determined
- The distance between each testing sample and all the training samples is calculated.
- The distance is sorted and nearest neighbors based on the k - th threshold is determined.
- The category (class) for each of the nearest neighbors is determined.
- The simple majority of the category of nearest neighbors is used as the prediction value of the testing sample classification.

1.1.4 Random Forest

Random forests is a traditional classifier based on a random subspace method. It includes various tree structures and randomly combines decision tree subspaces and bagging elements [14, 16].

Breiman defines random forests as: A random forest is a classifier consisting of a collection of tree_structured classifiers h(x, k), k=1, ... where the k are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x [17].

This structure can be used for binary classification problems, such as customer predicting customer purchase behavior or customer retention [16].

1.1.5 XG Boost

XGBoost is a supervised classification machine learning algorithm. The top reason for using XGBoost is execution speed and model performance. It is applied for tabular or structured datasets. XGBoost produces output as a single pattern using a combination of different algorithms. It also uses gradient descent to locate the minima or reduce the value of loss function. It supports parallel and distributed computing while offering efficient memory usage [14, 16, 18].

1.1.6 Extra Tree Classifier

The Extra Trees classifier is a machine learning technique that shares likenesses with the Random Forest technique. In the Extra Trees classifier, the default behavior is to create decision trees without using bootstrapping, which means that each tree is trained on the entire dataset. Additionally, the Extra Trees classifier differs from traditional decision trees in the way it splits nodes [28].

Unlike standard decision trees that search the optimal split according to certain criteria, such as Gini impurity or information gain, the Extra Trees classifier selects splits randomly. This random splitting process introduces additional randomness into the tree construction, further diversifying the ensemble of trees. By doing so, the Extra Trees classifier aims to reduce overfitting and increase robustness, particularly when dealing with noisy or high-dimensional datasets.

In our study, the Extra Trees classifier is utilized as a machine learning algorithm for predicting customer churn in the Pay-TV industry. The unique characteristics of the Extra Trees classifier, such as its lack of bootstrapping and random splitting, make it suitable for handling diverse and complex datasets commonly encountered in churn prediction tasks.

1.1.7 AdaBoost

AdaBoost is short for Adaptive Boosting. It is a machine learning ensemble method that unites a series of weak classifiers to construct a high-performance prediction rule. It operates by consecutively applying a classification algorithm to reweighted training data versions. The process involves assigning higher weights to instances that are misclassified by the previous classifiers, thereby focusing on the more challenging samples. The final prediction is obtained by taking a weighted plurality vote or averaging of the predictions made by the individual classifiers [29].

In our research, AdaBoost is employed as a machine learning algorithm for predicting customer churn in the Pay-TV industry. The strengths of AdaBoost lie in its ability to leverage multiple weak classifiers to create a strong ensemble model. By iteratively adjusting the weights of the training data, AdaBoost focuses on the difficult instances, thereby improving the overall predictive performance.

1.2 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a essential stage in data-driven analytics, where visualization techniques play a pivotal role in summarizing and interpreting data. By exploring and visualizing the rows and columns of data, EDA aims to uncover hidden patterns, relationships, and insights that can inform the subsequent stages of supervised and unsupervised machine learning modeling [14, 15].

Visualization serves as a powerful tool in EDA, allowing researchers and analysts to obtain an in-depth understanding of the dataset. Complex relationships between variables can be revealed through graphical representations, such as scatter plots, histograms, box plots, and heat maps. Visualizing the data helps in identifying outliers, understanding the distribution of variables, detecting trends, and identifying potential patterns that may inform the modeling process.

The iterative loop of EDA involves an interactive and iterative process of exploration and analysis. Researchers iteratively examine the data, generate visualizations, and draw insights, leading to further refinements and adjustments in the analysis. This iterative approach allows for an in-depth understanding of the data and the ability to uncover meaningful patterns that may have been initially overlooked [21].

In our study, EDA plays a fundamental role in gaining insights into the customer churn dataset in the Pay-TV industry. By employing various visualization techniques, we aim to identify relevant features, assess data quality, and detect any anomalies or patterns that may influence customer churn. Through EDA, we can make informed decisions about data preprocessing, feature selection, and the selection of appropriate machine learning algorithms.

Furthermore, EDA assists in establishing a solid foundation for subsequent modeling stages. By understanding the data characteristics, distributions, and relationships between variables, we can make informed decisions about feature engineering, data transformations, and model selection. EDA provides valuable insights that guide the development of accurate and robust customer churn prediction models.

1.3 Feature Engineering

Feature engineering is a crucial method in data processing that plays a vital role in developing the performance of machine learning methods. It involves transforming or creating new features from the original dataset to better align with the requirements and characteristics of the machine learning algorithm being employed. When applied effectively, feature engineering not only enhances the algorithm's performance as well as facilitates the exploration of valuable insights [19, 20].

In the context of our study on predicting customer churn in the Pay-TV industry, feature engineering is a critical step in preparing the dataset for modeling. By carefully selecting, transforming, and creating features, we aim to improve the predictive power of our models and uncover significant patterns and relationships related to customer churn. By applying these feature engineering techniques, we aim to create a refined dataset that maximizes the predictive power of the machine learning models and facilitates the extraction of actionable insights related to customer churn.

In summary, feature engineering is a crucial phase in data processing that enhances the performance of machine learning algorithms. By transforming, creating, and selecting features, we tailor the dataset to better align with the requirements of the algorithms and uncover valuable insights. In our study, feature engineering plays a pivotal role in Pay-TV industry's customer churn prediction leading to accurate models and actionable recommendations for customer retention.

1.4 Performance Metrics

Performance metrics are an integral component of every machine learning process. Regression models have continuous output. Following metrics could be used for the regression models [22]:

- MAE (Mean Absolute Error),
- MSE (Mean Squared Error),
- RMSE (Root Mean Squared Error),
- R^2 (R-Squared)

Following metrics could be used for Classification Models [22]:

- Accuracy
- Confusion Matrix (not a metric but fundamental to others)
- Precision and Recall
- F1-score
- AU-ROC

The confusion matrix succinctly encapsulates the outcomes of evaluating the performance of a machine learning classification model by contrasting the predicted values against the true values [23].

Table 1.1: The confusion matrix applies to a binary classification model

Confusion Matrix		**Actual Values**	
		0	**1**
Predicted Values	**0**	True Positive(TP)	False Positive(FP)
	1	False Negative(FN)	True Negative(TN)

TP (True Positive); indicates that the sample with a true class value of 1 is correctly predicted as 1.

TN (True Negative); indicates the situation where the sample with a true class value of 0 is correctly predicted as 0.

FN (False Negative); indicates that the sample with a true class value of 1 is incorrectly evaluated as 0 as a result of the prediction.

FP (False Positive); indicates that the sample with a true class value of 0 is incorrectly evaluated as 1 as a result of the prediction [23, 24].

Classification Models are computed as [22, 24, 25]

Accuracy refers to the ratio of the number of correct prediction samples to the total number of samples.

$$Accuracy = \frac{TP + TN}{TN + FP + FN}$$

Precision refers to the ratio of true positive samples to the samples predicted to be positive.

$$Precision = \frac{TP}{TP + FP}$$

Recall; is essentially the ratio of true positives to the number of all positive samples

$$Recall = \frac{TP}{TP + FN}$$

F1 – score; is calculated by using precision and recall values. Actually, the F1 score is the harmonic mean of the two.

$$F1 - score = \frac{2. Recall. Precision}{Recall + Precision}$$

AUC-ROC stands for Area Under the Curve of the Receiver Operating Characteristic. It quantifies the total area beneath the ROC curve, which is constructed using False Positive Rate (FPR) and True Positive Rate (TPR) values derived from the classification model. The AUC value spans from 0 to 1, with a value closer to 1 signifying enhanced model accuracy.

$$AUC = \int_{0}^{1} \quad TPR(t_i)dFRP(t_i)$$

where TPR(ti) and FRP(ti) show the true positive rate and the false positive rate .

2. MODEL DEVELOPMENT AND VALIDATION

2.1 Dataset

The data set includes the last 6 months subscription-based detail and summary data of a Pay-Tv company customers. This data set was obtained from various data sources in the company database. It contains many information such as Subscription, Commitment, Package & Tariff, Invoice fees, Communication Permit, Cancellation and Failure records. Individual customers are taken into account.

2.2 Exploratory Data Analysis

In this stage, the relationship between categorical variables and churn was examined.

It seems in Figure 2.1 that while 15.8% of the dataset appears to be Churn, the remaining 84.2% of the customers seem active.

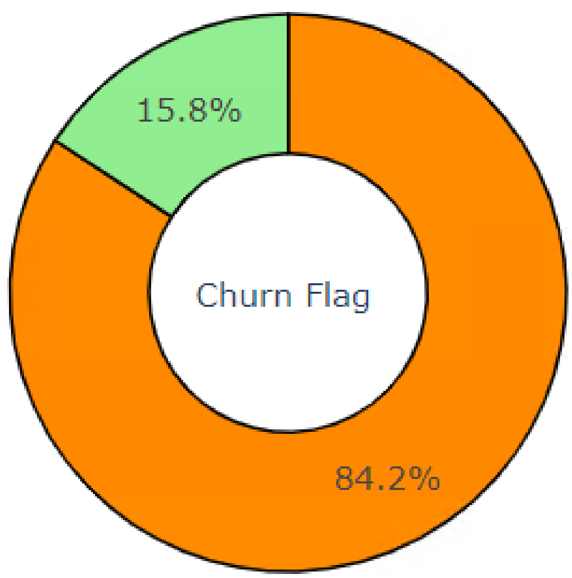

Figure 2.1: The Churn Distrubution

The following figure shows Pay-Tv Main Product distribution. The company has four main Tv package type.

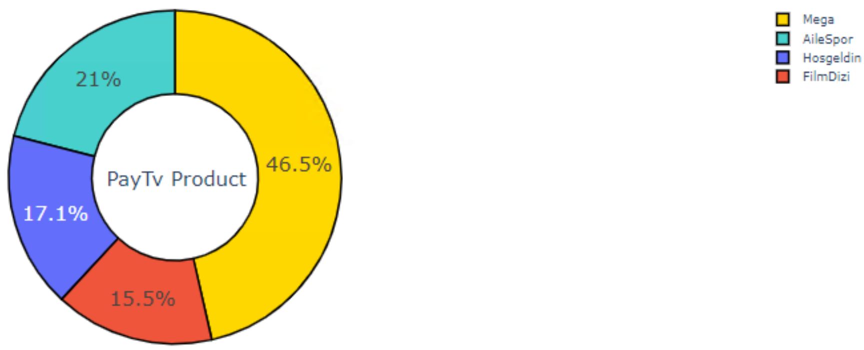

Figure 2.2: PayTv Products Distrubutions

Figure 2.3 shows the distribution of Paytv Products with churn. “Mega” subsription has a probability of 15.87 % churn. “Hosgeldin” Subscription has a probability of 16.66 % churn “FilmDizi” subsription has a probability of 14.34 % churn. “AileSpor” subscription has a probability of 16.0 % churn.

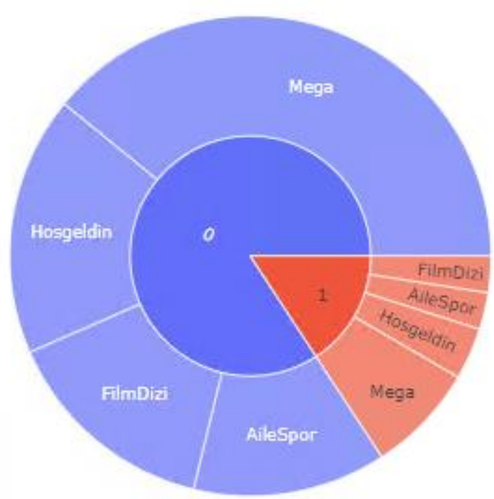

Figure 2.3: PayTv Products and Churn Distributions

The following figure shows Commitment Status and Churn distributions.

- “Taahhüt İçi” has a probability of 10.99 % churn
- “Taahhütsüz” has a probability of 19.06 % churn
- “Taahhüt +30” has a probability of 76.67 % churn
- “Taahhüt +60” has a probability of 42.32 % churn
- “Taahhüt +90” has a probability of 34.87 % churn
- “Taahhüt -30” has a probability of 38.48 % churn
- “Taahhüt -60” has a probability of 17.31 % churn
- “Taahhüt -90” has a probability of 10.1 % churn

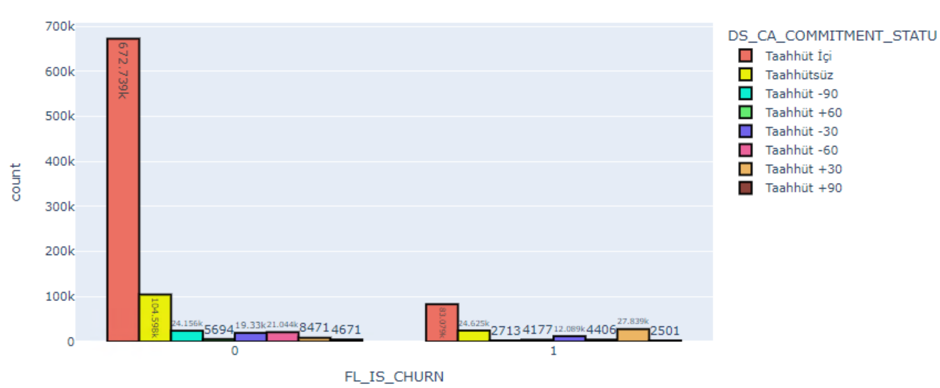

Figure 2.4: Commitment Status and Churn Distributions

Subscription Type and Churn distributions graph is shown below. Subscription of “Tv + Go + Internet” has a probability of 17.13 % churn. “Subscription of Tv + Go” has a probability of 15.26 % churn. “Subscription of Tv” has a probability of 3.77 % churn.

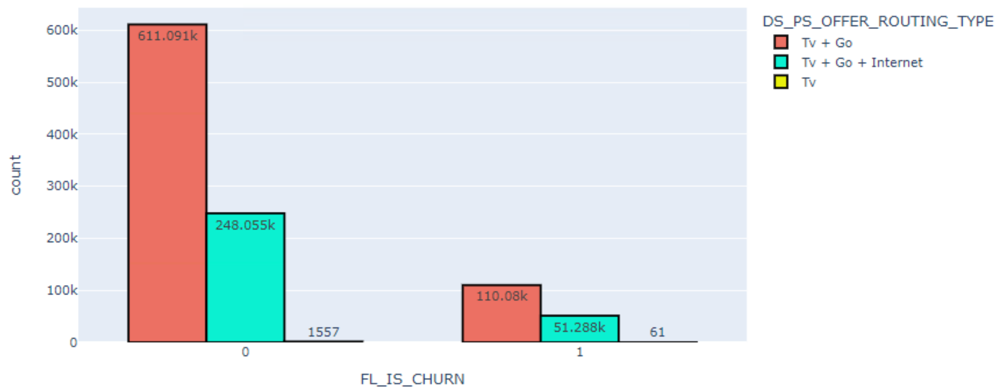

Figure 2.5: Subscription Type and Churn Distribution

FL_ZERO_PAID signs that Subscriptions that have not been paid since the moment they subscribed. In the FL_ZERO_PAID, a value of 0 corresponds to “No”, and a value of 1 corresponds to “Yes”. Figure 2.6 shows the distribution of subscriptions that do not pay at all. Zero Paid Subscriptions has a probability of 32.99 % churn. No Zero Paid Subscription has a probability of 15.23 % churn.

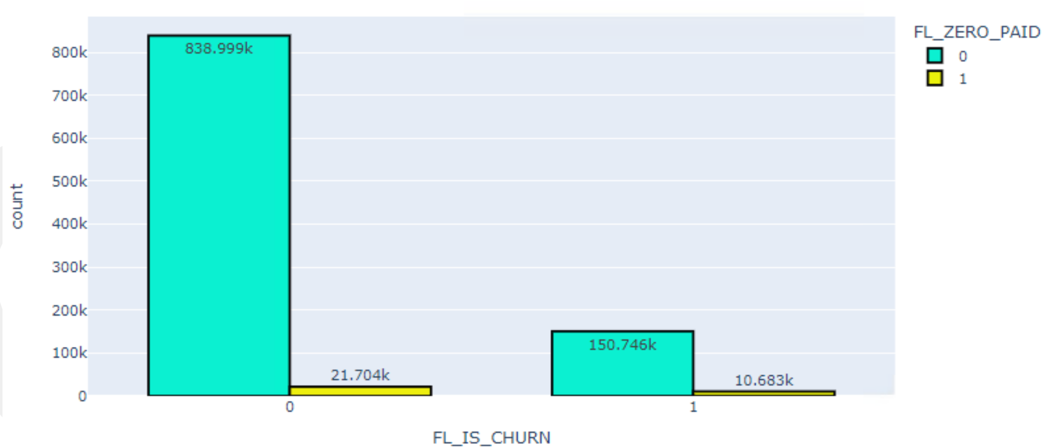

Figure 2.6: Zeroers and Churn Distribution

FL_PAID_IN_DUE flags that Subscriptions have paid their bills on or before the due date in the last 12 months. In the FL_PAID_IN_DUE, a value of 0 corresponds to “Yes”, and a value of 1 corresponds to “No”. When Figure 2.7 is analyzed, it shows that;

- Regular Payers has a probability of 9.88 % churn.
- Non Regular Payers has a probability of 18.47 % churn.

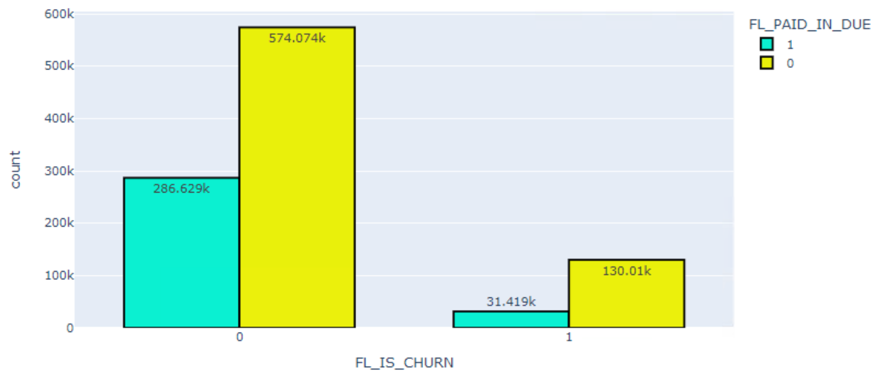

Figure 2.7: Regular Payers and Churn Distribution

Paid After Due Flag and Churn distribution graph is shown below. Fl_Paid_After_Due flags subscriptions that have been paid once in the last 12 months. Figure 2.8 shows;

- Paid After Due Subscriptions has a probability of 15.44 % churn.
- Not Paid After Due Subscriptions has a probability of 16.37 % churn.

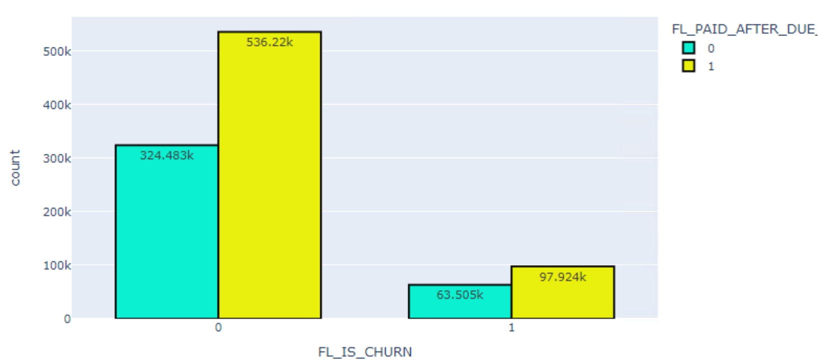

Figure 2.8: Paid After Due Flag and Churn Distributions

Fl_Paid_After_Suspend_For_Debt keeps subscriptions that were closed no more than 2 times in the last 12 months. The figure below shows the distribution of this variable with churn. In the feature, a value of 0 corresponds to “No”, and a value of 1 corresponds to “Yes”. Subscriptions Paid After Debt Suspension has a probability of 19.86 % churn. Non Subscriptions Paid After Debt Suspension has a probability of 15.57 % churn.

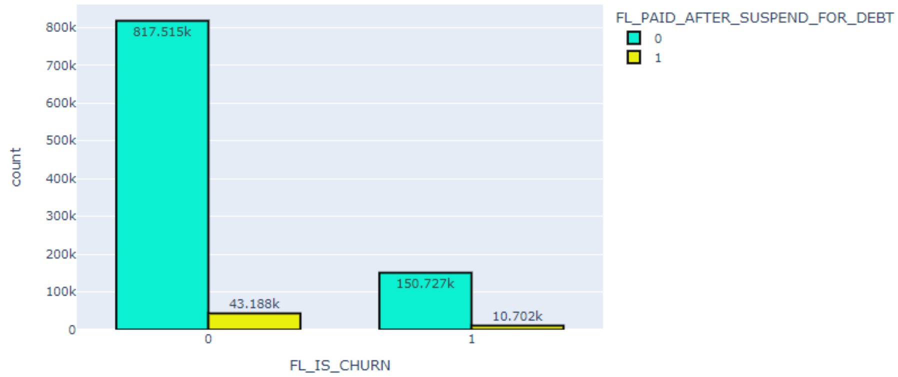

Figure 2.9: Subscriptions Paid After Debt Suspension and Churn Distributions

Ds_Block_Call’s value of 0 corresponds to “No”, and a value of 1 corresponds to “Yes”. According to Figure 2.10, Those who have permission to call have a probability of 13.79 % churn. Those who do not have permission to call has a probability of 30.08 % churn.

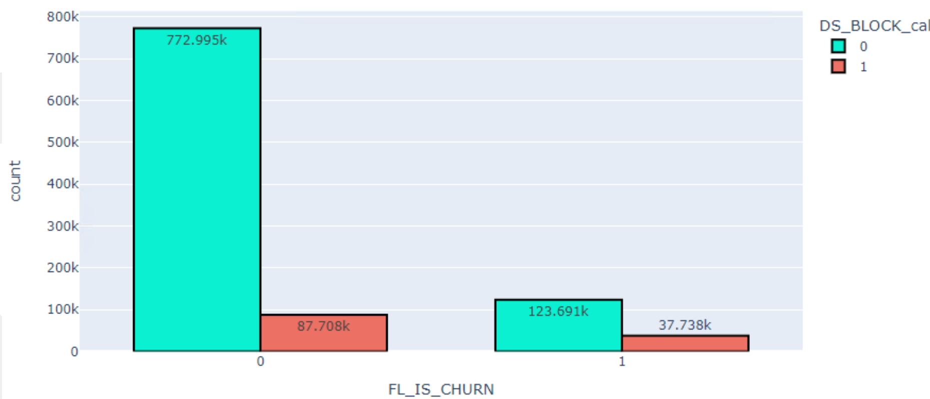

Figure 2.10: Call Permission for Communication and Churn Distributions

Figure 2.11 seems that those who have permission to sms communication have a probability of 15.43 % churn. Those who do not have permission to sms communication have a probability of 20.17 % churn. 0 is ‘No’, 1 is ‘Yes’.

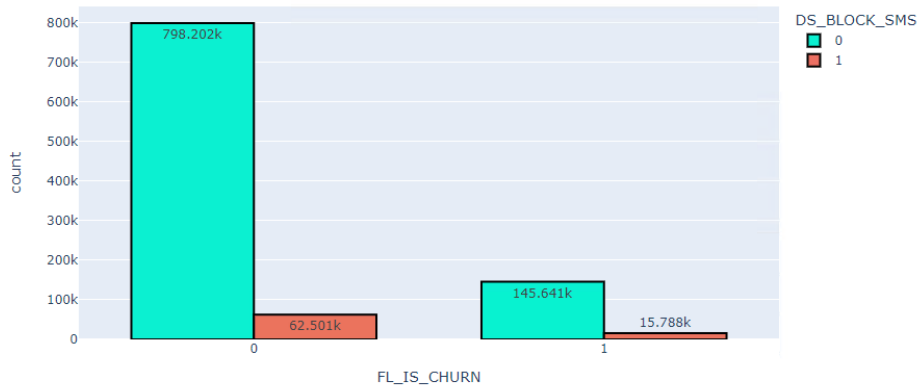

Figure 2.11: SMS Communication Permission and Churn Distributions

Those who have permission to e-mail communication have a probability of 15.5 % churn. Those who do not have permission to e-mail communication has a probability of 19.82 % churn. 0 is ‘No, 1 is ‘Yes.

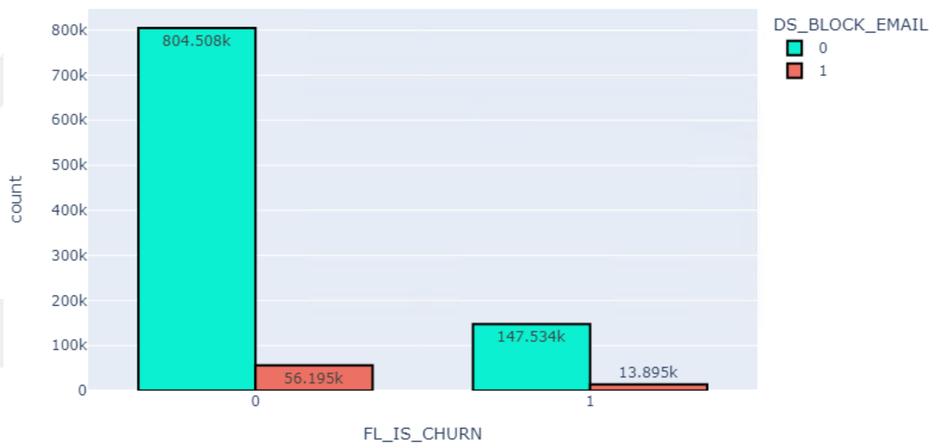

Figure 2.12: E-mail Communication Permission and Churn Distributions

Figure 2.13 seems that those who have permission to b-mail communication have a probability of 15.48 % churn. Those who do not have permission to b-mail communication have a probability of 20.07 % churn. 0 is ‘No’, 1 is ‘Yes’.

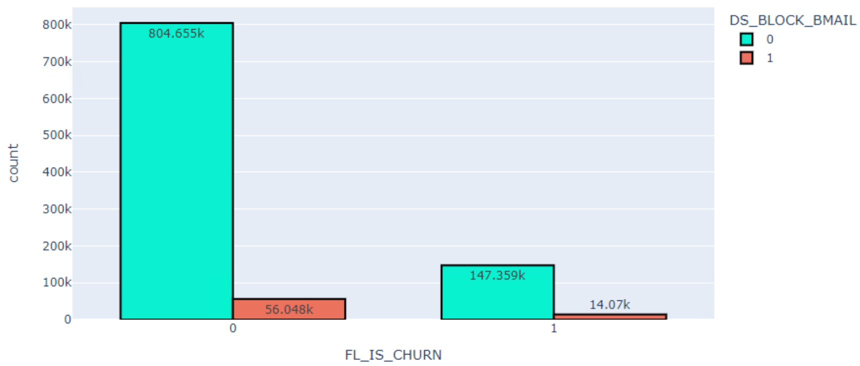

Figure 2.13: B-mail Communication Permission and Churn Distributions

2.3 Model Preparation

In this part, the application of the selected machine learning methods to the data model, the preparations made beforehand, the assessment of the performance and accuracy of these models, as well as the performance improvement studies are mentioned. All these studies were implemented using the Python programming language. Pandas, numpy, scikit-learn, plotly, xgboost and imblearn python libraries were used.

2.3.1 Categorical Variables Encoding

At this stage, the categorical data was converted into a suitable format for the machine learning model.

2.3.2 Missing Values

Numeric values from null are filled with 0.

```
data['MT_DAY_COUNT_PASSED_FROM_LAST_COMMITMENT_RENEWAL'].fillna(0,inplace=True)
data['MT_DAY_COUNT_REMAINED_FOR_COMMITMENT_RENEWAL'].fillna(0,inplace=True)
data['MT_INVOICE_AMOUNT_AVG3'].fillna(0,inplace=True)
data['MT_COUNT_CANCEL_REQUEST1'].fillna(0,inplace=True)
data['MT_COUNT_CANCEL_REQUEST2'].fillna(0,inplace=True)
```

Figure 2.14: Missing Data Transformation

Categorical values from null are defined as other.

```
data=df.drop(['DS_PS_OFFER_ROUTING_TYPE','DS_PRODUCT_PAYTV_T0','DS_CA_COMMITMENT_STATU','DS_CA_TENURE','MT_COUNT_COMMITMENT_F
data.head()
```

2]:

UE_1	...	Is_Fourth_Commitment_Renewal	Is_Five_Plus_Commitment_Renewal	Is_Nakil	Is_Other	Is_Single_Tv_Donusum	Is_Tv_To_Internet	Is_Yeni_Satis	Is_
1	...	0	0	0	0	0	0	1	
1	...	0	0	0	0	0	0	1	
0	...	0	0	0	0	0	0	1	
0	...	0	0	0	0	0	0	1	
0	...	0	0	0	0	0	0	1	

Figure 2.15: Missing Data Transformation for Categorical Variables

2.3.3 Dataset Splitting

The training and test data are split as follows.

```
In [23]: from sklearn.model_selection import train_test_split

In [24]: X=data.drop(['FL_IS_CHURN','ID_CU_CUSTOMER','ID_CA_CUSTOMER_ACCOUNT'], axis=1)
         y=data['FL_IS_CHURN']

In [25]: X_train, X_test, y_train, y_test=train_test_split(X,y, test_size=0.2,stratify=y, random_state=42)# y deki dağılımıca ayrımı s
```

Figure 2.16: Train-Test Split

2.3.4 Min-Max Scaling

In this section, numerical features are scaled.

```
In [39]: scaler=StandardScaler()

         X_train[['MT_DAY_COUNT_PASSED_FROM_LAST_COMMITMENT_RENEWAL','MT_DAY_COUNT_REMAINED_FOR_COMMITMENT_RENEWAL','MT_INVOICE_AMOUNT
         X_test[['MT_DAY_COUNT_PASSED_FROM_LAST_COMMITMENT_RENEWAL','MT_DAY_COUNT_REMAINED_FOR_COMMITMENT_RENEWAL','MT_INVOICE_AMOUNT_
```

Figure 2.17: Train-Test Split

2.4 Oversampling

As shown in Figure 2, we have an unbalanced data set. In addition, we see that although the accuracy values of the machine learning techniques we tried are high, their recall values are lower. Since the number of observations in the minority class was not sufficient, data were created using the Borderline-SMOTE technique. After this technique was applied to the entire data set, it was separated as train and test data.

```
In [54]: from imblearn.over_sampling import BorderlineSMOTE
```

```
In [55]: oversample = BorderlineSMOTE()
         X_train, y_train = oversample.fit_resample(X_train, y_train)
```

Figure 2.18: Borderline-SMOTE

2.5 Feature Selection

In this section, a feature selection technique was used to identify unnecessary and low-contributing features.

```
In [37]: from sklearn.feature_selection import RFECV
         from sklearn.linear_model import LogisticRegression
         from sklearn.model_selection import StratifiedKFold
         from sklearn.ensemble import RandomForestClassifier
         from sklearn.feature_selection import RFE

         min_features_to_select = 1  # Minimum number of features to consider
         clf = RandomForestClassifier()
         cv = StratifiedKFold(5)

         rfecv = RFECV(
             estimator=clf,
             step=1,
             cv=cv,
             scoring="accuracy",
             min_features_to_select=min_features_to_select,
             n_jobs=-1,
         )
         rfecv.fit(X_train.values, y_train)

         print(f"Optimal number of features: {rfecv.n_features_}")

         Optimal number of features: 51
```

Figure 2.19: Optimal Number of Feature

```
In [38]: import matplotlib.pyplot as plt

n_scores = len(rfecv.cv_results_["mean_test_score"])
plt.figure()
plt.xlabel("Number of features selected")
plt.ylabel("Mean test accuracy")
plt.errorbar(
    range(min_features_to_select, n_scores + min_features_to_select),
    rfecv.cv_results_["mean_test_score"],
    yerr=rfecv.cv_results_["std_test_score"],
)
plt.title("Recursive Feature Elimination \nwith correlated features")
plt.show()
```

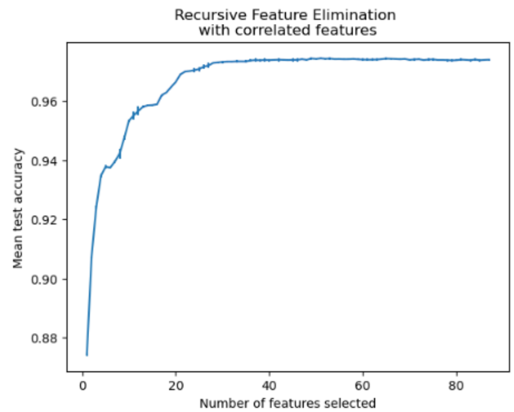

Figure 2.20: Recursive Feature Elimination with Correlated Features

2.6 Parameter Optimization

We re-evaluate the results by performing parameter optimization for the Random Forest model, which is one of the best-performing models of the machine learning techniques we apply.

```
In [ ]: RF_S = RandomForestClassifier(random_state = 42)
params_RF = {'n_estimators': list(range(150,451,100)),
             'min_samples_leaf': list(range(5,10)),
             'min_samples_split': list(range(5,10)),
             "criterion" : ["gini", "entropy", "log_loss"],
             "max_depth" : list(range(6,12)),
            # "max_features" : ["sqrt", "Log2"],
             "class_weight": ["balanced", "balanced_subsample"]
             }
grid_RF = RandomizedSearchCV(RF_S, param_distributions=params_RF, cv=10, n_jobs=-1, n_iter=50, random_state=42, return_train_scoi
grid_RF.fit(X_train, y_train)
print('Best parameters:', grid_RF.best_estimator_)

Best parameters: RandomForestClassifier(class_weight='balanced_subsample', max_depth=11, min_samples_leaf=7, min_samples_split=9,
n_estimators=250, random_state=42)
```

Figure 2.21: Parameter Optimization for Random Forest Model

2.7 Model Performances

Random Forest, Logistic Regression, K-Nearest Neighbors, Decision Tree, AdaBoost, XGBoost, Extra Tree Classifier algorithms were run for 5 different models according to scaling method and feature selection.

The models and applied methods are listed below.

Model 1:

- Variables determined by Feature selection were used
- Used the Borderline-SMOTE technique for train dataset
- Min Max scaling was implemented separately for training, testing and validation data

Model 2:

- All variables were used in the model
- Used the Borderline-SMOTE technique for train dataset
- Standard Scaler was applied to testing and validation data based on training data

Model 3:

- All variables were used in the model
- Used the Borderline-SMOTE technique for train dataset
- Min Max scaling was applied to testing and validation data based on training data

Model 4:

- All variables were used in the model
- Used the Borderline-SMOTE technique for train dataset
- Min Max scaling was implemented separately for training, testing and validation data

Model 5:

- Variables determined by Feature selection were used
- Used the Borderline-SMOTE technique for train dataset
- Min Max scaling was applied to testing and validation data based on training data

Random Forest, Logistic Regression, K-Nearest Neighbors, Decision Tree, AdaBoost, XGBoost, Extra Tree Classifier models were fitted with the training data and their performances were compared. The performance results of the machine learning algorithms applied in the research for 5 models are shown in tables.

Table 2.1: Churn Prediction Results of Model 1

Model 1				
ML Algorithms	**Precision**	**Recall**	**f1-score**	**Accuracy**
Logistic Regression	0.74	0.92	0.82	0.94
Ada Boost	0.7	0.95	0.81	0.93
KNN	0.78	0.89	0.83	0.94
Decision Tree	0.47	0.85	0.61	0.83
Extra Tree Classifier	0.98	0.99	0.98	0.98
XG Boost	0.4	0.96	0.57	0.77
Random Forest	0.71	0.92	0.8	0.93
Random Forest (parameter optimization)	0.67	0.98	0.8	0.92

Table 2.2: Churn Prediction Results of Model 2

Model 2				
ML Algorithms	**Precision**	**Recall**	**f1-score**	**Accuracy**
Logistic Regression	0.69	0.95	0.8	0.93
Ada Boost	0.71	0.95	0.82	0.93
KNN	0.75	0.94	0.84	0.94
Decision Tree	0.86	0.89	0.87	0.96
Extra Tree Classifier	0.88	0.93	0.91	0.97
XG Boost	0.89	0.94	0.91	0.97
Random Forest	0.88	0.95	0.92	0.97

Table 2.3: Churn Prediction Results of Model 3

Model 3				
ML Algorithms	**Precision**	**Recall**	**f1-score**	**Accuracy**
Logistic Regression	0.71	0.96	0.82	0.93
Ada Boost	0.74	0.93	0.82	0.94
KNN	0.77	0.92	0.84	0.94
Decision Tree	0.86	0.88	0.87	0.96
Extra Tree Classifier	0.88	0.94	0.91	0.97
XG Boost	0.9	0.94	0.92	0.97
Random Forest	0.88	0.95	0.91	0.97

Table 2.4: Churn Prediction Results of Model 4

Model 4				
ML Algorithms	**Precision**	**Recall**	**f1-score**	**Accuracy**
Logistic Regression	0.71	0.95	0.81	0.93
Ada Boost	0.71	0.94	0.81	0.93
KNN	0.77	0.88	0.82	0.94
Decision Tree	0.19	0.84	0.31	0.4
Extra Tree Classifier	0.89	0.89	0.89	0.97
XG Boost	0.54	0.94	0.69	0.86
Random Forest	0.75	0.92	0.83	0.94

Table 2.5: Churn Prediction Results of Model 5

Model 5				
ML Algorithms	**Precision**	**Recall**	**f1-score**	**Accuracy**
Logistic Regression	0.67	0.96	0.79	0.92
Ada Boost	0.71	0.94	0.81	0.93
KNN	0.79	0.93	0.86	0.95
Decision Tree	0.86	0.89	0.87	0.96
Extra Tree Classifier	0.88	0.94	0.91	0.97
XG Boost	0.89	0.94	0.91	0.97
Random Forest	0.88	0.95	0.92	0.97
Random Forest(parameter opt)	0.7	0.99	0.82	0.93

When we compare the model performances, we obtain the following conclusions for this study.

- Oversampling, scaling and feature status do not cause differences in Logistic Regression results.
- Random forest gave the best performance in the 2nd, 3th and 5th models.
- Adaboost gives similar results in all models.
- KNN gives similar results in all models

- Decision Tree gives the best results in the 2nd, 3th and 5th models.
- Model performance is very low in the 4th model. 1th model results are also behind compared to the 2nd and 3th models.
- Extra Tree Classifier provides high performance in all models.2nd, 3th, 4th and 5th models have better validation data.
- There is no difference between 1th model and 5th model in random forest.

In addition, model performances were observed with validation data consisting of up-to-date data in the same data structure. Accuracy values were in the range of 0.72 to 0.96.

3. DISCUSSION

The competition within the Pay-TV sector drives companies to continually strive for advancement. Customer churn, which leads to revenue loss and increased marketing costs, poses a significant challenge. Therefore, understanding the reasons behind customer subscription termination is crucial for retaining existing customers. Developing an accurate customer churn model that effectively utilizes customer data can give companies a competitive edge in this race [2, 1, 8].

Extensive literature studies have demonstrated the superior effectiveness of data science methods compared to traditional approaches in customer churn prediction [27, 3, 5]. These data-driven techniques provide higher accuracy and improved predictive capabilities, making them invaluable tools for businesses in the Pay-TV industry [14, 17].

In this project, a diverse range of variables extracted from various aspects such as Subscription, Product, and Invoice were utilized, leveraging the company's comprehensive data. The inclusion of features related to Cancellation Requests was found to be particularly significant in enhancing the predictive performance of the churn model [11, 8]. Furthermore, prior customer segmentation efforts play a pivotal role in enhancing the success of the churn estimation model. Segmentation enables a better understanding of the distinct customer groups, allowing for more targeted retention strategies and more accurate churn predictions [11, 14, 16].

To further augment the study's effectiveness, additional variables such as Sales, Address, and customer demographic information could be incorporated. These variables have been shown to provide valuable insights into customer behavior and enhance churn prediction models [8, 23]. Additionally, the utilization of the Fuzzy c-means method, known for its high success rate in handling imbalanced datasets, could further enhance the performance of the churn prediction model [22].

In summary, the competition within the Pay-TV sector necessitates effective strategies for addressing customer churn. Data science methods have proven to be highly effective in this regard, enabling accurate churn prediction and retention efforts.

Leveraging a diverse range of variables and incorporating customer segmentation techniques can further improve the success of churn prediction models. Additionally, enriching the models with additional relevant variables and utilizing advanced techniques like Fuzzy c-means can contribute to even better outcomes [20, 21].

CONCLUSIONS AND FURTHER WORK

In this study, various machine learning algorithms commonly used in churn prediction, including Random Forest, Logistic Regression, KNN, Decision Tree, Ada Boost, XGBoost, and Extra Tree Classifier, were employed to develop predictive models. Among these algorithms, the Random Forest, Extra Tree Classifier and XGBoost algorithms gave the best results when the recall and f1-score values were taken into account as well as the accuracy value [11, 14, 16]. When examining the study conducted by Ülkü et al., we observe that they experimented with various data classification algorithms using the Weka application on two separate datasets to find the most suitable dataset and algorithm. They suggested the best dataset using a decision tree-based algorithm. However, in this study, we proceeded with a single dataset. Additionally, the algorithms were implemented using the Python programming language [8].

To address the issue of class imbalance in the dataset, the Borderline-SMOTE technique was applied, resulting in a more balanced dataset and improved recall values. It is important to note that achieving high accuracy alone does not guarantee the detection of churn cases at a higher rate [8, 15]. Based on the findings of this research, we recommend employing the Random Forest, Extra Tree Classifier and XGBoost methods for customer churn prediction in the Pay-TV industry. By proactively identifying customers who are likely to cancel their subscriptions, companies can implement targeted measures to retain these customers [8, 14]. This enables businesses to allocate their marketing resources more effectively and tailor their strategies to specific customer segments [11, 17].

It is important to recognize that customer churn prediction is a complex task influenced by various factors, and there is not only one solution that fits all. Further research and exploration of other advanced techniques, such as ensemble methods or deep learning, may provide additional insights and improve churn prediction performance [1, 18].

In conclusion, this study highlights the effectiveness of machine learning algorithms in churn prediction for the Pay-TV industry. The Random Forest, Extra Tree

Classifier and XGBoost algorithms stand out as a recommended approach, offering accurate predictions and valuable insights. By leveraging these predictions, companies can proactively engage with customers at risk of churn and apply targeted retention strategies, leading to enhanced customer retention and business success.

REFERENCES

[1] Customer Churn Prediction for Subscription Businesses Using Machine Learning: Main Approaches and Models. [https://www.altexsoft.com/blog/business/customer-churn-prediction-for-subscription-businesses-using-machine-learning-main-approaches-and-models/], (Last Accessed Date 01.10.2023)

[2] Gold, C. (2019). Fighting Churn with Data. New York, Manning Publications Co., pp.2.

[3] Bhattacharya, C. (1998). When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science 26 (1), 31–44.

[4] Sundararajan, A. and Gursoy K. (2020). Telecom customer churn prediction. Rutgers University.

[5] Coussement, K., Lessmann, S. and Verstraeten, G. (2017). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. Decision Support Systems, 95, 27–36.

[6] Şadi Evren ŞEKER, Müşteri Kayıp Analizi (Customer Churn Analysis),YBS Ansiklopedi, v. 3, is. 1, March, 2016, pp. 26 – 27.

[7] Lopez, V., Egea, R., Museros, L. and Sanz, I. (2021). A Customer Churn Detection Model for the Pay-TV Sector. Ios Press, Volume 339: Artificial Intelligence Research and Development, 159 - 166.

[8] Ulku, I., Uney, F., Yilmaz, O., Aktas, M. U., and Akbalik N. (2021). Churn Prediction in a Pay-TV Company via Data Classification. International Journal of Artificial Intelligence and Machine Learning,Volume 11, Issue 1, Jan, pp 39–53.

[9] Hadden, J., Tiwari, A., Roy, R., and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research, 34(10), 2902–2917.

[10] Brmez, S., & Znidarsic, M. (2019). A case of churn prediction in telecommunications industry. Ipsi Bgd Transactions on Internet Research, 15(2, SI).

[11] Zhang T., Moro S. and Ramos R. F. (2022). A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation. Future Internet 14(94):1-19.

[12] Decision Tree Classification Algorithm. [https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm], (Last Accessed Date 01.10.2023).

[13] Bach, M., Pivar J., Jakovi B. (2021). Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees. Journal of Risk and Financial Management 14(11):544.

[14] Lalwani, P., Mishra, M. K., Chadha, J. S. and Sethi, P. (2021). Customer churn prediction system: a machine learning approach. Computing, 1-24.

[15] Kantardzic, M. (2011). DATA MINING Concepts, Models, Methods, and Algorithms.United States of America, IEEE Press, pp.121.

[16] F. Ehsani (2022). Customer churn prediction from Internet banking transactions data using an ensemble meta-classifier.

[17] Blank, C. and Hermansson, T. (2018). A Machine Learning approach to churn prediction in a subscription based service.

[18] Kavitha, V., Kumar, H. G., Mohan Kumar, S. V. and Harish, M. (2020). Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms. International Journal of Engineering Research and V9(05), 181-184.

[19] Kim, S., Choi, D., Lee, E., & Rhee, W. (2017). Churn prediction of mobile and online casual games using play log data. PLOS ONE, 12(7), e0180735.

[20] Dai, D., Xu, T., Wei, X., Ding, G., Xu, Y., Zhang, J., & Zhang, H. (2020). Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. Computational Materials Science, 175, 109618.

[21] Wickham, H. and Grolemund G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. United States of America, O'Reilly Media,81-82.

[22] A. Bajaj (2023). Performance Metrics in Machine Learning. [https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide], (Last Accessed Date 01.10.2023).

[23] Guliyev, H. and Yerdelen, F. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. Journal of Applied Microeconometrics 1(2):85-99.

[24] Wu, X. and Meng S. (2016). E-commerce customer churn prediction based on improved SMOTE and AdaBoost. 2016 13th International Conference on Service Systems and Service Management (ICSSSM).

[25] Jabeur, S. B., Gharib, C., Mefteh-Wali, S., & Arfi, W. B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. Technological Forecasting and Social Change, 166, 120658.

[26] Shah, M., Shah, D., Bhat, S., Vyeth, V. (2019). Prediction and Causality Analysis of Churn Using Deep Learning. CMLA 2019,153-165.

[27] Umayaparvathi, V. and Iyakutti, K. (2016) A Survey on Customer Churn Prediction in Telecom Industry: Datasets, Methods and Metrics. International Research Journal of Engineering and Technology, 3, 1065-1070.

[28] He, Y., Xiong, Y. and Tsai, Y. (2020). Machine Learning Based Approaches to Predict Customer Churn for an Insurance Company. 2020 Systems and Information Engineering Design Symposium (SIEDS).

[29] Wu, X. and Meng S. (2016). E-commerce customer churn prediction based on improved SMOTE and AdaBoost. 2016 13th International Conference on Service Systems and Service Management (ICSSSM).