

Establishing Standard Rules for Choosing Best KPIs for an E-Commerce Business based on Google Analytics and Machine Learning Technique

Haris Ahmed¹, Dr. Tahseen Ahmed Jilani², Waleej Haider³, Mohammad Asad Abbasi⁴, Shardha Nand⁵
Saher Kamran⁶

^{1,3,5,6}Sir Syed University of Engineering and Technology, Karachi, Pakistan

²University of Nottingham, Nottingham, UK

⁴National University of Computer and Emerging Sciences, Karachi, Pakistan

Abstract—The predictable values that indicate the performance of any company and determine that how well they are performing in order to achieve their objective is referred by the term called as “key performance indicators”. The key performance indicator techniques and other methods that are similar to KPI are usually implemented in the businesses that are running online, but for an e-commerce business, it is always difficult to select the right KPI. As long as the KPIs are concerned, the biggest blunder that an online business can make is that they calculate everything along with the KPIs. But whatever they are calculating cannot be referred as the “key” because they are measuring each and everything, so this can immediately become devastating. The need is to only measure certain specific keys in order to calculate the performance of a business. The main aim of this research is to establish the set of standard rules that must be adopted in order to identify the best KPIs for an e-commerce business website based on google analytics and machine learning technique.

Keywords—E-commerce KPI; Google Analytics; Machine Learning; C4.5 Decision Tree; Weka J48

I. INTRODUCTION

This paper will briefly give the overview of the Google analytics and also highlight the Google analytics KPIs for the websites based on e-commerce. It will also elaborate the techniques of machine learning in order to find the right KPI for the business based on e-commerce. Google analytics is the most useful and common tool of measuring and monitoring the performance of any website [1]. Key performance indicators are the method by which we can calculate the performance of something. As long as the e-commerce websites are concerned the KPIs are the multiple features that provide assistance to the webmasters or owners to determine the performance of their websites. In e-commerce websites, the KPIs have their own importance because these are the tools to measure the success of any website. There are many KPIs for the e-commerce website, some of which are: Website

traffic, the rate of conversion, bounce rate, purchase time, repetition of the visits, abandon rate of the cart, conversion cost. For the success of any website, it is important for the webmaster to keep track of all of the above mentioned KPIs [2]. In this paper, the machine learning technique is described briefly in order to find the right KPIs for the online business. There are various machine learning techniques but in this research paper, we will focus on the decision tree in order to find out the right KPI on the e-commerce business. A decision tree is a simple tree on which the non-terminal nodes depict the test on one or more than one traits whereas the terminal nodes depict the decision results. The initial decision tree that was algorithm induction ID3 [3] was promoted by C4.5. This paper provides the extensive study on the topic of Creating standard Rules for Choosing Best KPIs for an e-Commerce business based on Google Analytics through Machine Learning.

II. GOOGLE ANALYTICS INSTALLATIONS

To track the KPIs of the e-commerce sites by using the GA software it is necessary to first install the tracking codes on such websites.

A. Google Analytics Tracking Code Installation Steps

- 1) On your analytics account first sign in.
- 2) Click on admin tab.
- 3) In the drop down menu on the column of an account click on account.
- 4) In the property column go to the drop down menu and select property.
- 5) Click on Tracking info and then tracking code under the property tab.
- 6) GA tracking ID and code will be displayed on screen as shown in Figure 1.
- 7) Add tracking code or tracking ID to your site or app to collect the data.

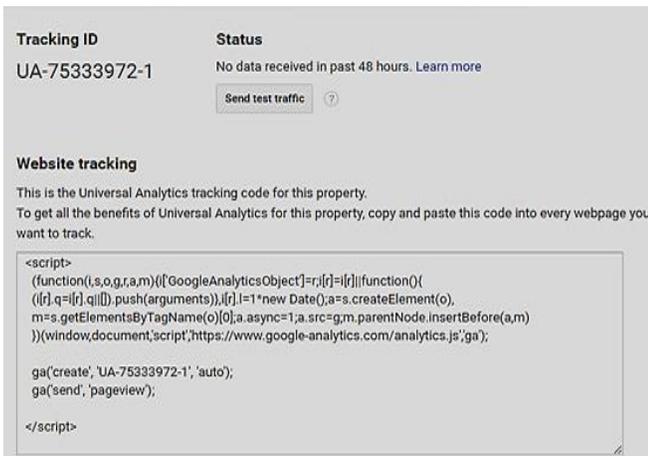


Fig. 1. Google Analytics Tracking ID & Code

B. Google Analytics Audience Overview

In Figure 2, the report on GA audience overview is depicted which clearly shows the overview of the performance of the website of the online business including, session numbers, the number of users; both new and those who are returning to the website, the duration of the average sessions, the bounce rate and the sessions that are new[7].

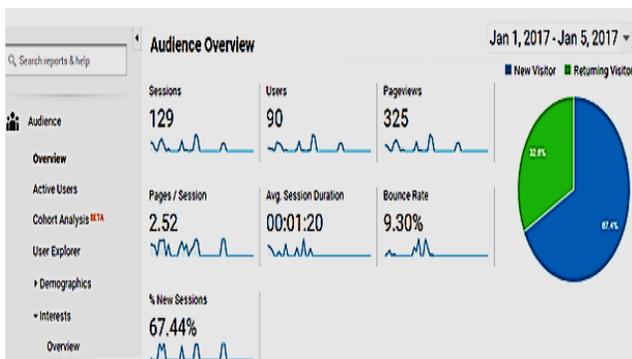


Fig. 2. Google Analytics Audience Overview Report

III. SET UP E-COMMERCE TRACKING

To see E-commerce data in Google Analytics reports, enable E-commerce for each view in which we want to see the data.

A. Enable E-Commerce in GA Reports

- 1) First on analytic account you need to sign in.
- 2) Click Admin then navigate to the view.
- 3) Select e-commerce setting in the column of view.
- 4) Click the Enable e-commerce toggle ON.
- 5) Click on the Next step.
- 6) Click on Submit.

B. Google Analytics Ecommerce Overview

The report that is shown in Figure 3 depicts the report of e-commerce that enables you to examine the activity of purchase on your application. You will be able to see the information about the product and the transaction, average value of the order, the rate of e-commerce conversion purchase time or any other useful information.



Fig. 3. Google Analytics E-commerce Overview Report

IV. KPI FOR AN E-COMMERCE BUSINESS

Common KPIs for an e-commerce business can be categorised into three dimensions presented in Table 1.

- 1) Sales KPIs.
- 2) Marketing KPIs.
- 3) Customer Service KPIs.

TABLE I. E-COMMERCE KPIs DIMENSION

| DIMENSION | KPI |
|----------------|--|
| SALES KPIs | Sales/revenue : Hourly, daily, weekly, monthly |
| | Shopping cart abandonment rate |
| | Conversion rate |
| | Average order size |
| MARKETING KPIs | Brand or display advertising click-through rates |
| | Time on site |
| | Page views per visit |
| | Unique versus returning visitors |
| CUSTOMER KPIs | Bounce rate |
| | Customer service email count |
| | Customer service chat count |

KPIs can help the e-commerce business to make well versed- decisions. The indicators like pages that are visited and the time spent on a certain site help to determine the views of the visitors about your product, whether they are just steering the page or do the window shopping or they are interested in your product [8].

V. RESEARCH METHODOLOGY

The research methodology includes: at the starting point we first locate all the KPIs that are associated with the e-commerce business; then with the help of Google analytics tool we determine the score of KPIs. After the determination of the KPI score, we then find the Correlation between the score of KPIs and the revenue that is generated on monthly basis. The last step is to apply the decision tree C4.5 algorithm to develop the rules for best KPI selection for an e-commerce business. In Figure 4 the steps of research methodology are indicated.

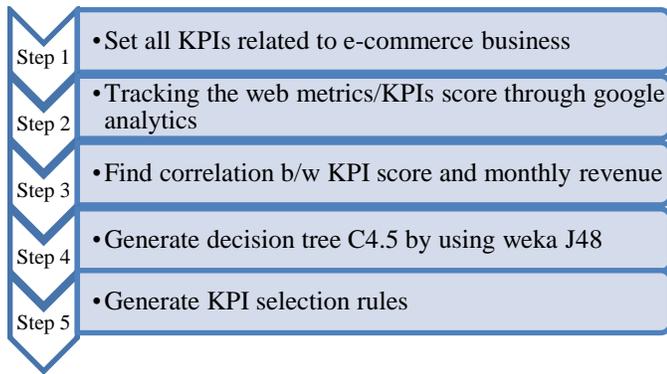


Fig. 4. Research Methodology Steps

VI. E-COMMERCE WEBSITE KPI SCORE

With the help of GA and medium sized online shopping store, the twelve KPIs scores were determined. The Table 2 shows the average scores of each KPI in different duration of time.

TABLE II. E-COMMERCE WEBSITE KPI SCORE

| Sr. No | KPI | May 1, 2016 -Jul 31, 2016 Avg. KPI Score | Aug 1, 2016 -Oct 31, 2016 Avg. KPI Score | Nov 1, 2016 -Jan 31, 2017 Avg. KPI Score |
|--------|--|---|---|---|
| 1 | Revenue | \$15.7K | \$27.5K | \$21.3K |
| 2 | Shopping cart abandonment rate | 26% | 16% | 15% |
| 3 | Conversion rate | 5% | 8% | 3% |
| 4 | Average order size | \$1.5K | \$2.2K | \$1.90K |
| 5 | Brand or display advertising click-through rates | 29% | 22% | 26% |
| 6 | Average Time on site | 189 seconds | 212 seconds | 203 seconds |
| 7 | Average Page views per visit | 3.44 | 2.12 | 2.01 |
| 8 | Unique visitors | 36.76% | 46.4% | 44% |
| 9 | returning visitors | 63.24% | 53.6% | 55% |
| 10 | Bounce rate | 12.76% | 17.87% | 19.84% |
| 11 | Customer service Chat count | 22 | 10 | 17 |
| 12 | Customer service email count | 36 | 28 | 22 |

The major goal of the business of e-commerce is to elevate the success of its sales and to enhance the generation of revenue [9]. As indicated in Table 2, we can find the revenue that is generated on the monthly basis by the use of Google analytics, the statistics show that for the month of May to July, August to October and November to July the average revenue is \$15.7K, \$27.5K and \$21.3K respectively. In order to find the association between the scores of each e-commerce KPI and the revenue; we calculate the correlation coefficient between each e-commerce KPI score and revenue.

VII. THE CORRELATION COEFFICIENT

A correlation coefficient is an arithmetical measure of the amount to which variations to the value of one variable

forecasts variation in the value of another [10]. If the variables are positively correlated with each other than their value increase or decrease simultaneously on the other hand if the values of these variables are negatively correlated than the value of one increases decreasing the value of others.

It is easy to understand the calculation of two coefficients that are correlated with each other. Imagine that these coefficients are **X** and **Y** respectively. The **zX** and the **zY** are the standardised version of the **X** and **Y** and let us suppose that the mean of **zX** and **zY** is 0 and the standard deviation is 1 respectively. In Equation 1 and 2, the re-expressions are used in order to find the identical scores.

$$zX_i = [X_i - mean(X)]/s.d.(X) \tag{1}$$

$$zY_i = [Y_i - mean(Y)]/s.d.(Y) \tag{2}$$

The definition of the correlation coefficient is stated as the mean product of the standardised score that is clearly shown in the equation number 3.

$$r_{X,Y} = sum\ of\ [zX_i \times zY_i]/(n - 1) \tag{3}$$

n shows the size of the sample.

TABLE III. CORRELATION COEFFICIENT B/W KPI SCORES & REVENUE

| Sr. No | KPI | May 1, 2016- Jul 31, 2016 Correlation b/w KPI_Score & Monthly_ Revenue | Aug 1, 2016- Oct 31, 2016 Correlation b/w KPI_Score & Monthly_ Revenue | Nov 1, 2016- Jan 31, 2017 Correlation b/w KPI_Score & Monthly_ Revenue |
|--------|--|---|---|---|
| 1 | Shopping cart abandonment rate | -0.93 | -0.89 | -0.91 |
| 2 | Conversion rate | 0.67 | 0.77 | 0.87 |
| 3 | Average order size | 0.54 | 0.49 | 0.50 |
| 4 | Brand or display advertising click-through rates | 0.30 | 0.34 | 0.32 |
| 5 | Average Time on site(Seconds) | 0.73 | 0.87 | 0.79 |
| 6 | Average Page views per visit | 0.23 | 0.32 | 0.29 |
| 7 | Unique visitors | 0.22 | 0.61 | 0.67 |
| 8 | returning visitors | 0.40 | 0.38 | 0.22 |
| 9 | Bounce rate | -0.23 | -0.62 | -0.69 |
| 10 | Customer service Chat count | 0.23 | 0.32 | 0.73 |
| 11 | Customer service email count | 0.43 | 0.36 | 0.23 |

Table 3 presented the value of correlated coefficient between scores of each KPI and the revenue that is generated in the different duration of time.

As a part of pre-processing the continuous data set of KPI width of the desired intervals, as shown in the Table 4. is transformed into the categorical form by the estimated

TABLE IV. CATEGORICAL PARTITIONING OF E-COMMERCE KPI DATA SET

| Sr. No | KPI | Partitioned data | |
|--------|--|--|---|
| | | Avg. KPI Score | Correlation b/w KPI_Score & Monthly_ Revenue |
| 1 | Shopping cart abandonment rate | {low, medium, high} { <15%,15-20%,>20% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 2 | Conversion rate | {low, medium, high} { <5%,5-10%,>10% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 3 | Average order size (\$) | {low, medium, high} { <1k,1k-3k>3k } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 4 | Brand or display advertising click-through rates | {low, medium, high} { <30%,30-40%,>40% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 5 | Average Time on site(Seconds) | {low, medium, high} { <200,200-400,>400 } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 6 | Average Page views per visit | {low, medium, high} { <2,2-5,>5 } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 7 | Unique visitors | {low, medium, high} { <30%,30-40%,>40% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 8 | returning visitors | {low, medium, high} { <40%,40-55%,>55% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 9 | Bounce rate | {low, medium, high} { <10%,10-15%,>15% } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 10 | Customer service Chat count | {low, medium, high} { <15,15-20,>20 } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |
| 11 | Customer service email count | {low, medium, high} { <25,25-30,>30 } | {weak, moderate, strong} (-ve or +ve) {0.20-0.39 , 0.40-0.59 , 0.60-1.0} |

The Table 5 clearly shows the e-commerce KPIs' data that is converted into the categorical form.

TABLE V. CATEGORICAL E-COMMERCE KPI DATA SET

| Sr. No | KPI | May 1, 2016-Jul 31, 2016 | | Aug 1, 2016-Oct 31, 2016 | | Nov 1, 2016-Jan 31, 2017 | |
|--------|--|--------------------------|--|--------------------------|--|--------------------------|--|
| | | Avg. KPI Score | Correlation b/w KPI_Score & Monthly_ Revenue | Avg. KPI Score | Correlation b/w KPI_Score & Monthly_ Revenue | Avg.KPI Score | Correlation b/w KPI_Score & Monthly_ Revenue |
| 1 | Shopping cart abandonment rate | High | Strong | Medium | Strong | Medium | Strong |
| 2 | Conversion rate | Medium | Strong | Medium | Strong | Low | Strong |
| 3 | Average order size | Medium | Moderate | Medium | Moderate | Medium | Moderate |
| 4 | Brand or display advertising click-through rates | Low | Weak | Low | Weak | Low | Weak |
| 5 | Average Time on site(Seconds) | Low | Strong | Medium | Strong | Medium | Strong |
| 6 | Average Page views per visit | Medium | Weak | Medium | Weak | Medium | Weak |
| 7 | Unique visitors | Medium | Weak | High | Strong | High | Strong |
| 8 | returning visitors | High | Moderate | Medium | Weak | Medium | Weak |
| 9 | Bounce rate | Medium | Weak | High | Strong | High | Strong |
| 10 | Customer service Chat count | High | Weak | Low | Weak | Medium | Strong |
| 11 | Customer service email count | High | Moderate | Medium | Weak | Low | Weak |

In the next stage the categorical data is provided as an input to Weka J4.8 to generate Decision tree C4.5.

VIII. DECISION TREE C4.5

The controlled approach of classification is represented by the decision tree. A decision tree is a type of a tree that is simple to describe on which the non-terminal nodes depict the test on one or more than one traits whereas the terminal nodes depict the decision results. The initial decision tree that was algorithm induction ID3 [3] was promoted by C4.4 [4, 5]. The package of WEKA classifier has its own form of C4.5 that is

referred as J4.8. The C4.5 uses the measures of information gain and ratio of gain as the criteria of splitting respectively [11]. The decision tree follows the steps that are stated below:

Step 1: The assumption is being made that n will be the output test and set T is acting as a tuple training sample for the class label, the training sample T then categorised into the various subsets $\{T1, T2, \dots, Tn\}$. So that we will be able to measure the entropy of the sample T (in bits):

$$info(T) = - \sum_{i=1}^k ((freq(C_i, T)/|T|) \times \log_2(freq(C_i, T)/|T|)) \quad (4)$$

Step 2: According to the particular value of property the sample T is divided, Then the property T 's information entropy is:

$$infox(T) = - \sum_{i=1}^n ((|T_i|/|T|) \times info(T_i)) \quad (5)$$

Step 3: The difference between the original requirement of the information and the new one [6] is referred as the information gain. With the help of Eq. (4) and Eq. (5), we may be able to find a gain standard that is given as:

$$Gain(X) = info(T) - infox(T) \quad (6)$$

Step 4: There is a flaw associated with the gain standard that the examination has many variations from the different

situations of output but on the other hand the gain standard is useful to develop the compact decision tree. So it must be given by standardisation that is indicated as:

$$Split - info(X) = - \sum_{i=1}^n ((|T_i|/|T|) \log_2(|T_i|/|T|)) \quad (7)$$

The gain standard that we find is:

$$Gain - ratio(X) = gain(X)/split - info(x) \quad (8)$$

In Table 6 the KPI training data set is depicted and it is presented as the input to Weka J48

TABLE VI. THE TRAINING KPI DATA SET

| Sr. No. | KPI | Avg. KPI_Score | Correlation b/w KPI_Score & Monthly_Revenue | Class |
|---------|--|----------------|---|-------|
| 1 | Shopping Cart Abandonment Rate | High | Strong | Yes |
| 2 | Shopping Cart Abandonment Rate | Medium | Strong | Yes |
| 3 | Conversion Rate | Medium | Strong | Yes |
| 4 | Conversion Rate | Low | Strong | No |
| 5 | Average Order Size | Medium | Moderate | No |
| 6 | Brand Or Display Advertising Click-Through Rates | Low | Weak | No |
| 7 | Average Time On Site(Seconds) | Low | Strong | No |
| 8 | Average Time On Site(Seconds) | Medium | Strong | Yes |
| 9 | Average Page Views Per Visit | Medium | Weak | No |
| 10 | Unique Visitors | Medium | Weak | No |
| 11 | Unique Visitors | High | Strong | Yes |
| 12 | Returning Visitors | High | Moderate | No |
| 13 | Returning Visitors | Medium | Weak | No |
| 14 | Bounce Rate | Medium | Weak | No |
| 15 | Bounce Rate | High | Strong | Yes |
| 16 | Customer Service Chat Count | High | Weak | No |
| 17 | Customer Service Chat Count | Low | Weak | No |
| 18 | Customer Service Chat Count | Medium | Strong | Yes |
| 19 | Customer Service Email Count | High | Moderate | No |
| 20 | Customer Service Email Count | Medium | Weak | No |
| 21 | Customer Service Email Count | Low | Weak | No |

IX. STEPS TO GENERATE DECISION TREE C4.5 IN WEKA J48

- 1) Develop a set of data by using MS Excel, MS Access or any other tool and then save it in the format of CSV.
- 2) Weka explorer is then started in the next step.
- 3) The next step is to Open your CSV file and then save it in the format of ARFF.
- 4) From the choose button select the J48 after clicking on the classify tab.

- 5) Any suitable test option is selected in the second last step.
- 6) The result will be shown after clicking on the start button this is the last step.

To view the graphical form of a decision tree, click on the option "visualise tree" which is present in pop-up menu of result list. In Figure 5, the graphical view of a tree is shown which is created by Weka J48. [12]

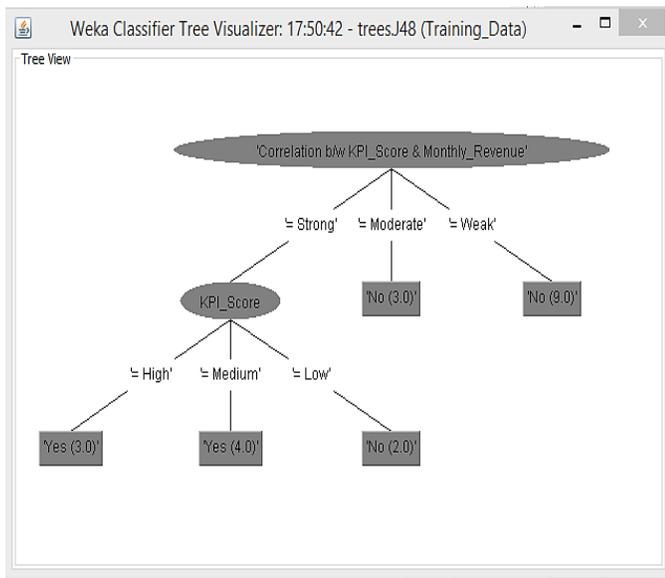


Fig. 5. Graphical Representation of Decision Tree C4.5 Using Weka J48

X. RULE GENERATION USING DECISION TREE FOR CHOOSING BEST KPI FOR AN E-COMMERCE BUSINESS

The Corresponding Rules Are:

R1: IF (Correlation b/w KPI_Score & Monthly_Revenue=Weak) THEN Adopt KPI= No

R2: IF (Correlation b/w KPI_Score & Monthly_Revenue=Moderate) THEN Adopt KPI= No

R3: IF (Correlation b/w KPI_Score & Monthly_Revenue=Strong) AND (Avg. KPI_Score=Low) Then Adopt KPI= No

R4: IF (Correlation b/w KPI_Score & Monthly_Revenue=Strong) AND (Avg. KPI_Score=Medium) Then Adopt KPI= Yes

R5: IF (Correlation b/w KPI_Score & Monthly_Revenue=Strong) AND (Avg. KPI_Score=High) THEN Adopt KPI= Yes

There are two classifications of the rules that are “YES” or “NO”. The following study reveals only one of the decision rule for each of the class and they are stated as below

A. “NO” Class Rule:

R1: IF (Correlation b/w KPI_Score & Monthly_Revenue=Weak) THEN Adopt KPI= No

It specifies that when Correlation between KPI Score and Monthly Revenue is equal to weak then respective KPI is not nominated for calculating the performance of the e-commerce business. In addition to this, nine examples of training data set support the rule.

B. “YES” Class Rule:

R4: IF (Correlation b/w KPI_Score & Monthly_Revenue=Strong) AND (Avg. KPI_Score=Medium) Then Adopt KPI= Yes

It specifies that when Correlation between KPI Score and Monthly Revenue is equal to Strong and avg. score of KPI is equal to Medium then respective KPI is nominated for calculating the performance of the e-commerce business. In addition to this, four examples of training data set support the rule.

XI. CONCLUSION

The tracking of KPI is a great technique to monitor and manage an e-commerce website that connects the consumers to thousands of different e-commerce sellers. Every separate e-commerce website is dissimilar so to enhance the functioning performance with the help of KPIs, each e-commerce seller should choose KPIs which are best related to their online trade. There is no standard rule for choosing correct KPIs for an e-commerce business. In this research, we proposed a technique to develop the standard rules for choosing the best KPI for an e-commerce website through the use of Google analytics and decision tree method.

This research will support online dealers to create more profits by understanding clients’ inclination to purchase and level of satisfaction and to enhance the trust of the client.

REFERENCES

- [1] Rebecca Sentance, (2016).Google Analytics: a guide to confusing terms
- [2] Kohavi, R., & Parekh, R. (2003). Ten supplementary analyses to improve e-commerce web sites. In *Proceedings of the Fifth WEBKDD workshop*.
- [3] Quinlan, J. Ross. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [4] Rokach, Lior, and Oded Maimon. *Data mining with decision trees: theory and applications*. World scientific, 2014.
- [5] J.R. Quinlan, Bagging, Boosting and C4.5, In Proc. 13th National Conf. Artificial Intelligence (AAAI’96), pp. 725-730. Portland, (Aug, 1996).
- [6] Kotsiantis, Sotiris B. "Decision trees: a recent overview." *Artificial Intelligence Review* 39.4 (2013): 261-283.
- [7] Pakkala, Heikki, Karl Presser, and Tue Christensen. "Using Google Analytics to measure visitor statistics: The case of food composition websites." *International Journal of Information Management* 32.6 (2012): 504-512.
- [8] Mistry, Jamshed. "Performance Measurement In The eCommerce Industry." *Journal of Business & Economics Research (JBER)* 1.11 (2011).
- [9] H Ahmed, TA Jilani, S Nand. "Fuzzy Classification Techniques for Online Advertisement Based on User’s Perception in Social Networks." *International Journal of Computer Science and Software Engineering* 5, no. 4 (2016): 49-57.
- [10] Sedgwick, Philip. "Pearson’s correlation coefficient." *Bmj* 345.7 (2012).
- [11] Quinlan, J. Ross. *C4. 5: programs for machine learning*. Elsevier, 2014.
- [12] Bhargava, Neeraj, et al. "Decision tree analysis on j48 algorithm for data mining." *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering* 3.6 (2013).