

# Research and Application of Machine Learning in Fake News Detection on Social Media



Van Son Tran<sup>1</sup>, Kim Quang Hoang<sup>2</sup>  
Hanoi University of Industry, No.298 Cau Dien Street, Bac Tu Liem District, Hanoi, Vietnam

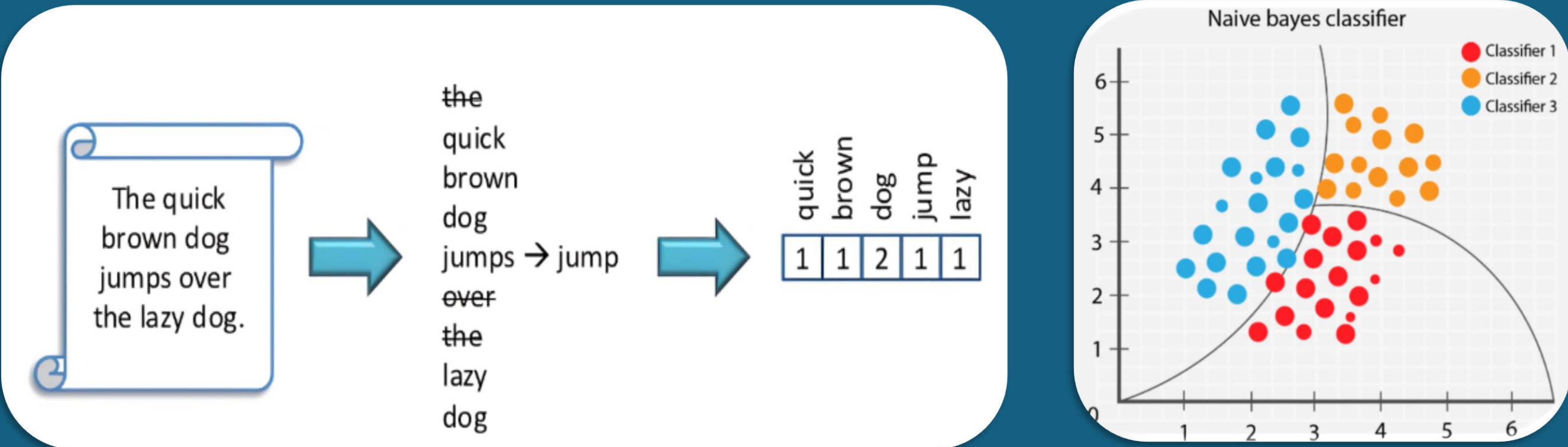
## 1. Introduction

In the era of digital information and social media, detecting and preventing the spread of fake news on social networks has become a significant challenge. Fake news, also known as misinformation, can cause serious harm to user trust and have negative impacts on society. In the effort to address this phenomenon, machine learning applications have been widely employed to detect fake news on social networks.

In this topic, we will explore the applications of Natural Language Processing (NLP) and three popular machine learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Naïve Bayes in detecting fake news on social networks.

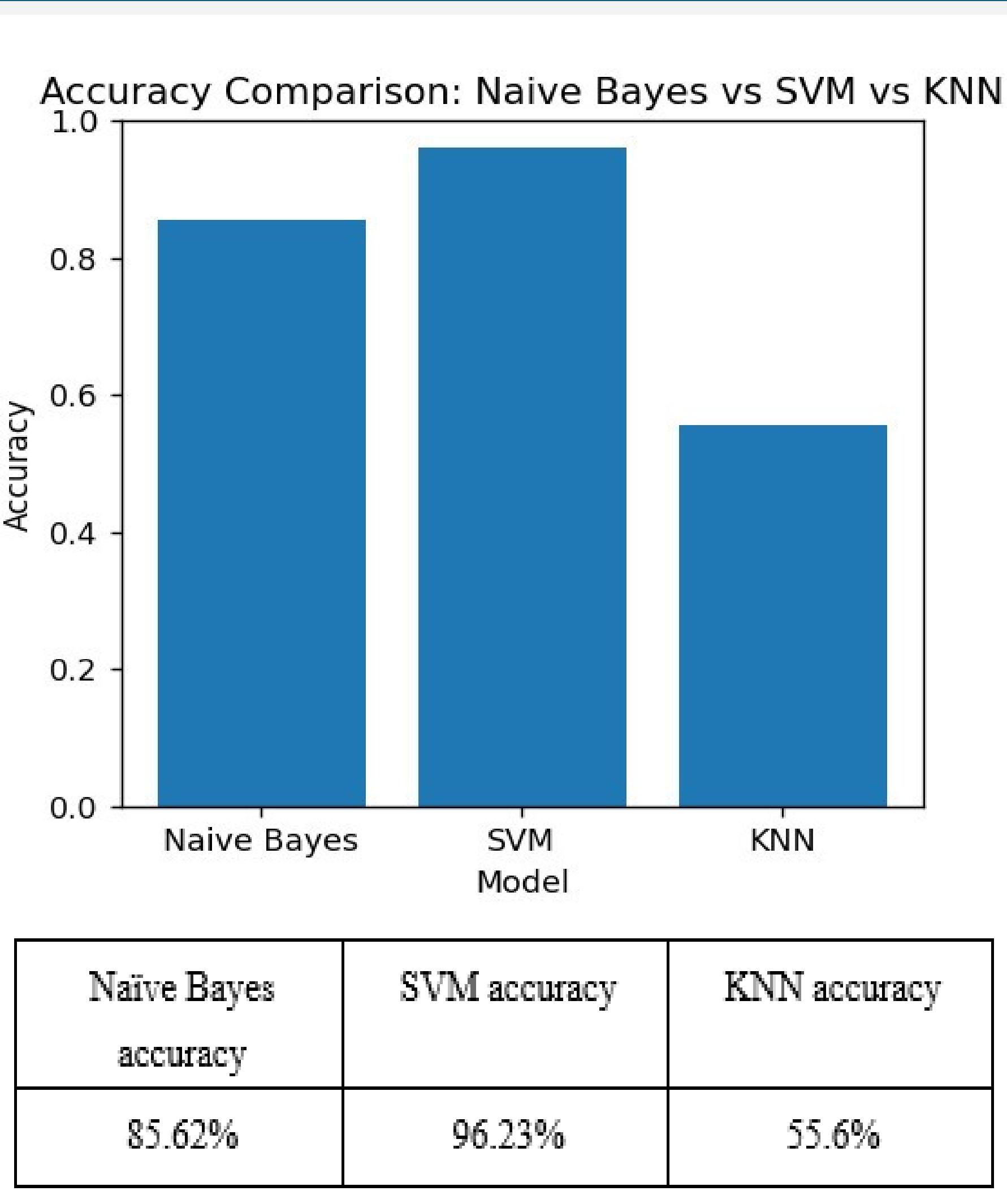
## 2. Materials and Methods

Kaggle data is an online data resource provided by Kaggle, a popular community platform for data scientists and experts in machine learning and data science. We selected a dataset consisting of 20,800 news articles from that data resource, which includes both fake and real news. We then processed the news articles using natural language processing techniques and used the preprocessed data in machine learning models such as Naïve Bayes, SVM, and KNN. We compared the model evaluation metrics to determine the best-performing models and provided observations on which model performed the best.



## 3. Results and Discussion

The comparison of machine learning models' evaluations reveals interesting insights. Naïve Bayes, although providing relatively accurate results with an accuracy rate of 85.62%, falls short compared to SVM and KNN. SVM, on the other hand, stands out as the top-performing model with an impressive accuracy rate of 96.23%, surpassing both Naïve Bayes and KNN. KNN, unfortunately, exhibits the lowest accuracy rate of 55.6%, indicating its limitations in effectively classifying the data. Overall, the results indicate that SVM is the most suitable model for this particular task, offering the highest accuracy among the evaluated models.



## 4. Conclusion and Recommendation

In this report, we utilized Natural Language Processing (NLP) and applied three machine learning models: Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) to address the problem of detecting fake news on social media. We performed data preprocessing by removing stop words, normalizing the text, and using TF-IDF to represent the texts as numerical vectors. Subsequently, we built three machine learning models for news classification, each employing a different approach. The results indicate that all three models show potential in accurately classifying news articles. However, the choice of model depends on the specific requirements of the problem and the characteristics of the data.