Asthma Recurrence and Associated Risk for Omnivida S.A.S

Codebase Summary

Colombia | Team #59

Andrés García, Andrés Mejía, Ariel García, Catalina Moreno, Gilberto A. Morales Zamora, Sergio Gómez

Codebase contents:

The content of this document lists all the code files that were used in the different analyzes carried out on the project, as well as the code with which the dashboard in AWS was developed. the document is divided into three parts. The first is associated with all the initial analysis and data cleaning of the databases received by Omnivida, the second part has the development file of the model and the third part has the file with the codebase for the development and execution of the dashboard.

1. Data Preparation & Cleaning.

This section list all the files used to the first exploratory analysis.

- **1. Explotatory Time Analysis.ipynb:** Contains a time analysis of the data. The information provided by Omnivida has the records of the patient at different points od time.
- **2. Analysis Continuous Covariates.ipynb:** This code provides some basic EDA to specific continuous variables.

After an initial data set review, each one of the 18 data sets where adjusted, by cleaning and modifying some information, also, some rules were defined to rename the column names of the files. this will guarantee the different merges and joins of de data. the next Jupiter notebook files contain all the work done:

- **3.1 Adherence Clean.ipynb:** Information about adherence records.
- **3.2 Family Records.ipynb:** Information about family background.
- **3.3 Quality Life.ipynb:** Test of quality life aspects.
- **3.4 Biological Clean.ipynb:** Info about the dispensation of biological drugs.
- **3.5 Farmaco Ram Clean.ipynb:** negative drug reaction history.
- **3.6 Med Clean.ipynb:** Info about treatment drugs for asthma.
- **3.7 ACT Clean.ipynb:** The ACT results of the patients.
- **3.8 Clean Others.ipynb:** Additional information about the patients.
- **3.9 Vacc Hosp Clean.ipynb:** History record of vaccines and hospitalization.
- **3.10 Clean Medical Records.ipynb:** Additional medical information.

The next files contain different procedures built to joins the clean data, create new variables, and some prior functions to generate patient reports.

- **4.1 Adherence Count Procedure.py:** Create some count variables related to the number of adherence in different ranges of time.
- **4.2 SQL Export.py:** Procedure to export the generated data in the RDS database.
- **4.3 Med Procedure Function.py:** Function to show the record of the adherence for each one of the most used medicaments to asthma treatment.
- **4.4 Train_test_split_Adherence.py:** Script that split the information built in two sets, one for estimation model an the other to validate it.
- **4.5 Join Clean Data Function.ipynb:** This procedure performs the joins of the different clean data generated by the previous work.
- **4.6 Adherence Mark .ipynb:** Creates an alternative adherence mark based on the persistence of claim (medications). This is the objective variable in our proposed Persistence Score.

2. Model Development.

After all the cleaning process and the consolidation of the information, the estimation of a model that predicts the probability of adherence is developed by using machine learning algorithms and the code contained in the file, have the step by step process of this exercise:

- **5. Models Develop Code.html**

Our project contemplates the construction of two models, one for the adherence and the second one for the persistence of claim. The file above contains the construction of the adherence score, but for the persistence score, we use the same methodology.

- 6. **Models Develop Code M2.html**

3. Application Base Code - Dash.

The project contemplates the development of a dashboard that contains two informative modules aimed at different types of users (medical personnel/health entities). The creation of this dashboard was developed with the tools included in Dash and it was deployed in a cloud environment on AWS. The following document contains all the code built for the development of the application:

- **7. Dash_documentation.html**