RESEARCH ARTICLE   OPEN ACCESS

Analysis of Product Comments are Examined in e-Commerce Systems with Text Mining Application

Gökay Yılmaz*, Prof. Dr. Zafer Aslan **
(Department of Computer Engineering, Istanbul Aydin University,
** (Department of Computer Engineering, Istanbul Aydin University,
Corresponding Author: Gökay Yılmaz

ABSTRACT
The general purpose of the study is to classify the comments made on products in e-commerce systems by type using text mining application. As the e-commerce system, OpenCart e-commerce infrastructure was used and the installation steps were explained by installing on the "ecommercecomments.store" domain. In particular, the average internet usage in both Turkey and the world for 15 years, including web-based applications has rapidly increased. The quality of the services offered online continues to increase in parallel. In this context, companies have started to use technology and software for their own interests and have taken steps to increase their service quality by interpreting data that will benefit them through methods such as data mining. With this study, the classification process of these comments/evaluations made on the products was carried out using text mining. In this way, big data can be transformed into more meaningful and smaller data and can be used as a faster method to identify users who have problems. Although it seems difficult to examine the interpretations made especially subjectively with this method, it was emphasized how this situation can be facilitated by text mining applications.
Keywords - e-commerce, text mining, product reviews, classification

---

I. INTRODUCTION

The concept of the internet, which emerged for defense purposes in the early days, spread rapidly with the development of technology, and has taken its place in our lives as an indispensable part of our lives, appearing almost everywhere in our daily lives. Today, almost all needs and activities can be accessed on the internet. The Internet is widely used in new business activities because of its contribution to productivity in commerce and its use of less resources. This commerce and shopping on the Internet are popularly known as electronic commerce (E-commerce). Companies are trying to gain and maintain their place in the market with the increase of the competitive environment should plan their production systems more efficiently and effectively both dynamically and economically [1]. That is why, companies have started to use the strengths of technology and software for their own interests and have taken steps to increase their service quality by interpreting data that will benefit them from complex customer pools through methods such as data mining.

II. PURPOSE

The most general purpose of data mining is to produce information from very large databases that is not understandable at first glance but is useful. Text mining, on the other hand, is a data mining technique that accepts the text itself as the data, allowing to automatically extract previously unknown and potentially useful information from large amounts of unstructured text data in a scalable and repeatable manner.

The aim of this study is to combine the concept of text mining, which is one of the strengths of the software, with the widespread e-commerce systems, to quickly interpret the comments and rating made on the products in the e-commerce systems and to enable the studies to increase the service quality with this information. This will both increase customer confidence and satisfaction and provide profit for the company by minimizing the use of resources.

III. E-COMMERCE

Although E-Commerce first appeared in 1979 in the world, the e-commerce as we know it today, emerged in the mid-1990s with the use of the internet. Amazon and eBay were founded in 1995. We can say that the foundations of modern e-

commerce were laid about 22 years ago. [2]. In its most general terms, e-commerce; it can be defined as "the exchange of goods / services carried out over computer networks in electronic environments designed to receive and place orders" [3]. E-commerce is one of the most popular tools for retail business today.

B2B (Business to Business)

B2B (business to business) allows a business to interact with another business electronically, especially over the internet. The B2B solution has many benefits, some of which are: increasing productivity, reducing potential staff expense, and audit trail. [4].
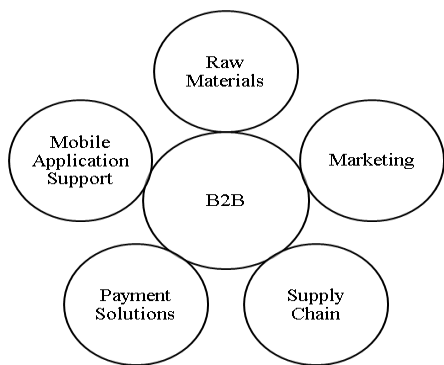

Figure 1. Some Sales Channels in B2B

B2C (Business to Consumer)

Business-to-consumer (B2C) (or business-to-customer) describes the activities of businesses that serve end consumers with products or services. To give an example of the B2C process, a person who buys shoe from a retailer can be a very simple example.

C2C (Consumer to Consumer)

Consumer-to-consumer (C2C) (or citizen-to-citizen) electronic commerce involves transactions facilitated electronically between consumers through certain third parties. Web pages are merely intermediaries and making consumers match each other. They don't need to check the quality of the products that offered.

Figure 2. C2C E-commerce Cycle

Effects of Comments Made on Products in E-Commerce Systems on Sales

One of the important areas of e-commerce systems is consumer ratings and comments.

Consumer reviews help potential customers choose good products and reliable sellers. In addition, it provides benefits to sellers as feedback for improvement.
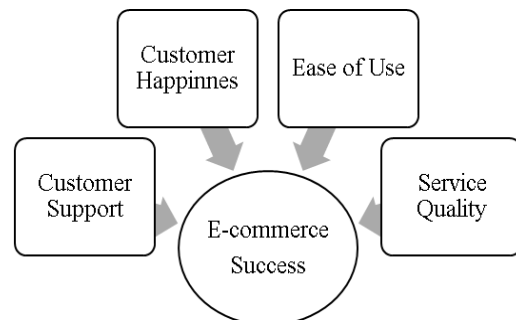

Figure 3. E-commerce Relationships for Success

Consumers often consider previous reviews and comments when purchasing new products. With the widespread use of e-commerce systems, online consumer reviews and ratings made by customers who have previously purchased the product have become an important resource for consumers who want to discover the quality of the product they will buy.

IV. TEXT MINING

Text mining is the discovery of new, previously unknown information by automatically extracting information from different written sources by computers. In text mining, the goal is to discover unknown information, something that no one knows and write yet.
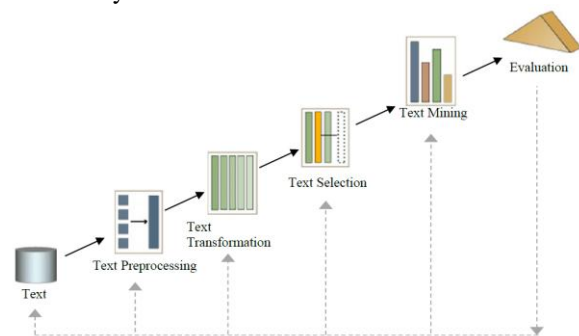

Figure 4. Steps of Text Mining [5]

There are programs that, with reasonable accuracy, can extract information from incomplete organized structure. For example, some programs can achieve 80% accuracy by reading resumes and get people's names, addresses, job skills, etc. It is estimated that approximately 85% of the data in the business world is in text format. [6].

MonkeyLearn is a text analysis software that automates text mining to save hours in manual data processing with the help of machine learning

features. It provides an easy-to-understand and user-friendly graphical interface where users can create custom text classification and sentiment analysis etc.

Application	Feature
-Topic Labeling	Title extraction from text.
-Sentiment Analysis	To find out whether the text is negative, positive or nuetral.
-Intent Detection	Determining the intention of the user within the text.
-Feature Extraction	Extract specific features from text.
-Keyword Extraction	Extract the keywords of the text.
-Entity Extraction	To extract information such as person, company, location from the text.

Figure 5. Some Text Mining Apps Supported by MonkeyLearn

V. SYSTEM IMPLEMENTATION AND CONCLUSIONS

In order to realize the text mining application, first of all, the e-commerce system where products and comments will be found should be set up. In this study, an e-commerce system has been set up using OpenCart e-commerce infrastructure. Then, the data were extracted from the database table containing the comments made to the products added to the system and transferred to MonkeyLearn online text mining application. The topics of the comments were extracted automatically by text mining.

The text mining application used in this study can be accessed on the web page "www.monkeylearn.com". Topic labeling has been used in this application as the type of the text mining.

Complaint	Discount	Fraud	Missing Item
Order Problem	Product Availability	Return or Replace	Shipping Problem

Figure 6. Tags Supported by MonkeyLearn Classifier API

We will classify the texts which are comments about the sample products on our e-commerce website. Certain words used in the text will affect the confidence rate of the topic labeling process. As this confidence rate approaches to 1, we will know the probability of the correct classification is increasing.

```
{
"text": "You should change your shipping company. I had a problem with shipping. It was awfull.",
"external_id": null,
"error": false,
"classification": [
{
"tag_name": "Shipping_Problem",
"tag_id": 60174006,
"confidence": 0.726
```

Figure 7. Sample of Classification Result

If a sample review is made over the "You should change your shipping company. I had a problem with shipping. It was awful" comment made on a product; as a result of the classification, confidence value of 0.726 was found. This ratio shows that the confidence in the accuracy of the classification result is high. For API (Application Programming Interface) integration, first of all, the API ID number must be obtained. For this, you must be a member by logging into the MonkeyLearn website, which we will use for text mining. All applications on the web page can be easily integrated into all text mining areas using the API service.

Model ID: cl_GhPhiVYE
Your API Key: 7ceb789883ca4f114c2149fb6867445b4fa180fe

Tags

Complaint Discount Fraud Missing_Item Order_problem

API Request

To classify make a POST request to the following URL:

https://api.monkeylearn.com/v3/classifiers/cl_GhPhiVYE/classify/

Figure 8. MonkeyLearn API Page

The relevant API link should be used to send the data from the table with comments in the database as POST. In order to perform this process,

Gökay Yılmaz, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 6, (Series-I) June 2021, pp. 01-05

firstly, a connection to the database must be established. In this study, providing the necessary connection, extraction the data from table, sending it to the API as POST and printing the results on the screen as a table were done by using PHP language.

```php
<?php
setlocale(LC_ALL, 'tr_TR');
$servername = "localhost";
$username = "veritabani_kullanici_adi";
$password = "veritabani_kullanici_sifresi";
$dbname = "veritabani_adi";

$conn = new mysqli($servername, $username, $password, $dbname);
if ($conn->connect_error) {
    die("Connection failed: " . $conn->connect_error);
}

$data = array();
$conn->query("SET NAMES UTF8");
$sql = "SELECT * FROM gky_review";
$sonuc = $conn->query($sql);
$sonucstr = "";

if ($sonuc->num_rows > 0) {
    while($row = $sonuc->fetch_assoc()) {
        array_push($data,strip_tags($row["text"]));
    }
} else {
    echo "";
}
$conn->close();
```

Figure 9. Database Connection with PHP and Extraction of Comment Data

After writing the necessary information for the database connection, all the data in the table containing the product comments were taken by using the code "SELECT * from gky_review". These data will be taken into an array variable. Then, with API, a POST process can be sent with the data we have.

```php
//API URL
$url = 'https://api.monkeylearn.com/v3/classifiers/cl_GhPhiVYE/classify/';

$ch = curl_init($url);
$payload = json_encode(array("data" => $data));


curl_setopt($ch, CURLOPT_POSTFIELDS, $payload);


curl_setopt($ch, CURLOPT_HTTPHEADER, array('Authorization:Token 7ceb789883ca4f
curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);

$sonuc = curl_exec($ch);

curl_close($ch);
$sonucArr = json_decode($sonuc);
echo "<table>";
foreach ($sonucArr as $rowsonuc){
    echo "<tr>";
    echo "<td>".$rowsonuc->text."</td>";
    echo "<td>".$rowsonuc->classifications[0]->tag_name."</td>";
    echo "<td>".$rowsonuc->classifications[0]->confidence."</td>";
    echo "</tr>";
}
echo "</table>";
?>
```

Figure 10. Posting Data with API

When the request returns, printing only the API values desired to be displayed makes the data result more understandable. The API that we used provides many data types such as text, internal id, error, tag name, tag id, confidence as a response. In this example, the data we need will be "the text, tag_name and confidence tags" we sent.

Text	Tag	Confidence
What a wonderfull product that is. I just loved it.	Product_Availability	0,39
Product quality is kind of okay. But I had a problem with the transportation. It took one week to get it.	Return_or_Replace	0,548
You should change your shipping company. I had a problem with shipping. It was awfull.	Shipping_Problem	0,726
The package says delivered but I never received it.	Shipping_Problem	1
I did not like the product. I will probably return it.	Return_or_Replace	0,961
I couldn't find this product in stock. You should add it.	Product_Availability	0,999
I'm waiting for months for discount, c'mon guys. Make a discount already.	Discount	0,956
I have a missing item in my shipment. Where should I write for help.	Missing_Item	0,873
I can not create a order for this product. It says "You can't order this right now. Please try again later." I don't know what to do.	Order_problem	0,66

Figure 11. Part of the Text Mining Results

These results have printed on the screen using an HTML table. The resulting data are as in Figure 11.

In this study, the classification of the comments made on the products were extracted by using text mining. Thus, big data can be transformed into more meaningful and smaller data and can be used as a faster method to identify users who have problems. Even the user who has a problem with the product can give a high score despite the problem. For example, the customer can comment as "Although the box delivered two weeks late, the product is really nice." and give 4 or 5 stars as they like the product. However, this does not eliminate the fact that the user is not satisfied with the shipping [7]. Although it normally seems very difficult to examine the subjective comments made in this way, it will be much more effective and faster with text mining applications. Acquiring new customers, delighting and retaining existing customers, and predicting buyer behavior will improve the availability of products and services and hence the profits [8].

There have been 264 million European e-shoppers whereas the number of online retailers was estimated to be 645,000 [9]. Organizations conducting e-commerce can greatly benefit from the insight that data mining of transactional and clickstream data provides [10]. For example, IBM estimated a savings of $2 billion in costs in the year 2000 by offering support information to customers on the Web [11].

The only data needed by this application is the comment itself. Therefore, it can be easily integrated into all systems. The system will not only reduce unnecessary work force, but it will also enable the customer satisfaction studies to be performed faster by easily calculating the frequency of problems experienced by customers.

REFERENCES

[1]. E Demir & S Dinçer, Üretim Sektöründe Veri Madenciliği Uygulamaları: Literatür Taraması, *Anadolu Üniversitesi İşletme Fakültesi Dergisi (İŞFAD), Volume 2(1),* 2020, 1-2.

[2]. N Aydın, Dünya'da E-Ticaret Nereye Gidiyor, *Reforma*, *Volume 1(73),* 2017, 52-60.

[3]. E Yılmaz & S Karakadılar, Türkiye'deki Elektronik Ticaret Uygulamalarına Müşteri Gözünden Bakış ve İyileştirme Önerileri, *İstanbul Gelişim Üniversitesi Sosyal Bilimler Dergisi, Volume 6(1),* 2019, 53-75.

[4]. J Yang & M Papazoglou, Interoperation support for electronic business, *Association for Computing Machiner, Volume 43(6),* 2000, 39-47.

[5]. Y Zohar, Introduction to Text Mining, *Automated Learning Group, National Center for Supercomputing Applications,* 2002.

[6]. A Hotho, A Nurnberger, G Paaß, A Brief Survey of Text Mining, *Ldv Forum, Volume 20(1),* 2002, 19-62.

[7]. K Bhargava, T Gujral & M Chawla, Comment Based Seller Trust Model for E-Commerce, *International Conference on Computational Techniques in Information and Communication Technologies,* 2016, 387-391.

[8]. R. Raghavan, Data mining in e-commerce: A survey, *Sadhana*, 2005, 275-289.

[9]. B Nagelvoort, R Van Welie, P Van den Brink, A Weening, & J Abraham, Europe B2C E-commerce reports, 2015.

[10]. R Kohavi, Mining E-Commerce Data: The Good, the Bad, and the Ugly, *KDD01: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001.

[11]. P Burrows, The Era of Efficiency, *Business Week*, 2001, 92-99.