Heart Attack Prediction Project Report CO21332

by Karandeep Singh

Chandigarh College Of Engineering & Technology

(Degree Wing) Sector-26, Chandigarh

DATA MINING
&
ANALYSIS

Heart Attack Prediction Project Report

CS-655C

Academic Year

2024-25

Submitted By:
Karandeep Singh

Submitted To:
Dr. Varun Gupta

CO21332
C.S.E – 6th Semester

Professor
C.S.E Department

Heart Attack Prediction Project Report

Introduction:

Heart disease is a serious global concern, causing significant loss of life worldwide. Early detection and prevention of heart attacks are crucial for improving patient outcomes and saving lives. That is why this project is so important—it is all about using machine learning to create a smart system that can predict the risk of heart attacks based on different health factors.

With heart-related mortality rates still alarmingly high, there is a pressing need to embrace advanced technologies to strengthen preventive healthcare measures. By tapping into the power of machine learning, we aim to give doctors and public health officials a powerful tool to identify individuals at higher risk of heart attacks.

We are blending cutting-edge computational techniques with deep knowledge of cardiology to pave the way for more precise medical interventions. Our goal is to equip healthcare providers with a tool that can assess risks and plan interventions proactively, shifting the focus from reacting to health issues to preventing them altogether.

So, how does it work? Well, we are using something called the XGBoost classifier. It's like having an expert on your team who's really good at spotting patterns in data. We've trained this expert to learn from past data and make predictions about whether someone might be at risk of having a heart attack, based on factors like age, blood pressure, and more.

By bringing together experts in data science, healthcare, and technology, we are working towards a future where heart disease is not as big of a threat, and where everyone can benefit from personalized healthcare strategies tailored to their unique needs. This project is a step towards making that vision a reality, ultimately improving public health and well-being for people everywhere.

Technologies Used:

Frontend:

- **HTML:** Think of HTML as the skeleton of a web page. It provides the structure and basic elements, like headings, paragraphs, forms, and more, organizing the content in a way that is easy for users to understand. In our project, HTML creates the form where users can input their medical data to predict heart attacks.
- **CSS:** Cascading Style Sheets give your web page its look and feel. They let you customize fonts, colors, layouts, and spacing, making your page visually appealing and easy to navigate. In our project, CSS is used to style the form and background, making everything look nice and cohesive.
- **JavaScript:** JavaScript adds life to your web page, making it interactive and dynamic. It is like the behind-the-scenes magic that responds to user actions, like scrolling or clicking. In our project, JavaScript creates a cool parallax effect on the background image, adding depth and visual interest as users scroll.

Figure 1: Frontend Technologies used

Backend:

- **Flask:** Flask is like the backbone of our web application. It handles the behind-the-scenes work, like managing URLs, processing user input, and rendering pages. It is flexible and easy to use, making it perfect for building web apps. In our project, Flask takes care of all the server-side stuff, making sure everything runs smoothly.

Figure 2: Flask Logo

- **Pickle:** Pickle is a handy tool in Python that helps us save and load our machine learning model. It is like putting our model in a box so we can store it for later and use it whenever we need it. In our project, Pickle saves our trained model to disk, ready to be unpacked and used for predictions.

Figure 3: Pickle tool logo

- **Python**: It is chosen for its simplicity and readability, which are crucial for smooth web development. In our project, Python, along with Flask, manages tasks like processing user input, making predictions using our machine learning model, and delivering the results to the user's browser. Its flexibility and vast library support make it an ideal choice for creating web applications that can handle complex tasks efficiently.

Figure 4: Python Logo

Machine Learning:

- **pandas:** Pandas is like our data assistant. It helps us wrangle and manipulate our dataset, making it easier to work with. In our project, pandas load our heart attack dataset, splits it into features and target variables, and gets it ready for training.

Figure 5: Pandas Logo

- **scikit-learn:** Scikit-learn is like our toolbox for machine learning. It is packed with algorithms and tools to help us train and evaluate our model. In our project, scikit-learn splits our data into training and testing sets, trains our XGBoost classifier, and evaluates its accuracy.

Figure 6: Scikit-learn Logo

- **XGBoost:** XGBoost is our superstar machine learning model. It's powerful, efficient, and great at predicting outcomes. In our project, XGBoost learns from our data to make predictions about whether someone is at risk of a heart attack. It's the heart of our heart attack prediction system.

Figure 7: XGBoost Logo

- **numpy**: NumPy acts as the powerhouse for numerical computations in Python. It is like having a versatile toolkit for handling large datasets and performing mathematical operations effortlessly. In our project, NumPy plays a vital role in converting data into arrays, preprocessing data, and executing mathematical computations required by our machine learning algorithms. Its speed and user-friendly interface make it indispensable for data manipulation and analysis in machine learning projects.

Figure 8: Numpy Logo

In summary, our technology stack combines the power of Python, web technologies, Flask, XGBoost, and Pickle to create a sophisticated and user-friendly heart attack prediction system. Each component plays a vital role in different aspects of our project, from data processing and model training to interface design and deployment. Together, these technologies empower us to build a robust and scalable solution that empowers healthcare professionals and individuals to make informed decisions and take proactive steps towards preventing heart attacks.

Data Acquisition:

For this study, we obtained a comprehensive dataset tailored for heart attack prediction. The dataset includes a wide range of patient information, including medical history, lifestyle habits, and demographic details. The dataset was obtained from the Cleveland database and has been extensively curated to ensure data quality and relevance to cardiovascular risk assessment.

About the dataset:

1. ***Age***:
   - *Age denotes the chronological age of the individual under consideration.*
   - *It is included in our analysis due to its recognized association with an increased susceptibility to cardiovascular ailments with advancing age.*
2. ***Sex***:
   - *Sex refers to the biological classification of individuals as male or female.*
   - *Gender (Male: 1, Female: 0).*
3. ***Chest Pain Type*:**
   - *This attribute encapsulates the varied sensations experienced during episodes of chest discomfort, encompassing descriptors such as sharpness, dullness, or pressure.*
   - *This term can be categorized in 4 categories: 0: Typical angina, 1: Atypical angina, 2: Non-anginal pain, 3: Asymptomatic.*
4. ***Resting Blood Pressure***:
   - *Resting blood pressure represents the force exerted by circulating blood on arterial walls during periods of cardiac quiescence.*
   - *It is calculated in mm/HG.*
5. ***Serum Cholesterol***:
   - *Serum cholesterol levels quantify the concentration of lipid molecules traversing the bloodstream.*
   - *It is calculated in mg/dl.*
6. ***Fasting Blood Sugar***:
   - *Fasting blood sugar measurements ascertain the glucose concentration in the bloodstream subsequent to a period of abstinence from food intake.*

- *FBS: 120 mg/dl represents 1 in case of true and 0 as false.*

7. ***Resting Electrocardiographic Results:***
   - *Resting electrocardiographic examinations capture the electrical activity of the heart during periods of rest.*
   - *0: Normal, 1: Having ST - T Wave abnormality, 2: Showing probable or definite left ventricular hypertrophy by Este's criteria.*
8. ***Maximum Heart Rate Achieved:***
   - *Maximum heart rate denotes the highest pulse rate attained during vigorous physical exertion.*
9. ***Exercise Induced Angina:***
   - *Exercise induced angina refers to chest discomfort experienced during physical activity.*
   - *0 depicts NO, 1 depicts YES.*
10. ***Oldpeak (ST Depression):***
    - *Oldpeak, or ST depression, denotes the magnitude of deviation in the ST segment of an electrocardiogram during exercise compared to rest.*
11. ***Slope of Peak Exercise ST Segment:***
    - *The slope of the peak exercise ST segment characterizes the trajectory of ST segment changes during exercise.*
    - *0: Up sloping, 1: Flat, 2 Down sloping.*
12. ***Number of Major Vessels Coloured by Fluoroscopy (0-3):***
    - *This attribute quantifies the number of coronary arteries exhibiting pathological features upon fluoroscopic examination.*
    - *Increased vessel involvement correlates with heightened cardiovascular risk and disease progression.*
13. ***Thal:***
    - *Thal assessment entails myocardial perfusion imaging with thallium-201, facilitating the identification of areas of impaired blood flow.*
    - *0: NULL, 1: Normal blood flow, 2: Fixed defect (no blood flow in some part of the heart), 3: Reversible defect (a blood flow is observed but it is not normal)*
14. ***Target:***
    - *The target variable indicates the likelihood of experiencing a heart attack, categorized as either low or high risk.*
    - *0 means the patient is normal, 1 means the patient is suffering from heart disease.*

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
1	53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
2	70	1	0	145	174	0	1	125	1	2.6	0	0	3	0
3	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
4	62	0	0	138	294	1	1	106	0	1.9	1	3	2	0

Figure 9: Dataset Head

Just 14 of the 1025 traits in this database are used in the research that have been published. More specifically, up until now, ML researchers have solely used the Cleveland database. The patient's cardiac condition is indicated in the "target" field. Its value is an integer, with 0 denoting no or minimal probability of a heart attack and 1 denoting greater likelihood.

ML Model Used:

XGBoost (Extreme Gradient Boosting):

XGBoost, an influential gradient boosting algorithm, builds upon decision trees as its core framework. It functions by iteratively incorporating trees into an ensemble, each rectifying errors from preceding ones.

Training the Model: With the dataset prepared, we proceed to train the XGBoost model. During training, the model learns from the patterns and relationships present in the training data. It iteratively builds a set of decision trees, each attempting to correct the errors of the previous trees. This process continues until a predefined stopping criterion is met, such as reaching a maximum number of trees or achieving a certain level of performance.

XGBoost is like the superhero of machine learning algorithms. It's super popular because it's really good at making accurate predictions, and it's flexible enough to handle all sorts of different tasks.

Here's how it works: Imagine you have a bunch of weak learners, like little decision trees. Each one is pretty good at making predictions, but none of them are perfect. XGBoost takes these weak learners and combines their predictions in a smart way to create a super strong model.

What makes XGBoost so special is that it's really good at learning from its mistakes. It starts by making some initial predictions, then looks at where it went wrong and tries to fix it in the next round. It keeps doing this over and over again, gradually improving with each iteration until it gets really good at making predictions.

But XGBoost isn't just about making accurate predictions—it's also super fast and efficient. It's like the speedster of the machine learning world, able to train complex models on massive datasets in no time at all. Plus, it can handle missing data like a pro, so even if some of your data is incomplete, XGBoost can still make accurate predictions.

Another cool thing about XGBoost is that it's really flexible. You can use it for all sorts of different tasks, like classification, regression, and ranking. And with all its built-in tools for evaluating models and fine-tuning parameters, you can really customize it to fit your specific needs.

In short, XGBoost is like the ultimate sidekick for data scientists and machine learning enthusiasts. With its incredible accuracy, speed, and versatility, it's no wonder why it's become the go-to choose for so many predictive modeling tasks.

Procedure Followed:

Data Collection and Exploration

First, we gathered a comprehensive dataset containing various factors related to heart health, like age, sex, and cholesterol levels. This dataset formed the backbone of our project. We then delved into the data to understand its structure, distribution, and any potential connections between different variables. This exploration guided us in preparing the data for analysis.

Data Preprocessing

Next, we focused on getting the data ready for training our model. We took care of tasks like handling missing information, converting categorical variables into a format suitable for analysis, and ensuring that all numerical features were scaled appropriately for consistency across the dataset.

Model Selection

When it came to selecting the right model for our task, we opted for the XGBoost Classifier, considering its track record of performing well in similar scenarios. Since we were dealing with predicting whether someone is likely to have a heart attack or not, XGBoost's capabilities and resistance to overfitting made it a solid choice.

Training the Model

After dividing our prepared data into training and testing sets, we trained the XGBoost model on the training data. This involved the model gradually improving its predictions by analyzing patterns in the data through a series of decision trees.

```
▾                         XGBClassifier
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, criterion='squared_error', device=None,
              early_stopping_rounds=None, enable_categorical=False,
              eval_metric=None, feature_types=None, gamma=None,
              grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=0.1, loss='log_loss',
              max_bin=None, max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaf_nodes=10,
              max_leaves=None, min_child_weight=None, missing=nan,
              monotone_constraints=None, multi_strategy=None, n_estimators=100, ...)
```

Figure 10: XGBoost Model

Model Evaluation

Once trained, we rigorously tested the model's performance using the testing set, using metrics like accuracy, precision, recall, and F1-score to gauge how well it could predict heart attack risks for new data. This step helped us understand where the model excelled and where it might need adjustments.

```
Accuracy: 0.9853658536585366
```

We Achieved an accuracy of 98.53% by using the XGBoost which utilizes gradient boosting techniques to sequentially enhance model predictions, known for its efficiency and performance.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       102
           1       1.00      0.97      0.99       103

    accuracy                           0.99       205
   macro avg       0.99      0.99      0.99       205
weighted avg       0.99      0.99      0.99       205
```

Figure 11: Classification Report

Fine-Tuning Hyperparameters

To fine-tune the model further, we tweaked its hyperparameters—like the learning rate and tree depth using techniques such as grid search or random search. This optimization aimed to boost the model's accuracy and its ability to generalize well to new data.

```
# Initialize the XGBoost classifier
model = XGBClassifier( loss='log_loss', learning_rate=0.1, n_estimators=100,
                       criterion='squared_error', random_state=None, max_leaf_nodes=10)
```

Figure 12: Hyperparameters Used

Deployment and Integration

With a well-trained and validated model, we moved on to deployment. We integrated the model into a user-friendly web application using Flask, allowing individuals to input their medical information and receive instant predictions about their risk of having a heart attack. This ensured that our tool was accessible and practical for anyone concerned about their heart health.

Results:

Visualizing the data:

1. **Plotting % of people who are prone to heart attack and who are not.**

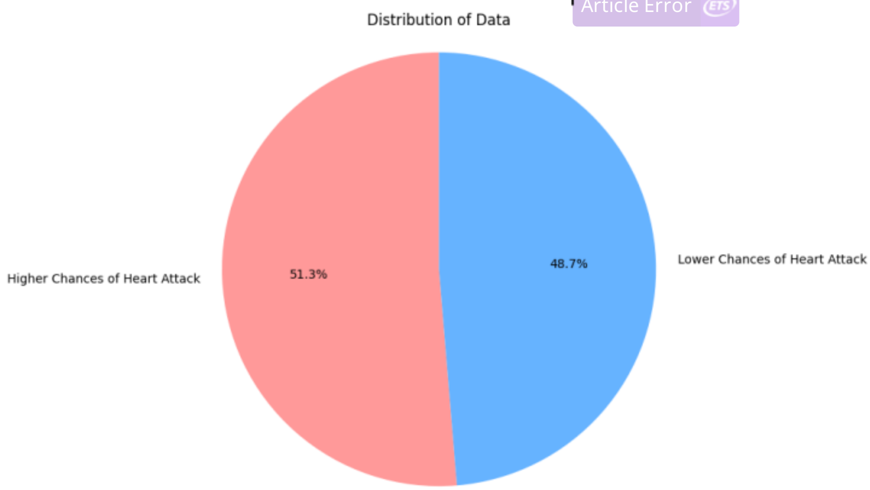

Figure 13: Heart Attack distribution

2. **Plotting % of people who are prone to heart attack above and under the age of 50.**

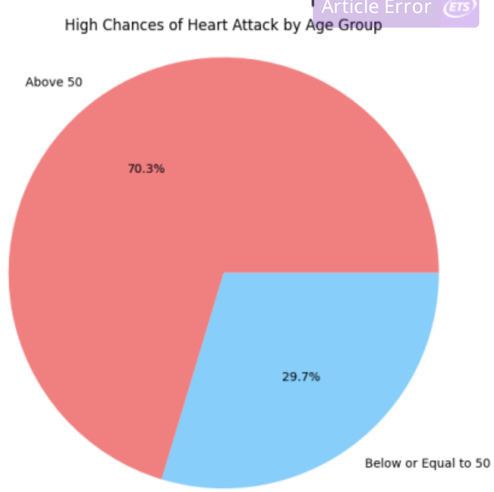

Figure 14: Heart attack distribution by age

3. Plotting the confusion matrix.

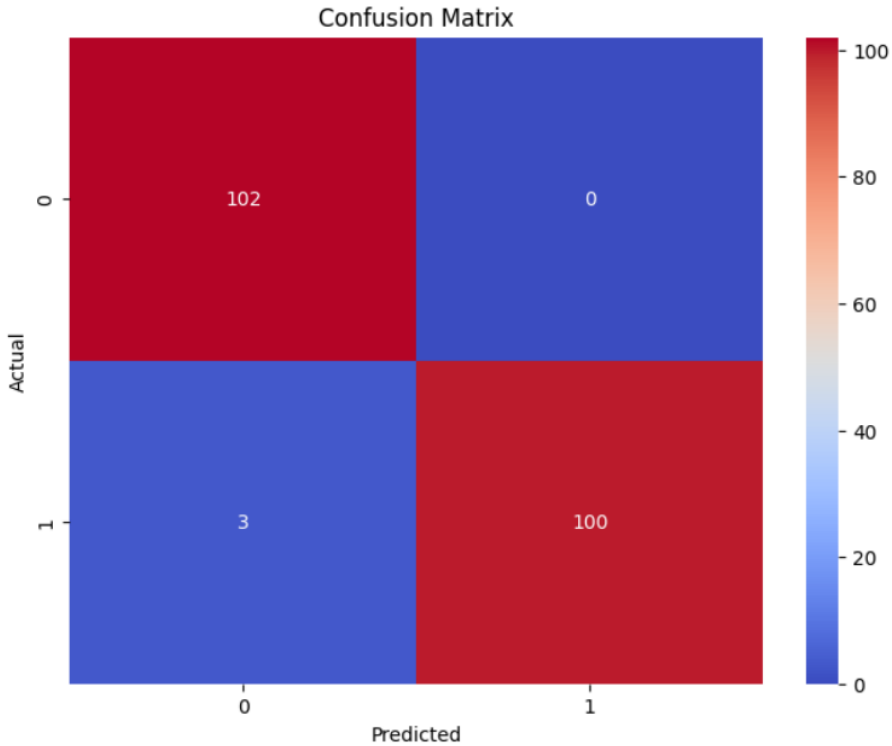

Figure 14: Confusion Matrix Plotting

4. Plotting the ROC Curve.

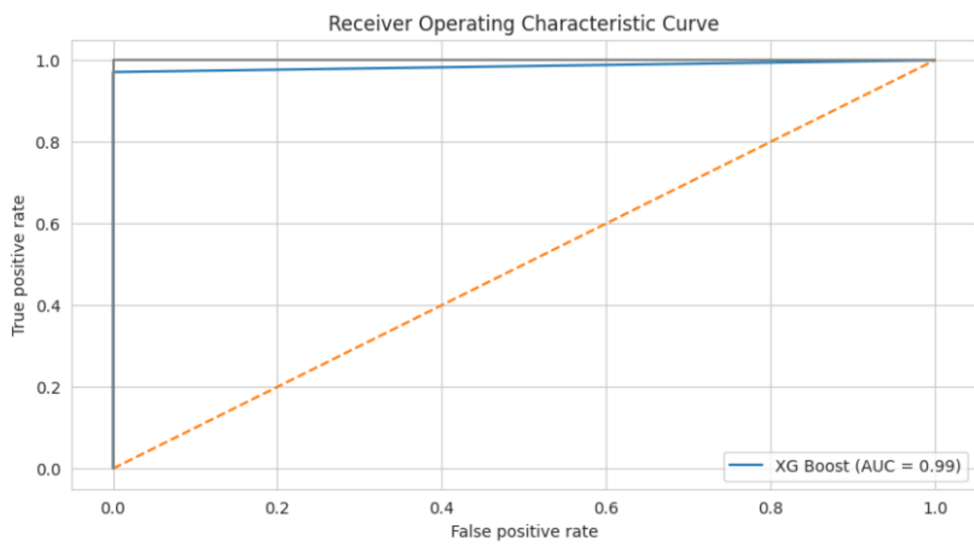

Figure 15: ROC curve

5. Web Page Outputs:

Input Screen:

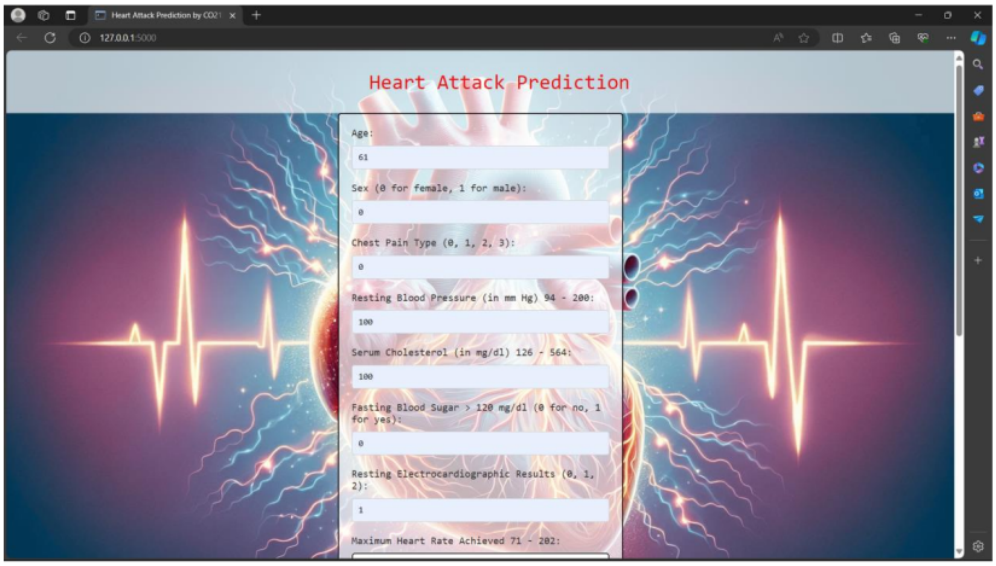

Figure 16: Input 1

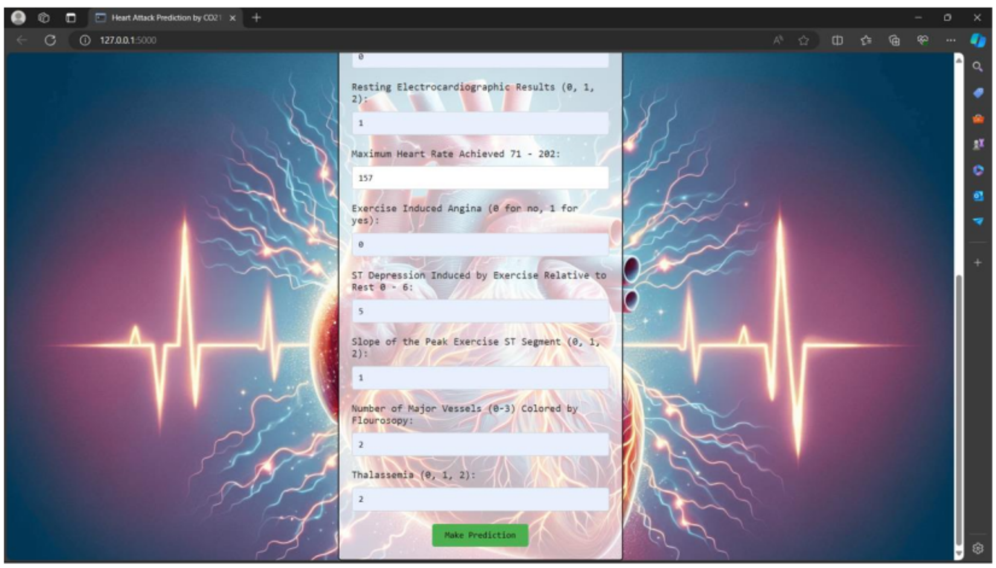

Figure 17: Input 2

Result Screen:

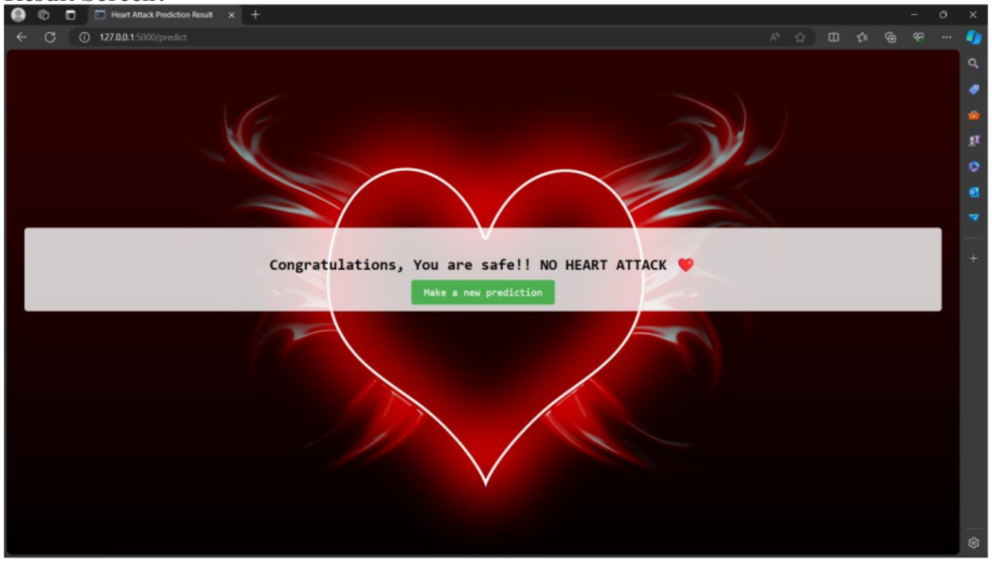

Figure 18: Output Screen

Appendix:

Link to Kaggle Notebook:

https://www.kaggle.com/code/kkaranismm/heart-attack-prediction

Link to code (Github):

https://github.com/KarandeepSinghCo21332/Heart_Attack_Prediction.git

Dataset Link:

https://github.com/KarandeepSinghCo21332/Heart_Attack_Prediction/blob/main/heart_dataset%20(1).csv

Heart Attack Prediction Project Report CO21332

ORIGINALITY REPORT

12%	8%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Georgetown University Student Paper	1%
2	Submitted to Chandigarh College of Engineering & Technology , CCET Student Paper	1%
3	Submitted to University of Greenwich Student Paper	1%
4	Submitted to Liverpool John Moores University Student Paper	1%
5	"Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020)", Springer Science and Business Media LLC, 2021 Publication	1%
6	Submitted to University of Abertay Dundee Student Paper	1%
7	Submitted to University of East London Student Paper	1%
	Submitted to University of North Texas	

8	Student Paper	1%
9	psecommunity.org Internet Source	1%
10	Submitted to University of Malaya Student Paper	1%
11	Submitted to Australian National University Student Paper	1%
12	Submitted to University of Sunderland Student Paper	1%
13	www.kluniversity.in Internet Source	<1%
14	Fadi Almohammed, Mohindra S. Thakur. "Forecasting compressive strength of concrete with basalt and polypropylene fiber by using ANN, RF and RT models", Asian Journal of Civil Engineering, 2023 Publication	<1%
15	www.ijirset.com Internet Source	<1%
16	www.imedpub.com Internet Source	<1%
17	"Proceedings of the 2nd International Conference on Big Data, IoT and Machine	<1%

Learning", Springer Science and Business Media LLC, 2024
Publication

Heart Attack Prediction Project Report CO21332

PAGE 1

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 2

PAGE 3

ETS **Article Error** You may need to use an article before this word.

ETS **Missing ","** Review the rules for using punctuation marks.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

ETS **Prep.** You may be using the wrong preposition.

ETS **Missing ","** Review the rules for using punctuation marks.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

ETS **Possessive** Review the rules for possessive nouns.

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 4

ETS **Sp.** This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Article Error You may need to use an article before this word.

Article Error You may need to use an article before this word.

Confused You have used either an imprecise word or an incorrect word.

PAGE 5

Proofread This part of the sentence contains an error or misspelling that makes your meaning unclear.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Confused You have used either an imprecise word or an incorrect word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 6

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 7

Article Error You may need to use an article before this word. Consider using the article **the**.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Missing "," Review the rules for using punctuation marks.

Run-on This sentence may be a run-on sentence.

PAGE 8

S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Missing "," Review the rules for using punctuation marks.

S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

Article Error You may need to use an article before this word.

Sp. This word is misspelled. Use a dictionary or spellchecker when you proofread your work.

PAGE 11

Article Error You may need to use an article before this word. Consider using the article **a**.

Article Error You may need to use an article before this word. Consider using the article **a**.

Article Error You may need to use an article before this word. Consider using the article **a**.

PAGE 12

Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 13

Wrong Form You may have used the wrong form of this word.

PAGE 14