Beyond the Pulse: A Healthcare-based Application of Machine Learning

Ayush Gupta, Emily Armstrong & Sarika Gandigude

ADMN5016-F23-101: Applied Artificial Intelligence and Machine Learning

St. Lawrence College, School of Business

Professor: Sujoy Paul

Dec: 15th, 2023

Table of Contents

Delving into the Heart Failure Prediction Model: A Comprehensive Analysis 3

Executive Summary 3

Dataset 3

Assessing the Problem and Market Size 4

Benchmarking the Model's Performance 5

1. Accuracy: 5
2. Precision: 5
3. Recall: 5
4. F1 Score: 5

Comparing Classifiers: Random Forest vs. K-Nearest Neighbors 6

Quantifying the Economic Impact 7

1. Prevention of Complications: 7
2. Treatment Costs: 7

Addressing Potential Risks: False Positives and False Negatives 8

1. False Positives: 8
2. False Negatives: 8

Beyond Monetary Benefits: Enhancing Patient Outcomes and Healthcare Efficiency 9

1. Early Intervention 9
2. Resource Optimization: 9
3. Quality of Life: 9

Navigating Ethical and Regulatory Considerations for Responsible Implementation 10

1. Ethical Considerations: 10
2. Model Generalization 10
3. Regulatory Compliance: 10

Harnessing the Potential of the Heart Disease Prediction Model 11

Final Thoughts 12

References 13

Delving into the Heart Disease Prediction Model: A Comprehensive Analysis

Executive Summary

In 2020, nearly 68,000 Canadians died from heart disease (Statistics Canada, 2022). This comprehensive report delves into the intricacies of building heart disease prediction models, encompassing technical performance, potential impact, and associated risks. The research highlights the effectiveness of machine learning in identifying individuals at risk of heart disease, offering a valuable tool for early intervention and improved patient outcomes. However, it also emphasizes the need for careful consideration of ethical concerns and model generalization to ensure the responsible and equitable implementation of this technology.

Dataset

The dataset is small, consisting of just over 900 data points. As such, it would be challenging to make inferences about a larger population based on our results. However, we believe that the insights gained from this analysis provide a strong foundation for future research. The results of the Principal Component Analysis highlight the complexity of heart disease and left us asking questions about how specific variables are or are not associated with one another.

Given more time, more domain knowledge, and more powerful computational capabilities, we believe our model stands as an excellent blueprint for future model development.

Assessing the Problem and Market Size

Heart disease's pervasiveness and its significant impact on individuals' lives necessitate the development of effective preventive healthcare strategies. The ability to accurately predict an individual's risk of developing heart disease holds immense promise for early intervention and improved patient outcomes.

The extensive dataset used in this study, encompassing 11 critical risk factors, provides a robust foundation for developing an effective predictive model. These risk factors, including age, sex, cholesterol levels, fasting blood sugar levels, resting blood pressure, and type of chest pain, have been extensively studied and are known to significantly influence the likelihood of developing heart disease.

The market for heart disease prediction models is vast, catering to individuals with prevalent cardiovascular risk factors such as hypertension, diabetes, and hyperlipidemia. The vast nature of these risk factors extends the model's applicability across diverse populations and regions. By providing personalized and proactive healthcare solutions, the model contributes to improved overall well-being and societal health.

Benchmarking the Model's Performance

The evaluation of the heart disease prediction model employs a comprehensive set of metrics to assess its performance:

1. Accuracy: The overall correctness of the model's predictions, which measures the proportion of correct predictions made. A high accuracy score indicates that the model accurately identifies individuals at risk of heart failure.

2. Precision: The proportion of positive predictions that are actually correct. A high precision score ensures that the model identifies high-risk individuals without generating excessive false alarms, which can lead to unnecessary medical interventions and patient distress.

3. Recall: The proportion of true positive cases that are correctly identified. A high recall score highlights the model's ability to detect individuals at risk effectively, minimizing the risk of missing critical cases.

4. F1 Score: A balanced measure of precision and recall, providing a holistic assessment of the model's performance. An F1 score closer to 1 indicates a well-balanced model that excels in both precision and recall.

Comparing Classifiers: Random Forest vs. K-Nearest Neighbors

Two machine learning classifiers, Random Forest and K-Nearest Neighbors (KNN) were evaluated for their performance in predicting heart disease risk. Ultimately, upon hyperparameter tuning, both models performed satisfactorily in predicting instances of heart disease.

While the KNN model had an accuracy of 89.5% after hyperparameter tuning (using GridSearchCV), KNN can be prone to overfitting and is more sensitive to outliers and noise in data. With an initial accuracy of 85%, precision of 86%, recall of 90%, and an F1 Score of 85%, the Random Forest classifier exhibited a well-rounded performance, making it a more suitable choice for heart failure prediction. RandomizedSearchCV was used for hyperparameter tuning, after which the Random Forest Classifier model produced an accuracy of 88%. Principle Component Analysis was done for the KNN model and greatly decreased the performance of the model, indicating that more variables are necessary for optimal KNN model performance.

The prediction of heart disease risk is a challenging problem, as it involves learning complex relationships between a variety of factors, such as medical history, demographics, and lifestyle choices. One of the advantages of the Random Forest classifier is that it is able to handle high-dimensional data with many features. This is important for the heart disease prediction problem, as there are many factors that can influence heart failure risk. Another advantage of the Random Forest classifier is that it is relatively insensitive to noise and outliers in the data. This is important for the heart disease prediction problem, as medical data can often be noisy and incomplete.

Overall, the Random Forest classifier is a well-suited choice for predicting heart disease risk. It is a robust and efficient algorithm that can handle high-dimensional data with noise and outliers. The combined computational time for the KNN and Random Forest Classifier models, the Principal Component Analysis, and hyperparameter tunings is approximately ten minutes.

Quantifying the Economic Impact

The potential economic benefits of the heart disease prediction model are significant, encompassing both direct and indirect healthcare cost savings:

1. Prevention of Complications: Early detection and intervention can avert costly hospitalizations, surgeries, and long-term treatments associated with advanced heart disease stages. These interventions can reduce the need for expensive medical procedures and prolong patient lives, leading to significant cost savings.

2. Treatment Costs: Proactive interventions, preventive medications, and lifestyle modifications can reduce both direct and indirect medical expenses. By addressing risk factors early on, the model can prevent the progression of heart disease, reducing the need for costly treatments later in the illness' course.

Addressing Potential Risks: False Positives and False Negatives

While the heart disease prediction model holds immense promise, it is crucial to acknowledge and address potential risks associated with its use:

1. False Positives: Unnecessary treatments stemming from false positives can lead to unwarranted medical expenses, potential side effects, and psychological distress for individuals. Overdiagnosis can lead to unnecessary anxiety, disruptions in daily life, and potential harm from unnecessary medications.

2. False Negatives: Missed opportunities for early intervention due to false negatives can result in delayed diagnosis, disease progression, and escalation of treatment costs. Failure to identify high-risk individuals early can delay appropriate treatment and exacerbate the severity of heart failure, leading to higher healthcare costs and potentially worsening patient outcomes.

Managing these risks is crucial to ensure the responsible implementation of the heart disease prediction model.

Beyond Monetary Benefits: Enhancing Patient Outcomes and Healthcare Efficiency

The heart disease prediction model offers significant benefits beyond its economic implications:

1. Early Intervention: Timely detection enables proactive interventions, potentially reducing the severity of heart failure and improving long-term patient outcomes. By identifying individuals at risk early, the model can facilitate lifestyle changes for a longer, healthier life.

2. Resource Optimization: Efficient allocation of medical resources to high-risk individuals optimizes healthcare resource utilization, reducing strain on healthcare systems. By focusing preventive efforts on those most likely to benefit, the model can help healthcare providers allocate resources more effectively, reducing the burden on healthcare systems and enabling the prioritization of critical care needs.

3. Quality of Life: Preventive measures not only save costs but also enhance individuals' quality of life by promoting healthier lifestyles and reducing the burden of chronic conditions. By empowering individuals to take action to reduce their risk factors, the model can contribute to improved overall well-being and reduced reliance on healthcare services.

Navigating Ethical and Regulatory Considerations for Responsible Implementation

To fully realize the benefits of the heart disease prediction model, its implementation must be guided by ethical principles and adherence to regulatory frameworks:

1. Ethical Considerations: Protecting sensitive health data and adhering to ethical standards are paramount to maintain patient trust and compliance with regulations. The model should be developed and deployed with respect for patient privacy and informed consent, ensuring that individuals' data is handled responsibly and securely.

2. Model Generalization: It is essential to evaluate the model's ability to generalize across diverse populations and demographic groups to ensure its effectiveness in varied scenarios and prevent biases. The model should be tested on a representative sample of the population and its performance should be assessed across various subgroups to ensure its general applicability.

3. Regulatory Compliance: Strict adherence to healthcare regulations, including data protection laws and patient confidentiality, is crucial for maintaining trust and compliance with legal standards. The model's development, deployment, and use should align with relevant regulations and guidelines, ensuring that patient data is protected, and privacy is respected.

Harnessing the Potential of the Heart Disease Prediction Model

The heart failure prediction model holds immense potential to revolutionize heart disease prevention and contribute to a healthier Canadian population. By combining advanced machine learning techniques with a deep understanding of cardiovascular disease risk factors, the model can offer a powerful tool for early detection, improved patient outcomes, and reduced healthcare costs.

However, its successful implementation requires careful consideration of ethical, regulatory, and practical considerations. Ethical principles must be upheld to protect patient privacy and ensure informed consent. Model generalization must be thoroughly evaluated to ensure its effectiveness across diverse populations. Regulatory compliance must be strictly maintained to safeguard patient data and align with legal standards.

By addressing these challenges and embracing its potential, the heart failure prediction model can revolutionize heart disease prevention and pave the way for a healthier and more resilient Canada.

Final Thoughts

Heart disease is the second leading cause of death in Canada, and early detection and management of existing medical conditions can help reduce risk. Our application, while limited in scope in its current stage, has the potential to be adapted to larger datasets from which a more robust model could develop. We are acutely aware of the importance of high-performance models and, as such, would conduct further research and model testing prior to deploying our model.

We saw during the EDA that sex and chest pain type were strong predictors for the presence of heart disease. We would want to explore these variables further. Deeper analysis of respective variables would allow for a clearer idea of what indicators need to be looked for when predicting heart disease. The correlation matrix did not indicate any significant relationships - is this because of the variety within our dataset? How would we see this change when applied to a dataset exclusive to male, or female? Or age? This report has forged new avenues to explore the predictability of heart disease and empower Canadians with knowledge.

Link to GitHub page: https://github.com/eepag/ADMN-5016

References

Government of Canada, Statistics Canada. (2022, February 10). Heart and Stroke month. Statistics Canada. https://www.statcan.gc.ca/o1/en/plus/353-heart-and-stroke-month

Public Health Agency of Canada. (2022, July 28). Heart disease in Canada. Canada.ca. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/heart-disease-canada.html

Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data

University Hospital, Zurich, Switzerland: William Steinbrunn, M.D. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data

University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data

V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D. https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction/data