Contribution Principal Component Analysis to Optimizing Data by Reducing Product Data on Transaction

Dewi Sartika Br Ginting¹, Fuzy Yustika Manik¹ and Fanindia Purnamasari²

¹Department of Computer Science, Universitas Sumatera Utara, Jl. Alumni No. 9-A, Medan 20155- Indonesia
²Department of Information Technology, Universitas Sumatera Utara, Jl. Alumni No. 9-A, Medan 20155- Indonesia

dewiginting@usu.ac.id, fuzy.yustika@usu.ac.id, fanindia@usu.ac.id

Abstract. Principal component analysis is to analyze the observation data table into a new data table that has the same correlation. And the aim is to simplify the previously complex observation data so that it is easier to process or analyze. The dataset used is transaction data which is often used by the association method in sales analysis, where the data taken consists of 1397 types of products sold in 1200 transactions. In this data, there are products that have very small sales, which means that the percentage of these products has very little effect on the future process, namely sales analysis with the association method. Therefore the authors try to optimize the data to become ready to use data by reducing products that have a small percentage value that affects research for the dataset. And on this occasion the author uses the main component analysis method to reduce products or form products that can represent the entire dataset without reducing the quality of the data for analysis. From the results of research conducted on transaction data, there was a product decline of 65.21%, where the products totaling 1397 were reduced to 486 products that could represent without reducing their value.

1. Introduction
Principal Component Analysis (PCA) is a technique that analyzes an observation data table into a new data table that has similar correlation. It is used to compress maximum information into first two columns of the transformed matrix known as the principal components by neglecting the other vectors that carries the negligible information or redundant data [1]. The objective of principal component analysis is to simplify the previously complex observation data so that it is easier to process or analyze. According to researchers, principal component analysis is a statistical technique that is processed linearly by replacing the pattern of a group of original variables into simpler variables that are not correlated but can represent information from the original set of variables [2]. In another case PCA can also be combined with least squares support vector machine, where PCA will reduce the dimension of the input variable X, which can reduce the number of inputs for LSSVM [2] and another thing is also PCA which is used to represent the image by developing into two dimensions of PCA [3]. There is a need for data reduction to optimize the data processing process, in this case principal component analysis is able to reduce data from such a large scale to data that can represent all data for processing by forming a matrix, calculating the variance covariance value, looking at the resulting eigen values and viewing the scree plot. The function of covariance matrix as weighting in model estimation to avoid correlation
between errors [4]. If the data contains more than one correlated response, it is possible that error correlation will occur.

2. Research Method
Data reduction is the reduction of product data from transaction data in supermarkets for 3 months with the number of transactions of 1200 and the number of products 1397. Reduction of data in this case will reduce data or very small products, the percentage of sales from the 1200 transactions that have occurred. Where the data resulting from reduction is data that can represent original data to be implemented or data that is ready to use for subsequent processing. In other words, the selected products can represent the value of the entire product in the data. It can be explained that the number of attributes or products in this case will decrease but the quality of the data processing results does not decrease if the original data is used. In the raw data, there are many blank product sales data for each transaction. Therefore it takes steps such as removing some products that are not determinants in this research process. The researcher used the principal component to reduce the product and in the PCA process the SVD technique was implemented using positive semi-definite eigenic decomposition matrix to get a similar decomposition that applies to all a rectangular matrix consisting of real numbers [5], as in [6] SVD is a general method for a change of basis and is used in the principal component analysis and according to SVD, we can factorize any matrix $A \in \mathbb{R}^{m \times n}$. Process stages for the completion of the principal component analysis with singular value decomposition method can be seen in Figure 1:

![Diagram of Principal Component Analysis Method]

**Figure 1. The Stages of Principal Component Analysis Method**

Based on figure 1, the calculation phase using the principal component analysis method can be explained as follows:

1. Initial data are prepared in an axb size matrix. Later the number of variable b will be reduced to c the number of principal component being maintained.
2. In the pre-principal component analysis process, the variance covariance matrix value is searched, and to get the variance covariance value, it is necessary to first obtain the deviation matrix value.
3. The Singular Value Decomposition stage is the process stage of the principal component analysis which processes the matrix value from the variance covariance results.

Covariance matrix is related to weighting which is impacted to error. The result of the analysis show that the error which is generated from data simulation in [4] is distinguished based on weighting. The error of the simulated data without weighted has better value than the estimated error of the
nonparametric bi-response regression model with the covariance matrix. It produces the average error of covariance matrix rate has smaller error range is approximately to zero.

In 2019, the study has done by Nema Salem and Saher Hussein [7]. They performed data reduction using principal component analysis on iris datasets [7]. In a previous study [8] modified the a priori algorithm for sales data analysis. From the previous research, the author will optimize the sales data used in the association method, one of which is a priori, by reducing sales data on preprocessing data using principal component analysis.

There are 1200 transactions with 1397 products in transactions taken from sales data at a supermarket. This data is not optimal for use in sales analysis if the sales product is not reduced first, because in the raw data there are products whose sales are very small so that it is considered not having a big influence on the need for product sales analysis in that period. As an example of a product that has only been sold once a month, it requires the role of principal component analysis in optimizing the data to be processed into ready-to-use data for sales analysis data which is generally used by the association method.

In sales data there are 1397 types of products sold out of 1200 transactions as shown as table 1:

| No. | No. Transaction   | Product 1 | Product 2 | ... | Product 1397 |
|-----|------------------|-----------|-----------|-----|--------------|
| 1   | 2018SA18019001501| 1         | 0         |     | 0            |
| 2   | 2018SA18006005628| 0         | 1         |     | 0            |
| ... | ...              | ...       | ...       |     | ...          |
| 1200| 2018SA18006005619| 1         | 1         |     | 0            |

Below is a calculation on the singular value decomposition algorithm process used by principal component analysis to analyze large-scale data matrices, where the matrix (m x n) will be transformed into a matrix (m x k) without reducing data variations. As in [9], a dataset, X with m rows represents a variable and n columns represent observations represented in a matrix with m vector row vectors, each length n. The initial stage is to initialize the matrix of the types of products sold in the transaction:

\[
\text{Initial Matrix} = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
... & ...
\end{bmatrix}
\]

\[
\text{Unit Matrix} = \begin{bmatrix}
1 & 1 & 1 \\
1 & 1 \\
... & ...
\end{bmatrix}
\]

From the matrix processing, the deviation score of the matrix will be sought by first looking for the value of the transpose deviation matrix:

\[
\begin{bmatrix}
0,333333 & -0,666666 \\
-0,666666 & 0,333333 \\
0,333333 & 0,333333
\end{bmatrix}
= \begin{bmatrix}
0,333333 & -0,666666 \\
-0,666666 & 0,333333 \\
0,333333 & 0,333333
\end{bmatrix}
\]

Deviation Score = transpose deviation matrix x deviation matrix values

\[
= \begin{bmatrix}
0,333333 & -0,666666 & 0,333333 \\
-0,666666 & 0,333333 & 0,333333 \\
0,333333 & 0,333333 & 0,333333
\end{bmatrix}
X
= \begin{bmatrix}
0,333333 & -0,666666 & 0,333333 \\
-0,666666 & 0,333333 & 0,333333 \\
0,333333 & 0,333333 & 0,333333
\end{bmatrix}

= \begin{bmatrix}
0,666666 & -0,333333 \\
-0,333333 & 0,666666
\end{bmatrix}
\]
After the deviation matrix value and deviation matrix score are obtained, the next step is to find the variance covariance value by:

\[
\text{Variance Covariance} = \left[ \frac{0.666666 - 0.33333}{-0.33333} \right] \quad (2)
\]

\[
\text{Variance Covariance} = \left[ \frac{0.33333 - 0.1666}{-0.1666} \right]
\]

After the pre-principal component analysis stage is complete, it is continued with the principal component analysis process where the variance covariance matrix results are processed to get the singular value decomposition value. The value in the variance covariance matrix will be used as the initial matrix value to process data in the singular value decomposition process symbolized by A and get a singular value decomposition matrix in the principal component analysis process, namely:

\[
A = \begin{bmatrix}
0.33333 & -0.1666 \\
-0.1666 & 0.33333
\end{bmatrix}
\]

Step 1. Calculate Transpose \( A^\top \) and \( A^\top A \):

\[
A^\top = \begin{bmatrix}
0.33333 & -0.1666 \\
-0.1666 & 0.33333
\end{bmatrix}
\]

\[
A^\top A = \begin{bmatrix}
0.33333 & -0.1666 \\
-0.1666 & 0.33333
\end{bmatrix} \times \begin{bmatrix}
0.33333 & -0.1666 \\
-0.1666 & 0.33333
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0.1385 & -0.1106 \\
-0.1106 & 0.1385
\end{bmatrix}
\]

Step 2. Find the eigenvalues of \( (A^\top A) \), and sort in descending order, in absolute terms, and take the square root to get a singular value from the matrix A:

\[
\begin{vmatrix}
0.1385 - c & -0.1106 \\
-0.1106 & 0.1385 - c
\end{vmatrix} = (0.1385 - c)(0.1385 - c) - (-0.1106)(-0.1106) = 0
\]

and from the above equation, the characteristic equation is:

\[
C^2 - 0.277C + 0.007 = 0
\]

This quadratic equation produces 2 roots of the equation, in this case called eigenvalues, namely:

\[
C = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\]

with the formula for finding the root value above, the eigenvalues are as follows: \( C_1 = 0.249 \) and \( C_2 = 0.028 \). From this eigenvalue, you can get the S value, which is the singular value \( S_1 = \sqrt{0.249} = 0.499 \) and \( S_2 = \sqrt{0.028} = 0.1673 \).

Step 3. Construct a diagonal S or singular value matrix by placing the single values in descending order along the diagonal line and calculating the inverse:

\[
S = \begin{bmatrix}
0.499 & 0 \\
0 & 0.1673
\end{bmatrix}
\]

Then, calculate the inverse of the matrix S above,

\[
S^{-1} = \begin{bmatrix}
2.016 & 0 \\
0 & 6.012
\end{bmatrix}
\]

Step 4. Use the eigenvalues ordered from step 2 and calculate the eigenvectors from \( A^\top A \). Place the eigenvalues of this vector along column \( V \) and compute the transpose, \( V^\top \):
For the eigenvalue $C_1 = 0.249$, then

$$A^\top \cdot A_{\text{cl}} = \begin{bmatrix} 0.1385 - 0.249 & -0.1106 \\ -0.1106 & 0.1385 - 0.249 \\ -0.1106 & -0.1106 \\ -0.1106 & -0.1106 \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} -0.1106 \\ -0.1106 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(-0.1106 X_1) + (-0.1106 X_2) = 0$$

Solve for $X_2$ in the two equations above, $X_2 = - X_1$

$$X_1 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ -X_1 \end{bmatrix} \quad (6)$$

Then divide the result by the length of the matrix, therefore first calculate the length:

$$\text{Length (L)} = \sqrt{X_1^2 + X_2^2} = X_1 \sqrt{2}$$

$$X_1 = \begin{bmatrix} X_1/L \\ -X_1/L \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix} \quad (7)$$

Eigen Value $C_2 = 0.028$, then

$$A^\top \cdot A_{\text{cl}} = \begin{bmatrix} 0.1385 - 0.028 & -0.1106 \\ -0.1106 & 0.1385 - 0.028 \\ -0.1106 & -0.1106 \\ -0.1106 & -0.1106 \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} -0.1106 & -0.1106 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$(-0.1106 X_1) + (-0.1106 X_2) = 0$$

Solve for $X_2$ in the two equations above, $X_2 = X_1$

$$X_2 = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_1 \end{bmatrix} \quad (8)$$

Then divide the result by the length of the matrix, therefore first calculate the length:

$$\text{Length (L)} = \sqrt{X_1^2 + X_2^2} = X_1 \sqrt{2}$$

$$X_2 = \begin{bmatrix} X_1/L \\ X_1/L \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}$$

after obtaining the $X_1$ and $X_2$ values at this stage, then calculate the vector value (V) and transpose the matrix results ($V^\top$):
\[ V = [X_1 \quad X_2] = \begin{bmatrix} 0.7071 & 0.7071 \\ -0.7071 & 0.7071 \end{bmatrix} \]

\[ V^T = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix} \] (9)

Step 5. Calculate the value of the matrix \( U \) or the so-called vector eigenvalues with the solution \( U = A V S^{-1} \), and to complete the full singular value decomposition (singular value decomposition) using \( A = USV^T \):

\[ U = A V S^{-1} \] (10)

\[ U = \begin{bmatrix} 0.333 & -0.166 \\ -0.166 & 0.333 \\ -0.166 & 0.333 \\ -0.7113 & 0.7099 \\ -0.7113 & 0.7099 \end{bmatrix} \begin{bmatrix} 0,7071 & 0,7071 \\ -0.7071 & 0.7071 \\ 1,4255 & 4,2511 \\ -1,4255 & 4,2511 \end{bmatrix} \begin{bmatrix} 2,016 & 0 \\ 0 & 6,012 \\ 1,4255 & 4,2511 \\ -1,4255 & 4,2511 \end{bmatrix} \]

\[ A = USV^T \] (11)

\[ A = \begin{bmatrix} 0.7113 & 0.7099 \\ -0.7113 & 0.7099 \\ 0,499 & 0 \\ 0 & 0,1673 \end{bmatrix} \begin{bmatrix} 0,7071 & -0,7071 \\ 0 & 0,1673 \end{bmatrix} \begin{bmatrix} 0,7071 & -0,7071 \\ 0 & 0,1673 \end{bmatrix} \begin{bmatrix} 0,3528 & -0,3528 \\ 0,1183 & 0,1183 \end{bmatrix} \]

\[ \begin{bmatrix} 0,3348 & -0,167 \\ -0,167 & 0,3348 \end{bmatrix} \]

The value above is normalized to be:

\[ A = USV^T = [1 \quad 0] \begin{bmatrix} 0 \\ 1 \end{bmatrix} \]

The value of the matrix \( A \) above is the full value of the principal component analysis process in the singular value decomposition which will later be normalized so that it can be continued in the next process. It appears that there is data transformation in the matrix.

3. Results and Analysis
The eigen value, variability and cumulative values of the principal component analysis process is shown as Table 2.

| Table 2. Eigen analysis of the Correlation Matrix |
|-----------------------------------------------|
| Component | Eigen value | Variability | Cumulative |
|-----------|-------------|-------------|------------|
| F1        | 32,240      | 0,5872      | 0,5872     |
| F70       | 19,532      | 0,3736      | 0,5891     |
| F140      | 9,448       | 0,178       | 0,8076     |
| F210      | 4,364       | 0,431       | 0,9165     |
| F280      | 2,015       | 0,172       | 0,9808     |
| F350      | 1,015       | 0,185       | 0,9962     |
| F420      | 1,001       | 0,142       | 0,9970     |
| F490      | 0,456       | 0,083       | 0,9979     |

| Table 3. Total Variance |
|-------------------------|
| Component | Eigen Value | % Variance of Component | % Variance of Cumulative |
|-----------|-------------|-------------------------|-------------------------|
| F1        | 32,240      | 58,72                   | 58,72                   |
| F70       | 19,532      | 37,36                   | 58,91                   |
| F140      | 9,448       | 17,8                    | 80,76                   |
| F210      | 4,364       | 43,1                    | 91,65                   |
| F280      | 2,015       | 18,5                    | 98,08                   |
The result is interpreted in Table 3 and is explained as below:

a. There are 486 component that have an eigen value greater than one, namely component 1 to 486.

b. Component 1 can explain as much as 58.72% of the variability of all initial variables (Item / Product Types) in all transaction data.

c. The size of the relationship of each item in the transaction can be represented by 486 selected component, which as a whole can explain 99.70% of the variability of sales of 1397 items.

| Product Code | F350 | F420 | F490 |
|--------------|------|------|------|
|               | 1,015| 1,001| 0,456|
|               | 17,2 | 14,2 | 8,3  |
|               | 99,62| 99,70| 99,79|

Figure 2. Scree Plot Principal Component Analysis

Figure 2 shows the scree plot of Principal Component Analysis. Judging from the Scree-Plot based on the Eigenvalue criterion, which is greater than one, the number of component obtained is 486. So it can be said that the number of component produced is 486 component.

Table 4. Pre-Reduction Product

| Product Code | Product Code |
|--------------|--------------|
| F1           | 1722290101   |
| F70          | 1690380101   |
| F140         | 1447890102   |
| F210         | 1741900101   |
| F280         | 630080109    |
| F350         | 2285870101   |
| F420         | 1979390101   |
| F490         | 304900106    |
| F560         | 303180101    |
| F630         | 303520118    |
| F700         | 27590104     |
| F770         | 1809060102   |
| F840         | 1901380102   |
| F910         | 2183100102   |
| F980         | 123270103    |
| F1050        | 161310106    |
| F1120        | 204500101    |
| F1190        | 00217770101  |
Table 4 shows the types of products in the transaction data that occurred in 1200 transactions. And the product above will be reduced by principal component analysis so that it can get a representative product to be processed or used. Product data that can represent in this case is product data that has a high transaction rate, and principal component analysis will play a role in analyzing this.

Table 5. Initial Matrix of Principal Component Analysis

| No. | F1    | F2    | F3    | F4    | F5    | F6    | F7    | F8    | F9    | ... | F1397 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|
| 1   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 2   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 3   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 4   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 5   | 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 6   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 7   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 8   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| 9   | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |
| ... |       |       |       |       |       |       |       |       |       |     |       |
| 1200| 0     | 0     | 1     | 0     | 0     | 0     | 0     | 0     | 0     |     | 0     |

The matrix is displayed on table 5 is the initial matrix in the principal component analysis process, where the normalized dataset is entered into the matrix, and this matrix is 1200 x 1397 in size according to the number of transactions and the number of items in the dataset. This matrix will be calculated singular value decomposition in the principal component analysis process as described in the previous chapter. When the principal component analysis process is running, namely calculating the value of the matrix deviation, deviation score, variance covariance matrix and finally the singular value decomposition, then the final result matrix of this principal component analysis can be seen in table 6.

Table 6. Matrix After principal component analysis Process

| No. | F1    | F2    | F3    | F4    | F5    | F6    | F7    | F8    | F9    | ... | F486 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|
| 1   | 0.003529 | -0.01530 | -0.05236 | -0.01425 | -0.05369 | -0.07837 |
| 2   | -0.00148 | 0.004625 | 0.027598 | 0.002767 | 0.005651 | -0.03776 |
| 3   | 0.001173 | 0.012491 | -0.08608 | 0.019514 | 0.005876 | -0.04409 |
| 4   | -0.00019 | 0.024039 | -0.00004 | 0.015188 | -0.014793 | -0.01706 |
| 5   | 0.001747 | 0.001392 | 0.120140 | 0.013003 | 0.021286 | 0.043584 |
| 6   | -0.00287 | 0.021208 | -0.08431 | -0.00051 | 0.019891 | -0.00201 |
| 7   | 0.006614 | -0.02438 | -0.00887 | -0.00847 | -0.02909 | -0.06540 |
| 8   | -0.00226 | 0.021142 | 0.078784 | 0.005090 | 0.015887 | -0.04137 |
| 9   | 0.022180 | 0.012088 | -0.13310 | 0.037155 | 0.043673 | 0.096113 |
| ... |       |       |       |       |       |       |       |       |       |     |       |
| 1200| -0.00419 | 0.054871  | 0.300704 | 0.000799  | 0.043884  | 0.034105 |

Table 7. Post Reduction Products

| Component | Eigen Value | % Variance of Component | % Variance of Cumulative | Code of Product |
|-----------|-------------|-------------------------|--------------------------|-----------------|
| F1        | 32,240      | 58,72                   | 58,72                    | 1722290101      |

8
The results of the principal component analysis run by the system according to the process steps described in the previous chapter can be seen in tables 6 and table 7, where the matrix value is visible, namely the singular value decomposition. Based on the information, we conclude that this principal component analysis process has reduced the value of the matrix, which in this case is the product. Moreover, the main component of the products formed are 486 products that can represent the next research process. The reduction of items in this case has been trimmed by around 65.21% of the original data. The data reduction that occurs is shown in Figure 3.

Table 6 and Table 7

|   | Original Value | Reduced Value |
|---|----------------|---------------|
| F70 | 19,532         | 37,36         |
| F140 | 9,448          | 17,8          |
| F210 | 4,364          | 43,1          |
| F280 | 2,015          | 18,5          |
| F350 | 1,015          | 17,2          |
| F420 | 1,001          | 14,2          |
| F490 | 0,456          | 8,3           |

Figure 3. Product Reduction Result Graph

4. Conclusion
There was a cut or reduction of 65.21% attributes, namely from 1397 products to 486 which had a big influence on the sales analysis process at a later stage, this happened because the frequency of product appearance in all transactions was very small which was explained by the singular value decomposition value below 0.1 or after normalization the value is 0. The data resulting from this reduction are ready for use in sales analysis using the association method.

References
[1]. Sehgal S, Singh H, Agarwal M, Bhasker V and Shantanu 2014 Data analysis using principal component analysis Int. Conf. on Medical Imaging, m-Health and Emerging Communication Systems (MedCom) pp. 45-48 doi: 10.1109/MedCom.2014.7005973.
[2]. George H. Dunteman, Principal Component Analysis, 1989 Series: Quantitative applications in the social science.
[3]. Yang J , Zhang D, Frangi A F, and Yang J Y 2004 Two-Dimensional PCA: A New Approach to Appearance-based Face Representation and Recognition IEEE Transaction on Pattern Analysis and Machine Intelligence
[4]. Islamiyati A, Fatmawati and Chamidah N 2018 Estimation of Covariance Matrix on Bi-Response Longitudinal Data Analysis with Penalized Spline Regression J. Phys.: Conf. Ser. 979 012093
[5]. Abdi H. Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD). In: Salkind NJ, ed. Encyclopedia of Measurement and Statistics. Thousand Oaks: Sage Publications 2007 907-912
[6]. David P 2015 *Linear Algebra: A Modern Introduction* 4th ed. Cengage Learning

[7]. Salem N, and Hussein S 2019 Data dimensional reduction and principal components analysis. Procedia Computer Science 163 p. 292–299. https://doi.org/10.1016/j.procs.2019.12.111

[8]. Ginting D S, Mawengkang H, and Efendi S 2018 Modification of A priori Algorithm focused on confidence value to association rules. IOP Conference Series: Materials Science and Engineering, 420 1. https://doi.org/10.1088/1757-899X/420/1/012

[9]. Jolliffe I T 2002 *Principal Component Analysis* 2nd ed. Springer series in statistics.