Predicting outcomes in Multiple Sclerosis through Machine Learning using data from Pharmaceutical Consultation

LUZ SAÚDE

Cardoso, P (1); Santos, C (1); Costa, F (1)

1) Luz Saúde, Lisbon, Portugal

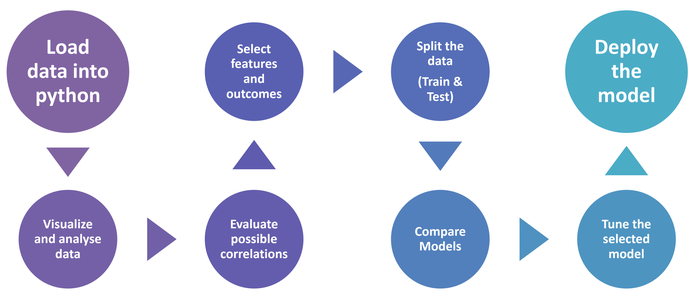
Introduction

Pharmaceutical consultation is part of the clinical pathway of all Multiple Sclerosis (MS) patients in our hospital. This event occurs at the start of the new drug and, recurring, every 3 months. For this purpose, a dedicated database was created in Microsoft Excel and Microsoft Visual Basic, to allow the register of clinical information (Date of diagnosis, number of previous relapses, Expanded Disability Status Scale (EDSS) and type of disease), demographic information (Sex and Birth Date) and Pharmaceutical Information (Number of previous treatments and Drug Related Problems). This process generates structured data that can be easily analyzed.

The process of therapeutic approval at the local level is done by the Pharmacy and Therapeutic committee is usually relies only on clinical trials data. Using real world evidence, even at a small scale, could improve de decision process.

Primary Objective

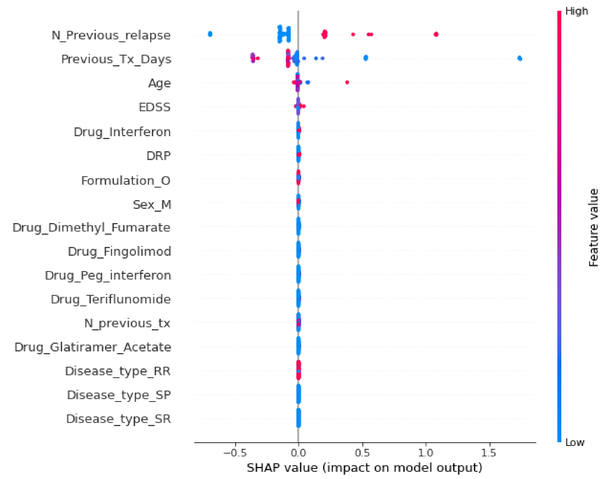
Use pharmaceutical consultation data to predict relevant outcomes using Machine Learning algorithms.



Methodology

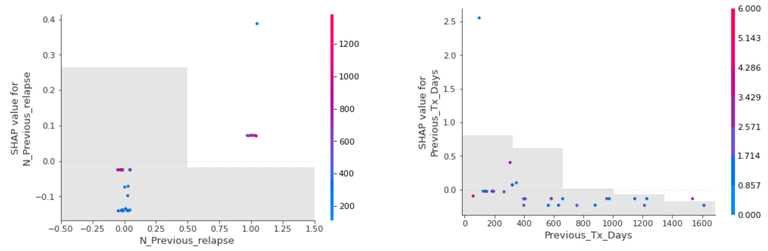
- Collected data between 2016 and 2020;
- Data was loaded and analyzed using VSCode (Microsoft) on an environment running python 3.8.8 and Jupyter Notebook kernel;
- Selected **outcome to predict was Annualized Relapse Rate (ARR);**
- Compared 4 Regression Models: Lasso, Logistic, Support Vector Machine and XGBoost;
- Selected metrics for model comparison were Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and $R^{2;}$
- Hyperparameters were tuned to minimize overfitting.

Final Model

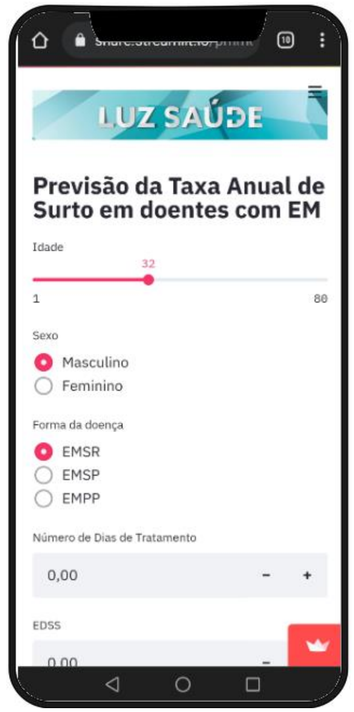


Feature Importance using SHAP methodology

- XGBoost was the best performing model;
- Most important feature was the number of previous relapses (magnitude 0.3);
- **The final model has a MAE of 0.06, a RMSE 0.01 and R^2 of 0.9**.



SHAP analysis for the Number of previous Relapses and the duration of previous treatment



Final model deployed as a Web-app

Conclusions

Despite the small dimension of the dataset, it is possible to use machine learning algorithms to predict relevant outcomes with good performance and adequate fit. The use of this model will allow for tailored clinical intervention and better decision support.