

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications



ISSN: 0005-1144 (Print) 1848-3380 (Online) Journal homepage: <https://www.tandfonline.com/loi/taut20>

A real-time social network-based knowledge discovery system for decision making

Asım Sinan Yüksel & Fatma Gülşah Tan

To cite this article: Asım Sinan Yüksel & Fatma Gülşah Tan (2018) A real-time social network-based knowledge discovery system for decision making, *Automatika*, 59:3-4, 261-273, DOI: 10.1080/00051144.2018.1531214

To link to this article: <https://doi.org/10.1080/00051144.2018.1531214>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Oct 2018.



Submit your article to this journal [↗](#)



Article views: 782



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



A real-time social network-based knowledge discovery system for decision making

Asım Sinan Yüksel ^a and Fatma Gülşah Tan ^b

^aComputer Engineering Department, Süleyman Demirel University, Isparta, Turkey; ^bVocational School Computer Programming Department, Celal Bayar University Kirkağaç, Manisa, Turkey

ABSTRACT

The increasing amount of data in social networks has complicated data processing and interpretation. Therefore, intelligent decision-support mechanisms that have the ability to automatically extract meaning from data and interpret the opinions of people in real time have become inevitable. In this study, an intelligent multilingual decision support system was implemented, and a new algorithm that employs text mining and sentiment analysis techniques was developed to automatically interpret the opinions of social network users about the places they plan to visit. The system can be used as a baseline for sentiment analysis in social networks and can be adapted to build new systems. In this study, we set our main focus on Turkish language and show the applicability of our approach for other languages through the experiments for English language. The dataset required for the implementation of text mining techniques was created based on the venue recommendations shared on Foursquare social media platform. As a result, a contribution was made to the way the social network users make decisions without reading thousands of recommendations. Our results show that the developed system achieves classification accuracy of 84.49% for Turkish and 95% for English. Finally, the most liked or disliked foods/beverages are correctly identified for 107 out of 128 venues.

ARTICLE HISTORY

Received 9 March 2018

Accepted 29 September 2018

KEYWORDS

Decision support system; knowledge discovery; social networks; text mining; natural language processing; sentiment analysis

1. Introduction

With the widespread use of social networks and the developments in information technology growing at a fast pace, access to information has become easier, and it has become increasingly important to obtain useful and clear information from such networks. The rapid increase in the amount of data has also caused an increase in the amount of non-structural data. As long as it is not processed, data is stored as a meaningless mass in databases, and it becomes difficult for people to discover knowledge [1,2].

Data mining is a part of knowledge discovery research and collaborates with research areas such as statistics, machine learning and pattern recognition [3]. The difference that separates text mining from data mining emerges here. Specifically, text mining uses uniform texts that are obtained through natural language processing rather than databases with a known or well-formed dataset. The language of the text and the meaning that it carries may vary according to purpose. Text mining methods alone do not produce results that understand the content. Therefore, expressions involving complex structures should be classified in a meaningful way using natural language processing techniques. The inferred view may be the mood, the decision, or the thought that the user wants to express about a subject.

Natural language processing research includes artificial intelligence studies, often integrated with linguistic knowledge [4]. Based on these studies, the aim can be to analyze and summarize research articles, automatically participate in a conversation or answer questions, and carry out translations between languages [5]. Besides general information about the language, there is also a need for a domain or task-specific knowledge that should be perceived independent of the general structure of a language [6].

With the rise of the Internet, social networks have become the most data-rich platforms. Thus, it has become necessary to carry out data and text mining studies on these networks [7]. One of the most popular social network platforms is Foursquare. It is an environment where people can share their complaints and pleasures about the venues they visited [8]. In this study, recommendations about the venues on the Foursquare platform were used as the data source. A web-based, multilingual, real-time knowledge discovery and decision support platform that identifies positive, negative and neutral user thoughts about foods, beverages and venues was developed by using natural language processing and text mining techniques.

In cases where there are no decision support systems, evaluating people's feelings and thoughts can take a very long time. For this reason, there is a need for

mechanisms that will help the user make decisions very quickly and almost instantaneously. It is an inconvenient process for a social media user to decide whether or not to go to a venue by reading hundreds or even thousands of recommendations. The development of an intelligent decision support system will eliminate such problems, enabling faster and more efficient analysis of sentiments and the automatic identification of opinions. In our system, as the first step, the words in users' recommendations are organized, and misspelling is corrected by applying natural language processing pipeline. The data is organized through the pipeline that is composed of diacritic restorer, vowelizer, spelling corrector and normalizer components. In the restoration phase, ASCII characters are converted to proper Turkish representations. However, the restored text may contain shortened words. Therefore, as the next phase, vowelizer restores the omitted vowels. Finally, misspelled words are corrected by spelling corrector and normalization process is applied by using state of the art text normalizer for Turkish. The organized data are then classified as positive, negative, or neutral according to feelings and thoughts by employing text mining techniques. As the final step, the system identifies whether a venue near the user's location was liked; which food and/or beverage in a venue was most liked by analyzing the data. Thus, the user is given the opportunity to make relatively quick and informed choice based on the developed system instead of reading all the recommendations from nearby venues.

The rest of this paper is organized as follows. In Section 2, we review the recent work. In Section 3, we present our methodology regarding the developed decision support system and our Real-Time Social Knowledge Discovery Algorithm (RT-SKDA). In Section 4, we share our results. In Section 5, we conclude with recommendations for future work.

2. Related work

Most of the recent work in this field has been developed for major languages such as English. Morphologically rich languages are addressed sparsely [9]. The one-to-one translation of the rules and algorithms used in the English language is not quite possible in the Turkish language or other languages due to the formal differences. Therefore, the studies to be done in other languages should be carried out with collaborative efforts of native linguists or native or foreign engineers who are very knowledgeable about the language they are working on. Natural language processing (NLP) does not include fixed algorithms, and it is necessary to understand a given problem, find the solution, and develop the most appropriate tools. When developing our tool, we mainly focused on Turkish language and researched into the recent studies conducted for this language. However, to show the applicability of our approach in

other languages, we also researched into the studies involving English language and conducted experiments for this language.

In this context, Delibas et al. [10] focused on the formal analysis of Turkish by using natural language processing techniques. The aim was to find the spelling mistakes in the Turkish text and to sort and correct these mistakes. Their results showed that a higher success rate was achieved than in previous studies.

İlhan et al. [11] developed a question-and-answer system and used natural language processing and text mining techniques. In this way, the questions that were obtained were subjected to the preliminary process of data mining, and keywords were determined. The weights of these keywords in the text were established, and an appropriate answer was generated according to the usage rate in the text.

Pang et al. [12] suggested a method for the subjective areas of a document using text classification techniques. This technique was adapted for sentiment analysis, and a sentiment classification was carried out. The most effective result was obtained using the Support Vector Machine (SVM) classification algorithm.

Vural et al. [13] presented a framework for sentiment analysis of Turkish text documents. Researchers customized SentiStrength sentiment analysis library by translating its lexicon to Turkish. Movie reviews were used to classify the polarity. For performance evaluation, they used a large corpus of Turkish movie reviews obtained from a popular Turkish social media site. The results showed that they achieved a fairly good classification accuracy that approached the performance of supervised classification methods.

Go et al. [14] used 8,00,000 positive and 8,00,000 negative Twitter messages as a dataset. They performed sentiment analysis using distant supervision method. They employed Naive Bayes, SVM, and Maximum Entropy methods of machine learning algorithms for this purpose. They combined and analyzed the data using bigrams and unigrams and achieved 83% accuracy through the Maximum Entropy method.

Trieu et al. [15] proposed a novel method for classifying news from social media. They proposed to vectorize articles with TD2V and applied the Doc2Vec approach for Twitter-based universal document representation. Furthermore, they applied retrieval and automatic query expansion to obtain the most relevant labelled documents. The results showed that they achieved better classification accuracy than existing methods on different datasets.

Agarwal et al. [16] used the Twitter social media platform as a data source and tagged the obtained data by hand. They performed a binary (positive, negative) and ternary (positive, negative, neutral) classification. They applied Tree Kernel, Senti-Feature, and Unigram models on the data. According to their experiments, the Senti-Feature and Unigram models gave the best

result in the binary classification, and Tree Kernel and Senti-Feature in the ternary classification.

Saif et al. [17] used Twitter comments as a data source. They categorized 60,000 Twitter comments that they collected by classifying them in a binary (positive, negative) manner. They also performed classification studies with Naive Bayes (NB) and N-grams, and the best result was obtained in the binary classification with a success rate of 86.3%.

Habernal et al. [18] used Facebook social media platform to collect 10,000 set of comments and grouped them as positive, negative, and neutral. They tried to classify the comments as binary (positive, negative) and ternary (positive, negative, neutral) by using traditional classification algorithms. They achieved a 90% success rate in the binary classification and a 69% success rate in the ternary classification.

Katz et al. [19] used TripAdvisor as a dataset that consists of 30,000 comments, including 24,930 positive and 5070 negative, and used a total of 2000 comments from IMDB, which included 1000 comments with positive polarity and 1000 comments with negative polarity. They created ConSent, a content-based sentiment analysis model. They extracted the features of keywords by identifying them in the text and classified documents according to these features in a second stage. Authors stated that the method they developed had almost the same success rate as the SVM method.

Yıldırım et al. [20] stated that sentiment analysis studies generally focused on English texts. They carried out natural language processing studies on social media texts for the Turkish language, which has a very rich morphological structure. They manually tagged 12,790 comments collected via Twitter (3541 positive, 4249 negative and 5000 neutral). They had 78.83% success on Turkish texts in their sentiment analysis process that was conducted through ternary classification.

Kilic et al. [21] used academic articles as a dataset and measured the success of classifying the articles by employing text mining techniques. The dataset created for feature selection was composed of abstracts of 2000 articles from 50 different journals. They categorized the dataset into Science and Engineering, and Social Sciences and Humanities. The Confusion Matrix was used to evaluate the success of the classification algorithms. Using the K-nearest neighbours (k-NN) algorithm, the category to which the other publications belonged was detected with an accuracy of 96.67%. This ratio was 91.32% and 86.53%, respectively in Naive Bayes and J48 algorithms.

Kürkçü et al. [22] used 81,174 comments from the Twitter accounts of news agencies as a dataset. In order to determine the interaction levels of users, they carried out an analysis in terms of the number of tweets, retweets, and likes. In their analysis, authors analyzed the interaction levels of the news tweets within 48 hours, how often the tweets that were sent at certain

times of the day were retweeted and liked by users, and the ratios of the total number of retweets.

3. Method

Sentiment analysis and natural language processing studies have been intensively developed for the English language. Due to the differences in the structures of languages, it is harder to apply these rules in other languages. Therefore, the most efficient method to be employed when building sentiment analysis systems is the employment of tools with language-specific features. Although building resources for each language is costly, it is the most accurate way to be followed [23].

Within the scope of this study, we aimed at developing a flexible, language independent model. A special attention is paid to Turkish and applicability of our model to other languages (mainly English) was showed through experiments. Furthermore, the developed method was compared with other classification algorithms adapted for Turkish.

3.1. Dataset collection

Foursquare is a social media tool that allows mobile users to read and offer recommendations on venues that are near them. The required dataset was obtained from recommendations about venues that were listed on the Foursquare social media platform. A total of 7086 Turkish and 6943 English recommendations were collected from different venues. Figure 1 shows our data collection process.

The recommendations were filtered by using “query = food” parameter when consuming Foursquare’s Application Programming Interface (API). A web application¹ was developed within the scope of this study. Figure 2 shows the home page of our application where the sample venues and available operations are presented. Sample recommendations related to a venue is shown in Figure 3.

3.2. Feature selection

Feature selection is the most commonly used method for data classification such as data mining and machine learning. Before the classification process, the meaningless words need to be identified and eliminated. The feature selection and pre-processing stages that are used in the field of text mining are applied to remove unnecessary words and convert them to the appropriate form [24,25].

Feature selection is employed to determine the most valuable and deterministic variables from the data obtained from Foursquare. The features that are specified in the data are actually the words of that text. Before selecting a feature, and during the cleanup stage of pre-processing, frequently used words, punctuation

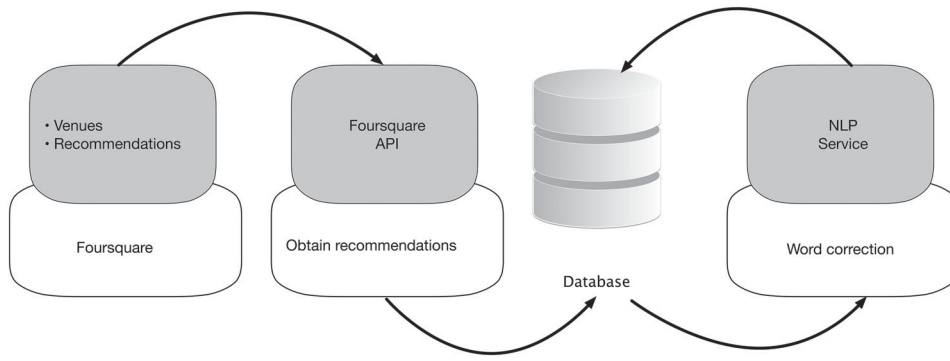


Figure 1. Data collection process.

Venue ID	Venue Name ↕	Operations
524468e811d2f84bb7a87c4b	A Plus Spor Center	Edit Analyze 🗑️
51b5b44c498e525ee6b5926c	Akççek Kozmetik	Edit Analyze 🗑️
52d95908498e89abf8e39859	Ali usta dondurmaları	Edit Analyze 🗑️
513de977e4b0f8b10c6bb0be	Altınış Kuyumculuk	Edit Analyze 🗑️
51250518e4b0855623fc233e	Anadolu İletişim	Edit Analyze 🗑️
50a53474498eef72a2e5db95	Arena Tekel Bayi	Edit Analyze 🗑️
51ae2cf7498e4599d30ff611	atlantik fast food	Edit Analyze 🗑️
51c499ab498e8dcd9651b097	Atlantis Pet Shop	Edit Analyze 🗑️
5141d4fbe4b02b62911e6347	Avea Zirve İletişim	Edit Analyze 🗑️
54f0ae42498e31ed2e3fa9f9	Ayvalık Tost	Edit Analyze 🗑️

1-10/88 10 ▾ 1 2 3 4 5 6 7 8 9 >

Figure 2. Sample venues and available operations.

Comment	Number Of People Who Agree	Number Of People With Different Opinions	Operations
Love the hazelnut and goat cheese.	2	0	🗑️
home made pasta is delicious	3	0	🗑️
you have to try donuts	0	0	🗑️
honey and goatcheese outstanding	5	0	🗑️

1-20/29 1 2 > >>

Figure 3. Sample recommendations related to a venue.

marks, adverbs, and verbs are searched in the dataset. By doing so, the words that do not contribute to the sentiment analysis to make interpretations are eliminated. When the unnecessary data are eliminated, it is easier to detect the words that change the result of interpretation. At the same time, the collected data are reduced

to the level that the computer can handle. The higher the consistency rate in the feature selection method, the higher the success rate will be.

In our study, the dataset was manually processed, unnecessary words were removed, and the misspelled words were corrected. Each recommendation was

Table 1. Sample binary classification.

Recommendation	Positive/Negative
Sausage with grilled onions & sweet peppers is my favorite!	1
The cornbread isn't too great.	0
Must try the fried chicken!	1
Lasagna isn't great.	0

Table 2. Sample ternary classification.

Recommendation	Positive/Negative/Neutral
Meatball is neither good nor bad	2
Sandwich is awesome! :)	1
Cheese is not tasty	0

separated into individual sentences according to the meaning of the sentence. The result of this action is stored as a separate dataset and is used for the purpose of establishing the ground truth when comparing the success of our approach. Binary (positive, negative) classifications were carried out for the English sentences, and binary (positive, negative) and ternary (positive, negative, neutral) classifications for the Turkish sentences. Furthermore, the names of specific food items and beverages were tagged in the sentences. Table 1 shows a sample binary classification and Table 2 shows the ternary classification. During the tagging process, 0 was used for negative sentences, 1 for positive sentences, and 2 for neutral sentences.

3.3. Text mining

Text mining is the process of obtaining structured data from the data source. Operations such as subject extraction, sentiment analysis, text classification and summarization can be performed [26]. In this research, we applied the text mining process in five steps as follows:

- Data collection and selection
- Removing unnecessary expressions from the data
- Feature extraction
- Data structuring

In the first step, the data were obtained in real time from the venue. The most significant problem in text mining is that the data are not structural. Therefore, the data cannot be processed directly by the computer because of missing and incorrect information, broken content, informal or slang words and advertisements. In the second step, the data were pre-processed and cleaned in order to make a meaningful interpretation. In the third step, feature selection and feature extraction techniques were applied to each record of collected data. With this process, unnecessary expressions were removed from the collected data, and the data were tagged. For the fourth step, text mining techniques were carried out simultaneously with natural language processing studies. For this purpose, Istanbul Technical University Turkish Natural Language Processing (ITU NLP)

Web Service [27] was used to detect misspelled words in the data that was made usable, and structural data were obtained.

3.4. Natural language processing

Language has great importance for humanity because it enables people to communicate with each other. New words are frequently added to or subtracted from languages, but the structure and rules of languages have remained unchanged for a long time [5].

Natural language processing (NLP) focuses on computer-assisted language processing. The main task of natural language processing is to design and implement computer systems that will analyze and interpret a natural language [28]. People who are working in the field of natural language processing endeavour to process language using the results linguists have produced based on their studies [5]. Application areas where natural language processing is used are as follows:

- Grammar analysis
- Database management
- Speech recognition
- Translation
- Text summarization
- Document management

In this stage, our goal was to find spelling errors using natural language processing techniques and correct these errors with the highest success rate. As shown in Figure 4, the following procedures were carried out:

- Checking for spelling errors in texts
- Suggesting correct words for misspelled words
- Correcting the spelling errors that were found

Many problems were encountered when these texts were processed such as:

- Irregular and misspelled words
- Text slicing
- Semantical ambiguities
- Slang words

Furthermore, we observed that the recommendations collected from social media often had grammar or punctuation errors. Use of abbreviations, the presence of local speech, use of words with different meanings, and incomplete sentences complicated the natural language processing step. In order to correct a spelling error, the misspelled word is first identified, and then the word that is most similar in spelling to this word is selected. This new word is checked to determine whether there are missing letters or whether the letters are in the wrong places. Misspelled words are edited according to the rules of the words.

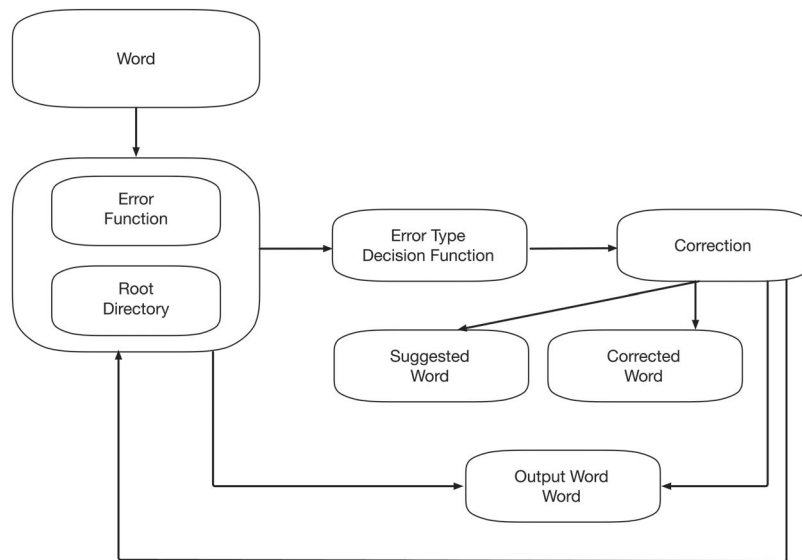


Figure 4. Word correction process.

Figure 5. Input sentence and edited sentence.

In this context, we utilized the ITU Turkish NLP Web Service which contains native language processing tools and APIs for Turkish. Figure 5 shows the result of an edited sentence in the sentence editing menu of our application that consumes ITU Turkish NLP Web Service API.

3.5. Sentiment analysis

Sentiment analysis is a field of study that evaluates people's feelings and opinions and analyzes their opinions about a subject through a written language [29]. With the advancement in information technology, the increased use of the Internet has led to an increase in the number of users in social media environments. People are now expressing their feelings and thoughts on social networks. Sentiment analysis is used to classify these thoughts according to their positive and negative status [30]. Due to the increased data flow worldwide, sentiment analysis has become one of the most active research fields of natural language processing and text mining.

Sentiment analysis in a text or the sentimentality of that text is generally represented by binary classifications. For example:

- Positive/Negative
- Like/Dislike
- Good/Bad

Furthermore, sentiment analysis is applied in many applications with different features [12]:

- Question and answer systems
- Suggestion systems
- Summarization and citation analysis
- Applications in different fields such as politics and sociology
- Practices in business and government intelligence

In the literature, two main approaches are highlighted for sentiment analysis. The first one is based on machine learning algorithms. The second approach utilizes dictionary of words to identify the sentiment

Table 3. Sample adjectives in the dictionary.

Word in Turkish	English Translation	Positive/Negative
muhteşem	awesome	1
lezzetli	delicious	1
lezzetsiz	flavorless	0
güzel	yummy	1
iğrenç	disgusting	0

Table 4. Sentiment analysis using Google APIs.

Turkish Sentence	English Translation	Sentiment Score
Kahve çok lezzetliydi	The coffee was very tasty	0.9
Kahve çok lezzetliydi	Coffee was touched	0
Kahve çok lezzetliydi	Very delicious	N/A
Kahve çok lezzetliydi	Coffee is very tasty	0.9
Kahve çok lezzetliydi	Coffee is very lezzetlydy	0

polarity [31]. Machine learning methods require labelled training set to build the classifier and the performance of the classifier is improved by using language dependent sentiment lexicons.

In this research, in order to perform multilingual sentiment analysis, we applied dictionary-based approach and took advantage of domain specific knowledge. Furthermore, we applied commonly employed machine learning algorithm, Naïve Bayes, for comparing the results with our approach. For sentiment analysis, it was first necessary to determine the words that contain emotions. For this purpose, we created an extensive dictionary for adjectives for the target language, tagged the adjectives as positive and negative

and assigned a score for each adjective. If the words contained in the sentences were included in our dictionary, the emotions of the sentence were determined by adding the scores of the words and taking their average. Table 3 lists some of the adjectives in our dictionary used in sentiment analysis.

To see the effect of machine translation on sentiment analysis, the unstructured data were translated from Turkish to English using the Google Translate API [32]. For sentiment analysis process, Google Cloud Natural Language API [33], which does not support Turkish, was used. Table 4 shows the result of a sample sentiment analysis using these services. Different forms (containing typos or local speech) of the Turkish sentence “Kahve çok lezzetliydi”/ “The coffee was very tasty” were created and tested.

As concluded from the table above, misspelled words and local speech directly affect the result of sentiment analysis. Although all sentences had the same meaning, the API failed to assign the same sentiment score to each one. This shows that natural language analysis is a necessary step. Furthermore, the API does not check whether the sentence is related to a food or beverage, meaning that domain-specific knowledge is ignored. In our system, we not only attempted to find out whether the sentence was positive or negative but to also extract domain specific knowledge. In order to do this, an extensive list of foods and beverages was created in addition to the list of adjectives. Creating a

Edit Sentence

Sentence to Edit

Sentence to Edit

kahvesii çok güzel. Köfte iğrenç. Manti berbat

Edit

Edited Sentence

Edited Sentence

Kahvesi çok güzel

Köfte iğrenç

Manti berbat

Scoring and Sentiment Analysis

Meal-Adjective-Score

Meal & Drink Name	Adjective	Positive-Negative
Kahve-	güzel-	Pozitif-
Köfte-	iğrenç-	Negatif-
Manti-	berbat-	Negatif-

Score Order

Meal & Drink Name	Score
Kahve	1
Köfte	-1
Manti	-1

Sentiment Analysis Results

Most Liked Meals-Drinks

Kahve

About Venue

Mekan beğenilmemiş

Figure 6. Sentence editing and sentiment analysis.

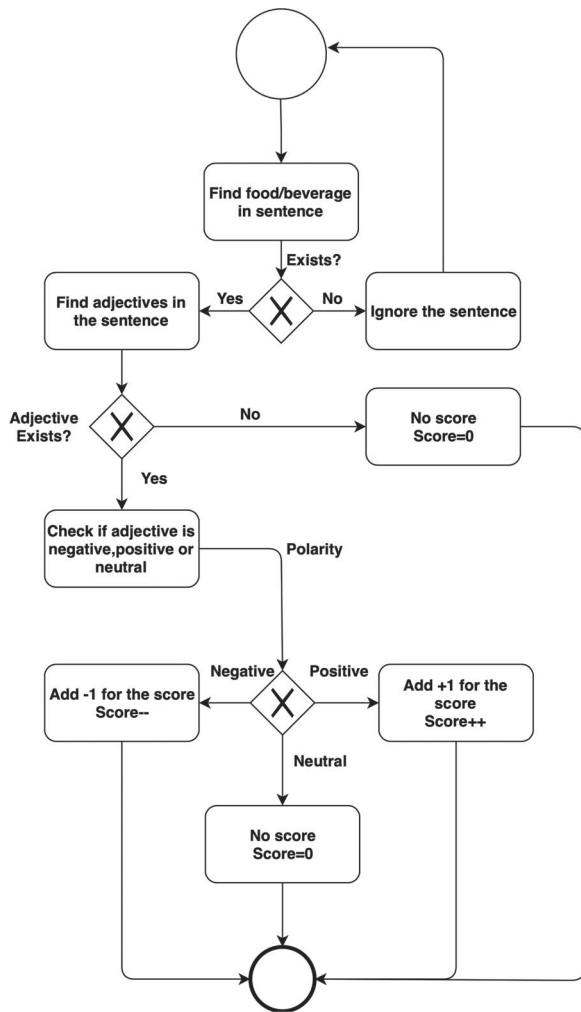


Figure 7. Sentiment analysis and scoring flow.

food and beverage list is mandatory because of cultural and cuisine differences, no direct translation of traditional foods, and nonexistence of these foods in other cultures.

If there was a word from the list of food and beverage items in the sentence, such sentences were identified before performing sentiment analysis step. In this way, unnecessary recommendations about the venue were eliminated. Sentiment analyses were conducted on the filtered sentences to find out which food or beverage was mentioned and whether it was liked or not. Figure 6 shows the results of sentence editing and sentiment analysis processes for a given recommendation in our system. The recommendation consists of three separate sentences. Each sentence contains misspelled words and food/beverage information.

The scoring was done in a way that positive word received +1, negative words received -1, and neutral ones received 0. A classification was performed according to the names of the food and beverage items. The food or beverage that was mentioned the most and whether it was liked or not was determined. As seen in the above figure, all foods/beverages were correctly identified, and all misspelled words were successfully

corrected. At the end of the page, the most liked food or beverage was shown. Additionally, the sentiment about the venue was also presented. The data used during sentiment analysis was listed under the menu named Adjectives. The list of food and beverage items used during the interpretation stage was listed on the Meals & Drinks menu. The open and collaborative nature of the system allows anyone to add adjectives or food and beverage items. The larger the number of items in the data list, the better the accuracy of classification and sentiment analysis studies. The flow of the sentiment analysis and scoring stages are shown in Figure 7.

3.6. Sentiment API generation

Our Sentiment API Generation Module makes it easier to build the sentiment analysis process as a service by employing RT-SKDA and turning user specifications into code for different languages. Figure 8 shows the flow of the API generation module.

The flow of the system is as follows:

- (1) User enters the system, clicks “Generate API” menu and selects the language that he wants to build the sentiment analysis service for.
- (2) User adds related dictionaries in Dictionary Creation Module or uses pre-built dictionaries. The dictionaries can be updated; new words can be added.
- (3) User plugs his own language specific NLP service or any other service developed by typing the web service URL. The service must return the results in string format.
- (4) Web Service Generation Module uses the information from previous stages and passes them to RT-SKDA as parameters. Finally, RT-SKDA is built as a web service and a URL is generated for further use.
- (5) User consumes the generated web service and obtains the sentiment analysis results.

Developing an API for sentiment analysis purpose can be difficult task when there are many internal and external dependencies for text mining, natural language processing, social network crawling processes. Therefore, our system saves time for social network-based sentiment analysis and enables faster service development and better collaboration with researchers who are working in this field.

4. Discussion and results

The Success Rate menu shows the results of the RT-SKDA in addition to Naive Bayes (WEKA) and manual observation. This is shown in Figure 9.

In order to test the real-time performance of the developed system, experiments were carried out on

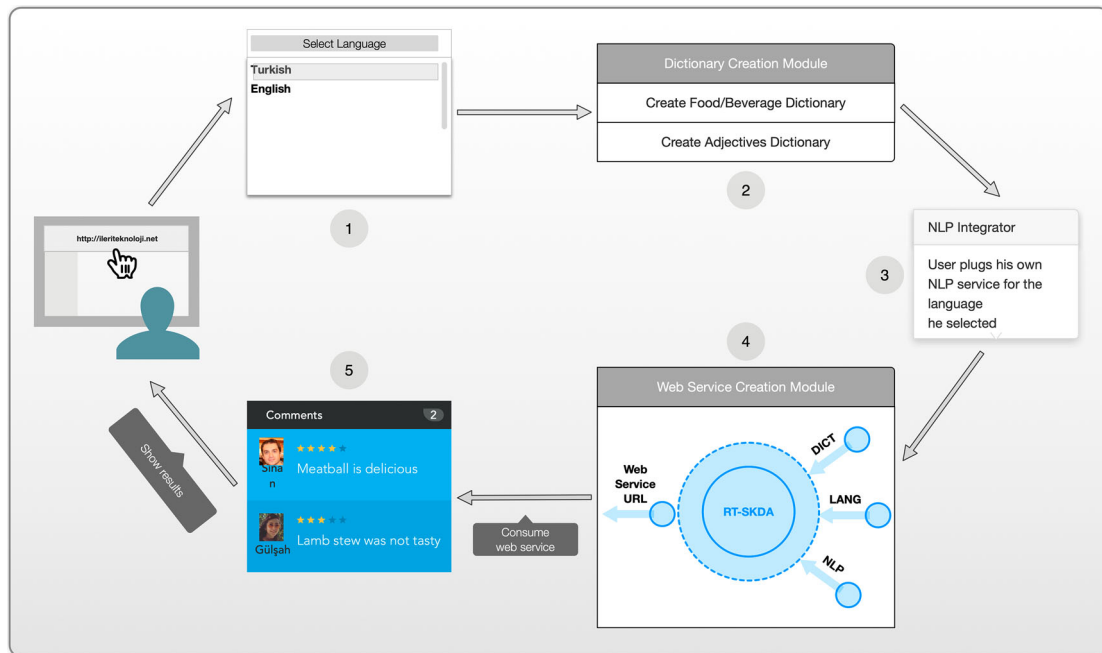


Figure 8. Sentiment API generation flow.

Search Criteria				
Venue Name				
<input type="text"/>				
Venue Id	Venue Name ▲	RT-SKDA	Naive Bayes	Manual Observation
56d44377498e02f9072b1898	Mariento	kahve	çay	kahve
4fc0e680e4b0d6554a32c0a4	Martı restaurant	meze	meze	meze
56f408a7498e3933d7c255b9	Mekan Kafe	çay	çay	çay
4d91e2a1f5388cfaafeb3d	Melbo	ekler	tavuk	salata
4cd5ce95b6962c0f24a52e96	Meşhur Köfteci Doktor	köfte	köfte	köfte
4bf40f9e94af2d7f361f3a72	Meşhur Mahmutbey Kanatçısı	kanat	kanat	kanat
535ac9d8498e0f645e2fde78	midye diyari aziz halil usta	midye	midye	midye
559c7938498ee81bc5c584c5	Mini Lila Anne Usulü Gözleme Manti Kahvaltı Evi	ekler	ekler	çay
57fb768d498e1f78e0975840	MİSKİN	sahlep	sahlep	nargile
4c46ea4976d72d7f7a713b4d	MR Tantuni Döner	tantuni	tantuni	tantuni

Figure 9. Success rate menu.

real-time data. For this purpose, the Nearby Venues menu was added as shown in Figure 10.

Tests can be conducted according to the location coordinates selected from the Nearby Venues menu. When any one of the listed venues is selected, the recommendations are gathered from Foursquare, RT-SKDA is applied to the recommendations, and the results are displayed on the screen.

In addition to the sentiment analysis, a food and beverage dataset was created to find out the most liked item of the venue. Ultimately, it is possible to create datasets for many different subjects or areas, where RT-SKDA can be applied and sentiment analyses can be carried out.

In order to test the accuracy of the classification, binary (positive, negative) and ternary (positive,

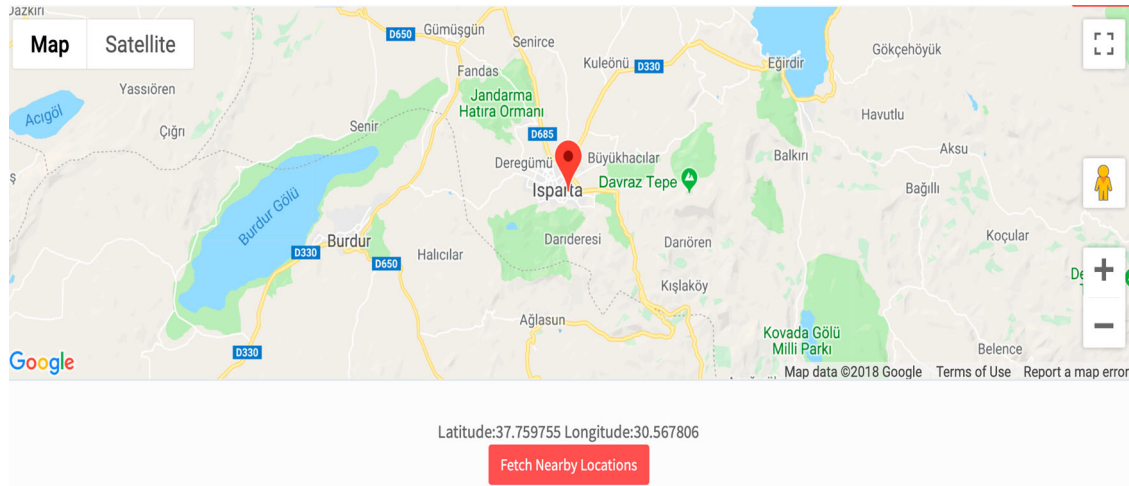


Figure 10. Nearby venues menu.

negative, neutral) classification experiments were conducted. In addition, the Naive Bayes algorithm was applied. Binary and ternary classification results for Turkish are shown in Tables 5 and 6. We applied the Confusion Matrix method to find the sentiment classification accuracy and the results are presented in Table 7.

A high accuracy was obtained when comparing RT-SKDA with other traditional machine learning methods that have been implemented for the Turkish language. Table 8 shows the comparison of our method with similar studies applied for the Turkish language.

Table 5. Binary classification results for Turkish.

Method	Positive	Negative
Manual Observation	6028	1058
RT-SKDA	5204	1882
Naive Bayes	4519	2567

Table 6. Ternary classification results for Turkish.

Method	Positive	Negative	Neutral
Manual Observation	5594	996	526
RT-SKDA	5203	851	1031
Google APIs	3308	1592	2186
Naive Bayes	4680	760	1646

Table 7. Classification accuracy for Turkish.

Method	Binary Classification (%)	Ternary Classification (%)
RT-SKDA	81,97	84,49
Naive Bayes	73,80	78
Google APIs	–	59,38

Table 8. Comparison with similar studies for Turkish.

Study	Data Source	Method	Accuracy (%)
[20]	Twitter comments	Support Vector Machine	78.05
[34]	E-commerce websites	k-Nearest Neighbours	83.87
[35]	Twitter comments	Naive Bayes	62.04
[36]	Movie reviews	Support Vector Machine	75
[37]	Twitter comments	Support Vector Machine	73.63
[38]	Twitter comments	Sequential Minimal Optimization	65.5
[39]	Twitter comments	Maximum Entropy	76.78

Furthermore, to show the applicability of our approach in other languages, we conducted our experiments for English. Google APIs and Naive Bayes approach are applied. Binary and ternary classification results for English are shown in Tables 9 and 10. Table 11 shows our classification accuracy.

Finally, we conducted a separate experiment to see the effect of up votes and down votes on classification performance. As recommendations were obtained from Foursquare, the number of people who agreed and who did not agree with these recommendations were also obtained. These numbers were added to the score based on whether the sentence was positive or not. We observed that votes did not have an effect on the classification accuracy. Additionally, these up/down votes were not available in every recommendation.

Along with the classification, we aimed to find which food or beverage was liked the most in the venue. Accordingly, the food and beverage items that were liked were correctly identified for 107 out of 128 venues. This number dropped to 91 when the NB method was

Table 9. Binary classification results for English.

Method	Positive	Negative
Manual Observation	6715	228
RT-SKDA	6577	366
Naive Bayes	6445	498

Table 10. Ternary classification results for English.

Method	Positive	Negative	Neutral
Manual Observation	4945	228	1770
RT-SKDA	5140	366	1437
Google APIs	4098	756	2089
Naive Bayes	5340	498	1105

Table 11. Classification accuracy for English.

Method	Binary Classification (%)	Ternary Classification (%)
RT-SKDA	95	89
Naive Bayes	78	75
Google APIs	–	71

used. We note that increasing the number of venues and recommendations will increase the accuracy.

5. Conclusions and future work

Social network-based intelligence is an indispensable technique for venue owners to measure customer satisfaction. In this way, whether the venue is liked or not is found as well as which food and beverage items are liked most. The effect created on social media can be closely monitored, and correct steps can be taken when necessary. We anticipate that our tool will be an important analytical tool for venue owners to determine how their venue is perceived on social media, to intervene quickly to correct deficiencies at times of crisis, to determine the success rates of products, and to better position their businesses against competitors.

We can think of the result of this study – being carried out on social media – as being directly proportional to customer satisfaction. Recommendations made about a venue do not remain private between the user and the venue owner but instead spread across the entire social network. Such data become impossible to follow by the user after a while. Therefore, both users and venue owners need decision support systems by which they can quickly, automatically, and accurately analyze the recommendations made on social platforms. Decision support systems are crucial for accelerating the decision-making process, achieving superiority in a competitive environment, rapidly identifying data to be used for performance enhancement, and creating reports that can be used to increase productivity. In this direction, a multilingual decision support system was developed, which is flexible, easily adaptable, able to improve itself, easy to use, responds quickly, and eliminates the need for the user to read thousands of recommendations one by one. With the developed decision support system, recommendations about a venue are grouped into three categories – positive, negative, and neutral. At the same time, the food or beverage that was liked the most is found and shown to the user, along with whether the venue was liked or not. RT-SKDA can be further developed to work domain independently, support many languages so that it provides successful results for different domains and languages. Finally, the developed system can be transformed into an application that will work on mobile devices.

Note

1. Developed web application: <http://ileriteknoloji.net>.

Acknowledgments

This work was supported by Scientific Research Project Coordination Unit of Süleyman Demirel University. Project Number: 4853-YL1-17. The dataset created for this study can

be accessed from: <https://data.mendeley.com/datasets/29tbvwwkdp/2>



Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Scientific Research Project Coordination Unit of Süleyman Demirel University [Project Number: 4853-YL1-17].

ORCID

Asım Sinan Yüksel  <http://orcid.org/0000-0003-1986-5269>
Fatma Gülşah Tan  <http://orcid.org/0000-0002-2748-0396>

References

- [1] García S, Luengo J, Herrera F. Data preprocessing in data mining [Internet]. Cham: Springer International Publishing; 2015 [cited 2018 Feb 13]. (Intelligent Systems Reference Library; vol. 72). Available from: <http://link.springer.com/10.1007/978-3-319-10247-4>.
- [2] Birant D, Yıldırım P. A framework for data mining and knowledge discovery in cloud computing. Data Science and Big Data Computing [Internet]. Cham: Springer International Publishing; 2016 [cited 2018 Feb 13]. p. 245–267. Available from: http://link.springer.com/10.1007/978-3-319-31861-5_11.
- [3] Witten IH, Ian H, Frank E, et al. Data mining: practical machine learning tools and techniques [Internet]. [cited 2018 Feb 13].
- [4] Shaalan K, Hassanien AE, Tolba F. editors. Intelligent natural language processing: trends and applications [Internet]. Cham: Springer International Publishing; 2018 [cited 2018 Feb 13]. (Studies in Computational Intelligence; vol. 740). Available from: <http://link.springer.com/10.1007/978-3-319-67056-0>.
- [5] Adalı E. Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgi Bilim ve Mühendisliği Derg [Internet]. 2012 Jun 1;5(2) [cited 2018 Feb 13]. Available from: <http://dergipark.gov.tr/tbbmd/issue/22245/238797>.
- [6] Ferrari A, Dell'Orletta F, Esuli A, et al. Natural language requirements processing: a 4D vision. IEEE Softw [Internet]. 2017 Nov;34(6):28–35 [cited 2018 Feb 13]. Available from: <http://ieeexplore.ieee.org/document/8106888/>.
- [7] Irfan R, King CK, Grages D, et al. A survey on text mining in social networks. Knowl Eng Rev [Internet]. 2015 Mar 25;30(02):157–170 [cited 2018 Feb 13]. Available from: http://www.journals.cambridge.org/abstract_S0269888914000277.
- [8] Hollander JB, Graves E, Renski H, et al. A (Short) history of social media sentiment analysis. Urban Social Listening [Internet]. London: Palgrave Macmillan UK; 2016 [cited 2018 Feb 13]. p. 15–25. Available from: http://link.springer.com/10.1057/978-1-137-59491-4_2.
- [9] Kincl T, Novák M, Přibil J, et al. Language-independent sentiment analysis with surrounding context extension. Cham: Springer; 2015 [cited 2018 Jul 24]. p. 158–168. Available from: http://link.springer.com/10.1007/978-3-319-20367-6_17.
- [10] Delibas A. Doğal Dil İşleme İle Türkçe Yazım Hatalarının Denetlenmesi [Thesis (MSc)]. İstanbul Tech Univ Inst Sci Technol [Internet]; 2008 Jul 7 [cited 2018

- Feb 13]. Available from: <https://polen.itu.edu.tr/handle/11527/321>.
- [11] İlhan S, Duru N, Karagöz Ş, et al. Metin Madenciliği ile Soru Cevaplama Sistemi. [cited 2018 Feb 13]. Available from: http://www.emo.org.tr/ekler/a782e81b5bf2698_ek.pdf.
- [12] Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends Inf Retr* [Internet]. 2008;2(1-2):1-135 [cited 2018 Feb 13]. Available from: <http://www.nowpublishers.com/article/Details/INR-011>.
- [13] Vural AG, Cambazoglu BB, Senkul P, et al. A framework for sentiment analysis in Turkish: application to polarity detection of movie reviews in Turkish. *Computer and Information Sciences III* [Internet]. London: Springer London; 2013 [cited 2018 Feb 13]. p. 437-445. Available from: http://link.springer.com/10.1007/978-1-4471-4594-3_45.
- [14] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. *Processing* [Internet]; 2009 [cited 2018 Feb 13];150(12):1-6. Available from: <https://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>.
- [15] Trieu LQ, Tran HQ, Tran M-T. News classification from social media using Twitter-based Doc2Vec model and automatic query expansion. *Proceedings of the Eighth International Symposium on Information and Communication Technology - SoICT 2017* [Internet]; 2017. New York, USA: ACM Press [cited 2018 Feb 13]. p. 460-467. Available from: <http://dl.acm.org/citation.cfm?doid=3155133.3155206>.
- [16] Agarwal A, Xie B, Vovsha I, et al. Sentiment analysis of Twitter data. *Proc Work Lang Soc Media* [Internet]; 2011 [cited 2018 Feb 13]. p. 30-38. Available from: <https://dl.acm.org/citation.cfm?id=2021114>.
- [17] Saif H, He Y, Alani H. Alleviating data sparsity for twitter sentiment analysis. *CEUR Workshop Proceedings* [Internet]. CEUR Workshop Proceedings (CEUR-WS.org); 2012 [cited 2018 Feb 13]. p. 2-9. Available from: <http://oro.open.ac.uk/38501/>.
- [18] Habernal I, Ptáček T, Steinberger J. Supervised sentiment analysis in Czech social media. *Inf Process Manag* [Internet]. 2014 Sep;50(5):693-707 [cited 2018 Feb 13]; Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0306457314000399>.
- [19] Katz G, Ofek N, Shapira B. Consent: context-based sentiment analysis. *Knowledge-Based Syst* [Internet]. 2015 Aug;84(C):162-178 [cited 2018 Feb 13]. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0950705115001471>.
- [20] Yıldırım E, Çetin FS, Eryiğit G, et al. The impact of NLP on Turkish sentiment analysis. *Türkiye Bilişim Vakfı Bilgi Bilim ve Mühendisliği Derg* [Internet]. 2015 Apr 16;7(1) [cited 2018 Feb 13]. Available from: <http://dergipark.gov.tr/download/article-file/395136>.
- [21] Kilic E, Ates N, Karakaya A, et al. Two new feature extraction methods for text classification: TESDF and SADF. 2015 23rd Signal Processing and Communications Applications Conference (SIU) [Internet]. IEEE; 2015 [cited 2018 Feb 13]. p. 475-478. Available from: <http://ieeexplore.ieee.org/document/7129862/>.
- [22] Kürkçü DD. Haber Tweetleri Üzerinden Kullanıcı Etkileşiminin Analizi. *Erciyes İletişim Derg* [Internet]. 2017;5(1) [cited 2018 Feb 13]. Available from: <http://dergipark.gov.tr/download/article-file/275854>.
- [23] Korayem M, Aljadda K, Crandall D. Sentiment/subjectivity analysis survey for languages other than English. *Soc Netw Anal Min* [Internet]. 2016 Dec 9;6(1):75 [cited 2018 Jul 24]. Available from: <http://link.springer.com/10.1007/s13278-016-0381-6>.
- [24] Aggarwal CC, Zhai C. An introduction to text mining. *Mining Text Data* [Internet]. Boston, MA: Springer US; 2012 [cited 2018 Feb 14]. p. 1-10. Available from: http://link.springer.com/10.1007/978-1-4614-3223-4_1.
- [25] Nenkova A, McKeown K. A survey of text summarization techniques. *Mining Text Data* [Internet]. Boston, MA: Springer US; 2012 [cited 2018 Feb 14]. p. 43-76. Available from: http://link.springer.com/10.1007/978-1-4614-3223-4_3.
- [26] Weiss SM, Indurkha N, Zhang T. *Fundamentals of predictive text mining* [Internet]. London: Springer London; 2015 [cited 2018 Feb 14]. (Texts in Computer Science). Available from: <http://link.springer.com/10.1007/978-1-4471-6750-1>.
- [27] Eryiğit G. ITU Turkish NLP web service. *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics* [Internet]. Stroudsburg (PA, USA): Association for Computational Linguistics; 2014 [cited 2018 Feb 14]. p. 1-4. Available from: <http://aclweb.org/anthology/E14-2001>.
- [28] Chowdhury GG. *Natural language processing*. *Annu Rev Inf Sci Technol* [Internet]. 2005 Jan 31;37(1):51-89 [cited 2018 Feb 14]. Available from: <http://doi.wiley.com/10.1002/aris.1440370103>.
- [29] Feldman R. Techniques and applications for sentiment analysis. *Commun ACM* [Internet]. 2013 Apr 1;56(4):82 [cited 2018 Feb 14]. Available from: <http://dl.acm.org/citation.cfm?doid=2436256.2436274>.
- [30] Feldman R, Sanger J. *The text mining handbook: advanced approaches in analyzing unstructured data* [Internet]. Cambridge University Press; 2007 [cited 2018 Feb 14]. p. 410. Available from: <http://www.cambriidge.org/gb/academic/subjects/computer-science/knowledge-management-databases-and-data-mining/text-mining-handbook-advanced-approaches-analyzing-unstructured-data?format=HB&isbn=9780521836579#skV7P5o6W9Mvmza1.97>.
- [31] Balahur A. Sentiment analysis in social media texts. 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis [Internet]; 2013 [cited 2018 Jul 27]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.310.4764>.
- [32] Google Inc. Google cloud translation API documentation | Translation API | Google cloud platform [Internet]; 2018 [cited 2018 Feb 15]. Available from: <https://cloud.google.com/translate/docs/>.
- [33] Google Inc. Cloud natural language API | Google cloud platform [Internet]; 2018 [cited 2018 Feb 15]. Available from: <https://cloud.google.com/natural-language/>.
- [34] Kaşıkçı T, Gökçen H. Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi. *Bilişim Teknoloji Derg* [Internet]; 2013 Oct 30;7(1) [cited 2018 Jul 16]. Available from: <http://dergipark.gov.tr/gazibtd/issue/6630/88021>.
- [35] Coban O, Ozyer B, Ozyer GT. Sentiment analysis for Turkish Twitter feeds. 2015 23rd Signal Processing and Communications Applications Conference (SIU) [Internet]. IEEE; 2015 [cited 2018 Jul 16]. p. 2388-2391. Available from: <http://ieeexplore.ieee.org/document/7130362/>.
- [36] Kaynar O, Yıldız M, Görmez Y, et al. Sentiment analysis with machine learning techniques. *International Artificial Intelligence and Data Processing Symposium*; 2016.

- [37] Eliaçık AB, Erdogan N. Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi. Ulusal Yazılım Mühendisliği Sempozyumu; 2015.
- [38] Cetin M, Amasyali MF. Supervised and traditional term weighting methods for sentiment analysis. 2013 21st Signal Processing and Communications Applications Conference (SIU) [Internet]. IEEE; 2013 [cited 2018 Jul 16]. p. 1–4. Available from: <http://ieeexplore.ieee.org/document/6531173/>.
- [39] Kaya M, Fidan G, Toroslu IH. Sentiment analysis of turkish political news. Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01 [Internet]. IEEE Computer Society; 2012 [cited 2018 Jul 16]. p. 174–180. Available from: <https://dl.acm.org/citation.cfm?id=2457679>.