

Market Basket Analysis with Enhanced Support Vector Machine (ESVM) Classifier for Key Security in Organization



P.Yoganandhini, G.Prabakaran

ABSTRACT: Market Basket Analysis is considered to be one among the highly popular and efficient sort of data analysis exploited in the marketing and retailing field. The objective of market basket analysis lies in deciding the products purchased together by the customers. Its name has originated from the concept of customers filling into a shopping cart everything of all they had purchased (a "market basket") while doing shopping in the grocery. Having a knowledge of the products that customers buy in group can be quite useful for a retailer or to any other organization. A store could make the best use of this information to keep the products that are often sold together in the same place, whereas a catalog or World Wide Web (WWW) merchant could utilize it for deciding the structure of their catalog and order form. Since several applications such as market basket analysis, fraud detection in web, medical diagnosis, census data, Customer Relationship Management of business that makes use of association rules exists, the process involving Decision making can be improved. Security is also regarded to be an important facet for transactions done individually and frequent itemsets for database that are horizontally partitioned. In order to render security for lastly bought often used itemsets for transaction purposes, this research work introduces a novel key security algorithm that uses RSA cryptographic technique which is classifier based. The classifier makes use of information about several often utilized itemsets and it provides a key value to the actual company. For instance, in case if there are any reliance users, only the valid users can obtain that market info. The rest of the users belonging to the reliance organization are not allowed to select the data's key value. First, the frequent itemsets are mined with the help of association rule mining employing Probabilistic Graphical Model techniques. Then the Enhanced Support Vector Machine (ESVM) classifier checks the key values of the mined frequent itemsets.

Keywords: Association rules, Customer relationship management, ESVM, Frequent item set mining, key values, Market basket analysis.

I. INTRODUCTION

Due to the progress made in the technological domain of information and communication, retail companies and enterprises can rightly get and store the transactional and geographic information on different customers with reasonable expenses involved [1,2]. Therefore, one among the problems faced by corporations and the retailers is that there has been heavy investment in data collection about customer and there is

no clue on the ways of extracting the significant information from their massive customer repositories for achieving a contentious edge. Market basket analysis (also called as association rule mining) is an approach employed for the discovery of end user buying patterns. Also, one among the important tasks of retail dealers is the management of their product portfolios to increase the profit obtained of the store or retail chain on the whole. Making use of the marketing mix mechanisms for influencing the purchases for a particular product generally affects both the recommended category and associated groups. In addition, a retailer prefers not just advertising for one group, but in several at the same time [3,4,5].

By extracting associations co-occurrences from stores' transactional databases. The discovery of an aspect, for instance, that the customers of supermarket possibly will tend to buy products such as milk, bread, and cheese jointly, or the bank's clients can possibly make use of a category of services together, which can aid the concerned managers in the design of the store setup, web sites, product blend and grouping, and other tactics of marketing. Due to the association rules being helpful and convenient to interpret, several useful business applications, such as, finance, telecommunication, marketing, retail, and web analysis have been developed [6,7].

The technique has also gained considerable research focus, and multiple extensions have been introduced recently, inclusive of (1) improvements in the algorithm; (2) fuzzy rules; (3) multi-level and generalized rules; (4) quantitative rules; (5) spatial rules; (6) inter-transaction rules; (7) intriguing rules; and (8) temporal association rules. Security is also considered to be an important aspect in transactions done individually and frequent itemsets for databases that are horizontally partitioned. With the aim of providing security for the frequently utilized itemsets that is purchased finally for the transaction, this research work introduces a novel key security algorithm that employs RSA cryptographic technique in which the classifier optimizes the key values [8,9].

The classifier makes use of the information about several often utilized itemsets and thereafter provides a key value to the actual company. For instance, in case if there are any reliance users, only the valid users can obtain that market info. The rest of the users belonging to the reliance organization are not allowed to select the data's key value. First, the frequent itemsets are mined with the help of association rule mining employing Probabilistic Graphical Model techniques. Then the Enhanced Support Vector Machine (ESVM) classifier checks the key values of the mined frequent itemsets [10].

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

P.Yoganandhini*, Research Scholar, Department of Computer and Information Science, Annamalai University, Chidambaram (Tamil Nadu), India

Dr. G.Prabakaran, Associate Professor, Department of Computer Science and Engineering, Annamalai University, Chidambaram (Tamil Nadu), India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

II. LITERATURE REVIEW

Waghmare and Mukhopadhyay [11] designed a location-oriented, mobile-based shopping application useful for shops dealing with bake house products. The deployment of the entire application is done over the cloud foundation. The three-tiered architecture comprises of a front-end, middle-ware and a back-end. The front-end stage involves a location-oriented, mobile-based shopping application developed for mobile devices running on android, for buying the bake house products from areas in the neighbourhood. The front-end stage also shows the relationship between the products bought. The middle-ware stage renders a web service for generating the JSON (JavaScript Object Notation) as the result from the relational database. It helps in exchanging the information between the application on the mobile and servers present on the cloud infrastructure. The back-end stage includes the Apache Tomcat Web server and My SQL database. Google Cloud Messaging (GCM) is used by the application for the generation and transmission of notification of the orders to the shop owner.

Hussain and Hussein[12]used a data mining approach, involving the market basket analysis, using the data on student attendance. This analysis contributes towards identifying the student groups who have nearly matching absence records. This kind of similarity may highlight that the students miss classes owing to peer pressure, instead of acceptable reasons. This technique has been tested through the analysis of the student attendance data for more than two thousand students attending a public senior educational institution for a period of one semester. The obtained results were useful in finding the students who are missing classes just because their friends are missing the classes.

Sherly and Nedunchezhian[13] introduced distributed FPM approaches including two kinds of count distributed and compressed data distributed concurrent techniques. Both the algorithms are experimentally analysed and their performances are compared with commonly used distributed FPM algorithms employing benchmarked datasets. Then their performances are compared and assessed for providing a market analysis and applications involving online credit card fraud activity detection.

Gatzioura and Sánchez-Marrè[14]suggested a new scheme, which deals with the above mentioned problems by means of a case-oriented recommendation strategy. The technique’s scope extends to the generation of useful recommendations that depend on co-occurrence patterns of products and to offer more perspective into the purchasing behaviour of customers, and the recommender deployed makes use of a hierarchical model for the products and searches for matching sets of products, with the aim of recommending those, which possibly can fulfil users’ desires.

Varma and LijiP[15] suggested the usage of Advanced Encryption Standard (AES) for encrypting the data items prior to their outsourcing so as to avoid them being susceptible to ‘Known Plaintext’ attack invokedon the system available already. Imaginary transactions are added to the databases employing k-anonymity technique for counteracting against the frequency analysis attack. In accordance with the experimental tests and analysis, even though the execution time of the newly introduced approach is somewhat more than the available systems, it yields better

protection to the data products. As third party server carry out the computational tasks, there is very less resource consumed at the data proprietors’ end.

Sweatband Ramachandran [16] highlighted on investigating an effective cryptographic algorithm used for the protection of cloud information over homomorphic encryption strategy and also the Key management and system performance for market basket analysis on customer bought product information are explored.

Trnka, [17] employed GRI (General Rule Induction) algorithm for the generation of association rules between the items present in the market basket. These relationships exhibit the diversity between the items. In order to depict the dependency between the items, a Web plot is used. C5.0 was the final algorithm used in analysis. This algorithm was applied for the development of rule-dependent profiles.

| s. no | Name of the Author | Technique | Advantages | Drawbacks |
|-------|-------------------------------------|---|--|---|
| 1. | Waghmare and Mukhopadhyay [2014] | Three-tier architecture. | Limits the depletion of battery. | Consumes lot of time. |
| 2. | Hussain and Hussein[2019] | Data mining approach. | Students who are absent for classes | It does not contain adequate number of records. |
| 3. | Sherly and Nedunchezhian [2015] | Prefix based equivalence class partitioning approach. | Limits the usage of memory. | Extremely costly. |
| 4. | Gatzioura and Sánchez-Marrè [2014] | New scheme | Superior Performance | Second stage of processing is required. |
| 5. | Varma and LijiP [2017] | AES | Maximum security. | Consumes lot of time. |
| 6. | Swetha and Ramachandram [2016] | Key management | Better performance | Security needs improvement |
| 7. | Trnka [2010] | GRI | Increases the degree of Six Sigma performance. | Extremely costly. |

III. PROPOSED METHODOLOGY

This section explains about the newly introduced secured transactional model based market basket analysis for improving the profit and sales. It comprises of two stages. The primary one includes frequent itemset mining employing Association Rule Mining Algorithm depending on Probabilistic Graphical Model and the next one includes the key generation employing RSA cryptographic technique in which Enhanced support vector machine classifier optimizes the key values.



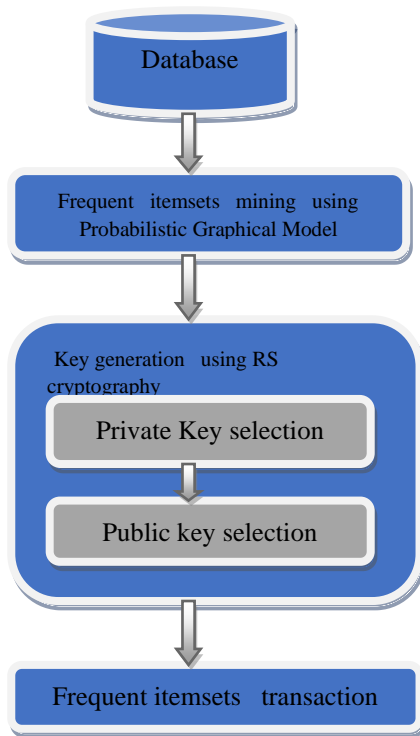


Figure 1. Proposed system

It starts with the mining of the frequent item set out of the database employing association rule mining for providing this information to the company.

A. Association Rule Mining:

The objective of association rule mining is to discover the correlation between various objects present in a database. Several essential concepts exist in this processing process, listed as below:

1. T represents the dataset used in the mining of the association rule, i.e. $T = \{t_1, t_2, \dots, t_n\}$, where $t_k (k = 1, 2, \dots, n)$ is called as the 'transaction' $t_k = \{i_1, i_2, \dots, i_p\}$, and 'item' represents $i_m (m = 1, 2, \dots, p)$. Every transaction is represented by a distinct identifier, known as TID.

2. Supposing that $I = \{i_1, i_2, \dots, i_p\}$ refers to a set composed of all the data objects present in T, a random subset X of I is called as the item set of T. If $|X| = k$, X indicates a k-item set. The association $X \in t_k$ specifies that the transaction t_k has the item set X.

3. σ_X represents the number of transactions, which consists of the item set X present in the dataset (T) is called as the support number of the item set X. The support coefficient of the item set X is given by the ratio of σ_X and |T| and expressed as Support(X)

$$\text{Support}(X) = \frac{\sigma_X}{|T|} \times 100\% \quad (1)$$

where |T| indicates the number of items present in T. min Sup represents a threshold limit for the support coefficient that could be pre-assumed. In case, support(X) is not less compared to min Sup, X is considered to be a frequent item set. Else, X is called as an infrequent item set.

4. Supposing that X and Y refer to the two object sets present in T, If $X \cap Y \neq \emptyset$, the relation $X \Rightarrow Y$ that holds shall be known as an association rule, and X and Y indicate the premise and conclusion related to the association rule $X \Rightarrow Y$, correspondingly. The support coefficient of the item set $X \cup Y$ is called as the support coefficient of the association rule $X \Rightarrow Y$ and expressed as support $(X \Rightarrow Y)$:

$$\text{Support}(Y \Rightarrow X) = \text{support}(X \cup Y) \quad (2)$$

The confidence coefficient of the association rule given by $X \Rightarrow Y$ is represented as confidence $(X \Rightarrow Y)$, and expressed as in (9).

$$\text{Confidence}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} \times 100\% \quad (3)$$

Min Conf refers to the assumed threshold level for confidence coefficient. In the case of association rule mining, the statistical importance of an association rule is weighed with the help of its support coefficient, and the confidence coefficient provides the measure of its confidence degree.

5. In case an association rule $X \Rightarrow Y$ meets the condition of both support $(X \Rightarrow Y) \geq \text{minConf}$, the association rule $X \Rightarrow Y$ will be considered a stronger one. Else, it is regarded to be a weaker one.

The process of association rule mining involves finding each one of the better association rules present in T. This process of mining involves two stages as follows.

Step :1 Get each one of the frequent item sets present in the data set in accordance with the minimum support coefficient, given by minsup.

Step :2 Create the association rules depending on the frequent item sets and the minimum confidence coefficient, given by minConf.

B. Association Rule Mining Algorithm Based on Probabilistic Graphical Model

The association rule mining algorithm that depends on the probabilistic graph integrates both the Apriori and probabilistic graph model. The initial stage of the technique finds each one of the frequent sets of two-item employing the Apriori algorithm, and thereafter the frequent sets shall be represented with the help of a probabilistic graph. Also the association rules can be extracted in accordance with the probabilistic graph. In addition, the support coefficient and confidence coefficient pertaining to every association rules can be computed together. By using the probabilistic graph, one drawback encountered is that during the search for the frequent items, all the data items need to be scanned cyclically is avoided.

The steps followed in the association rules mining for power transformers state parameters depending on the probabilistic graphical model are described as below.

Step :1. With the help of Apriori, get all the frequent two-item sets. Apriori makes use of the iterative technique defined by layer after layer search, creates the frequent item sets out of the candidate item sets and generates the association rules using the frequent item sets [13]. But, just one and two-item sets get searched in this process. Hence, the complexity involved in the retrieval can be significantly reduced.

The process of frequent two-item sets' generation is conducted with the aid of the Apriori algorithm and the steps given below are followed.

1. The frequent one-item sets L_1 is generated. In this step, all of the transactions are read through and the occurrence frequency of every transition is computed.

2. The frequent two-item sets L_2 is generated depending on the one-item sets L_1 found.

1) The candidate 2-item sets C_2 is generated. In this stage, each of the two modes that share one object are linked to generate C_2 . Then depending on the anti-monotonicity property of frequent item sets, pruning of C_2 is carried out.

2) The support coefficient of the candidate two-item sets C_2 is decided so it can be added to the transaction t subset (C_2, t) , and then it is saved to a hash table. During this step, the database has to be scanned.

3) The generation of the frequent two-item sets L_2 is done by eliminating the item sets whose support coefficients are lesser compared to minsup .

Step 2. Choose the frequent two-item sets L_2 depending on the interestingness factor.

When $(X \cup Y) \approx \text{support}(X) \text{ support}(Y)$, the two item sets (X, Y) are considered to have mutual independence, which implies that the association rule $X \Rightarrow Y$ is not useful.

The P-S interestingness of two-item sets X and Y can be expressed as (10)

$$\text{Interest}(X, Y) = \frac{\text{support}(XUY)}{\text{support}(X)\text{support}(Y)} - 1 \quad (4)$$

$$= \frac{P(X/Y) - P(X)}{P(X)} \quad (5)$$

(1) If $\text{interest}(X, Y) > 0$, X and Y are positively associated.

(2) If $\text{interest}(X, Y) \approx 0$, X and Y are mutually independent and it needs to be removed out of the frequent two-item sets.

(3) If $\text{interest}(X, Y) < 0$ X and Y are considered to be negatively associated.

Step :3. Form the probabilistic graph. The edges provide conditional probability distributions in L_2 .

Step :4. Create the association rule set R on the basis of the generated probabilistic graph.

Step :5. Compute the probability table of the association rules.

Here, the frequently bought itemsets are extracted from the store database and this information has to be securely sent to the organization and for this, in the next subsequent step, the encryption of the frequent itemsets is done with the help of RSA cryptography key.

C. Key Generation Using RSA Cryptography

PKC mechanism makes use of a single individual key for the encryption process and another key for decryption process. The advanced PKC was first defined with the help of a two-key crypto system in which the users and companies could securely communicate over an insecure communications channel with no necessity of sharing a secret key. RSA is one among the first and yet the most popular implementation of PKC that it has been remained in use till recent times for exchanging keys. In PKC, one among the keys is designated to be the public key (p) and may be publicised as extensively as the owner desires. The

other key is designated to be the private key (d) and it is never disclosed to a third party.

Generally, arbitrary values are chosen for the private keys. But, these may lead to the generation of infinite number of values in the computations involving public key when their selection is done correctly. Moreover, it is necessary to optimize the selection of the private keys for generating useful public keys and also in accordance with other limitations like size. ESVM is used separated as two cases keeping these objectives into consideration.

D. Private Key Selection using ESVM

The standard SVM approach was presented in the form of a tool for binary classification, however it can be updated to be an n -class classification technique by considering a series of n or $n(n-1)/2$ binary classification operations. SVM extends the input data onto the Hilbert space where the classification uses an optimal isolating hyperplane. Through the maximization of the hyperplane isolating the two classes, binary SVM reduces the higher limit of generalization error.

This capacity can be considered to be the rough realization of Structural Risk Minimization (SRM), which boost the generalization performance of SVM, with no regard to the data distribution. A training dataset composed of n data points, which are differentiable into two classes can be denoted by

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \in X \times Y, X \in R^n, Y \in \{1, -1\}.$$

The objective of the classification technique is to get a classifier $y = f(x)$, which in turn is a projection of X onto Y depending on data present in T . Any of the points external to the training set T but within $X \times Y$ will be rightly classified by the projection specified.

Consider the problem involving isolating the training set into two classes of $+1$ and -1 , then $H_1 : W \cdot x + b = +1$ and

$$H_2 : W \cdot x + b = -1 \quad (X \in R^n, b \in R)$$

would become the probable hyperplanes so that a major section of class 1 instances fall up above $H_1 (w \cdot x + b > 1)$ and a major section of class -1 are below $H_2 (w \cdot x + b < -1)$, where the points positioned on H_1 and H_2 are indicated as support vectors and are accountable for deciding the optimal isolating hyperplane $H : w \cdot x + b = 0$. The distance between H_1 and H_2 can be denoted by therefore the maximization of distance between H_1 and H_2 can be obtained by minimizing the norm of w , resulting in a restrained optimization problem.

Figure 2 points out that, not all the training sample data may have a linear hyperplane separation exactly. In order to take the misclassification errors into consideration, a penalty parameter c for the instances slipping out off the margin, along with slack variables ξ_i (nonnegative) are included to the problem. Slack variables indicate the distances computed between the miscategorised points and the initial hyperplane. The penalty parameter c achieves a balance between the margin size and the number of training points that are misclassified; while a bigger c yields lesser number of misclassifications, and also leads to a margin size that is small.



Consequently, it becomes a restrained optimization problem, and also a quadratic programming problem having inequality conditions, as expressed by Eq. 6,7

$$\text{Minimize } f(w, \xi) = \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n \xi_i \quad (6)$$

$$\text{Subject to } y_i [w \cdot x_i + b] + \xi_i - 1 \geq 0, \xi_i \geq 1, i = 1, \dots, n \quad (7)$$

Where ξ_i refers to the positive slack variables and c stands for the penalty parameter. But, the objective is to get an optimal hyperplane to reduce the number of errors occurring due to misclassification and increase the margin size at the same time. The popular generic means of tackling with such issues that are difficult to be solved in a straightforward manner, is by using Lagrange multipliers for projecting the problem from the primal space onto a dual space. Inclusion of n nonnegative Lagrange multipliers $\alpha_1, \alpha_2, \dots, \alpha_n \geq 0$ related to the inequality conditions specified in Eq. 1 generates Eq. 8,9

$$\text{Maximize } L(a) = \sum_{i=1}^n \alpha_i / \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (8)$$

$$\text{Subject to } : \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq c, i = 1, \dots, n \quad (9)$$

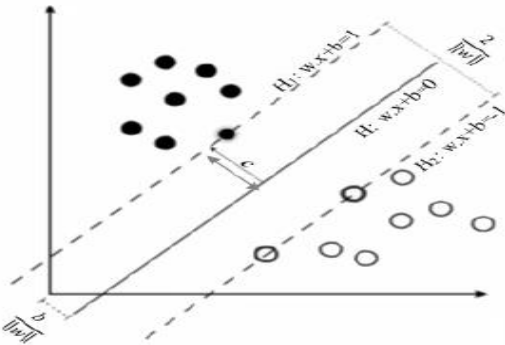


Figure .2 : A linear, binary SVM classifier and the optimal separating hyperplane H , present in between and lying parallel to H_1 and H_2

In order to tackle non-linearity, the mapping of the data can be done onto a higher dimensional space generated with the help of an analytical projection and called as the kernel trick. This is done as in this problem of optimization, just the dot product of two vectors shows up in the feature space, and by substituting x with its mapping done in the feature space, the kernel function k shall be expressed as . Making use of a kernel function, the optimization function helps in the maximization of Eq. 10

$$L(a) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (10)$$

Where the general kernel functions include the linear function $k(x, y) = x_i \cdot x_j$ radial basis function (RBF) $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ and polynomial function $k(x, y) = (1 + x_i \cdot x_j)^q$ where γ and q refer to the kernel parameters.

Therefore ESVM chooses a private key for improving the security. In a similar manner, ESVM can also choose an optimal random number, R .

Just before the encryption, both the user A and organization B must have the knowledge on determining the RSA parameters and also translate the text into numerical

data. In this, the ASCII values of numbers are used for converting the text into numbers. Then, A and B can send the transaction data securely by adopting steps given below.

1. Two users A and B choose the RSA parameters.
2. User A chooses the Key Generation Point KGP.
3. Private key number P_{TA} is chosen by User A with the help of ESVM algorithm
4. Key generation point and public key calculated from the private key are transmitted to B organization .
5. Organization B **ciphers the text** utilizing the public key P_{UA} of User A.
 - a. Get the text message that has to be transmitted
 - b. Compute the respective ASCII value
 - c. Obtain value R (any random number) employing ESVM.
 - d. Calculate the cipher parts using point multiplication
 - e. Compute the next cipher part utilizing point addition or point doubling
 - f. Compute cipher text as given in equation to give User A
6. The **decryption procedure carried out** by A is given as.
 - a. Obtain the cipher text CP.
 - b. Obtain the left segment C1 and right segment C3 of the CP individually.
 - c. Multiply with PrA to the left segment and then subtract it from the right segment to obtain MP
 - d. Translate it back to the list of ASCII values.
 - e. ASCII values are then mapped onto respective characters

IV. RESULT AND DISCUSSION

This section explains the results achieved of the tests carried out on the newly introduced model. The implementation of the model is done with the help of NS2. The comparison of existing GRI, AES algorithm and the newly introduced RSA cryptographic with Enhanced support vector machine (RSESVM) are performed in terms of the metrics of Precision, Execution time, Accuracy and Error rate for multi store database.

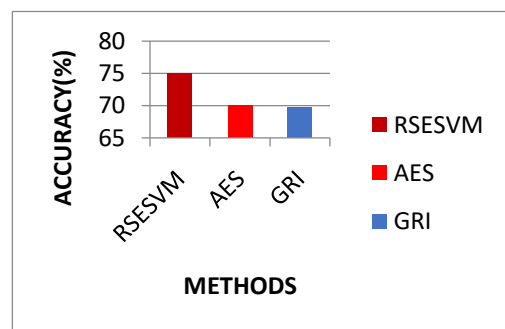


Figure:3. Accuracy results comparison vs. Classifiers

Figure:3 demonstrates the results obtained of the performance comparison between three different techniques inclusive of GRI, AES and RSESVM in terms of the Accuracy metrics.

It can be concluded from the results that the novel RSESVM technique yields much better accuracy of 75%, while other techniques including AES and GRI techniques yields just 70.00% and 69.76% values correspondingly.

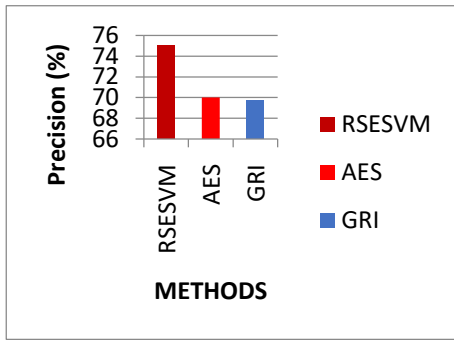


Figure :4. Precision results comparison vs. Classifiers

Figure 4 demonstrates the results of the performance comparison between three different techniques including GRI, AES and RSESVM in terms of the precision metrics. It can be concluded from the results that the novel RSESVM technique yields much better precision results of 85%, while other techniques including AES and GRI techniques yields just 81.5% and 76% values correspondingly.

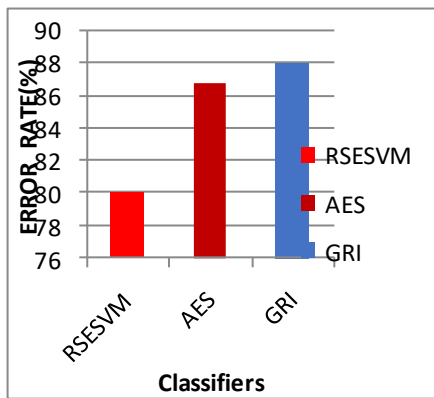


Figure:5. Error rate results comparison vs. Classifiers

Figure 5. depicts the results obtained of the performance comparison analysis between three different classifiers inclusive of GRI, AES and RSESVM in terms of the Error rate metric. It can be concluded from the results that the novel RSESVM technique yields much better results of error rate of 80%, while other techniques including AES and GRI techniques yields just 86.75% and 88% values correspondingly.

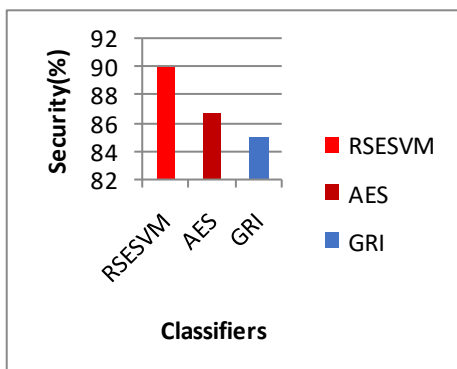


Figure: 6. Security results comparison vs. classifiers

Figure: 6 illustrate the results achieved of the performance comparison analysis between three different classifiers inclusive of GRI, AES and RSESVM in terms of the security metric. It can be concluded from the results that the novel RSESVM technique yields much better results of security value of 90.91%, while other techniques including AES and GRI techniques yields just 86.75% and 85% values correspondingly.

V. CONCLUSION AND FUTURE WORK

An idealistic automatic market basket analysis system stands for one which includes a right kind of analysis system operating behind the scene to help the organization in improving their market. In this research work, the demonstration of different internet technologies and novel algorithms is capable of rendering solutions to the problem involving the analysis of customer purchasing habits. At first, the mining of frequent itemsets are done with the help of association rule mining employing Probabilistic Graphical Model techniques. After this, the generation of the key will be done with the help of RSA algorithm and Enhanced Support Vector Machine (ESVM) classifier is employed for the private key selection. The works intended for the future will provide the implementation of specific features that the work was not capable of exploring by using a more reliable algorithm in the system, which in turn would facilitate the system in operating rapidly and with more efficiency. Efforts on improving the search techniques can also be useful in boosting the market and profitability.

REFERENCES:

- Santarcangelo, V., Farinella, G.M., Furnari, A. and Battiato, S., 2018. Market basket analysis from egocentric videos. Pattern Recognition Letters, 112, pp.83-90.
- Santarcangelo, V., Farinella, G.M., Furnari, A. and Battiato, S., 2018. Market basket analysis from egocentric videos. Pattern Recognition Letters, 112, pp.83-90.
- Dhawan, G., 2018. Data Mining Techniques and its Uses in different fields: A Review Paper. Journal of Electronic Research and Application, 2(4).
- Valle, M.A., Ruz, G.A. and Morrás, R., 2018. Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97, pp.146-162.
- Raja, B., Pamina, J., Madhavan, P. and Kumar, A.S., 2019. Market Behavior Analysis Using Descriptive Approach. Available at SSRN 3330017.
- Sharma, N., 2019. A Review on Yield Prediction of Various Techniques and Features.
- Maske, A. and Joglekar, B., 2018, August. Survey on Frequent Item-Set Mining Approaches in Market Basket Analysis. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.
- Tatiana, K. and Mikhail, M., 2018. Market basket analysis of heterogeneous data sources for recommendation system improvement. Procedia Computer Science, 136, pp.246-254.
- Srivastava, N., Gupta, K. and Baliyan, N., 2018, April. Improved Market Basket Analysis with Utility Mining. In Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT) (pp. 26-27).
- Reyes, R., Valenzuela, S. and de Manila, A., 2018. Shopping for Politicians: Insights from Market Basket Analysis of Senatoriables. Building Inclusive Democracies InAsean, p.333.
- Waghmare, V. and Mukhopadhyay, D., 2014, December. Mobile Agent based market basket analysis on cloud. In 2014 International Conference on Information Technology (pp. 305-310). IEEE.



12. Hussain, M. and Hussein, A., 2019, April. Market Basket Analysis of Student Attendance Records. In 2019 IEEE Global Engineering Education Conference (EDUCON) (pp. 1198-1203). IEEE.
13. K. K. Sherly and R. Nedunchezian, 2015. A Improved Incremental and Interactive Frequent Pattern Mining Techniques for Market Basket Analysis and Fraud Detection in Distributed and Parallel Systems. Indian Journal of Science and Technology, Vol 8(18), pp(1-12).
14. Gatzoura, A. and Sánchez-Marrè, M., 2014. A case-based recommendation approach for market basket data. IEEE Intelligent Systems, 30(1), pp.20-27.
15. Varma, S. and LijiP, I., Secure Outsourced Association Rule Mining using Homomorphic Encryption, 207. International Journal of Engineering Research and Science, 3(9).
16. N.Swetha, Prof. S Ramachandram, Effective Crypto System for Achieving Security and Performance over Market Basket Data Analysis, 2016. IOSR Journal of Computer Engineering (IOSR-JCE) Volume 18, Issue 6, PP 13-18.
17. Trnka, A., 2010, June. Market basket analysis with data mining methods. In 2010 International Conference on Networking and Information Technology (pp. 446-450). IEEE.
18. Yuan, X., 2017, March. An improved Apriori algorithm for mining association rules. In AIP conference proceedings (Vol. 1820, No. 1, p. 080005). AIP Publishing.
19. Li, L., Lu, R., Choo, K.K.R., Datta, A. and Shao, J., 2016. Privacy-preserving-outsourced association rule mining on vertically partitioned databases. IEEE Transactions on Information Forensics and Security, 11(8), pp.1847-1861.
20. Kaur, M. and Kang, S., 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. Procedia computer science, 85, pp.78-85.
21. Yi, X., Rao, F.Y., Bertino, E. and Bouguettaya, A., 2015, April. Privacy-preserving association rule mining in cloud computing. In Proceedings of the 10th ACM symposium on information, computer and communications security (pp. 439-450). ACM.
22. Hazay, C., Mikkelsen, G.L., Rabin, T., Toft, T. and Nicolosi, A.A., 2019. Efficient RSA Key Generation and Threshold Paillier in the Two-Party Setting. Journal of Cryptology, 32(2), pp.265-323.
23. Thangavel, M., Varalakshmi, P., Murali, M. and Nithya, K., 2015. An enhanced and secured RSA key generation scheme (ESRKGS). Journal of information security and applications, 20, pp.3-10.
24. Guenther, N. and Schonlau, M., 2016. Support vector machines. The Stata Journal, 16(4), pp.917-937.
25. Suthaharan, S., 2016. Support vector machine. In Machine learning models and algorithms for big data classification (pp. 207-235). Springer, Boston, MA

AUTHORS PROFILE



P.YOGANANDHINI, M.C.A., M.Phil., working as Assistant Professor in Dept. of Computer Technology, Vellalar College for Women, Erode and currently pursuing her Ph.D in the Dept. of Computer and Information Sciences, Annamalai University, Annamalai Nagar. She has published one paper in an UGC refereed International Journal. she has five years

of teaching experience and two and a half years of Industrial experience.



Dr. G.PRABAKARAN working as Associate Professor in Dept. of Computer Science and Engineering, Annamalai University, Annamalai Nagar. He has published two papers in National Journals and four papers in International refereed Journals. He has published more than 8 papers in National Conference and 13 papers in International Conferences. He has 19

years of teaching experience and 4 years of Research experience. He is a member of CSI and ISTE. His main research work focuses on Image processing, computer graphics.