International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

# Performance Analysis of SVM and KNN

**Ms. Shree Shathveekaa k v**

III-B. Sc CT

Department of CT Dr. N.G.P Arts and Science College Coimbatore

Email:shathveekaa0904@gmail.com

**Dr. D. Geethamani**

Assistant Professor Department of CT Dr. N.G.P Arts and Science College Coimbatore

https://doi.org/10.55041/ijsmt.v2i3.304

**Cite this Article**: v, S. S. K. (2026). Performance Analysis of SVM and KNN. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.304



## ABSTRACT

Machine learning is one of the fastest-growing fields in computer science that enables systems to learn automatically from data and make intelligent decisions without being explicitly programmed. It plays a major role in real-world applications such as healthcare, banking, fraud detection, education, and recommendation systems. The primary objective of machine learning is to identify meaningful patterns and relationships in large datasets and use them to predict future outcomes. Among different learning approaches, supervised learning is widely used because it trains models using labeled data, and classification is one of its most important tasks. Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are two commonly used supervised learning algorithms for classification and prediction problems. SVM works by finding an optimal hyperplane that separates classes with maximum margin, while KNN classifies new instances based on the majority class among the nearest neighbors using distance measures. This review paper presents an overview of machine learning concepts, discusses the importance of classification, explains the working principles of SVM and KNN, and compares their performance using evaluation metrics such as accuracy, precision, recall, F1-score, training time, testing time, scalability, and memory usage. The study concludes that SVM is more suitable for high-dimensional and complex datasets, whereas KNN is effective for smaller datasets due to its simplicity and ease of implementation.

## 1.INTRODUCTION

Machine learning is a major branch of artificial intelligence that focuses on developing systems capable of learning from data and improving performance without explicit programming. In the modern digital era, a huge amount of data is generated continuously through social media platforms, e-commerce systems, healthcare records, banking transactions, industries, and educational applications. Since manual processing of such large-scale data is difficult and time-consuming, machine learning provides automated methods to analyze datasets, extract meaningful patterns, and generate accurate predictions. Machine learning techniques are broadly classified into supervised learning, unsupervised learning, and reinforcement learning, where supervised learning is widely used when labeled data is available. Classification is one of the most important supervised learning tasks, as it helps in predicting categories such as spam or not spam, diseased or healthy, fraud or genuine, and pass or fail. Among the various classification algorithms, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are widely used due to their effectiveness and reliability in solving real-world problems.

These algorithms have applications in medical diagnosis, text categorization, pattern recognition, customer prediction, and fraud detection. This review paper focuses on explaining the fundamental concepts of machine learning and provides a detailed study of SVM and KNN algorithms along with their comparative performance analysis.

## 2.LITERATURE REVIEW

In earlier years, classification systems mainly depended on traditional statistical approaches and rule-based expert systems, where decision-making relied heavily on manually designed rules and handcrafted feature extraction techniques. Although these methods were effective for small-scale problems, they faced major limitations when handling large datasets and complex patterns due to poor scalability, high computational requirements, and reduced accuracy in noisy environments. With the rapid development of artificial intelligence and advanced computing technologies, machine learning-based classification techniques gained popularity because they can automatically learn from data and improve predictive performance. Several research studies have demonstrated that Support Vector Machine (SVM) performs effectively in applications such as text mining, image recognition, medical diagnosis, and fraud detection due to its strong generalization ability and efficiency in handling high-dimensional datasets. Researchers also highlight that kernel-based SVM methods can successfully manage non-linear classification problems by transforming the input data into higher-dimensional feature spaces. Similarly, K-Nearest Neighbor (KNN) has been widely applied in recommendation systems, pattern recognition, and disease prediction because of its simplicity and ease of implementation, especially for small datasets. However, studies also report that KNN becomes computationally expensive for large datasets since it requires repeated distance calculations during prediction. Therefore, based on previous research findings, both SVM and KNN provide strong classification performance, but the selection of the most suitable algorithm depends on dataset size, feature complexity, and computational requirements.

## 3. SUPPORT VECTOR MACHINE (SVM)

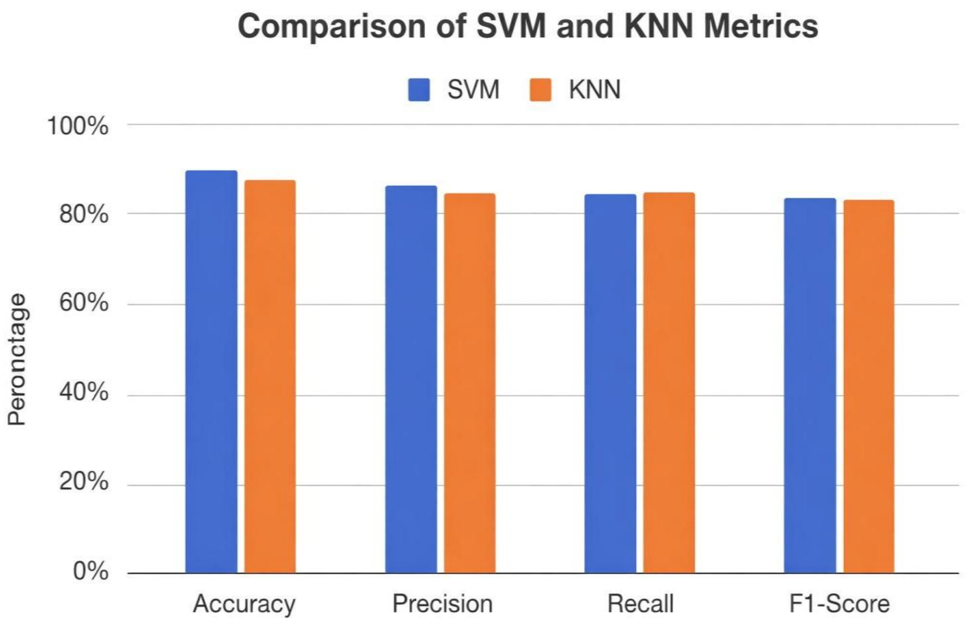
Support Vector Machine (SVM) is a supervised machine learning algorithm widely used for classification and prediction tasks. The main objective of SVM is to identify an optimal decision boundary, known as a hyperplane, that separates different classes in the dataset with maximum margin. The data points that lie closest to the hyperplane are called support vectors, and they play a crucial role in defining the position and orientation of the separating boundary. SVM is highly effective for both linear and non-linear classification problems. When the dataset is not linearly separable, SVM applies kernel functions such as linear, polynomial, radial basis function (RBF), and sigmoid kernels to transform the input data into a higher-dimensional feature space, making classification easier. Due to its strong generalization capability and efficiency in handling high-dimensional datasets, SVM is widely applied in real-world applications such as medical diagnosis, handwriting recognition, face recognition, text classification, and fraud detection.

## 4. K-NEAREST NEIGHBOR (KNN)

K-Nearest Neighbor (KNN) is a supervised learning algorithm mainly used for classification problems and is considered one of the simplest machine learning techniques. KNN works based on similarity and distance measurement by classifying a new data instance according to the class labels of its nearest neighbors in the training dataset. The algorithm determines the K closest training samples using distance measures such as Euclidean distance, Manhattan distance, or Minkowski distance, and assigns the final class label based on majority voting among the selected neighbors. KNN is referred to as a lazy learning algorithm because it does not build a predictive model during training; instead, it stores the entire dataset and performs computation only during the testing phase. KNN is commonly used in pattern recognition, recommendation systems, image classification, and medical prediction applications. However, its performance strongly depends on selecting an appropriate value of K, and it becomes computationally expensive for large datasets due to increased prediction time and higher memory requirements.

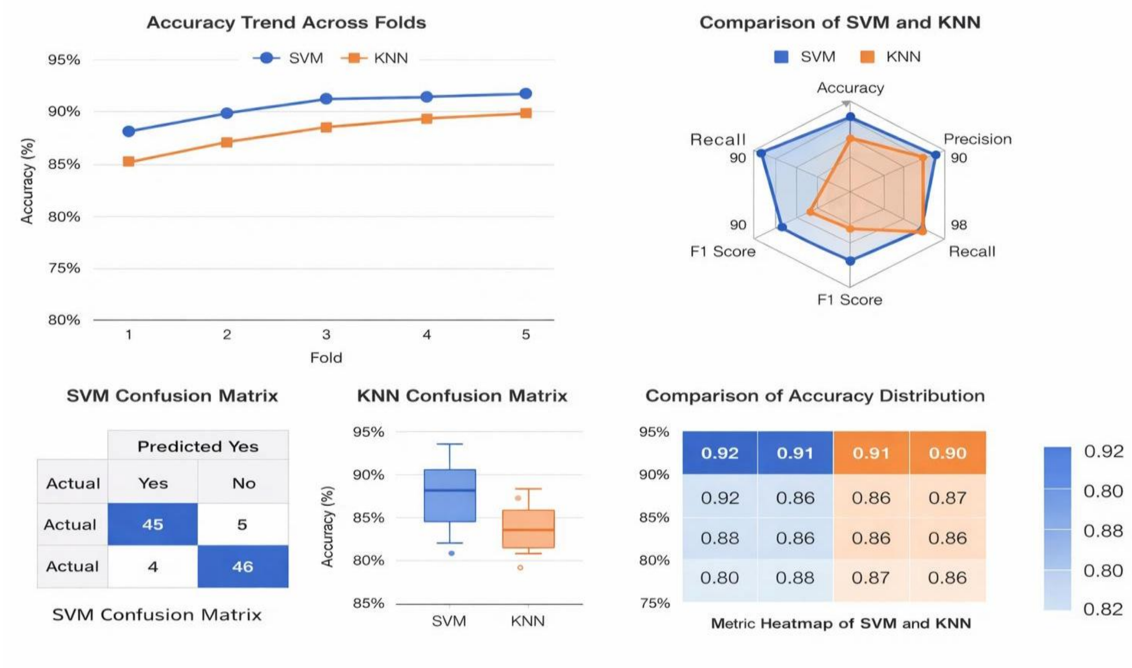
## 5.COMPARISON OF SVM AND KNN USING METRICS

The performance of classification algorithms is commonly evaluated using metrics such as accuracy, precision, recall, F1-score, training time, testing time, scalability, and memory usage. Support Vector Machine (SVM) generally provides higher accuracy and better generalization performance because it constructs an optimal hyperplane with maximum margin, making it highly suitable for high-dimensional and complex datasets, especially when kernel functions are applied for non-linear classification. However, SVM requires more computational effort and time during training since the model must be optimized before prediction. In contrast, K-Nearest Neighbor (KNN) is a simple instance-based algorithm that does not require an explicit training phase, as it stores all training data and performs classification only during testing, resulting in low training time but higher prediction time due to repeated distance calculations. KNN also consumes more memory because it must store the entire dataset, and its performance may reduce when handling large datasets. Overall, SVM is more effective for large-scale and high-dimensional classification problems requiring strong accuracy and stability, while KNN is more suitable for small datasets where simplicity and ease of implementation are preferred.




## COMPARATIVE ANALYSIS OF SVM AND KNN PERFORMANCE METRICS:

The comparative analysis of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms based on key performance metrics such as accuracy, precision, recall, and F1-score. The accuracy trend graph across multiple folds shows that SVM consistently achieves higher accuracy compared to KNN, indicating better stability and generalization performance. The radar chart clearly highlights that SVM performs better overall in most evaluation parameters, especially in recall and F1-score, which shows its effectiveness in correctly identifying positive cases. The confusion matrix representation also indicates that SVM produces fewer misclassifications compared to KNN, leading to improved prediction reliability. Additionally, the box plot and heatmap visualization confirm that SVM has a higher and more consistent accuracy distribution across different test conditions, while KNN shows slightly lower performance with more variation. Overall, the comparison concludes that SVM is more efficient and accurate than KNN for classification tasks, especially when handling complex and high-dimensional datasets.




## 6. CONCLUSION

This review paper presented an overview of machine learning and discussed two important supervised learning algorithms, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN), which are widely used for classification tasks. SVM is a powerful algorithm that identifies an optimal hyperplane with maximum margin, making it highly effective for high-dimensional and complex datasets, and it provides strong generalization performance for unseen data. In contrast, KNN is a simple and efficient algorithm that classifies instances based on similarity and nearest neighbor distance measures, making it suitable for small datasets due to its ease of implementation. However, KNN becomes computationally expensive for large datasets because it requires repeated distance calculations during prediction and demands higher memory usage for storing training samples. Based on the comparative analysis, SVM is more appropriate for complex and large-scale classification problems where accuracy and stability are essential, while KNN is more suitable for smaller datasets where simplicity and quick implementation are required. Both algorithms remain significant in machine learning research and continue to play an important role in various real-world applications.

## 7. REFERENCES

[1] T. M. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.

[2] V. N. Vapnik, The Nature of Statistical Learning Theory. New York, NY, USA: Springer, 1995.

[3] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[4] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA, USA: Elsevier, 2012.

[5] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, U.K.: Cambridge University Press, 2000.

[6] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” Machine Learning, vol. 6, no. 1, pp. 37–66, Jan. 1991.

[7] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[8] S. Raschka and V. Mirjalili, Python Machine Learning, 2nd ed. Birmingham, U.K.: Packt Publishing, 2017.

[9] S. Lata and D. Singh, “A hybrid approach for cloud load balancing,” in Proc. 2nd Int. Conf. Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022, pp. 548–552.

[10] S. Lata and R. Kumar, “A hybrid approach for ECG signal analysis,” in Proc. IEEE Int. Conf. Advances in Computing, Communication Control and Networking (ICACCCN), 2018, doi: 10.1109/ICACCCN.2018.8748858.

[11] Arbor Networks, “Annual security report,” 2015. [Online]. Available: https://www.arbornetworks.com/□. Accessed: Feb. 2026.

[12] DataFlair, “Machine Learning Tutorial.” [Online]. Available: https://data-flair.training/□. Accessed: Feb. 2026.