Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.15 : 2024
ISSN : **1906-9685**

Paper ID: ICRTEM24_113 **ICRTEM-2024 Conference Paper**

AN EXTENSIVE SENTIMENT ANALYSIS AND OVERVIEW OF THE RECOMMENDER SYSTEM

[#1] **BOORLA SANTHOSH,** ***Research Scholar,***
[#2]**Dr. ANUPAM DESHPANDE,** ***Associate Professor & Guide,***
[#3]**Dr. T. SRINIVAS,** ***Professor & Co-Guide,***
***Department of Electronics & Communication Engineering*,**
SHRI JAGDISHPRASAD JHABARMAL TIBREWALA UNIVERSITY, RAJASTHAN.

ABSTRACT-

Many fields use recommender systems, which are valuable in many ways. User ratings are available in many areas, but most classic recommender systems use numerical ratings to express users' opinions on ingested things. To compensate for the lack of evaluations, the recommendation technique uses user-generated reviews, which supply new information. Feedback in the sentiment analysis section might reveal a lot about a product or service. This paper's detailed summary will aid recommender system and sentiment analysis researchers. The stages, strategies, and performance measures of recommender systems are explained. In the next section, sentiment analysis' level, technique, and focus on aspect-based sentiment analysis are explained.

Keywords: recommender system, ratings, user reviews, sentiment analysis, aspects.

1. INTRODUCTION

There has been an explosion of internet data, and that figure will only climb as more people use the web to research and book housing, dining, and other services. Despite the benefits of data, the massive flow of information makes it tough for users to choose from many options. Information overload complicates decision-making (Liu et al. 2011). To help users and customers choose, restrict the data set to their current preferences (Hdioud et al. 2013). Recommended system (RS) algorithms filter information for consumers by adapting service (i.e. product) recommendations to their tastes to reduce information overload.

Since its founding decades ago, RS has been crucial to academia, business, and industry. They are used in commerce, music, movies, and periodicals as well as Amazon, Pandora, Netflix, TripAdvisor, Yelp, Facebook, and TED speeches. E-commerce websites have shown that RSs help users/customers identify relevant items that meet their needs and potentially preferences. Amazon made 35% of its 2015 sales on product recommendations.

The most frequent recommendation approaches are hybrid, collaborative, and content-based. Content-based (CB) suggestions are based on the user's recent actions, such as what they liked, bought, or watched (Pazzani&Billsus 2007). CF makes a proposal by comparing the user to other users who have showed interest in similar subjects (Aciar et al. 2007). Finally, Danilova and

Ponomarev (2016) suggest the hybrid approach, which incorporates various recommendation components or algorithm implementations into one recommendation system. However, classical RS approaches advocate a single criterion rating (the overall rating). A recommendation with only one criterion is inaccurate since aggregate scores do not represent the fine-grained analysis behind users' behaviors. As a result, the RS became a broad research topic, resulting in many studies to improve its functionality.

Social media customer experience sharing is growing rapidly. Many consumers base their service choice on peer opinions. This is partly responsible for the dramatic surge in user reviews and other internet discussion. Each review expresses a customer's opinion of a product, movie, or hotel reservation. Companies and customers benefit from these reviews. Despite their merits, such evaluations are too large and unique to extract useful information from (Chen et al. 2015). Most recommendation systems avoid reviews because machines struggle to interpret natural language (Musat et al. 2013). Opinion mining, text mining, and natural language processing are used to analyze textual reviews and extract meaningful information. Sentiment analysis identifies positive and negative objects.

2. RECOMMENDER SYSTEM

People frequently rely on recommendations from others, whether by word of mouth, book and movie reviews, letters of recommendation, guides to local restaurants and motels, or another way (Resnick & Varian 1997). The recommender system enhances this natural social function. As a result, the following sections provide a thorough explanation of the recommender system, including its characteristics, stages, aims, methodologies, numerous algorithms, and performance metrics.

OVERVIEW

The recommender system (RS) is a crucial component of modern digital infrastructure. As a data filtering method, RS has been effective in addressing the problems of information overload for the last 20 years (Resnick et al. 1994). There are two types of data processing: "filtering in," which involves looking for relevant information, and "filtering out," which involves removing irrelevant data (Resnick et al. 1994). Social media and online shopping are just two of the many platforms that have adopted and made use of RSs since its inception (Lu et al. 2018). Also, RS is a hot issue in academia, with innumerable research both finished and in progress covering the subject. Helping consumers and users locate products that fulfill their requirements and, more importantly, their preferences is one way to address the RS issue (Adomavicius&Tuzhilin 2005). RSs have been described in several ways; one being as a "data mining" tool that provides individualized suggestions to users and aids their search (Hidioud et al., 2013; Kermany&Alizadeh, 2017). In order to help users cope with the deluge of data they encounter online, content filtering software can analyze their preferences and product details to provide tailored product recommendations (D'addio & Manzato 2015; Lakiotaki et al. 2011; Wang et al. 2018). A method for managing information overload that employs data collection, user customization, and preference-based suggestion-making (Chen et al. 2015). There are three main parts to recommendation systems (RS): data about items (such as specs and features), data about people (such as interests, purchasing history, ratings, and reviews), and data about things (such as specs and features). By making use of this data, filtering algorithms narrow the results to only those items that are highly relevant to the user's tastes. A high-level overview of the RS architecture is shown in Figure 1.

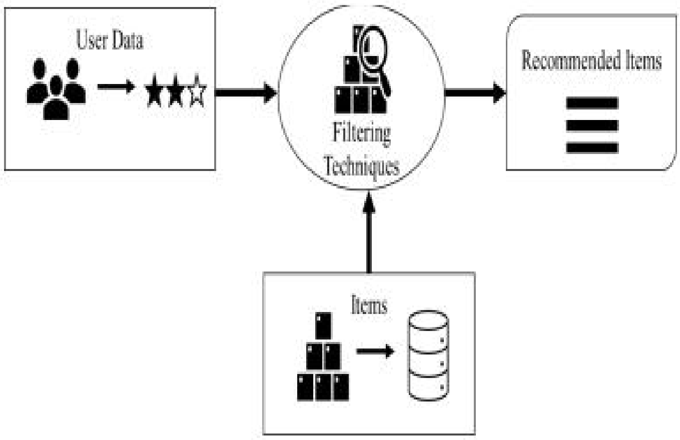

Figure 1 General RS architecture

To be more precise, an RS model consists of two sets (Items and Users) and a utility function. The Users set contains all users, and the Items set contains all things that can be suggested to users. For each given user u and item i in the Users and Items classes, the utility function checks to see if the suggestion is appropriate. According to Adomavicius et al. (2011), this is represented as R: Users×Items→R0, where R0 might be either a positive integer or a real value falling inside a specific range.

Figure 2 depicts the typical three-stage RS process, as reported by Adomavicius et al. (2011), Adomavicius et al. (2005), and Ricci et al. (2011). The procedures are as follows:

Modeling Stage: The major aim of this phase is to prepare the data for the next two stages. A rating matrix, in which users are records and items are attributes, is one of the three potential implementations; each cell in the matrix stores the rating that a user has assigned to an object. Additionally, when you create a profile for a user, you basically end up with a vector that represents their preferences for a product, either overall or based on specific attributes. Thirdly, creating a profile from a comprehensive description of an item.

Prediction Phase: This step estimates a user's rating or score for unseen items using data gathered in the modeling phase and a utility function.

Stage of Suggestion: This section expands upon the prediction section by utilizing multiple approaches to weed out the most suitable items to assist the user in making a choice. It recommends new products to the user based on his past purchases and what he might be interested in. In order of aesthetic appeal, these things are listed below.

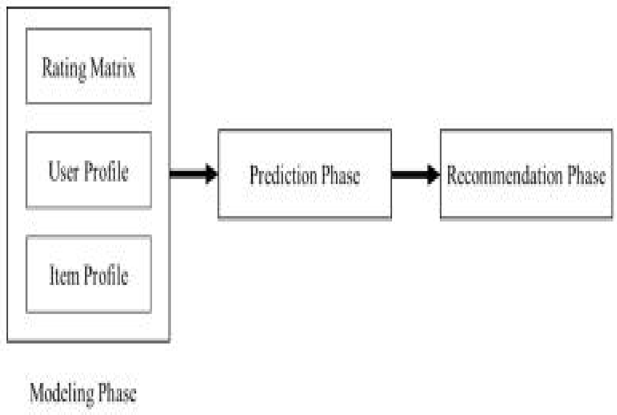

Figure 2 The phases of the recommender system

Anyone who uses an e-commerce website or does business online can profit from RS. Many individuals now use RSs to acquire specialized product recommendations and save money on product selection, thanks to the meteoric development of online shopping (Isinkaye et al. 2015). Figure 3 depicts a product suggestion bubble that occurs when you shop online at Amazon. Typically, customers use these applications to obtain a specific service, such as making a hotel or restaurant reservation, purchasing something, watching a movie, etc. The service in this situation is the purchasing of a face mask. Those who have tried this product (purchased this face mask) will definitely have a variety of opinions and views about it. The red circles represent customer feedback; consumers can assess the product as a whole (on a scale of 1 to 5) or a specific attribute (for example, comfort, thickness, value for money, etc.).

3. RECOMMENDER SYSTEM APPROACHES

Common methods for making suggestions fall into one of three broad classes: content-based, collaborative filtering, or hybrid. Shown in Figure 3 are the three types of RS models.

CONTENT-BASED APPROACH

Based on the user's previous actions, including what they liked, bought, or watched, a content-based (CB) strategy can provide them with appropriate recommendations (Pazzani & Billsus 1997). The approach described by GarcíaCumbreras et al. (2013) builds a profile of the user using their past product choices and then suggests items that are similar to their preferences in terms of quality. It makes each user unique without requiring them to compare their choices to others. This does not take into account the preferences or resemblance of other users, as stated by Aciar et al. (2007) and García-Cumbreras et al. (2013). A simplified explanation of the CB approach is shown in Figure 5. (Blanco-Fernandez et al., 2008; Cantador et al., 2008; Hdioud et al., 2013) The following steps comprise the CB approach:

Product Illustration: In order to construct the structured item representation, the object's attributes (i.e. features) are extracted from the item description, which serves as the information source.

Study the profile of the user: All of a user's ratings, comments, likes, and dislikes go into making their profile.

Make some suggestions: The user's profile and the item's characteristics are used to compile a list of things that the user is likely to like. Products with the highest probability of pleasing the user are added to the list (top-ranked items).

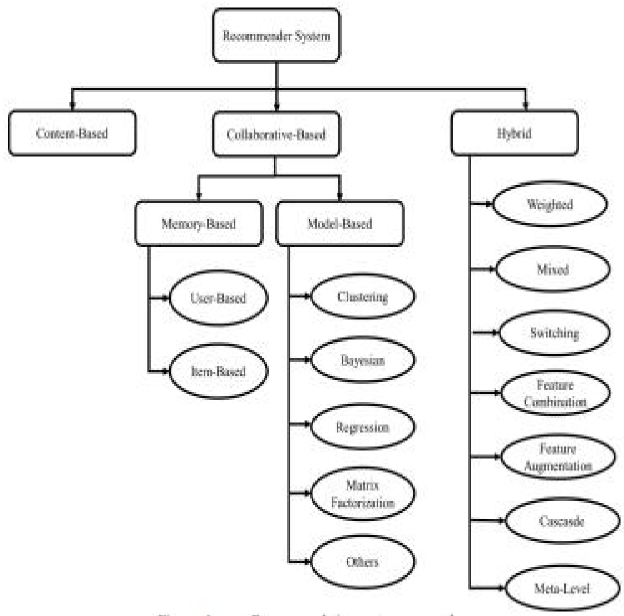

Figure 3 Recommendation system approaches

Web pages, news stories, and articles are just a few examples of the various places this method has found use (Chen et al., 2015). It also suggests things to do like going on vacation, shopping online, and watching television (García-Cumbreras et al. 2013). For things of a moderate size, this method works best.

The CB method has several benefits, including:

By outlining some characteristics of the information, it can demonstrate how to make personalized recommendations (i.e., explain the reasoning behind the suggestions). According to Aciar et al. (2007), this can make the user feel more confident in the RS that suits his preferences.

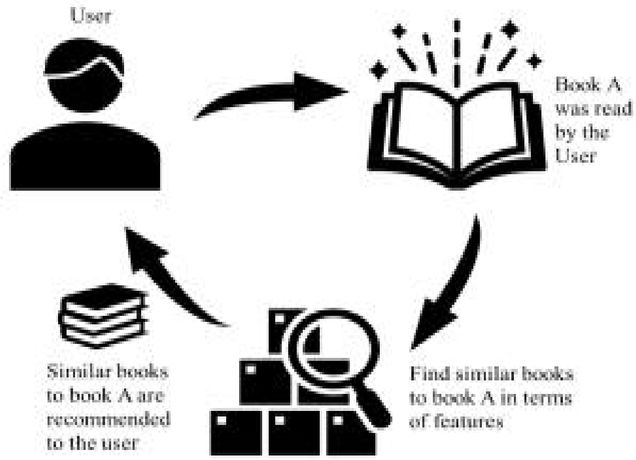

Figure 4 Simple illustration for CB approach

COLLABORATIVE-BASED APPROACH

Yang et al. (2016) states that the collaborative filtering approach (CF) is the most often used RS technique. Based on the user's interactions with others who share their interests and preferences, it generates a recommendation. Users will keep getting support from those who have already agreed with them, which is the main idea behind this strategy (Aciar et al. 2007). Figure 6 provides a simplified description of the CF approach that aids in understanding the previous concept. Yang et al. (2016) states that CF analyzes user-thing interactions to find the new user-item link. It depends on the implicit knowledge of the community's users to find connections between used objects and other users who haven't used or seen them (García-Cumbreras et al. 2013). You may think of this matrix as a user × items matrix, where each cell reflects the user's rating of a certain item.

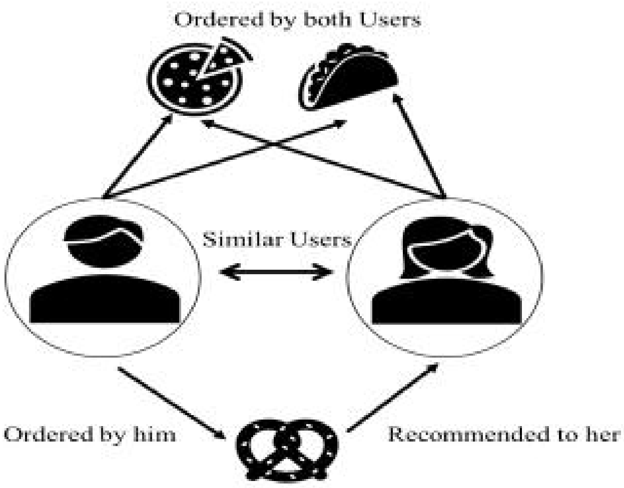

Figure 5 Simple illustration for CF approach

In 1994, Resnick et al. developed GroupLens, the first CF framework for RS. It pushes material to Netnews clients through a rating server called Better Bit Bureaus (BBB), based on the CF hypothesis, which claims that clients who have approved to article ratings in the past are likely to agree to them again.

Chen et al. (2015) states that memory-based and model-based CF are the two primary varieties. The memory-based CF type uses user ratings as a heuristic to predict the item's rating. The first method, which is based on ratings, finds users who have similar interests, or "neighbors," and then proposes products to those users (Adomavicius&Tuzhilin, 2005). Things are given more attention in the second method. In the second approach, products are suggested to other users based on their previous purchases, views, or likes. Figure 5 compares item-by-item CF suggestions with recommendations based on user behavior. 'User 1' and 'User 3' are similar according to the user-based technique, since they both had product preferences. That being the case, 'User 3' ought to partake in the activities that 'User 1' desires. The item-to-item method, on the other hand, compares products based on the amount of likes each did. 'Apple' and 'pear' are similar as 'User 2' and 'User 1' both appreciated them. Since 'pear' is 'user 3's favorite fruit,' we will recommend 'apple' to them..

4. SENTIMENT ANALYSIS

We act in large part based on our opinions, which can take many forms such as mood, feeling, or attitude, and are thus essential to nearly every aspect of human life. According to Liu (2012) and Schouten and Frasincar (2016), individuals' perceptions and assessments of the cosmos greatly influence their own emotions, ideas, and behavior. So, it's human nature to want to hear other people's thoughts and feelings anytime a choice is imminent. Sentiment analysis is centered around opinions. Since the new millennium, this area has been leading the charge in data mining, text mining, and natural language processing research because of the deluge of subjective data that is accessible online (Liu 2012). In light of the fact that it defines sentiment analysis, outlines its levels, duties, methodologies, and even aspect-based sentiment analysis, this section represents a thorough introduction to the subject.

SENTIMENT ANALYSIS APPROACHES

In sentiment analysis, there are primarily two schools of thought: one focused on lexicon-based techniques and the other on machine learning. The objective of these techniques is to classify the user's emotional response to each word in the given text. When it comes to document and sentence level sentiment classification, machine learning approaches are widely used. As Ravi and Ravi (2015) point out, aspect-based sentiment classification relies heavily on the lexicon-based approach. Figure 9 displays the primary distinctions among the SA approaches.

MACHINE LEARNING APPROACH

The primary goal of this method is to use machine learning for SA; selecting appropriate features for sentiment and view detection is critical to the strategy's effectiveness (Liu 2012; Serrano Guerrero et al. 2015). That is why NLP (Natural Language Processing) technologies are crucial to this approach. Teaching computers to detect if a user is feeling good, neutral, or negative is the essence of this approach. Supervised and unsupervised learning are the two main types, as stated by Serrano-Guerrero et al. (2015) and Ravi and Ravi (2015).

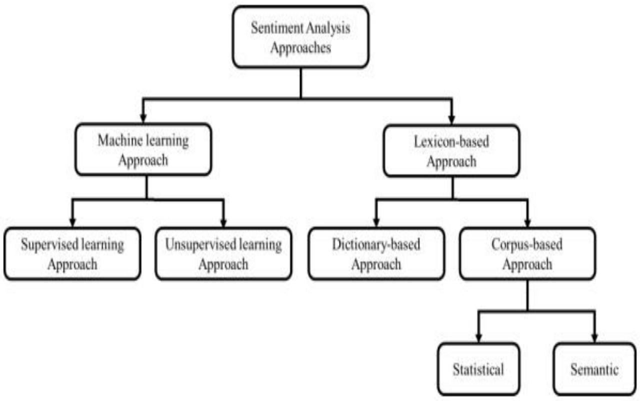

Figure 6 Sentiment analysis approaches

Supervised Learning Approach

Almost all academics view SA as a classification challenge, and sentiment categorization is SA's mainstay, as we mentioned earlier. Sentiment classification is defined as a two-class classification issue by Liu (2012), who notes that most works exclude the neutral class to make classification easier. Similar to sentiment classification, text classification is a natural language processing (NLP) problem that seeks to assign unclassified information to one or more specified categories (themes) (AlGhuribi & Alshomrani 2013). The most crucial aspect of text classification that ensures the information is appropriately classified is the presence of words related to each topic. In contrast, the most crucial component of sentiment classification are opinion words, which are defined as words that convey a positive or negative opinion, such as great and dreadful. Any existing supervised learning methods can be applied to sentiment classification since it is fundamentally a text classification problem (Liu 2012). Ye et al. (2009), Support vector machines, Decision trees, Naive Bayes, Maximum entropy, and Neural Networks are the most popular methods, according to Al-Ghuribi and Alshomrani (2013), Hemmatian and Sohrabi (2019), and Others.

Unsupervised Learning Approach

Sentiment classification often relies on opinion and sentiment terms, making this domain a good fit for the unsupervised learning technique (Liu 2012). The use of annotated or labelled data in supervised learning is labor-intensive and expensive, as pointed out by Hemmatian and Sohrabi (2019). A way to fix this is with unsupervised learning. It is suggested in situations where there is a lack of labelled data due to its excellent accuracy in multiple applications. An unsupervised approach to sentiment classification, clustering algorithms were first proposed by Liu and Zhang (2012). Hemmatian and Sohrabi (2019) state that clustering algorithms can generate reasonably accurate analytical results without requiring any human intervention, training, or language knowledge.

LEXICON-BASED APPROACH

One employs a sentiment word to express an opinion, whether positive or negative. terms that reflect ideas, polar terms, or words that are important for SA analysis are those that bear this name (Liu 2012). "Good," "fabulous," and "beautiful" are words that make people feel good. Use of derogatory adjectives such as "bad," "ugly," and "awful" indicates that the user feels negatively about the reviewed item. A sentiment lexicon is a list of words that describe a certain emotion. According to Serrano-Guerrero et al. (2015), lexicon-based approaches rely on sentiment lexicons, which are basically compilations of recognized and precompiled sentiment phrases. Both dictionary-based and corpus-based approaches can be used to build a vocabulary (Liu 2012).

Dictionary-based Approach

An alphabetical list of words with their definitions and possible synonyms is the conventional layout for a dictionary. So, it follows that the dictionary-based method of sentiment word compilation will be an easy operation (Liu 2012). In the beginning, when building the sentiment words dictionary, a seed list is defined. This is a tiny collection of emotion words that have been carefully collected and annotated. As new antonyms and synonyms of the terms in this collection are added following a search in a certain dictionary, it is expanded. We keep searching until we run out of new words in the dictionary, as stated by Liu (2012), Ravi and Ravi (2015), and Serrano-Guerrero et al. (2015). (Hu & Liu 2004; Strapparava & Valitutti 2004) used the previous simple method for the dictionary approach. The next step is to add more processes to the strategy; for instance, Kim and Hovy (2004) wanted to make sure that all the created words were error-free and that each word in the list had a probability-based strength. One big problem with this method is that it can't deal with domains and orientations that are exclusive to a single scenario, according to Serrano-Guerrero et al. (2015) and Schouten and Frasincar (2016). A popular SA dictionary, SentiWordNet, was developed by Baccianella et al. (2010) using the Word Net (Miller 1995) lexicon as an example.

Corpus-based Approach

New corpus-based approaches have arisen to overcome the shortcomings of traditional dictionary-based approaches by creating dictionaries tailored to certain domains (Serrano-Guerrero et al. 2015). The technique finds similar terms in a huge collection of writings called a corpus by using a seed list of opinion words and grammatical rules or patterns (Medhat et al. 2014). Two basic scenarios have seen the implementation of this method: the first involves the provision of a seed set of opinion words (a general-purpose set), and the second involves the search of a domain corpus for recall sentiment words. As an alternative, you can make use of a domain corpus to tailor a generic sentiment lexicon to a given domain (Liu 2012). Ravi and Ravi (2015) state that this tactic can be applied in two ways: statistically and semantically.

Finding the opinion words that occur together in a corpus is the purpose of the statistical method (Medhat et al. 2014). Read and Carroll (2009) and Serrano-Guerrero et al. (2015) state that calculating the frequency of a word's occurrence in a corpus allows us to determine its polarity or sentiment score. To put it simply, the frequency of use in positive textual content indicates that a word has positive polarity, whereas the frequency of use in negative textual material indicates that it has negative polarity. So, it's reasonable to think that two words with a high frequency of occurrence in the same text will have the same polarity. According to Medhat et al. (2014), one way to find out the polarity of a word is to look at how often it appears with another word. Pointwise mutual information (PMI), which finds the degree of statistical dependence between two words, could help with this. By deducting the strength of the word's association with a set of negative terms from the strength of its association with the word, the PMI finds the word's alignment with a set of positive words. According to Serrano-Guerrero et al. (2015), one statistical method associated with SA is Latent Semantic Analysis, which looks at the relationship between a collection of texts and the words used in them.

The semantic method, which uses word similarity to explicitly give sentiment values to sentiment terms (Ravi & Ravi 2015), is a useful tool for this purpose. According to Medhat et al. (2014), this technique assigns similar sentiment scores to words that have similar meanings. Both the Wordnet and the Sent Wordnet dictionaries provide a variety of semantic relations between words, which can be useful for this strategy. As mentioned by Al-Ghuribi et al. (2020) and Maks and Vossen (2012), among the many potential SA applications of this technology is the creation of a lexicon model describing nouns, verbs, adjectives, and adverbs. By combining this method with the statistical methodology, more precise results can be achieved.

5. CONCLUSION

At present, sentiment analysis is used to transform user-generated evaluations from their unstructured form into a format that RSs can understand and use, hence increasing the RSs's accuracy. This article provides a thorough overview of sentiment analysis and the recommender system. An overview of the recommender system is provided at the beginning, outlining its steps, methodologies, and important performance indicators. Following this, the article delves into the nature and operation of sentiment analysis, covering ground such as levels, techniques, and aspect-based sentiment analysis, among other subjects. Academics' understanding of sentiment analysis and recommendation systems is our goal in administering this survey.

REFERENCES

1. Aciar, S., Zhang, D., Simoff, S. & Debenham, J. 2007. Informed Recommender: Basing Recommendations on Consumer Product Reviews. IEEE Intelligent Systems 22(3):
2. Adamopoulos, P. & Tuzhilin, A. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. ACM Transactions on Intelligent Systems and Technology (TIST) 5(4): 1-32.
3. Adomavicius, G., Manouselis, N. & Kwon, Y. 2011. Multi-Criteria RecommenderSystems. Dlm. Ricci, F., Rokach, L., Shapira, B. & Kantor, P. B. (pnyt.). Recommender Systems Handbook, hlm. 769-803. Boston, MA: Springer US.
4. Adomavicius, G., Sankaranarayanan, R., Sen, S. & Tuzhilin, A. 2005. Incorporating Contextual Information in Recommender Systems Using a Multidimensional Approach. ACM Transactions on Information Systems (TOIS) 23(1): 103-145.
5. Adomavicius, G. & Tuzhilin, A. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE transactions on knowledge and data engineering 17(6): 734-749.
6. Al-Ghuribi, S. M. & Alshomrani, S. 2013. A Simple Study of Webpage Text Classification Algorithms for
7. Arabic and English Languages. 2013 International Conference on IT Convergence and Security (ICITCS), hlm. 1-5.
8. Al-Ghuribi, S. M., Noah, S. A. & Tiun, S. 2020. Various Pre-Processing Strategies for Domain-Based Sentiment Analysis of Unbalanced Large-Scale Reviews. International Conference on Advanced Intelligent Systems and Informatics, hlm. 204-214.
9. Al-Ghuribi, S. M. & Noah, S. a. M. 2019. Multi-Criteria Review-Based Recommender System–the State of the Art. IEEE Access 7(169446-169468.
10. Al-Ghuribi, S. M., Noah, S. a. M. & Tiun, S. 2020. Unsupervised Semantic Approach of Aspect-Based Sentiment Analysis for Large-Scale User Reviews. IEEE Access 8(218592-218613.