Improving heart attack prediction

General metrics

26,679	3,684	242	14 min 44 sec	28 min 20 sec
characters	words	sentences	reading time	speaking time

Score

96

This text scores better than 96% of all texts checked by Grammarly

Writing Issues

67	9	58
Issues left	Critical	Advanced

Plagiarism

9 %

30 sources

9% of your text matches 30 sources on the web or in archives of academic publications

Writing Issues

43	**Clarity**
14	Wordy sentences
13	Passive voice misuse
12	Unclear sentences
1	Hard-to-read text
3	Intricate text
11	**Correctness**
7	Misspelled words
1	Text inconsistencies
2	Ungrammatical sentence
1	Incorrect citation format
6	**Engagement**
6	Word choice
7	**Delivery**
6	Inappropriate colloquialisms
1	Tone suggestions

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

26%

unique words

Rare Words

44%

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

rare words

Word Length

5.8

Measures average word length

characters per word

Sentence Length

15.2

Measures average sentence length

words per sentence

Improving heart attack prediction

Improving Heart Attack Prediction with Feature Selection and Stacking Ensemble Method

Sachin Girawale, Janhavi Kahare, Aneesh Kanhere

Under the Guidance of Dr. K . Rajeshwari

Pimpri Chinchwad College of Engineering, Pune -411044, India

Abstract: Heart attacks, or myocardial infarctions, remain a significant global health concern, contributing to substantial morbidity and mortality. Early detection and prevention are paramount in reducing the burden of cardiovascular diseases, necessitating the development of effective predictive models. This paper comprehensively explores machine learning techniques for heart attack prediction, aiming to enhance preventive healthcare strategies. Leveraging a diverse range of classifiers, including logistic regression, decision trees, random forests, support vector machines, and gradient boosting, we evaluate their performance in predicting heart attack occurrences. Through rigorous analysis and comparison, we highlight the potential of machine learning in accurately assessing an individual's likelihood of experiencing a heart attack based on demographic, lifestyle, and health-related factors. Additionally, we discuss the contribution of heart attack prediction to healthcare analytics, emphasizing its role in proactive management, personalized medicine, resource optimization, patient empowerment, and ongoing research efforts in cardiovascular medicine. Our findings underscore

the significance of machine learning in shifting healthcare delivery from reactive to proactive, ultimately leading to improved health outcomes and resource utilization.

Keywords: Heart attack prediction, machine learning, preventive healthcare, cardiovascular disease, healthcare analytics

1. INTRODUCTION

Heart attacks, or myocardial infarctions, represent acute and often life-threatening events within the spectrum of cardiovascular diseases. Globally, heart attacks stand as a leading cause of mortality, underscoring the critical need for accurate risk prediction and early intervention strategies [1]. Despite advancements in medical science, heart attacks pose significant challenges, necessitating innovative approaches for timely detection and preventive measures.

Coronary artery disease, characterized by the narrowing or blockage of coronary arteries, serves as the primary precursor to heart attacks, making it imperative to identify individuals at heightened risk [2]. The ability to predict an individual's susceptibility to a heart attack not only facilitates proactive management but also offers opportunities to mitigate the disease's devastating consequences.

In recent years, the integration of machine learning techniques has emerged as a promising avenue for enhancing heart attack risk prediction. Machine learning algorithms, fueled by vast datasets and computational power, hold the potential to discern complex patterns and relationships within heterogeneous patient data. Leveraging this technology, researchers endeavor to develop robust predictive models capable of accurately assessing an individual's likelihood of experiencing a heart attack.

The motivation to explore heart attack prediction transcends mere academic inquiry in this context. It represents a proactive stance in combating a formidable public health challenge, aiming to empower individuals, healthcare providers, and policymakers with actionable insights for preventive healthcare. By harnessing the predictive power of machine learning, we aspire to shift the paradigm from reactive to proactive healthcare delivery, ultimately saving lives and alleviating the burden on healthcare systems.

This study embarks on a journey to explore the efficacy of machine learning models in predicting heart attacks, leveraging comprehensive patient data encompassing demographic, lifestyle, and health-related factors. Through rigorous analysis and evaluation, we seek to delineate the landscape of heart attack prediction, shedding light on the potential of machine learning to revolutionize cardiovascular risk assessment and management.

Objectives

The objective of this research paper is to explore the role of heart attack prediction models in enhancing preventive healthcare strategies by investigating their effectiveness in identifying individuals at high risk of heart attacks, facilitating early intervention, promoting personalized medicine, optimizing resource allocation, empowering patients through education, and driving ongoing research and development efforts in cardiovascular medicine.

Contribution to Healthcare Analytics

Predicting heart attacks is a critical application of healthcare analytics that can save lives by identifying high-risk individuals before a cardiac event occurs. Here is an explanation of how heart attack prediction contributes to healthcare analytics:

1. Data Collection and Integration: Healthcare analytics relies on collecting and integrating various data types, including patient demographics, medical history, lifestyle factors, genetic information, and diagnostic test results. This data can come from electronic health records (EHRs), wearable devices, imaging tests, and patient-reported outcomes.

2. Feature Selection and Engineering: Once the data is collected, healthcare analysts and data scientists identify relevant features or variables that could predict a heart attack. These features may include age, gender, blood pressure, cholesterol levels, smoking status, family history of heart disease, and exercise habits. Feature engineering techniques may also create new variables or transform existing ones to improve predictive performance.

3. Model Development: Healthcare analytics involves building predictive models that can learn from historical data to predict future events, such as heart attacks accurately. Standard machine learning algorithms used[1] for heart attack prediction include logistic regression, decision trees, random forests, support vector machines, and neural networks. These models are trained[2] on labeled data, where the outcome (i.e., whether a patient had a heart attack) is known.

Overall, heart attack prediction is a valuable application of healthcare analytics. By leveraging data-driven approaches, healthcare providers can proactively identify and intervene with high-risk patients, ultimately[3] leading to better health outcomes and a more efficient healthcare system.

2. LITERATURE REVIEW

Paper 1: A Review: Heart Disease Prediction in Machine Learning & Deep Learning

The paper emphasizes the increasing prevalence of heart disease globally and highlights the importance of early diagnosis and prevention. It discusses the potential of machine learning and deep learning in predicting heart disease, noting variations in accuracy among different algorithms. The study concludes that deep learning outperforms machine learning in healthcare analytics, offering consistency and high accuracy in predicting heart disease.

Paper 2: Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques

This study addresses the significance of early detection in reducing mortality rates due to heart disease. It proposes a machine-learning model for predicting heart disease at an early stage, employing feature selection techniques and various machine-learning algorithms. The research finds that random forest achieves the most optimistic performance in predicting heart disease using selected feature subsets. The study suggests the potential clinical use of the proposed model for early-stage heart disease prediction with low cost and time.

Paper 3: Heart disease prediction using machine learning algorithms

The paper highlights the increasing cases of heart disease and the importance of predicting such diseases beforehand[4]. It presents a heart disease prediction system utilizing machine learning algorithms such as logistic regression and KNN. The research demonstrates the effectiveness of the proposed model in accurately predicting the likelihood of heart disease in individuals based on their medical history. The study concludes that the developed model enhances medical care, reduces costs, and provides significant knowledge for predicting heart diseases.

Comparison:

All three papers focus on the use of[5] machine-learning techniques for heart disease prediction. They highlight the importance of early detection and emphasize the potential of machine learning models in improving accuracy and efficiency. While Paper 1 and Paper 3 primarily discuss the effectiveness of machine learning algorithms, Paper 2 specifically addresses feature selection techniques and compares the performance of different machine learning models. Overall, these studies contribute valuable insights into the application of healthcare analytics in predicting heart disease and improving patient outcomes.

3. METHODOLOGY

The methodology section outlines the systematic approach and procedures employed to achieve the objectives of the study[6]. It provides a detailed description of[7] the research design, data collection methods, analytical techniques, and any tools or instruments utilized. This section serves as a roadmap for replicating the study and ensures the validity and reliability of the findings.

Figure 1: Flowchart describing the overall methodology process

A. Dataset Description

The heart attack prediction dataset contains 1025 entries and 14 columns, representing various attributes related to heart health and risk factors for heart attacks. The 'age' column indicates the age of the patient[8], while the 'sex' column represents the gender (0 = female, 1 = male). The 'cp' column categorizes chest pain types (0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic). 'trestbps'[9] stands for resting blood pressure, 'chol' for cholesterol level, and 'fbs'[10] for fasting blood sugar (>120 mg/dl). The 'restecg'[11] column indicates resting electrocardiographic results (0 = Normal, 1 = Abnormality, 2 = Probable or definite left ventricular hypertrophy), and 'thalach'[1] represents the maximum heart rate achieved. 'exang'[13] denotes exercise-induced angina (1 = Yes, 0 = No), while[14] 'oldpeak'[15] indicates ST[16] depression induced by exercise relative to rest. The 'slope' column represents the slope of the peak exercise ST[16] segment (0 = Upsloping, 1 = Flat, 2 = Downsloping), and 'ca' denotes the number of major[17] vessels colored by fluoroscopy. 'thal' represents the thallium stress test result (0 = Normal, 1 = Fixed defect, 2 = Reversible defect). The 'target' column is the target variable indicating the presence of a heart attack (1 = Yes, 0 = No). The dataset contains no missing values, with all columns having 1025 non-null entries.

B. Data Preprocessing:

Data preprocessing is a[18] crucial step[18] in preparing a dataset for analysis or machine learning. In the initial stages,[19] the dataset is checked for null values and duplicate rows, ensuring its integrity. Following this, an augmentation process is employed to enhance the dataset's diversity and, subsequently, the machine learning model's performance. This[20] involves randomly selecting rows without duplicates and adding them back to the dataset, maintaining its original size.

Identifying correlations

The correlation matrix reveals relationships between features in the heart attack prediction dataset. Key observations include a negative correlation between age and maximum heart rate (-0.41), indicating that as age increases,[21] the maximum heart rate tends to decrease. There's[22] a moderate positive correlation (0.43) between chest pain type and the target variable, suggesting certain chest pain types may be more indicative of a heart attack. Additionally, the number of major[23] vessels colored by fluoroscopy shows a moderate negative correlation (-0.45) with the target, implying that a higher number of vessels might be linked[24] to a lower likelihood of a heart attack. The presence of exercise-induced angina also shows a moderate negative correlation (-0.45) with the target, indicating a potential association with a lower risk of heart attack. These insights can inform feature selection and aid in understanding patterns in the dataset.

Figure 2:Heatmap showing a correlation between features:

Feature Selection

The SelectKBest method from the sci-kit-learn library is used[25] for feature selection based on the f_classif score function, which computes the ANOVA F-value for the features. This technique ranks the features by their importance in predicting the target variable.

Figure 3:The[26] bar chart shows the feature selection scores, indicating the relative importance of each feature. The higher the score, the more important[27] the feature is considered[28] for the model.

In this specific[29] case, the top 5 selected features are 'exang',[30] 'thalach',[30] 'ca',[30] 'oldpeak',[30] and 'cp'.[30] These features are deemed most relevant for predicting the target variable based on their ANOVA F-values. For instance, 'exang'[31] represents exercise-induced angina, 'thalach'[31] indicates the maximum heart rate achieved during exercise, 'ca' stands for the number of major[32] vessels colored by fluoroscopy, 'oldpeak'[31] is the ST[16] depression induced by exercise relative to rest, and 'cp' denotes the chest pain type. These features are essential in determining the likelihood of a heart attack, making them key[33] factors in the predictive model.

E. Building the Model

Several machine learning algorithms were evaluated[34] for their ability to predict heart attack risk based on the selected features. The following algorithms were considered[35]: Logistic Regression, Decision Tree, Random Forest.

1. Random Forest Classifier: This Classifier constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees[36]. It's[37] an ensemble learning method known for its robustness and accuracy.

2. Logistic Regression: Despite its name, logistic regression is a linear model for binary classification that predicts the probability of occurrence of an event by fitting data to a logistic curve.

3.SVM(Support Vector Machine)

Support Vector Machine (SVM) is a classification algorithm that can be used[38] to predict heart attacks. It works by finding the optimal hyperplane to separate patients into different[39] risk categories based on features like[39] age, sex, cholesterol levels, and blood pressure. It's[40] trained on preprocessed data and evaluated using metrics like[41] accuracy, precision, recall, and F1-score. Overall, SVM is a powerful tool for accurately classifying patients' risk of heart attacks.

4. Gradient Boosting

Gradient Boosting is an ensemble learning technique that builds a strong[42] predictive model by combining multiple weak models sequentially. In the context of heart attack prediction, Gradient Boosting can be used to create a predictive model.

5. Naive Bayes

Naive Bayes is a simple yet effective machine learning algorithm based on Bayes' theorem with the "naive" assumption of feature independence[43]. It works well for large datasets with many features, making it computationally efficient and scalable. Despite its simplicity, Naive Bayes often performs well in practice, particularly in text classification and spam filtering tasks. However, its performance may suffer when the assumption of feature independence is[44] violated[45].

6. Multi-layer Perceptron (Neural Network)

MLPClassifier represents a feedforward artificial neural network composed of multiple layers of nodes, allowing for nonlinear relationships between inputs and outputs. It can model complex patterns but requires careful tuning of hyperparameters and may be prone to overfitting.

7. K-Nearest Neighbors (KNN)

KNN is a non-parametric lazy learning algorithm that classifies instances based on their[46] similarity to nearby instances in the feature space. It is simple to implement and effective in practice but may suffer from the curse of dimensionality in high-dimensional spaces.

Each Classifier has its strengths and weaknesses, making them suitable for different types of data[47] and modeling objectives.

F.Hybrid Models

Hybrid models, such as the Voting Classifier and Stacking Classifier, were implemented to improve the overall predictive performance of the heart attack prediction model. These models combine the strengths of multiple base models, each trained on the same dataset, to produce a more accurate and robust prediction.

i.Voting Classifier

The Voting Classifier combines the predictions of multiple base models to improve overall performance. In this case, the Voting Classifier uses Logistic Regression and Gradient Boosting as base models with soft voting. Soft voting considers the weighted average probability of each class predicted by the individual classifiers, providing more reliable predictions.

The Voting Classifier achieved a training accuracy of approximately 90.52% and a validation accuracy of about 86.36%. It performed well on the validation set, with a balanced precision and recall for both classes. This[48] indicates that the model generalizes well to unseen data and does not overfit.

On the test set,[49] the Voting Classifier achieved an accuracy of around 88.96%. It demonstrated consistent performance across all metrics, with high precision and recall for both classes. This[50] suggests that the model is reliable and can effectively classify instances into the correct classes.

Overall, the Voting Classifier shows promise as an effective ensemble method for this heart attack prediction task, offering a balance between simplicity and performance.

Table 1: Training and testing accuracy after Voting Classifier (LR+GB).

Training Accuracy

90.51%

Testing Accuracy

88.96%

Figure 4: Confusion matrix for Voting Classifier

ii. Stacking Classifier

The Stacking Classifier combines the predictions of multiple base models (Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine)[51] to make final predictions. It uses a Logistic Regression meta-model to learn how to best combine the predictions of the base models[52].

On the validation set, the Stacking Classifier achieved an accuracy of approximately 95.45%. It demonstrated high precision and recall for both classes, indicating its ability to effectively classify instances into the correct classes[53].

On the test set, the Stacking Classifier achieved an even higher accuracy of around 97.40%. It maintained high precision and recall for both classes, suggesting that the model is robust and generalizes well to unseen data.

Overall, the Stacking Classifier shows significant promise as an ensemble method for this heart attack prediction task, outperforming individual base models and providing a reliable means of predicting the likelihood of a heart attack.

Table 2: Training and testing accuracy after Stacking Classifier (LR+GB+RF+SVM).

.

Training Accuracy

95.45%

Testing Accuracy

97.40%

Figure 5:Confusion Matrix for Stacking Classifier

4. RESULT

In the conducted research, a variety of[54] classifiers were employed[55] to predict heart disease outcomes using a dataset sourced from relevant medical records. The classifiers utilized encompass a broad spectrum of machine learning algorithms, including Random Forest, Gradient Boosting, Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, Multi-layer Perceptron (Neural Network) and K-Nearest Neighbors (KNN). These classifiers were chosen[56] for their distinct methodologies and capabilities in handling classification tasks. Through rigorous experimentation and evaluation, it was observed[57] that each Classifier exhibited varying levels of accuracy in predicting heart disease outcomes. The results of this study provide valuable insights into the performance of different classification techniques in the context of cardiovascular health prediction. Such findings contribute to the advancement of medical research by offering practitioners a deeper understanding of the efficacy of machine learning models in aiding clinical decision-making processes.

The training process involved fitting each model to the training dataset, allowing it to learn the underlying patterns and relationships between input features and the target variable.

Following training, the models were evaluated[58] using both[59] the training and testing datasets to measure their performance accurately. Evaluation metrics accuracy was utilized[60] to provide a comprehensive assessment of[61] each model's predictive capabilities. The results of the evaluation are as follows:

Table 3: Training and testing accuracies of all applied classifiers.

ML Classifier

Training Accuracy

Testing Accuracy

Random Forest

96.10%

98.05%

Logistic Regression

81.11%

82.46%

Naïve Bayes

79.80%

79.22%

Support Vector Machine

85.71%

88.96%

K-Nearest Neighbours

88.33%

87.66%

Neural Network

90.20%

90.90%

Gradient Boosting

91.55%

91.55%

Voting Classifier

(LR+GB)

98.2%

94.16%

Proposed Approach

Stacking Classifier

(LR+GB+RF+SVM)

95.45%

97.40%

Figure 6: Comparison of accuracies between classifiers

These results provide valuable insights into the performance of the proposed model and other classifiers, highlighting their strengths and weaknesses in predicting obesity levels. Overall, the training and evaluation process underscores the potential of machine learning algorithms in addressing public health challenges through data-driven approaches.

These findings offer valuable insights into the performance of both the proposed model and alternative classifiers[62], illuminating their respective capabilities and limitations. Through the training and evaluation process, it becomes evident that machine learning algorithms possess significant potential in tackling public health challenges by leveraging data-driven methodologies.

5. FUTURE SCOPE

The heart attack prediction task has demonstrated the effectiveness of various machine learning models and ensemble techniques. To[63] further enhance the predictive accuracy and applicability of the models, several future directions can be considered[63,64]. Firstly, incorporating more advanced deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), could potentially[65] capture complex patterns in the data and improve prediction performance. Additionally, integrating more diverse and comprehensive features, such as genetic data or lifestyle factors, could provide a more holistic understanding of heart attack risk. Furthermore, exploring the use of explainable AI[16] techniques to interpret model predictions could enhance the trust and adoption of these models in clinical practice. Overall, continued research and development in these areas hold the potential to significantly advance the field of heart attack prediction and improve patient outcomes in the future.

6. CONCLUSION

The heart attack prediction task utilized various machine learning models, including Random Forest, Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Neural Network, and Gradient Boosting. A Voting Classifier combining Logistic Regression and Gradient Boosting, as well as[66] a Stacking Classifier with LR[16], GB[16], RF[16], and SVM,[66] were also employed. Among the individual models, Random Forest achieved the highest training accuracy of 96.10%, while the Voting Classifier (LR+GB) achieved the highest testing accuracy of 94.16%. However, on the testing set, Random Forest performed the best, with an accuracy of 98.05%.

The Stacking Classifier, combining Logistic Regression, Gradient Boosting, Random Forest, and Support Vector Machine, achieved a training accuracy of 95.45% and a testing accuracy of 97.40%, demonstrating strong generalization to unseen data. In conclusion, combining traditional machine learning models and ensemble techniques led to a competitive performance in heart attack prediction, showing promise for real-world applications in healthcare.

7. REFERENCES

[1] Md. I. [67] Hossain et al., "Heart disease prediction using distinct artificial intelligence techniques: performance analysis and comparison," Iran Journal of Computer Science, vol. 6, no. 4, pp. 397–417, Dec. 2023, doi: 10.1007/s42044-023-00148-7.

[2] A. Yazdani, K. D. Varathan, Y. K. Chiam, A. W. Malik, and W. A. Wan Ahmad, "A novel approach for heart disease prediction using strength scores with significant predictors," BMC Med Inform Decis Mak, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01527-5.

[3] R. Indrakumari, T. Poongodi, and S. R. Jena, "Heart Disease Prediction using Exploratory Data Analysis," in Procedia Computer Science, Elsevier B.V. [16], 2020, pp. 130–139. doi: 10.1016/j.procs.2020.06.017.

[4] C. A. ul Hassan et al., "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers," Sensors, vol. 22, no. 19, Oct. 2022, doi: 10.3390/s22197227.

[5] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms, vol. 16, no. 2, Feb. 2023, doi: 10.3390/a16020088.

[6] M. S. Raja, M. Anurag, C. P. Reddy, and N. R. Sirisala, "Machine Learning Based Heart Disease Prediction System," in 2021 International Conference on Computer Communication and Informatics, ICCCI 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402653.

[7] Noorul Islam Centre for Higher Education. Department of Electrical and Electronics Engineering, IEEE Electron Devices Society. India Chapter, and Institute of Electrical and Electronics Engineers, Proceedings of IEEE International Conference on Circuits, Power and Computing Technologies : ICCPCT-2016 on 18th & 19th March 2016.

[8] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in IOP Conference Series: Materials Science and Engineering, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1757-899X/1022/1/012072.

[9] W. A. W. A. Bakar, N. L. N. B. Josdi, M. B. Man, and M. A. B. Zuhairi, "A Review: Heart Disease Prediction in Machine Learning & Deep Learning," in 2023 19th IEEE International Colloquium on Signal Processing and Its Applications, CSPA 2023 - Conference Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 150–155. doi: 10.1109/CSPA57446.2023.10087837.

[10] N. Biswas et al., "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," Biomed Res

Int, vol. 2023, 2023, doi: 10.1155/2023/6864343.

1.	~~used~~	Wordy sentences	Clarity
2.	*are trained*	Passive voice misuse	Clarity
3.	~~ultimately~~	Wordy sentences	Clarity
4.	~~beforehand~~	Wordy sentences	Clarity
5.	~~the use of~~ → using	Wordy sentences	Clarity
6.	study's objectives	Wordy sentences	Clarity
7.	*It provides a detailed description of the research design, data collection methods, analytical techniques, and any tools or instruments utilized.*	Unclear sentences	Clarity
8.	patient's age	Wordy sentences	Clarity
9.	~~trestbps~~ → treetops	Misspelled words	Correctness
10.	~~fbs~~ → FBS	Misspelled words	Correctness
11.	~~restecg~~ → resting	Misspelled words	Correctness
12.	~~thalach~~ → thali	Misspelled words	Correctness
13.	~~exang~~ → exacting, sang	Misspelled words	Correctness
14.	~~, while~~ → . In contrast	Hard-to-read text	Clarity
15.	~~oldpeak~~ → old peak	Misspelled words	Correctness
16.	*ST; AI; LR; GB; RF; B.V.*	Text inconsistencies	Correctness
17.	~~major~~ → significant	Word choice	Engagement
18.	*Data preprocessing is a crucial step in preparing a dataset for analysis or machine learning.*	Unclear sentences	Clarity

19.	*In the initial stages, the dataset is checked for null values and duplicate rows, ensuring its integrity.*	Unclear sentences	Clarity
20.	*This*	Intricate text	Clarity
21.	*Key observations include a negative correlation between age and maximum heart rate (-0.41), indicating that as age increases, the maximum heart rate tends to decrease.*	Unclear sentences	Clarity
22.	~~There's~~ → There is	Inappropriate colloquialisms	Delivery
23.	~~major~~ → significant	Word choice	Engagement
24.	*be linked*	Passive voice misuse	Clarity
25.	*is used*	Passive voice misuse	Clarity
26.	~~3:The~~ → 3 The	Misspelled words	Correctness
27.	~~more important~~ → more influential	Word choice	Engagement
28.	~~considered~~	Wordy sentences	Clarity
29.	~~specific~~	Wordy sentences	Clarity
30.	*In this specific case, the top 5 selected features are 'exang', 'thalach', 'ca', 'oldpeak', and 'cp'.*	Ungrammatical sentence	Correctness
31.	*For instance, 'exang' represents exercise-induced angina, 'thalach' indicates the maximum heart rate achieved during exercise, 'ca' stands for the number of major vessels colored by fluoroscopy, 'oldpeak' is the ST depression induced by exercise relative to rest, and 'cp' denotes the chest pain typ...*	Ungrammatical sentence	Correctness

32.	~~major~~ → significant	Word choice	Engagement
33.	~~key~~ → critical	Word choice	Engagement
34.	*were evaluated*	Passive voice misuse	Clarity
35.	*were considered*	Passive voice misuse	Clarity
36.	*1. Random Forest Classifier: This Classifier constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees.*	Unclear sentences	Clarity
37.	~~It's~~ → It is	Inappropriate colloquialisms	Delivery
38.	*be used*	Passive voice misuse	Clarity
39.	*It works by finding the optimal hyperplane to separate patients into different risk categories based on features like age, sex, cholesterol levels, and blood pressure.*	Unclear sentences	Clarity
40.	~~It's~~ → It is, It has	Inappropriate colloquialisms	Delivery
41.	*It's trained on preprocessed data and evaluated using metrics like accuracy, precision, recall, and F1-score.*	Unclear sentences	Clarity
42.	~~strong~~ → robust	Word choice	Engagement
43.	feature independence assumption	Wordy sentences	Clarity
44.	feature independence assumption	Wordy sentences	Clarity
45.	*is violated*	Passive voice misuse	Clarity
46.	~~their~~	Wordy sentences	Clarity

47.	~~types of data~~ → data types	Wordy sentences	Clarity
48.	*This*	Intricate text	Clarity
49.	*On the test set, the Voting Classifier achieved an accuracy of around 88.96%.*	Unclear sentences	Clarity
50.	*This*	Intricate text	Clarity
51.		Incorrect citation format	Correctness
52.	to combine the predictions of the base models best	Inappropriate colloquialisms	Delivery
53.	to classify instances into the correct classes effectively	Inappropriate colloquialisms	Delivery
54.	~~a variety of~~ → various	Wordy sentences	Clarity
55.	*were employed*	Passive voice misuse	Clarity
56.	*were chosen*	Passive voice misuse	Clarity
57.	*was observed*	Passive voice misuse	Clarity
58.	*were evaluated*	Passive voice misuse	Clarity
59.	~~both~~	Wordy sentences	Clarity
60.	*was utilized*	Passive voice misuse	Clarity
61.	*Evaluation metrics accuracy was utilized to provide a comprehensive assessment of each model's predictive capabilities.*	Unclear sentences	Clarity

62.	*These findings offer valuable insights into the performance of both the proposed model and alternative classifiers, illuminating their respective capabilities and limitations.*	Unclear sentences	Clarity
63.	*To further enhance the predictive accuracy and applicability of the models, several future directions can be considered.*	Unclear sentences	Clarity
64.	*be considered*	Passive voice misuse	Clarity
65.		Tone suggestions	Delivery
66.	*A Voting Classifier combining Logistic Regression and Gradient Boosting, as well as a Stacking Classifier with LR, GB, RF, and SVM, were also employed.*	Unclear sentences	Clarity
67.	*I*	Inappropriate colloquialisms	Delivery
68.	*logistic regression, decision trees, random forests, support vector machines, and*	User abnormal behavior recommendation via multilayer network	Originality
69.	*ultimately leading to improved health outcomes and resource utilization.*	Patients' views on incidental findings from clinical exome sequencing	Originality
70.	*Coronary artery disease, characterized by the narrowing or blockage of coronary arteries,*	What are the common diseases at age 60 - Extra Large As Life | General Blog https://extralargeaslife.com/what-are-the-common-diseases-at-age-60/	Originality
71.	*The ability to predict an individual's susceptibility to*	Deploying next-generation sequencing in a hospital setting	Originality

72.	*machine learning techniques has emerged as a promising avenue for*	Editorial: Spectroscopy, imaging and machine learning for crop stress	Originality
73.	*By harnessing the predictive power of machine learning,*	Decision Intelligence – this is how most businesses will adopt AI - TechHQ https://techhq.com/2022/07/ai-artificial-decision-intelligence-machine-learning-predictive-data-analytics-business/	Originality
74.	*the efficacy of machine learning models in predicting*	Assessing and Validating the Ability of Machine Learning to Handle Unrefined Particle Air Pollution Mobile Monitoring Data Randomly, Spatially, and Spatiotemporally	Originality
75.	*effectiveness in identifying individuals at high risk of*	Incidence and predictors of neck and widespread pain after motor vehicle collision among US litigants and nonlitigants	Originality
76.	*This data can come from electronic health records*	Ehsan Ghanbari - Big Data Analytics in Healthcare https://ehsanghanbari.com/post/blog/big-data-analytics-in-healthcare	Originality
77.	*leading to better health outcomes and a more efficient healthcare system.*	PCNOK Healthcare Network : Improving Patient Care https://smartkela.com/revolutionizing-healthcare-exploring-the-benefits-of-pcnok/	Originality
78.	*machine learning algorithms such as logistic regression and*	https://ijaecs.iraj.in/abstract.php?paper_id=12530	Originality
79.	*Overall, these studies contribute valuable insights into the*	재생에너지 확대가 탄소배출 및 지역경제에 미치는 영향 연구	Originality

80.	*and procedures employed to achieve the objectives of the study.*	Financial inclusion of rural smallholder farmers in Nigeria : measurement issues, impact on livelihood and implications for policy interventions	Originality
81.	*It provides a detailed description of the research design, data collection*	A strategy analysis of HQCC, a competency center in SLB	Originality
82.	*Data Preprocessing: Data preprocessing is a crucial step in preparing*	Speech funding https://speechdat.org/category/speech-funding/	Originality
83.	*a potential association with a lower risk of*	High flow nasal oxygen therapy to avoid invasive mechanical ventilation in SARS-CoV-2 pneumonia: a retrospective study | Annals of Intensive Care | Full Text https://annalsofintensivecare.springeropen.com/articles/10.1186/s13613-021-00825-5	Originality
84.	*the mode of the classes (classification) or the mean prediction (regression) of the individual trees.*	Adaptive Machine Learning Based Distributed Denial-of-Services Attacks Detection and Mitigation System for SDN-Enabled IoT	Originality
85.	*is a linear model for binary classification that*	Feature Analysis of Smart Shoe Sensors for Classification of Gait Patterns	Originality
86.	*predicts the probability of occurrence of an event by fitting data to a logistic*	A General Framework Based on Machine Learning for Algorithm Selection in Constraint Satisfaction Problems	Originality
87.	*Support Vector Machine) Support Vector Machine (SVM) is a classification algorithm*	Intelligent Identification of Cavitation State of Centrifugal Pump Based on Support Vector Machine	Originality

88.	*Gradient Boosting Gradient Boosting is an ensemble learning technique that*	Numerical Simulation and Design of Ensemble Learning Based Improved Software Development Effort Estimation System	Originality
89.	*to learn how to best combine the predictions of the base models.*	Ensemble Methods in Supervised Learning: Review towards an application in a model for predictions about Ecology	Originality
90.	*These results provide valuable insights into the performance of the*	How Gore-Tex Test Products Before Certifying Them https://totalhiker.com/gore-tex-certification/	Originality
91.	*These findings offer valuable insights into the performance of*	Survival, growth, and leaf area index of annual Moringa cuttings (Moringa oleifera) in different diameter classes	Originality
92.	*Random Forest, and Support Vector Machine, achieved a*	Computer Vision and Machine Learning Based Grape Fruit Cluster Detection and Yield Estimation Robot	Originality
93.	*A novel approach for heart disease prediction using strength scores with significant predictors," BMC Med Inform Decis Mak,*	Modified Self-Adaptive Bayesian Algorithm for Smart Heart Disease Prediction in IoT System	Originality
94.	*Analysis," in Procedia Computer Science, Elsevier B.V., 2020, pp.*	Improving fake news classification using dependency grammar	Originality
95.	*C. A. ul Hassan et al., "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers," Sensors, vol. 22, no. 19,*	Machine Learning for Cardiovascular Disease Risk Assessment: A Systematic Review	Originality

96.	*H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms,*	Data Mining for Heart Disease Prediction Based on Echocardiogram and Electrocardiogram Data | Jurnal Online Informatika https://join.if.uinsgd.ac.id/index.php/join/article/view/1027	Originality
97.	*Institute of Electrical and Electronics Engineers Inc., 2023, pp.*	Adaptive Neuro-Fuzzy Inference System-Based GPS-IMU Data Correction for Capacitive Resistivity Underground Imaging with Towed Vehicle System	Originality