THE DATA MINING TECHNIQUE TO DEVELOP THE FINANCIAL DEVELOPMENT ON FOREIGN DIRECT INVESTMENTS BY APRIORI OVER K MEANS

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

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

Abstract

Aim: The objective of this research paper is the data mining technique to develop the financial development on foreign direct investments by using Apriori algorithm in comparison with K-means clustering. **Materials & Methods:** The dataset in this paper utilizes the publicly available dataset from National financial development to prove the effectiveness of the approach. The sample size of the data mining technique to develop the financial development on foreign direct investments was sample 280 (Group 1=140 and Group 2 =140) and calculation is performed utilizing G-power 0.8 with alpha and beta qualities are 0.05, 0.2 with a confidence interval at 95%. The data mining technique to develop the financial development on foreign direct investments is performed by Apriori algorithm whereas number of samples (N=10) and K-means clustering where number of samples (N=10).
Results: The Apriori algorithm classifier has a 95.68 higher accuracy rate when compared to the accuracy rate of the K-means clustering is 91.346. The study has a significance value $p=0.029$ ($p<0.05$), which shows that two groups are statistically significant. **Conclusion:** Apriori algorithm provides better outcomes in accuracy rate when compared to K-means clustering for the data mining technique to develop the financial development on foreign direct investments.

Keywords: Financial Development, Novel Apriori algorithm, K-means clustering, 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 direct investment is a versatile concept based on the fact that it plays an important role in a country's development, and there are many aspects related to this issue that need to be carefully examined (N. Prasanna 2010). Foreign investments are a key development factor in the modern economy, and jointly with the trade, represent the most important leverage of an enterprise, organization of production, supplying goods and services on a global scale (Malhotra 2014). The main objective of this paper is to analyze the short and long-run interactions between financial development and Foreign direct investments inflows using a novel Apriori algorithm and compare the results with the conventional K-means clustering algorithm (S. Prasanna and Ezhilmaran 2016). The results indicated that the proposed Apriori method outperformed the K-means clustering method. Foreign Direct Investments is one of the most widely studied and challenging problems, attracting researchers from many fields including economics, history, finance, mathematics and computer science (Tuomi 2011).

A review of some of those works is provided. IEEE Explore published 78 research papers, and Google Scholar found 102 articles. Qodmanan et al. (Qodmanan, Nasiri, and Minaei-Bidgoli 2011) have proposed a multi-objective method for mining association rules. They have mainly concentrated on association rule mining in their work. Gebrehiwot et al. (Gebrehiwot, Esfahani, and Sayim 2016) analyzed the connection between financial development and Foreign direct investments in eight African countries between 1991–2013 employing Granger causality tests and panel regression and revealed a two-way causality between the variables. Bayar and Ozturk (Hatemi-J and Shamsuddin 2016) also examined the causal interaction between financial development and Foreign direct investments inflows in Turkey over the 1974–2015 periods with the bootstrap Granger causality tests of Hacker and Hatemi-J (Hacker and Hatemi-J 2006) and determined a one-way causality from the development of financial sectors to foreign direct investment inflows. 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)

Fauzel (Fauzel 2016) also analyzed the relation between Foreign direct investments and financial development in small island developing states during the 1990–2013 period, using a panel vector autoregressive model and found a bi-causal relationship. Enisan (Enisan 2017) investigated the major determinants underlying Foreign direct investments inflows in Nigeria employing the Markov regime-switching approach and revealed that the development level of a financial sector is one of the main determinants for Foreign direct investments attraction. Popescu and Andreica (Popescu et al. 2016) found that within 28 E.U. countries, labor productivity indicator determines whether the country is classified among high economic performance countries or lower economic performance. Divya and Agarwal (Agarwal 2011) used other techniques to classify 70 countries. The investigated dataset contains macro-economic indicators such as Economic Freedom Index (EFI) and HDI, It allowed the author to divide the countries into three groups; developed, developing and under development countries. Another study realized by Cheng (Cheng 2014) applied the k-means cluster algorithm in the regional economy for a comparative analysis on 30 areas of China.

2. Materials and Methods

This work was carried out in the DBMS Laboratory, Department of Computer Science Engineering, Saveetha School of Engineering. In this paper, the dataset contains a time series with yearly measurements of Foreign direct investments as percentage of Gross domestic product (GDP) for 40 countries during the period 2005 – 2020. Out of 40 sample countries, a total of 20 are the developed countries and 20 are the developing countries. The database is divided by the amount of 75% training and 25% testing. Group 1 was a K-means clustering algorithm and Group 2 was a novel Apriori algorithm. The calculation is performed utilizing G-power 0.8 with alpha and beta qualities 0.05, 0.2 with a confidence interval at 95% (S. Singh 2019). These datasets, required a monitor with resolution of 1024×768 pixels (CPU 7th gen, i5, 4 8GB RAM, 500 GB HDD), and Python software with required library functions and tool functionsThe testing setup for the proposed system to implement with the IBM SPSS V26.0tool used for statistical analysis.

K-Means Clustering

The K-means algorithm is known for being one of the simplest clustering algorithms. The concept of this algorithm is to divide the given data into a predefined number of groups i.e., k groups or k clusters. Each cluster has a center, far from the other centers. Each center defines the characteristics of the data belonging to this cluster, therefore, it helps distribute every other record in the nearest convenient cluster. The function of partitioning for the objects $(x1, x2, \dots, xn)$, and n

points that should be distributed in k clusters, it calculates the distance between the point and the centroid, and assigns the object to the smallest distance.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \|x_i^{(j)} - c_j\|^2 \qquad (1)$$

Where$\|x_i^{(j)} - c_j\|^2$ is the distance between a point (j) and the centroid cj. Once all the objects are distributed, it recalculates the position of centroids until the k clusters centers are not changing.

Pseudocode:
Input:
M: Dataset with 'n' instances
K: number of clusters.
Output:
Dataset partitioned into 'K' clusters
Algorithm:
1. Choose arbitrarily 'K' random points from M as the cluster centers
2. repeat
3. Reassign each object to the clusters based on calculation of mean value using Euclidean distance measure.
4. Revise the cluster means by recalculating the mean value of each cluster
5. Until
6. there is no change in the clusters obtained
7. Evaluation using silhouette coefficient: Calculate silhouette coefficient for an object as below:
Si = (bi – ai) / max (ai , bi)
Where,
Si - silhouette coefficient
ai - is the mean value of an intra-distance cluster to all objects in a cluster.
bi – minimum average distance from ith object to other clusters.
8. Determine the mean silhouette coefficient over all objects.

Apriori Algorithm

Apriori association rule is used to harness the frequent patterns in databases. Support and confidence are the normal methods used to measure the quality of association rule. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). The algorithm attempts to find subsets which are common to at least a minimum number C (the cutoff, or confidence threshold) of the item-sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time, a step known as candidate generation, and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. Apriori uses breadth-first search and a hash tree structure to count candidate itemsets efficiently.

Pseudocode :
Input: Database of Transactions D= {t1, t2, …, tn}
Set if Items I= {I1, I2,…., Ik}
Frequent (Large) Itemset L Support, Confidence.
Output: Association Rule satisfying Support & Confidence
Method:
1. C1 = Itemsets of size one in I;
2. Determine all large itemsets of size 1, L1;
3. i = 1;
4. Repeat
5. i = i + 1;
6. Ci = Apriori-Gen(Li-1);
7. Apriori-Gen (Li-1)
- Generate candidates of size i+1 from large itemsets of size i.
- Join large itemsets of size i if they agree on i-1.
- Prune candidates who have subsets that are not large.

8. Count Ci to determine Li;
9. until no more large itemsets found;

. The performance of these Apriori based data mining models was found to be far too superior to that of the K-means clustering model. The study has conclusively elicited the fact that Apriori models have much higher capability in extracting and learning the features of a time series data than their corresponding k-means counterparts. Accuracy is defined as the ratio of the number of cases which are correctly classified to the total number of cases expressed as a percentage.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \qquad (2)$$

The time taken for generating the stock rules by apriori algorithm was better than using K-means clustering algorithm because by apriori algorithm, the superior stock rules are only tolerated in the database. Apart from this, these rules were eliminated in each frequent itemset database.

Statistical Analysis

The output is obtained by using Google Colab (P. Singh and Manure 2020). To train these datasets, required a monitor with resolution of 1024×768 pixels (CPU 7th gen, i5, 4 8GB RAM, 500 GB HDD), and Python software with required library functions and tool functions. For statistical implementation, the software tool used here is IBM SPSS V26.0 (Frey 2017). The independent sample t test was performed to find the mean, standard deviation and the standard error mean statistical significance between the groups, and then comparison of the two groups with the SPSS software will give the accurate values for the two different parameters which will be utilized with the graph to calculate the significant value with maximum accuracy value (95.68%), mean value (95%) and standard deviation value (0.21929).

Dependent variables are accuracy and independent variables are defined as Apriori and K-means algorithms.

3. Results

Figure 1 shows the simple Bar graph for Apriori algorithm accuracy rate is compared with K-means clustering. The Apriori algorithm is higher in terms of accuracy rate 95.68 when compared with K-means clustering 91.346. Variable results with its standard deviation ranging from 100 lower to 150 higher Apriori algorithms where K-means clustering standard deviation ranging from 200 lower to 300 higher. There is a significant difference between Apriori algorithm and K-means clustering ($p<0.05$ Independent sample test). X-axis: K-means clustering accuracy rate vs Apriori algorithm Y-axis: Mean of accuracy rate, for identification of keywords ± 1 SD with 95 % CI.

Table 1 portrays the Evaluation Metrics of Comparison of Apriori algorithm and K-means clustering Classifier. The accuracy rate of the Apriori algorithm is 95.68 and K-means clustering has 91.346. In all aspects of parameters Apriori algorithm provides better performance compared with the K-means clustering of the data mining technique to develop the financial development on foreign direct investments.

Table 2 shows the statistical calculation such as Mean, standard deviation and standard error Mean for Apriori algorithm and K-means clustering. The accuracy rate parameter used in the t-test. The mean accuracy rate of Apriori algorithm is 95.68 and K-means clustering is 91.346. The Standard Deviation of Apriori algorithm is 0.21029 and K-means clustering is 1.81738. The Standard Error Mean of Apriori algorithm is 0.10291 and K-means clustering is 0.71893.

Table 3 displays the statistical calculations for independent samples test between Apriori algorithm and K-means clustering. The results achieved with $p=0.029$ ($p<0.05$), which shows that two groups are statistically significant. Independent samples T-test is applied for comparison of Apriori algorithm and K-means clustering with the confidence interval as 95% and level of significance as 0.32132. This independent sample test consists of significance as 0.001, significance (2-tailed), Mean difference, standard error difference, and lower and upper interval difference.

4. Discussion

In this study, the proposed model exhibits the Apriori algorithm and K-means clustering, in which the Apriori algorithm has the highest values. The accuracy Rate of Apriori algorithm is 95.68% higher compared with the K-means clustering that has an accuracy rate of 91.346% in analysis of the data mining technique to develop the financial development on foreign direct investments..

According to Wu and Kumar (Wu et al. 2008), the K-means algorithm is simple, easily understandable, and can be readily adjusted to deal with different scenarios, such as semi-supervised learning or streaming data. Similarly, Sohn and Kim (Ahn and Sohn 2009) claim that some clustering algorithms generate too many clusters, however the K-means algorithm can possibly be used to derive a set of desirable clusters. The K-means algorithm has been increasingly applied in the financial area, and helps financial institutions to effectively manage portfolio and asset selection. According to Nanda et al. (Nanda, Mahanty, and Tiwari 2010), the K-means algorithm can reduce the time for selection of stock, as compared to fuzzy C-means and the Self Organizing Map (SOM). Liao et al. (Liao, Ho, and Lin 2008) suggest that the confidence level should be set above 70 and a support level above 80 is considered meaningful. Liang and Christos (Dong and Tjortjis 2003) point out that a number of frequent itemsets decrease as the minimum support increases. As a result, fewer rules are found and thus the time spent in each phase decreases.

In future, work can be carried out with some other data mining algorithm in terms of time factor.The limitation of the proposed Apriori algorithm is very low and inefficient when memory capacity is limited with a large number of transactions.

5. Conclusion

The proposed model exhibits the Apriori algorithm and K-means clustering, in which the Apriori algorithm has the highest values. The accuracy Rate of Apriori algorithm is 95.68% higher compared with the K-means clustering that has an accuracy rate of 91.346% in analysis of the data mining technique to develop the financial development on foreign direct investments.

Declaration

Conflicts of Interest

No conflict of interest in this manuscript

Authors Contributions

Author A.R was involved in data collection, data analysis & manuscript writing. Author S.A.K was involved in conceptualization, data validation, and critical review of manuscripts.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for successfully carrying out this work.

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

5. References

Agarwal, Sonali. 2011. "Classification of Countries Based on Macro-Economic Variables Using Fuzzy Support Vector Machine." *International Journal of Computer Applications in Technology* 975: 8887.

Ahn, Jin Sook, and So Young Sohn. 2009. "Customer Pattern Search for after-Sales Service in Manufacturing." *Expert Systems with Applications* 36 (3, Part 1): 5371–75.

Cheng, Zhuang. 2014. "Regional Economic Indicators Analysis Based on Data Mining." In *2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications*, 726–30.

Dong, Liang, and Christos Tjortjis. 2003. "Experiences of Using a Quantitative Approach for Mining Association Rules." In *Intelligent Data Engineering and Automated Learning*, 693–700. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg.

Enisan, Akinlo A. 2017. "Determinants of Foreign Direct Investment in Nigeria: A Markov Regime-Switching Approach." *Review of Innovation and Competitiveness: A Journal of Economic and Social Research* 3 (1): 21–48.

Fauzel, Sheereen. 2016. "Modeling the Relationship between FDI and Financial Development in Small Island Economies: A PVAR Approach." http://repository.embuni.ac.ke/bitstream/handle/123456789/1718/OJS20120300005_28255412.pdf?sequence=1.

Frey, Felix. 2017. "SPSS (Software)." *The International Encyclopedia of Communication Research Methods*, November, 1–2.

Gebrehiwot, Aklilu, Naser Esfahani, and Mustafa Sayim. 2016. "The Relationship between FDI and Financial Market Development: The Case of the Sub-Saharan African Region." *International Journal of Regional Development* 3 (1): 1–64.

Hacker, R. Scott, and Abdulnasser Hatemi-J. 2006. "Tests for Causality between Integrated Variables Using Asymptotic and Bootstrap Distributions: Theory and Application." *Applied Economics* 38 (13): 1489–1500.

Hatemi-J, Abdulnasser, and Mrittika Shamsuddin. 2016. "The Causal Interaction between Financial Development and Human Development in Bangladesh." *Applied Economics Letters* 23 (14): 995–98.

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.

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.

Liao, Shu-Hsien, Hsu-Hui Ho, and Hui-Wen Lin. 2008. "Mining Stock Category Association and Cluster on Taiwan Stock Market." *Expert Systems with Applications* 35 (1): 19–29.

Malhotra, Bhavya. 2014. "Foreign Direct Investment: Impact on Indian Economy." *Global Journal of Business Management and Information Technology* 4 (1): 17–23.

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.

Nanda, S. R., B. Mahanty, and M. K. Tiwari. 2010. "Clustering Indian Stock Market Data for Portfolio Management." *Expert Systems with Applications* 37 (12): 8793–98.

Popescu, Madalina Ecaterina, Amalia Cristescu, Larisa Stanila, and Maria Denisa Vasilescu. 2016. "Determinants of Undeclared Work in the EU Member States." *Procedia Economics and Finance* 39 (January): 520–25.

Prasanna, N. 2010. "Direct and Indirect Impact of Foreign Direct Investment (FDI) on Domestic Investment (DI) in India." *Journal of Economics* 1 (2): 77–83.

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.

Qodmanan, Hamid Reza, Mahdi Nasiri, and Behrouz Minaei-Bidgoli. 2011. "Multi Objective Association Rule Mining with Genetic Algorithm without Specifying Minimum Support and Minimum Confidence." *Expert Systems with Applications* 38 (1): 288–98.

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.

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.

Singh, Shikha. 2019. "Foreign Direct Investment (FDI) Inflows in India." *Journal of General Management Research* 6 (1): 41–53.

Tuomi, Krista. 2011. "The Role of the Investment Climate and Tax Incentives in the Foreign Direct Investment Decision: Evidence from South Africa." *Journal of African Business* 12 (1): 133–47.

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, X., V. Kumar, J. Ross Quinlan, and J. Ghosh. 2008. "Top 10 Algorithms in Data Mining." *And Information Systems*. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/s10115-007-0114-2&casa_token=c5Hw2pbQCdcAAAAA:xv07RfJagCnxitQrduYRxS56uBApywcVApvXMJljGB-j1rna3H1srstBrEcDAiwspRciwbdUZzdT_lc3Gg.

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 K-Means

SI.No.	**Test Size**	**ACCURACY RATE**	
		Apriori algorithm	K-means clustering

1	Test1	93.23	90.10
2	Test2	93.54	90.23
3	Test3	92.36	91.19
4	Test4	94.34	90.92
5	Test5	93.12	89.92
6	Test6	94.56	91.01
7	Test7	95.35	90.85
8	Test8	92.36	91.28
9	Test9	93.35	90.58
10	Test10	94.54	91.34
Average Test Results		95.68	91.346

Table 2. The statistical calculation such as Mean, standard deviation and standard error Mean for Apriori algorithm and K-means clustering. The accuracy rate parameter used in the t-test. The mean accuracy rate of Apriori algorithm is 95.68 and K-means clustering is 91.346. The Standard Deviation of Apriori algorithm is 0.21029 and K-means clustering is 1.81738. The Standard Error Mean of Apriori algorithm is 0.10291 and K-means clustering is 0.71893.

Group		**N**	**Mean**	**Standard Deviation**	**Standard Error Mean**
Accuracy Rate	K-MEANS CLUSTERING	10	91.346	1.81738	0.71893
	APRIORI ALGORITHM	10	95.68	0.21029	0.10291

Table 3. The statistical calculations for independent samples test between Apriori algorithm and K-means clustering. The significance value p=0.029 ($p<0.05$) shows that two groups are statistically insignificant. Independent samples T-test is applied for comparison of Apriori algorithm and K-means clustering with the confidence interval as 95% and level of significance as 0.32132.

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	5.443	0.029	5.070	18	.000	5.4300	1.0709	3.18005	7.67995
	Equal variances not assumed			5.070	15.754	.000	5.4300	1.0709	3.15683	7.70317

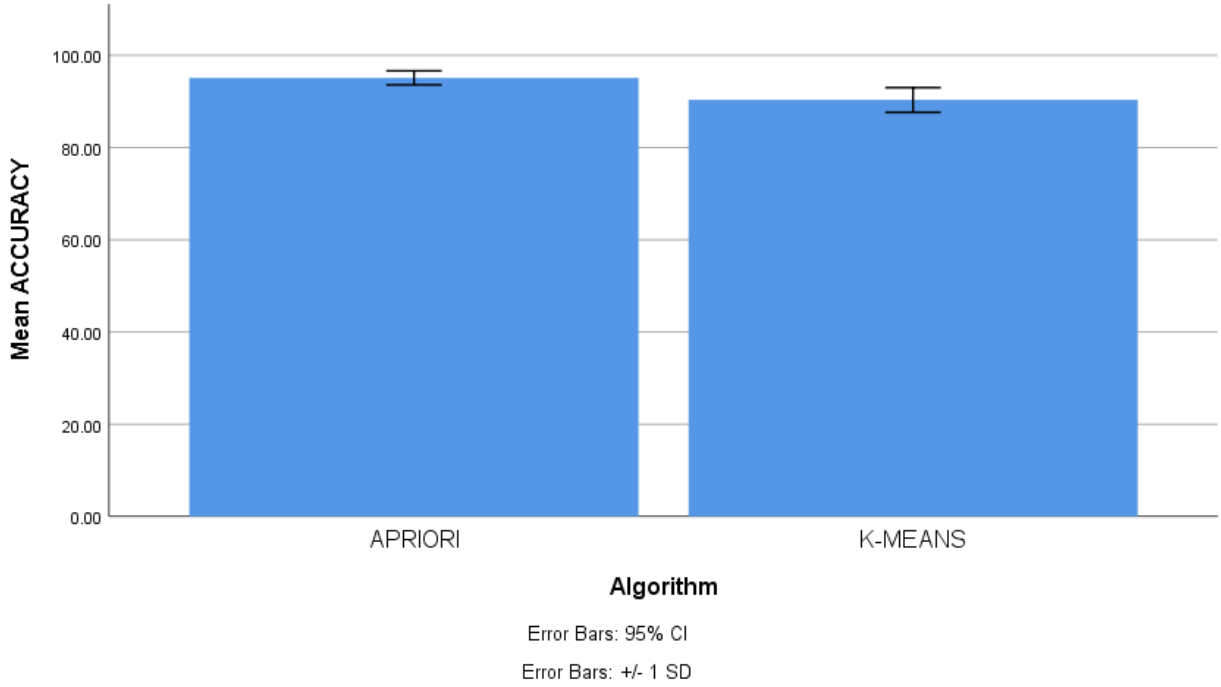

Fig. 1. Simple Bar graph for Apriori algorithm accuracy rate is compared with K-means clustering. The Apriori algorithm is higher in terms of accuracy rate 95.68 when compared with K-means clustering 91.346. Variable results with its standard deviation ranging from 100 lower to 150 higher Apriori algorithm where K-means clustering standard deviation ranging from 200 lower to 300 higher. There is a significant difference between Apriori algorithm and K-means clustering ($p<0.05$ Independent sample test). X-axis: K-means clustering accuracy rate vs Apriori algorithm Y-axis: Mean of accuracy rate, for identification of keywords ± 1 SD with 95 % CI.