Project 1: Heart Risk Prediction Model

Manuel Duran

DSC 680

Bellevue University

- **Overview**
    - The leading cause of death within the United States of America is heart disease\failure (Stats of States). This fatal disease takes a life of an individual every 36 seconds in the USA today (CDC.gov). Heart Failure is generated from various conditions and activities that are performed during an individual's day to day lifestyle. Some conditions include Diabetes, High Blood Pressure, and Anemia, while other lifestyle conditions that help prevent Heart Failure is lack of movement and smoking. Understanding the various issues that occur witinh heart failure, as a user we would like to create a way to predict if a patient is at risk of dying. Utilizing a dataset from Kaggle, I created a model using Logistic Regression to attempt to try a predict if a patient will be a risk within the near future.
- **Data Understanding**
    - The dataset that acquired from Kaggle.com. It was an all-numerical dataset that features various information of patients. This includes their age, sex, conditions of the patient, and whether the patient died from heart failure or is still alive

- **Data Preparation**
    - Data preparation for this project included searching for null values. The dataset fortunately did not contain any Null values, and there was no need for imputation for the project. Additionally, there was no need to adjust the data type of the dataset since all the columns featured numerical data easier use for analysis.
- **EDA**
    - After performing my data prep, I focused on exploring the dataset the best I could within python. I generated a distribution of age group (see Appendix A). Reviewing the distribution, I could see that dataset was more focused on individuals forty years old and above. Additionally, majority of the patients were between the ages of fifty and sixty years old. To attempt to gain more insight on my dataset, I created heatmap (See Appendix B). Lastly, using Pandas I created crosstab bar charts to compare death events to various activities\conditions of the patients within the dataset (See Appendix C). By performing this I discovered that women have a slightly higher mortality rate than men when it comes to heart failure. Lastly, comparing diabetes appears to have a higher effect on death events than high blood pressure according to my crosstab charts.

- **Method\Model Results**
    - The method selection for model selection was logistic regression, with an additional usage of SVC (Support Vector Classifier). From the dataset I utilized the different variables to provide my model the best inputs to allow use to find the best precision for death event risk possible. After careful trial and error, I was discovered that Age, Ejection Fraction (Heart Failure Measurement), Serum Creatine Level, Serum Sodium Level, and Time (Follow Up Days from Patient Check Up) were the best possible variables for my model. After, the variable selection then proceeded to create my training and testing sets from the dataset itself. After completing the task of creating my training and testing sets, I then moved forward to creating the model with the sklearn package within my notebook. The results produced an accuracy of 90% to predict risk of heart failure which could occur into a death event. Additionally, my mode produced a Recall and Precision Score of 78%.
- **Conclusion**
    - Overall, the model performance appears to be very well at 90%. I considered that it maybe too high; however, the precision and recall scores are very high so I believed I selected the correct variables during the feature selection. Additionally, it is very intriguing from the list of variables, the best heart failure prediction variables were specific scientific measurements, not daily activities or previous diseases discovered with an individual. A reasoning for this is because

most of the variables were specified scientific measurements that were catered to each patient. This could have possibly helped increase the model precision when testing the accuracy. In conclusion, the model seems to be functional to a point; however, there is room for growth, such as specifying gender risk of heart failure.

References

1. **Heart Disease Facts | cdc.gov. (2020, September 8). Centers for Disease Control and Prevention. https://www.cdc.gov/heartdisease/facts.htm**

2. **Stats of the States - Heart Disease Mortality. (2021). Cdc.Gov. https://www.cdc.gov/nchs/pressroom/sosmap/heart_disease_mortality/heart_disease.htm**

3. **Heart Failure: Types, Symptoms, Causes & Treatments. (2021). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/17069-heart-failure-understanding-heart-failure**

4. **Diabetes and Your Heart. (2021, May 7). Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html**

5. **Tidy, C. (2017, July 10). Hyperlipidaemia. Patient. https://patient.info/heart-health/high-cholesterol/hyperlipidaemia#:%7E:text=Hyperlipidaemia%20means%20a%20high%20level%20of%20cholesterol%20or%20triglycerides%20in%20your%20blood.**

6. **Felman, A. (2019, July 22). Everything you need to know about hypertension. MedicalNewsToday. https://www.medicalnewstoday.com/articles/150109**

7. **Heart failure cases soar globally. (n.d.). ScienceDaily. Retrieved September 6, 2021, from https://www.sciencedaily.com/releases/2021/02/210211195326.htm**

8. **Cover Story | One Size Does Not Fit All: The Role of Sex, Gender, Race and Ethnicity in Cardiovascular Medicine. (n.d.). American College of Cardiology. Retrieved September 6, 2021, from https://www.acc.org/latest-in-cardiology/articles/2018/10/14/12/42/cover-story-one-size-does-not-fit-all-sex-gender-race-and-ethnicity-in-cardiovascular-medicine**

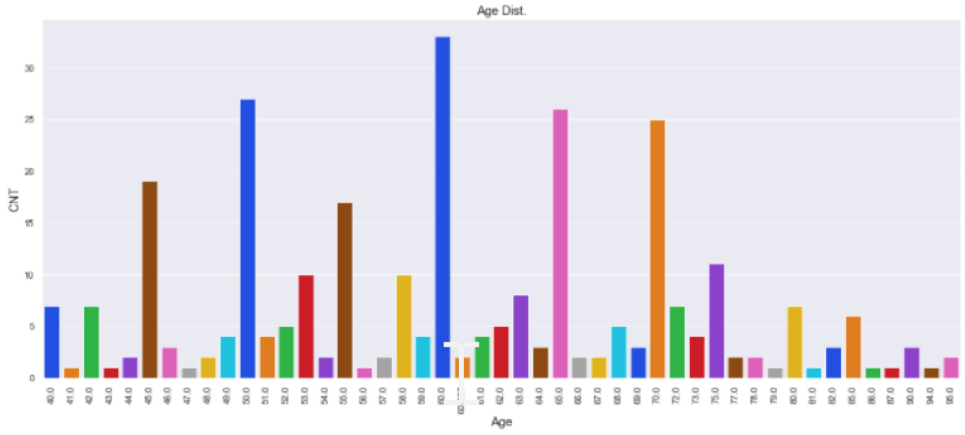
9. **Harvard Health. (2015, July 17). Race and ethnicity: Clues to your heart disease risk? https://www.health.harvard.edu/heart-health/race-and-ethnicity-clues-to-your-heart-disease-risk**

10. **The Heart (Human Anatomy): Picture, Definition, Location in the Body, and Heart Problems. (2009, June 30). WebMD. https://www.webmd.com/heart/picture-of-the-heart**

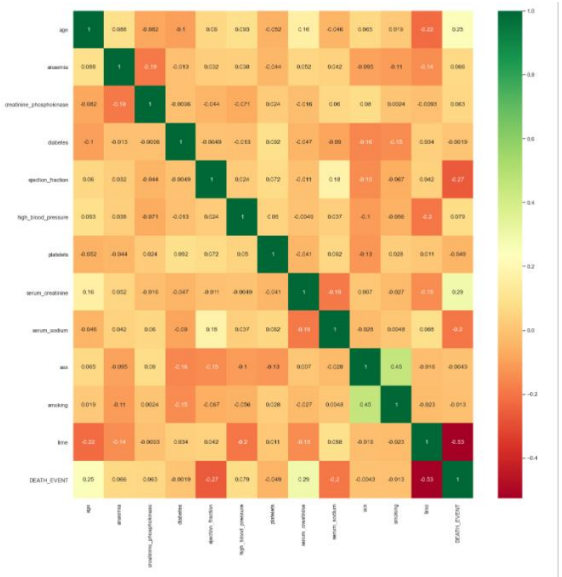
11. **https://www.kaggle.com/andrewmvd/heart-failure-clinical-data**

Appendices

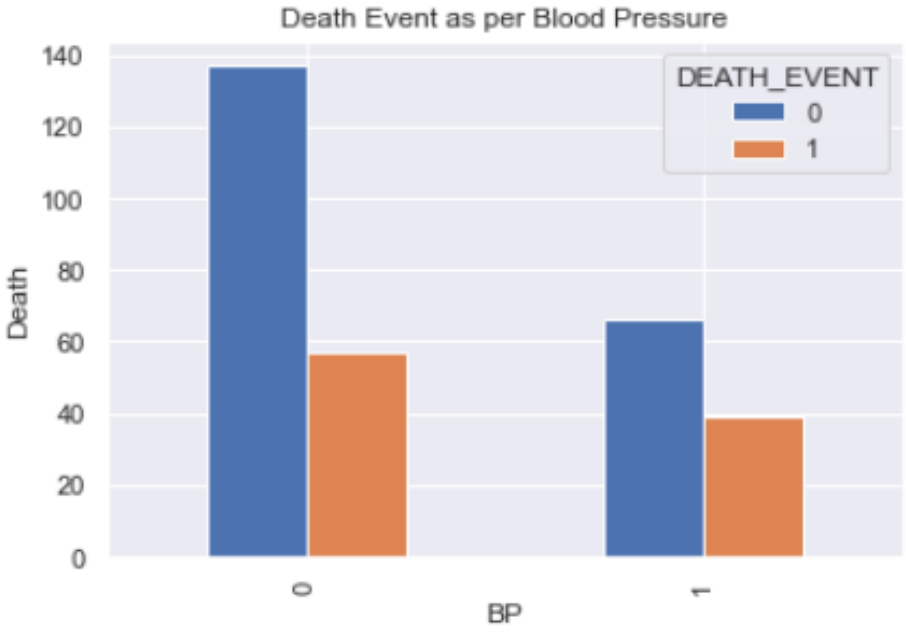
Appendix A



Appendix B



Appendix C


Death Event as per Blood Pressure
DEATH_EVENT
0
1
Death
140
120
100
80
60
40
20
0
0
1
BP