30 APRIL, 2023

HEALTH CARE - PERSISTENCY OF A DRUG

Group Name: DG23
Names of Members: Rahul Saini, Fatima Ujunwa, Chinmay Arvind

BACKGROUND - HEALTH CARE - PERSISTENCY OF A DRUG

- Problem Description : One of the challenge for all Pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification

- Objective : To gather insights on the factors that are impacting the persistency of the drug as per physician description

- Business understanding: Building a classification ML algorithm that learns insights on the factors that are impacting the persistency of the drug as per physician description

DATA COLLECTION

- Type of data:

The dataset contains the following information as show below

Total number of observations	3424
Total number of files	1
Total number of features	68
Base format of files	csv
Size of data	901KB

DATA EXPLORATION AND PREPROCESSING

Problems of the data :

- The data had no missing data points and so there were no problems identified in the data.
- The column names were also devoid of white spaces
- Each column contained accurate data types.

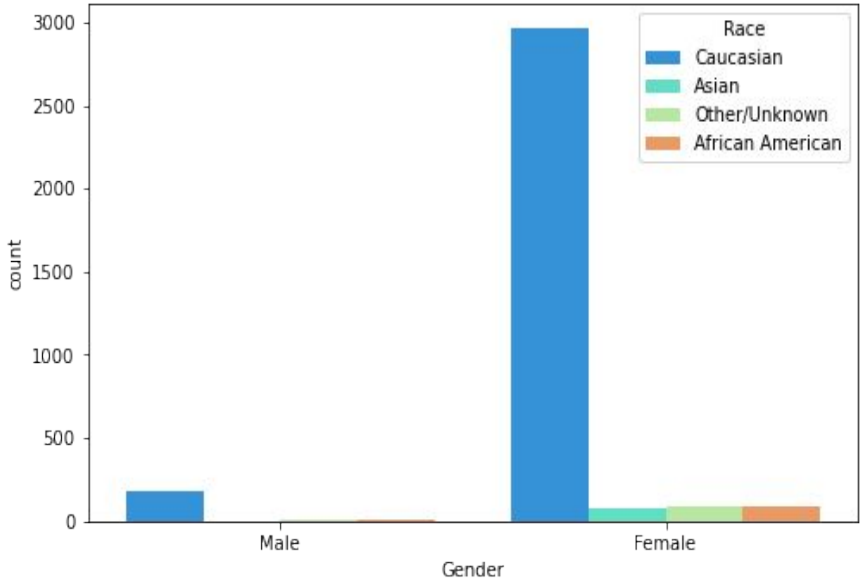
Approaches to overcome problems in data:

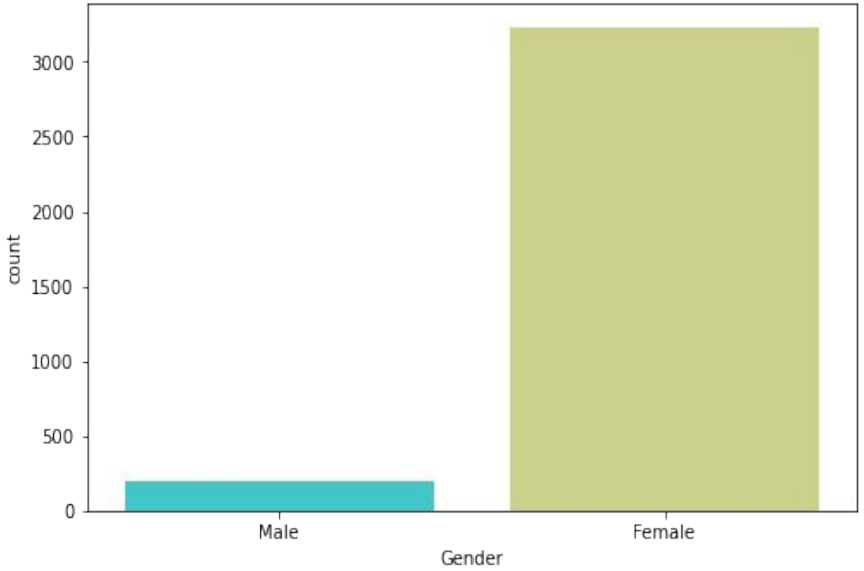
- There were no approaches used to overcome problem in data since there were no problems in the data

DATA EXPLORATION

From the figures below we can establish that:

- There is a very high margin between the number of women and men in the data set.
- Caucasian Female has the most captured data on the data set.
-

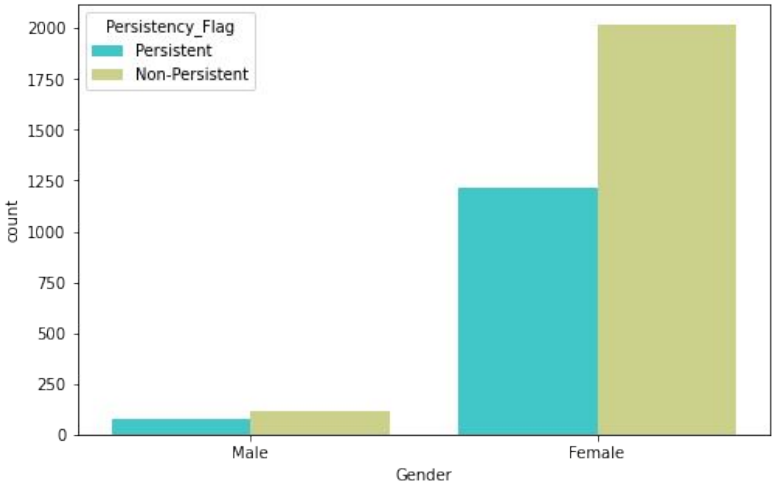


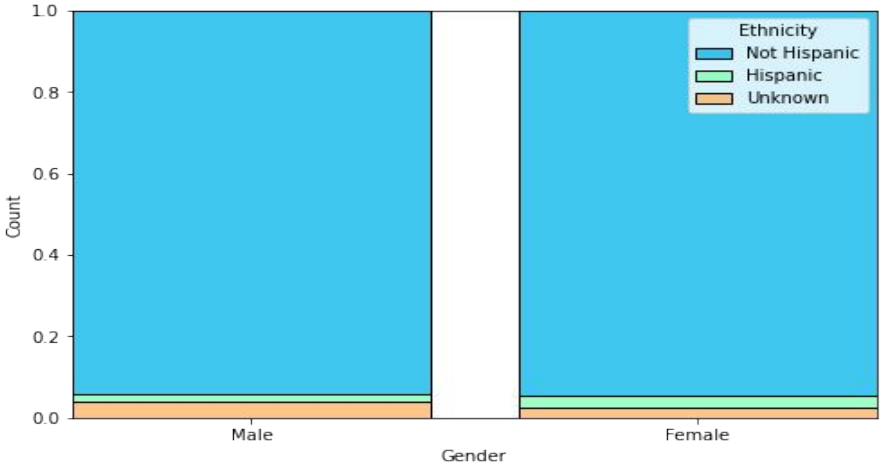


DATA EXPLORATION

From the figures below we can establish that:

- female patients are generally more persistent with the NTM Rx than their male counterparts
- Non -hispanics hold the highest population in the data set

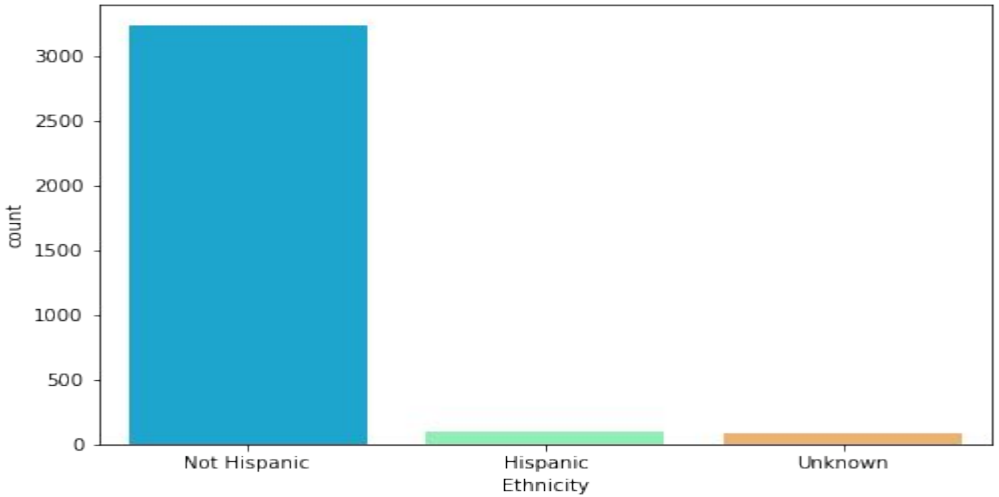


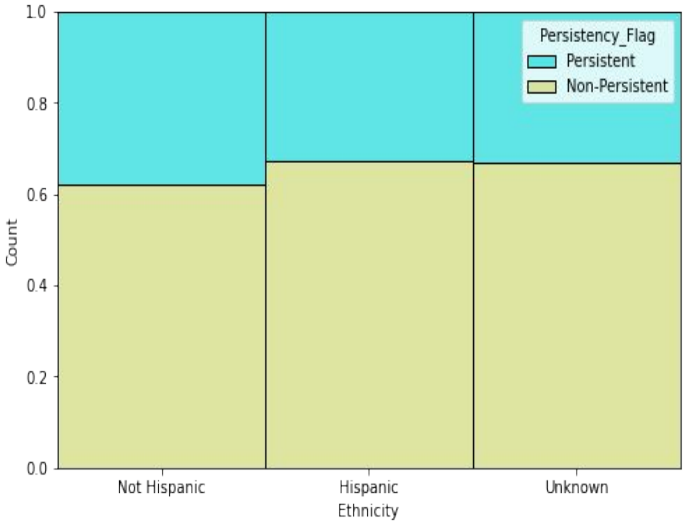


DATA EXPLORATION

From the figures below we can establish that:

- individuals who are not hispanic by ethnicity are more persistent in taking the NTM Rx than individuals who are hispanics which is a good thing seeing that the number of non-hispanics exceeds that of hispanics

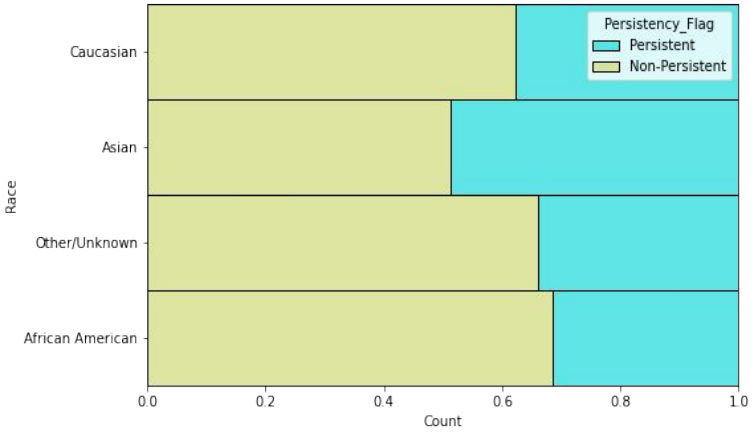


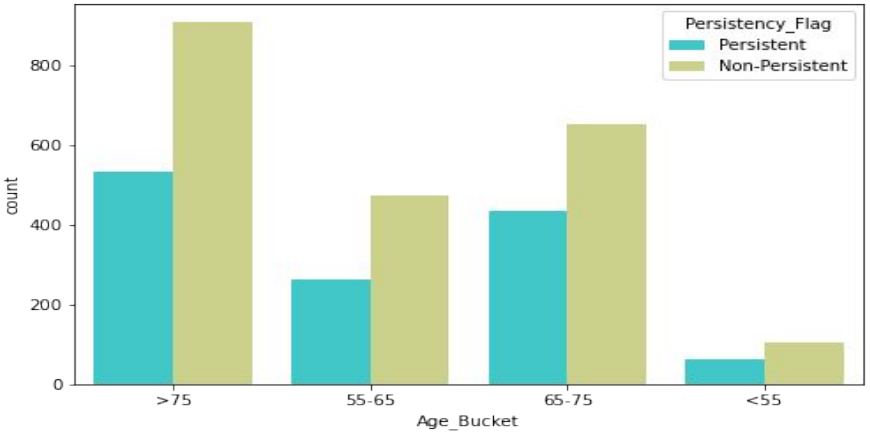


DATA EXPLORATION

From the figures below we can establish that:

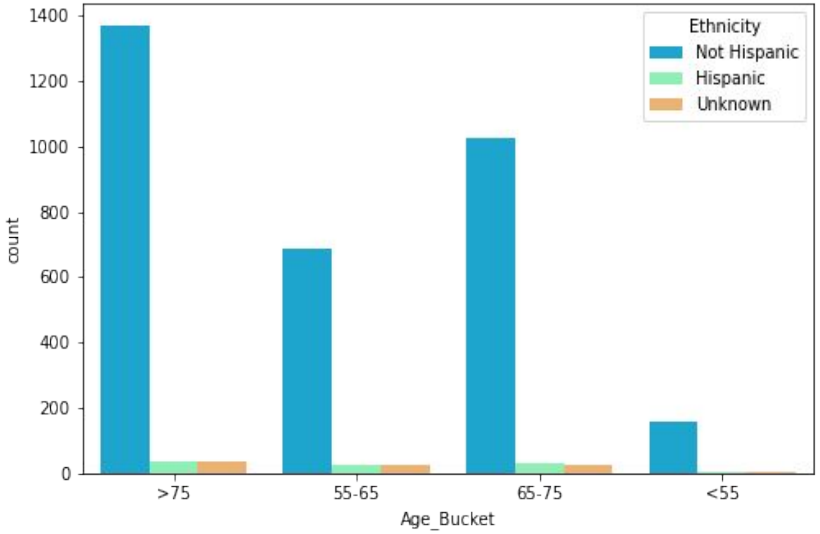
- According to the dataset, Asians are more persistent in taking NTM Rx than all of the other races. They are followed by the caucasians, with the African americans being the least.
- We see that the age bucket greater than 75 years old are the most persistent with NTM Rx given. This is in contrast to the age bucket less than 55. However we also see that when it comes to non-persistency, age bucket >75 has the highest number of individuals who are no consistent with the NTM Rx while the age bucket <55 has the least number of non-persistent patients

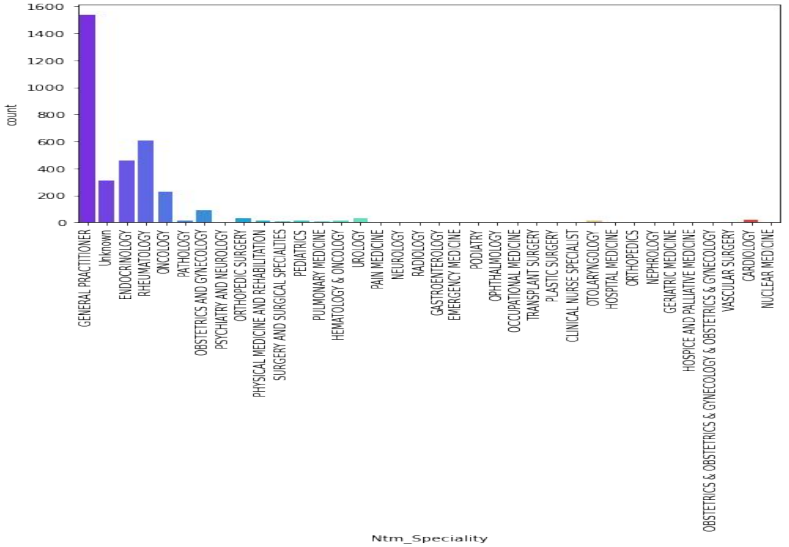




From the figures below we can establish that:

- Most patient that were prescribed the NTM Rx by HCP with Specialty of a General Practitioner; followed by Endrocrinologist and Rheumatologist. Very few Patients that were administered same therapy by other HCP Specialist other than Endrocrinologist, Rheumatologist, General Practitioner, Oncologist, Obsetrics and Gynecologist

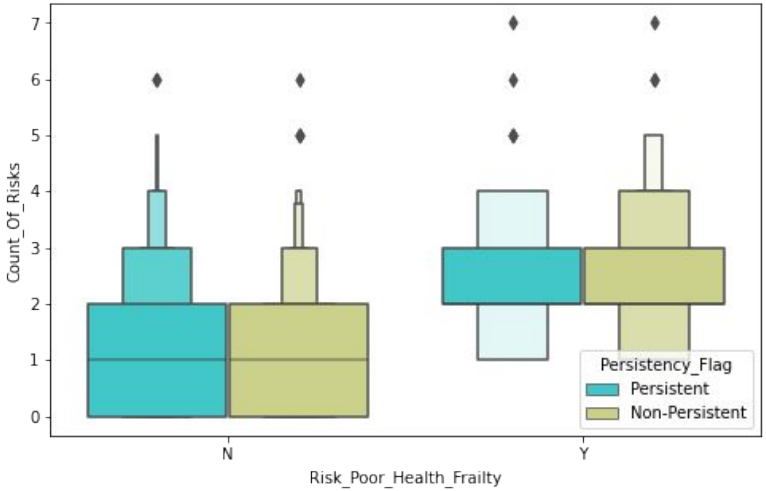


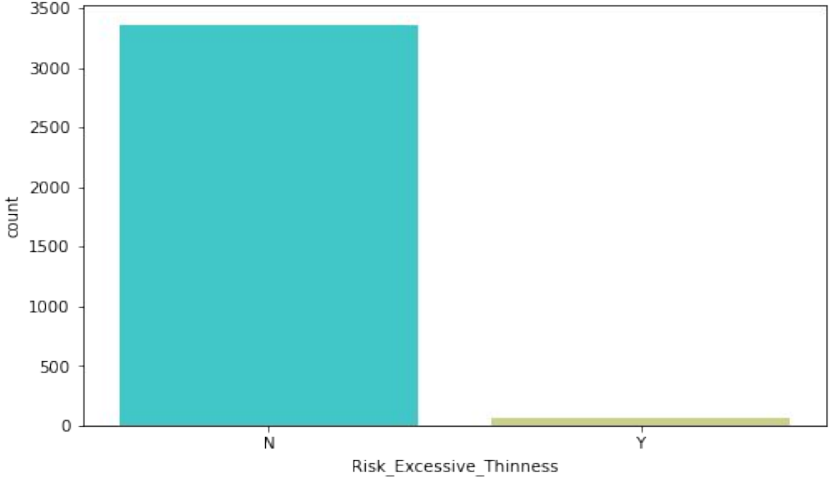


DATA EXPLORATION

From the figures below we can establish that:

- From the figure the above we see that we see that patients with Risk of Poor Health Frailty have a higher risks counts than patients without Risk of Poor Health Frailty
- There are Lower number of patients with Risk of Excessive Thinness the highest population in the data set





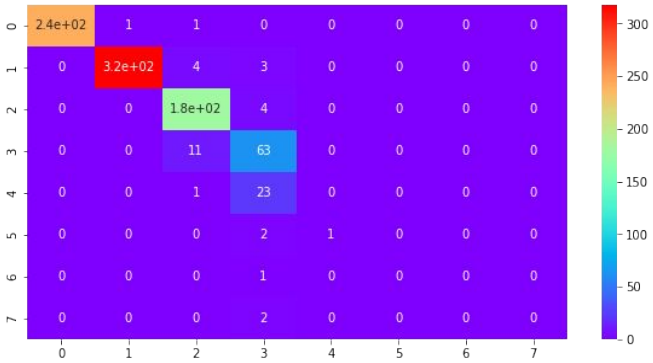
MODEL BUILDING AND EVALUATION

For Building the model building, the following algorithms were used:

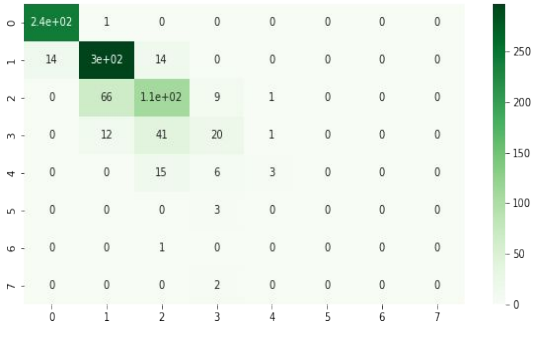
- Grid Search Algorithm
- Random forest tree
- Xg Boost Algorithm

MODEL BUILDING AND EVALUATION

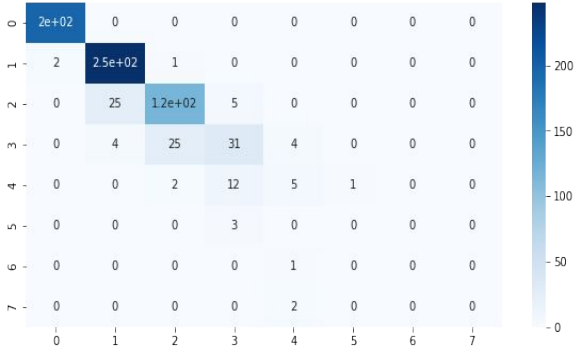
- For the grid search Algorithm, an accuracy score of 93% was realised
- For the grid search Algorithm, an accuracy score of 78% was realised
- For the grid search Algorithm, an accuracy score of 87% was realised



Grid search algorithm



Random forest tree algorithm



Xg boost classification

CONCLUSION

In conclusion, the analysis of the persistency of a drug using data science techniques has provided valuable insights into the factors that influence patients' adherence to a prescribed medication regimen. Through the analysis of patient demographic data, prescription history, and health outcomes, we were able to identify significant predictors of persistency, such as age, comorbidities, and medication type. We developed three models using the grid search algorithm, the forest tree algorithm and the Xg boost algorithm of which the grid search performed the best at 93%

THANK YOU!