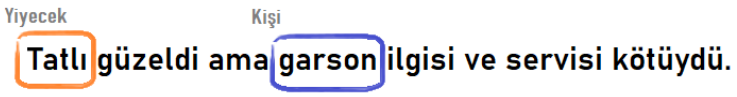
ASPECT – BASED SENTIMENT ANALYSIS FOR TURKISH

1. INTRODUCTION

Sentiment Analysis is a text analysis technique using Natural Language Processing and aims to determine the class that the given text wants to express emotionally. The first studies of sentiment analysis are referred to as emotional polarity and aim to classify the given text as positive, negative and neutral. In the literature, all these studies work under the field of information retrieval. Thanks to the widespread use of social media and e-commerce, individuals are no longer passive users who consume the content offered to them, but they have become active individuals who can produce and share content themselves. In this way, analyzes will provide positive changes in areas such as trade, politics and social events, and the predictions that can be derived from the strategic plans to be made in these areas by using the data to be obtained will be role-changing. People often care about ideas expressed by others because they represent value for themselves as individuals, or rather for organizations and institutions. Analyzing large volumes of data with natural human skills is either impossible or highly inefficient to arrive at a valuable conclusion for these purposes.

Sentiment Analysis; The document is divided into three categories: sentence-based and goal-based. Document-level and sentence-level sentiment analyzes cannot provide sufficient information that is important for decision-making. This type of analysis is a very effective way of examining people's opinions, but it is not enough. Such information can be obtained with Goal-Based Sentiment Analysis. For example, if a restaurant customer gives feedback about the company, the customer usually comments on some aspect of that restaurant. This does not mean that the reviewer completely liked or disliked that restaurant. While the reviewer's overall opinion of the restaurant may be positive or negative, this person can often write both positive and negative about different aspects of the restaurant. From the comment "Dessert was good, but the waiter's attention and service was bad", positive conclusions can be drawn about the desserts, but negative conclusions about the waiter. This is where Goal Based Sentiment Analysis differs from standard sentiment analysis. The main task in Goal-Based Sentiment Analysis is to extract and summarize the views people express about entities and their aspects. As seen in Figure 1, from the example sentence, first of all, the expressions "Dessert" and "waiter" are determined as an aspect of the restaurant, that is, the objectives to be found as a result of the analysis, and "Food" and "Person" as their categories. Then, the goal-based sentiment analysis is concluded by performing a standard sentiment analysis for each of the objectives.



Şekil 1. Örnek cümle.

Named Entity Recognition (NER) is a Natural Language Processing technique frequently used in Goal Based Sentiment Analysis. NER is an information extraction method that aims to find entities mentioned in unstructured text and classify them into predefined categories such as persons. Names, organizations, places, time expressions, amounts, monetary values etc. categories are examples. In the example in Figure 1, the "Food" and "Person" categories were determined by this method.

In this project, it is aimed to obtain high performance analysis results with modern Natural Language Processing approaches where previous solutions are insufficient. According to studies, current approaches clearly support other common languages other than Turkish, especially English. In this respect, Turkish is a low-resource language due to research deficiencies. Although the success rates are low, there are some methods for Turkish. However, these solutions mostly rely on Machine Learning algorithms and standard Sentiment Analysis. In order to obtain high-performance analysis results, modern Natural Language Processing techniques based on Deep Learning methods, which offer very high performance in various tasks from image processing to natural language processing, as well as large and complex language models such as BERT, DistilBERT, ELECTRA will be used.

In the proof-of-concept (POC) studies conducted by us so far, we have achieved a success rate (F-Score) of 70% in the NER model and 75% in the Sentiment Analysis model. The main goal of the project is to increase these success rates and provide more useful results to product and service suppliers.

2. AIMS and GOALS

The aim of the Target-Based Sentiment Analysis project in Turkish is to analyze the emotions and ideas of the target audience in various sectors with scientific methods according to different target-categories and to make new business plans with the data obtained. In order to enable these business plans to be made, it will be developed in a web application format due to its easy accessibility, and an interface will be designed that the user can easily understand and interpret by using various visualizations of the analysis results within this web application. Achievement targets of existing models;

- At least 5,000 comments will be collected from different areas (restaurant, hotel, etc.),
- Collected comments will be tagged target-based (5,000 tagged comments),
- 80% F-Score in NER model,
- 85% F-Score in Sentiment Analysis model.

3. TECHNIQUES

Techniques are listed below. Step by step each technique will be used for this project to achieve aims and goals.

- Data Collection
- Data Cleaning and Tagging
- Training the Deep Learning Models
- Development of Web Application
- Connection of Model and Web Application

- **PROJECT MEMBERS : Furkan ÖCALAN and Mustafa Tayyip BAYRAM**
- **PROJECT ADVISOR : Assist. Prof. Dr. Volkan TUNALI**