Towards churn prediction on TELCO operators

Churn; Top-up; Machine learning; SHAP; Customer retention

White paper

Version 1.0, March 2022

Introduction

Despite their stable maturity, current telecommunication operators have some herculean challenges in acquiring and retaining clients. For example, in countries with market penetration sometimes higher than 100% of the population, it is very important to retain customers. Additionally, studies show that the cost to retain a customer is lower than the cost associated with acquiring new ones [1].

The goal of this work is to detect customers about to leave the TELCO operator and create a predictor model. Furthermore, identifying the underlying causes of the client's departure is considered a major advantage. It opens the model black box, adding knowledge about the problem and allowing for customized campaigns, using this knowledge to create the best offer or campaign to target specific clients. To answer these questions, we resort to a dataset of raw events of a European TELCO operator. Using this data and performing some feature engineering to capture patterns in customer behavior, different models were created to predict customers that will not make any kind of activity in the next 30 days.

This work focuses mainly on pay-as-you-go clients with uneven top-ups during their lifetime. Throughout this article, the methodology used is described along with the results and how to attain insights into why some clients will leave the company. The framework relies on cutting-edge state-of-the-art algorithms to predict and explain the reason behind clients leaving an operator. Results show that the prediction algorithm achieved good performance in major metrics. Therefore, this algorithm could be an important add-on to current TELCO operation systems, being able to predict and mitigate clients' departure.

Related work

We are, by no means, the first to explore churn aiming to gain insights related to customers abandoning a company. In this section, we present some articles that are related to (and also inspired) the work presented in this article.

Neural networks and images

Hotels can automate and fully digitalize the inIn [2], a variant approach of usual churn prediction was studied, where customers and their actions are codified in images. Each image is associated with a distinct customer and is composed of three different parts: the first one contains the customer's ID to identify each customer; the second one has the events converted to red-green-blue (RGB) code; the third one contains the information whether that customer is a churner or a non-churner. The image is codified as follows: a table containing the values of each day per row, and the number of events (calls and messages) and recharges in each column. The values are then codified to RGB code (different values have a different codification and color).

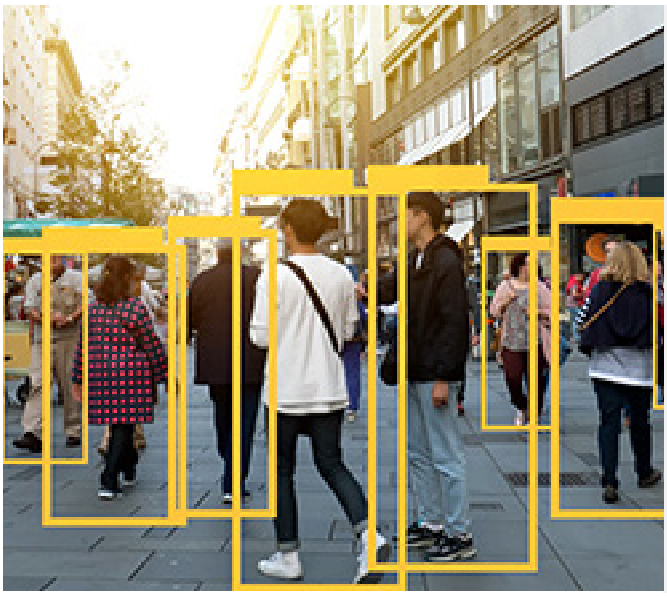
Then, a neural network is implemented, with the images as the inputs of this network. This method is different from standard approaches, where it is used tabular information from datasets. Although the authors present good results, the heaviness of the algorithm and the need for computational resources are unclear.

SHAP

In [3], a new approach to interpreting machine learning models is proposed, making the impact of each variable on the model and on each output clearer.

In this article, the shapley additive explanations (SHAP) method is presented to understand the behavior and contribution of variables globally and individually. Globally, a graphic is represented with all variables, arranged in a decreasing form of impact, and with the positive or negative impact of this variable on the model. A graph is also made for each variable, where it is possible to see the impact of the fluctuation of the values taken by that variable. Individually, it is shown how to represent the score associated with each customer's ID with the impact that each variable had to lead the customer to have that score.

This approach was used in this work and is one of the foundations of explainable machine learning, leading to new insights viable to downstream systems to make the best offer to the client.



Sliding window

In [4] it is proposed a new study with a novel approach through a sliding window.

In this work, the sliding window is studied by varying the training set in the time window both in period and size. Thus, it is possible to understand how the model is affected when the training set period is longer or shorter and when the window of this set is slid. It is also possible to see the impact on the model when doing some pre-processing of the data and some statistical analysis.

Proposed methodology

This section describes the data sets used in this proof-of-concept, the methodology used, and finally, the feature engineering made over the raw events.

Churners

In a concise definition, churn corresponds "to the loss of customers for a given service in a certain period of time" [5]. Another important definition is the churn rate that measures the percentage of customers leaving the company and consequently the end of their journey with that specific product or service. Bearing this in mind, the identification of churners is an important aspect for any company delivering some sort of product.

In this specific research problem - TELCO companies - the goal is to detect and predict which customers are more likely to leave the service. Altice Labs' Active Campaign Manager product [6] will be responsible for all the campaigns carried out to retain these customers, as it presents all the functionalities needed to interact and give products and services to final customers.

Companies tend to distinguish between voluntary churn and involuntary churn. Voluntary churn corresponds to a voluntary departure on the client's part, either to change to another service or to move to another company. Involuntary churn is used when the reasons for departing are external to the customer, e.g., change of home location or death. In this case, involuntary churn is not included in the analyses and in the studies carried out, focusing on voluntary churn, which usually occurs due to factors related to the downfall of the customer-company relationship and which the company is generally capable of reacting.

Another important aspect of churn is the division into two major groups:

A group of customers with a loyalty contract, generally designated as postpaid clients, with a recurrent monthly invoice being charged to the customer.

A group of customers with pay-as-you-go mobile phone cards, generally defined as prepaid clients, that have non-regular events (e.g., recharges, voice calls, or SMS).

This work will focus on prepaid clients since postpaid clients' churn is a more straightforward problem to solve: there is a loyalty contract between the customer and the TELCO, with a clear termination date. Prepaid clients, on the other hand, provide us with a more challenging problem. The group is more heterogeneous and showcases an irregular lifecycle with higher uncertainty associated with their revenue, which will test the limits of our approach.

Although it is possible to pinpoint clients that will leave a company after several months without events (e.g., 90 days) with high accuracy, this approach may sometimes identify them too late. Thus, it seems more appealing to detect churn in earlier stages, as downstream systems responsible for interacting with the clients will not be able to retain them since they already have abandoned the company.

Considering the reasons mentioned above, the decision was to consider a prepaid customer as a churner if he/she had not made any recharge or event in a certain period of time, i.e., 30 days.

Data used for churn detection

This study used three types of customer data: historical data of top-ups, events (calls, messages, and mobile internet), and the customer profile.

The data is a daily aggregation of events and recharges. It contains information related to the amount of the top-up, calls made, and duration, amongst others. All sensitive fields were anonymized for this work, complying with the European General Data Protection Regulation (GDPR), ensuring no personal information was used. Afterward, we added other features related to the customers' profiles, and applied filtering and aggregation to build the characteristics of those profiles.

Feature engineering

Data preparation and quality play an important role in machine learning-based solutions, and the adoption of data-centric strategies can substantially leverage model performance. Thus, a feature engineering approach was taken starting with the datasets described above and finishing with a dataset containing many features calculated from these records. These new features have different granularities (1 month, 3 months, and 12 months) and aim to capture changes in customer behavior over time.

Feature importance [7] was used to measure the benefit of adding the new variables. Weight of evidence and information value [8] were also used for feature selection [9]. The weight of evidence measures the contribution of each individual feature for the target (in this work, churn) and ranks their predictive strength. This method is usually performed pre-modeling, enabling to inspect variables for bias, strange patterns, and measuring the effect of missing values, thus identifying the best strategy in each situation. This approach allowed us to identify the top 27 more relevant features for the problem, which significantly improved the results, as shown in the next section.

Methodology

The methodology used in this work to train and test the models consists of two steps: firstly, we used the data described in the 'Data used for churn detection' section, selecting month M to train the model. The classification of non-churner or churner is made according to the activity or inactivity of the customer in the next month (M+1); secondly, data from month M+1 is used to test the model, where predictions are validated using churners and non-churners from month M+2. This methodology is illustrated in **Figure 1**.

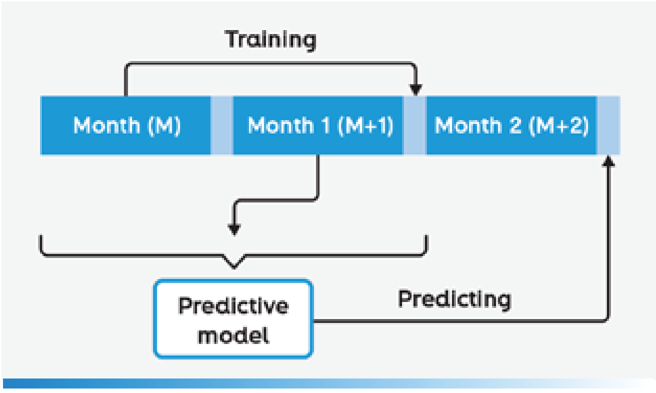


Figure 1 – Methodology used in this study

As an example, thinking in M as June, M+1 as July and M+2 as August, the model is trained with the data of June and the information of July to classify each customer as churner or non-churner. After that, this model is used with the data of July to predict the churners of August. This methodology can be used as a sliding window to predict the following months.

To evaluate this approach and to understand the models' trustworthiness in predicting churn customers, several classification metrics were used as precision, recall, and f-score [10]. These metrics compare the model result with the observed outcome (classification in churn vs. non-churn), resulting in true positives, false positives, and false negatives present in the list of churners. The threshold (TH) used to separate the two classes was chosen to optimize one or more metrics mentioned before. The classification of true positives, false positives, false negatives, and true negatives was done according to the rules present in **Table 1**.

	Predicted churner >= TH	Predicted churner < TH
True churner	True Positive	False Negative
True non-churner	False Positive	True Negative

Table 1 – Classification of true churners and predicted churners for a specific threshold

Results

This section presents the results obtained using four different models, along with the results using SHAP values to interpret the final model and determine which variables have the most impact on the results. It is important to notice that we are only presenting the final results, for the sake of conciseness.

Models

Choosing which models to use was a task that started with a broad list, from which models were culled if they did not show good enough performance (precision, recall, F1-Score, accuracy). In the end, we were left with four different algorithms: XGBoost [11], LightGBM [12], Random Forest Classifier, and Gradient Boosting Classifier. We used the same data to train and test all algorithms, obtaining comparable results, as presented in **Table 2**.

Algorithm	Precision	Recall	F1-Score	Accuracy
XGBoost	65%	61%	63%	89%
LightGBM	48%	84%	61%	83%
Random Forest Classifier	63%	65%	64%	89%
Gradient Boosting Classifier	54%	78%	64%	86%

Table 2 – Metric values of the different algorithms with a threshold default (50%)

With these results, a graphical representation was built to evaluate the impact of different thresholds in precision, recall, and accuracy for all four models so that, when varying the threshold value, we are able to understand the behavior of these metrics and which models would be the best to continue this study. This is shown in **Figure 2**. In this phase, a model is considered to have a good performance if it presents linear and stable recall and precision in all the spectrum of the threshold. Hence, the marketeer can vary the threshold depending on the campaign purpose.

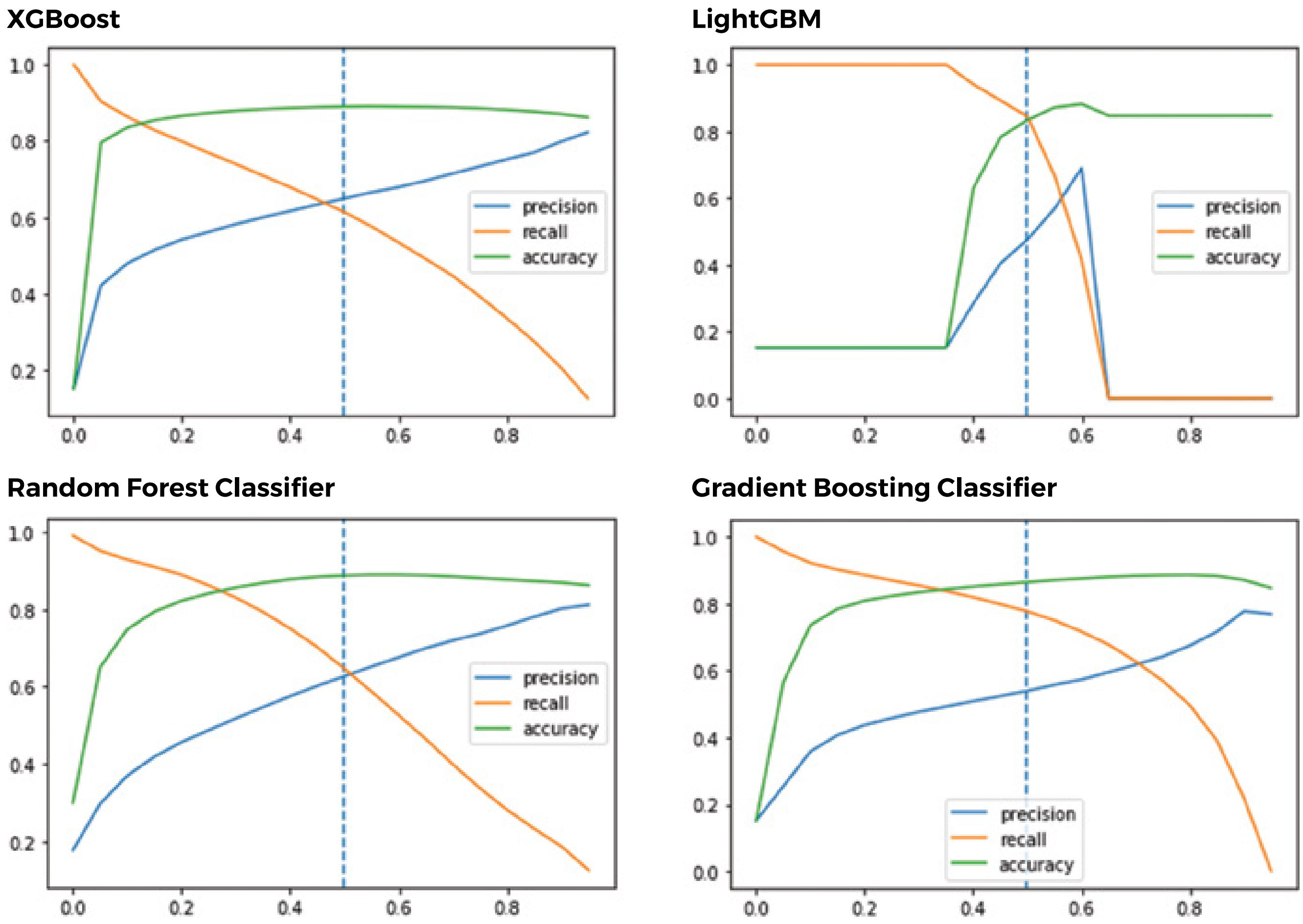


Figure 2 – Model performance across different thresholds

According to the results obtained in **Table 2** and **Figure 2**, XGBoost and Random Forest Classifier were chosen to move on. Both presented encouraging results, and therefore it was necessary to understand which of the two is the best model. Thus, the selected models were compared in three different and consecutive months, and the results are presented in **Table 3**, which shows that the results of the XGBoost and the Random Forest Classifier models are very similar.

Since we have a higher slope for lower thresholds, which allows having a wider range of results to suit each campaign and following the principle of Occam's razor or the law of parsimony ("when you have two competing theories that make exactly the same predictions, the simpler one is the better") [13] we decided to choose the XGBoost model. This is due to the fact that this model allows to apply a threshold that presents linear results on the combination of recall and precision.

XGBoost

Month to predict	Precision	Recall	F1-Score	Accuracy
February	65%	61%	63%	89%
March	67%	67%	67%	90%
April	69%	70%	70%	90%

Random Forest Classifier

Month to predict	Precision	Recall	F1-Score	Accuracy
February	63%	65%	64%	89%
March	66%	69%	67%	90%
April	67%	73%	70%	90%

Table 3 – Metric values of the two best algorithms with a threshold default (50%)

Model explainability

In this study, the need arose to interpret the model and understand how the variables impacted our model and outputs. SHAP was used to explain our model in a global and individual way. Globally, the model can be explained by **Figure 3**, where it is possible to observe the variables represented in decreasing form of impact on the model.

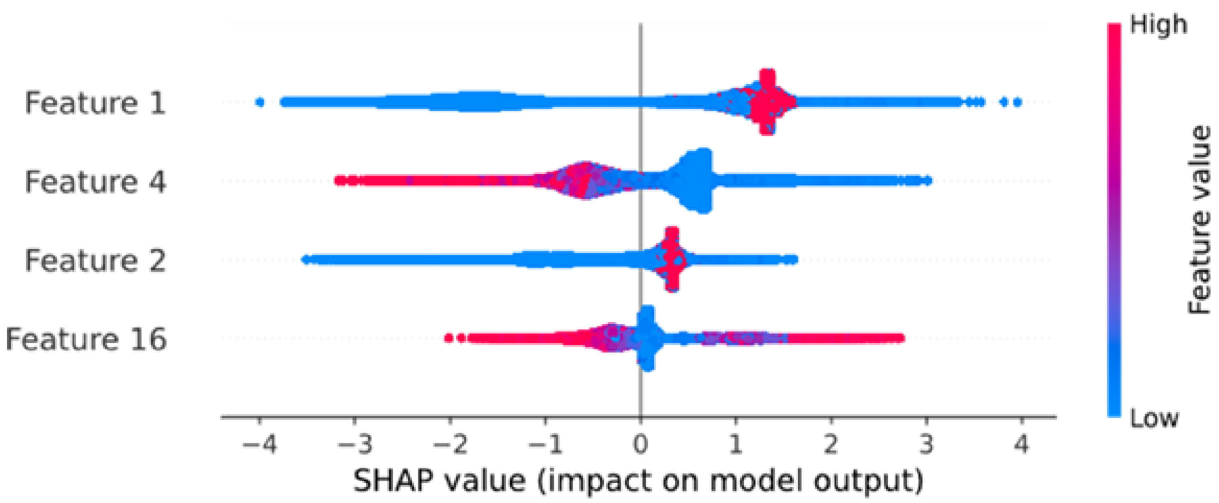


Figure 3 – Global impact of the features on the model using SHAP values

Can also see that some features have a very clear impact on the model: those that are more reddish points on the right, which positively impact our model, i.e., the higher the value of this feature, the greater the probability of a customer to do churn; also, those that are more reddish points on the left, which negatively impact our model, i.e., the higher the value of this feature, the lower the probability of a customer to do churn.

The impact of a feature on the output can be seen in **Figure 4**. In this figure, it is possible to see that lower values of the feature are associated with lower values of the final output. Higher values of the feature are associated with higher values of the final output.

Individually, it is possible to understand which features contributed the most to the customers' scores. For example, in **Figure 5**, it is possible to see a customer whose score starts at the average score of all customers. Its value increases or decreases when we add the features associated with this customer. Red features increase the score value, and blue features decrease the score value. In the end, we arrive at the probability of this customer to do churn. It is also possible to see which features had the most positive and negative influence, and which had the greatest impact on the final score.

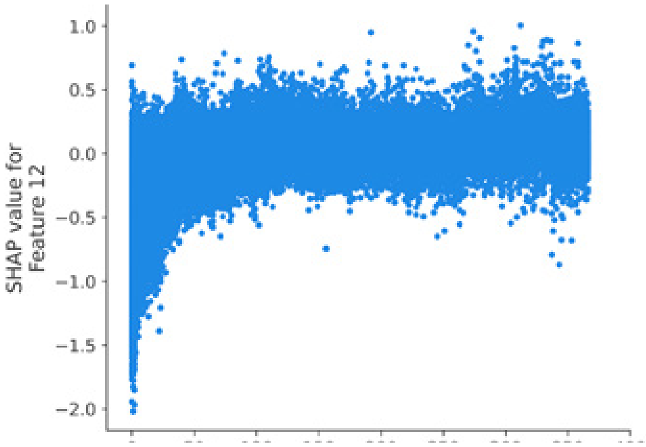


Figure 4 – The impact of a feature on the output

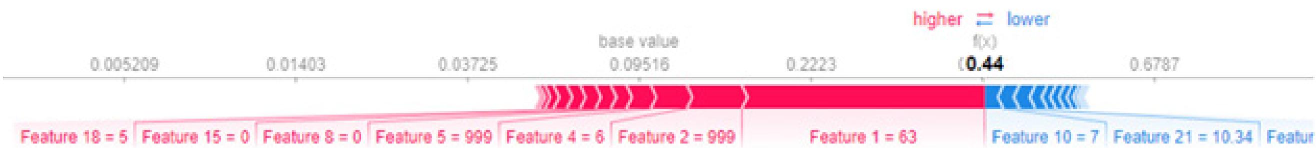


Figure 5 – Individual impact of the features on the model using SHAP values

Conclusion and future work

In a world full of services and with the cost of acquiring a new customer much greater than retaining customers already in the operator, detecting who is about to leave a company proves to be a crucial differentiator factor in TELCO companies. This is a trend on the rise with company investment in this field increasing.

This study clarifies that churn prediction is possible and very feasible through raw datasets and feature engineering. The built model reveals a great performance, and the obtained results allow us to choose the strategy that better suits the campaign purpose. The fact that the model is supported by previous studies with an academic base already demonstrated and consolidated shows even greater viability for this project.

In the future, we pretend to develop and build a pilot project with a European TELCO operator to understand better different campaign strategies. Using A/B tests to test different models and test which ones provide better results will further increase the knowledge about this method and its advantages, drawbacks, and lifts for the TELCO company.

We also intend to validate the results, the model, and our strategy with other operators outside of Europe with a vast prepaid client base, to prove the robustness and adaptability of our model in different social contexts.

References

[1] S. Syed, H. Halaoui, F. Hicham, M. Mahmoud, and M. Mansour, “Analytical imperatives for telecom marketers in emerging markets,” vol. 11, no. 2, pp. 19–36, 2014

[2] A. Wangperawong, C. Brun, O. Laudy, and R. Pavasuthipaisit, “Churn analysis using deep convolutional neural networks and autoencoders,” pp. 1–6, 2016, [Online]. Available: http://arxiv.org/abs/1604.05377

[3] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” ***Adv. Neural Inf. Process. Syst.***, vol. 2017-December, no. Section 2, pp. 4766–4775, 2017

[4] A. K. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” ***J. Big Data***, vol. 6, no. 1, p. 28, Dec. 2019, doi: 10.1186/s40537-019-0191-6

[5] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, “New insights into churn prediction in the telecommunication sector: A profit driven data mining approach,” Eur. J. Oper. Res., vol. 218, no. 1, pp. 211–229, Apr. 2012, doi: 10.1016/j.ejor.2011.09.031

[6] Altice Labs, “ACM Active Campaign Manager,” 2020. https://www.alticelabs.com/products/campaign-management/

[7] Fabian Pedregosa et al., “Scikit-learn: Machine Learning in Python,” ***J. Mach. Learn. Res.***, vol. 12, pp. 2825–2830, 2011, [Online]. Available: https://scikit-learn.org/stable/

[8] L. Zhang, H. Ray, J. Priestley, and S. Tan, “A descriptive study of variable discretization and cost-sensitive logistic regression on imbalanced credit data,” ***J. Appl. Stat.***, vol. 47, no. 3, pp. 568–581, Feb. 2020, doi: 10.1080/02664763.2019.1643829

[9] K. Larsen, “Information Package Vignette,” ***The Comprehensive R Archive Network***, 2016. https://cran.r-project.org/web/packages/Information/vignettes/Information-vignette.html

[10] D. L. Olson and D. Delen, ***Advanced Data Mining Techniques***. Springer, 2008

[11] XGBboost Developers, “XGBoost Documentation,” 2021. https://xgboost.readthedocs.io/en/latest/

[12] “LightGBM,” Microsoft Corporation, 2021. https://lightgbm.readthedocs.io/en/latest/

[13] W. M. Thorburn, “Occam’s Razor,” ***Mind***, vol. XXIV, no. 2, pp. 287–288, 1915, doi: 10.1093/mind/XXIV.2.287

Authors

Bernardo Duarte

Data Scientist

Altice Labs, Portugal

bernardo-x-duarte@alticelabs.com

https://www.linkedin.com/in/bernardoxnduarte/

Ana Mestre

Data Analyst

MEO, Portugal

ana-m-mestre@telecom.pt

https://www.linkedin.com/in/ana-maria-mestre-48b2293/

Ricardo Filipe

Data Scientist | Software Engineer

Altice Labs, Portugal

ricardo-a-filipe@alticelabs.com

https://www.linkedin.com/in/ricardo-filipe-0a589787/

Nuno Pimenta

Data Team Leader

MEO, Portugal

nuno-pimenta@telecom.pt

https://www.linkedin.com/in/nspimenta/

Contacts

Address

Rua Eng. José Ferreira Pinto Basto
3810 - 106 Aveiro (PORTUGAL)

Phone

+351 234 403 200

Media

contact@alticelabs.com
www.alticelabs.com