MANTILLA Omar

MEDINA MARTINEZ Juan Jose

COLOMBIER Remi

Professor: Philipp Borchert

Date: 11/18/2021

COMMUNICATION SKILLS GROUP PROJECT

Table of Contents

1. Insight into the company

To date, the telecommunications company has 4250 customers. Of these customers 598 have unsubscribed from the service, this is around the 14% of the customers. One of the challenges that companies are facing is to reduce the number of churns. The churn rate is a very important measure because it is much more profitable to retain existing customers than to acquire new ones. This is because it saves marketing costs and sales costs. You will get a return on retention because you will gain the trust and loyalty of the customer. We will identify through our model why these customers churn and quantify the features that influence churn.

2. Linear model: Logistic regression

Logistic regression is a very simple model that will predict the probability that a customer will unsubscribe from the service. This is exactly our business case today: we need to identify and predict the customers who intend to unsubscribe from the service in the coming days or months. To do this we will use a customer scoring algorithm. This algorithm will use all existing data on customers who have already unsubscribed and those who are still subscribed. The algorithm will target the individuals with the highest probability of churn and identify the characteristics that influence customer churn.

Logistic regression will not predict a numerical value. It is currently the most widely used method for constructing scores. Logistic regression is very simple to interpret, we obtain coefficients associated with each explanatory variable, these coefficients allow us to analyze the impact of these variables on the fact that a customer will churn or not. A brief reminder of the explanatory and explained variables of our logistic regression.

- 1 explained variable: churn
- 19 explanatory variables: state, account_length, area_code, international_plan, voice_mail_plan, number_vmail_messages, total_day_minutes, total_day_calls, total_day_charge, total_eve_minutes, total_eve_calls, total_eve_charge, total_night_minutes, total_night_calls, total_night_charge, total_intl_minutes, total_intl_calls, total_intl_charge, number_customer_service_calls.

Finally, logistic regression is very useful on the predictive aspect of the model, but this model is not performing well in terms of interpretability. Therefore, our team will propose other prediction models with different interpretations.

3. Neural Network model

Neural network model is the most accurate of the black box models. It is a model that is more accurate than the Random Forest and logistic regression. Here are the accuracy coefficients of the different models:

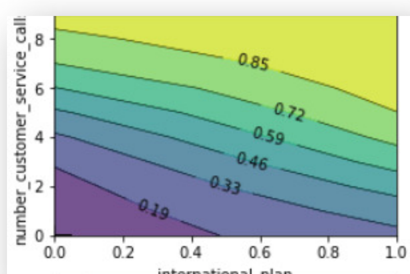
	Linear model	**Black box model**	
	Logistic regression	Neural Network	Random Forest
Train: AAC	0.8691	0.8926	0.8874
Test: AAC	0.8565	0.8824	0.8894

We set up an algorithm to identify the most important features. We will explain two of these techniques to you to show you the results: PDP and ICE because the latter are easy to understand and very relevant for making decisions and reducing the churn rate. We then selected the 6 most significant features according to the k highest scores. We measured the impact of these features on customer churn using three different techniques: Partial Dependent plot, ICE, and Shapley value. Here are the 5 highest score for the features selected.

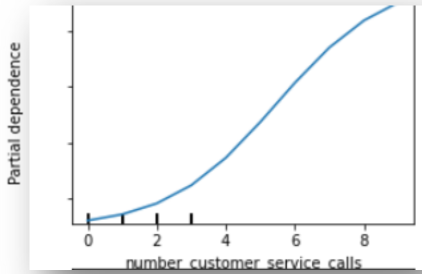
Specs	Score
total_day_minutes	3186.752474
number_vmail_messages	1012.645374
total_day_charge	541.695120
total_eve_minutes	333.275929
international_plan	258.635794
number_customer_service_calls	229.385859

A. Partial dependence plot:

These graphs make it possible to study how the variable weighs on the model's prediction. This graph measures the intensity of the overall impact of the variable on churn. This model allows you to use 2 features maximum.

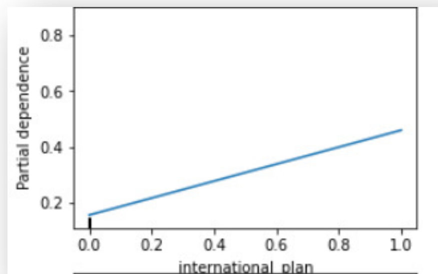


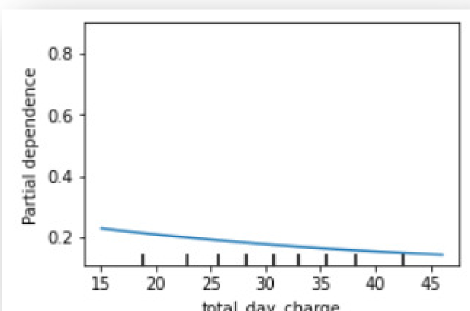
On the right we have a heatmap that visualizes the importance of the two variables **number of customers service call** and **international plan**. When looking at light colors, there is a high probability> 0.85 that a customer will unsubscribe. Unlike dark colors, it is observed that customers may be lucky to unsubscribe when there is no international plan and the number of telephone calls to customer service is low.



This graph measures the probability that a customer will churn based on the **number of customer service calls**. It is clearly observed that the more the customer calls customer service, the more likely it is that the latter will churn, which is logical.

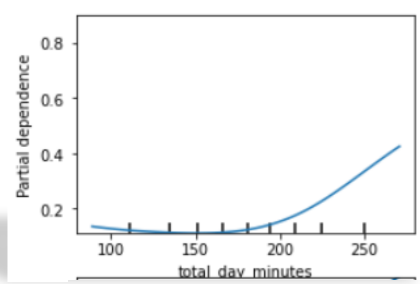
Customers who have an **international plan** have a 45% chance of unsubscribing. This is an alarming figure, and it may be worth revisiting the price and format of the international plan. It is important to identify the flaws in the international plan, you have to make sure that the telephone network is good internationally and if this is not the case then establish partnerships with the best international networks.

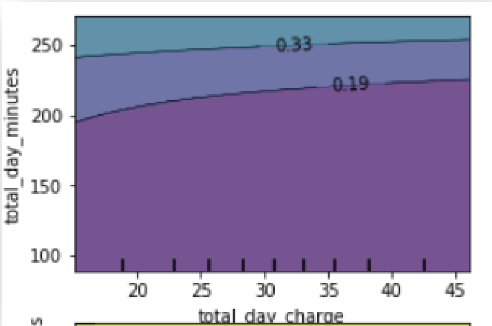




This graph shows us that the more **day charge** the customer pays, the less likely the customer will be to unsubscribe from the service. It is not logical we admit it but it is what the model tells us. This is why it might be interesting to identify the factors that influence this decline in probability of the churn rate. For example, if the total day charge is 15$ then we have 22% probability that the customers will churn whereas when the customers paid 45$ then there is less thant 10% thant the customers will churn.

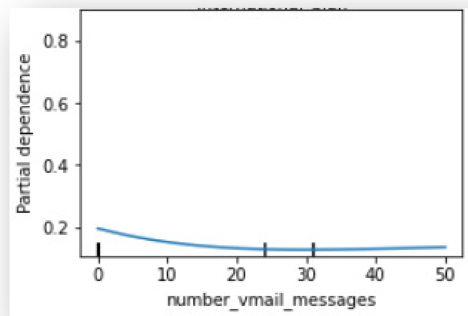
Here, the graph tells us that the more **minutes** the customer spends on the phone, the higher the likelihood of their churn. Up to 3 hours of calling, customers are satisfied with their plan and have a < 5% probability of unsubscribing. Whereas when the customer exceeds the 3 hours of phone call, he is then more apt to unsubscribe, for example when a customer spends 4h30 on the phone per day there is a 50% chance that he will unsubscribe. It is imperative to set up an unlimited plan for customers spending more than 3 hours on the phone.

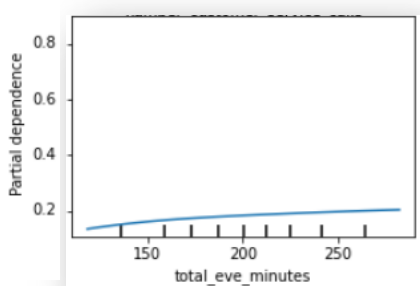




This 3-dimensional graph allows us to measure how the probability of a customer unsubscribing based on the 2 variables: **total day charge** and **total day minutes**. We can see an interesting relationship. For example, if a customer pays a $ 30 charge and spends an hour and a half on the phone then their odds of a churn are 19%. If, on the other hand, he pays the same price but spends 4h30 on the phone then his probability of churning is 33%. The total day charge variable does not affect whether a customer will churn at all. While the number of minutes will affect the customer's behavior.

A customer who has a lot of **voicemail** will tend to stay loyal with a churn rate> 10%. Whereas a customer who doesn't have a lot of voicemail is going to have an almost 20% chance of unsubscribing. It makes you think that some customers should not listen to their voicemail or use their phone very little. It might be interesting to identify with customers the reason why they are not listening to their voicemail to improve the voicemail mailbox.





Customers who spend the most **evening minute** on the phone are more likely to unsubscribe. For example, a customer who spends just 1 or 2 hours on the phone in the evening will have a <5% probability of churning. Whereas a customer who spends more than 3 hours on the phone in the evening will have a churn rate> 20%. The fees need to be revisited if we are to keep our customers who make a lot of calls at night.

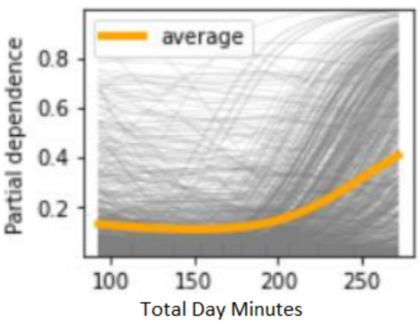
Understanding what happens in our model is very important because knowing how to understand the predictive mechanism of the model allows us to identify the reasons why customers unsubscribe from the service, and this gives us ways to improve behavior. The more total day minutes there are, the higher the probability of churning. We can clearly see that the probability of unsubscribing increases considerably when the customer spends a lot of minutes on the phone. Thus, it could be interesting to make an unlimited plan for customers who phone a lot.

B. Individual Conditional Expectations (ICE)

ICE plot is easier to understand than partial dependent plot because each line represents the prediction that a customer churn or not depending on the features, we are interested in. The big difference with partial dependence plots and that ICE plot can highlight heterogeneous relationships.

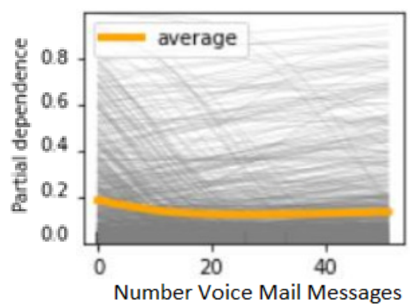
The ICE method displays one line per observation, how the sample impacts customer churn when the feature changes. The difference with the PDP is that there is one line per sample compared to an overall line for the partial dependent plot. Here we observe the same result as the partial dependent plot. The more customer calls there are, the higher the probability that the churn customer will be.

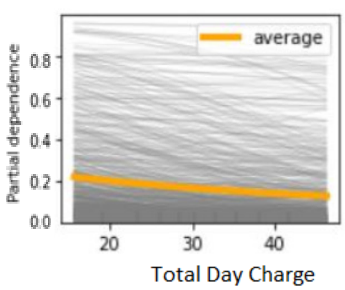
Here each line corresponds to a customer. As in our previous technique, we observe that the more minutes the customer spends on the phone, the more likely he is to unsubscribe. It is imperative to make an unlimited package for these customers in order to keep them. When we analyze the trajectory of all the customers, we observe that all the curves seem to follow the same course.



It is easy to understand that the users who communicate the most during the day are the people who most want to churn. We can read from this trend these are the people who have the most problems with their communications. It would be a good practice for the company to identify these users to further guarantee their communications.

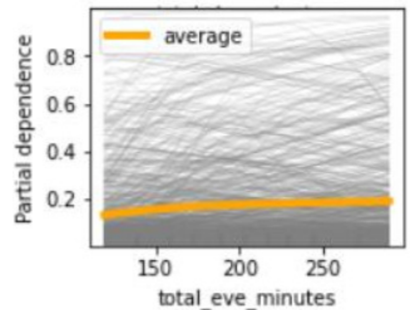
We can interpret that the users who use their cell phone the least and who keep it turned off most of the time are the people who want to unsubscribe the least. It may be because they are simply not very interested in the service or because they have coverage problems. It would be a good practice for the customer care team to contact these people and find out why this behavior and avoid future churn.

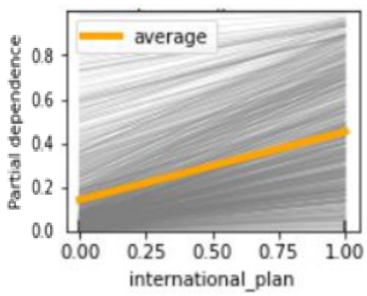




We can read from this graphic that users with the least number of minutes per day are those who most want to unsubscribe. This could be due to the fact that these people are not satisfied with the service and want to cancel it, or the company could simply approach them and evaluate their experience, thus they are able to offer packages to increase their loyalty and prevent them from leaving completely.

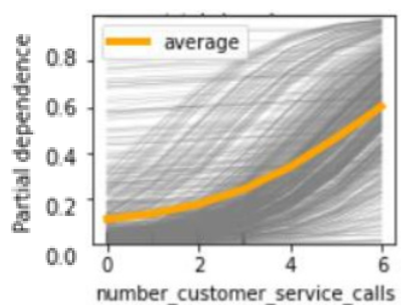
It is difficult to identify a trend on this graph. Although we can slightly see how by increasing the number of even minutes, the probability of churn minimally increases.





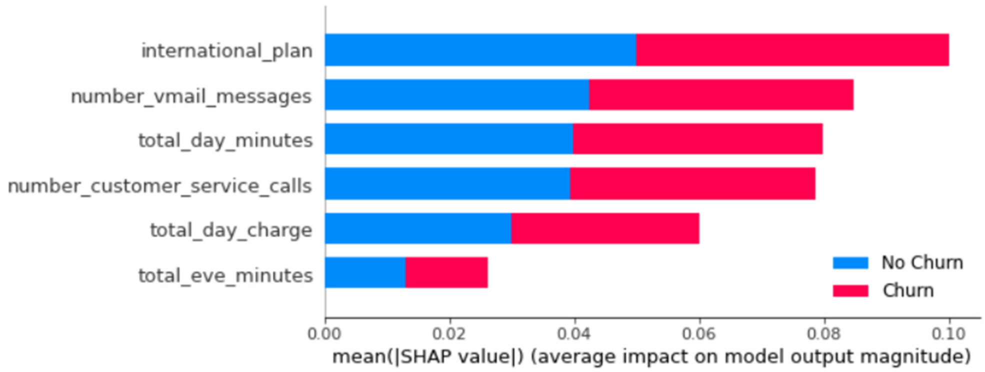
A marked trend is reflected with the users of international plans. Since the more its use increases, so does the tendency to unsubscribe. The company can act proactively with these users and evaluate the service that is currently being offered. It may be that possible service failures or uncompetitive rates are causing users to think about churn.

This graph shows a connection between the subscribers who call the support lines reporting possible issues with their services, this looks a close connection with their decision to unsubscribe. It would be a good strategy from the company to start retraining the retention staff to prevent users with service issues to finally churn.

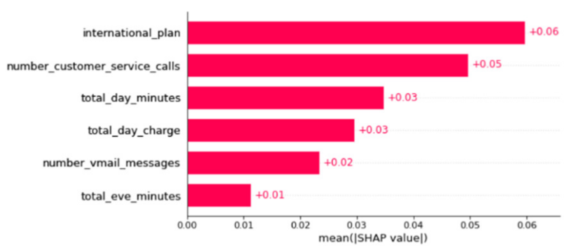


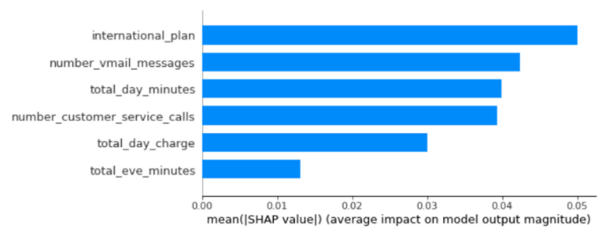
C. Shapley values

The features with large absolute Shapley values are important. To obtain the feature importance globally, we take the mean of each feature and then we sort them from the most to the least important feature.

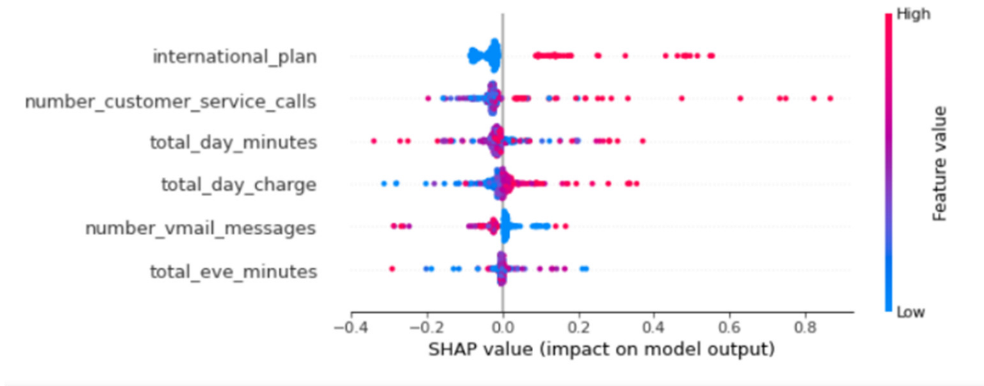


In this this graph we can see how the international plan feature is the most important to identify a clear trend for churn. This is followed by number_vmail_messages, total_day_minutes, number_customer_service_calls, total_day_charge and total_eve_minutes.



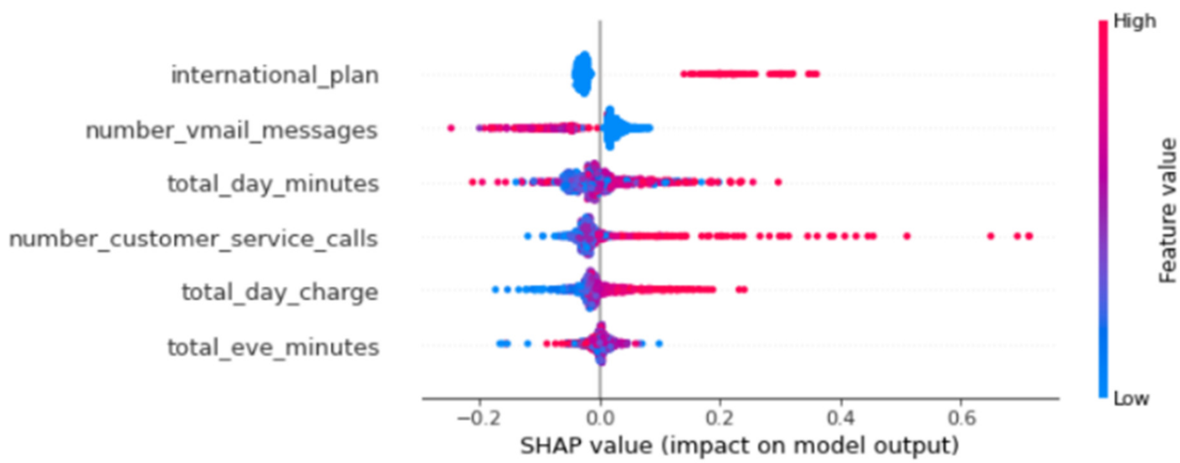


This technique also identifies the international_plan feature and the contacts made to customer service, as one of the factors that influence users to unsubscribe. This clearly indicates that this telecommunications company must strengthen and improve its international call service, whether in price or operability, and additionally, it must focus on improving its customer service to improve this experience.

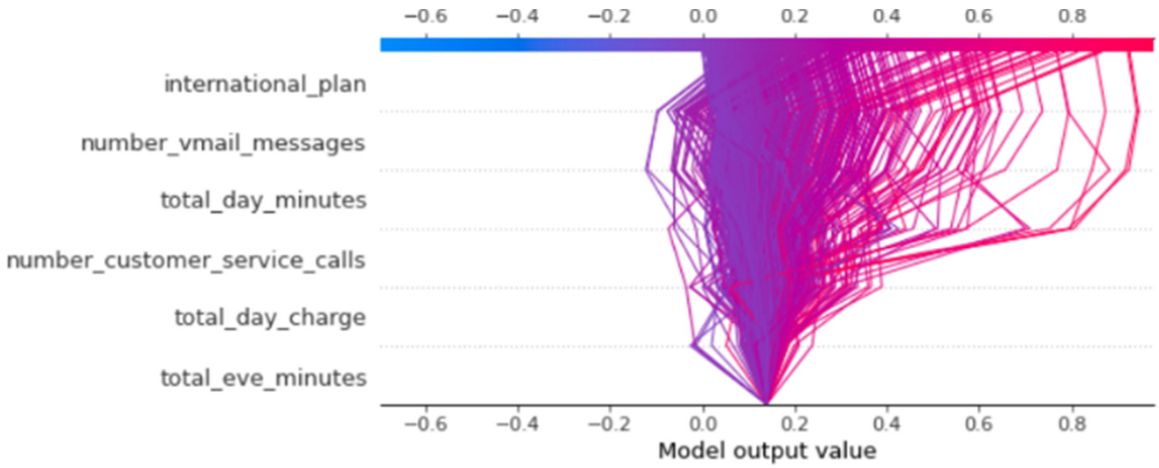


This graph shows us the features ordered by importance. To the right of this graph, we can see the feature value. On the left we have the features located according to the importance. The different dots blue and red color show us the number of observations taken by each feature and that based on this determine its importance.

Based on the information shown, we see that people using international plans greatly influence trends and probabilities so that they eventually turn into a potential churn. On the other hand, we see that users that do not use this service, do not have a significant impact on the tendency for it to be churn. In this way we can identify that the international call service is showing clear shortcomings in its performance.

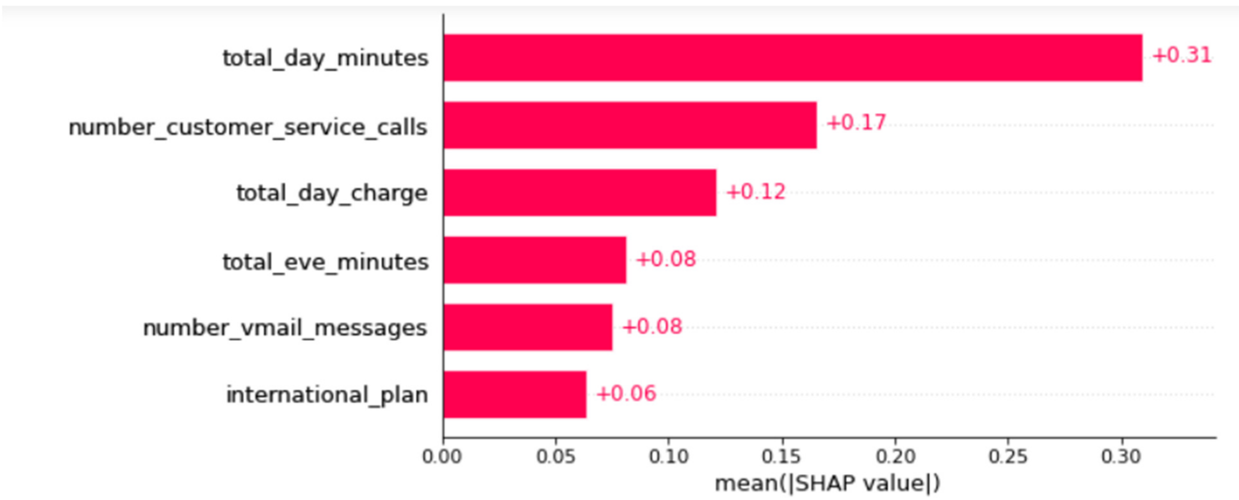


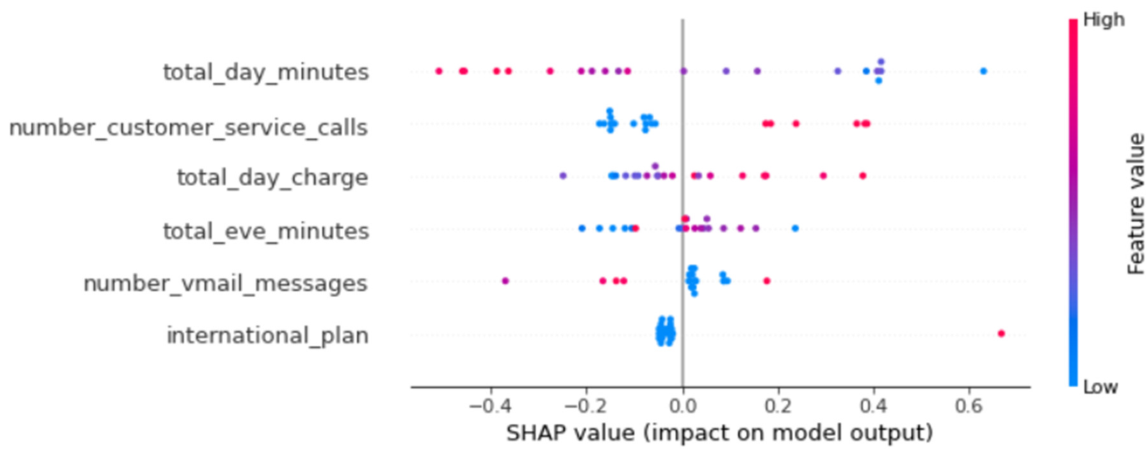
Analyzing the voice mail message feature, we appreciate that although users have a large number of this service, it reduces the probability that they will churn. Continuing with the analysis, we see in the feature of total day minutes. Identifying that the people who use this service the most are effectively the ones who most affect the probability that they become churn.



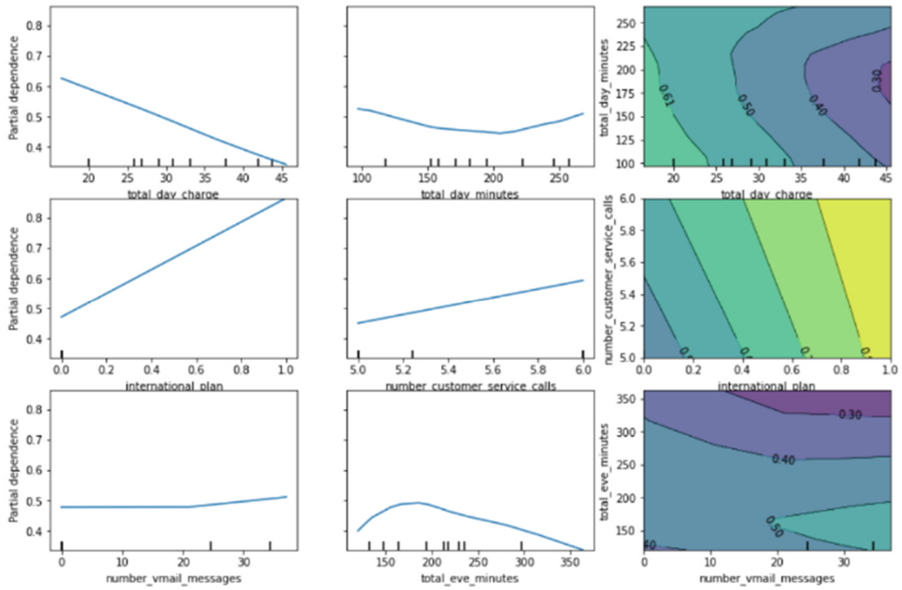
The graph above shows us how each observation or user is affected in their probability of churn or not according to the feature and its importance. In this way, we can explain that the international plan feature is the one that increases the trend the most so that it becomes churn. This once again confirms that this service shows us shortcomings in its provision and its intervention is imperative to contain the number of churners.

C.1 Subseting a population of users that call more that 4 times to the customer service line



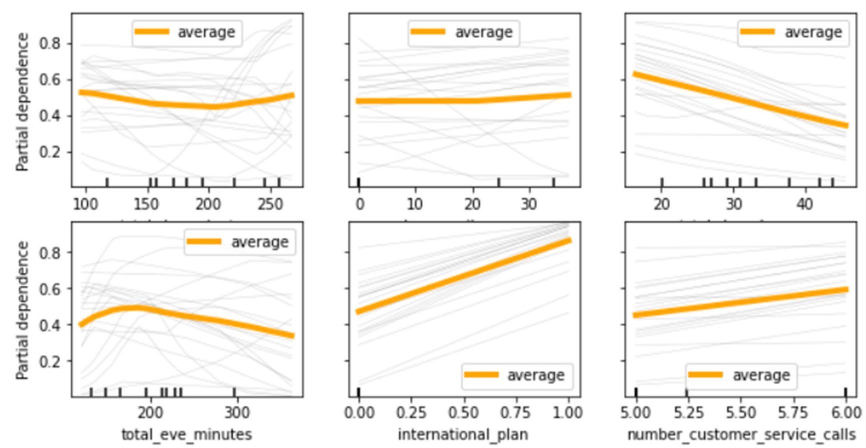


We made this subset based on users who call the customer service line more than 4 times. This shows us how this population of users has a 49 percent chance of unsubscribing. Concluding in this way that users are not having the best of solutions with a single call. The fact that a user calls repeatedly, may indicate that his requirements are not being properly addressed or that he is presenting more than one problem in his service.



Due to the feature that relates the calls to the customer service line, with the probability that a user has to unsubscribe, we want to see how each feature behaves in a subseted population of people. We could say that by doing this population subset, our information to predict would not have enough base information to give a more accurate result. For purposes of comparison with the general population, we see that certain trends are still maintained, as for example in the feature of international_plan. This is still a factor to consider since there is a high probability that these interactions with the customer service line are due to a possible failure in the provision of this service.

Additionally, we see how the trend continues with this subset of 4 people. We can say that the customer service line is one of the most important departments of each service provider company, since it is the interface between the end user and the service provider. We see that this tendency to churn is maintained in this population study. We can suggest that together with the international call service, the customer service is one of the areas that must be intervened to improve the customer's experience in this service. This line is the first line of attention of the end user and is where the problems of easy solution should be solved in the first contact.



4. Conclusion

To conclude, we have 3 features that have a strong impact on the customer to unsubscribe. These are the call time in minutes made per day, the subscription to an international plan and the number of service calls.

The suggestions of our teams are as follows:

1. Create an advantageous unlimited plan for customers making a lot of daily calls
2. Identify why customers call for customer service and increase the quality of customer service. This requires the training of customer relations staff. And also, by setting up a chat box to answer frequently asked questions and focus more on specific questions.
3. Review the international plan imperatively, you lose too many customers because of your international plan. To do this, you need to identify with your customers whether they are happy with their international plan and what the needs are. When you have identified the needs of your customers then you will have won everything because the latter will feel listened to and will therefore be retained.

5. Dashboard Description

We have presented a visual and interactive tool to interpret this report. This dashboard is made up of 9 tabs in which we highlight an abstract to detail the current situation regarding the linear prediction model used by the company.

Then you will find an insights tab where we highlight the information highlighted by this model.

Subsequently we have added a tab for each linear model used: Logistic Regression, Random Forest and Neural Networks

Finally, you will be able to find the different interpretation techniques applied after selecting the Neural Networks model since it indicated an accuracy value greater than the others.

6. References

https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e

https://discuss.streamlit.io/t/this-is-how-to-use-sweetviz-with-streamlit/10897