

Comparative Analysis of Machine Learning Algorithms for Heart Attack Classification: K-Nearest Neighbors vs. Naïve Bayes

Basineni Sri Manikanta¹, G V Ramesh Babu²

¹PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

²Associate Professor, Dept of Computer Science, SV University, Tirupati

Abstract— Cardiovascular diseases, including heart attacks, remain a significant global health concern. Early detection and accurate classification of individuals at risk are essential for effective prevention and intervention. In this research paper, we evaluate the performance of two popular machine learning algorithms, K-Nearest Neighbors (K-NN) and Naïve Bayes, for the classification of heart attack cases. Using a dataset comprising 1319 instances and 9 features, we compare the accuracy, precision, and recall of these algorithms.

I. INTRODUCTION

Cardiovascular diseases, including heart attacks, represent a pervasive and critical global health concern, responsible for a significant portion of morbidity and mortality worldwide [1][4]. The timely identification and accurate classification of individuals at risk of heart attacks are vital for implementing preventive measures and providing appropriate medical interventions [10][11]. With the advent of machine learning and data analytics, there has been a growing interest in leveraging these technologies to aid in the early detection and classification of heart attack cases.

In this research paper, we delve into the realm of machine learning for healthcare by conducting a comparative analysis of two widely used algorithms: K-Nearest Neighbors (K-NN) and Naïve Bayes. Our study aims to assess the performance of these algorithms in classifying heart attack cases using a dataset specially curated for this purpose, known as the Heart Attack Classification dataset. With 1319 instances and nine relevant features, this dataset presents a challenging and realistic scenario for evaluating the algorithms' effectiveness in medical diagnosis.

The objective of this research is to provide insights into the capabilities of K-NN and Naïve Bayes in accurately identifying individuals who may be at risk of experiencing a heart attack. By measuring key metrics such as accuracy, precision, and recall, we aim to determine which algorithm exhibits superior performance in this critical task.

II. METHODOLOGY

This study holds immense significance for healthcare practitioners, as a robust and accurate classification model can aid in the timely identification of patients at risk, facilitating timely interventions and potentially saving lives. Additionally, it contributes to the broader field of machine learning for healthcare, showcasing the potential of these algorithms in assisting medical professionals in their decision-making processes.

III. NAÏVE BAYES

The Naïve Bayes is a lively strategy for game plan of quantifiable farsighted models. NB depends upon the Bayesian hypothesis [2][3]. This calculation utilizes class restrictive autonomy and has capacity to rapidly change. This depiction method evaluations the relationship between each property and the class for each manual for choose a restrictive likelihood for the relationship between the brand name qualities and the class [5]. Amidst setting up, the likelihood of each class is enlisted by checking how routinely it happens in the plan dataset. This is known as the "earlier likelihood" $P(C=c)$. Regardless of the earlier likelihood, the calculation moreover enrolls the likelihood for the occasion x given c with the presumption that the attributes are free. This likelihood changes into the eventual outcome of the probabilities of each single quality. The probabilities would then have the choice to be assessed from the frequencies of the occasions in the arranging set.

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.

IV. 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.

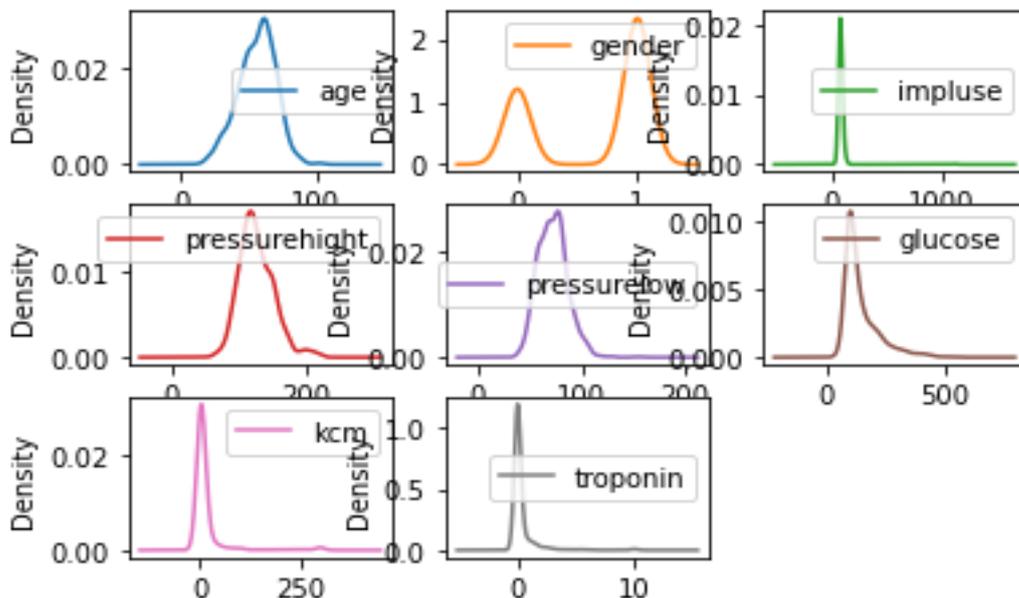


Figure-1: Density plot of Dataset

4.1 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 table-1 and same shown in the figure-2.

Table-1
Results of Classifiers

Algorithm	Accuracy	Precision	Recall
K-NN	89.86	89.8	90
Naïve Bayes	93.43	93.4	93.5

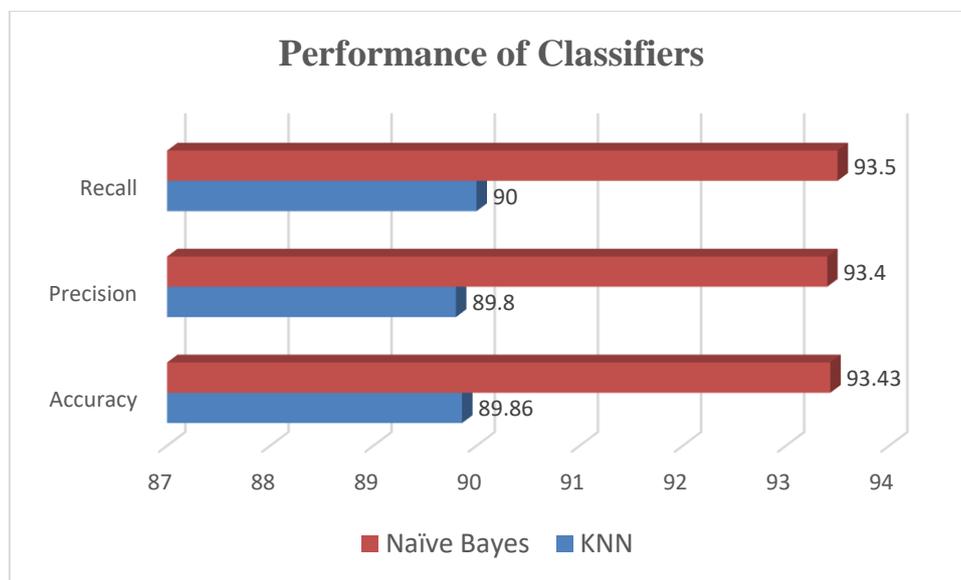


Figure-2: Classifiers results

4.2 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 (93.43%) indicates its ability to correctly classify a significant proportion of positive and negative cases. Precision and recall scores of 93.4% and 93.5%, respectively, further emphasize the algorithm's effectiveness in minimizing false positives and false negatives, critical factors in the context of heart attack prediction.

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.

V. CONCLUSION

In this research paper, we compared the performance of K-NN and Naïve Bayes algorithms in classifying heart attack cases using the Heart Attack Classification dataset. Our findings suggest that Naïve Bayes outperforms K-NN in terms of accuracy, precision, and recall. These results highlight the potential of machine learning algorithms, particularly Naïve Bayes, in assisting healthcare professionals in identifying individuals at risk of heart attacks.

Further research can explore the incorporation of additional features or more advanced machine learning techniques to improve classification accuracy and to account for potential data imbalances. Additionally, real-world validation and clinical studies are necessary to assess the practical applicability of these algorithms in a healthcare setting.

REFERENCES

- [1] Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", <http://mllearn.ics.uci.edu/databases/heartdisease/>, 2004.
- [2] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, "An Improved Decision Tree Classification Approach for Expectation of Cardiocogram", 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
- [3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [4] 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.
- [5] 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.
- [6] <https://www.kaggle.com/code/nimapourmoradi/heartattack-classification>
- [7] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005
- [8] P.N.Tan, M.Steinbach and V.Kumar "Introduction to Data Mining", A: Addison-Wesley, 2005.
- [9] 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
- [10] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
- [11] "The Atlas of Heart Disease and Stroke", [online]. http://www.who.int/cardiovascular_diseases/resources/atlas/en/