CUSTOMER RETENTION AND CHURN PREDICTION: IBM TELCO CUSTOMER CHURN

MODEL BUILDING IN RAPIDMINER

by Sarina Gurung

Machine Learning Final Project: CSBU 5420

Webster University: Department of Computer Science and Information Technology

Master of Science

In

Business Analytics

October 2025

Table of Contents

Introduction

1.1 Project Overview

In today's data-driven business landscape, customer retention has become a strategic priority across industries. As acquiring new customers is often more expensive than retaining existing ones, predicting customer churn has become increasingly critical for organizations seeking to sustain long-term growth. The telecommunications industry faces intense competition and high switching behavior among customers, making churn predictions both a business necessity and an analytical challenge.

Machine learning (ML) offers a powerful approach for detecting behavioral patterns that indicate customer dissatisfaction or intent to leave. By leveraging large volumes of structured and unstructured data, ML models can generate actionable insights that support marketing, service design, and customer experience management. This study applies supervised learning techniques, Logistic Regression, Decision Tree, and Random Forest, to predict customer churn, evaluate performance, and identify the key drivers influencing customer retention.

Beyond its technical implementation, this project demonstrates the broader potential of machine learning in solving real-world business problems by aligning statistical modeling with interpretability and business impact. This project bridges the gap between academic research and practical analytics applications relevant to today's data-driven landscape in any industry.

1.2 Objective

The primary goal of this project is to develop and evaluate multiple machine learning models to predict whether a customer will churn. The project also aims to identify the most influential variables driving customer retention behavior and precisely forecast the future churn of the telecommunication customers. This study aims to use RapidMiner, a data science platform that automates the process of machine learning and predictive models.

1.3 Significance

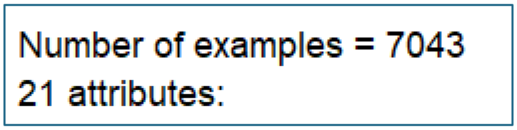
Customer churn is a persistent issue for organizations that rely on recurring revenue streams. In telecommunications, where annual churn rates can exceed 30%, even small improvements in retention translate into substantial financial gains (Chang et al., 2024). Predictive analytics powered by machine learning enables companies to proactively identify customers at risk of leaving and design interventions to prevent attrition. The significance of this research lies in its focus on developing interpretable, accurate, and scalable models that can guide data-driven retention strategies.

Research by Barsotti et al. (2024) highlights that while churn prediction has advanced considerably over the past decade, critical gaps remain in ensuring model interpretability and generalization. This study contributes to addressing those gaps by comparing multiple algorithms, Logistic Regression, Decision Tree, and Random Forest, to assess the trade-off between accuracy and transparency. Moreover, studies such as Explaining Customer Churn Prediction in the Telecom Industry Using Explainer Models (ScienceDirect, 2024) emphasize that understanding why customers churn is as critical as predicting who will churn.

Accurate churn prediction enables organizations to implement proactive retention strategies, improve customer satisfaction, and reduce revenue loss. This study will help identify the optimal models to predict future customer churn in telecommunication, aiming for increased accuracy. The findings also demonstrate practical applications of machine learning in customer analytics, allowing the administration to implement data-driven decision-making and strategies.

Data Description and Preparation

2.1 Dataset Information



Number of examples = 7043
21 attributes:

Figure 1: Dataset shape

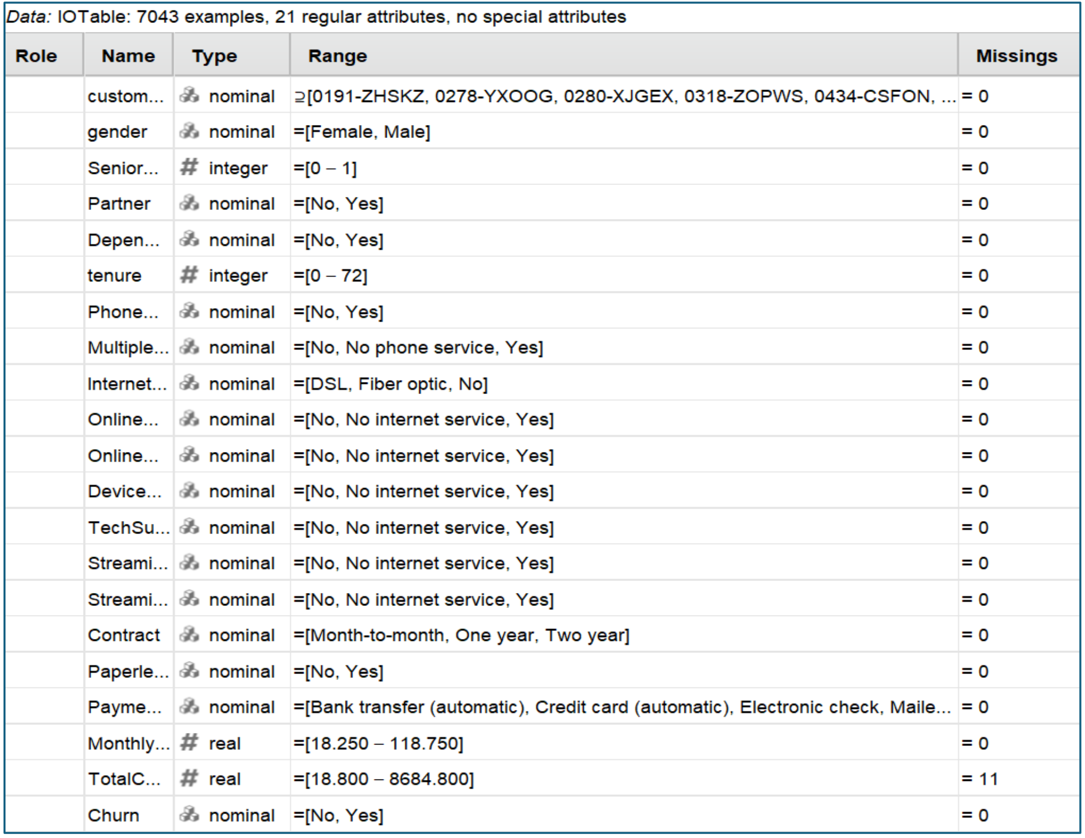
Dataset Source: Telco Customer Churn (Kaggle / IBM Watson)

Records: 7,043 customers

Attributes: 21 original variables (demographic, service, billing, and account information)

Target Variable: Churn (Yes/No)

i. **Data Types**
The dataset contains 21 attributes, among them, two (Monthly Charges and Total Charges) are real, two (Senior Citizen and Tenure) are integer, while the rest are nominal. The figure also states the range of values in different attributes.

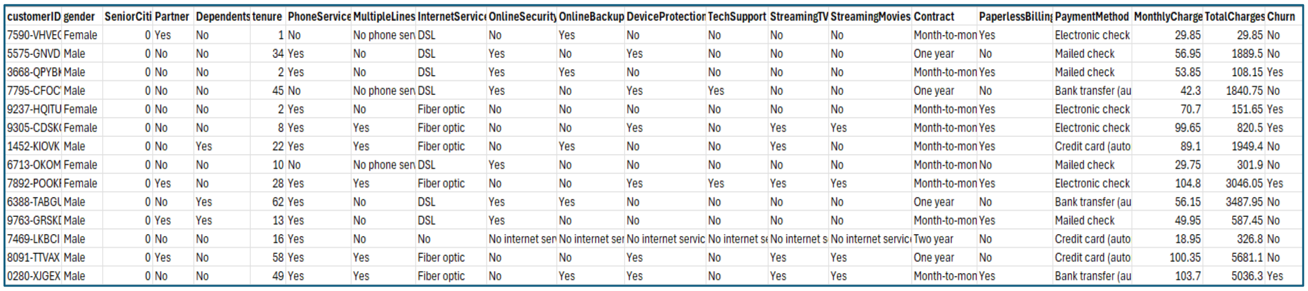


Data: IOTable: 7043 examples, 21 regular attributes, no special attributes

Role	Name	Type	Range	Missings
	custom...	nominal	⊇[0191-ZHSKZ, 0278-YXOOG, 0280-XJGEX, 0318-ZOPWS, 0434-CSFON, ...	= 0
	gender	nominal	=[Female, Male]	= 0
	Senior...	# integer	=[0 – 1]	= 0
	Partner	nominal	=[No, Yes]	= 0
	Depen...	nominal	=[No, Yes]	= 0
	tenure	# integer	=[0 – 72]	= 0
	Phone...	nominal	=[No, Yes]	= 0
	Multiple...	nominal	=[No, No phone service, Yes]	= 0
	Internet...	nominal	=[DSL, Fiber optic, No]	= 0
	Online...	nominal	=[No, No internet service, Yes]	= 0
	Online...	nominal	=[No, No internet service, Yes]	= 0
	Device...	nominal	=[No, No internet service, Yes]	= 0
	TechSu...	nominal	=[No, No internet service, Yes]	= 0
	Streami...	nominal	=[No, No internet service, Yes]	= 0
	Streami...	nominal	=[No, No internet service, Yes]	= 0
	Contract	nominal	=[Month-to-month, One year, Two year]	= 0
	Paperle...	nominal	=[No, Yes]	= 0
	Payme...	nominal	=[Bank transfer (automatic), Credit card (automatic), Electronic check, Maile...	= 0
	Monthly...	# real	=[18.250 – 118.750]	= 0
	TotalC...	# real	=[18.800 – 8684.800]	= 11
	Churn	nominal	=[No, Yes]	= 0

Figure 2: Data Types Overview

ii. Sample Dataset



customerID	gender	SeniorCiti	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetServic	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBillin	PaymentMethod	MonthlyCharge	TotalCharges	Churn
7590-VHVEC	Female	0	Yes	No	1	No	No phone sen	DSL	No	Yes	No	No	No	No	Month-to-mon	Yes	Electronic check	29.85	29.85	No
5575-GNVD	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-mon	Yes	Mailed check	53.85	108.15	Yes
7795-CFOC	Male	0	No	No	45	No	No phone sen	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (au	42.3	1840.75	No
9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-mon	Yes	Electronic check	70.7	151.65	Yes
9305-CDSK	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-mon	Yes	Electronic check	99.65	820.5	Yes
1452-KIOVK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-mon	Yes	Credit card (auto	89.1	1949.4	No
6713-OKOM	Female	0	No	No	10	No	No phone sen	DSL	Yes	No	No	No	No	No	Month-to-mon	No	Mailed check	29.75	301.9	No
7892-POOKI	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-mon	Yes	Electronic check	104.8	3046.05	Yes
6388-TABGU	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank transfer (au	56.15	3487.95	No
9763-GRSKI	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-mon	Yes	Mailed check	49.95	587.45	No
7469-LKBCI	Male	0	No	No	16	Yes	No	No	No internet ser	No internet ser	No internet servic	No internet se	No internet s	No internet servic	Two year	No	Credit card (auto	18.95	326.8	No
8091-TTVAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	One year	No	Credit card (auto	100.35	5681.1	No
0280-XJGEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-mon	Yes	Bank transfer (au	103.7	5036.3	Yes

Figure 3: Sample of the Dataset

Figure 1.2 shows a sample overview of the dataset and its contents. The dataset contains 21 attributes, including demographic service, billing, and account information.

Data Cleaning and Preprocessing

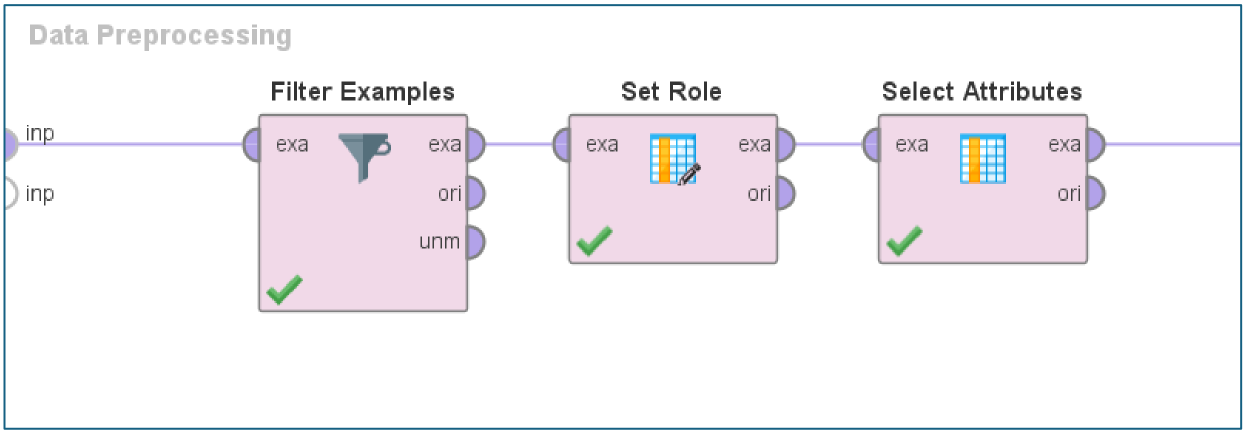
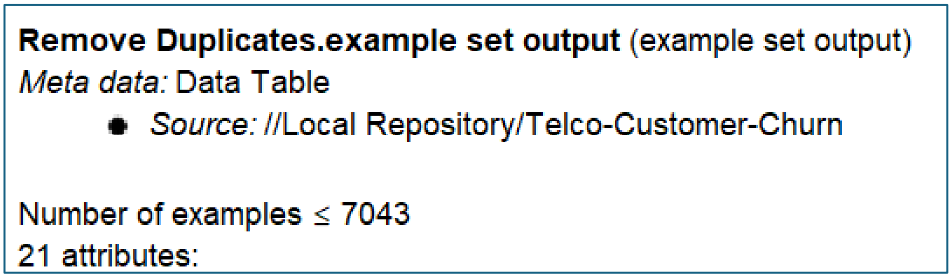


Figure 4: Data Preprocessing Steps

i. Identifying Duplicates

To ensure that there are no duplicates, the Remove Duplicates operator was used, which identified that there were no duplicates in the dataset. Since there were no duplicates, the operator was removed from the process.



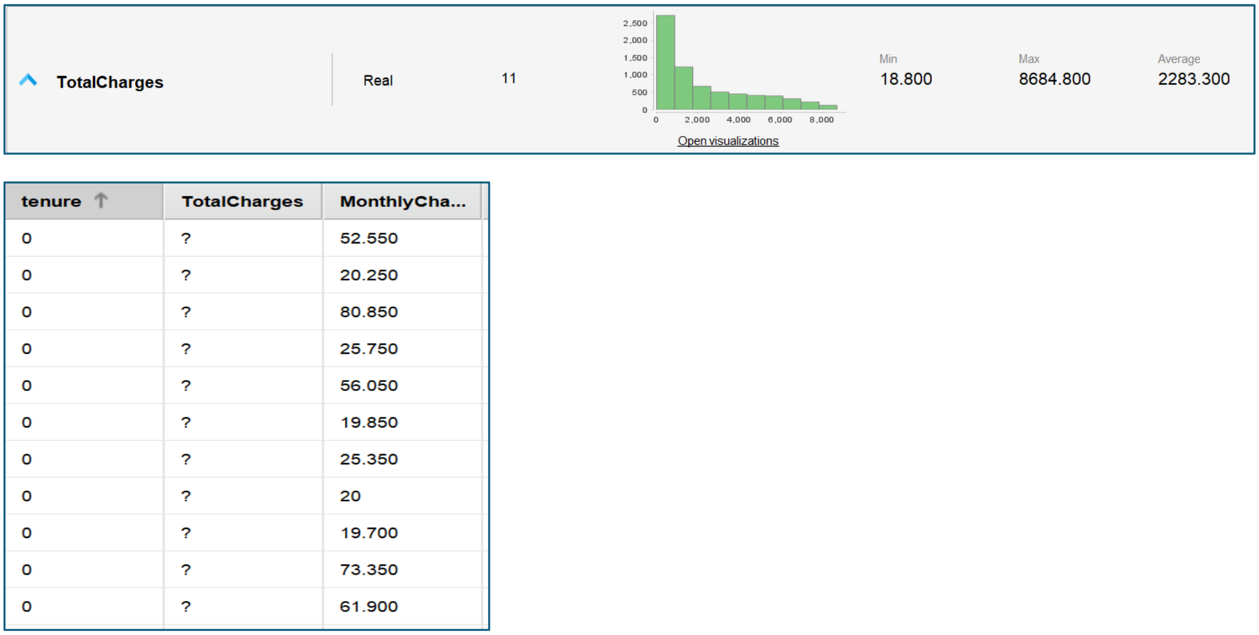
Remove Duplicates.example set output (example set output)
Meta data: Data Table

- *Source:* //Local Repository/Telco-Customer-Churn

Number of examples ≤ 7043
21 attributes:

Figure 5: Number of attributes after removing duplicates

ii. Handling Missing Values

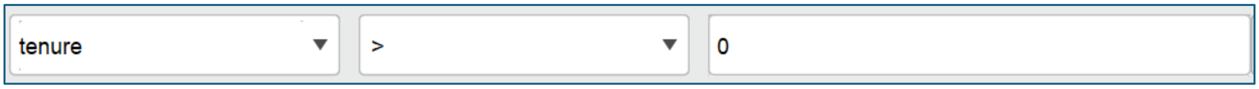


TotalCharges Real 11 Min 18.800 Max 8684.800 Average 2283.300

tenure ↑	TotalCharges	MonthlyCha...
0	?	52.550
0	?	20.250
0	?	80.850
0	?	25.750
0	?	56.050
0	?	19.850
0	?	25.350
0	?	20
0	?	19.700
0	?	73.350
0	?	61.900

Figure 6: Missing values identified

Upon inspection, 11 records were identified with a tenure value of 0 and missing total charges. These concerns newly acquired customers who have not yet completed a billing cycle. Since these records lack sufficient historical information, they were excluded from the modeling dataset using the Filter Example Operator.



tenure ▼ | > ▼ | 0

Figure 7: Filtering out the missing values

The Set Role operator was used to set Churn as a label attribute. The Select attributes operator was used to exclude CustomerID, as it does not serve any purpose in this analysis.

iii. Data Preprocessing for Regression Model

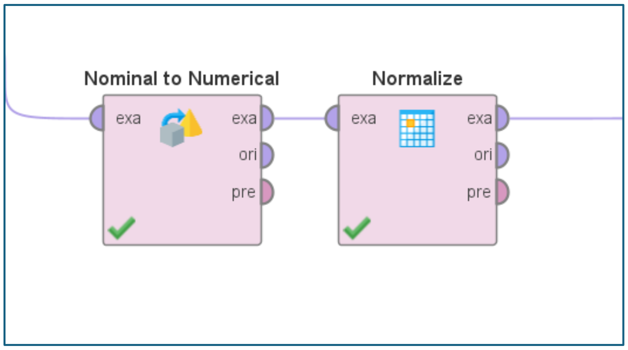


Figure 8: Data scaling

The polynomial attributes were converted into numerical attributes through dummy coding and using comparison groups. The numerical values were then normalized to scale the dataset into a standard set. This step is crucial before running any regression model.

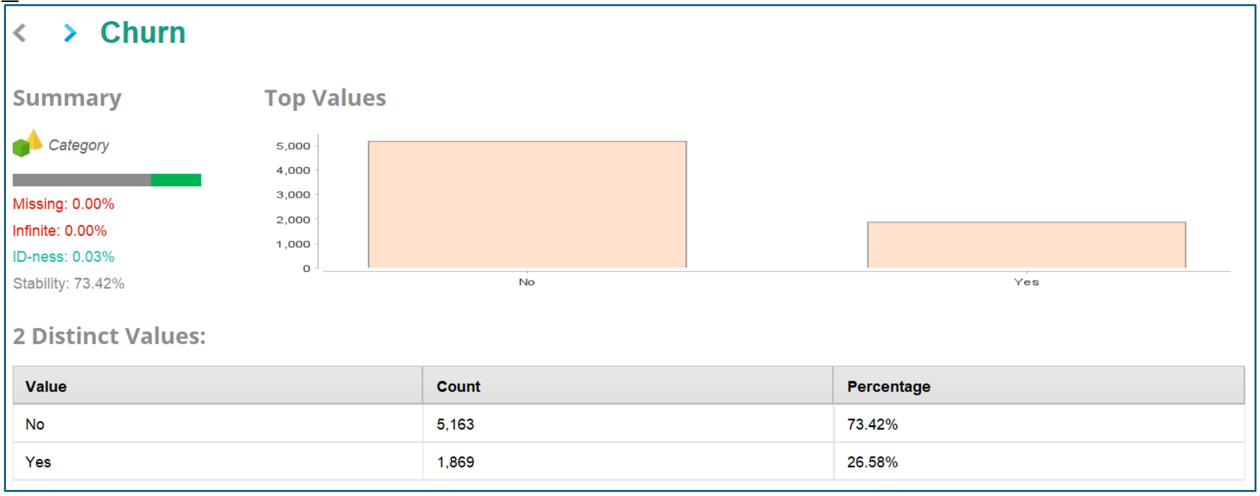
Exploratory Data Analysis (EDA)

The statistics operator was used to summarize the data and get an overview of the attributes and their values. Along with the statistics provided in the example set.

3.1 Univariate Analysis

i. Churn

Out of 7032 customers of IBM Telco, 26.58% of the customers churned.



Churn

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.03%

Stability: 73.42%

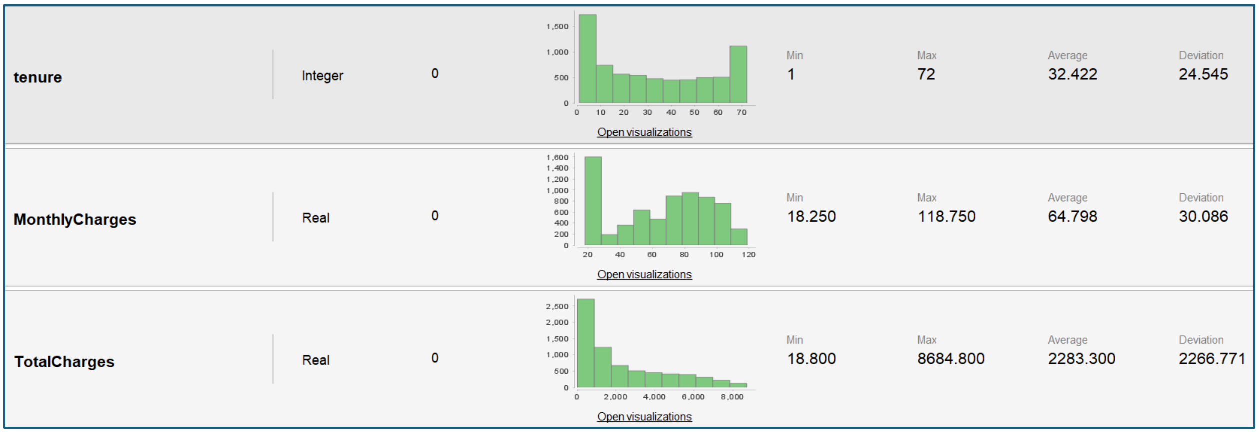
Top Values

2 Distinct Values:

Value	Count	Percentage
No	5,163	73.42%
Yes	1,869	26.58%

Figure 9: Churn Statistics

ii. Numeric Attributes



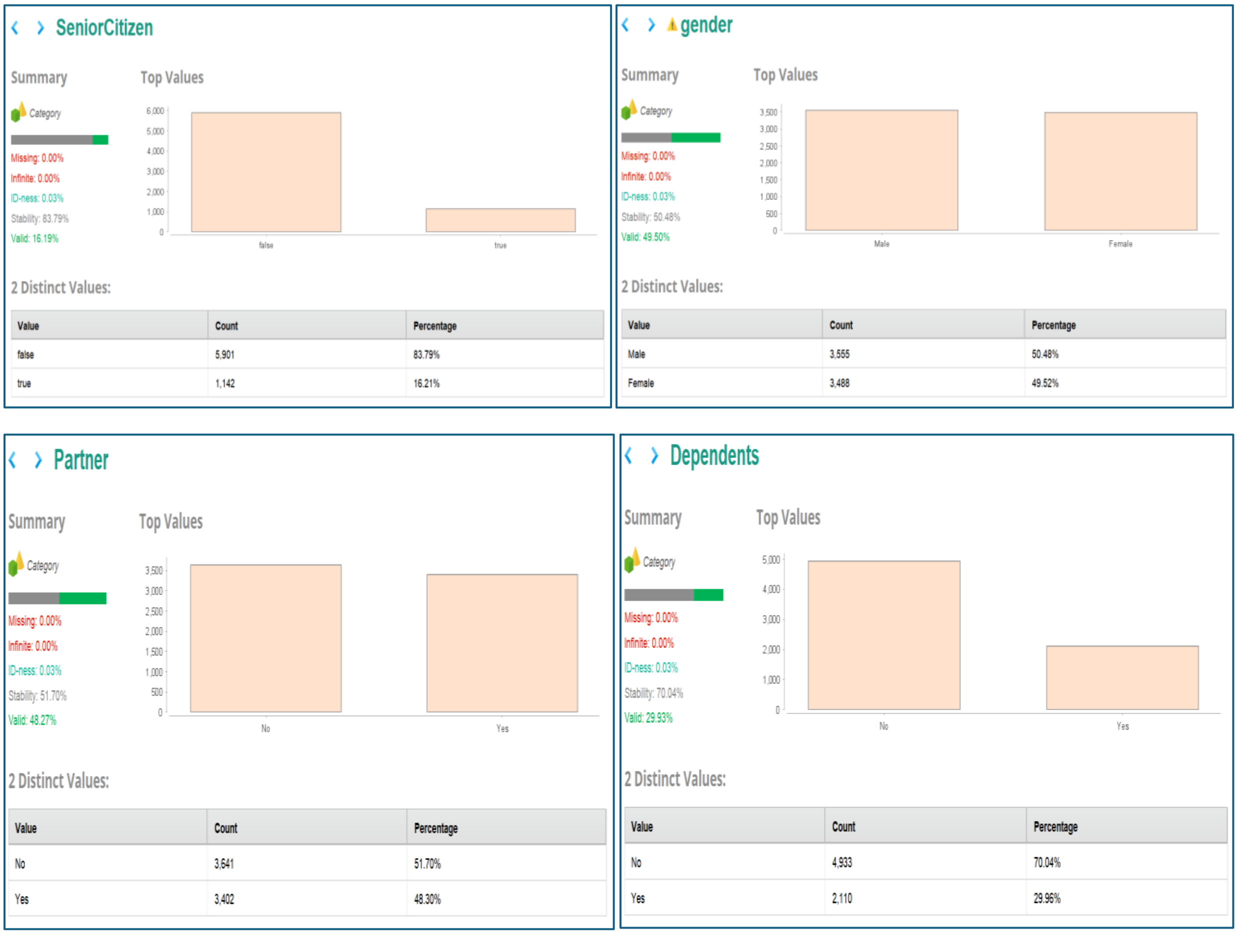
tenure	Integer	0	Open visualizations	Min 1	Max 72	Average 32.422	Deviation 24.545
MonthlyCharges	Real	0	Open visualizations	Min 18.250	Max 118.750	Average 64.798	Deviation 30.086
TotalCharges	Real	0	Open visualizations	Min 18.800	Max 8684.800	Average 2283.300	Deviation 2266.771

Figure 10: Numeric Attributes Statistics

Tenure ranged from 1 to 72 months. Monthly charges averaged $64.79, whereas Total charges averaged $2283.3.

iii. Demographic Attributes

This dataset contains four demographic attributes (Gender, Senior citizens, Partner, Dependents). The statistics show that the majority of the customers are not senior citizens (only 16.21% senior citizens). The distribution of males and females is almost the same. The same goes for the attribute Partner (51.70% of No, 48.30% of Yes). While 70.04% had no dependents and 29.96% did.



SeniorCitizen

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 83.79%
Valid: 16.19%

Top Values

2 Distinct Values:

Value	Count	Percentage
false	5,901	83.79%
true	1,142	16.21%

gender

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 50.48%
Valid: 49.50%

Top Values

2 Distinct Values:

Value	Count	Percentage
Male	3,555	50.48%
Female	3,488	49.52%

Partner

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 51.70%
Valid: 48.27%

Top Values

2 Distinct Values:

Value	Count	Percentage
No	3,641	51.70%
Yes	3,402	48.30%

Dependents

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 70.04%
Valid: 29.93%

Top Values

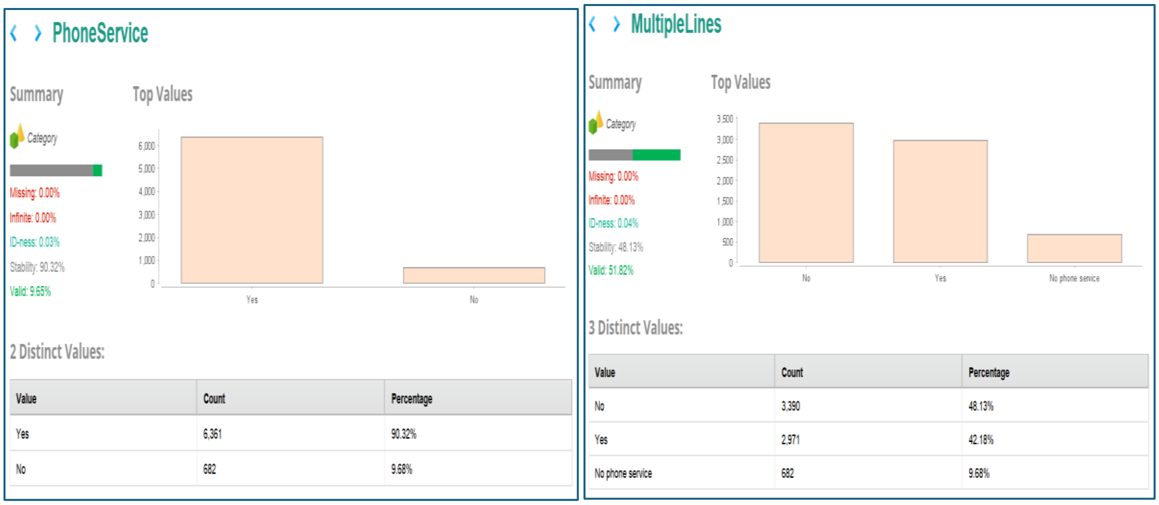
2 Distinct Values:

Value	Count	Percentage
No	4,933	70.04%
Yes	2,110	29.96%

Figure 11: Demographic attributes statistics

iv. Service Attributes

9.68% of the customers had no phone service. The majority of the customers had Fiber optics, while 21.67% had no internet connection.



PhoneService

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.03%

Stability: 90.32%

Valid: 9.65%

2 Distinct Values:

Value	Count	Percentage
Yes	6,361	90.32%
No	682	9.68%

MultipleLines

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.04%

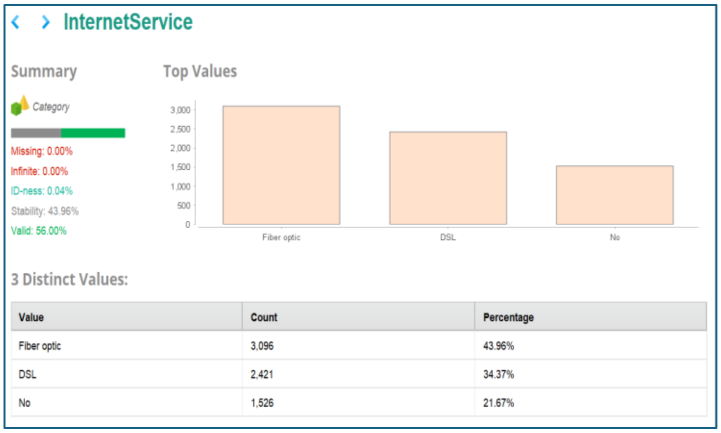
Stability: 48.13%

Valid: 51.82%

3 Distinct Values:

Value	Count	Percentage
No	3,390	48.13%
Yes	2,971	42.18%
No phone service	682	9.68%

Figure 12: Phone service stats



InternetService

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.04%

Stability: 43.96%

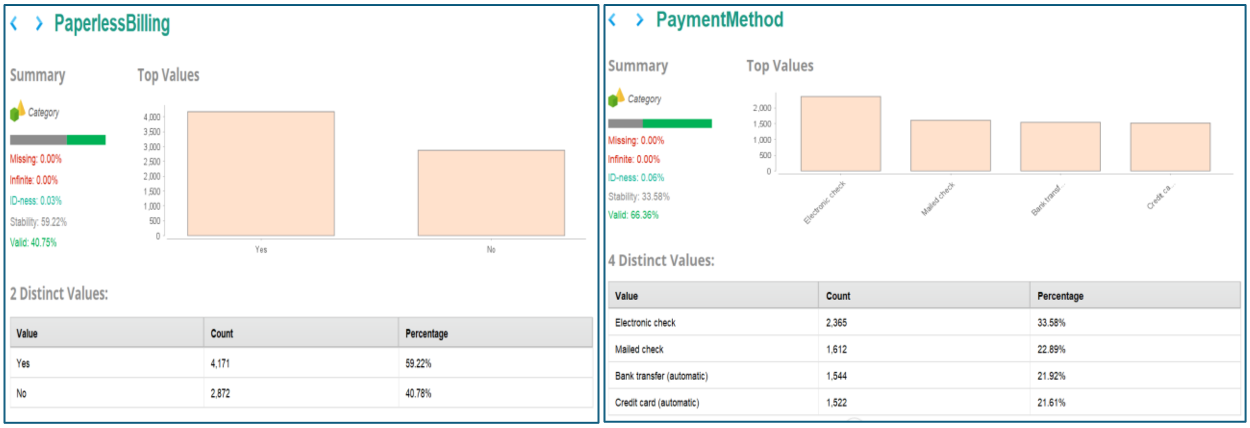
Valid: 56.00%

3 Distinct Values:

Value	Count	Percentage
Fiber optic	3,096	43.96%
DSL	2,421	34.37%
No	1,526	21.67%

Figure 13: Internet service stats

v. Payment Attributes



PaperlessBilling

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.03%

Stability: 59.22%

Valid: 40.75%

2 Distinct Values:

Value	Count	Percentage
Yes	4,171	59.22%
No	2,872	40.78%

PaymentMethod

Summary

Category

Missing: 0.00%

Infinite: 0.00%

ID-ness: 0.06%

Stability: 33.58%

Valid: 66.36%

4 Distinct Values:

Value	Count	Percentage
Electronic check	2,365	33.58%
Mailed check	1,612	22.89%
Bank transfer (automatic)	1,544	21.92%
Credit card (automatic)	1,522	21.61%

Figure 14: Payment methods stats

The majority of the customers used the traditional method (Electronic check) to pay. And 59.22% used paperless billing.

Bivariate Analysis

i. Correlation Matrix

The correlation matrix operator was used to calculate the correlation among the attributes to identify any multicollinearity. The correlation heatmap shows negative correlations between the attributes' values 'Yes' and 'No'. Fiber optics and monthly charges show a considerably high correlation of 0.787. Total charges and tenure also show a high correlation of 0.826. While the number of service add-ons seems to correlate with high monthly charges (approximately >= 0.6). The 'No internet service' values had perfect collinearity with the value 'No' of the same attribute. Hence, 'No internet service' values were added as baseline in the comparison group while converting nominal to numerical. The remaining attributes were not dropped because they capture different dimensions of the dataset.

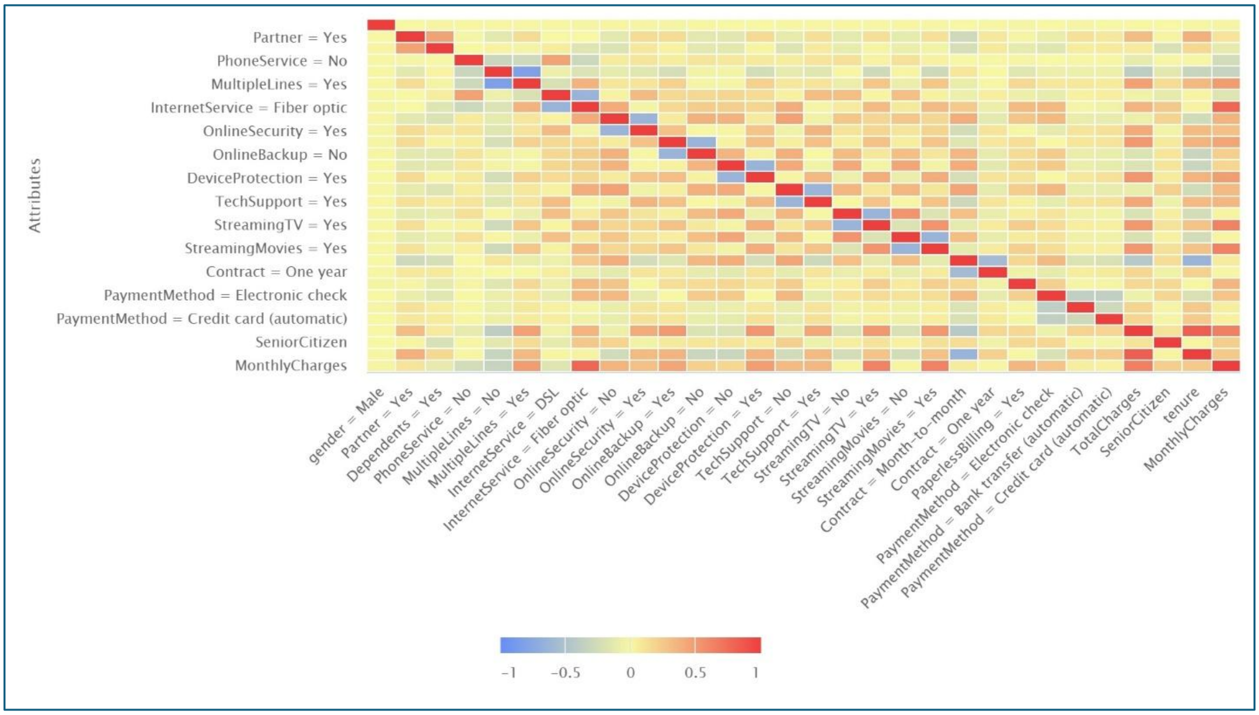
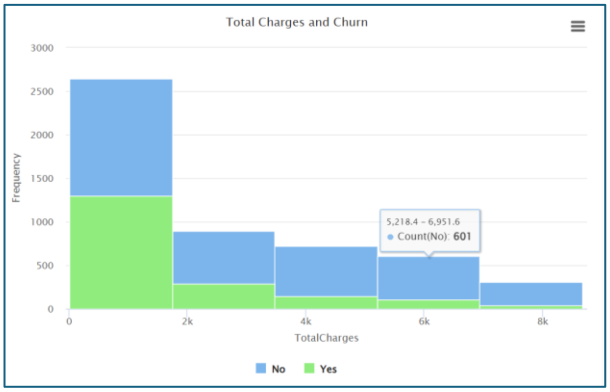


Figure 15: Correlation Heatmap

ii. Numeric attributes and Churn



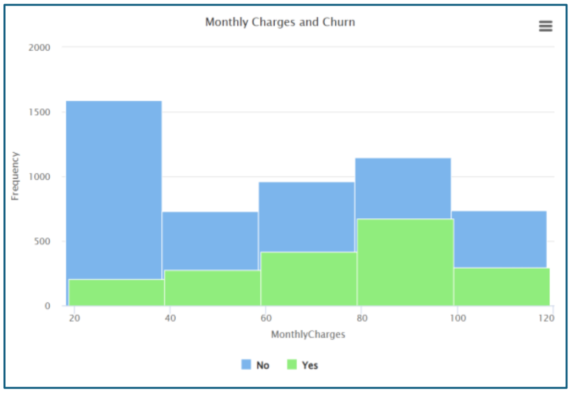


Figure 16: Charges and churn

Customers with lower total charges (below 2000) have the highest churn rate, whereas customers with higher monthly charges (78-99) have higher churn rates. Also, according to Figure 16, new customers or customers with a low tenure period (1-15 months) are more likely to churn.

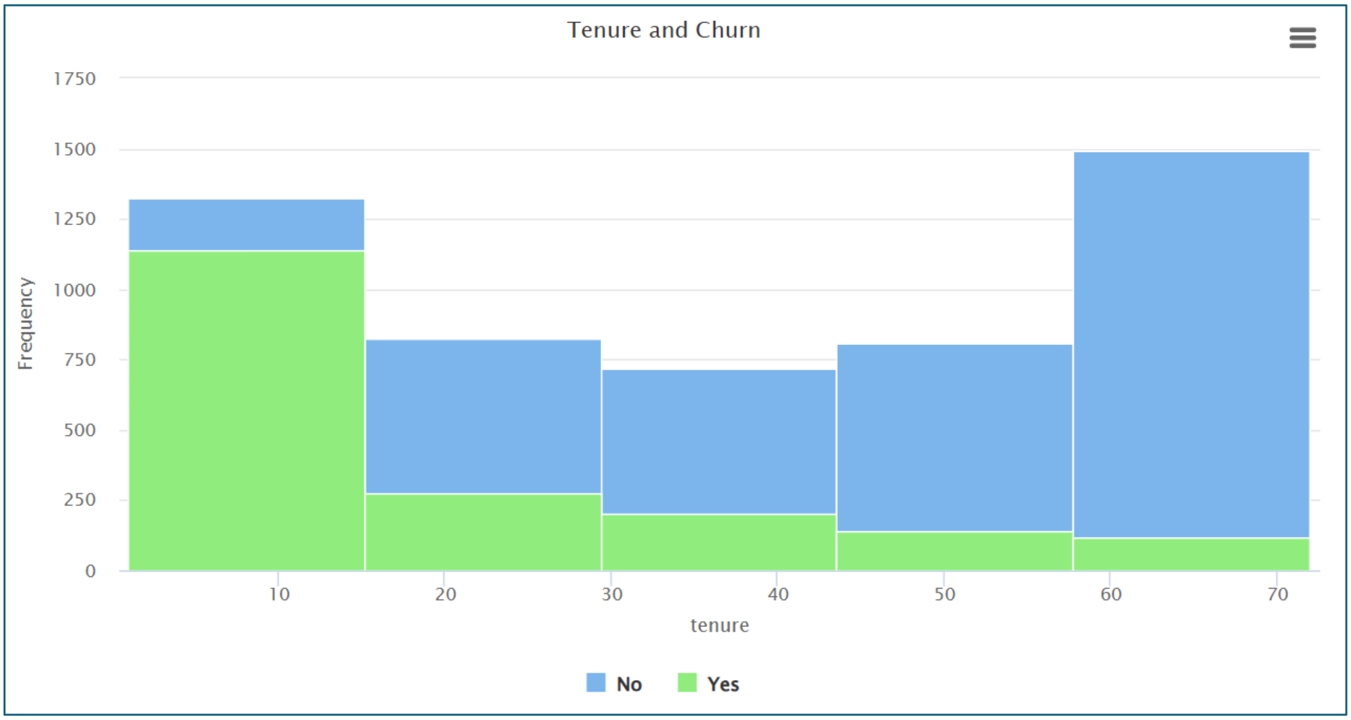


Figure 17: Tenure and Churn

Model Development

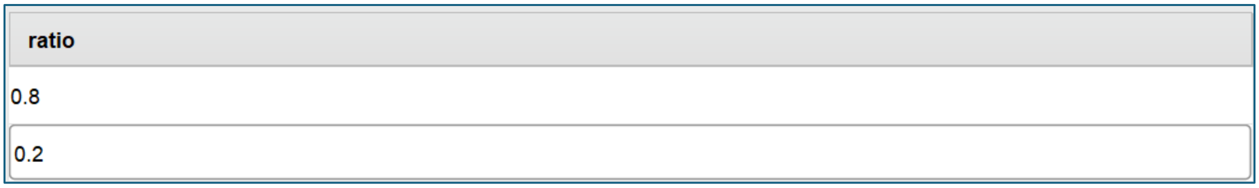
4.1 Models Used

In this study, to predict churn, 4 different models are used. The models are then to be compared based on their performance.

- Decision Tree
- Linear Regression
- Logistic Regression
- Random Forest

4.2 Split Data

The data is split into an 80% training set and a 20% holdout set.



ratio
0.8
0.2

Figure 18: Split Data

4.3 Data Balancing

The churn datasets are always greatly imbalanced, with not churn being the majority of the value. The dataset used for this project encounters a similar issue. Due to the significantly imbalanced nature of the dataset, with a larger proportion of customers labeled as “No” Churn compared to those labeled as “Yes” Churn, the result of the predictive model can be biased towards the majority class, reducing its ability to correctly identify churners. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data. SMOTE generates synthetic samples of the minority class by interpolating between existing minority instances, rather than simply duplicating them. This approach balances the class distribution and enables the model to better learn patterns associated with customer churn. As a result, it improves the model’s recall, ensuring more accurate identification of customers likely to churn while maintaining generalization performance across models such as Logistic Regression, Decision Tree, and Random Forest.

4.3 Decision Tree

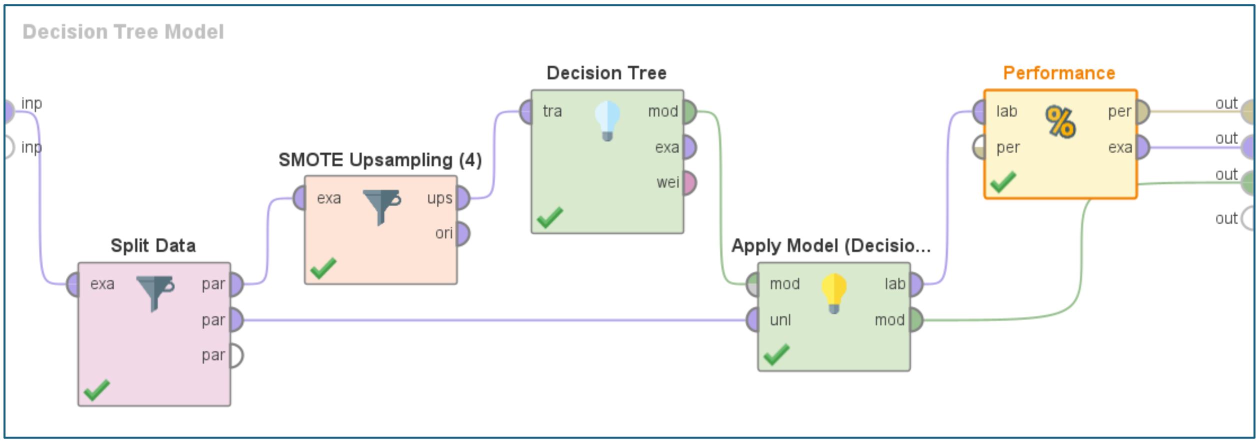
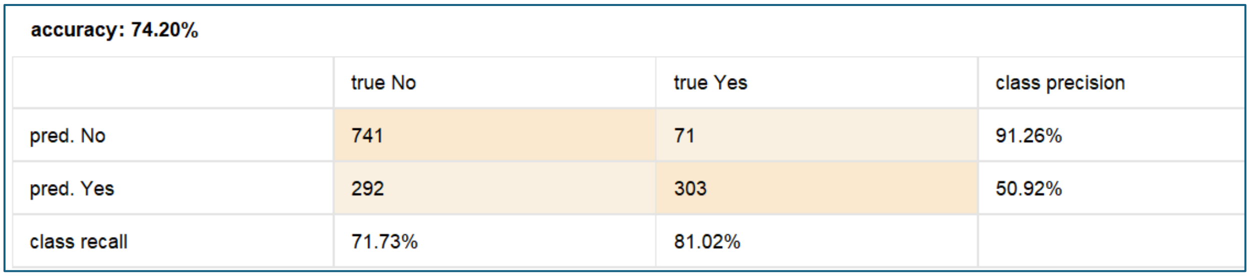


Figure 19: Decision Tree Model Process

The decision tree criterion is set to gain ratio with a maximal depth of 15, applying pruning and pre-pruning with a confidence of 0.25. The model was trained using the training set, and the holdout set was used to evaluate the model. The decision tree model recorded the accuracy of 74.20%, classification error of 25.80%, AUC of 80.1%, precision of 50.92%, recall of 81.02%, F score of 62.54%, sensitivity of 81.02%, and specificity of 71.73%.



accuracy: 74.20%

	true No	true Yes	class precision
pred. No	741	71	91.26%
pred. Yes	292	303	50.92%
class recall	71.73%	81.02%	

Figure 20: Decision Tree Performance

4.4 Linear Regression

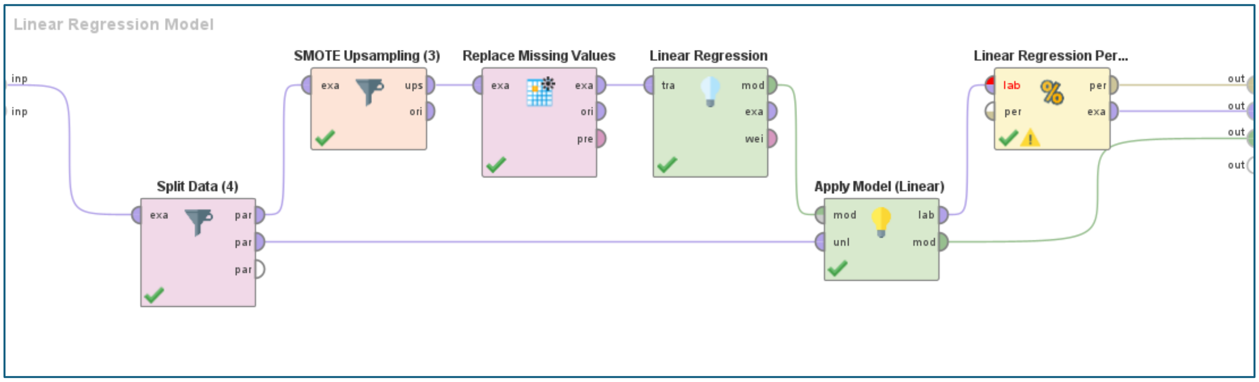
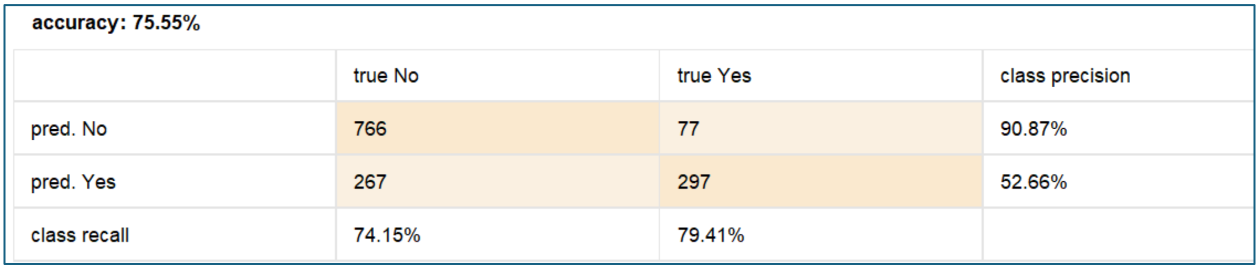


Figure 31: Linear regression process

Similarly, SMOTE Upsampling was used to balance the training set, which was then used to train the model, and after the model was evaluated on the holdout set. The linear regression model recorded the accuracy of 75.55%, classification error of 24.45%, AUC of 84.1%, precision of 52.66%, recall of 79.41%, F score of 63.33%, sensitivity of 79.41%, and specificity of 74.15%.



accuracy: 75.55%

	true No	true Yes	class precision
pred. No	766	77	90.87%
pred. Yes	267	297	52.66%
class recall	74.15%	79.41%	

Figure 22: Linear regression performance

4.5 Logistic Regression

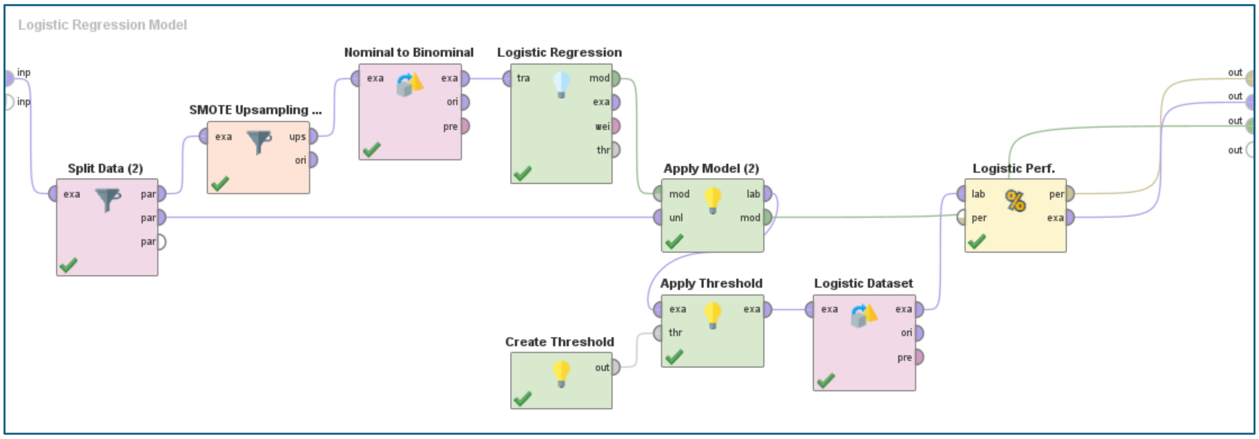
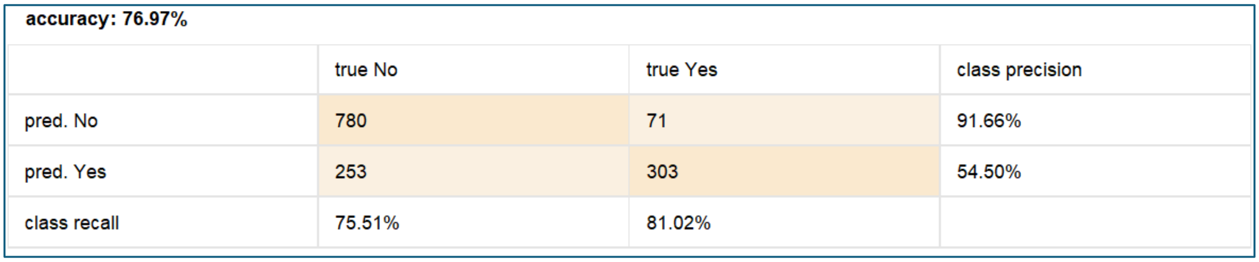


Figure 23: Logistic regression process

A threshold of 0.5 with the first class being 'No', and the second class 'Yes' was set up and applied to the model. The logistic regression model recorded an accuracy of 76.97%, a classification error of 23.03%, an AUC of 86.1%, a precision of 54.50%, a recall of 81.02%, an F score of 65.16%, a sensitivity of 81.02%, and a specificity of 75.51%.



accuracy: 76.97%

	true No	true Yes	class precision
pred. No	780	71	91.66%
pred. Yes	253	303	54.50%
class recall	75.51%	81.02%	

Figure 24: Logistic regression performance

4.6 Random Forest

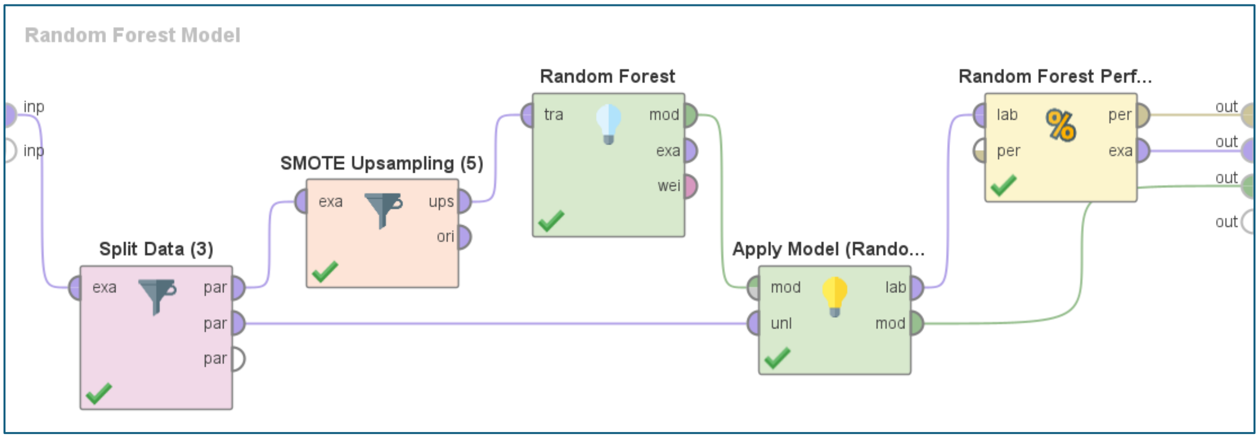
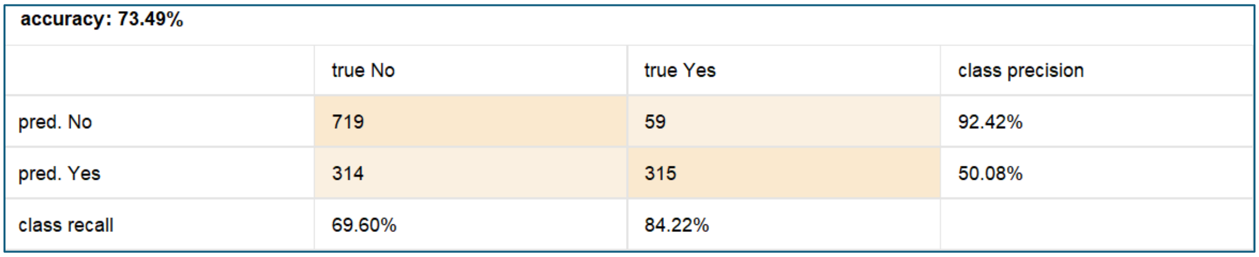


Figure 25: Random Forest Process

The random forest model recorded an accuracy of 73.49%, a classification error of 26.51%, an AUC of 83.9%, a precision of 50.08%, a recall of 84.22%, an F score of 62.81%, a sensitivity of 84.22%, and a specificity of 69.60%.



accuracy: 73.49%

	true No	true Yes	class precision
pred. No	719	59	92.42%
pred. Yes	314	315	50.08%
class recall	69.60%	84.22%	

Figure 26: Random forest performance

Results and Discussion

Model	**Accuracy**	**Recall**	**AUC**
Decision Tree	74.20%	81.20%	81.1%
Linear Regression	75.55%	79.41%	84.1%
Logistic Regression	**76.97%**	**81.02%**	**86.1%**
Random Forest	73.49%	84.22%	83.9%

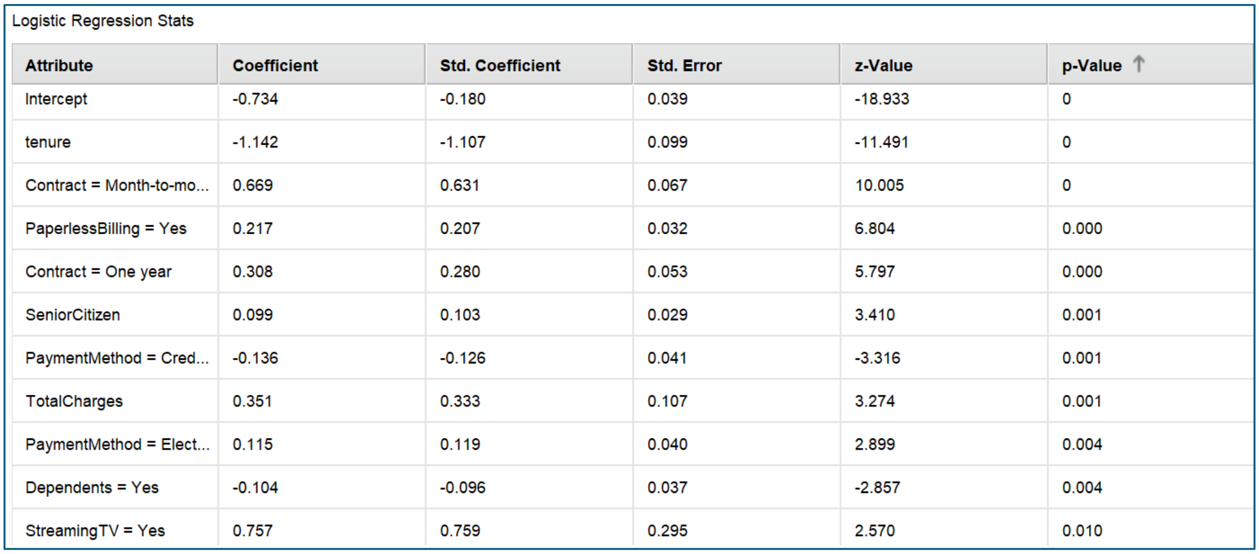
Figure 27: Model performance comparison

Logistic Regression was selected as the best-performing model for predicting customer churn. It achieved the highest AUC (0.861), the lowest classification error (23.03%), and the best F-measure (65.16%), indicating the strongest overall ability to distinguish between churners and non-churners. It maintains a strong recall of 81.02%, meaning it captures the majority of actual churners, while also delivering the highest precision (54.50%) and specificity (75.51%) among all models tested, minimizing false alarms that would waste retention resources. Additionally, Logistic Regression offers high interpretability, allowing stakeholders to understand which factors drive churn and take targeted action.

Identifying Significant Predictors

Along with the highly accurate predictive power of the churners, it is equally important to identify the drivers of the churn. Therefore, this study also aims to identify those predictors for future predictions. Figure 28 shows the list of statistically significant predictors in predicting churn. From the Linear Regression stats, we can interpret the following:

- The month-to-month contract and fiber optic service have the strongest positive influence on churn.
- Tenure and total charges act as stabilizing factors; loyal customers with longer service durations are less likely to churn.
- The model shows a mix of behavioral and contractual drivers, aligning with typical churn dynamics in telecom analytics.



Logistic Regression Stats

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Intercept	-0.734	-0.180	0.039	-18.933	0
tenure	-1.142	-1.107	0.099	-11.491	0
Contract = Month-to-mo...	0.669	0.631	0.067	10.005	0
PaperlessBilling = Yes	0.217	0.207	0.032	6.804	0.000
Contract = One year	0.308	0.280	0.053	5.797	0.000
SeniorCitizen	0.099	0.103	0.029	3.410	0.001
PaymentMethod = Cred...	-0.136	-0.126	0.041	-3.316	0.001
TotalCharges	0.351	0.333	0.107	3.274	0.001
PaymentMethod = Elect...	0.115	0.119	0.040	2.899	0.004
Dependents = Yes	-0.104	-0.096	0.037	-2.857	0.004
StreamingTV = Yes	0.757	0.759	0.295	2.570	0.010

Figure 28: Predictors Stats

Conclusion

This study applied multiple supervised machine learning algorithms, Decision Tree, Linear Regression, Logistic Regression, and Random Forest to predict customer churn using IBM's Telco Customer Churn dataset. After data preprocessing, feature selection, and class balancing through the Synthetic Minority Oversampling Technique (SMOTE), each model was trained and evaluated on an 80/20 train-test split.

From Figure 27, we can conclude that Logistic Regression has the best accuracy of 76.97%. However, Random Forest predicted the highest number of true predictions for churners. Among all the models tested, Logistic Regression demonstrated the most balanced and robust performance. The high AUC value indicates superior predictive power and a stronger ability to distinguish between churners and non-churners. Logistic Regression also maintained interpretability, allowing for a clear understanding of how variables such as tenure, contract type, payment method, and monthly charges influenced churn probability.

While Random Forest achieved a slightly higher recall (84.22%), it showed lower overall accuracy (73.49%), precision, and specificity, suggesting overfitting tendencies. Therefore, Logistic Regression is recommended as the optimal model for customer churn prediction in this study because it balances predictive performance, interpretability, and business applicability.

In practical applications, this model can help telecommunication companies proactively identify at-risk customers and design targeted retention strategies. Moreover, the same framework can be adopted across industries such as banking, insurance, and e-commerce, where understanding customer attrition is essential for sustaining long-term profitability. By integrating logistic regression-based churn models within customer relationship management (CRM) systems, organizations can make timely, data-driven decisions that minimize churn and strengthen customer loyalty.

Reference

Ahmad, A. K., Jafar, A., & Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0191-6

Barsotti, A., Gianini, G., Mio, C., Lin, J., Babbar, H., Singh, A., Taher, F., & Damiani, E. (2024). A Decade of Churn Prediction Techniques in the TelCo Domain: A Survey. *SN Computer Science/SN Computer Science*, *5*(4). https://doi.org/10.1007/s42979-024-02722-7

Chang, V., Hall, K., Xu, Q. A., Amao, F. O., Ganatra, M. A., & Benson, V. (2024). Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. *Algorithms*, *17*(6), 231–231. https://doi.org/10.3390/a17060231

Raut, N. V. (2020). *A study of ensemble machine learning to improve telecommunication customer churn prediction* (Doctoral dissertation, Dublin Business School). https://esource.dbs.ie/bitstream/10788/4237/1/msc_raut_nv_2020.pdf

Shaikhsurab, M. A., & Magadum, P. (2024). *Enhancing Customer Churn Prediction in Telecommunications: An Adaptive Ensemble Learning Approach*. ArXiv.org. https://doi.org/10.48550/arXiv.2408.16284

Sumana Sharma Poudel, Suresh Pokharel, & Mohan Timilsina. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning with Applications*, *17*, 100567–100567. https://doi.org/10.1016/j.mlwa.2024.100567