Myocardial Infarction Complications

SDS322E – Final Project

Kishen Narayan, Nick Read, Kathleen Leever, Diya Sharma,
Laura Nylykke, Yuvraj Uddanti, Sam Song

Problem Statement

Myocardial Infarction, more commonly known as a heart attack, is a severe condition known to plague many. MI may occur with or without a number of various complications.

Background Information

- A few of the most common complications linked to MI are sudden cardiac arrest, depression, anxiety, heart failure, arrhythmia, heart valve issues, ventricular septal defect, and free wall ruptures.

- Risk factors of MI caused by arterial plaque blockage include: obesity, smoking, hypertension, high cholesterol, diabetes, stress and anxiety, and a family history of heart disease.

- About 5% of all MI cases are caused by issues other than arterial plaque. Risk factors and causes for this type of MI are: coronary artery spasm, trauma to the coronary arteries, an embolism trapped in the coronary artery, electrolyte imbalance, eating disorders, and other rare medical conditions that may cause unusual narrowing of blood vessels (Damluji, A. A. *et al,* 2019.)

Motivations

MI is one of the most challenging problems of modern medicine.

- ➔ Acute myocardial infarction is associated with high mortality in the first year after it. The incidence of MI remains high in all countries.
  - ◆ This is especially true for the urban population of highly developed countries, which is exposed to
    - Chronic Stress Factors
    - Irregular and unbalanced nutrition.

In the United States, for example, more than a million people suffer from MI every year, and 200-300 thousand of them die from acute MI before arriving at the hospital.

Motivations (Continued)

Rise in Risk Factors

- A large cause of MI is arterial blockage, an outcome of cholesterol buildup due to obesity. With the rise in obesity in the recent years we found this dataset to hold interesting repercussions for a growing patient base.

Optimized Treatment

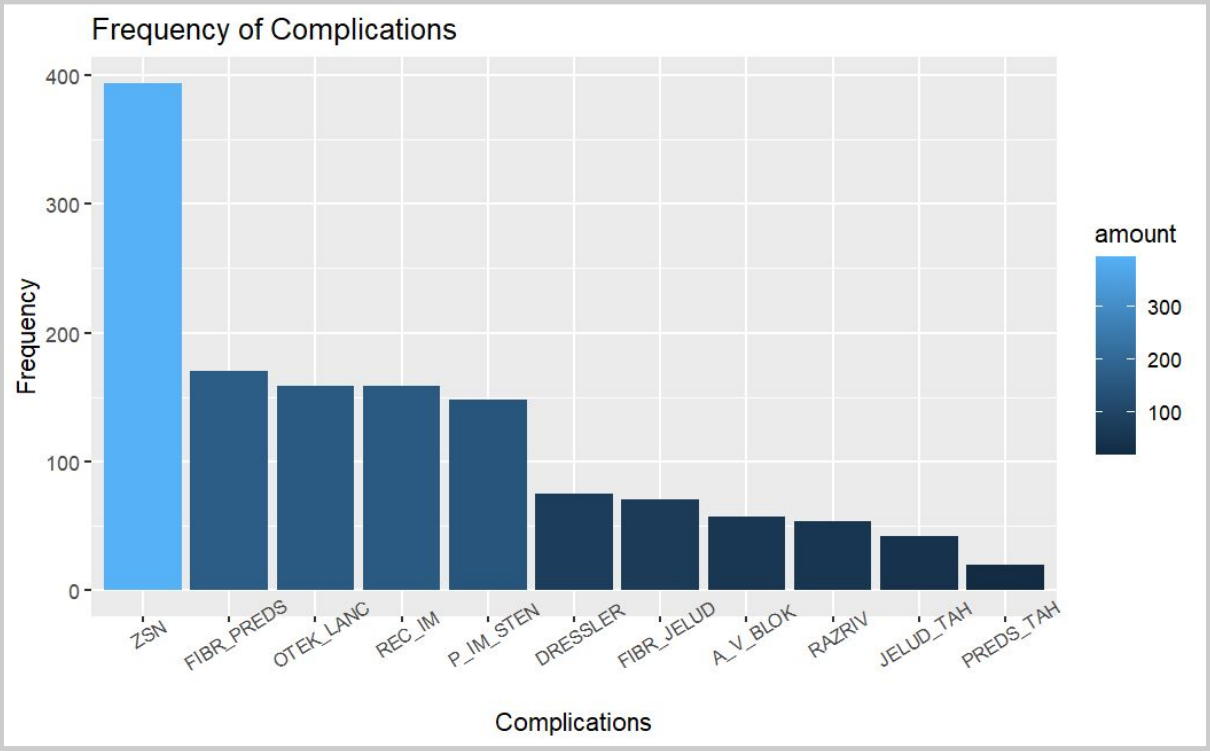
- The follow up treatment of patients who had MI, can be improved if professionals are able to predict possible complications before they occur based on certain observable attributes of the patient.
- Even an experienced specialist can not always foresee the development of these complications. In this regard, predicting complications of myocardial infarction in order to timely carry out the necessary preventive measures is an important task.

Data Cleaning

Data Cleaning was split into **3** rounds

- **Round 1**
  - Dropped all column comprised more than 50% of missing values.
- **Round 2 (Continuous Clean-Up)**
  - Replaced all the missing values in continuous columns with the means of respective columns.
- **Round 3 (Nominal & Ordinal Clean-Up)**
  - Replaces the nominal and ordinal variables columns with the mode of that respective

Exploratory Analysis: Complication Frequency



Exploratory Analysis

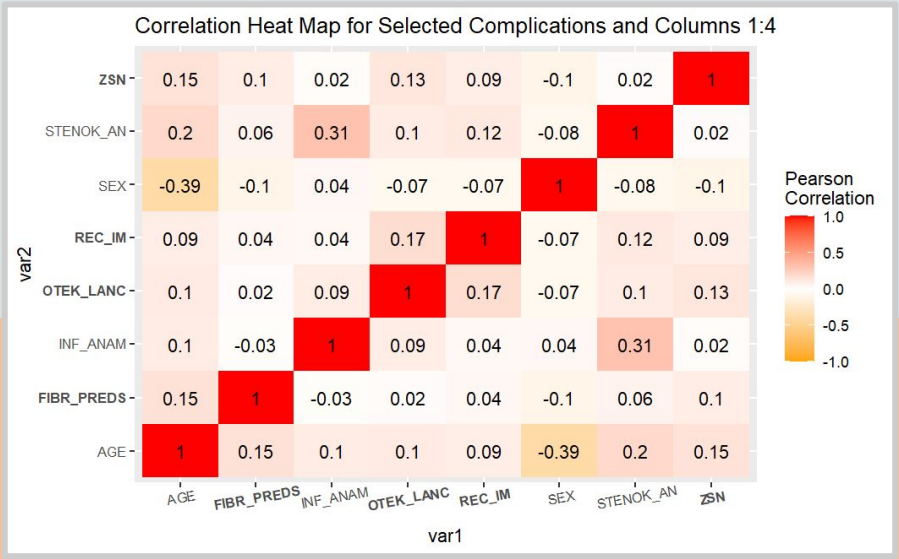
From the figure it is clear to see the four most prevalent complications by case count are –

- ZSN (Chronic Heart Failure) → 394 cases
  - 0.23176471 or 23.176471%
- FIBR_PREDS (Atrial Fibrillation) → 170 cases
  - 0.1 or 10%
- OTEK_LANC (Pulmonary Edema) → 159 cases
  - 0.09352941 or 9.352941%
- REC_IM (Relapse of the Myocardial Infarction) → 159 cases
  - 0.09352941 or 9.352941%

Our analysis specifically observed these complications of MI

Exploratory Analysis

- Correlation heat maps were made to observe correlation between input variables and the selected complications



Example Heat Map

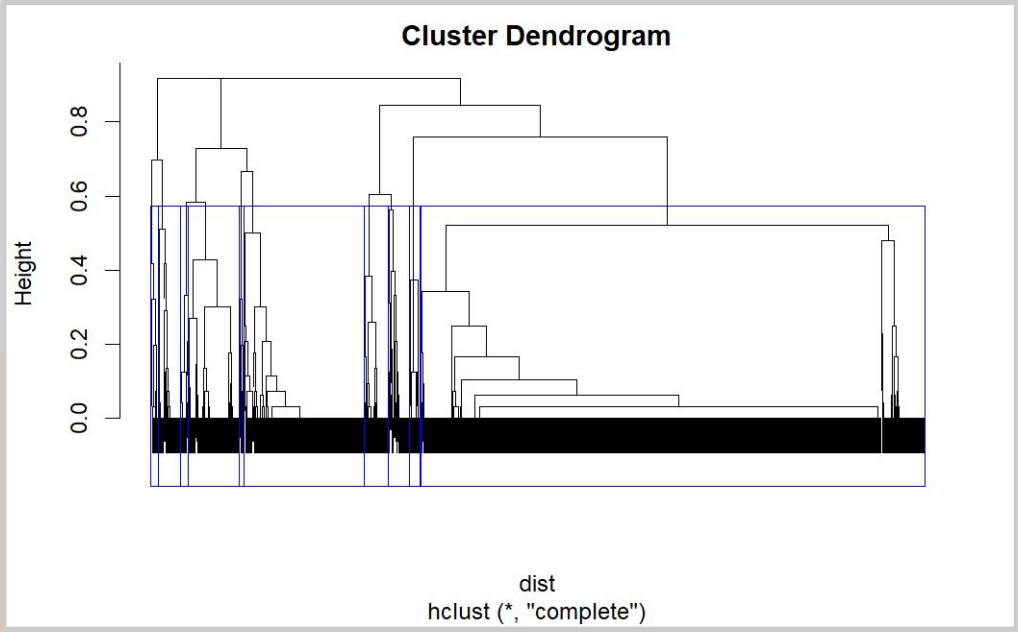
Exploratory Analysis

Highly Correlated Input Variable Information –

- **ZSN_A**
  - Presence of chronic Heart failure (HF) in the anamnesis
- **n_r_ecg_p_05**
  - Paroxysms of atrial fibrillation on ECG at the time of admission to hospital
- **NITR_S**
  - Use of liquid nitrates in the ICU
- **R_AB_3_n**
  - Relapse of the pain in the third day of the hospital period

Exploratory Analysis: Clustering

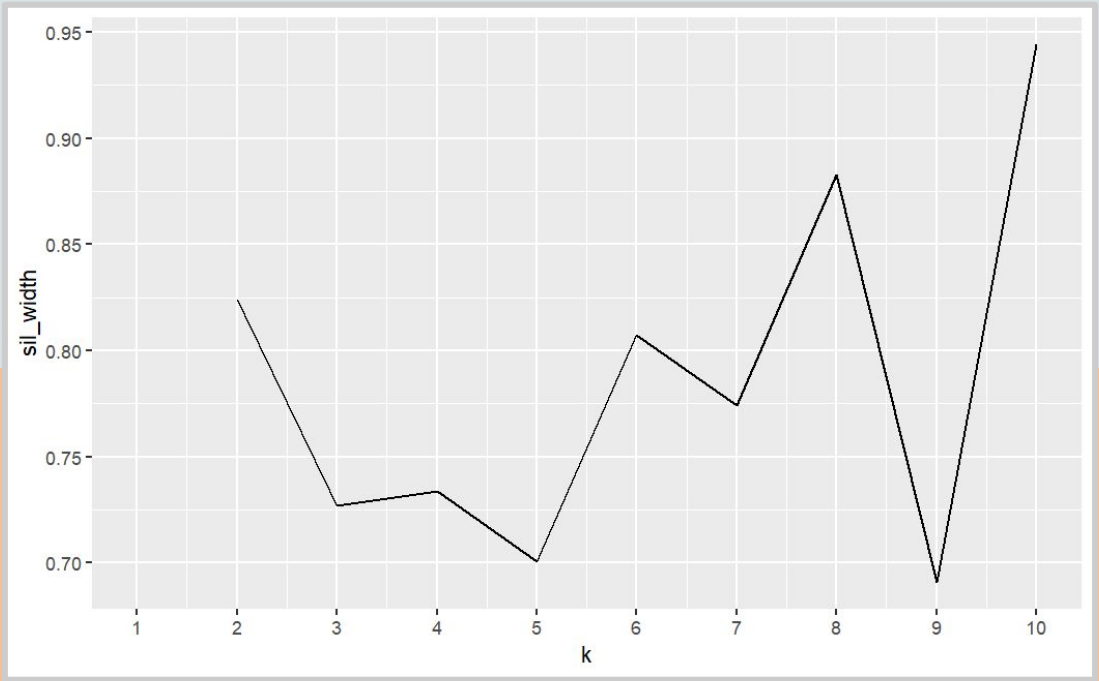
Hierarchical Clustering



Cluster Dendrogram

Exploratory Analysis: Clustering

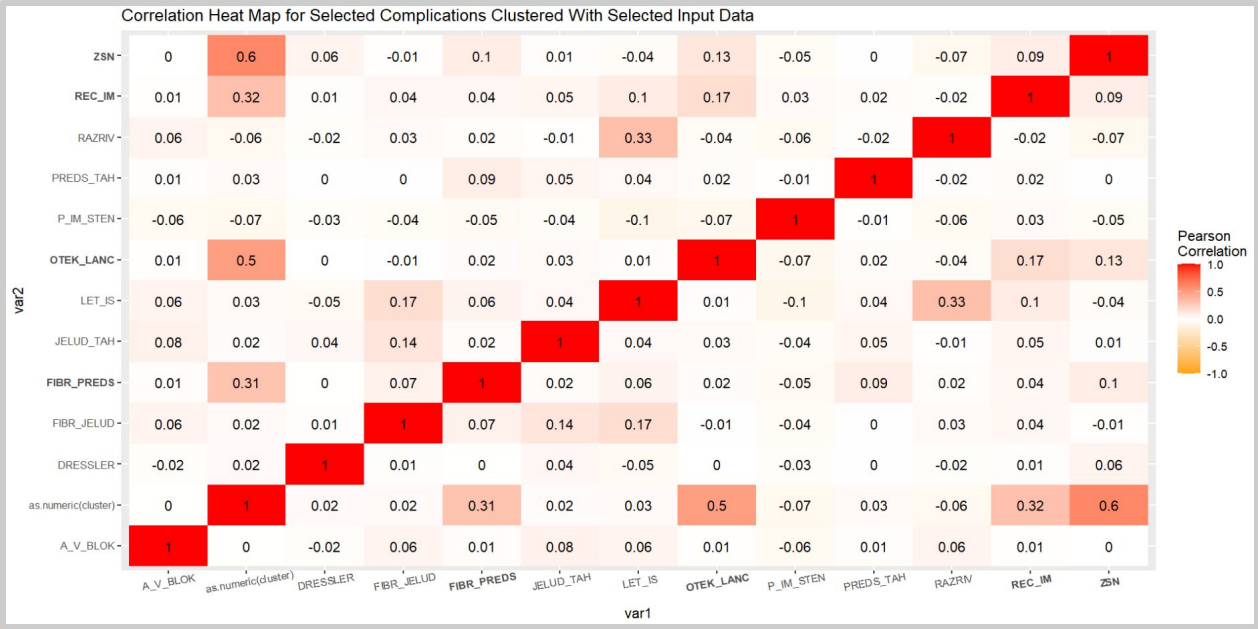
K-means clustering



Silhouette Plot

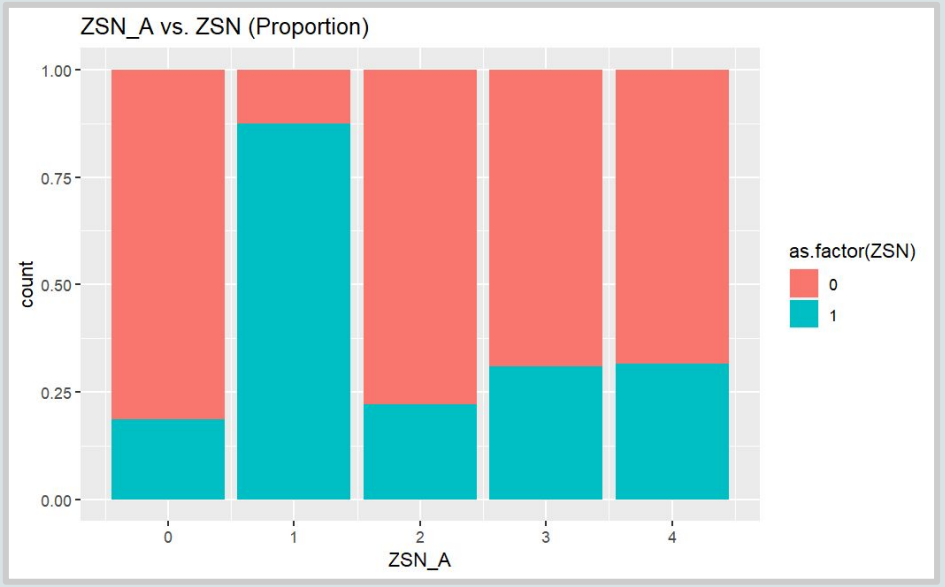
Exploratory Analysis: Clustering

Correlation Heat Map for Clustered Input Data and Selected Complications



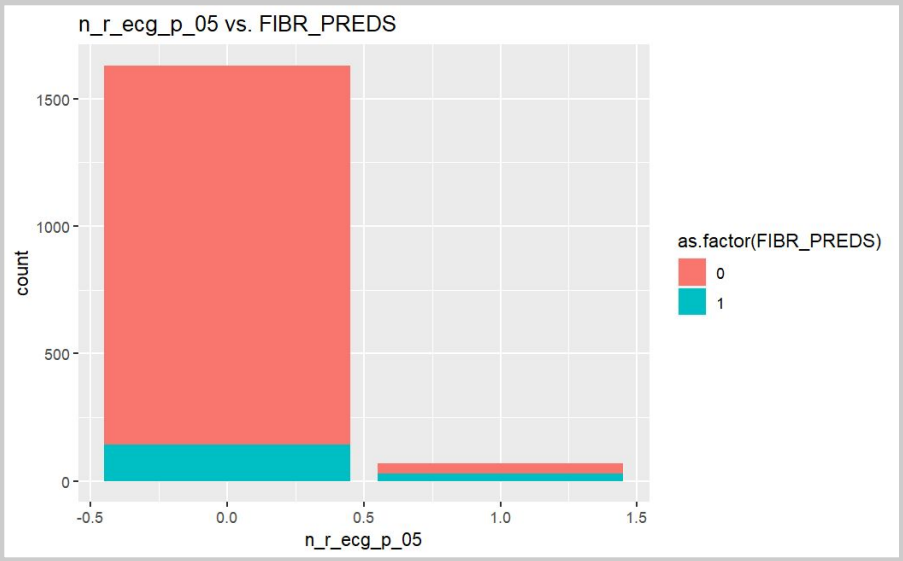
Correlation Heat Map

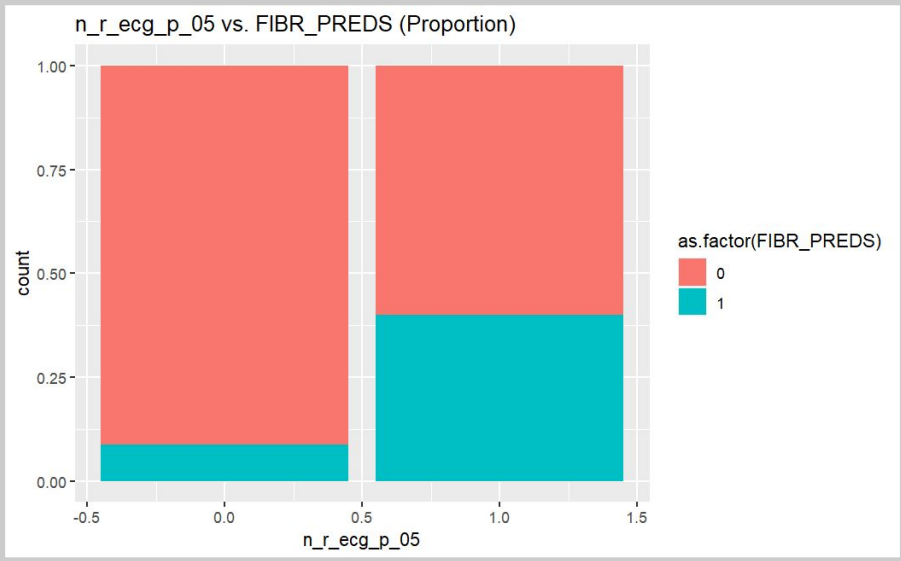
Exploratory Analysis: Hypotheses



Hypothesis 1 → If an individual has a severity level 1 of ZSN_A (Presence of chronic Heart failure (HF) in the anamnesis), then the person is more likely to develop ZSN (Chronic Heart Failure) as a complication of MI.

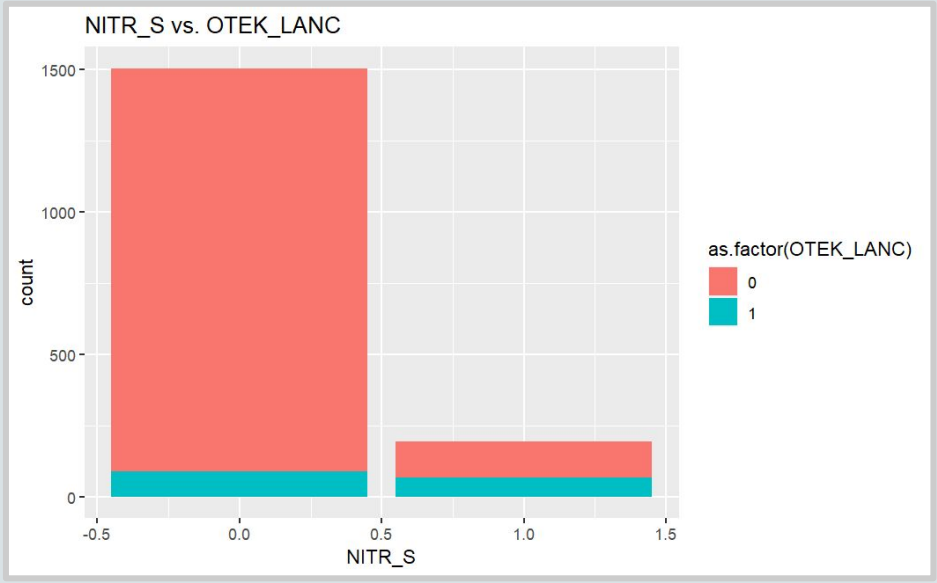
Exploratory Analysis: Hypotheses

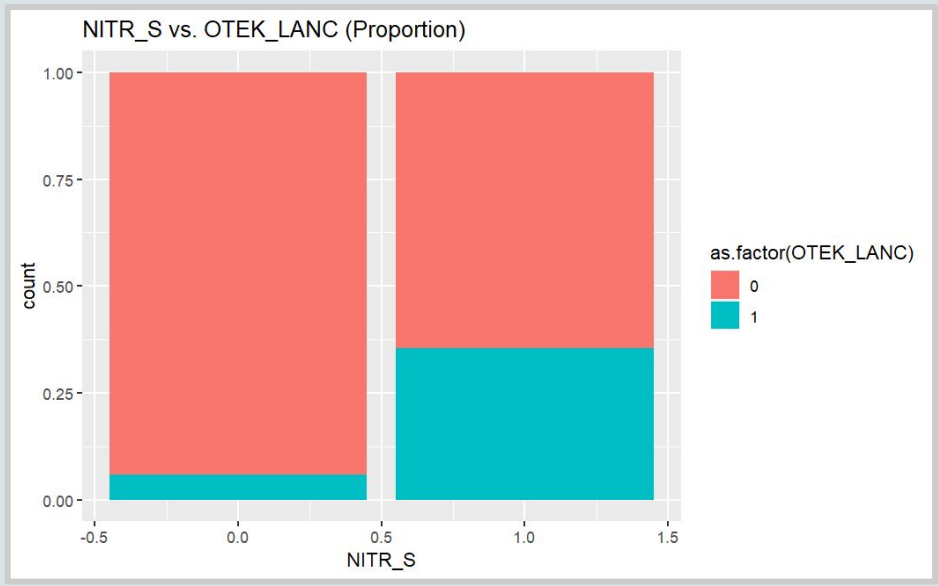




Hypothesis 2 → If a patient has had n_r_ecg_p_05 (paroxysms of atrial fibrillation on ECG at the time of admission to hospital), they are more likely to develop FIBR_PREDS (atrial fibrillation) as a complication of MI.

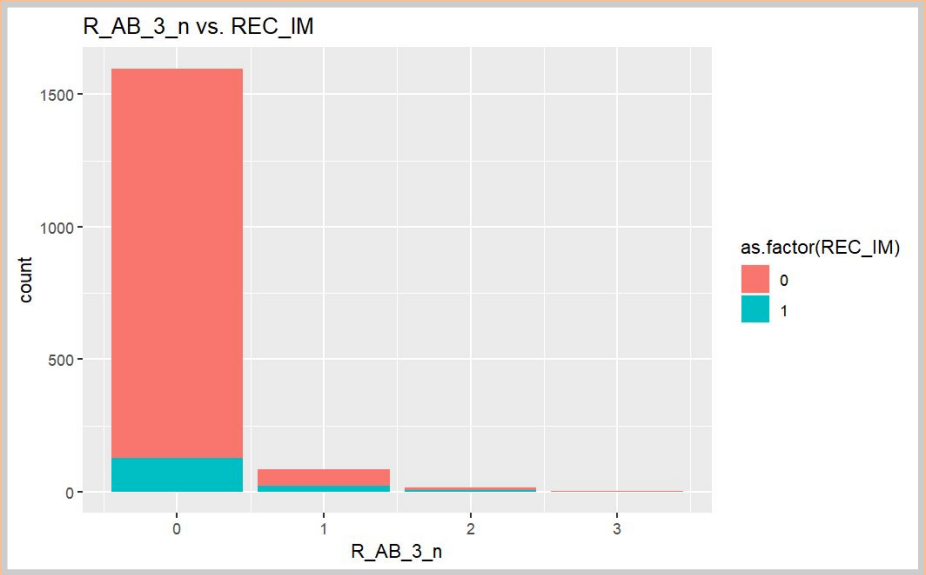
Exploratory Analysis: Hypotheses

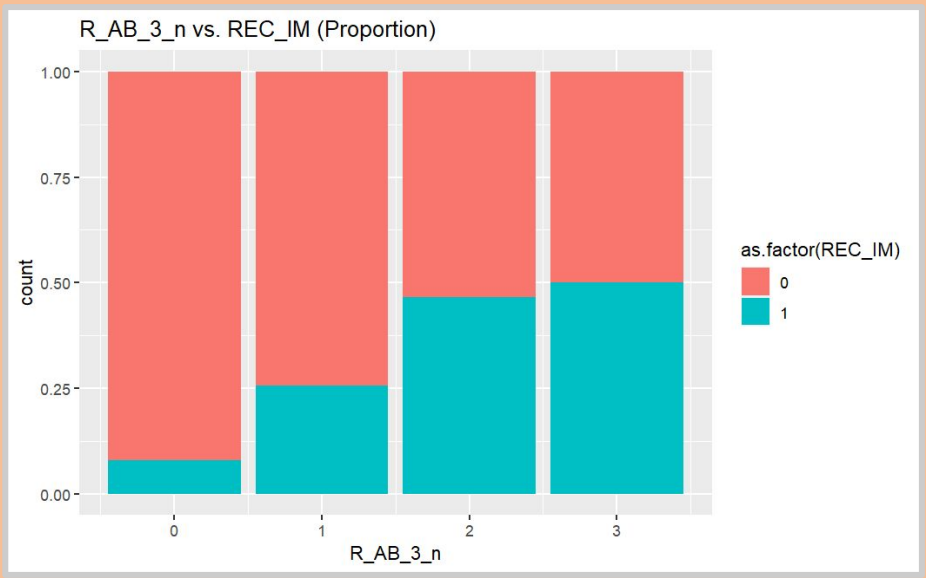




Hypothesis 3 → If a patient has had NITR_S (liquid nitrates used in the ICU) used for treatment, then there is a greater chance that the patient has OTEK_LANC (pulmonary edema) as a complication of MI.

Exploratory Analysis: Hypotheses





Hypothesis 4 → If the patient experiences R_AB_3_n (relapse of pain on the third day of the hospital period), then they are more likely to develop REC_IM (relapse of the myocardial infarction) as a complication of MI.

Classification Problems

Predictions of complications based on patient information on –

i. the time of admission to hospital:
- all input columns (2-112) except 93, 94, 95, 100, 101, 102, 103, 104, 105 can be used for prediction;

ii. the end of the first day (24 hours after admission to the hospital):
- all input columns (2- 112) except 94, 95, 101, 102, 104, 105 can be used for prediction;

iii. the end of the second day (48 hours after admission to the hospital):
- all input columns (2- 112) except 95, 102, 105 can be used for prediction;

iv. the end of the third day (72 hours after admission to the hospital):
- all input columns (2- 112) can be used for prediction

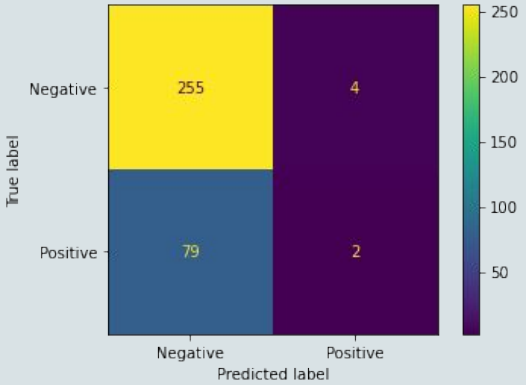
Classification Task & Results for Selected Complications

Predictions of each complication based on patient information for all of the present input variables

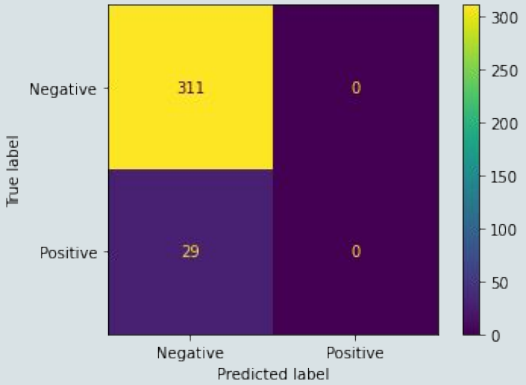
- A general K-Nearest Neighbors Classifier

Accuracy Scores for Selected Complications

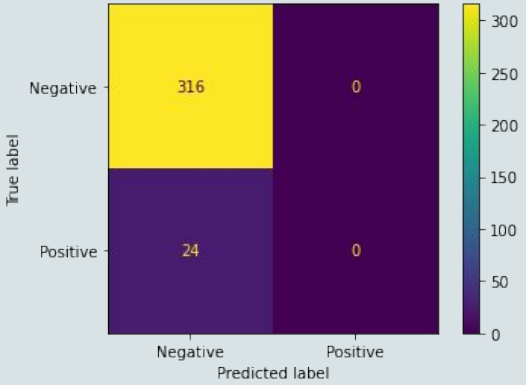
ZSN	0.7558823529411764
FIBR_PREDS	0.9147058823529411
OTEK_LANC	0.9294117647058824
REC_IM	0.9117647058823529



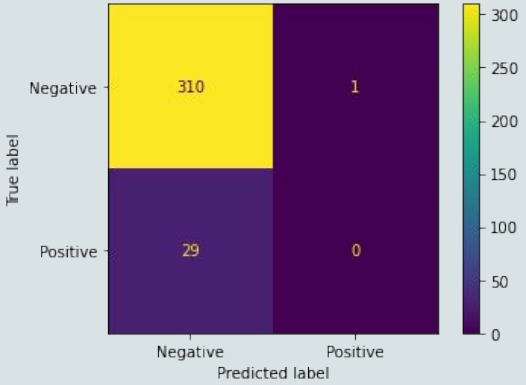
Chronic Heart Failure (ZSN)



Atrial Fibrillation (FIBR_PREDS)

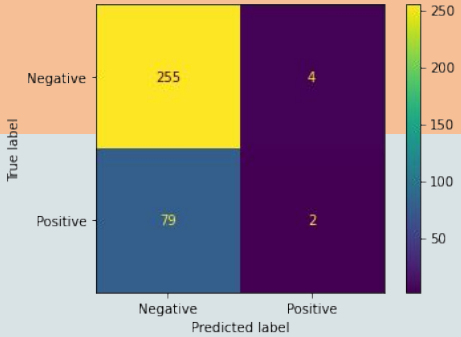


Pulmonary Edema (OTEK_LANC)



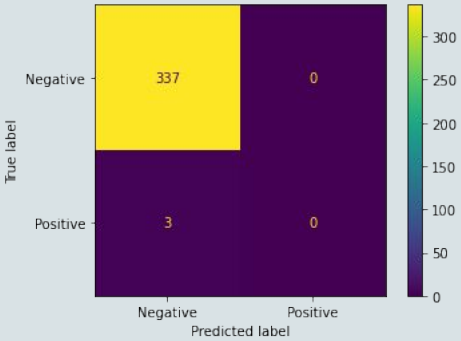
Relapse of MI (REC_IM)

Classification Task & Findings



Chronic Heart Failure *(ZSN)*

Complication with highest precision score (0.333)



Supraventricular Tachycardia *(PREDS_TAH)*

Complication with highest accuracy score (0.991)

Our Classifier:

- ➔ We used a K-Nearest Neighbors Classifier
- ➔ Used Cross-Fold Validation to determine best value of K (ended up being K = 8)
- ➔ Classifier is very good at predicting negative cases, but not very efficient at predicting positive cases (low overall recall)

Limitations & Discussion

Data Set & Cleaning

- There was a large amount amount of input columns (124) so it was difficult to fully understand the contribution of specific columns to the presence of a specific complication as a result of MI.
- Furthermore, the data was a mixed with continuous, nominal, and ordinal data, which made it difficult to properly clean and perform cluster analyses on the data
    - Loss of information after Round 1
    - Different imputations for Round 2 and 3 based upon continuous vs. nominal & ordinal data

Modeling

- Only used K-Nearest Neighbor Classifier
    - Other possible classifiers that could have been used as a way to improve accuracy and precision
        - Naive Bayes
        - Decision Tree
        - Neural Networks
- The data had many different types of input variables
    - In the future, SHAP could be utilized to evaluate the importance & contribution of the specific feature variables or combinations of the specific feature variables

Conclusions, Findings, & Discussion

The **precision score** tells us how good the chosen model was at predicting a specific outcome. The **accuracy score** tells us how many times the model itself was correct.

- ➔ The complication associated with MI with the highest precision score was (ZSN) chronic heart failure. This is in line with the aforementioned most common complications following diagnosis of MI.
- ➔ The complication associated with MI with the highest accuracy score was (PREDS_TAH) supraventricular tachycardia, which is also consistent with what is seen across all data with respect to MI outcomes / complications.

Thank You

References / works cited

Golovenkin, S. E., Shulman, V. A., Rossiev, D. A., Shesternya, P. A., Nikulina, S. Y., Orlova, Y. V., Krasnoyarsk, V. F. V.-Y., Gorban, A. N., & Mirkes, E. M. (2020, December 9). Myocardial infarction complications Data Set. Irvine. Retrieved from https://archive.ics.uci.edu/ml/datasets/Myocardial+infarction+complications.

Golovenkin, S. E., Bac, J., Chervov, A., Mirkes, E. M., Orlova, Y. V., Barillot, E., Gorban, A. N., & Zinovyev, A. (2020). Trajectories, bifurcations, and pseudo-time in large clinical datasets: Applications to myocardial infarction and diabetes data. GigaScience, 9(11). https://doi.org/10.1093/gigascience/giaa128

Damluji, A. A., van Diepen, S., Katz, J. N., Menon, V., Tamis-Holland, J. E., Bakitas, M., Cohen, M. G., Balsam, L. B., & Chikwe, J. (2021). Mechanical complications of acute myocardial infarction: A scientific statement from the American Heart Association. Circulation, 144(2). https://doi.org/10.1161/cir.0000000000000985