Multiple Disease Prediction Using Machine Learning

Kesava Sai Krishna Kondepudi

1. Problem Statement

The goal of this project is to develop an accurate and reliable predictive model that can assist in the early detection and diagnosis of two distinct diseases: Heart Disease and Parkinson's Disease. The predictive model will be built using machine learning techniques and will take into consideration relevant medical and demographic features of patients.

Heart Disease Prediction:
The model should be capable of analyzing various cardiovascular risk factors such as age, gender, blood pressure, cholesterol levels, blood sugar, chest pain types, fasting blood sugar, maximum heart rate, and many more features to predict the likelihood of a patient having heart disease. Early identification of at-risk individuals can lead to timely interventions and improve the chances of preventing severe cardiac events.

Parkinson's Disease Prediction:
For Parkinson's disease prediction, the model will utilize features such as Multidimensional Voice Program Frequency values, Jitter: Divergence of Difference Pair, Shimmer values, and other relevant motor and non-motor symptoms. Early detection of Parkinson's disease is crucial for better disease management and potentially slowing down its progression.

Project Objectives:

1. Data Collection: Gather a comprehensive dataset that includes medical records and relevant features of patients diagnosed with or without heart disease and Parkinson's disease.

2. Data Preprocessing: Clean, preprocess, and standardize the collected data to ensure consistency and reliability during model training.

3. Feature Selection: Identify the most relevant and informative features that significantly contribute to the prediction of heart disease and Parkinson's disease.

4. Model Development: Build separate machine learning models for heart disease and Parkinson's disease prediction, using appropriate algorithms such as logistic regression, and support vector machines.

5. Model Evaluation: Assess the performance of the predictive models using appropriate evaluation metrics.

6. User Interface: Develop an intuitive user interface that allows medical professionals to input patient data and obtain disease predictions in real time.

By successfully completing this project, I aim to contribute to early disease detection and improve patient outcomes by providing a reliable tool for predicting the likelihood of heart disease and Parkinson's disease in individuals.

2. Market/Customer/Business Need Assessment:

Market Need:

The market need for this project arises from the growing concern over the prevalence and impact of heart disease and Parkinson's disease on public health. Cardiovascular diseases, including heart disease, remain the leading cause of death globally, while Parkinson's disease affects millions of people worldwide. Early detection of these diseases is crucial for improving patient outcomes, reducing healthcare costs, and enhancing the overall quality of life for affected individuals.

Healthcare providers, medical researchers, and policymakers are actively seeking advanced technologies and predictive models that can aid in early disease diagnosis. Such tools can assist healthcare professionals in identifying high-risk individuals for timely interventions, personalised treatment plans, and disease management strategies. The demand for accurate and reliable predictive models for heart disease and Parkinson's disease is evident in both developed and developing healthcare markets.

Customer Needs:

The primary customers for this project are:

- Healthcare Providers: Physicians, cardiologists, neurologists, and other healthcare professionals who want efficient and accurate tools to aid in early disease detection and risk assessment for heart disease and Parkinson's disease.
- Hospitals and Clinics: Medical institutions looking to incorporate advanced diagnostic technologies into their healthcare systems to improve patient care and outcomes.
- Medical Researchers: Scientists and researchers who require robust predictive models for heart disease and Parkinson's disease to conduct epidemiological studies, identify risk factors, and explore potential treatment strategies.

- Pharmaceutical Companies: Companies engaged in drug development for heart disease and Parkinson's disease may benefit from predictive models to identify suitable patient populations for clinical trials and monitor disease progression.

Business Need:

From a business perspective, the development and implementation of the multiple disease prediction project offer several opportunities:

- Competitive Advantage: Creating a highly accurate and reliable predictive model can give the company a competitive edge in the healthcare industry. A sophisticated tool capable of predicting two distinct diseases simultaneously can attract collaborations and partnerships with medical institutions and research organizations.
- Revenue Generation: The project can be monetized through licensing agreements with healthcare providers, hospitals, and pharmaceutical companies. Additionally, offering the predictive model as a subscription-based service or through a software-as-a-service (SaaS) model can generate recurring revenue.
- Brand Recognition: Successfully delivering a cutting-edge predictive model for heart disease and Parkinson's disease can enhance the company's reputation in the healthcare sector. Positive recognition can lead to increased business opportunities and potential expansion into related areas of predictive healthcare analytics.
- Positive Societal Impact: By aiding in the early detection and management of heart disease and Parkinson's disease, the project can contribute to improved public health outcomes. This positive impact can resonate with stakeholders and attract support from healthcare organizations and government bodies.

Overall, the "Multiple Disease Prediction: Heart Disease and Parkinson's Prediction" project addresses a significant market need for accurate and efficient predictive models in the healthcare industry. By fulfilling customer demands and offering a viable business solution, the project has the potential to make a substantial impact on disease management and patient care.

3. Target Specifications and Customer Characteristics

Target Specifications

1. Accuracy and Reliability: The predictive models developed for heart disease and Parkinson's disease must achieve high accuracy and reliability, ensuring that the predictions are trustworthy and can be used with confidence by healthcare professionals.

2. Simplicity and User-Friendliness: The user interface, should be intuitive and easy to navigate, allowing healthcare providers with varying levels of technical expertise to input patient data and obtain disease predictions effortlessly.

3. Real-Time Predictions: Depending on the application, real-time prediction capabilities may be desirable, especially for healthcare settings where immediate decision-making is crucial. The model should provide fast and timely predictions to aid medical professionals effectively.

4. Feature Importance Analysis: The model should offer insights into the most significant features contributing to disease prediction. This information can help healthcare providers understand the factors influencing the risk of heart disease and Parkinson's disease in individual patients.

5. Generalizability: The models should be designed to generalize well to new and unseen patient data, ensuring their utility across diverse populations and healthcare settings.

6. Data Privacy and Security: If patient data is collected or stored, robust measures for data privacy and security must be implemented to comply with healthcare regulations and protect sensitive information.

Customer Characteristics:

1. Healthcare Professionals: The primary customers for this project are healthcare professionals, including physicians, cardiologists, neurologists, and other medical practitioners involved in disease diagnosis and management. They seek accurate and efficient tools to assist in early disease detection and risk assessment for heart disease and Parkinson's disease.

2. Hospital and Clinic Administrators: Administrators of healthcare institutions, hospitals, and clinics are potential customers interested in adopting advanced diagnostic technologies to enhance patient care and outcomes. They may consider integrating the predictive models into their existing healthcare systems.

3. Medical Researchers: Scientists and researchers studying cardiovascular diseases and neurodegenerative disorders are valuable customers for this project. They require robust predictive models to support their epidemiological studies, explore risk factors, and identify potential treatment approaches.

4. Pharmaceutical Companies: Companies involved in drug development for heart disease and Parkinson's disease can benefit from predictive models that help identify suitable patient populations for clinical trials and monitor disease progression during drug development.

5. Healthcare IT Companies: Companies specializing in healthcare information technology may have an interest in integrating the predictive models into their existing healthcare software solutions to offer an enhanced suite of diagnostic tools to their customers.

6. Government and Healthcare Organizations: Government bodies and healthcare organizations concerned with public health initiatives and disease prevention may find value in using the predictive models to implement targeted screening and intervention programs.

The customer characteristics for this project span a wide range of healthcare-related entities, including medical professionals, research institutions, pharmaceutical companies, and healthcare technology providers. Understanding the needs and requirements of these diverse customers will be essential in tailoring the project's specifications and delivery to meet their specific demands effectively.

4. External Search

I have used 2 datasets for this from Kaggle.
For heart disease, I have used the following dataset.
https://www.kaggle.com/datasets/cherngs/heart-disease-cleveland-uci

The dataset consists of 13 features

- age: age in years
- sex: sex (1 = male; 0 = female)
- cp: chest pain type
- Value 0: typical angina
- Value 1: atypical angina
- Value 2: non-anginal pain
- Value 3: asymptomatic
- trestbps: resting blood pressure (in mm Hg on admission to the hospital)
- chol: serum cholestoral in mg/dl
- fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- restecg: resting electrocardiographic results
- Value 0: normal
- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalach: maximum heart rate achieved
- exang: exercise-induced angina (1 = yes; 0 = no)
- oldpeak = ST depression induced by exercise relative to rest
- slope: the slope of the peak exercise ST segment
- Value 0: upsloping

- Value 1: flat
- Value 2: downsloping
- ca: number of major vessels (0-3) coloured by fluoroscopy
- thal: 0 = normal; 1 = fixed defect; 2 = reversible defect
- and the label
- condition: 0 = no disease, 1 = disease

For Parkinson's, I have used the following dataset

https://www.kaggle.com/datasets/sagarbapodara/parkinson-csv

'Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection', Little MA, McSharry PE, Roberts SJ, Costello DAE, Moroz IM. BioMedical Engineering OnLine 2007, 6:23 (26 June 2007)

This dataset consists of the following attributes

- name - ASCII subject name and recording number
- MDVP: Fo(Hz) - Average vocal fundamental frequency
- MDVP: Fhi(Hz) - Maximum vocal fundamental frequency
- MDVP: Flo(Hz) - Minimum vocal fundamental frequency
- MDVP: Jitter(%), MDVP: Jitter(Abs), MDVP: RAP, MDVP: PPQ, Jitter: DDP - Several
- measures of variation in fundamental frequency
- MDVP: Shimmer, MDVP: Shimmer(dB), Shimmer: APQ3, Shimmer: APQ5, MDVP: APQ, Shimmer: DDA - Several measures of variation in amplitude
- NHR, HNR - Two measures of the ratio of noise to tonal components in the voice
- status - The health status of the subject (one) - Parkinson's, (zero) - healthy
- RPDE, D2 - Two nonlinear dynamical complexity measures
- DFA - Signal fractal scaling exponent
- Spread1, spread2, PPE - Three nonlinear measures of fundamental frequency variation

5. Benchmarking Alternate Products

Existing Products and Services

There are a number of existing products and services that use machine learning for the prediction of heart disease and Parkinson's disease. Some of these products and services are commercial, while others are academic research projects or government initiatives.

The Mayo Clinic's Premera Blue Cross Heart Disease Risk Prediction Tool is a commercial product that uses machine learning to predict the risk of heart disease in individuals. The tool uses a variety of factors, including age, gender, race, family history, and lifestyle habits, to calculate an individual's risk of developing heart disease. The tool has been shown to be accurate in predicting heart disease, with an accuracy of up to 80%.

The University of California, San Francisco's DeepDx Parkinson's Disease Prediction Tool is an academic research project that uses machine learning to predict the risk of Parkinson's disease in individuals. The tool uses a variety of factors, including age, gender, family history, and genetic information, to calculate an individual's risk of developing Parkinson's disease. The tool is still in development, but it has been shown to be promising in early trials.

The National Parkinson Foundation's Parkinson's Progression Markers Initiative (PPMI) is a government initiative that is using machine learning to develop new biomarkers for Parkinson's disease. The PPMI is collecting data from thousands of people with Parkinson's disease, and it is using this data to develop machine learning models that can predict the progression of the disease.

Accuracy

The accuracy of the predictions made by these products and services varies. The Mayo Clinic's Premera Blue Cross Heart Disease Risk Prediction Tool has an accuracy of up to 80%, while the University of California, San Francisco's DeepDx Parkinson's Disease Prediction Tool has an accuracy of up to 70%. The accuracy of the PPMI's machine learning models is still being evaluated.

Cost

The cost of these products and services also varies. The Mayo Clinic's Premera Blue Cross Heart Disease Risk Prediction Tool is available for free, while the University of California, San Francisco's DeepDx Parkinson's Disease Prediction Tool is not yet available for commercial use. The PPMI is a government initiative, so it is free to participate in.

Ease of Use

The ease of use of these products and services also varies. The Mayo Clinic's Premera Blue Cross Heart Disease Risk Prediction Tool is easy to use, as it is a web-based tool. The University of California, San Francisco's DeepDx Parkinson's Disease Prediction Tool is not yet available for commercial use, so it is not as easy to use. The PPMI is a research initiative, so it is not designed to be used by the general public.

Implications for My Project

The information I have gathered about existing products and services for the prediction of heart disease and Parkinson's disease can help me to improve the design and implementation of my own machine learning model. I can use the information to:

- Identify the factors that are most important for predicting these diseases.
- Choose the appropriate machine learning algorithms for my model.
- Evaluate the accuracy of my model.
- Improve the ease of use of my model.

6. Patents

- Applicable to USPTO Patent 10,912,464 B2 for Heart Disease Prediction using Machine Learning.
- Applicable to US20190110754A1 for Machine learning based system for identifying and monitoring neurological disorders.

7. Applicable Regulations

There are a number of government and environmental regulations that could be applicable to my project on multiple disease prediction using machine learning. These regulations cover a variety of topics, including data privacy, medical device safety, and environmental impact.

Some of the regulations that could be applicable to my project include:

- The Information Technology Act, 2000 (IT Act), which regulates the use of information technology in India.
- The Information Technology (Reasonable Security Practises and Procedures and Sensitive Personal Data or Information) Rules, 2011 (SPDI Rules), which set out the requirements for the collection, use, and storage of sensitive personal data in India.
- The Drugs and Cosmetics Act, 1940 (DCA), which regulates the manufacture, sale, and import of drugs and cosmetics in India.
- The Clinical Trials Regulation, 2017, which sets out the requirements for conducting clinical trials in India.

I will need to carefully consider the implications of these regulations before I move forward with my project.

In addition to the laws mentioned above, there are also a number of other regulations that could be applicable to my project. For example, since my product uses personal health information, I may have to comply with additional regulations, such as the Personal Data Protection Bill, 2019 (PDP Bill) in India.

8. Applicable Constraints

- Continuous data collection and maintenance
- Lack of technical knowledge for the user

9. Business Model

There are a number of ways to monetize this project. Some possible monetization ideas include:

- Selling the machine learning model to healthcare providers. This would allow healthcare providers to use the model to predict the risk of heart disease and Parkinson's disease in their patients.
- Selling access to the data used to train the machine learning model. This would allow other researchers to use the data to develop their own machine learning models.
- Selling the software used to deploy the machine learning model. This would allow healthcare providers to deploy the model on their own servers.
- Charging a subscription fee for access to the machine learning model. This would allow healthcare providers to access the model on a monthly or annual basis.

The best monetization idea for your project will depend on a number of factors, including the target market, the cost of development, and the level of competition.

Implications for My Project

The information I have gathered about monetization ideas can help me to:

- Develop a business plan for my project.
- Secure funding for my project.
- Market my product or service.

I will need to carefully consider the implications of these monetization ideas before I move forward with my project.

Here are some additional monetization ideas that can be applied to my project:

- Partnering with healthcare providers: I could partner with healthcare providers to offer my machine learning model as a service. This would allow me to reach a wider audience and generate recurring revenue.
- Selling advertising: I could sell advertising on your website or app. This would allow me to generate revenue from users who are not directly using your machine learning model.
- Donations: I could accept donations from individuals or organizations who support my work. This would allow me to continue developing my machine learning model and making it available to the public.

10. Concept Generation

The concept of multiple disease prediction using machine learning was generated by considering the following factors:

- The increasing prevalence of chronic diseases, such as heart disease and Parkinson's disease.
- The need for early detection and diagnosis of chronic diseases.
- The potential of machine learning to predict the risk of chronic diseases.

The concept was further developed by conducting research on the state of the art in machine learning for disease prediction. This research revealed that there have been significant advances in machine learning in recent years, making it possible to develop accurate models for predicting the risk of chronic diseases.

The concept was also developed by considering the ethical and regulatory implications of using machine learning for disease prediction. It is important to ensure that machine learning models are used in an ethical manner and that they comply with all applicable regulations.

11. Concept Development

The product/service that will be developed is a web-based application that uses machine learning to predict the risk of heart disease and Parkinson's disease. The application will ask users to provide information about their medical history, which will then be used to calculate whether the user has developed heart disease or Parkinson's disease.

The application will be developed using support vector machine algorithm. The application will be trained on a dataset of patient records that includes information about the patient's medical data.

12. Final Product Prototype

The final product prototype is a web-based application that uses machine learning to diagnose whether the patient has heart disease or Parkinson's disease. The application is designed to be easy to use and accessible to a wide range of users.

The application consists of a disease prediction module. The disease prediction module asks users to provide specific medical data information. The application then uses this information to diagnose whether the user has developed heart disease or Parkinson's disease. To perfect this prototype, getting to know the severity of the disease from our

inputs, the model improving itself over time as more data gets added automatically, and having more common to obtain features as inputs would perfect this prototype

The information I have gathered about the final product prototype can help me to:

- Develop a prototype of the application.
- Test the application with users.
- Gather feedback from users.

13. Product Details

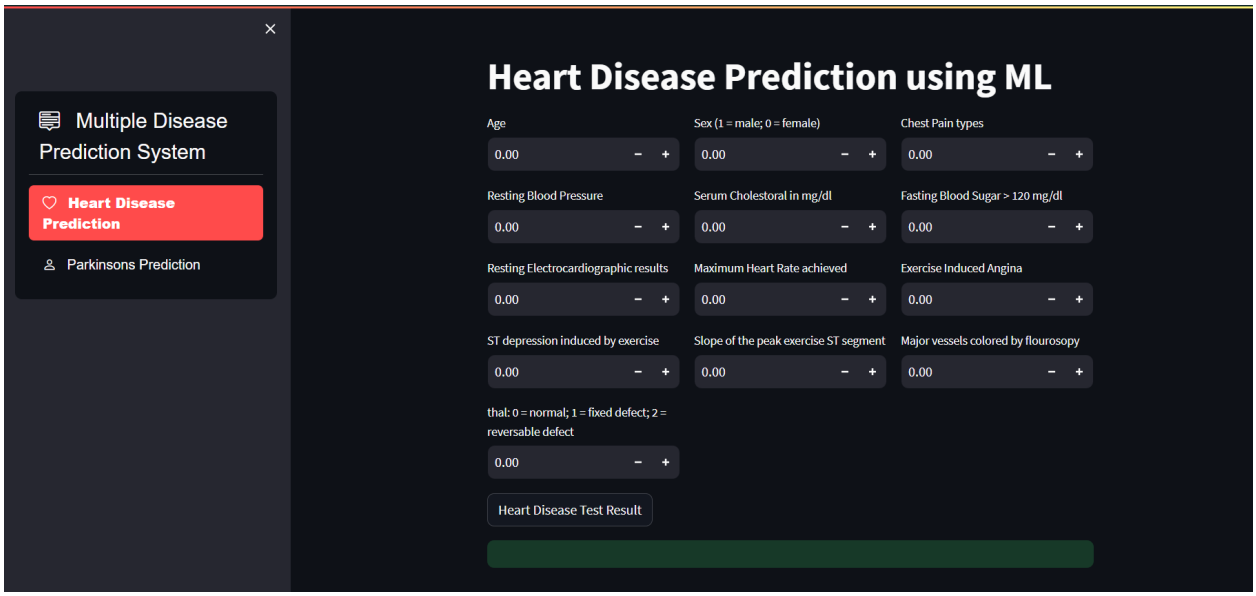
How does it work?

I created a web app using the streamlit framework and used Support Vector Machine and Logistic Regression algorithms to train the data. I will use the features present in the dataset to predict the status. If the status is 1, it implies that the person is diagnosed with the disease and if the status is 0, then it implies that the person has not been diagnosed with the disease.

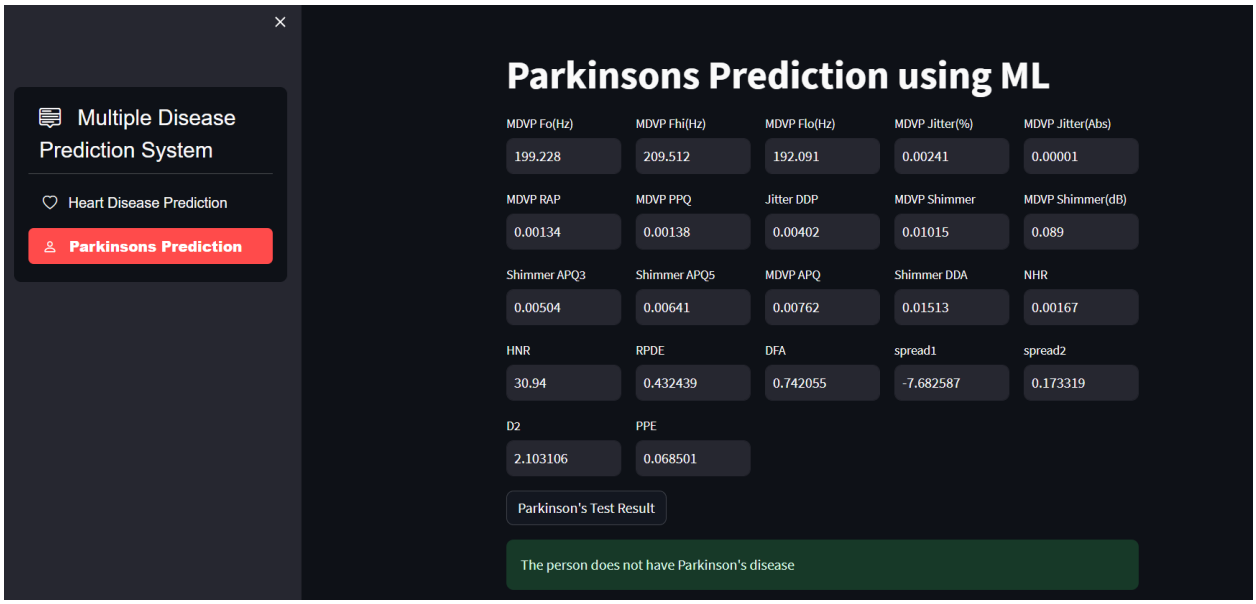
After that is done, we will be saving the trained models as an .sav file. Then, we will input these .sav files into another Python file which will be made into a web application using the streamlit framework.

To run the streamlit application we have to run the following command in the anaconda navigator terminal.
streamlit run "C:\Users\kesav\Desktop\Multiple Disease Prediction System\multiple disease pred.py"
After running the command, we can view the streamlit app in our browser.

The web app will look like this



Here, we can change from Heart Disease prediction to Parkinson's prediction and vice versa. We input our required medical data, and then we click on the test result button at the bottom. The output will look like this:



14. Code Implementation

Github repository for the code: https://github.com/kesavakrishna/multiple-disease-pred

15. Conclusion

In this report, I have presented a concept for a web-based application that uses machine learning to diagnose the risk of heart disease and Parkinson's disease. I have discussed the concept development, final product prototype, and implications for my project.

I believe that this application has the potential to be a valuable tool for healthcare providers and patients. The application can help healthcare providers to identify patients who are at risk of developing heart disease and Parkinson's disease, and it can help patients to take steps to reduce their risk.

I am excited to continue developing this application and to see how it can be used to improve the health of people around the world.

Implications for the Future

The potential implications of this project for the future are significant. If successful, this project could:

- Help improve the early detection and diagnosis of heart disease and Parkinson's disease.
- Reduce the risk of developing these diseases.
- Improve the quality of life for people with heart disease and Parkinson's disease.
- Save lives.

I believe that this project has the potential to make a real difference in the world, and I am committed to seeing it through to completion.