An Analysis of Machine Learning Algorithms for Early Prediction of Heart Attack in Stroke Patients

C. Naga Jyothi

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Predicting heart attacks early in stroke patients through data analysis is crucial to reducing the high mortality rate associated with these conditions. However, accurately predicting heart attacks in stroke patient data poses a challenge. Early detection of stroke-related diseases is beneficial for prevention or early intervention. Machine learning and data mining play important roles in stroke prediction. This paper proposes an effective method for identifying stroke and compares the performance of three machine learning algorithms: Decision Tree, Naïve Bayes and K-Nearest Neighbors. The study analyzes the performance of these algorithms on a stroke dataset. The preliminary results demonstrate that the Decision Tree algorithm achieves the highest accuracy of 95.75%, outperforming the Naïve Bayes and K-Nearest Neighbors algorithms.*

I. INTRODUCTION

A stroke refers to the interruption of blood supply to the brain caused by the rupture of one or more small blood vessels. This interruption deprives the brain of essential oxygen and nutrients, leading to neuronal death. Ischemic stroke, the most common type of stroke, occurs when a blood vessel supplying the brain becomes blocked by a blood clot, often due to factors such as smoking and high cholesterol [1][5]. While many people believe that strokes primarily affect older individuals due to their higher risk of heart disease and diabetes, strokes are caused by blockages in the brain's blood vessels that carry oxygen and nutrients. This arterial dysfunction can be caused by various factors, including smoking, diabetes, high blood pressure, cholesterol, and heart disease, with variations based on age and gender.

The clinical symptoms of stroke include sudden collapse, mental coma, impaired speech, and hemiplegia [6]. A heart attack is myocardial decay caused by acute and persistent ischemia and hypoxia of the coronary artery, presenting symptoms such as arrhythmia, shock, or heart failure, which can be fatal [9]. Stroke complicated by a heart attack refers to a cerebral infarction accompanied by a heart attack. It is reported that 30% of stroke cases are complicated by a heart attack, with a high mortality rate of 54% [5]. The main causes of death are ventricular arrhythmia, acute left heart failure, and cardiogenic shock. Troponin is an effective marker for detecting heart attacks [9], commonly used in clinics. However, a drawback of troponin is that it starts changing only four hours after a heart attack has occurred, resulting in a time delay for detection. In contrast, heart attacks have a rapid onset, and sudden deaths can occur easily in heart attack patients.

II. MACHINE LEARNING

Machine Learning (ML) is a field of artificial intelligence focused on developing algorithms that enable computers to learn and make decisions based on observational data. It aims to understand complex patterns and make intelligent predictions using data. ML has a wide range of applications, including web search engines, medical diagnosis, text and handwriting recognition, image classification, demand forecasting, and sales prediction. ML models can be predictive, generating forecasts, or descriptive, extracting insights from data [10][11]. In ML, two primary approaches are commonly used: supervised learning, where the program predicts predefined classes or labels for new observations, and unsupervised learning, where the classes or labels are not predefined during the training process.

III. METHODOLOGY

A wide range of sorts of order procedures have been proposed in writing that incorporates Choice Trees, Naive Bayesian techniques, Brain Organizations, Strategic Relapse, SVM and KNN and so on. In this paper, we assess the exhibition of the KNN calculations on Understudies' Scholastic Execution expectation informational index was utilized for the characterization contrasted and the Decision Tree calculation.

3.1 Naïve Bayes

Naïve Bayes characterization is a famous AI calculation that depends on Bayes' hypothesis with a supposition of freedom between the highlights. It is a straightforward yet compelling probabilistic model utilized for order undertakings.

The calculation is designated "credulous" in light of the fact that it expects that the presence or nonappearance of a specific component is irrelevant to the presence or nonattendance of different elements. As such, it accepts that all highlights are autonomous of one another, which isn't generally obvious in genuine situations. Notwithstanding this improving on presumption, Naïve Bayes frequently performs well by and by and can give solid outcomes [8].

The Naïve Bayes calculation works by computing the probabilities of an example having a place with every conceivable class in view of the noticed component values. It then, at that point, appoints the example to the class with the most noteworthy likelihood. The computation of these probabilities includes assessing the probability of each element given each class and the earlier likelihood of each class.

The calculation is especially valuable while working with high-layered datasets and when the suspicion of component autonomy is sensible. It is known for its computational proficiency and is much of the time utilized in message characterization, spam separating, opinion examination, and other comparable tasks [2][3].

3.2 Decision Tree

Choice Trees are directed learning calculations utilized for both grouping and relapse undertakings. The calculation fabricates a tree-like model by dividing the info information in light of element values. The tree is built utilizing a hierarchical methodology, where each interior hub addresses an element test, and each leaf hub compares to a class name or relapse esteem [8][10].

The fundamental thought behind Choice Trees is to make segments that augment the division of various classes. This is accomplished by recursively dividing the information in view of element values, utilizing measures like Gini File or Data Gain. Choice Trees are not difficult to decipher and can deal with both straight out and mathematical elements. In any case, they might experience the ill effects of overfitting, and their exhibition can be delicate to little changes in the information.

3.3 K-Nearest Neighbors (KNN)

The KNN estimation consigns class imprints to lines inside a dataset considering the class names of planning data that are relative [2]. The KNN estimation works through examining the readiness data for k planning tuples that are closest to the test data tuple and gives out the test tuple a class mark considering the class names of those closest planning tuples. The closeness of a planning tuple to a test tuple is settled using a distance capacity, as Euclidean distance [8]

The central thought for KNN depends resulting to choosing the distances between the tried, and the arrangement data tests to see its nearest neighbors. The tried model is then moved to the class of its nearest neighbor [10].

The KNN is a sensible regardless convincing philosophy for outline. The KNN evaluation is a technique for social event objects subject to closest organizing models in the part space. KNN is a kind of event based learning, or disengaged perceiving where the end is essentially approximated locally and all computation is yielded until get-together.

IV. EXPERIMENTAL RESULTS

The examinations have been coordinated by using Python programming lingo. The Python Scikit-learn is a pack for data portrayal, gathering and portrayal. The information on heart stroke was procured from the Kaggle [7]. This dataset includes of 10 independent variables as features and one dependent variable as the class label that is used to predict stroke disease. The Variables name is like gender, age, hypertension, heart_disease, ever_married, work_type, residence_type, avg_glucose_level, bmi, smoking status and stroke. There were two values for class label which is: 0 for absence of stroke; another is 1 for presence of stroke. The dataset contains 5110 observations with 12 attributes and two class labels i.e., absent class contains 4861 instances and present class contains 249 instances. The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%).

In this study, three machine learning algorithms, namely Naïve Bayes, Decision Tree, and K-Nearest Neighbors (KNN), were applied to predict heart stroke. The performance of these algorithms was evaluated based on accuracy, precision, and recall metrics. The experimental results are shown in the figure-1 as follows:

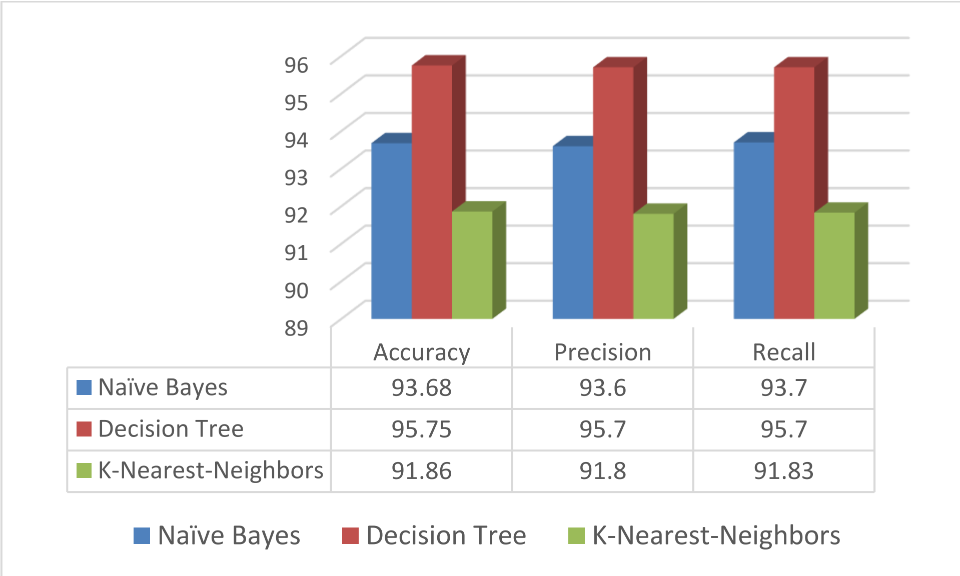

Figure-1: Experimental Results

The results indicate that all three machine learning algorithms performed reasonably well in predicting heart stroke. The Decision Tree algorithm achieved the highest accuracy, precision, and recall scores, closely followed by Naïve Bayes. K-Nearest Neighbors (KNN) showed slightly lower performance compared to the other two algorithms.

These findings suggest that machine learning algorithms have the potential to assist in the early prediction of heart stroke, which can be crucial for reducing the high mortality rate associated with this condition. By accurately identifying individuals at risk, healthcare professionals can intervene early and provide appropriate preventive measures.

It is important to note that the choice of algorithm may depend on specific requirements and considerations. While Decision Tree and Naïve Bayes demonstrated superior performance in this study, the performance of algorithms can vary depending on the dataset and problem domain. Further research and experimentation are needed to validate these findings on larger and more diverse datasets.

V. CONCLUSION

In conclusion, the results of this study demonstrate the effectiveness of machine learning algorithms in predicting heart stroke. These algorithms can serve as valuable tools for healthcare professionals in making informed decisions and implementing preventive strategies. Continued advancements in machine learning techniques can contribute to further improvements in heart stroke prediction and ultimately lead to better patient outcomes.

Based on the experimental results, it can be observed that all three machine learning algorithms showed promising performance in predicting heart stroke. The Decision Tree algorithm achieved the highest accuracy, precision, and recall scores, followed closely by Naïve Bayes. K-Nearest Neighbors (KNN) algorithm exhibited slightly lower performance compared to the other two algorithms.

REFERENCES

[1] D. Bhatnagar, I. Kaur, and A. Kumar, ''Ultrasensitive cardiac troponini antibody based nanohybrid sensor for rapid detection of human heart attack,'' Int. J. Biol. Macromolecules, vol. 95, pp. 505–510, Feb. 2017.

[2] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)

[3] G. Ravi Kumar, S. Rahamat Basha, Surya Bhupal Rao, "A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 324-332, ISSN:0973-8975.

[4] G. Ravi Kumar, K. Tirupathaiah and B. Krishna Reddy, "Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques", International Journal of Computer Sciences and Engineering, Volume-7, Issue-6, e-ISSN: 2347 — 2693, PP: 871-875, June 2019,

[5] G. V. Dous, A. C. Grigos, and R. Grodman, ''Elevated troponin in patients with acute stroke—Is it a true heart attack,'' Egyptian Heart J., vol. 69, no. 3, pp. 165–170, 2017.

[6] Hager Ahmed, Sara F Abd-el Ghany, Eman Younis, Nahla Omran "Stroke Prediction using Distributed Machine Learning Based on Apache Spark" International Journal of Advanced Science and Technology Vol. 28, No. 15, (2019), pp. 89-97

[7] Healthcare Dataset Stroke Data from Kaggle.

[8] J.Han and M.Kamber,"Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[9] L. Junfan, Observation of Clinical Effect of 'Stroke Integration' in the Treatment of Ischemic Stroke. Beijing, China: Beijing Univ. Chinese Medicine, 2019.

[10] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005.

[11] Surya Bhupal Rao, S.Rahamat Basha, G. Ravi Kumar, "A Comparative approach of Text Mining: Classification, Clustering and Extraction Techniques", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 120-131, ISSN: 0973-8975.