

Using Decision Tree and Grid Search Optimization to Predict Customer Interest in Purchasing Travel Insurance

Otávio Ligeiro¹, Bruno Guedes¹, Luis Toledo¹ e Leonardo Meira¹

¹ Escola Nacional de Ciências Estatísticas (ENCE), Rio de Janeiro, Brazil

With the guidance of Eduardo Corrêa

Doctor's degree in Computer Science at UFF and professor at ENCE

Abbreviated abstract: Machine Learning appears as a powerful tool to overcome the crisis in the tourism sector caused by COVID-19. In this work we address the following problem: a travel agency wants to offer travel insurance packages to its 1,987 customers, who have information registered in the company's database. However, the agency would like to conduct a deeper analysis to find out which people have an initial interest to avoid unnecessary spending on the budget for marketing campaigns. Therefore, to predict the customer's interest in acquiring the service, we used the Decision Tree algorithm and Grid Search Optimization technique, which automates the process of finding the best set of hyperparameters, improving model performance.



Related publications:

- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, v. 12, p. 2825–2830, 2011.
- ZAKI, M. J.; MEIRA JR., W. Data Mining and Machine Learning: Fundamental Concepts and Algorithms. 2. ed. Cambridge: Cambridge University Press, 2020.

brunoguedes.nave@gmail.com - 1



3rd Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
October 28-30, 2024

Problem and Motivation

- The COVID-19 pandemic could cause a loss of over \$4 trillion to the global GDP for the years 2020 and 2021.¹
- The most optimistic scenario shows it would take at least two and a half years for the tourism sector to have the same number of tourist arrivals as 2019.²

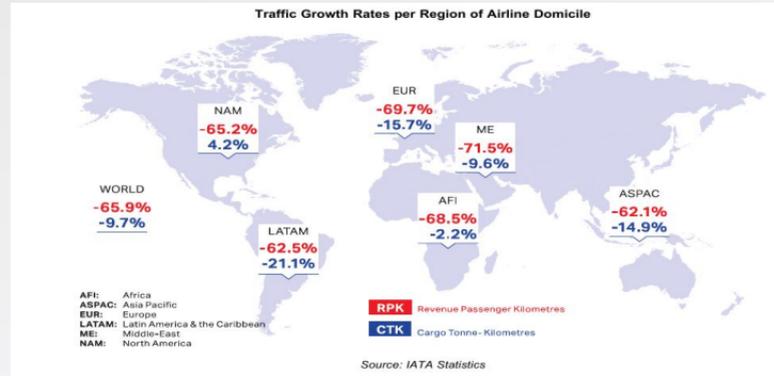
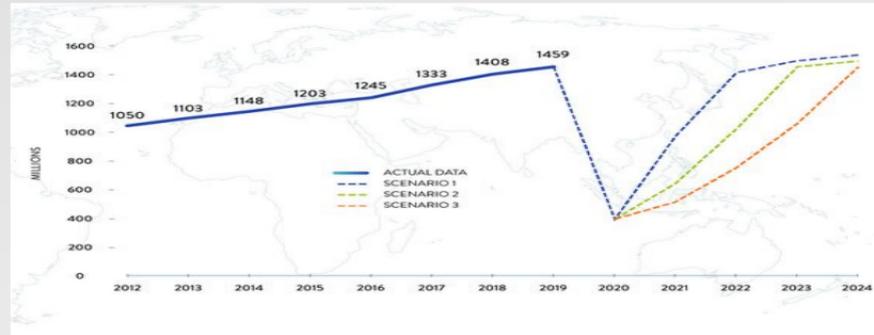
In this work, we created a model to predict customer's interest in acquiring a travel insurance using Decision Tree and Grid Search Optimization.

Source¹ and Source²: United Nations Conference on Trade and Development (UNCTAD)

Source³: [UWNTO](http://www.uwnto.org)



2021 - 2024 SCENARIOS³
2.5 TO 4 YEAR TO RECOVER 2019 LEVELS OF INTERNATIONAL TOURIST ARRIVALS

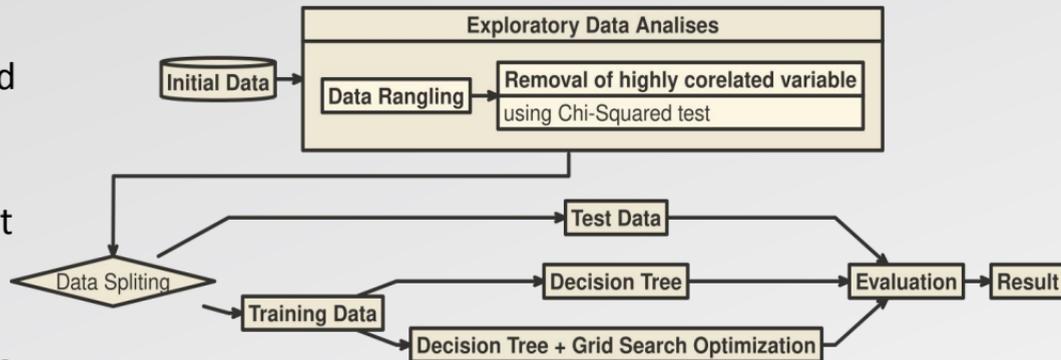


3rd Conference on
**Statistics and
Data Science**
Salvador, Brazil (online)
October 28-30, 2021

Methodology

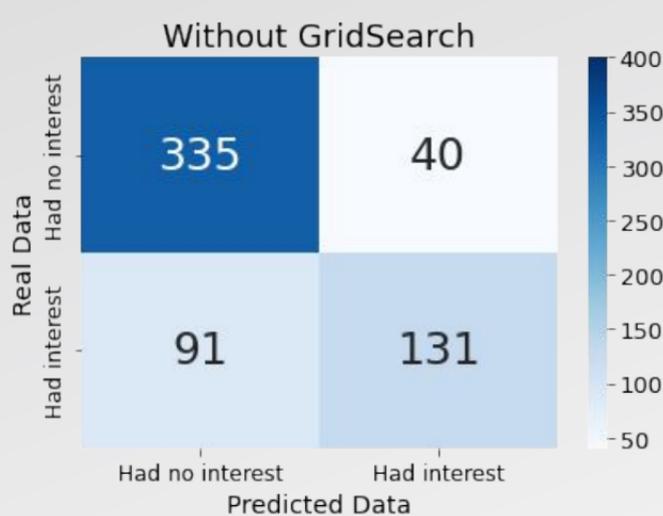
The Decision Tree Algorithm is a rule based classification method that works by splitting the data into those observations that follow an arbitrary rule and those that doesn't and, at the same time, evaluating those rules using a loss function, which is, in our case, one of the parameters that the model can receive.

In our methodology, the main objective was to show how the use of Grid Search Optimization can bring better results than those found without the use of it. The Decision Tree algorithm has numerous parameters that we can use to avoid overfitting or underfitting the model. The technique of Grid Search Optimization will perform several iterations that testing all values we passed in which the parameters can take to find the best possible match and achieve a better result than before.

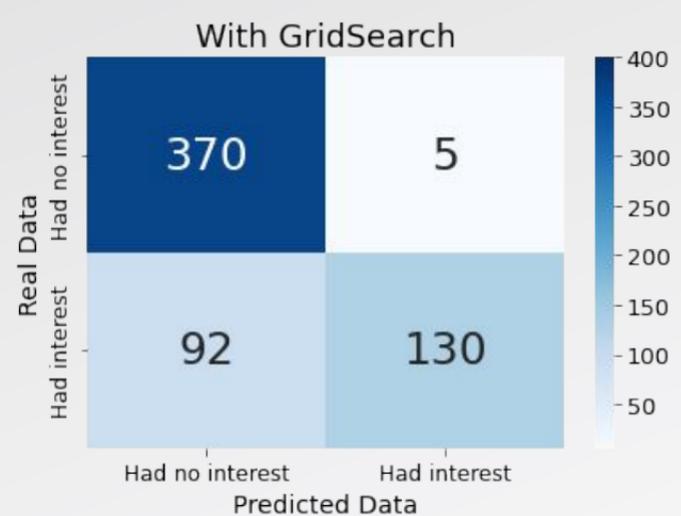


Results

To evaluate the model, the most relevant metric was Precision, since our main goal was to mitigate the number of false positives.



Without the use of Grid Search Optimization, the model achieved an Precision of 76.61%.



With Grid Search Optimization, the Precision increased by 19.69%, reaching 96.3%.

