Data Glacier

Your Deep Learning Partner

Development of a Prediction Model

Healthcare: Persistency of a Drug

Date: August 30, 2022
Batch Code: LISUM 10
Team Healthy Bones - Jeeyeon Shon (shon.jeeyeon@gmail.com)
Company: Data Glacier
Country: the United States
Specialization: Data Science
Submitted to: Data Glacier
Link to the Repo: https://github.com/shonjeeyeon/DG_Week_13

Data Glacier
Your Deep Learning Partner

Agenda

- Executive Summary
- Problem Description
- Data Cleansing/Preprocessing
- Model Development
- Logistic Regression: Confusion Matrix
- Logistic Regression: Learning Curves
- Application Development: Overview
- Application Development: Preview
- Links to the Repo/App
- Conclusion
- References

Executive Summary

Data Glacier
Your Deep Learning Partner

- Using a dataset of 3,424 records and 69 features, Exploratory Data Analysis (EDA) was performed to analyze persistency of a certain medication.
- The Logistic Regression model was able to classify persistent vs. non-persistent patients with AUC of 0.8049 and accuracy of 0.8189.
- An app was developed using the model above.
  - Link: https://persistency.herokuapp.com/

Problem Description

- Medication persistence refers to completing the medication treatment using the duration set by the prescriber. (Cramer et al., 2008)
- Therefore, persistence is important in patients' positive outcomes as well as in pharmaceutical industries' profits.
- Developing a model to automate prediction process will contribute to save time and cost spent by the company, and the prediction results can be used for marketing, patient education, or R&D purposes.

Data Cleansing/Preprocessing

Data Glacier
Your Deep Learning Partner

- Imputation of Missing Values
- Outliers and skews in numeric columns were reduced
- Each patient’s total number of risks, comorbidities, and concomitant therapies were calculated.
- Using Recurrent Feature Elimination (RFE), seven most important features were selected for model development.
- SMOTE was used to oversample to address balance issues.

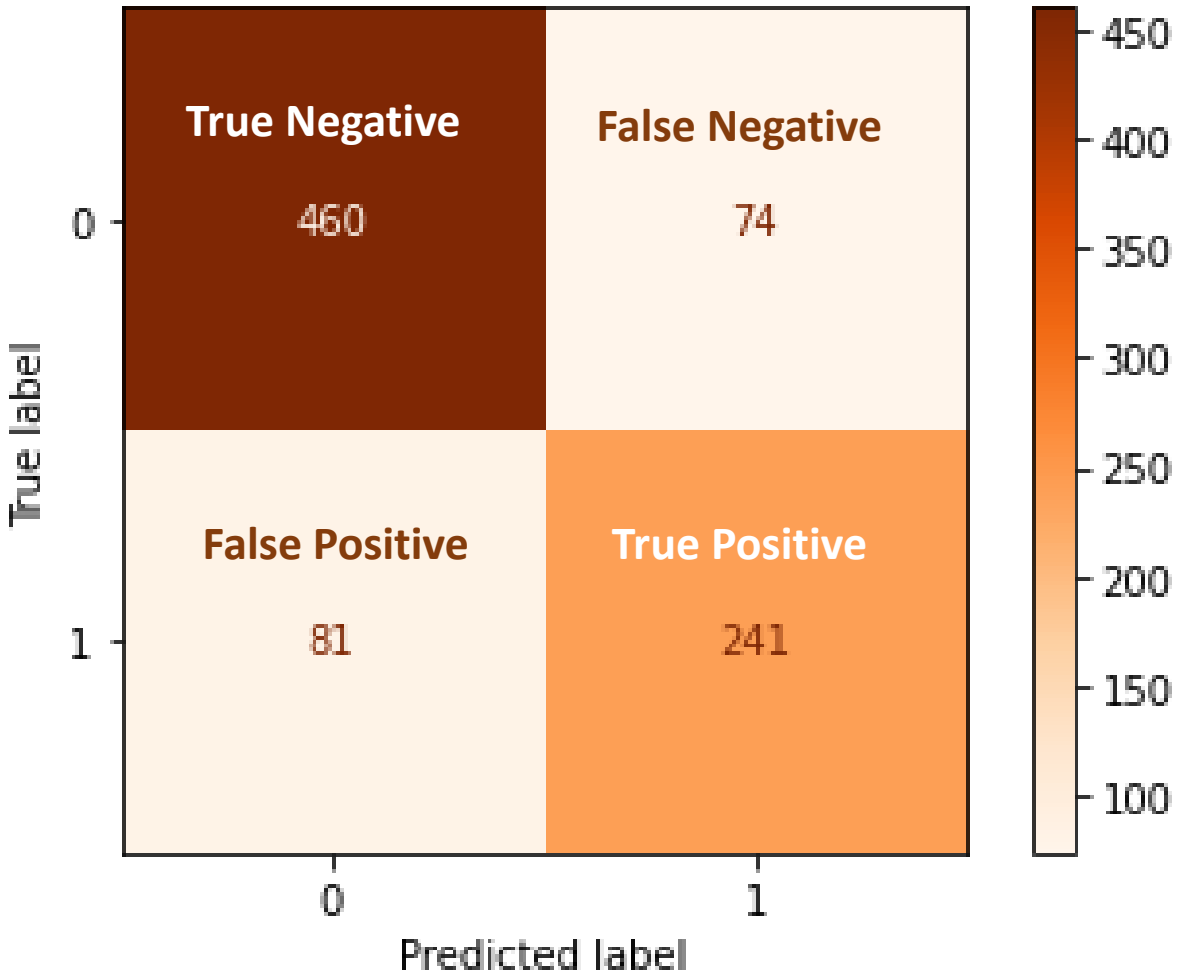
Model Development

Data Glacier
Your Deep Learning Partner

Model	AUC	Accuracy	Recall
Logistic Regression	**0.8049**	**0.8189**	**0.7484**
Random Forest	0.7667	0.7850	0.6925
XGBoost	0.7877	0.7944	0.7609
Multi-Layer Perceptron	0.7959	0.8061	0.7547

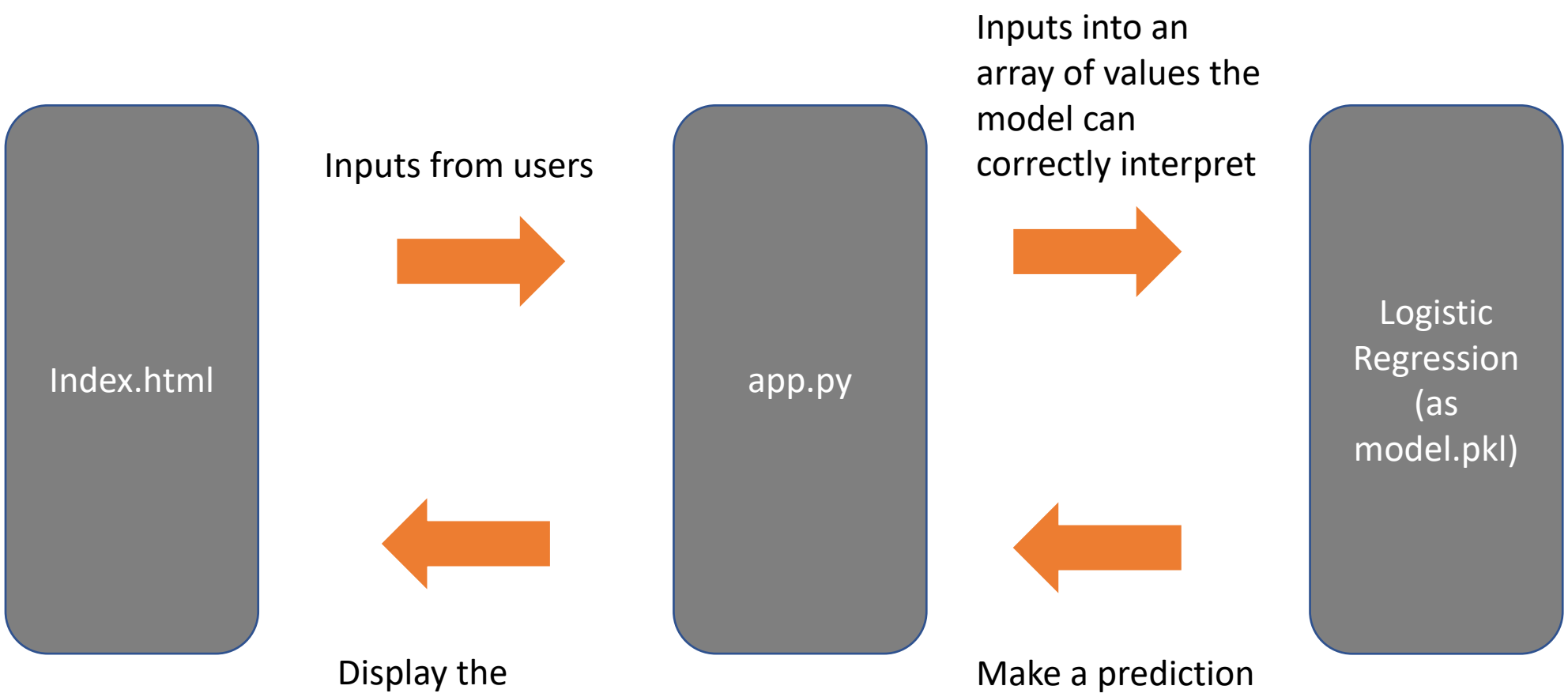
- Four model were tested using optimized parameters via GridSearchCV.
- Logistic Regression model had the highest AUC and accuracy on the test set.
  - AUC=0.8049
  - Accuracy=0.8189

Logistic Regression Model: Confusion Matrix

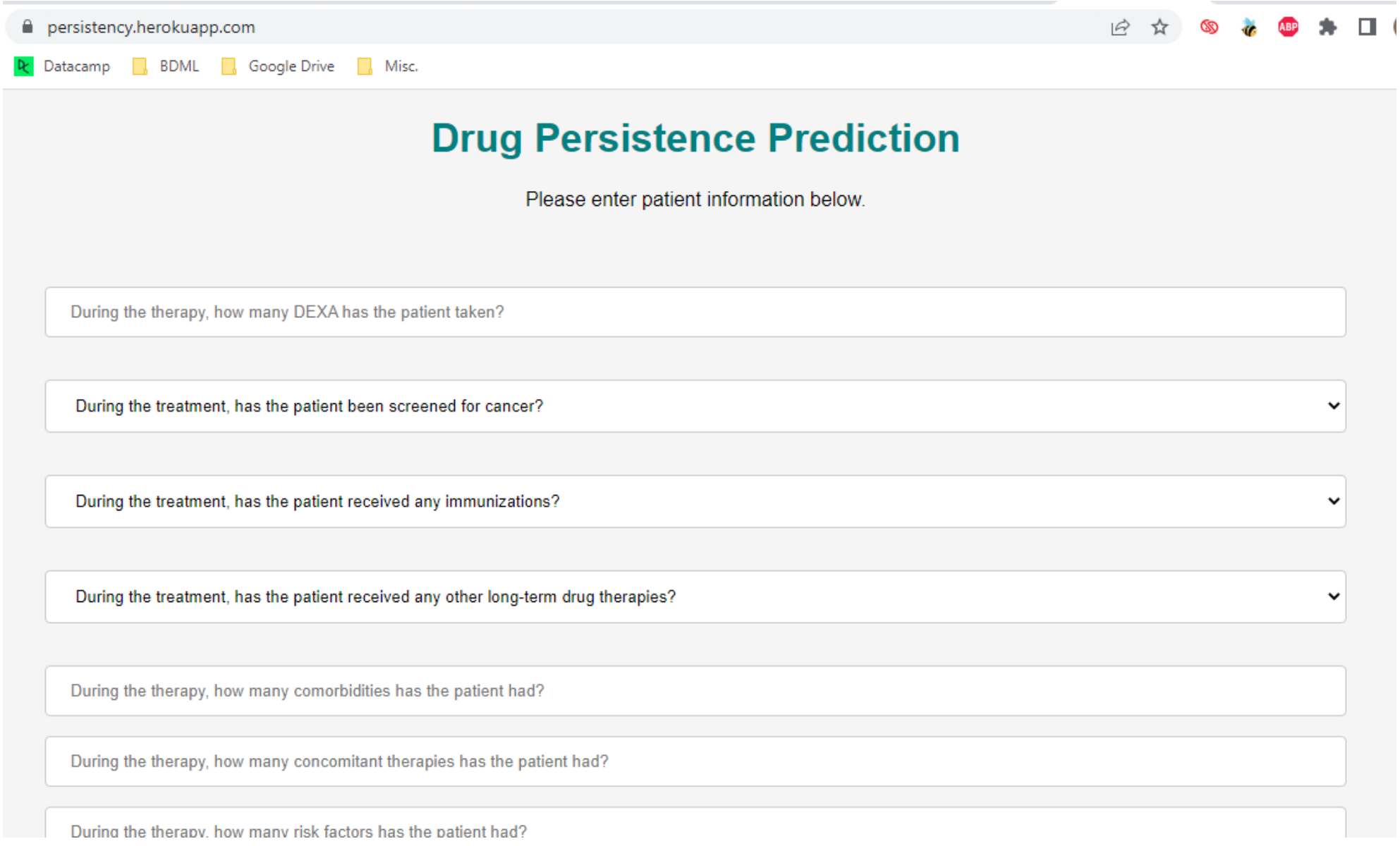


- True Positive Rate (TPR) = 0.7484
- True Negative Rate (TNR) = 0.8337

Application Development: Overview



Application Development: Preview



Links to the Repo/App

- Link to the repo: https://github.com/shonjeeyeon/DG_Week_13
- Link to the app: https://persistency.herokuapp.com/

Conclusion

Data Glacier
Your Deep Learning Partner

- The Logistic Regression model was able to predict a patient's persistency of a drug. Seven select features were used for prediction.
- The AUC of the model was approximately 80% and accuracy was 82%.

Reference

- Cramer, J.A., Roy, A., Burrell, A., Fairchild, C. J., Fuldeore, M.J., Ollendorf, D.A., Wong, P.K. (2008). Medication compliance and persistence: terminology and definitions. Value Health. 11(1), 44-47

Thank You