Comparative Analysis of Naive Bayes Algorithms for Date Fruit Classification

Ambalathandhi Palani Bhagyalakshmi

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

Abstract— *Date organic products are the most widely recognized natural product in the Center East and North Africa. There are a wide assortment of dates with various kinds, colors, shapes, tastes, and healthy benefits. Grouping, distinguishing, and perceiving dates would assume an essential part in the farming, business, food, and wellbeing areas. Deciding the range of natural products by taking a gander at their outer appearance might require mastery, which is tedious and requires incredible exertion. The aim of this study is to classify the types of date fruit, that are, Barhee, Deglet Nour, Sukkary, Rotab Mozafati, Ruthana, Safawi, and Sagai by using two different machine learning methods. This study presents a comparative analysis of two Naive Bayes algorithms, namely Naive Bayes and Naive Bayes Multinomial, for the classification of date fruits. The experimental results evaluate the performance of these algorithms in terms of accuracy, precision, and recall. The findings contribute to understanding the effectiveness of different Naive Bayes variants in the context of date fruit classification.*

I. INTRODUCTION

Dates are the natural product created from palm trees. It is a late spring natural product that is far reaching in the Middle Easterner world. The qualities of dates incorporate taste and surface, which contrast and must be anticipated in the event that they have been tasted previously. Many dates have various sorts, colors, shapes, tastes, and healthy benefits [1]. For ranchers to protect the right assortments, the shoots are shipped, appropriated, and planted once more. Curiously, in regards to palm trees, manors use seeds to develop palms, and another assortment seldom becomes like the mother.

Date natural product, which has numerous assortments all through the world, is utilized in the development of food, clinical, and restorative items. Well-qualified assessment is expected to recognize date natural product assortments because of various healthy benefit, different utilization times, various costs, and quality contrasts [2] .

Every district of the realm is renowned for a couple of kinds of dates. There might be interesting species that don't have business esteem and are not boundless among ranchers. This type might have a high dietary benefit and quality, and this assortment is excluded from the data set. There is no examination with some other organic product as well as how much creation of dates. With this block, there is practically no exploration in the characterization, recognizable proof, and acknowledgment of dates [9]. Improvement of this examination might twofold the amount of creation and the worth of deals and give a reasonable logical impression of the dietary benefit of dates, which should be followed during utilization. Building data sets for dates experimentally considers examination into improvement and coherence, and in this manner for a logical, financial advantage.

II. METHODOLOGY

Order models are a technique for high significance utilized in different fields. In class assurance, grouping models are utilized to figure out which class the information has a place with. The order model is a model that works by making expectations. In our review, models were made involving Naïve Bayes and Multinomial Naïve Bayes frameworks to group date fruit as per their qualities.

2.1 Naïve Bayes

Naive Bayes classification is a popular machine learning algorithm that is based on Bayes' theorem with an assumption of independence between the features. It is a simple yet effective probabilistic model used for classification tasks [3].

The algorithm is called "naive" because it assumes that the presence or absence of a particular feature is unrelated to the presence or absence of other features. In other words, it assumes that all features are independent of each other, which is not

always true in real-world scenarios. Despite this simplifying assumption, Naive Bayes often performs well in practice and can provide reliable results [4][6].

The Naive Bayes algorithm works by calculating the probabilities of a sample belonging to each possible class based on the observed feature values. It then assigns the sample to the class with the highest probability. The calculation of these probabilities involves estimating the likelihood of each feature given each class and the prior probability of each class [7][8].

The algorithm is particularly useful when working with high-dimensional datasets and when the assumption of feature independence is reasonable. It is known for its computational efficiency and is often used in text classification, spam filtering, sentiment analysis, and other similar tasks [5][10].

2.2 Multinomial Naïve Bayes

The Multinomial Naïve Bayes is one of the variations of the Naive Bayes calculation in AI. It is exceptionally valuable to use on a dataset that is circulated multinomially. This calculation is particularly liked in arrangement errands in view of regular language handling [7]. Spam discovery is one of the applications where this calculation can be utilized. The multinomial Naive Bayes is broadly utilized for doling out archives to classes in view of the factual examination of their items. It gives an option to the "weighty" Simulated intelligence based semantic examination and definitely improves on printed information order.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The investigations have been coordinated by using Weka. The Weka is an open-source software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problem. The Date Fruit dataset used in this review was procured from the UCI data repository [11]. The dataset under study consists of 898 samples and 35 elements recorded and 7 label identifying the species of the date fruits class. In total, 34 features were extracted for each date fruit based on 3 main features, which are morphological, shape, and color features. The standard dataset is distributed two sets one for preparing (70%) and one more set for testing (30%). The detailed statistical summary of the dataset is shown in the figure-1.

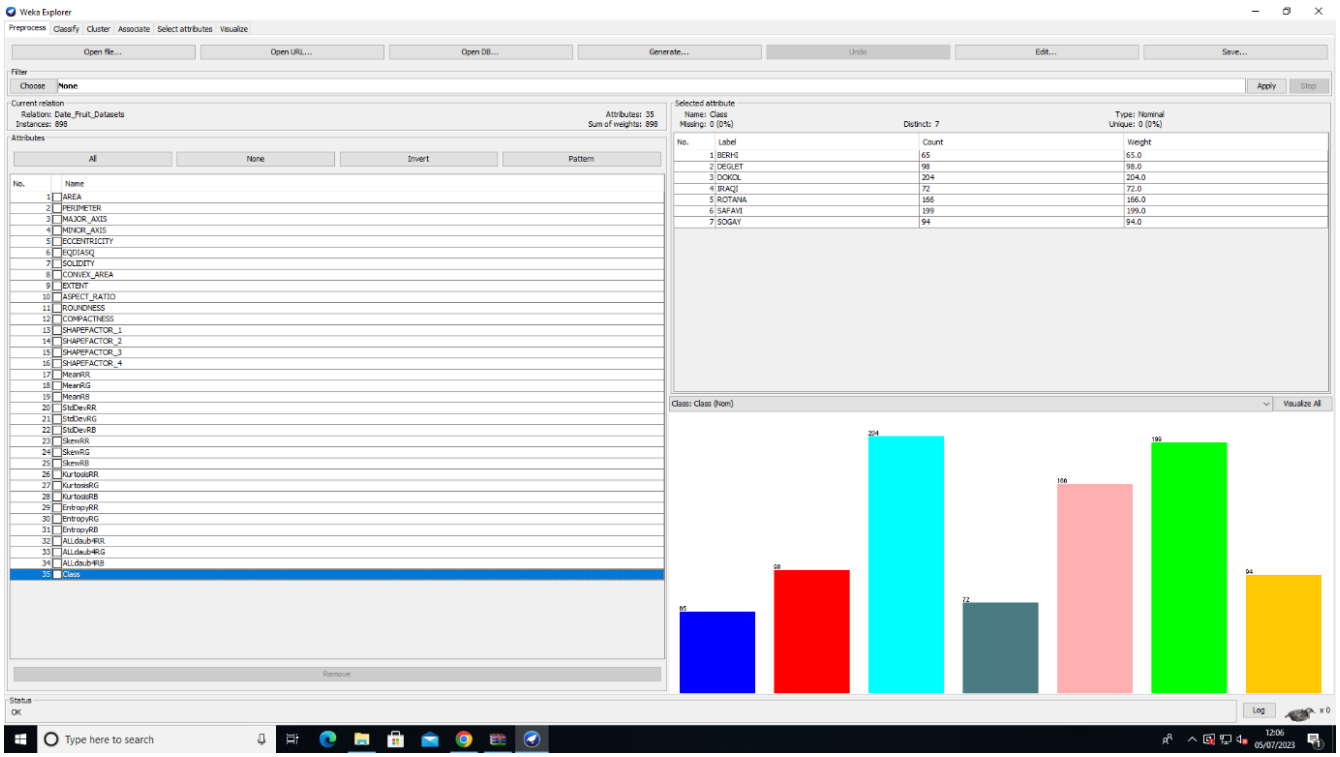

Figure-1: Statistical information of Date Fruit Datasets

In this study, we performed a comprehensive evaluation of Naive Bayes and Naive Bayes Multinomial algorithms for the task of classifying date fruits. The experiments were conducted on a dataset comprising various features and corresponding labels of date fruits. The performance of the algorithms was assessed using three evaluation metrics: accuracy, precision, and recall. The results obtained from the experiments are summarized are presented in the table-1 and same shown in the figure-2.

Table-1
Experimental Results

Algorithm	Accuracy	Precision	Recall
Naive Bayes	88.08	88.4	88.1
Naive Bayes Multinomial	87.75	88.2	97.8

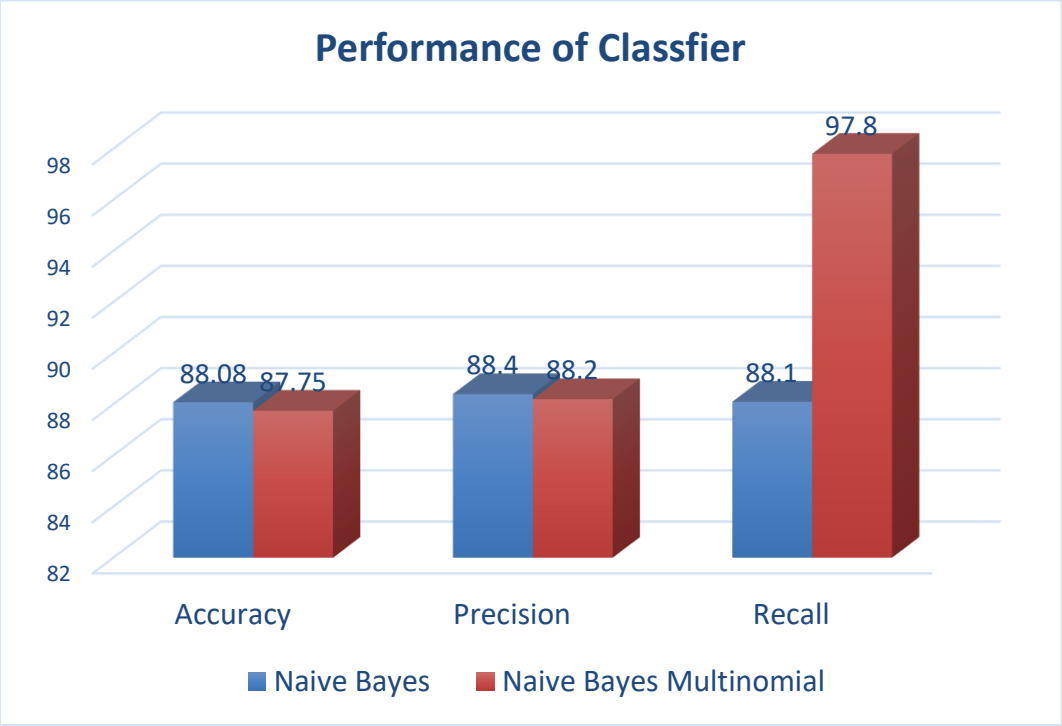

Figure-2: Experimental results of Classifiers

The results indicate that both algorithms achieved reasonably high accuracy rates, indicating their potential for date fruit classification. Naive Bayes exhibited slightly higher accuracy (88.08%) compared to Naive Bayes Multinomial (87.75%).

Looking at precision, Naive Bayes achieved a precision rate of 88.4%, indicating a low false positive rate in the classification of date fruits. On the other hand, Naive Bayes Multinomial achieved a precision rate of 88.2%. Although slightly lower than Naive Bayes, it still demonstrates a good ability to accurately identify date fruits.

In terms of recall, Naive Bayes achieved a recall rate of 88.1%, which signifies the algorithm's capability to correctly identify positive instances of date fruits. Naive Bayes Multinomial, however, showed a significantly higher recall rate of 97.8%, indicating its ability to capture a larger proportion of positive instances.

These results suggest that Naive Bayes Multinomial has a higher recall rate due to its ability to handle discrete data more effectively, which aligns with the assumption that features in the dataset were represented using a multinomial distribution. The results obtained from the experiments, screen shots are shown from figure-3 and figure-4.

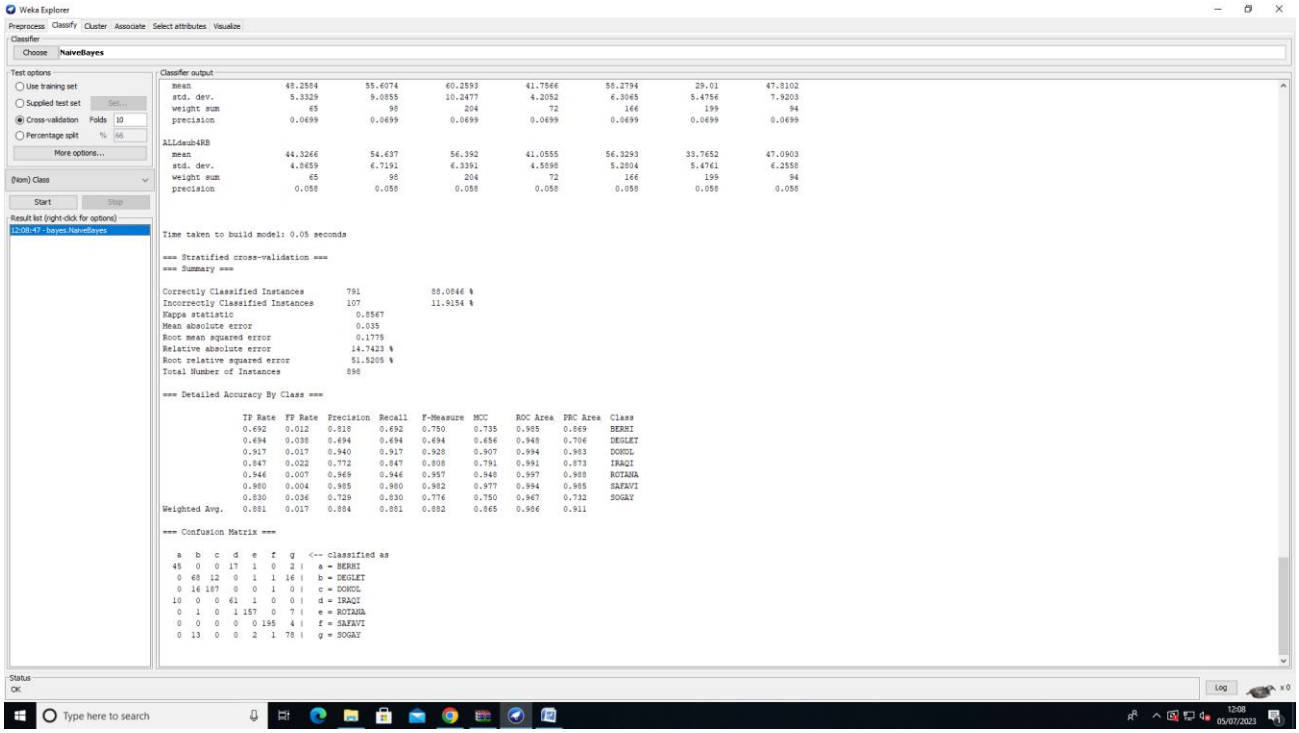

Figure-3: Experimental results of Naive Bayes

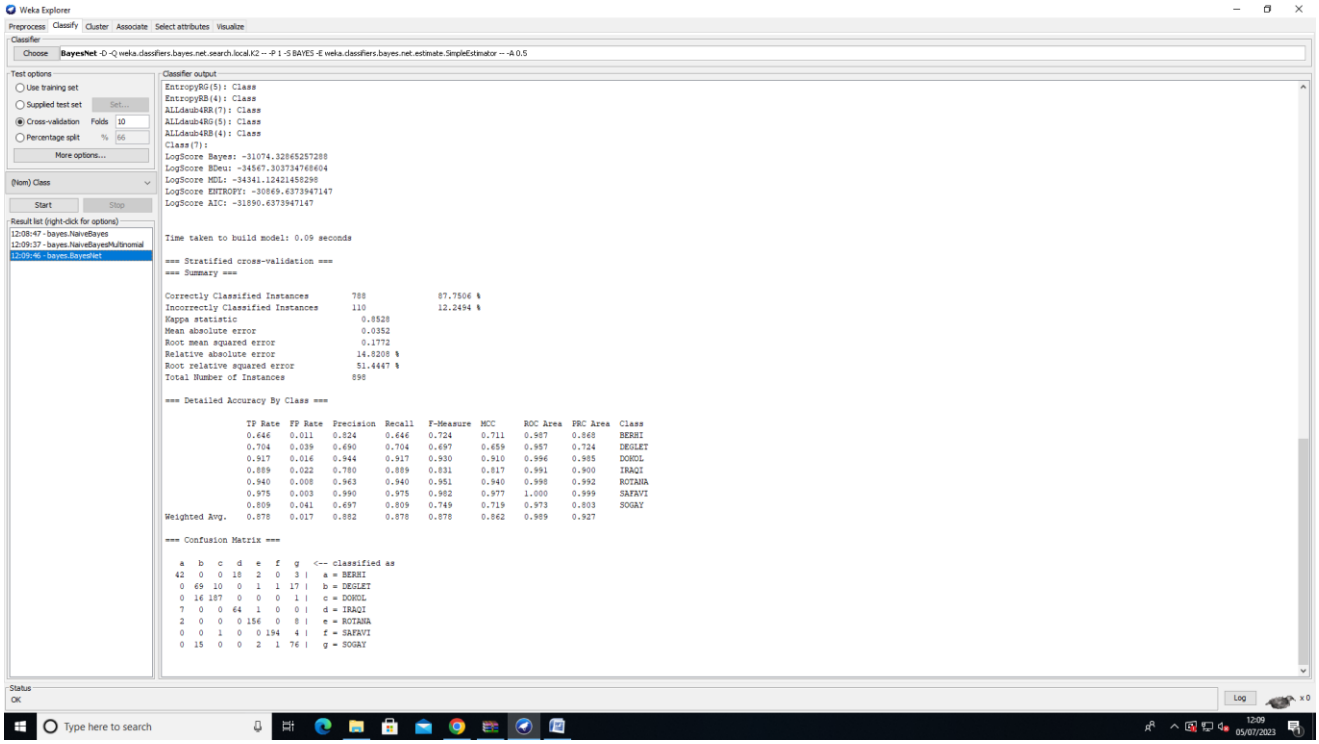

Figure-4: Experimental results of BayesNet

Overall, both Naive Bayes and Naive Bayes Multinomial algorithms demonstrate promising results for date fruit classification. However, the choice of algorithm should depend on the specific requirements of the application. If higher recall rates are crucial, Naive Bayes Multinomial may be preferred, while Naive Bayes could be suitable for applications with a greater emphasis on precision.

IV. CONCLUSION

The findings of this study contribute to the understanding of different Naive Bayes variants and their performance in the classification of date fruits. Future research could explore the incorporation of additional features or consider other classification algorithms to further enhance the accuracy and overall performance of date fruit classification systems.

In conclusion, this study highlights the comparative analysis of Naive Bayes and Naive Bayes Multinomial algorithms for date fruit classification, providing insights into their performance based on accuracy, precision, and recall metrics. These findings can serve as a valuable reference for researchers and practitioners working in the field of fruit classification and related applications.

REFERENCES

[1] Altaheri, H.; Alsulaiman, M.; Muhammad, G. Date Fruit Classification for Robotic Harvesting in a Natural Environment using Deep Learning. IEEE Access 2019, 7, 117115–117133.

[2] Ammari, A.; Khriji, L.; Awadalla, M. HW/SW Co-‘esign For Dates Classification on Xilinx Zynq Soc. In Proceedings of the IEEE International Conference on Open Innovations Association (FRUCT), Yaroslavl, Russia, 20–24 April 2020.

[3] D. Hand, H. Mannila, P. Smyth.: Principles of Data Mining. The MIT Press. (2001)

[4] G. Ravi Kumar, P. Murthuja, G. Anjan Babu, and K. Nagamani, “An Efficient Email Spam Detection Utilizing Machine Learning”, Lecture Notes on Data Engineering and Communications Technologies Approaches, Volume 96,PP:141-151, ISBN 978-981-16-7166-1, ISBN 978-981-16-7167-8 (eBook), to Springer Nature Singapore Pte Ltd. 2022.

[5] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Reddy, “An Improved Decision Tree Classification Approach for Expectation of Cardiotocogram”, Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, Springer Nature Singapore Pte Ltd. 2021, PP:327-333, ISBN 978-981-33-4967-4

[6] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques.2nd ed. San Francisco: Morgan Kaufmann, 2005.

[7] J. Han and M. Kamber,” Data Mining concepts and Techniques”, the Morgan Kaufmann series in Data Management Systems, 2 nd ed. San Mateo, CA; Morgan Kaufmann, 2006.

[8] M. V. Lakshmaiah, Dr. G. Ravi Kumar and Dr. 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, Issue 1(I), PP:6-9, ISSN 2394 -7780, January-March 2015

[9] Nasiri, A.; Garavand, A.; Zhang, Y. Image-Based Deep Learning Automated Sorting of Date Fruit. Postharvest Biol. Technol. 2019, 153, 133–141.

[10] N.Michael, “Artificial Intelligence – A Guide to Intelligent Systems”, 2nd Edition, Addison Wesley 2005

[11] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/ datasets.