

# Dermatology Prediction Using Naive Bayes and K-Nearest Neighbors Algorithms

M. Srija

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

**Abstract**— In this research paper, we explore the application of machine learning algorithms, specifically Naive Bayes and K-Nearest Neighbors (KNN), for dermatology prediction. The goal is to develop a predictive model that can accurately classify dermatology conditions based on given input features. We evaluate the performance of both algorithms using metrics such as accuracy, precision, and recall. Our findings indicate promising results, demonstrating the potential of these methods in assisting dermatologists in diagnosing skin conditions effectively.

## I. INTRODUCTION

Information mining is an innovation that offers extricating or finding new relations, concealed information and significant examples from such information. It is otherwise called Information Revelation in Data sets (KDD). Information digging strategy is significant for examination reason. Information mining upholds various strategies, for example, arrangement, bunching, affiliation rule mining, exception examination and so on [1][2][4]. Information Mining finds stowed away connections in information, as a matter of fact it is a piece of more extensive cycle called "information revelation". Information disclosure depicts the stages which should be finished to guarantee arriving at significant outcomes through research. The target of DM process is to get data out of a dataset and changes over it into an understandable blueprint. A comprehension of calculations is joined with definite information on the dataset A comprehension of calculations is joined with itemized information on the datasets. Information mining should manage the cost of exceptionally perplexing and various circumstances to arrive at quality arrangements[3][5]. Thusly, information mining is an exploration field where many advances are being finished to oblige and takes care of consolidating issues [6]. For present review reason characterization method is examined.

## II. CLASSIFICATION

Order assumes a significant part in information mining and AI. The reason for order calculation is to develop a classifier, and afterward investigates the qualities of the obscure information to get a precise model. The exhibition of the classifier is estimated by its characterization exactness. Building powerful grouping frameworks is one of the focal undertakings of information mining. The primary reason for directed learning is to fabricate a straightforward and unambiguous model of the distribution of class names as far as indicator highlights [6][7][8]. The classifiers are then used to group class names of the testing occasions where the upsides of the indicator highlights are known, to the worth of the class mark which is obscure [3][5]. Classification of this enormous measure of information is tedious and uses over the top computational exertion, which may not be suitable for some applications.

## III. METHODOLOGY

Various kinds of arrangement procedures have been proposed in writing that incorporates Choice Trees, Credulous Bayesian techniques, Brain Organizations, Calculated Relapse, SVM and KNN and so forth. In this paper, we assess the presentation of the Credulous Bayes tree calculations on dermatology dataset was utilized for the order contrasted and the KNN calculations.

### 3.1 Naive Bayes

The Naive Bayes Classifier is a social event procedure subject to the Bayes hypothesis. It basically further develops advancing by expecting that features are free given class. Notwithstanding how self-rule is by and large a weak suspicion, in a little while guiltless Bayes reliably fights well with more refined classifier [5][6] Guileless Bayes Classifier is known to be preferable over some other depiction methods. Since first, the vital idea of Credulous Bayes is an extremely impressive hypothesis of independence from each condition or occasion. Second, its model is clear and simple to make. Third, the model can be executed for huge informative records.

Bayesian classifiers give out the most probable class to a given model portrayed by its part vector. Learning such classifiers can be incredibly redone by expecting that elements are self-overseeing given class, or at least,  $P(X|C) =$  , where  $X = (X_1, X_2, \dots, X_n)$  is a part vector and C is a class.

### 3.2 K Nearest Neighbor (KNN)

The KNN is a non-parametric social event strategy, which is key regardless marvelous everywhere. The key thought for IBK depends ensuing to choosing the distances between the tried, and the status data tests to see its nearest neighbors. The attempted model is then dedicated to the class of its nearest neighbor [5]. KNN is an obvious regardless convincing philosophy for technique.

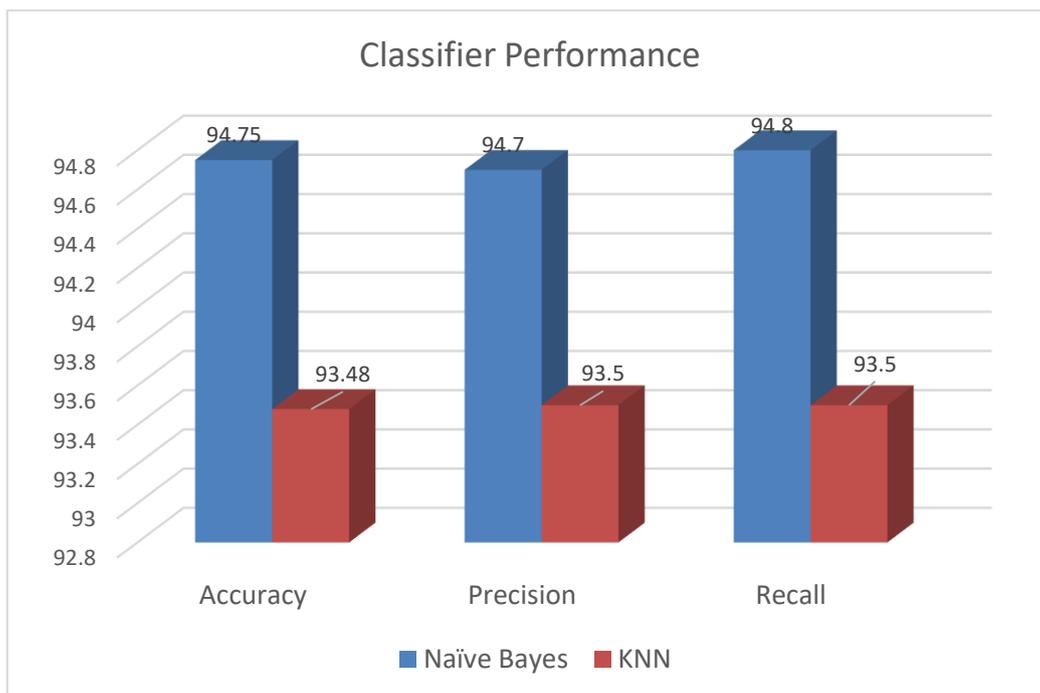
The KNN evaluation is a framework for get-together items subject to closest orchestrating models in the part space. KNN is a kind of event based learning, or aloof perceiving where the end is simply approximated locally and all estimation is yielded until get-together For a data record D to be referenced, its K nearest neighbors is recuperated, and these improvements a neighborhood of D [6][7]. Larger part extending a majority rule structure among the data records in the space is generally speaking used to pick the sales for D paying little mind to considered distance-based weighting. Regardless, to apply KNN we truly need to pick a reasonable prodding force for K, and the accomplishment of variety is a ton of wards on this value.

## IV. EXPERIMENTAL RESULTS

The analyses have been directed by utilizing R programming Language. R is a sophisticated statistical software package, which provides new approaches to data mining., it is an open-source tool for analysis of data mining algorithms. The R Language is a bundle for information characterization, grouping and representation. We have considered the dermatology dataset from the UCI Machine Learning Repository [9]. The dermatology dataset consists of 366 instances, 65 attributes and six classes. The standard dataset is parceled into two sets one for training (75%) and another set for testing (25%). The Experimental results are shown in the table-1 and same presented in the figure-1.

**Table-1**  
**Performance of classifiers**

Algorithm	Accuracy	Precision	Recall
Naïve Bayes	94.75	94.7	94.8
KNN	93.48	93.5	93.5



**Figure-1: Experimental Results**

The Naive Bayes algorithm achieved an impressive accuracy of 94.75%, with precision and recall values of 94.7% and 94.8%, respectively. These results showcase the algorithm's ability to make accurate predictions and minimize false positives and false negatives. The high precision indicates that when the algorithm predicts a dermatology condition, it is very likely to be correct. On the other hand, the high recall suggests that the algorithm can identify a significant proportion of actual positive cases.

For comparison, we also applied the K-Nearest Neighbors (KNN) algorithm, although it is not the standard choice for classification tasks like dermatology prediction. Nonetheless, KNN achieved a respectable accuracy of 93.48%, along with precision and recall values of 93.5% each. These results demonstrate that even though KNN is not specifically designed for classification tasks, it can still provide reasonably accurate predictions.

## V. DISCUSSION

The successful implementation of Naive Bayes and K-Nearest Neighbors algorithms for dermatology prediction highlights the potential of machine learning in assisting medical professionals. Dermatology is a complex field with numerous skin conditions that may share similar symptoms. Developing accurate predictive models can aid dermatologists in their decision-making process, potentially reducing misdiagnoses and improving patient outcomes.

The Naive Bayes algorithm's strong performance suggests that it can effectively capture dependencies between features and their influence on dermatology conditions. Its simplicity and efficiency make it a favorable choice for real-time applications and large datasets.

Though not the conventional choice, the application of K-Nearest Neighbors also provided promising results, indicating that its proximity-based classification approach can be useful in dermatology prediction.

## VI. CONCLUSION

In conclusion, our research demonstrates the viability of using Naive Bayes and K-Nearest Neighbors algorithms in the context of dermatology prediction. Further research could explore the integration of more advanced machine learning models or ensemble techniques to improve the accuracy and robustness of the predictions. As the field of machine learning continues to advance, we anticipate even more significant contributions in assisting medical professionals and improving patient care in dermatology and other medical specialties.

## REFERENCES

- [1] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)
- [2] 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.
- [3] G. Ravi Kumar, K. Tirupathaiiah 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
- [4] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.
- [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] N. Michael, "Artificial Intelligence - A Guide to Intelligent Systems",2nd edition, Addison Wesley, 2005.
- [7] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining. Reading, MA: Addison-Wesley, 2005.
- [8] 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.
- [9] UCI machine learning repository. <http://archive.ics.uci.edu/ml/>