# AN ENHANCED FINANCIAL DEVELOPMENT ON FOREIGN DIRECT INVESTMENTS USING NOVEL DATA MINING TECHNIQUE BY APRIORI OVER KNN

**Rajesh.A[1], S.Ashok Kumar[2*]**

**Article History: Received:** 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

**Abstract**

**Aim:**The main objective of this research article is an enhanced financial development on foreign direct investments using a data mining technique named novel Apriori algorithm in comparison with K-Nearest Neighbor (KNN).

**Materials & Methods:** The dataset in this paper utilizes the publicly available dataset from the National financial development to prove the effectiveness of the approach. The sample size of an enhanced financial development on foreign direct investments was sample 280 (Group 1=140 and Group 2 =140) and Calculation is done using G-power 0.8, and the alpha and beta values are 0.05 and 0.2, respectively, with a 95 percent confidence level.The enhanced financial development on foreign direct investments using data mining techniques is performed by novel Apriori algorithm whereas a number of samples (N=10) and K-Nearest Neighbor (KNN) where a number of samples (N=10).

**Results:** The Apriori algorithm classifier has a 95.68 higher accuracy rate when compared to the accuracy rate of K-Nearest Neighbor (KNN) is 92.46.Two groups are statistically significant in the study, as indicated by the significance value of p=0.035 (p0.05).

**Conclusion:** Apriori algorithm provides a better outcomes inaccuracy rate when compared to K-Nearest Neighbor (KNN) for an enhanced financial development on foreign direct investments using data mining techniques.

**Keywords:** Financial Development, Novel Apriori algorithm, K-Nearest Neighbor (KNN), Accuracy rate, Foreign Direct Investments, Data Mining Technique.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, pin: 602105

[2*]Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu, India, pin: 602105

## 1. Introduction

Foreign investment, often known as foreign direct investment (FDI), is the net inflow of funds used to purchase a long-term ownership stake in a business that operates in a country other than the investor's own (Bajpai 2009). One of the most effective strategies for developing nations to increase capital formation, create jobs, and support growth and development is to attract foreign investment (Lipsey 2001).The host country and the investment country are the two parties involved in FDI, and it is beneficial to both of them. When deciding which host country to invest in, certain host country characteristics are essential for the investing country to consider. To forecast the future trajectory of FDI inflows to the host country, it is crucial to look into prospective FDI factors and their relationships (Economou et al. 2017). For each nation, the FDI series is only partially available. Therefore this paper proposes a novel Apriori algorithm for the financial development on foreign direct investment (FDI) and compares the results with the K-Nearest Neighbour (KNN) algorithm (Sari, Triayudi, and Sholihati 2019). The experimental findings demonstrate that the suggested Apriori method outperforms the KNN method in terms of economic status prediction.Due to its practical significance, FDI has received extensive research in a number of disciplines, including economics, mathematics, and statistics.

Many recent publications in the fields of finance, economics, and data mining provide details and a context for our investigation.(Kunnathuvalappil Hariharan 2018; Chang, McAleer, and Wong 2015). IEEE Explore published 89 research papers, and Google Scholar found 93 articles. Arel-Bundock (Arel-Bundock 2017) used non-parametric machine learning models to investigate the impact of 31 political determinants on FDI inflow at the firm level. They also looked at the predictive power of adaptive boosting, extra tree, logistics, and RF regression in the context of FDI inflow and its political determinants. Using 58 economic variables, Bruneckiene et al. (Bruneckiene et al. 2019) examined the investment attractiveness of Europe from 2001 to 2018. They demonstrated that the ANN is a precise and quick method for handling a large number of variables more effectively when analyzing investment attractiveness.Jiménez and Herrero (Jiménez and Herrero 2019) explored the selection of FDI features in Spain and used RF regression to identify the characteristics that would encourage an international corporation to make an investment in another nation. The characteristics of FDI were described by Pekarskiene and Susniene (Pekarskiene and Susniene 2015) as one of the effects and driving forces of globalization. According to Niels Hermes and Robert Lensink (Hermes and Lensink 2003), the process of technical diffusion brought about by FDI benefits from a more developed financial system. To ascertain whether there is a causal connection between FDI and the degree of financial development, Zukarnain Zakaria (Zakaria 2007) conducts a thorough investigation. To investigate this causation in a multivariate context, the author analyses data from 37 developing nations.Ho et al. (Ho et al. 2012) in their work have developed a fuzzy-based association rule mining technique that was used for financial marketing. Their technique was intended to support the investors to derive rules for investing in the financial market, since predicting the volatile financial market was very difficult. The fuzzy-based (Jiang, Chen, and Wen 2014) association rule mining technique was applied on Hang Seng Index of Hong Kong which was associated with factors including the GDP, the CPI, the interest rate, and the value of the export of commodities. Wu and Huang (Wu and Huang 2011) have worked on developing an efficient data structure for mining association rules.Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

The fundamental issue with the current KNN approach is that it performs poorly when the training set is huge in terms of run-time. Because all features contribute to similarity and subsequently to classification, it is extremely sensitive to irrelevant or redundant features.This paper proposes a robust technique for the financial development of FDI using the novel Apriori algorithm in comparison with KNN algorithm. Apriori algorithm is to better understand the effectiveness of financial growth of FDI. The proposed Apriori method outperforms the KNN algorithm, according to performance analysis.

## 2. Materials and Methods

This study was done in the DBMS Laboratory of the Saveetha School of Engineering's Department of Computer Science Engineering. For the 40 nations included in this study's dataset, annual measures of FDI as a share of GDP were made from 2005 through 2020.Out of the 40 countries in the sample, 20 are developed and 20 are in the developing world.The database is divided by the amount of 75% training and 25% testing. Group 1 was a K-Nearest Neighbour (KNN) algorithm and

Group 2 was a novel Apriori algorithm.G-power 0.8, alpha and beta quality of 0.05 and 0.2, and a confidence interval of 95% are used in the calculation (Koyuncugil and Ozgulbas 2012). These datasets required Python software with the necessary library functions and tool functions, as well as a display with a resolution of 1024 x 768 pixels (CPU 7th generation, i5, 4 8GB RAM, 500 GB HDD). The system under test for the proposed implementation using the IBM SPSS V26.0 tool for statistical analysis

**K-Nearest Neighbor (KNN)**

When little or no prior knowledge of the distribution of the data is available, the KNN is the fundamental and most basic classification method. The complete training set is simply kept in memory throughout learning, and each query is given a class based on the training set's k-nearest neighbor's label that receives the most votes. When K = 1, the simplest variant of the KNN is the Nearest Neighbor rule (NN). Each sample in this method must be categorized in a manner that is consistent with the samples around it. As a result, if a sample's classification is unknown, its closest neighbor samples' classifications can be used to predict the classification of the sample in question.All the distances between an unknown sample and all the samples in a training set can be calculated given an unknown sample and a training set. The sample in the training set that is closest to the unknown sample is identified by the distance with the smallest value. As a result, the classification of the unknown sample could be predicated on that of its closest neighbour.(Domeniconi, Peng, and Gunopulos 2002). Unfortunately, a large value of K easily causes the estimate to exceed smoothing, and the classification accuracy suffers when outliers from other classes are included. The average of the K nearest neighbor result is the KNN prediction:

$$y = \frac{1}{K} \sum_{i=1}^{k} y_i \qquad (1)$$

where y is the expected result (outcome) of the query point and yi is the ith case of the example sample.

**Pseudocode:**

Input: An improved financial development on foreign direct investments _Input_Features
Assign Training and Testing for an improved financial development on foreign direct investments Dataset
Output: Classification on Fixed Number of an improved financial development on foreign direct investments (Class labels C)
Function: K-means Nearest Neighbor (Input features I)
Step 1: Initialize Training dataset T
Step 2: Initialize the Class labels C of T
Step 3: Initialize the Unknown data items U of T
Step 4: While (condition) do
Step 3: Compute distance of each foreign direct investments of Input features I with distance(C, U)
Step 4: End while
Step 5: Sort the correlated foreign direct investments of Input features I distances by ascending order
Step 6: Calculate the number of occurrences of each foreign direct investments of Input features I with class labels 'C' among the 'k' nearest neighbor
Step 7: Assign the Unknown samples of T to the most frequent class labels 'C'
Return Classification outcomes of Class labels 'C'

**Apriori Algorithm:**

Apriori is the well-known algorithm used for mining associations among items in the huge stock market databases (Prasanna and Ezhilmaran 2016). A general apriori is designed to operate on databases containing transactions. Initially the frequent itemsets for the entire databases are created by minsupport values. Then through these frequent itemsets and minconfidence values, the best stock rules were determined from the dataset. In our proposed work, the enhanced apriori algorithm has been applied for mining association rules and the number of frequent items set is 7 because the data was calculated for seven days. Each frequent item set is calculated from its prior values with minsupport threshold. So, unwanted rules are eliminated by this method in each transaction and the time computation complexity to find the stock rule was minimized compared to normal apriori algorithm. This resultant mined data will be fed to the GA for the purpose of generating fine stock rules.

**Pseudocode :**

Input: An improved financial development on foreign direct investments _Input_Features
Assign Training and Testing for an improved financial development on foreign direct investments Dataset
Output: Classification on Fixed Number of an improved financial development on foreign direct investments
Function: An Apriori algorithm (Input features F, Label vector V=[1.....n])
Step 1: Divide the cluster into groups; set the term to be n and the colors to k; this will cause the loop to be run k times.
Select each cluster individually now.
Step 2: Create items, the support for those items, and transaction ids.
Step3: For (k = 2; Fk-1 ; k++);
Step4: For (F1 = find frequent 1 itemsets (T)).
Step:Create the Ck from the Fk-1 .
Step 6: Ck represents candidates produced from Fk-1.

Step 7: Use F1, (1 w k), to get the item Iw in Ck with the least amount of assistance.
Step8: x = Get item min sup (Ck, F1).
Step9: Retrieve the target transaction IDs that contain item x.
Step10:Target = get_ Transaction_ ID(x);
Step11: Perform each transaction in Target for each transaction t.
Step12: Aim to increase the total number of goods in Ck that were discovered in Target.
Step13: Fk
The new Apriori algorithm exhibits numerous improvements over the traditional method. In comparison to the KNN technique, we find that the Apriori algorithm more accurately predicts FDI inflows, inbound and outward stock flows. Prediction accuracy is measured using a variety of metrics, including accuracy, mean absolute error, and root mean square error. The following is their mathematical formulation:

$$MSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (ai - pi)^2} \quad (2)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |ai - pi| \quad (3)$$

$$Accuracy = \frac{N_c}{N} \quad (4)$$

Here, N stands for the total number of projected records, ai for a record's actual value, pi for a record's predicting value, and Nc for the number of records that are correctly classified.
Determinants of FDI in developing and developed nations in Asia, Africa, Latin America, and Europe have been found in the literature (Mottaleb and Kalirajan 2010; Demirhan, Masca, and Others 2008; Addison and Heshmati 2003). Using information from the London, New York, and Karachi stock exchanges, Khan et al. (Khan et al. 2016) tried to forecast the stock market's future tendencies. KNN, Naive Bayes, and K-Nearest Neighbor were used by Khan et al (KNN). Khedr et al. (Khedr, Yaseen, and Others 2017) made an effort to create a two-stage model based on financial market sentiment analysis and historical stock market prices to forecast future trends in the stock market with a lower error and accuracy rate.The naive Bayes method is used in the initial analysis of news sentiment. A method for mining association regulations to enhance the calibre of the production process has been put out by Foguem et al. (Kamsu-Foguem, Rigal, and Mauget 2013).

**Statistical Analysis**

The output is obtained by using Google Colab (Singh and Manure 2020). Python programme with the necessary library functions and tool functions, as well as a display with a resolution of 1024 x 768 pixels (CPU 7th generation, i5, 4 8GB RAM, 500 GB HDD), were needed to train these datasets. According to Koyuncugil and Ozgulbas (2012), the computation is done using a G-power of 0.8, alpha and beta quality of 0.05 and 0.2, and a confidence interval of 96%. IBM SPSS V26.0 is the programme utilised in this instance to implement the statistics.The independent sample t test was used to determine the mean, standard deviation, and standard error mean statistical significance between the groups. After comparing the two groups using SPSS software, the significant value and standard deviation values were calculated with the graph's maximum accuracy value (95.68 percent) (0.21929).Dependent variables are accuracy and independent variables are defined as Apriori and K-Nearest Neighbor algorithms.

## 3. Results

Figure 1 shows the simple Bar graph for novel Apriori algorithm accuracy rate is compared with K-Nearest Neighbor (KNN). The Apriori algorithm is higher in terms of accuracy rate 95.68 when compared with K-Nearest Neighbor (KNN) 92.46. Apriori algorithms produce variable outcomes with standard deviations that range from 100 lower to 150 higher, whereas K-Nearest Neighbor (KNN) algorithms have standard deviations that range from 200 lower to 300 higher. The K-Nearest Neighbor (KNN) algorithm and the Apriori algorithm differ significantly (p0.05 Independent sample test). K-Nearest Neighbor (KNN) accuracy rate versus Apriori algorithm, on the X-axis. Y-axis: Mean accuracy rate for identifying keywords, with a 95% confidence interval.
Table 1 portrays the Evaluation Metrics of Comparison of novel Apriori algorithm and (KNN) Classifier. The Apriori algorithm's accuracy rate is 95.68 and K-Nearest Neighbor (KNN) has 92.46. In all aspects of parameters Apriori algorithm provides better performance compared with the K-Nearest Neighbor (KNN) of an enhanced financial development on foreign direct investments using novel data mining techniques.
Table 2 demonstrates the statistical calculations for the innovative Apriori method and K-Nearest Neighbor, including Mean, standard deviation, and standard error Mean (KNN).The accuracy rate parameter used in the t-test. The Apriori algorithm's average accuracy rate is 95.68 and K-Nearest Neighbor (KNN) is 92.46. The Standard Deviation of Apriori algorithm is 0.21029 and K-Nearest Neighbor (KNN) is 1.77283. The Standard Error Mean of Apriori algorithm is 0.10291 and K-Nearest Neighbor (KNN) is 0.88193.
Table 3 displays the statistical calculations for independent samples tested between novel Apriori algorithm and K-Nearest Neighbor (KNN). The significance value for accuracy p = 0.035, which shows that two groups are statistically significant..

## 4. Discussion

In this study, the proposed model exhibits the Apriori algorithm and K-Nearest Neighbor (KNN), in which the Apriori algorithm has the highest values. The accuracy Rate of Apriori algorithm is 95.68% higher compared with the K-Nearest Neighbour (KNN) that has an accuracy rate of 92.46% in analysis of an enhanced financial development on foreign direct investments using novel data mining techniques.

The Apriori method and K-Nearest Neighbor (KNN) were used in this study to analyse the effects of financial development variables on foreign direct investment into and out of different nations. Preliminary examination of the financial development variables that can be used to forecast the quantity of foreign direct investment was supplied by feature selection methodologies. We can forecast the range in which the Foreign Direct Investment values can lie given the range of the financial development variables to which it is connected using the association rules on applying the Apriori method. We gauge the health of the stock market using its capitalization as a proportion of GDP, the amount of money traded relative to GDP, and the turnover ratio.Regression study reveals a substantial and positive correlation between stock market capitalisation and FDI stock and inflows to the countries. The biggest obstacle is the expensive time wasted holding a lot of candidate sets with frequent itemsets, little minimum support, or big itemsets.

The complexity of this algorithm can be further reduced in the future by an order, and it can be used to implement mining tools, other data mining processes, and applications in the health care sector, libraries, and grocery stores. Additionally, the running time of this algorithm can be further reduced by employing a different strategy. In the second stage, KNN and SVM are used to combine recent financial data with historical stock market prices in order to forecast future stock market prices. SVM and KNN were utilised by Puspitasari and Rustam (Puspitasari and Rustam 2018) to forecast stock prices on the Indonesia Stock Exchange. To choose the most important financial indicators, the authors first used a feature selection method.In their article Applications of Improved Apriori Algorithm on Educational Information, Quiang Yang, Yanhong Hu, and others (Yang and Hu 2011) addressed the primary issues with the applications and presented an improved algorithm.

## 5. Conclusion

The proposed model exhibits the Apriori algorithm and K-Nearest Neighbour (KNN), in which the Apriori algorithm has the highest values. The accuracy Rate of Apriori algorithm is 95.68% higher compared with the K-Nearest Neighbour (KNN) that has an accuracy rate of 92.46% in analysis of an enhanced financial development on foreign direct investments using novel data mining techniques.

**Declaration**

**Conflicts of Interest**

No conflict of interest in this manuscript

**Authors Contributions**

Data collection, analysis, and manuscript writing were all done by author A.R. Conceptualization, data validation, and paper critical evaluation were all handled by author S.A.K.

**Acknowledgment**

In order to successfully complete this study, the authors would like to express their gratitude to Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University).

**Funding:** We thank the following organizations for providing financial support that enabled us to complete the study.

1. Chipontime Technologies Pvt. Ltd. Bangalore.
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

## 6. References

Chang, Chia-Lin, Michael McAleer, and Wing-Keung Wong. 2015. "Informatics, Data Mining, Econometrics and Financial Economics: A Connection." repub.eur.nl. https://repub.eur.nl/pub/79219/.

Domeniconi, C., Jing Peng, and D. Gunopulos. 2002. "Locally Adaptive Metric Nearest-Neighbor Classification." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (9): 1281–85.

Ho, G. T. S., W. H. Ip, C. H. Wu, and Y. K. Tse. 2012. "Using a Fuzzy Association Rule Mining Approach to Identify the Financial Data Association." *Expert Systems with Applications* 39 (10): 9054–63.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar

Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." *Environmental Research* 212 (Pt A): 113153.

Jiang, Feng, Fuyang Chen, and Changyun Wen. 2014. "Fuzzy Control for Road Intersection Signal with Bus Priority." *Alloy Journal of Soft Computing and Applications (AJSCA)* 1 (1).

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." *Computers and Electrical Engineering*. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." *Korean Journal of Chemical Engineering*. https://doi.org/10.1007/s11814-022-1120-5.

Kunnathuvalappil Hariharan, Naveen. 2018. "Applications of Data Mining in Finance." *OF DATA MINING IN FINANCE …*. https://papers.ssrn.com/abstract=3914436.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." *Carbohydrate Polymers* 287 (July): 119349.

Prasanna, S., and D. Ezhilmaran. 2016. "Association Rule Mining Using Enhanced Apriori with Modified GA for Stock Prediction." *International Journal of Data Mining, Modelling and Management* 8 (2): 195–207.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." *Chemosphere* 298 (July): 134341.

Sari, Yuli Nur Indah, Agung Triayudi, and Ira Diana Sholihati. 2019. "Implementation of Data Mining to Predict Food Sales Rate Method Using Apriori." *International Journal of Computer Applications in Technology* 975: 8887.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." *Chemosphere* 298 (July): 134265.

Singh, Pramod, and Avinash Manure. 2020. "Introduction to Tensorflow 2.0." In *Learn TensorFlow 2.0*, 1–24. Springer.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-022-06636-5.

Wu, Chieh-Ming, and Yin-Fu Huang. 2011. "Generalized Association Rule Mining Using an Efficient Data Structure." *Expert Systems with Applications* 38 (6): 7277–90.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113114.

## TABLES AND FIGURES

Table 1. Accuracy Values for Apriori and KNN.

| **SI.No.** | **Test Size** | **ACCURACY RATE** | |
|---|---|---|---|
| | | Apriori algorithm | K-Nearest Neighbor (KNN) |
| 1 | Test1 | 93.23 | 92.20 |
| 2 | Test2 | 93.54 | 92.53 |
| 3 | Test3 | 94.36 | 91.73 |
| 4 | Test4 | 94.34 | 90.32 |
| 5 | Test5 | 95.12 | 91.72 |
| 6 | Test6 | 94.56 | 92.01 |
| 7 | Test7 | 94.35 | 91.85 |
| 8 | Test8 | 95.36 | 91.28 |
| 9 | Test9 | 95.35 | 91.58 |
| 10 | Test10 | 95.54 | 92.34 |
| Average Test Results | | 95.68 | 92.46 |

Table 2. The statistical calculations for the Apriori method and K-Nearest Neighbor include mean, standard deviation, and standard error (KNN).The accuracy rate parameter used in the t-test. The mean accuracy rate of Apriori algorithm is 95.68 and K-Nearest Neighbor (KNN) is 92.46. The Standard Deviation of Apriori algorithm is 0.21029 and K-Nearest Neighbor (KNN) is 1.77283. The Standard Error Mean of Apriori algorithm is 0.10291andK-Nearest Neighbor (KNN) is equal to 0.88193.

| **Group** | | **N** | **Mean** | **Standard Deviation** | **Standard Error Mean** |
|---|---|---|---|---|---|
| Accuracy Rate | K-NEAREST NEIGHBOR (KNN) | 10 | 92.46 | 1.77283 | 0.88193 |
| | APRIORI ALGORITHM | 10 | 95.68 | 0.21029 | 0.10291 |

Table 3. Apriori method and K-Nearest Neighbor statistical computations for independent samples test (KNN).The sig. for signal to noise ratio is set at 0.035. unbiased samples with a confidence interval of 95% and a level of significance value of p=0.035 (p0.05), the T-test is used to compare the Apriori algorithm with K-Nearest Neighbor (KNN), demonstrating that the two groups are statistically significant.

| Group | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval (Lower) | 95% Confidence Interval (Upper) |
| Accuracy | Equal variances assumed | 6.72 | 0.035 | 17.182 | 18 | .000 | 11.617 | 0.8928 | 11.718 | 13.783 |
| | Equal variances not assumed | | | 17.672 | 12.827 | .000 | 11.437 | 0.7283 | 10.192 | 12.012 |

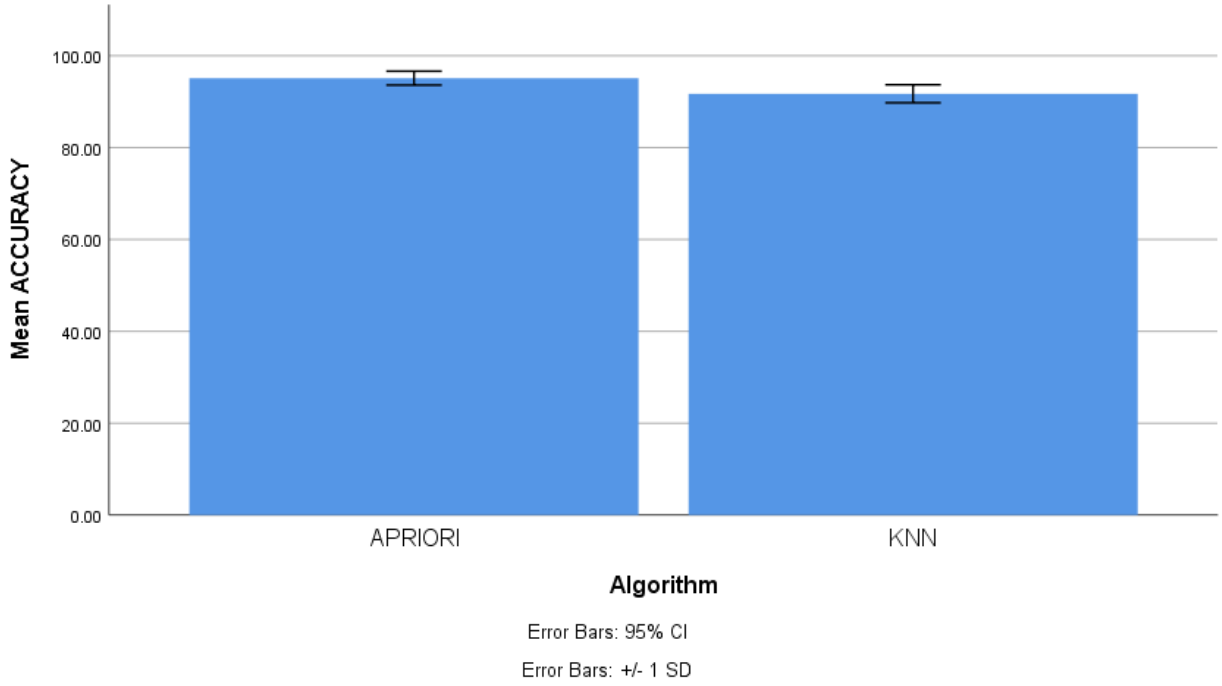


Fig. 1. Simple Bar graph for Apriori algorithm accuracy rate is compared with K-Nearest Neighbor (KNN). The Apriori algorithm is higher in terms of accuracy rate 95.68 when compared with K-Nearest Neighbor (KNN) 92.46. Apriori algorithm produces variable results with a standard deviation that ranges from 100 to 150 higher whereas K-Nearest Neighbor's (KNN) standard deviation is 200 to 300 higher.The K-Nearest Neighbor (KNN) algorithm and the Apriori algorithm differ significantly ($p<0.05$ Independent Sample Test).. X-axis: K-Nearest Neighbor (KNN) accuracy rate vs Apriori algorithm ,Mean accuracy rate on the Y-axis, ±1 SD with a 95% confidence interval for keyword identification.