Dr Preetha V, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Research on classification models in Machine Learning for IRIS Dataset

Dr Preetha V
Assistant Professor of Computer Science, Sri S.Ramasamy Naidu College, Sattur

Abstract- Clasification models in machine learning is used in various applications for prediction. It is a supervised learning model where the models can be trained with the required outputs. There are various types of classifiers such as Naïve Bayes, KNN, Decision trees and support vector machines. Neural network machine learning model is also used for classification and prediction. The IRIS dataset is most familiar and commonly used dataset for classification to predict the accuracy. The research focusses on the analysis of the KNN based classification and Naïve bayes classification algorithms for the performance analysis.

Keywords: Machine learning, Naïve Bayes, KNN, Classification.

I. INTRODUCTION

Machine Learning (ML) is a subfield of Artificial Intelligence that focuses on developing algorithms and models that enable computers to learn from data and make decisions or predictions without being explicitly programmed. With the rapid growth of data and computational power, machine learning has become a cornerstone of modern technology and scientific research. Research in machine learning aims to design efficient, accurate, and scalable models that can extract meaningful patterns from large and complex datasets. It combines concepts from statistics, mathematics, optimization, and computer science to solve real-world problems across various domains such as healthcare, finance, education, and transportation.

Machine Learning Approaches:

There are three primary type of machine learning approaches such as supervised learning, unsupervised learning and reinforcement learning. Classification and Regression will fall under supervised learning with labelled data. Unsupervised learning identify patterns in unlabelled data with the use of clustering algorithms.Reinforcement learning learns by interaction with the environment and maximizing rewards.

Modern ML models focusses on advanced models such as :

- Deep Learning and neural networks
- Natural Language Processing (NLP)
- Computer Vision
- Explainable AI (XAI)
- Ethical and fair AI systems

Machine learning has enabled breakthroughs in applications like speech recognition, image classification, recommendation systems, and autonomous vehicles. However, challenges such as data privacy, model bias, interpretability, and computational cost remain active areas of research.

Classification in Machine learning

The primary goal of classification is to predict a discrete label or category for a given input based on historical data.

1. Types of Classification

Classification can be of the following types such as:

- **Binary Classification:** The outputs may either true or false such as:
- **Examples:** Spam vs. Not Spam, Disease vs. Healthy.
- **Multiclass Classification:** There may be More than two categories, but each input may belong to only one.
- **Examples:** Identifying a digit (0–9), classifying types of fruit.
- **Multilabel Classification:** An input can belong to multiple categories simultaneously.
- **Example:** Tagging a news article as both "Politics" and "Economy."

K-Nearest Neighbors (KNN):
K-Nearest Neighbors (KNN) is a simple supervised machine learning algorithm used for classification and regression. It works by finding the K closest data points (neighbors) to a new input and assigning the most common class among them. It is also known as "lazy learner".

Naïve Bayes Classifier:
The Naive Bayes classifier is a supervised machine learning algorithm based on the concept of Bayes' Theorem. It is mainly used for classification tasks such as spam detection, text classification, and medical diagnosis.Naive Bayes calculates the probability that a data point belongs to a particular class based on prior knowledge and observed features. It assumes that all features are independent of each other (this is the "naive" assumption).

Peformance Analysis of Classification:
Accuracy of a model can be measured but still confusion matrix will be useful to see the output of the model where it gets confused. For example, if it the problem is to detect whether a mail is spam or not the following metrics will determine as:

TABLE 1. Performance Analysis

Metric	What it tells
Precision	"Of all the times I predicted 'Spam', how often was I right?"
Recall	"Of all the actual 'Spam' emails, how many did I catch?"
F1-Score	The balance between Precision and Recall.
AUC-ROC	How well the model distinguishes between classes at various thresholds.

Proposed work:
The Iris dataset is one of the most popular datasets in machine learning. It contains measurements of iris flowers: Sepal length ,Sepal width ,Petal length Petal width.These features are used to classify flowers into three species: Setosa ,Versicolor and Virginica.

Figure 1. Iris Dataset Species

The Iris dataset has been downloaded from kaggle dataset and the coding was done in python by using KNN-classifier.

Working of KNN:
1. Choose a value of K (e.g., K = 3 or 5).
2. Measure the distance (usually Euclidean distance) between the new data point and all training data.
3. Select the K nearest neighbors.
4. Assign the class that appears most frequently among those neighbors.

Working of Naïve Bayes:
Since the Iris dataset has continuous values, we typically use Gaussian Naive Bayes.
1. For each class (Setosa, Versicolor, Virginica), calculate:
2. Mean and variance of each feature
3. Use probability distribution (Gaussian) to compute likelihood
4. Apply Bayes' Theorem to find posterior probability
5. Assign the class with the highest probability

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5	3.6	1.4	0.2	Iris-setosa

Figure 2. Iris Dataset

OUTPUT:

Accuracy: 1.0
Confusion Matrix:
[[10 0 0]
[0 9 0]
[0 0 11]]

Figure 3. Output of the classification models

We observed that the both the classifiers predict the same output with accuracy and confusion matrix. The results are identical and no data preprocessing was also applied. Even though Naïve Bayes assumes data independence, it still works on iris dataset.

II. CONCLUSION

This research paper analysis the performance of K-NN and Naïve bayes in the Iris dataset and implemented the results.It shows that both classifiers perfomed well with the training and testing datasets.]

REFERENCES

1. Chadha, J., & Singh, R. (2025). A comparative analysis of decision tree, K-nearest neighbors and naive Bayes algorithms for classification tasks. International Journal of Engineering Research & Technology, 14(4). (IJERT)
2. Octavianto, M. D. A., & Subtari, T. (2025). Analisis perbandingan kinerja algoritma klasifikasi data menggunakan metode K-NN, naive Bayes, dan decision tree pada dataset UCI Iris. MEANS (Media Informasi Analisa dan Sistem), 10(1), 35–40. (slot thailand)
3. Akarshana, S., Annapoorani, S., Kaviya, S., Kousika, K. S., Ebenezer, S., Rajkumar, D., & Kumar, Y. (2025). Comparative analysis of naive Bayes, SVM, random forest, and KNN algorithms for iris flower classification. International Journal of Computer and Information Engineering, 19(6). (World Academy)
4. Yu, C. (2024). Analysis of naive Bayesian and back propagation algorithms in iris classification. Applied and Computational Engineering, 37. (Ewa Direct)
5. Hasan, M., & Ali, A. (2024). Implementing a naive Bayes classifier on iris data using MATLAB with grid parameter optimization. Wasit Journal for Pure Sciences, 3(4), 51–58. (Wasit Journal for Pure Sciences)
6. Riadi, I., Umar, R., & Anggara, R. (2024). Comparative analysis of naive Bayes and K-NN approaches to predict timely graduation using academic history. International Journal of Computing and Digital Systems, 16(1). (ResearchGate)