TEM Journal. Volume 11, Issue 3, pages 1399-1405, ISSN 2217-8309, DOI: 10.18421/TEM113-51, August 2022.

Association Rules Mining Regarding the Value of Business Intelligence Solutions

Petr Havel [1], Manomeet Gupta [1], Athanasios Podaras [2]

[1] Faculty of Economics, Department of Informatics, Technical University of Liberec, Voronezska 13, 46001, Liberec 1, Czech Rebublic,
[2] Faculty of Economics, Department of Informatics, Voronezska 13,46001, Liberec 1, Czech Rebublic

***Abstract* – The paper investigates the importance of business intelligence solutions in modern enterprises using association rule mining techniques. The research is based on a questionnaire addressed to different employee target groups regarding their age interval, their employment status, their domain of employment, their experience or inexperience with business intelligence tools and their positive or negative aspect regarding the importance of business intelligence in modern companies. 90 responses have been received and used for dataset formulation. Using the association rule induction standard procedure, the most popular rules with respect to different antecedent item combinations and business intelligence value as consequent item have been inferred setting as minimum confidence 50% and minimum support 0,1. The collected data have been prepared in common separated values format and the association rules have been inferred using the R- Package. In general, among other rules, a strong relation between BI experience and positive BI aspect can be reported which is also confirmed via simple Pearson X^2 statistical test in R. An investigation paradox which has been spotted is the negative opinion regarding the BI usefulness stemming from a minority of respondents familiar with BI tools.**

DOI: 10.18421/TEM113-51
https://doi.org/10.18421/TEM113-51

Corresponding author: Petr Havel,
Faculty of Economics, Department of Informatics, Technical University of Liberec, Czech Republic.
Email: havelpetr2@gmail.com

Received: 28 June 2022.
Revised: 12 August 2022.
Accepted: 18 August 2022.
Published: 29 August 2022.

***Keywords* – association rule mining, business intelligence importance, Apriori Algorithm, R-package.**

1. Introduction

Business intelligence solutions are nowadays the norm in data manipulation and the base for crucial decision making [3]. The existence of multiple data sources and formats, the complexity of the modern business environment, the speed of the required data processing, analysis and knowledge extraction for the best possible process implementation and customer service provision are only some of the various challenges that modern enterprises and organizations are asked to tackle in order to measure their performance, maintain their competitive position and generate a positive aspect of the customers and citizens.

The importance of the business intelligence has been nowadays realized and highlighted in many studies [2]. Multiple business intelligence academic researchers have been focused on the dominant role of BI solutions towards effective decision making [1]. A study conducted by [4] indicates the paramount importance of the BI adoption in higher education. The specific study is a literature review of the BI effectiveness and the sources used stem from multiple different scientific databases. Another study focuses on the efficiency of BI tools in the online marketing [5]. Other researchers deal with the adoption of BI systems in healthcare [7] for effective decision support in the given sector [6]. Thus, the importance of BI is evident by the available academic literature.

Nevertheless, some researchers refer to specific cases where doubts regarding the effectiveness of BI tools have been expressed by business managers. For example, in [5] refer to the fact that low number of executives in low number of business leaders are adopting BI solutions for marketing intelligence and small – medium enterprises hardly comprehend the value and the benefits of BI tools due to lack of skills in using novel information technology solutions. Despite this statement, [8] indicates that the majority of the most successful enterprises adopt BI tools for their core business decisions and activities.

From the above-mentioned contradictory statements, it has been considered necessary set as a primary goal for the present research the further investigation of the positive or negative employees' aspect of the BI importance and the role of different factors on the formulation of this opinion namely, the age, the BI experience, the employment domain, and the employment status. The research is based on data collected via an electronic questionnaire, prepared in google forms. The responses have been used to create the dataset which has been used as input for association rules mining in order to discover hidden knowledge regarding the influence of the involved demographic variables in the BI positive or negative opinion formulation. Association rules mining based on the Apriori Algorithm has been already used in previous demographic studies [10]. The data that have been collected for the current research has been processed in the R-Package [9] for association rule inference.

The rest of the present paper is organized as follows: the research procedure is modelled and preliminaries regarding the utilized tools and methods are provided. Then, the dataset formulation is explained. Additionally, the results and discussion section are devoted to the analysis and presentation of the conducted research and the analysis of the inferred rules. The conclusions section includes the summarized research achievement and the potential future research steps.

2. Research Procedure

As stated in the introduction, the primary motivation for conducting the current research has been the contradictory aspect regarding the importance of BI solutions by business managers. As a result, the available literature has been used as an inspiration to define the research variables. Specific research questions listed in Table 1. have been considered as categorical variables for association rule mining investigation.

The research procedure depicted in Figure 1. has been followed to infer potential association rules according to which the consequent item is the BI importance positive or negative aspect while the antecedent item combinations have been based on the rest of the dataset variables.

Table 1. The utilized questionnaire for the BI importance investigation

Research Question	Possible Answers
Which is your age?	18-23 24-28 29-35 >35
Which is your employment status? (Multiple selection was possible)	Employed, Part-Time Job, Studying Running own business,
Do you have BI experience?	Yes/No
Do you know what BI is?	Yes/No
Which is the domain of your employment? (Multiple selection was possible)	Education, Banking, Insurance, Manufacturing, Health, Other
Do you find BI Useful for modern enterprises?	Yes/No

Figure 1. The research procedure for Association Rule Mining regarding BI importance

The most important information which has been used for investigation is the influence of different dataset items on the decision whether a BI solution is valuable and important or not. As a consequence, association rule metrics recognizes concepts like *confidence*, *support* and *lift*. Basic definitions of the aforementioned concepts are below mentioned and are follow the definitions provided by [12].

- *Support* indicates the frequency a given item appears in a dataset.
- *Confidence* indicates how often a rule has been confirmed to be true, showing in this way the rule's reliability.
- *Lift* is a metric to measure the ratio of the confidence of items appearing together if they were not dependent in a statistical manner. The equations regarding the *support* (1) and *confidence* (2) measures are below mentioned [10]:

$$SUPPORT\{X => Y\} = P(X \cap Y) \qquad (1)$$

Where $P(X \cap Y)$ is the *proportion* of all transactions in a dataset with combined itemset X and Y.

$$CONFIDENCE\{X => Y\} = P(Y|X) = \frac{P(X \cap Y)}{P(X)} \qquad (2)$$

The *lift* measure can be estimated based on the following formula [13] (3):

$$LIFT\{X => Y\} = \frac{CONFIDENCE\{X=>Y\}}{SUPPORT(Y)} \quad (3)$$

The *lift value greater than 1* indicates strong relation between items X, Y.

3. Tools and Methods

Data Collection and Dataset Formulation

The questions included in Table 1. have been organized in an electronic google form questionnaire and have been submitted in various companies in Liberec including students from the Economics Faculty, Technical University of Liberec. 90 responses have been obtained. The google forms export the received responses to an Excel format so it has been easy to process the collected data and organize it into a .csv (comma separated values) format. The original questionnaire included more questions. Currently the questions of major importance have been included in this research. Furthermore, in the final dataset we applied duplicates removal which resulted into a dataset of 69 records. As a result, the present investigation relies exclusively on 6 categorical variables, 3 receive a binary decision while the responses regarding the rest can rely on multiple selection and include more than 3 categories.

Association Rules

```
install.packages("arules")
library("arules")
install.packages("arulesViz")
library("arulesViz")
```

From the collected data we attempted to report the top 10 (10 most popular) rules with minimum SUPPORT=0.1 and min Confidence=0.5 or 50%, where the consequent variable has been the *BI_ Usefulness* for which a positive opinion has been reported (YES).

In that case the script is formulated as follows:

```
BI_USEFULNESS_rules1 <- apriori(data=dataset, parameter=list (supp=0.1,conf = 0.5), appearance = list (rhs="BI_VALUE=Yes"))
inspect(head(sort(BI_USEFULNESS_rules1, by = "confidence"), 10))
```

In the case that the consequent item is the *BI_ Usefulness* for which a negative opinion has been reported (NO), the script is the following:

```
BI_USEFULNESS_rules1 <- apriori(data=dataset, parameter=list (supp=0.1,conf = 0.5), appearance = list (rhs="BI_USEFUL=No"))
inspect(head(sort(BI_USEFULNESS_rules1, by = "confidence"), 10))
```

In the aforementioned R code the *rhs* represents the consequent item with its value. The antecedent item is reported in the results of the executed code as *lhs*.

4. Results and Discussion

After the execution of the aforementioned code, two different reports have been obtained. The results are depicted in Table 2. and Table 3. The former table shows the positive opinion of the respondents, while the latter illustrates the negative opinion of the respondents regarding the utilization of BI tools in the modern enterprises and organizations. Both tables include the inferred rules based on the antecedent item, or combinations of items, and how often (count or support) the antecedent item appears together with a consequent item, with which confidence level and with which lift value. It can be observed in both tables that the positive or negative aspect of the respondents is mainly based on their experience (0.36 support, 0.92 confidence, 1.38 lift) or inexperience with BI tools (0.30 support and 0.50 confidence, 1.5 lift) respectively. However, the most frequent lhs item has been the BI_Knowledge in both tables (6/10 records in Table 2. and 6/7 records in Table 3.). We can thus confidently report that the positive or negative aspect of the respondents relied on whether the respondents know what BI tools are or whether they have any experience with BI Tools or not. Both the inferred rule categories are also represented via the corresponding plots in Figure 2. and Figure 3. For the respondents with the positive aspect 37 rules have been inferred, while for the respondents with negative aspect only 7 rules have been generated. In each case only the 10 top rules are reported.

Table 1. The inferred rules for the positive opinion regarding the BI Tools usefulness in modern companies
Source: Own work

Rule No	lhs	rhs	Support	Confidence	Lift
1.	{EMPLOYMENT=Studying, BI_KNOWLEDGE=Yes}	=> {BI_USEFUL=Yes}	0.10	1.00	1.50
2.	{AGE=18-23, BI_KNOWLEDGE=Yes}	=> {BI_USEFUL=Yes}	0.13	1.00	1.50
3.	{DOMAIN=Other, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.11	1.00	1.50
4.	{BI_KNOWLEDGE=Yes, DOMAIN=Other, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.11	1.00	1.50
5.	{BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.36	0.92	1.38
6.	{BI_KNOWLEDGE=Yes, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.36	0.92	1.38
7.	{AGE=24-28, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.17	0.92	1.38
8.	{AGE=24-28, BI_KNOWLEDGE=Yes, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.17	0.92	1.38
9.	{AGE=24-28, BI_KNOWLEDGE=Yes, DOMAIN=Other}	=> {BI_USEFUL=Yes}	0.13	0.90	1.35
10.	{DOMAIN=Manufacturing, BI_EXP=Yes}	=> {BI_USEFUL=Yes}	0.11	0.88	1.35

Table 2. The inferred rules for the negative opinion regarding the BI Tools usefulness in modern companies
Source: Own work

Rule No	lhs	rhs	Support	Confidence	Lift
1.	{BI_KNOWLEDGE=No}	=> {BI_USEFUL=No}	0.24	0.60	1.82
2.	{BI_KNOWLEDGE=No, BI_EXP=No}	=> {BI_USEFUL=No}	0.24	0.60	1.82
3.	{AGE=24-28, BI_KNOWLEDGE=No}	=> {BI_USEFUL=No}	0.13	0.60	1.80
4.	{AGE=24-28, BI_KNOWLEDGE=No, BI_EXP=No}	=> {BI_USEFUL=No}	0.13	0.60	1.80
5.	{BI_KNOWLEDGE=No, DOMAIN=Other}	=> {BI_USEFUL=No}	0.11	0.57	1.71
6.	{BI_KNOWLEDGE=No, DOMAIN=Other, BI_EXP=No}	=> {BI_USEFUL=No}	0.11	0.57	1.71
7.	{BI_EXP=No}	=> {BI_USEFUL=No}	0.30	0.50	1.50

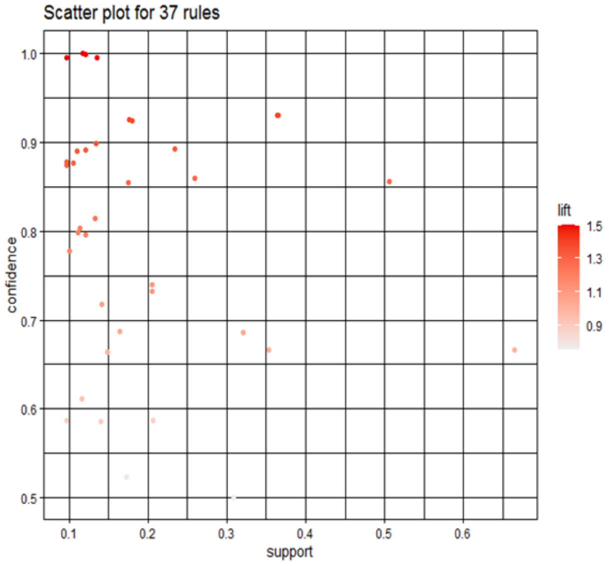

Figure 1. The scatter plot for the positive aspect

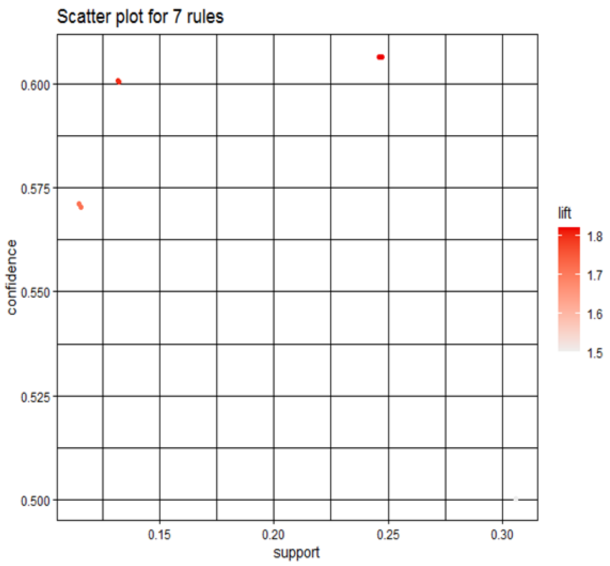

Figure 2. The scatter plot for the negative aspect regarding

The strong relation between the variables BI_Experience and BI_Usefullness or BI_Value has been also confirmed using Pearson's t-test x^2 in the R-package. The following hypothesis has been taken into account:

H_0(Null Hypothesis): No relation between Variable a = BI_Experience and Variable b (BI_Usefullness, or BI_Value Exists

H_1(Alternative Hypothesis): No relation between Variable a = BI_Experience and Variable b (BI_Usefullness, or BI_Value Exists

To infer the contingency, in Table 4. the following script in R has been executed:

```
table(dataset$BI_EXP, dataset$BI_VAL)
```

Result (0=No, 1=Yes):

Table 3. The contingency table stemming from the test statistic investigation in R regarding the relationship between BI_experience and BI_Value (or importance)

	0	1
0	25	24
1	2	39

To run Pearson's standard x^2 test, the following R code has been executed:

```
chisq.test(dataset$BI_EXP,
dataset$BI_VAL, correct=FALSE)
```

Result: (p-Value = 1.962e-06, Rejection of Null Hypothesis since p-value<0.05)

```
Pearson's Chi-squared test

data: dataset$BI_EXP and dataset$BI_VAL
X-squared = 22.632, df = 1, p-value =
1.962e-06
```

Finally, an investigation paradox is the fact that some respondents who, despite the fact they have been self-characterized as experienced with BI (Rule 1 in Table 5.) or they have declared their awareness of BI tools (Rule 13 in Table 5.), expressed a negative opinion about the usefulness of BI tools. However, this percentage is very small (support=0.01). As a result, the strong relation between BI experience and positive aspect about BI tools remains a fact. The observations of the Table 5. reveal that the respondents are either running their own business, or the majority is over 35 years old. An interesting remark, which requires further research, is the fact that a considerable number of the negative aspects stem from the banking sector. This is a research paradox since BI solutions are strongly supporting complex decisions in demanding business sectors such as the banking domain [11].

Table 4. Rules depicting the negative aspect of the respondents regarding BI tools with minor support (0.01 minimum support value). Surprisingly some experienced and familiar with BI solutions respondents expressed negative opinion regarding the BI Value (Rule 1, Rule 13)

Rule No	lhs	rhs	support	confidence
1	{Employment=Employed, Running Own Business, BI_Experience=Yes}	=> {BI_USEFUL=NO}	0.014	1.00
2	{Age=18-23, Employment= Studying, Running Own Business}	=> {BI_USEFUL=NO}	0.014	1.00
3	{Employment=Studying, Running Own Business, BI_Knowledge=No}	=> {BI_USEFUL=NO}	0.014	1.00
4	{Employment=Studying, Running Own Business, BI_Experience=No}	=> {BI_USEFUL=NO}	0.014	1.00
5	{Age=Over 35, Domain=Banking}	=> {BI_USEFUL=NO}	0.014	1.00
6	{Age=Over 35, Employment= Running Own Business}	=> {BI_USEFUL=NO}	0.014	1.00
7	{Age=Over 35, BI_Knowledge=No}	=> {BI_USEFUL=NO}	0.014	1.00
8	{Age=Over 35, Domain=Other}	=> {BI_USEFUL=NO}	0.014	1.00
9	{Age=Over 35, BI_Experience=No}	=> {BI_USEFUL=NO}	0.014	1.00
10	{Employment= Running Own Business, Domain=Banking}	=> {BI_USEFUL=NO}	0.014	1.00
11	{Employment= Studying, Part – Time Job, Domain=Banking}	=> {BI_USEFUL=NO}	0.014	1.00
12	{Age= 24-28, Domain = Banking}	=> {BI_USEFUL=NO}	0.014	1.00
13	{BI_Knowledge = YES, Domain = Banking}	=> {BI_USEFUL=NO}	0.014	1.00

5. Conclusions

The present article investigated the importance of BI solutions in modern enterprises. The motivation of the research mainly stemmed from the fact that in the literature, a surprisingly negative aspect regarding the value or the importance of BI tools from business managers is reported. Thus, using an electronic questionnaire, demographic research has been conducted including employees' BI aspect regarding the BI importance in modern companies. The gathered data stemmed from respondents from various domains, age intervals, occupation, BI knowledge, BI experience and business domains. Using association rule mining in the R package the strong connection between the BI experience and the positive aspect regarding the BI value or importance has been confirmed despite the negative opinion expressed by the minority of respondents who either run their own business, are older than 35 or stem from the banking sector. The strong relationship between the two variables has been also confirmed using Pearson's chi-square (X^2) test statistic. The research is ongoing and it will be further enriched with more variables, more data records and more data mining tools. A special focus on the banking domain is also considered.

References

[1]. Persson, J., & Sjöö, E. (2017). Business Intelligence– its impact on the decision-making process at higher education institutions: A case study at Karlstad University. Retrieved from: http://www.diva-portal.org/smash/get/diva2:1119142/FULLTEXT01.pdf [accessed: 10 June 2022].

[2]. Agarwal, R., & Dhar, V. (2014). Big data, data science, and analytics: The opportunity and challenge for IS research. *Information systems research*, *25*(3), 443-448.

[3]. Shollo, A. (2011, September). Using business intelligence in IT governance decision making. In *IFIP International Working Conference on Governance and Sustainability in Information Systems-Managing the Transfer and Diffusion of IT* (pp. 3-15). Springer, Berlin, Heidelberg.

[4]. Santi, R. P., & Putra, H. (2018, October). A systematic literature review of business intelligence technology, contribution and application for higher education. In *2018 International Conference on Information Technology Systems and Innovation (ICITSI)* (pp. 404-409). IEEE.

[5]. An, T. S., Fadilah, S. I., Keikhosrokiani, P., & Samsudin, N. H. (2021, December). Exploring the Business Intelligence Efficiency in Organizations Among Employees in Penang, Malaysia. In *2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (Vol. 6, pp. 1-7). IEEE.

[6]. Ally, S. S., & Khan, N. (2016, December). Data warehouse and BI to catalize information use in health sector for decision making: A case study. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 92-97). IEEE.

[7]. Ali, O., Crvenkovski, P., & Johnson, H. (2016, October). Using a business intelligence data analytics solution in healthcare. In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 1-6). IEEE.

[8]. Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88-98.

[9]. Rahlf, T. (2019). Data for everybody. In *Data Visualisation with R* (pp. 1-4). Springer, Cham.

[10]. Huang, M. J., Sung, H. S., Hsieh, T. J., Wu, M. C., & Chung, S. H. (2020). Applying data-mining techniques for discovering association rules. *Soft Computing*, *24*(11), 8069-8075.

[11]. Zelenka, M., & Podaras, A. (2021). Increasing the Effectivity of Business Intelligence Tools via Amplified Data Knowledge. *Studies in Informatics and Control*, *30*(2), 67-77.

[12]. Teoh, T. T., & Rong, Z. (2022). *Artificial Intelligence with Python* (pp. 3-328). Springer.

[13]. Han, J., Kamber, J. and Pei, M. (2011). *Data Mining: Concepts and Techniques.* Elsevier.