Customer Churn Prediction Using Machine Learning

Project members:

Emrullah Ayaz

Selahattin Koray Yıldız

Introduction

Customer churn rate is the rate at which a user using an application or an individual who is receiving any service abandons the service or application they are using. This rate is a very important issue for companies. Customer churn, also known as customer attrition, refers to the phenomenon where customers stop doing business with a company. Predicting customer churn is crucial for businesses as it helps them identify at-risk customers and take proactive measures to retain them. Retaining existing customers is often more cost-effective than acquiring new ones, making churn prediction a critical task for improving customer satisfaction and profitability. Machine learning plays a significant role in solving this problem by analyzing historical customer data to identify patterns and predict which customers are likely to churn. By leveraging machine learning algorithms, businesses can develop targeted retention strategies, such as personalized offers or improved customer service, to reduce churn rates. This project aims to build a machine learning model to predict customer churn in telecom sector based on customer behavior, demographics, and transaction history.

In our dataset, we have a total of 7013 customer records and 33 features. These data types include categorical numeric Geographical Textual data types. In the previous report, we performed data pre-processing steps such as handling missing values, coding categorical variables, and normalizing numeric features. We trained the model using Logistic Regression, Random Forest, XGBoost, Gradient Boosting, SVM, and KNN algorithms. After training these models, we tested them one by one. Among them, Gradient Boosting emerged as the best performing model by achieving 93% accuracy, 86% precision, 86% recall, and 0.016% F1 score after hyperparameter tuning. This report focuses on further improving the performance of the Gradient Boosting model by applying additional techniques and evaluating their effects.

Motivation for Model Explainability

There are several reasons why explainability is important in this particular problem. Here are some:

1- Sustainability: If a company can predict its customers' decision to stop being customers before they decide, it can take the necessary precautions to avoid losing them. This helps the company to focus more on customer satisfaction and create a sustainable culture.

2- Trust: Trust in a company comes from customer feedback. If trust in a company increases, the company will grow more and receive more investment. If a company constantly gains new customers and never loses any, the trust in the company in the market will increase.

3- Accountability: Accountability is a metric that shows how company employees spend their energy and how much of it they get in return. If company employees can predict in advance which customers are easier to win back, they will get a higher conversion rate and work more efficiently by not focusing on customers who are unlikely to return. Thanks to this model, employees will be able to communicate with the right customers and not waste time on customers who will leave anyway.

4- Business Insights: If a company knows what percentage of its customers will leave in the next year, and if the model can explain the reasons for their departure well, the company can change its business strategies, make new hires, or prioritize solutions for things that create customer dissatisfaction.

For these reasons, it is very important for the customer churn prediction model to be explainable.

There is a lot of competition in the telecom industry and it is very easy and inexpensive for a customer to switch to another service provider. It is very easy to lose a customer and very difficult to gain a new customer.

If the model can explain transparently why it predicts that the customer will leave, special campaigns can be made for the customer based on the reasons it explains, or if it can know the satisfaction with the service it receives before its decision is finalized, work can be done to increase the quality of the service the customer receives.

Model transparency is also very important for company employees to trust the model and make decisions in customer relations.

Another important reason is that if the model transparently indicates the reasons why customers want to leave, the company can develop new strategies for those reasons.

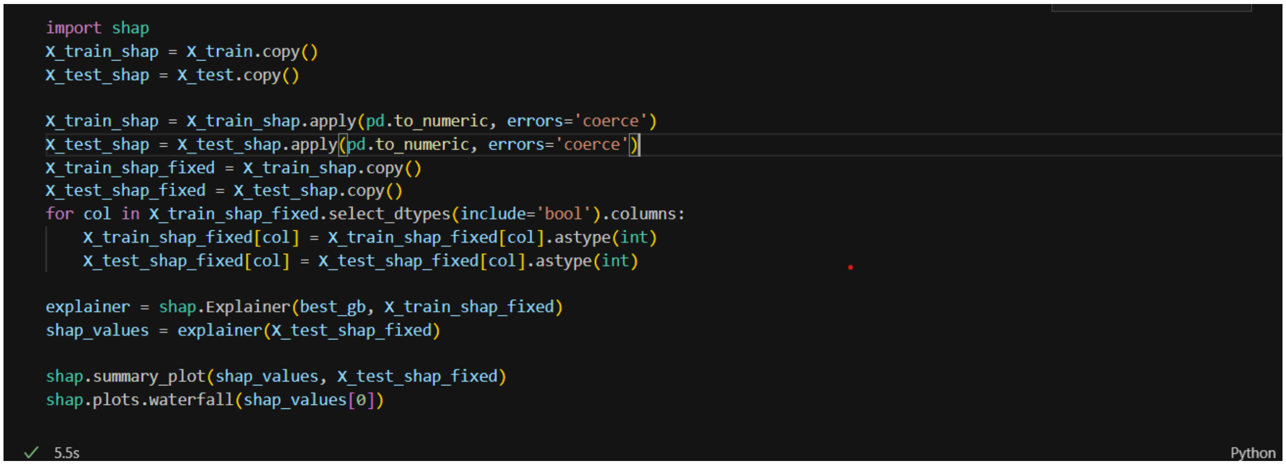
Another important point is that even when the company cannot get direct feedback from the customer, it allows the company to know the behavior of that customer and why they are unhappy.

Otherwise, if the model acts like a black box and only says that the customer will leave and does not specify why, the model becomes useless and does not work in terms of preventing customer loss. Model transparency is very important in the telecom industry for these reasons.

Explainability Techniques Used

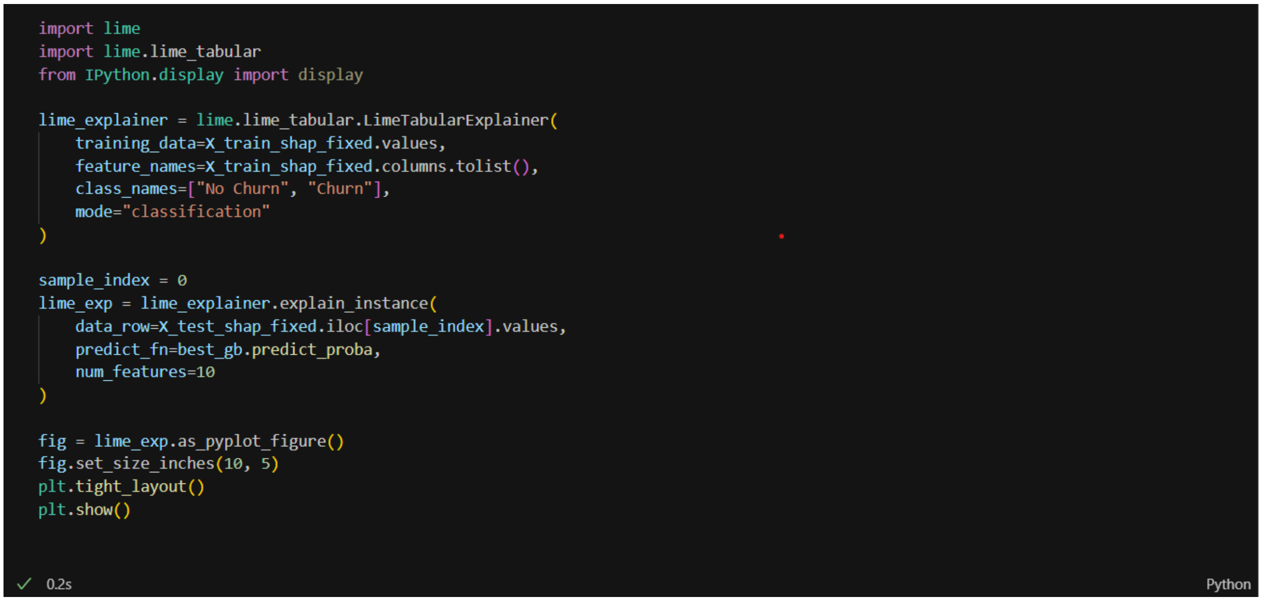
In our project, we used SHAP and LIME as XAI techniques. To briefly explain why we used them:

SHAP: The SHAP model gave us information about which features (e.g., tenure, contract type, monthly charges) affected our model's forecasting. It helped us interpret the model both globally and locally (individually). Thus, it clearly tells us which things to focus on when acting according to the data from the forecast.



```python
import shap
X_train_shap = X_train.copy()
X_test_shap = X_test.copy()

X_train_shap = X_train_shap.apply(pd.to_numeric, errors='coerce')
X_test_shap = X_test_shap.apply(pd.to_numeric, errors='coerce')
X_train_shap_fixed = X_train_shap.copy()
X_test_shap_fixed = X_test_shap.copy()
for col in X_train_shap_fixed.select_dtypes(include='bool').columns:
    X_train_shap_fixed[col] = X_train_shap_fixed[col].astype(int)
    X_test_shap_fixed[col] = X_test_shap_fixed[col].astype(int)

explainer = shap.Explainer(best_gb, X_train_shap_fixed)
shap_values = explainer(X_test_shap_fixed)

shap.summary_plot(shap_values, X_test_shap_fixed)
shap.plots.waterfall(shap_values[0])
```

LIME: The LIME model allows us to see in the simplest and most detailed way which features the churn value is derived from for a customer. Thus, it is important for the model to be explainable for only one customer and to be able to make a customer-based approach.

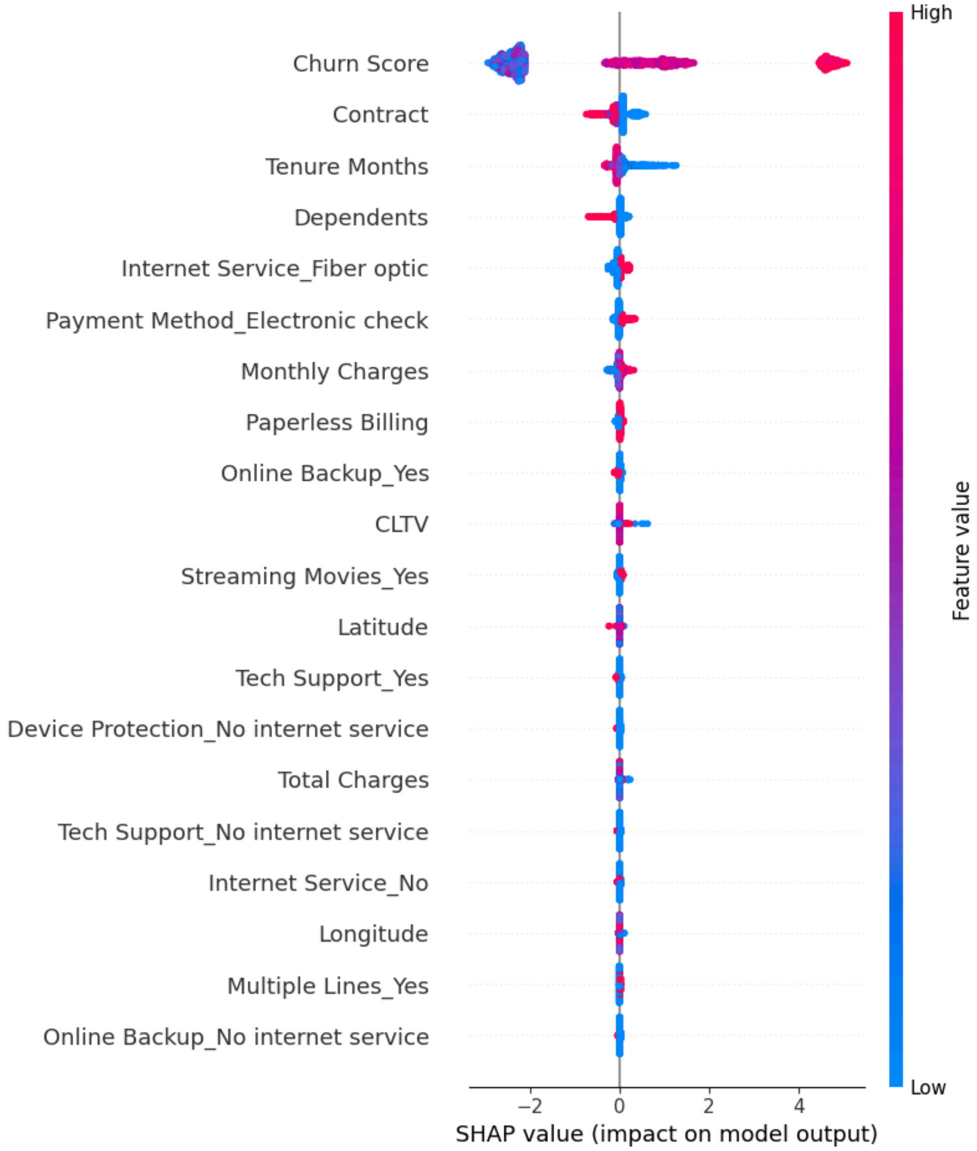


```python
import lime
import lime.lime_tabular
from IPython.display import display

lime_explainer = lime.lime_tabular.LimeTabularExplainer(
    training_data=X_train_shap_fixed.values,
    feature_names=X_train_shap_fixed.columns.tolist(),
    class_names=["No Churn", "Churn"],
    mode="classification"
)

sample_index = 0
lime_exp = lime_explainer.explain_instance(
    data_row=X_test_shap_fixed.iloc[sample_index].values,
    predict_fn=best_gb.predict_proba,
    num_features=10
)

fig = lime_exp.as_pyplot_figure()
fig.set_size_inches(10, 5)
plt.tight_layout()
plt.show()
```

Our Dataset is tabular (excel) and we need to know the general tendency and also which features individual customers want to leave. The model we chose was the GradientBoosting model, and this model is a tree-based model. The method that works best with this model is the SHAP method. The SHAP method explains the model in a consistent manner both globally and locally for GradientBoosting models. Another method we chose is the LIME method. This method is the model that best explains the factors that affect the target variable of a single customer individually. This is a model that will be used when presenting offers to individual customers as required by business strategy.

Global Explanations

SHAP Summary Plots:



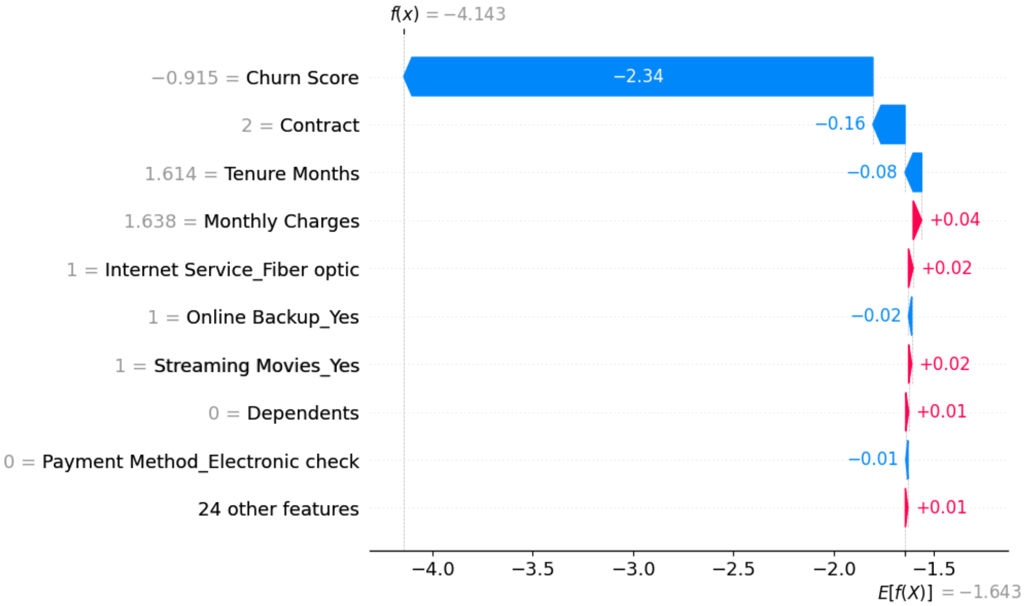
The above graph is the graph where the SHAP method provides information about the dataset globally. The horizontal axis of the SHAP graph explains the effect of the feature on the churn rate. Positive values represent features that increase the risk of churn, while negative values represent features that decrease it. We can say that the most effective feature in the graph is

Contract. It is the feature with the highest absolute SHAP value, and since it is negative, we conclude that the probability of customer departure decreases as the Contract period increases. The second most effective feature is the Tenure Months feature. This indicates that the probability of customer departure decreases as the service period increases because it has a negative value. One of the important features is the monthly charge, which has both positive and negative values. This indicates that high bills increase churn for some customers and decrease it for others. Of course, since the coefficient of each feature is different, it may not be as effective as the other features in absolute terms. The other features in the graph above can also be interpreted according to this logic. As a result, as seen in the first graph, the most effective features are Contract(negative), Tenure Months(negative) and Monthly Charges(positive).

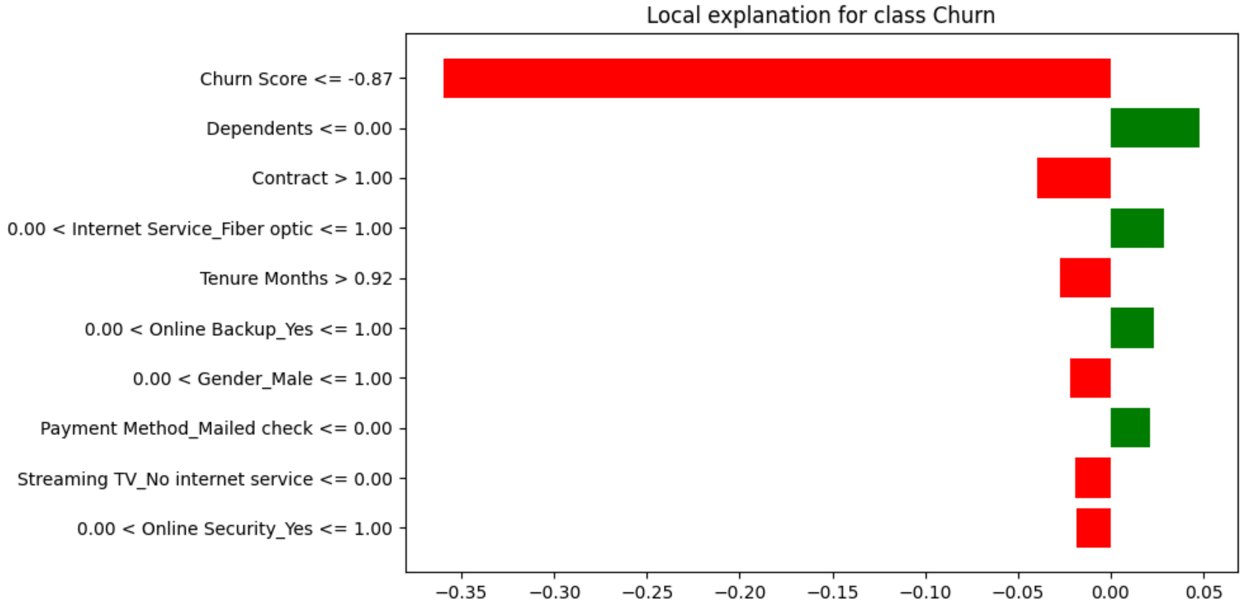
The SHAP model has shown that the information it gives us both globally and locally is consistent in the 2 graphs we examine above globally and below locally. The LIME method we evaluate below also says that it reduces the probability of separation of values such as Contract, Tenure Months. This shows that the results are consistent both globally and locally with different methods (LIME, SHAP).

Local Explanations

SHAP graph for a customer:



LIME graph for a customer:



SHAP graph is a customer SHAP analysis graph. It shows how SHAP method affects locally. F(x)=-4.143, this value is the model's estimate. E[F(x)]=-1.643 is the average estimate of all customers. Since the model estimate is much lower than the average, the model tells us that the probability of this customer leaving is quite low. We can comment that the Contract value is the feature that effectively reduces the probability of the customer churning the most. The Tenure month value also reduces the risk a little, although not as much as the contract. We see that the Monthly Charge value is very low, which tells us that the amount the customer pays monthly does not encourage them to leave that much.

The LIME graph is a graph that shows how the features of a customer according to the LIME method affect them positively and negatively. This shows that the customer has a contract of more than 1 year and the service period is also long. Since such factors are negative, it shows that the probability of the customer leaving is reduced. In the SHAP graph that we interpreted globally, these factors were again negative and reduced the probability of churn. The reason why the model gave the result that the probability of churn was low is due to these features.

Class-Wise or Subgroup Analysis

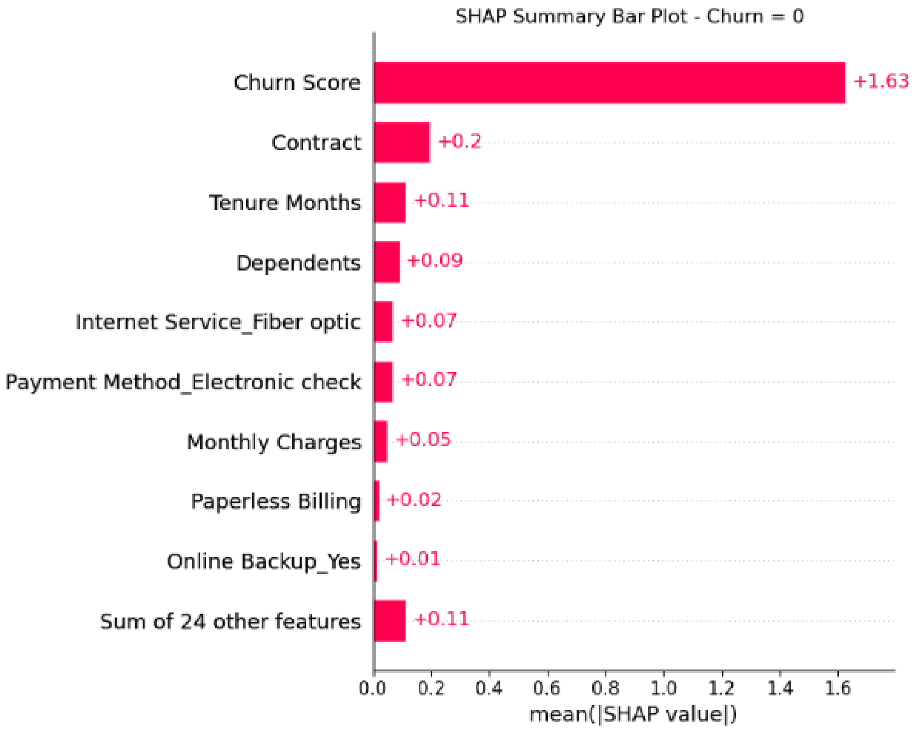
Here, we used SHAP to understand the behavior of the model we trained on different classes and demographic subgroups. Our main goal is to reveal whether it uses the same features for all users or whether it reaches different decisions for some users.

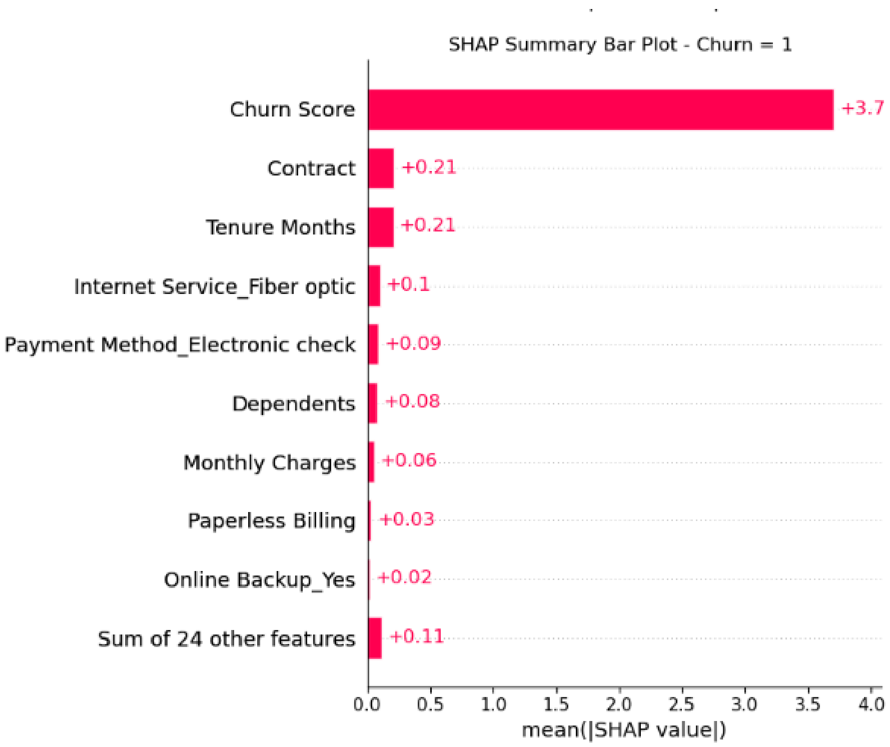
First, we categorized the test data according to the target variable churn value. As a result of this situation, our data was divided into 2 different classes.

Class 0: Customers who continue to receive service.

Class 1: Customers who stopped receiving service.

SHAP values were calculated separately for both classes and summary bar plots were created.



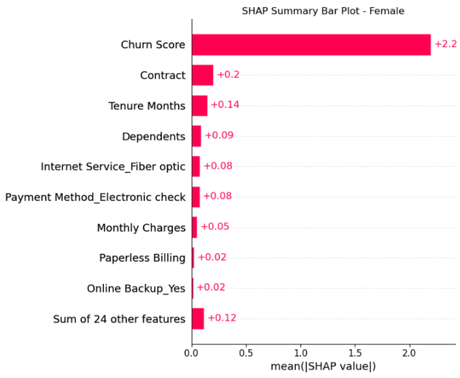


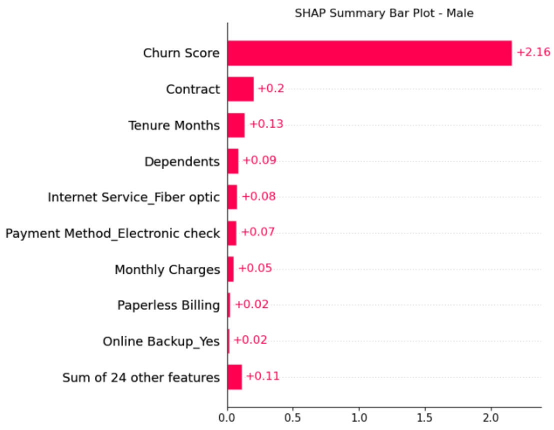
For churning customers (Class 1), the model pays most attention to the following features: Contract, Tenure Months, Internet Service_Fiber Optic. These features were evaluated as factors contributing to customer departure.

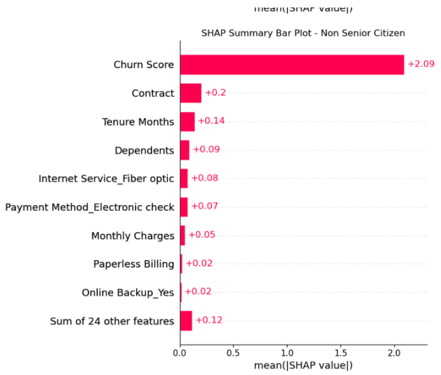
For non-churning customers (Class 0), the model interpreted the same features in different ways; for example, long-term contract and high Tenure value were effective in reducing the probability of churn.

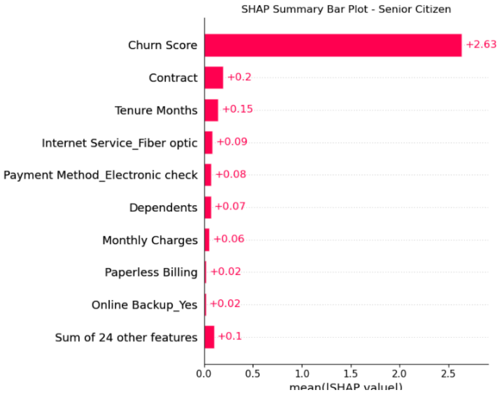
As a result, although the model focuses on some features, the effects of these features vary according to churn classes.

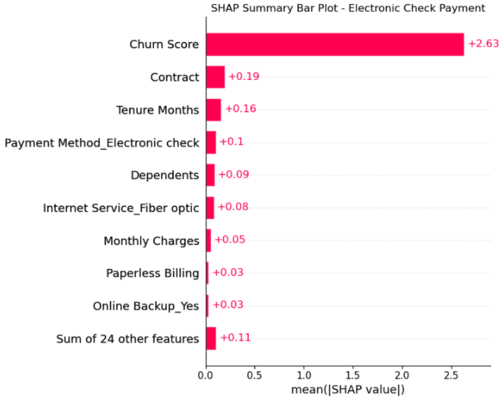
For Subgroups:

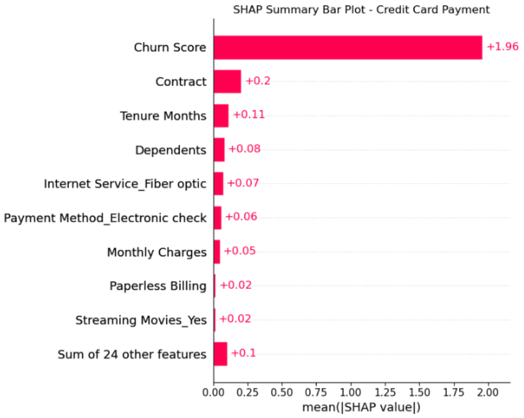


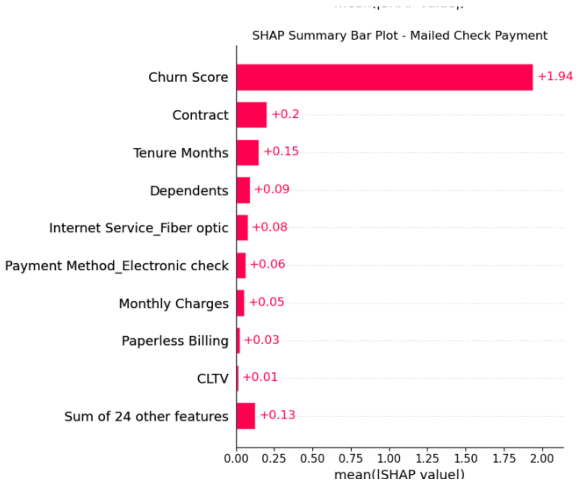












As a result, the analyses we conducted with SHAP revealed that the model works differently in different classes and user groups. Contract, Internet Service, Tenure Month and Dependents are important in each group, but their effects vary by group. Sometimes the most important is contract, sometimes the most important is tenure month. No significant discrimination was observed. However, small differences reflect the behavioral differences of the model.

Insights and Model Debugging

The model's analysis revealed some unexpected patterns and data problems. For example, it showed us that some features that were expected to have a strong effect did not have such a significant effect on model decisions. Some of the remaining features had a greater effect on the model's decision-making capacity than we expected. In addition, some inconsistencies in the data set increased our work. For example, since a feature that we expected to be numeric was marked as categorical, we had to re-convert them. When all these were taken into consideration, we observed that the model's accuracy rate increased when improvements were made to the model and data set. In addition, by re-observing the patterns created by the model, more meaningful patterns can be found and the model's accuracy rate can be increased.

Limitations of Explainability Methods Used

We used the gradient boosting algorithm in our model. We used the SHAP method for explainability. SHAP allowed us to observe the contribution of our model's decisions to each feature in detail. We used this method because we knew that our data set was small. Otherwise, SHAP reduces explainability analysis, efficiency and speed because it requires high computational cost and time in large data sets and models with many features. In addition, the Gradient Boosting model we use may sometimes contain complex structures. In this case, SHAP may not be able to capture all interactions. This leads to some restrictions and uncertainties.

Conclusion

Global Insight (From SHAP summary graph):
The results we found from the graph where the SHAP method provides a general trend on the entire dataset globally are as follows:
As the Contract period increases, the probability of customers leaving decreases significantly. (Most Critical) As the Tenure Months value increases, the probability of customers leaving

decreases. The Monthly Charges value has a mixed effect. When this value increases, the probability of some customers leaving decreases, while the probability of others increases.

The Fiber Optic Internet value increases the churn risk.

Local Insights (Both LIME and Local SHAP):

When we applied the LIME method and the SHAP method to a customer, when we applied these analyses that we saw in the SHAP graph globally to a customer, we observed that the features affected them consistently.

The impact of XAI techniques on the decision-making process has a very important role for the telecom industry. As a general trend, the Global SHAP method is one of the most important methods to be used when determining business strategy and making more general decisions. Local SHAP and LIME methods are of critical importance for the customer-based approach. When the customer relationship employee receives information that there is a high probability of churn for a specific customer, thanks to these local SHAP and LIME methods, he can predict in advance the reason why the customer may want to leave and can develop a strategy and take the relevant actions accordingly.

References

Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn prediction in telecommunication using logistic regression and logit boost. Procedia Computer Science, 167, 101-112. https://www.sciencedirect.com/science/article/pii/S1877050920306529

Lundberg, S. M., & Lee, S.-I. (2017).A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS)*. https://arxiv.org/abs/1705.07874

Molnar, C. (2022).Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.https://christophm.github.io/interpretable-ml-book/

Kayaalp, F. (2017). Review of customer churn analysis studies in telecommunications industry. Karaelmas Fen ve Mühendislik Dergisi, 7(2), 696 705.https://dergipark.org.tr/tr/download/article-file/1329508

Kumar, A. S., & Chandrakala, D. (2016). A survey on customer churn prediction using machine learning techniques. *International Journal of Computer Applications, 154*(10), 13–18.https://www.researchgate.net/profile/Saran-A/publication/310757545_A_Survey_on_Customer_Churn_Prediction_using_Machine_Learning_Techniques/links/5bb5fb8a299bf13e605e2ae9/A-Survey-on-Customer-Churn-Prediction-using-Machine-Learning-Techniques.pdf

Provost, F., & Fawcett, T. (2013). Data science for business. O'Reilly Media

Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. Results in Control and Optimization, 14, 100342. https://doi.org/10.1016/j.rico.2023.100342

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. Ch. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory, 55*, 1–9.https://www.sciencedirect.com/science/article/abs/pii/S1569190X15000386

Yean, Z. C. (n.d.). Telco customer churn (IBM) dataset. Kaggle. Retrieved April 4, 2025, from https://www.kaggle.com/datasets/yeanzc/telco-customer-churn-ibm-dataset