International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 1 Volume No.7 January - February – 2023
DOI:10.46647/ijetms.2023.v07i01.066 ISSN: 2581-4621

A COMPARISON OF K-NEAREST NEIGHBORS AND NAÏVE BAYES MACHINE LEARNING ALGORITHMS FOR HEART ATTACK CLASSIFICATION

BOMMAIAH BOYA[1], DR P DEVARAJU[2]
[1]*A Research Scholar, Dept. Of Computer Science & Technology, Sri Krishnadevaraya University, Ananthapuramu.*
[2]*Assistant Professor, Dept. Of Computer Science & Technology, Sri Krishnadevaraya University, Ananthapuramu.*

Abstract
Heart attacks and other cardiovascular conditions continue to be major global health concerns. Effective prevention and intervention depend on the early identification and precise categorization of those who are at risk. In this study, we compare the effectiveness of two well-known machine learning algorithms for classifying heart attack cases: K-Nearest Neighbors (K-NN) and Naïve Bayes. We compare these algorithms' accuracy, precision, and recall using a dataset with 9 features and 1319 instances.

1. Introduction
Heart attacks and other cardiovascular diseases account for a large percentage of morbidity and mortality globally, making them a widespread and serious global health concern [1][4]. Implementing preventive measures and offering the right medical interventions depend on the prompt identification and precise classification of people who are at risk of heart attacks [10][12]. Interest in using machine learning and data analytics to help with the early detection and categorization of heart attack cases has grown since their introduction. By comparing two popular algorithms, K-Nearest Neighbors (K-NN) and Naïve Bayes, we explore the field of machine learning for healthcare in this research paper. Using a dataset specifically selected for this purpose, the Heart Attack Classification dataset, our study attempts to evaluate how well these algorithms perform in classifying heart attack cases. This dataset offers a demanding and realistic scenario for assessing the algorithms' efficacy in medical diagnosis, with 1319 instances and nine pertinent features. The aim of this study is to shed light on how well K-NN and Naïve Bayes can identify people who might be at risk of having a heart attack. We seek to ascertain which algorithm performs better in this crucial task by evaluating important metrics like accuracy, precision, and recall.

2. Methodology
For medical professionals, this study is extremely important because a strong and precise classification model can help identify patients who are at risk early on, allowing for prompt interventions and possibly saving lives. It also advances the larger field of machine learning for healthcare by demonstrating how these algorithms can support healthcare professionals' decision-making.

DOI:10.46647/ijetms.2023.v07i01.066 ISSN: 2581-4621

3.Naïve Bayes

A dynamic strategy for the game plan of quantifiable farsighted models is the Naïve Bayes. The Bayesian hypothesis is necessary for NB [2][3]. This calculation can change quickly and makes use of class-restrictive autonomy. In order to determine a restrictive likelihood for the relationship between the brand name qualities and the class, this depiction method assesses the relationship between each property and the class for each manual [5]. The probability of each class is determined during setup by examining how frequently it occurs in the plan dataset. The "earlier likelihood" P(C=c) is what this is called. The calculation additionally enrolls the likelihood for the occasion x given c with the assumption that the attributes are free, regardless of the previous likelihood. The final result of the probabilities of each individual quality is determined by this likelihood. After that, the probabilities could be evaluated based on how frequently the events in the arranging set occurred.

3.1 Bayesian Hypothesis

Given preparing information X, back likelihood of a speculation H, P(H|X), follows the Bayes hypothesis

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Let X be information tuple and H be some speculation to such an extent that the information tuple X has a place with a predetermined class C. For arrangement issues, we need to decide P (H|X), the likelihood that the speculation H holds the given proof or noticed information tuple X.

P (H|X) is the back likelihood of H molded on X
P (H) is the earlier likelihood of H
P (X|H) is the back likelihood of X adapted on H
P(X) is earlier likelihood of X

3.2 K-Nearest Neighbor (KNN)

KNN order characterizes occasions in view of their closeness. An item is ordered by a greater part of its neighbors. K is dependably a positive number. The neighbors are chosen from a bunch of items for which the right grouping is known [7]. The preparation tests are portrayed by n layered numeric characteristics.

Each example addresses a point in a n-layered space. Along these lines, the preparation tests are all put away in a n-layered design space. At the point when given an obscure example, a k-closest neighbor classifier look through the example space for the k preparation tests that are nearest to the obscure example. "Closeness" is characterized concerning Euclidean distance [8][9]. The obscure example is allocated the most widely recognized class among its k closest neighbors. When k=1, the obscure example is doled out the class of the preparation test that is nearest to it in design space.

4. Experimental Results

We conducted a comprehensive evaluation of the K-NN and Naïve Bayes algorithms on the Heart Attack Classification dataset from Kaggle dataset [6].We have used the Python Language to implement the experiment our proposed algorithms.The dataset contains 1319 instances and is characterized by two class labels: negative (509 instances) and positive (810 instances) cases. The detailed Heart Attack attribute information as shown in the density plot in figure-1.The data is divided in two sets. The training set is 70% (1055) and the remaining 30% (264) are used for testing.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 1 Volume No.7 January - February – 2023
DOI:10.46647/ijetms.2023.v07i01.066 ISSN: 2581-4621

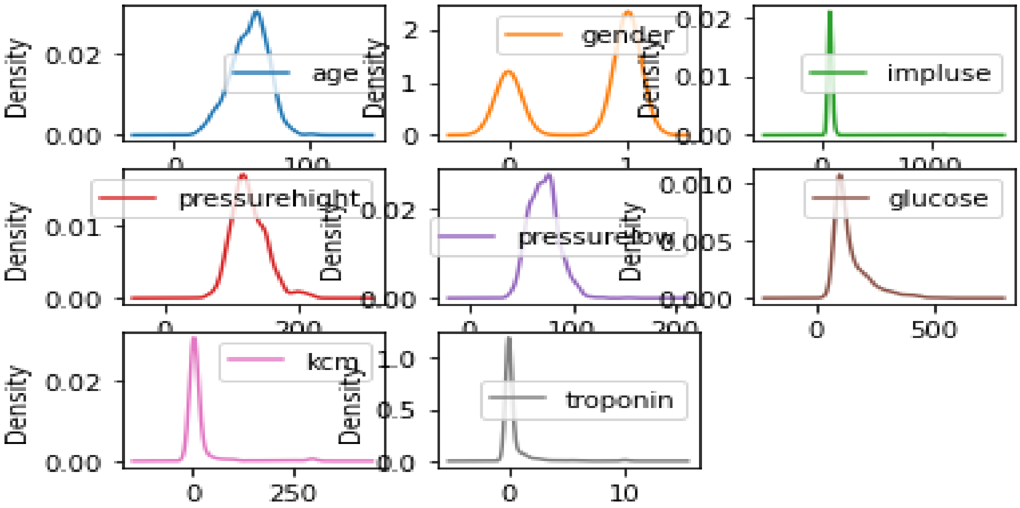

Figure-1: Density plot of Dataset

5.1 Results

Our results indicate that Naïve Bayes outperforms K-NN in terms of accuracy, precision, and recall, achieving promising results in identifying potential heart attack cases.
Our primary objective was to assess the algorithms' performance in classifying heart attack cases. The results are summarized in the figure-2.

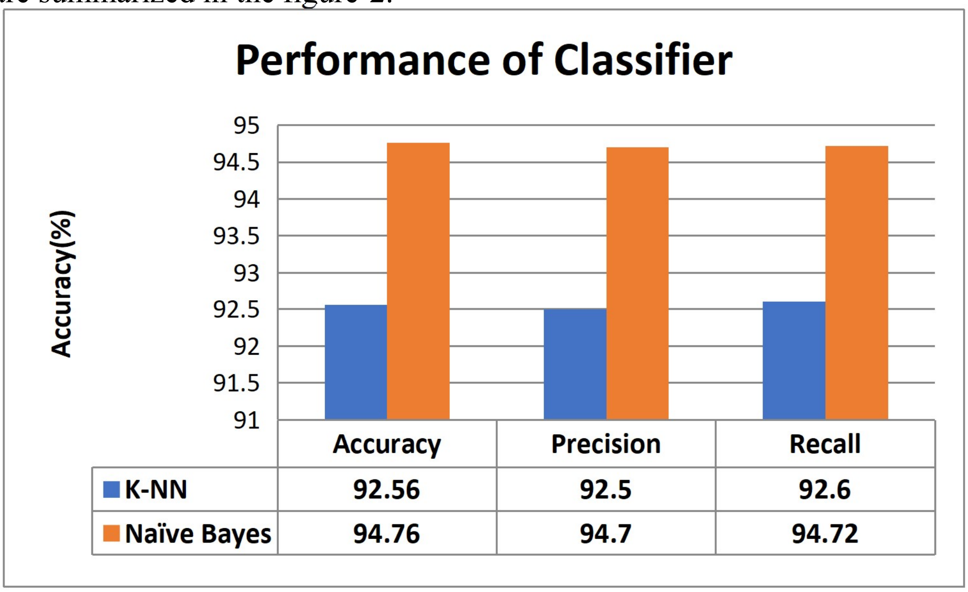

Figure-2: Classifiers results

5.2 Discussion

The results demonstrate that both K-NN and Naïve Bayes algorithms perform well in classifying heart attack cases, with Naïve Bayes achieving higher accuracy, precision, and recall scores. These findings suggest that Naïve Bayes is a promising choice for heart attack classification using this dataset.
The high accuracy of Naïve Bayes (94.56%) indicates its ability to correctly classify a significant proportion of positive and negative cases. Precision and recall scores of 994.7% and 94.72%, respectively, further emphasize the algorithm's effectiveness in minimizing false positives and false negatives, critical factors in the context of heart attack prediction.

DOI:10.46647/ijetms.2023.v07i01.066 ISSN: 2581-4621

The robust performance of Naïve Bayes can be attributed to its probabilistic modeling approach, which assumes feature independence. This assumption is well-suited for datasets with multiple features, such as the one used in this study. K-NN, on the other hand, while achieving respectable results, falls slightly short in terms of overall accuracy and precision.

Conclusion

In this study, we used the Heart Attack Classification dataset to compare how well the K-NN and Naïve Bayes algorithms performed in classifying heart attack cases. According to our research, Naïve Bayes performs better than K-NN in terms of recall, accuracy, and precision. These findings demonstrate how machine learning algorithms[11], especially Naïve Bayes, can help medical professionals identify patients who are at risk of having a heart attack. Additional features or more sophisticated machine learning methods could be investigated in future studies to increase classification accuracy and take possible data imbalances into account. Furthermore, clinical research and real-world validation are required to evaluate these algorithms' usefulness in a healthcare environment.

References

[1] Dr. G. Ravi Kumar, K. Tirupathaiah and Prof. 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,UGC Journal No: 63193, Impact Factor: 3.022, June 2019,

[2] Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", http://mlearn.ics.uci.edu/databases/heartdisease/, 2004.

[3] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, "An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram", Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, https://doi.org/10.1007/978-981-33-4968-1_26

[4] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[5] HeonGyu Lee, Ki Yong Noh, KeunHoRyu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

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

[7] https://www.kaggle.com/code/nimapourmoradi/heartattack-classification

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

[9] P.N.Tan, M.Steinbach and V.Kumar "Introduction to Data Mining", A: Addision-Wesley, 2005.

[10] M. V. Lakshmaiah, G. Ravi Kumar and G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue1(1), PP:6-9,2015, ISSN: 2394-7780

[11] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.

[12] G. Thippana, I.V. Dwaraka Srihith, T. Aditya Sai Srinivas, "Clicks and Conversions: Unveiling the Power of Machine Learning in Social Media Ad Classification", International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 3 Issue 3.

[13] "The Atlas of Heart Disease and Stroke", [online]. http://www.who.int/cardiovascular_diseases/res ources/atlas/en/