Heartrackr: A Heart Health App

Anubhav Sharma • Jyotsna Kumari • Kurapati Pavankumar • Sanhita Saxena

❖ GitHub Links for the Project:

- Anubhav Sharma: EDA (Heart Disease Indicator Data)
- Jyotsna Kumari: Prototype Development
- Kurapati Pavankumar: Web Application
- Sanhita Saxena: Market Segmentation

1. Abstract

This innovative Heart Disease Prediction web application serves as a predictive health companion, leveraging user-input data to assess the likelihood of heart disease and determine associated risk levels. Upon positive predictions, the app seamlessly integrates geolocation functionality, prompting users for their location to recommend nearby hospitals and suggest tailored medications. Beyond diagnostics, the platform extends its impact by providing personalized dietary guidance, empowering users to make informed and healthier food choices. With a commitment to user-friendliness, the app not only identifies potential health issues but also serves as a holistic guide, offering comprehensive insights into heart health and facilitating proactive healthcare decisions. The overarching goal is to create an accessible and informative tool that not only identifies health concerns but also guides individuals towards optimal care, ensuring a holistic approach to heart health management.

2. Problem Statement

Heart disease remains a leading cause of global mortality, often detected at later stages when intervention options are limited. Existing diagnostic approaches lack accessibility and user engagement, hindering early detection and proactive healthcare decisions. Additionally, individuals with positive predictions lack streamlined access to relevant healthcare facilities and personalized guidance, resulting in suboptimal health outcomes. There is a pressing need for an integrated and user-friendly solution that not only accurately predicts heart disease but also facilitates informed decision-making, leveraging geolocation for hospital recommendations and providing tailored dietary guidance. This project aims to address these gaps by developing a comprehensive web app that empowers users with predictive insights, guiding them towards timely healthcare interventions and healthier lifestyle choices.

3. Need Assessment

i. Market Need Assessment

The market assessment reveals a critical need for digital health solutions like the heart disease prediction app, driven by the global prevalence of heart disease and the lack of engaging early detection tools. The trend towards user empowerment and the demand for user-friendly health apps, coupled with geographical disparities in healthcare access, underscores the necessity for a tool that offers hospital recommendations based on user location. The rising interest in predictive health technologies and the gap in personalized dietary guidance present opportunities for comprehensive wellness tools. The ongoing shift to digital health solutions and a health-conscious consumer base further bolsters the market for apps that provide personalized guidance. The integration of technology in healthcare and the unique competitive edge of a holistic approach to health management highlight the app's competitive advantage.

ii. Customer Need Assessment

To effectively meet customer needs, the heart health app should be developed with a clear understanding of the target audience, represented through detailed user personas. It’s essential to address common pain points such as the need for early detection and accessible healthcare, while aligning with users’ goals for improved heart health and informed decision-making. Preferences for app design and functionality should be considered, ensuring accessibility for all users. Motivations for better health, feedback on existing apps, and desired features should guide the app’s development. Privacy concerns must be addressed transparently, and feedback mechanisms should be in place for continuous improvement. Cultural sensitivity and educational features are also crucial to empower users to take charge of their heart health.

iii. Business Need Assessment

A comprehensive business strategy for the heart health app should encompass a thorough market analysis to gauge demand and competition, alongside a robust revenue model. Cost estimation is crucial for development, launch, and maintenance, while regulatory compliance ensures adherence to health app standards. A solid technology stack is necessary, with strategic partnerships enhancing credibility. Marketing strategies should focus on user acquisition and

retention, supported by strong customer service and engagement plans. Data security and privacy are paramount, requiring clear communication to users. Plans for scaling and expansion must be adaptable, with risk analysis and feedback mechanisms integral for ongoing improvement and user trust.

4. Target Specification

Our health prediction application is tailored for adults residing in urban and suburban areas with convenient access to healthcare facilities. It specifically caters to professionals leading sedentary lifestyles, such as office workers, who may be more susceptible to heart health issues. Our target audience comprises individuals with a moderate to high level of technological proficiency, and comfortable navigating smartphones and web platforms. We prioritize language diversity and plan to offer multilingual support to ensure inclusivity. Our ideal users are moderately health-conscious individuals seeking proactive solutions for managing and enhancing their heart health. They typically have a moderate to high-income level, enabling them to afford healthcare services and technological advancements.

5. External Search

A concerning trend is emerging—a surge in heart problems across various demographics, particularly since the onset of the COVID-19 pandemic. This surge isn't limited to those with pre-existing risk factors; even young, healthy individuals are experiencing troubling cardiac issues. Experts are still unravelling the underlying reasons, but a combination of factors is likely at play. The pandemic undoubtedly heightened stress and anxiety levels due to social isolation, job uncertainties, and overall uncertainty, all of which can adversely affect heart health. Lifestyle adjustments, such as weight gain and unhealthy dietary habits during lockdowns, may also be contributing factors. Disruptions in preventive healthcare services due to lockdowns could lead to undetected or poorly managed conditions that exacerbate heart health issues. Additionally, some studies suggest that the COVID-19 virus itself may have a direct detrimental effect on the heart, even in mild cases, warranting further investigation. However, amidst this challenging scenario, there's a glimmer of hope—Artificial Intelligence (AI). With smartphones' widespread usage and connectivity, AI can become a potent tool in disease detection and management. Imagine AI-powered applications analyzing data from wearable health

trackers, monitoring heart rate, rhythm, and other vital signs. These applications could detect early warning signs, prompting users to seek medical attention. Furthermore, AI can sift through vast medical data to identify patterns and risk factors, enabling healthcare professionals to develop more targeted preventive strategies. By harnessing AI's potential and the widespread reach of smartphones, we can empower individuals to take a proactive stance on their heart health, potentially mitigating the alarming rise in cardiac issues.

6. Benchmarking Alternate Products

The investigation into the heart disease prediction web app space reveals several key competitors, including A-HealthPredict, CardioInsights, and HeartGuard Pro, each offering predictive analytics, geolocation-based hospital recommendations, and dietary guidance. While A-HealthPredict excels in user experience with its clean design, HeartGuard Pro struggles with usability. Our app aims to combine the best features to provide an intuitive interface. Competitors leverage geolocation for hospital recommendations, but our app seeks to refine this feature for precision. Similarly, our app aims to offer in-depth, personalized dietary recommendations, addressing disparities in competitors' offerings. All competitors prioritize privacy and security, but our app goes beyond compliance to enhance user trust. While CardioInsights leads in market presence, user feedback highlights the importance of a seamless experience, which our app strives to deliver. Our marketing strategy emphasizes targeted digital advertising, focusing on the app's unique predictive features. Finally, our app distinguishes itself through intuitive design, precise recommendations, and personalized guidance, with a future roadmap focusing on continuous updates based on user feedback.

7. Applicable Regulations

i. Health Data Protection Laws:

Comply with health data protection regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States or similar laws in other regions. Ensure the secure handling and storage of sensitive health information.

ii. Data Privacy Regulations:

Adhere to general data privacy regulations, such as the General Data Protection Regulation (GDPR) in the European Union, ensuring user consent, data transparency, and the right to be forgotten.

iii. Medical Device Regulations:

If the app qualifies as a medical device, comply with relevant medical device regulations, such as the U.S. Food and Drug Administration (FDA) regulations in the United States or the CE marking requirements in the European Union.

iv. Ethical Guidelines:

Follow ethical guidelines in the development of health-related applications, considering principles like fairness, transparency, and accountability in algorithmic decision-making.

v. Telemedicine Regulations:

If the app involves telemedicine or remote health monitoring, comply with applicable regulations governing telehealth services, ensuring legal and ethical standards are met.

8. Applicable Constraints

i. Technical Constraints
ii. Data Security Constraints
iii. User Accessibility Constraints
iv. Resource Constraints
v. Legal Constraints
vi. Regulatory Compliance Costs
vii. User Consent and Transparency

9. Business Opportunity and Monetization

The heart disease prediction and health guidance web application present several promising business opportunities:

i. Strategic Partnerships: Forge alliances with health and wellness platforms to extend our market reach and enhance user engagement.

ii. Health Insurance Collaboration: Partner with health insurance companies to provide our predictive health services as a value-added benefit to policyholders, promoting proactive health management.

iii. Corporate Wellness Programs: Approach businesses to implement our app as part of their corporate wellness initiatives, aiming to improve employee health outcomes and reduce healthcare costs.

iv. Healthcare Integration: Collaborate with healthcare providers to integrate our tool into patient portals, facilitating seamless access to predictive health insights and personalized guidance.

v. Pharmaceutical Partnerships: Explore partnerships with pharmaceutical companies to recommend appropriate medications based on predictive health data, enhancing treatment efficacy and patient outcomes.

vi. Educational Institution Engagement: Partner with educational institutions to offer health insights and guidance to students, promoting health literacy and proactive health behaviors from a young age.

vii. Government Health Initiatives: Collaborate with government health agencies on public health initiatives, leveraging our predictive health technology to support population-wide health monitoring and intervention efforts.

Monetization Strategy is as follows:

i. Subscription Plans:

   a. **Basic Plan**: Includes essential features such as heart disease risk assessment, personalized health recommendations, and access to community forums.

   b. **Standard Plan**: Includes all features of the Basic Plan plus integration with wearable devices, telemedicine consultations, and health challenges with rewards.

   c. **Premium Plan**: Includes all features of the Standard Plan plus advanced analytics, unlimited telemedicine consultations, and personalized nutrition plans.

d. **Family Plan**: This customizable option encompasses all the features offered in the standard plan. With a pay-as-you-add pricing model, users have the flexibility to tailor their plan according to their specific needs and preferences.

ii. Additional Revenue Streams:

a. **In-App Purchases**: Offers premium content, personalized health insights, and advanced analytics as optional add-ons for users across all subscription plans.

b. **Corporate Partnerships**: Collaborate with healthcare providers, insurance companies, and corporate wellness programs to offer discounted subscriptions or enterprise solutions.

10. Financial Equation

$$\boldsymbol{Total\ Revenue} = (\mathbf{300}\boldsymbol{x} + \mathbf{500}\boldsymbol{y} + \mathbf{800}\boldsymbol{z} + (\mathbf{200} + \mathbf{200}\boldsymbol{n}).\boldsymbol{f} + \mathbf{100}\boldsymbol{a}) \times (\frac{\boldsymbol{r}}{\mathbf{100}})^{\boldsymbol{t}} \times (\frac{\boldsymbol{p}}{\mathbf{100}})$$

$$\boldsymbol{Net\ Profit} = \boldsymbol{Total\ Revenue} - \boldsymbol{Total\ Cost}$$

$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ represent the number of subscribers with basic, standard, and premium plans respectively.

$\boldsymbol{n}, \boldsymbol{f}$ represent the number of members in the family and total families respectively.

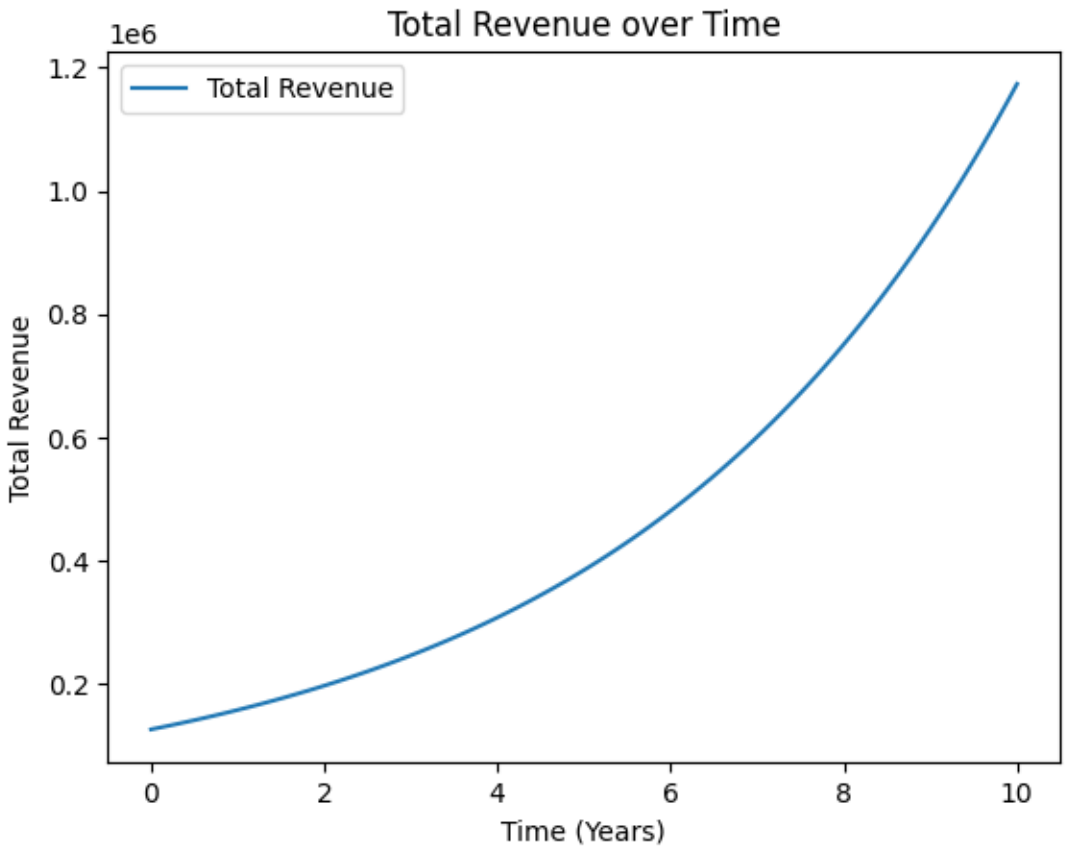
$\boldsymbol{a}$ represents number of customizable add-ons added.

$\boldsymbol{r}$ represents the market growth rate (in %)

$\boldsymbol{p}$ represents the web app market penetration rate (in %)

$\boldsymbol{t}$ represents time

$\boldsymbol{Total\ Cost}$ consists of includes all operational expenses incurred in the development, maintenance, and promotion of the app, including labor costs, workspace setup, database management, selling and distribution costs, advertisement costs, and any other operational expenses.



11. Final Product Prototype

i. Data Collection and Pre-processing

In this project, we aimed to develop a machine learning model for predicting heart disease using data collected from the Behavioral Risk Factor Surveillance System (BRFSS) conducted by the CDC. The dataset, sourced from Kaggle, comprised responses from 441,455 individuals and encompassed 330 features. Preprocessing and feature engineering steps were performed to ensure the dataset was suitable for training machine learning models to predict heart disease with optimal performance. These steps involved handling missing values and performing feature engineering to enhance the predictive power of the model. Our objective was to empower healthcare professionals with a tool capable of early identification of individuals at risk of heart disease, facilitating timely intervention and preventive measures. Through rigorous model development and evaluation, we aimed to provide a reliable predictive tool that could assist in the

management and prevention of heart disease, ultimately contributing to improved public health outcomes.

Feature Engineering

```
columns_to_drop = ['AnyHealthcare', 'NoDocbcCost', 'Education', 'Income']

df = data.drop(columns=columns_to_drop)
```

```
df.head()
```

	HeartDiseaseorAttack	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	Diabetes	PhysActivity	Fruits	Veggies	HvyAlcoholCo
0	0.0	1.0	1.0	1.0	40.0	1.0	0.0	0.0	0.0	0.0	1.0	
1	0.0	0.0	0.0	0.0	25.0	1.0	0.0	0.0	1.0	0.0	0.0	
2	0.0	1.0	1.0	1.0	28.0	0.0	0.0	0.0	0.0	1.0	0.0	
3	0.0	1.0	0.0	1.0	27.0	0.0	0.0	0.0	1.0	1.0	1.0	
4	0.0	1.0	1.0	1.0	24.0	0.0	0.0	0.0	1.0	1.0	1.0	

```
# Move the "HeartDiseaseorAttack" column to the last position
cols = list(df.columns)
cols.remove("HeartDiseaseorAttack")
cols.append("HeartDiseaseorAttack")
data = df[cols]
```

Data Preprocessing

```
data.isnull().sum()
```

```
HighBP                0
HighChol              0
CholCheck             0
BMI                   0
Smoker                0
Stroke                0
Diabetes              0
PhysActivity          0
Fruits                0
Veggies               0
HvyAlcoholConsump     0
GenHlth               0
MentHlth              0
PhysHlth              0
DiffWalk              0
Sex                   0
Age                   0
HeartDiseaseorAttack  0
dtype: int64
```

```
# Dataset information
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 253680 entries, 0 to 253679
Data columns (total 18 columns):
 #   Column                Non-Null Count   Dtype
---  ------                --------------   -----
 0   HighBP                253680 non-null  float64
```

ii. Model Development:

In the model development phase, we experimented with five different machine learning algorithms: logistic regression, random forest, support vector machine (SVM), gradient boosting machine (GBM), and neural networks. Each algorithm

was trained and evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Interestingly, during model evaluation, we observed that GBM, logistic regression, and neural networks achieved comparable results, displaying similar performance on the evaluation dataset.

However, when tested with new unseen data, GBM and neural networks exhibited suboptimal performance, particularly in cases where the dataset contained imbalanced proportions in the target classes. To address this issue, we applied oversampling and resampling techniques to rebalance the dataset and mitigate the effects of class imbalance. Subsequently, logistic regression emerged as the most reliable model, demonstrating superior performance on the new unseen data compared to GBM and neural networks. Its robustness in handling imbalanced datasets, coupled with its simplicity and interpretability, made logistic regression the preferred choice for our heart disease prediction model. Overall, the iterative process of model evaluation and selection underscored the importance of considering both performance metrics and practical considerations in choosing the most suitable algorithm for real-world applications.

```
import pandas as pd

# Assuming you have already calculated the evaluation metrics for each algorithm
# Replace the values below with the actual evaluation metrics
metrics_data = {
    'Algorithm': ['Logistic Regression', 'Random Forest', 'Gradient Boosting', 'SVM', 'Neural Network'],
    'Accuracy': [lr_accuracy, rf_accuracy, gb_accuracy, svm_accuracy, accuracy],
    'Precision': [lr_precision, rf_precision, gb_precision, svm_precision, precision],
    'Recall': [lr_recall, rf_recall, gb_recall, svm_recall, recall],
    'F1 Score': [lr_f1_score, rf_f1_score, gb_f1_score, svm_f1_score, f1]
}

# Create a DataFrame from the metrics data
metrics_df = pd.DataFrame(metrics_data)

# Display the DataFrame
print(metrics_df)
```

```
           Algorithm  Accuracy  Precision    Recall  F1 Score
0  Logistic Regression  0.906181   0.516634  0.110138  0.181568
1        Random Forest  0.896523   0.380252  0.151022  0.216184
2    Gradient Boosting  0.906693   0.530738  0.108052  0.179549
3                  SVM  0.905511   0.000000  0.000000  0.000000
4       Neural Network  0.906851   0.544271  0.087192  0.150306
```

Based on the above results and considering both performance and computational complexity, we can choose the Logistic Regression model

Resampling Techniques

```
from imblearn.over_sampling import RandomOverSampler

# Create an instance of RandomOverSampler
oversampler = RandomOverSampler(random_state=42)

# Resample the dataset
X_resampled, y_resampled = oversampler.fit_resample(X, y)
```

```
from imblearn.under_sampling import RandomUnderSampler

# Create an instance of RandomUnderSampler
undersampler = RandomUnderSampler(random_state=42)

# Resample the dataset
X_resampled, y_resampled = undersampler.fit_resample(X, y)
```

```
from sklearn.linear_model import LogisticRegression

# Create an instance of Logistic Regression with class weights
lr_model = LogisticRegression(class_weight='balanced', random_state=42)

# Train the model
lr_model.fit(X_train, y_train)
```

iii. Web Application Prototype:

In the Web Application Prototype phase, we aimed to create an intuitive and user-friendly interface for our heart disease prediction model. Leveraging Streamlit, a powerful framework for building interactive web applications with Python, we developed a streamlined application that allows users to input their health information and receive predictions regarding their risk of heart disease.

The application interface was designed to be simple and accessible, featuring user-friendly input fields and clear instructions for data entry. Users are prompted to provide information such as age, gender, blood pressure, cholesterol levels, and lifestyle habits. Upon submitting their data, the application processes the inputs using the trained logistic regression model and generates a prediction regarding the likelihood of the user developing heart disease.

To enhance user experience, we incorporated informative messages and tooltips throughout the application to guide users through the prediction process. Additionally, we implemented error handling mechanisms to ensure the robustness and reliability of the application, providing feedback to users in case of invalid inputs or technical issues.

Overall, the Web Application Prototype serves as a practical tool for individuals and healthcare professionals alike, offering a convenient means of assessing heart disease risk based on personal health data. With its intuitive interface and accurate predictions, the application has the potential to aid in early detection and preventive measures, ultimately contributing to improved public health outcomes.

```python
import streamlit as st
import numpy as np
import pickle

# Load the trained model
with open('train_new_mod.sav', 'rb') as model_file:
    model = pickle.load(model_file)

# Define function to make predictions
def predict_heart_disease(features):
    # Reshape the input features into the format expected by the model
    input_features = np.array(features).reshape(1, -1)
    # Make prediction
    prediction = model.predict(input_features)
    return prediction

# Define the web app
def main():
    st.title("Heart Health Predictor")

    # Add image
    # Add image with custom height and width using CSS styling
    image = 'https://pics.craiyon.com/2023-07-05/64c0ff3658c44869b4ad53dfed590f72.webp'  # Replace 'your_image.png' with the path to your image file
    custom_height = 300  # Adjust as needed
    custom_width = 500  # Adjust as needed
    st.image(image, width=custom_width, caption='Your Image', use_column_width=False)



    # Add sidebar for user input
    st.sidebar.header("User Input")
```

```python
                                        "80 or older"])
age_categories = {
    "18 to 24": 1,
    "25 to 29": 2,
    "30 to 34": 3,
    "35 to 39": 4,
    "40 to 44": 5,
    "45 to 49": 6,
    "50 to 54": 7,
    "55 to 59": 8,
    "60 to 64": 9,
    "65 to 69": 10,
    "70 to 74": 11,
    "75 to 79": 12,
    "80 or older": 13
}
age = age_categories[age_category]
high_bp = st.sidebar.radio("High Blood Pressure", ["No", "Yes"])
high_bp = 1 if high_bp == "Yes" else 0
high_chol = st.sidebar.radio("High Cholesterol", ["No", "Yes"])
high_chol = 1 if high_chol == "Yes" else 0
chol_check = st.sidebar.radio("Cholesterol Check ( Past 5 years)", ["No", "Yes"])
chol_check = 1 if chol_check == "Yes" else 0
bmi = st.sidebar.slider("BMI = weight (kg) / (height (m))^2")
smoker = st.sidebar.radio("Smoker ( atleast 100 cigarettes )", ["No", "Yes"])
smoker = 1 if smoker == "Yes" else 0
stroke = st.sidebar.radio("Stroke", ["No", "Yes"])
stroke = 1 if stroke == "Yes" else 0
diabetes = st.sidebar.radio("Diabetes", ["No", "Yes"])
diabetes = 1 if diabetes == "Yes" else 0
```

```
    # Collect the user input into a list
    user_input = [high_bp, high_chol, chol_check, bmi, smoker, stroke, diabetes,
                  phys_activity, fruits, veggies, heavy_alcohol,
                  gen_health, ment_health, phys_health, diff_walk,
                  sex, age]

    # Add a button to trigger the prediction
    if st.sidebar.button("Result"):
        # Predict the heart disease based on the user input
        prediction = predict_heart_disease(user_input)

        # Display the prediction result
        st.write("")
        st.write("## Result")
        if prediction[0] == 1:
            st.write("The model predicts that the person is likely to have heart disease.")
        else:
            st.write("The model predicts that the person is not likely to have heart disease.")

        # Feedback form
        st.header("Feedback Form")
        name = st.text_input("Name", "")
        email = st.text_input("Email", "")
        feedback = st.text_area("Feedback Message", "")
        satisfaction_rating = st.slider("Satisfaction Rating", 1, 5)
        improvement_suggestions = st.text_area("Improvement Suggestions", "")
        if st.button("Submit Feedback"):
            # Process feedback submission (you can add your code to handle feedback submission here)
            st.success("Thank you for your feedback!")

if __name__ == "__main__":
    main()
```

The Streamlit-based web application prototype provides a user-friendly interface for accessing the heart disease prediction model, allowing users to input their health information and receive personalized risk assessments. While logistic regression emerged as the optimal model for this task, the iterative process of model evaluation and selection underscored the importance of considering performance metrics, algorithmic complexity, and practical considerations in real-world applications.

Moving forward, further validation and refinement of the model, as well as integration into healthcare systems, will be essential to maximize its impact on early detection and prevention efforts for heart disease. By empowering healthcare professionals with a reliable predictive tool, this project aims to contribute to improved patient outcomes, reduced healthcare costs, and ultimately, a healthier society.

12. Minimum Viable Product Features

i. Inputting Information: Develop an intuitive interface enabling users to input their health-related data, including indicators like high blood pressure, high cholesterol, BMI, and smoking status.

ii. Prediction: Create an intelligent tool that utilizes this information to predict potential heart problems in the future.

iii. Providing Results: Inform users about their likelihood of developing heart issues based on the data they have provided.

iv. Soliciting Feedback: Allow users to provide feedback on the accuracy of the app's predictions to enhance its performance.

v. Ensuring Data Security: Implement robust measures to safeguard users' health data, ensuring privacy and security.

vi. Educational Resources: Integrate educational materials within the app to educate users about the significance of various health indicators related to heart disease and recommend preventive measures.

vii. User Profile Management: Enable users to create and manage profiles within the app, facilitating the tracking of health data over time and monitoring changes in risk factors.

viii. Reminder System: Send users reminders to regularly input their health information and review their results.

ix. Data Visualization: Utilize data visualization tools such as graphs or charts to present users' health data and risk assessments in a comprehensible format.

x. Compatibility: Ensure the app is compatible with various devices like smartphones and tablets, catering to diverse user needs.

xi. Offline Functionality: Provide offline functionality, allowing users to access certain features and resources even without an internet connection.

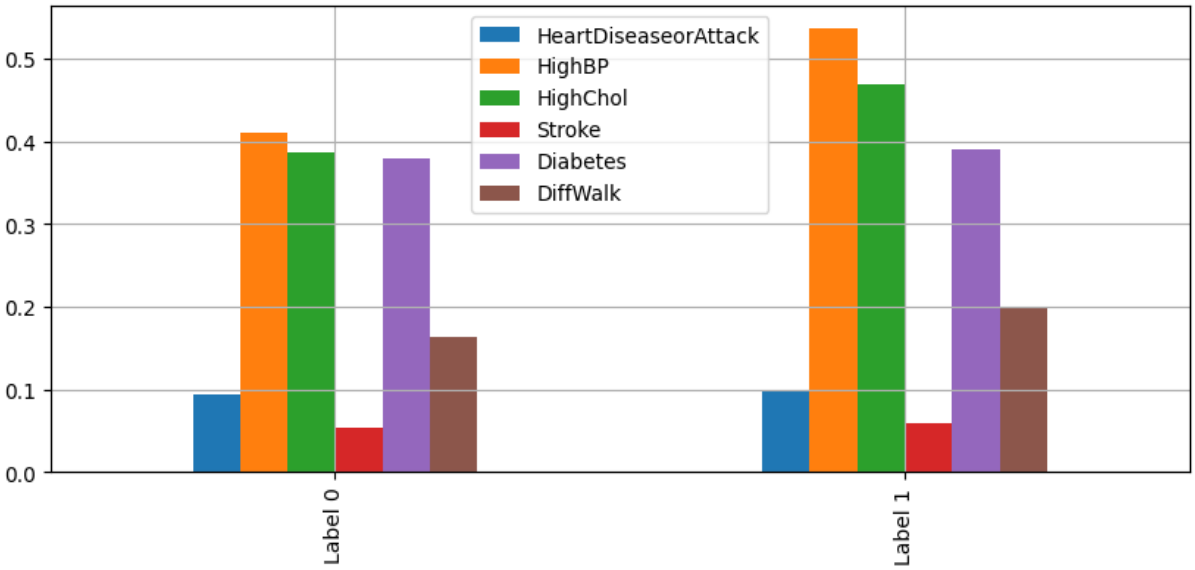
xii. Testing: Conduct thorough testing to ensure the app's functionality and usability before its release to the public.

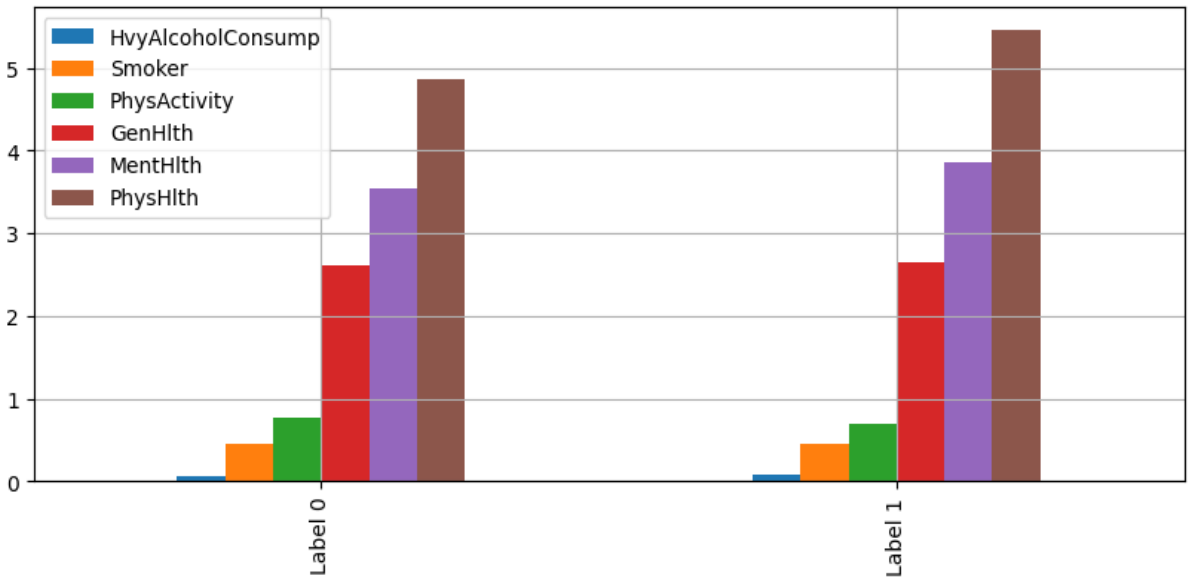
Conclusion: By prioritizing these essential features, we can develop a foundational version of the Heart Disease Health Indicators app that offers meaningful assistance to users. Subsequently, we can incorporate additional features and enhancements based on user feedback and preferences.

13. **Market Segmentation**

i. Demographic Segmentation

Segment *Characterization:*





a. Label 0 (Moderate Risk, Active Lifestyle):

- For Label 0, there is a relatively low prevalence of heart disease or heart attacks, high blood pressure, high cholesterol, stroke, and diabetes among individuals.
- For Label 0, there is a moderate prevalence of heavy alcohol consumption and smoking, with a relatively high level of physical activity. However, individuals report experiencing bad physical health for approximately 4.86 days within the past 30 days, alongside fair to moderate general health and mental health.

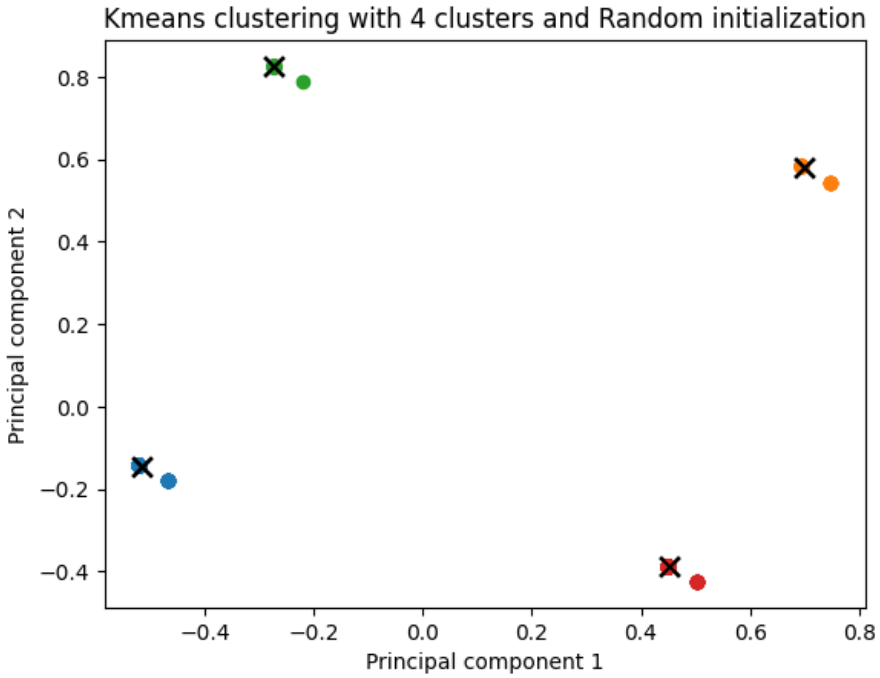
b. Label 1 (Higher Risk, Sedentary Lifestyle):

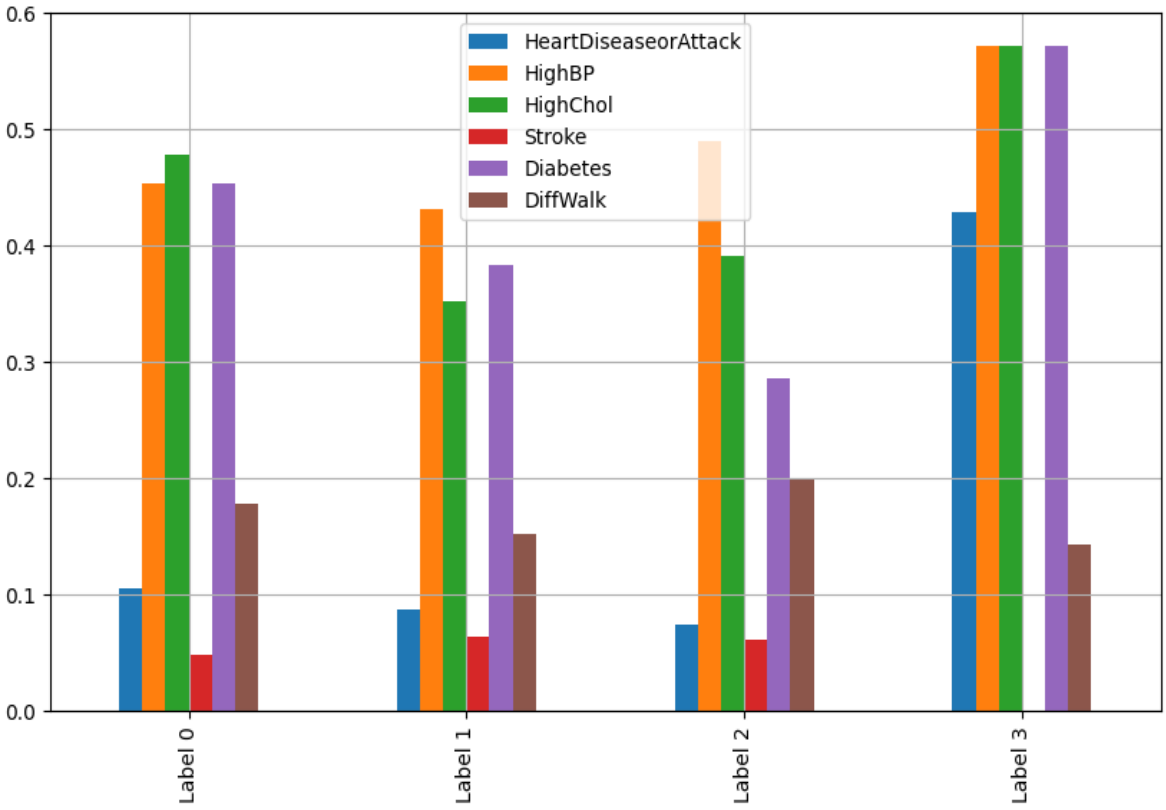
- For Label 1, there is a slightly higher prevalence of heart disease or heart attacks, high blood pressure, high cholesterol, stroke, and diabetes compared to Label 0, indicating a potentially higher risk profile for these health conditions.
- For Label 1, there is a slightly higher prevalence of heavy alcohol consumption and smoking compared to Label 0, with slightly lower physical activity. Individuals report experiencing bad physical health for approximately 5.46 days within the past 30 days, alongside fair general health, and moderate mental health.

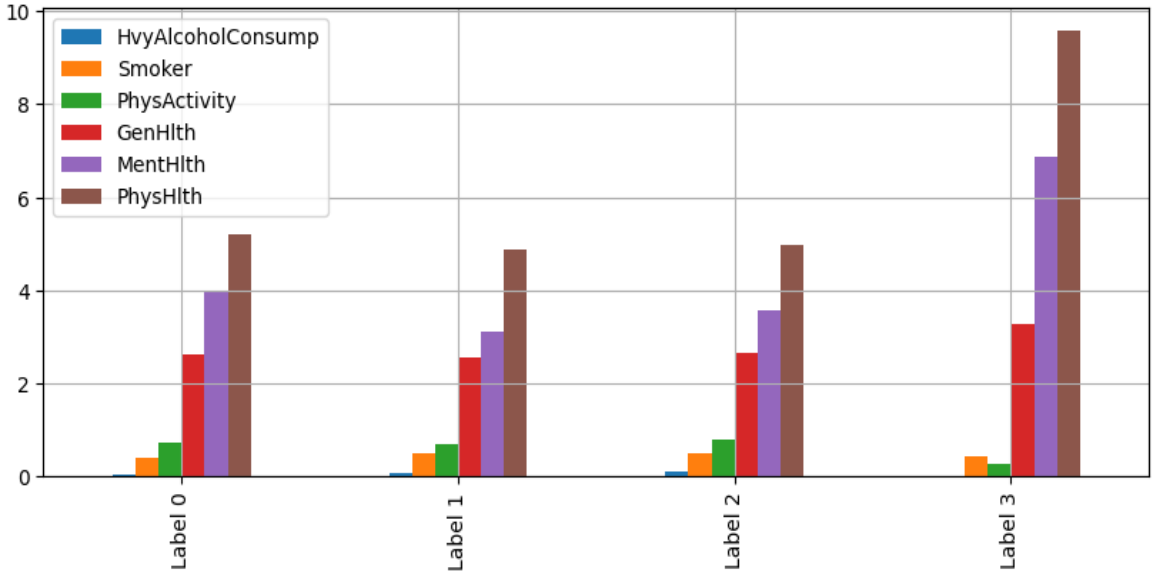
Selecting Target Segment:

- Both segments offer equal potential for targeting, with tailored experiences to address their unique needs. Initially focusing on the moderate risk, active lifestyle segment is strategic due to its larger population representation and proactive approach to health. Their commitment to a healthy lifestyle serves as a compelling incentive to adopt the app's services, providing a robust market foothold. By prioritizing this segment, we can establish a solid foundation while also resonating with individuals keen on proactive health management. As our expansion plans evolve, we can pivot towards targeting the higher-risk, sedentary lifestyle segment. This phased approach allows us to cater to a larger demographic initially, ensuring our offerings align with evolving needs. Strategically addressing each segment maximizes our reach, driving the app's success and fostering improved heart health outcomes for all users.

ii. Health Behavior Segmentation







Segment Characterization:

a. Label 0 (Low Risk):

- This group has a moderate prevalence of high blood pressure, high cholesterol, and diabetes, with a low occurrence of heart disease or

heart attacks and strokes. Additionally, individuals report experiencing some difficulties in walking.

- Individuals in this group have a relatively low prevalence of heavy alcohol consumption and smoking, along with a moderate level of physical activity. They report fair to moderate general health and mental health, with approximately 5.19 days of bad physical health within the past 30 days.

b. Label 1 (Moderate Risk):

- Individuals in this group exhibit a lower prevalence of heart disease or heart attacks, high blood pressure, high cholesterol, and diabetes compared to Label 0, but strokes occur slightly more frequently. Some individuals also report facing difficulties in walking.
- This group exhibits a slightly higher prevalence of heavy alcohol consumption and smoking compared to Label 0, with similar levels of physical activity. General health and mental health are slightly lower, and individuals report approximately 4.89 days of bad physical health within the past 30 days.

c. Label 2 (High Risk):

- There is a moderate prevalence of heart disease or heart attacks, high blood pressure, and high cholesterol, with a slightly lower occurrence of diabetes. The incidence of strokes is moderate, and individuals report experiencing difficulties in walking.
- There is a higher prevalence of heavy alcohol consumption compared to Labels 0 and 1, along with similar levels of smoking and physical activity. Individuals report fair to moderate general health and mental health, with approximately 4.98 days of bad physical health within the past 30 days.

d. Label 3 (Very High Risk):

- This group presents a significantly elevated risk profile for cardiovascular diseases, with a high prevalence of heart disease or heart attacks, high blood pressure, high cholesterol, and diabetes.

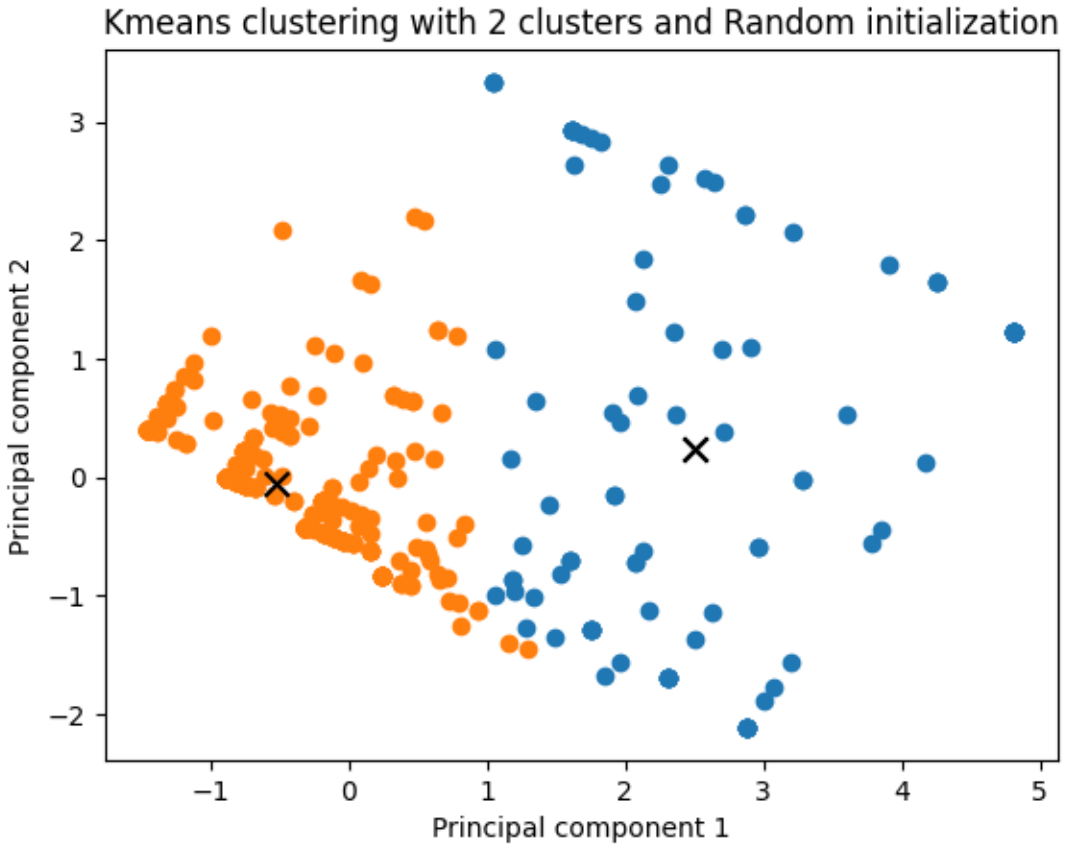
Strokes are infrequent, but individuals report considerable difficulties in walking.

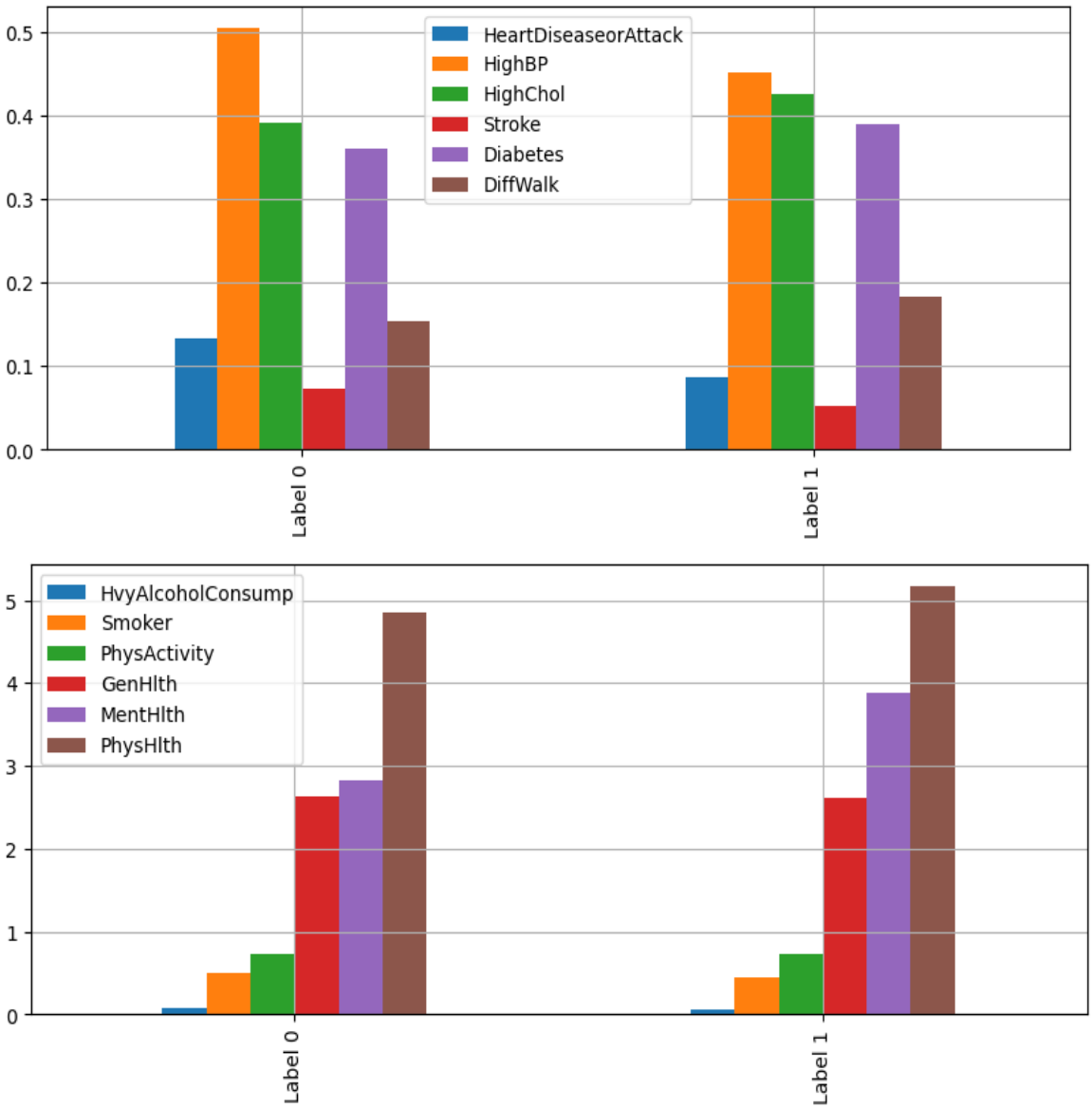
- Individuals in this group report the lowest level of physical activity among all labels, alongside the lowest prevalence of heavy alcohol consumption and smoking. However, they report the highest levels of bad physical health within the past 30 days, indicating significant health challenges.

Selecting Target Segment:

- Label 1 exhibits a lower prevalence of heart disease and associated risk factors compared to the higher-risk segments, yet individuals still face health challenges and may benefit from proactive health management. They have slightly higher levels of alcohol consumption and smoking, indicating potential areas for intervention and lifestyle modification. Additionally, their reported levels of physical activity suggest a willingness to engage in health-related behaviors.
- By targeting individuals in Label 1, we can position our app as a valuable tool for assessing and managing moderate heart health risks, offering personalized insights and guidance to support healthier lifestyles. This segment represents a sizable portion of the population that could benefit from early intervention and preventive measures to reduce their risk of developing heart-related conditions.

iii. Health Status Segmentation





Segment Characterization:

a. Label 0 (Moderate Risk, Active Lifestyle):

- Health Conditions: Individuals in Cluster 0 have a higher prevalence of heart disease or heart attacks, high blood pressure, and strokes compared to Cluster 1. However, they have a lower prevalence of high cholesterol and diabetes. They also report experiencing moderate difficulty in walking.
- Overall Health: Individuals in Cluster 0 exhibit a higher prevalence of heavy alcohol consumption and smoking compared to Cluster 1.

Despite having slightly lower levels of physical activity, they report fair to moderate general health and mental health. However, they experience bad physical health for approximately 4.86 days within the past 30 days.

b. Label 1 (Low Risk, Sedentary Lifestyle):

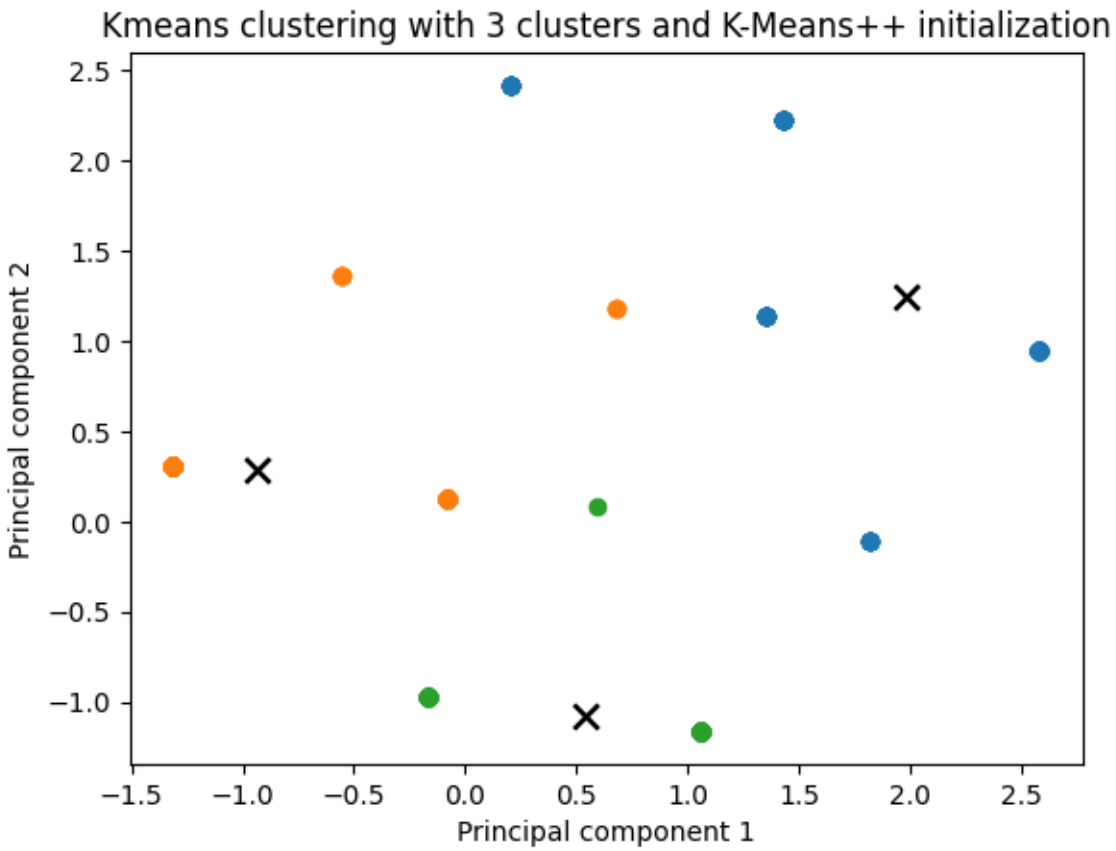
- Health Conditions: Individuals in Cluster 1 have a lower prevalence of heart disease or heart attacks, high blood pressure, strokes, and diabetes compared to Cluster 0. However, they exhibit a slightly higher prevalence of high cholesterol. They also report experiencing some difficulty in walking.
- Overall Health: Individuals in Cluster 1 have a lower prevalence of heavy alcohol consumption and smoking compared to Cluster 0. They exhibit similar levels of physical activity but report slightly better general health and mental health. They experience bad physical health for approximately 5.17 days within the past 30 days.

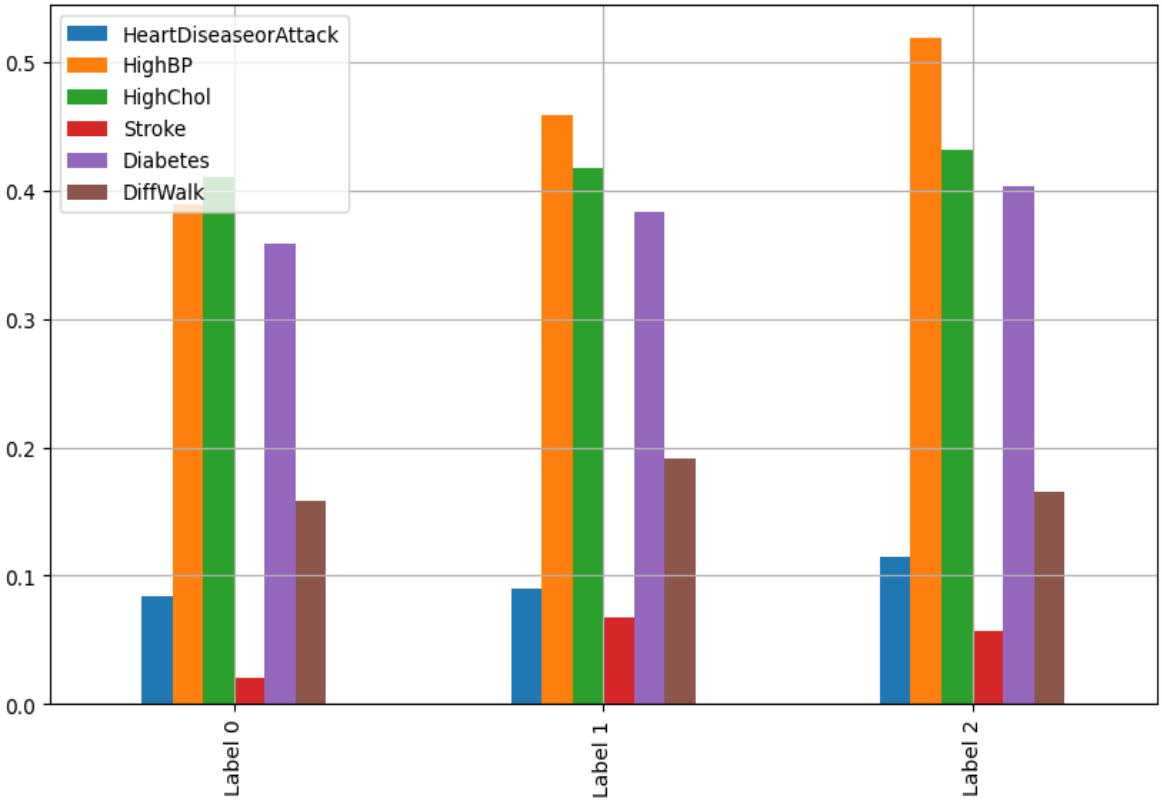
Target Selection

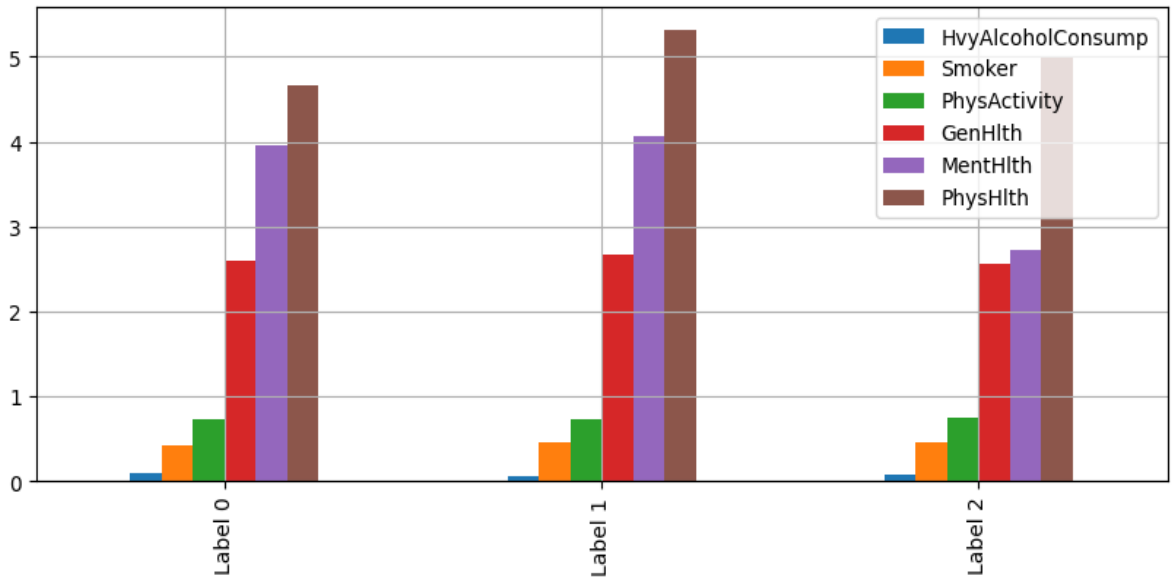
- Targeting **Label 0: Moderate Risk, Active Lifestyle** appears to be the optimal choice. Despite having a higher prevalence of heart-related conditions compared to **Label 1: Low Risk, Sedentary Lifestyle**, individuals in **Label 0** demonstrate a greater commitment to maintaining their health. They exhibit a proactive approach towards health management, with fair to moderate general and mental health, indicating receptiveness to health interventions. Thus, focusing on this segment aligns with their health-conscious behaviour and offers the potential for impactful engagement with our heart health prediction app.

iv. Prevalent Health Condition Segmentation

Segment Characterization:


Kmeans clustering with 3 clusters and K-Means++ initialization
Principal component 2
Principal component 1
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5


HeartDiseaseorAttack
HighBP
HighChol
Stroke
Diabetes
DiffWalk
0.5
0.4
0.3
0.2
0.1
0.0
Label 0
Label 1
Label 2



a. Label 0 (Moderate Risk, Sedentary Lifestyle):

- Moderate prevalence of cardiovascular risk factors, including heart disease, high blood pressure, high cholesterol, and diabetes.
- Individuals in this cluster report relatively lower levels of physical activity and fair to moderate overall health.

b. Label 1 (Low Risk, Active Lifestyle):

- Lower prevalence of cardiovascular risk factors compared to Cluster 0, with slightly lower incidences of heart disease, high blood pressure, high cholesterol, and diabetes.
- Individuals in this cluster exhibit similar levels of physical activity compared to Cluster 2 and report fair to moderate overall health.

c. Label 2 (High Risk, Active Lifestyle):

- Higher prevalence of cardiovascular risk factors compared to Cluster 0, with similar levels of incidence for heart disease, high blood pressure, high cholesterol, and diabetes.
- Individuals in this cluster report higher physical activity levels than Cluster 0 and fair overall health.

Target Selection:

- Selecting Low Risk, Active Lifestyle (Cluster 1) offers the best combination of lower prevalence of cardiovascular risk factors compared to Cluster 0 and similar physical activity levels to Cluster 2. Individuals in this cluster are likely to be more receptive to health interventions and proactive in maintaining their well-being. By targeting this segment, we can effectively engage individuals who are already leading active lifestyles and offer them personalized health insights and guidance to further optimize their cardiovascular health. This approach maximizes the potential for positive outcomes and user satisfaction.

14. Statistics

Heart disease is a major global health concern, with staggering statistics highlighting its prevalence and impact. According to the World Health Organization, it's responsible for approximately 17.9 million deaths worldwide annually. In the US, heart disease accounts for one in every four deaths, as reported by the Centers for Disease Control and Prevention.

Against this backdrop, the digital healthcare market, particularly mobile health applications (mHealth apps), is experiencing explosive growth. MarketsandMarkets projects the global digital health market to reach USD 636.6 billion by 2028, with a Compound Annual Growth Rate (CAGR) of 27.7%. Notably, within this market, mHealth apps are predicted to reach USD 340.5 billion by 2030, reflecting their effectiveness in addressing heart health concerns.

mHealth apps offer various tools to empower individuals in managing their heart health effectively. They enable remote monitoring of vital signs like heart rate and blood pressure, facilitate telemedicine consultations with cardiologists, provide personalized health information, and flag potential risk factors for early intervention.

The COVID-19 pandemic has further accelerated the adoption of mHealth apps, particularly for telemedicine consultations. As social distancing measures continue, the

convenience and effectiveness of digital healthcare solutions are becoming increasingly recognized.

Looking ahead, mHealth apps are poised to play a pivotal role in shaping the future of heart healthcare. By empowering individuals to take proactive steps in managing their heart health, these apps have the potential to significantly reduce the burden of heart disease and save lives. While challenges remain, the surge in mHealth app usage signals a digital revolution in heart health management, offering hope for improved health outcomes globally.

15. Conclusion

The heart disease prediction app is a game-changer in proactive health management, using cutting-edge machine learning and data science to deliver personalized insights into heart health. It's designed with users in mind, offering predictive analytics, wellness plans, telehealth services, and community forums—all aimed at empowering users to take control of their heart health. The app's business model combines a freemium approach, data monetization, and strategic partnerships to make it accessible and financially sustainable. It prioritizes data privacy, security, and regulatory compliance, ensuring ethical practices. Beyond just predicting heart disease, the app fosters overall well-being through lifestyle advice and community engagement. Its development process is dynamic, responding to user feedback and market trends to provide a practical roadmap for implementation. With a focus on continuous improvement and user satisfaction, the app aims to raise awareness and improve heart health outcomes significantly. Offering a range of plans, from Basic to Family, with customizable add-ons, and corporate collaborations, the app caters to diverse user needs, making it easy for individuals and families to manage their heart health proactively.

16. References

i. https://pubmed.ncbi.nlm.nih.gov/38278564/

ii. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

iii. https://www.cdc.gov/heartdisease/index.htm

iv. https://www.marketsandmarkets.com/Market-Reports/digital-health-market-45458752.html