T.C. Maltepe University

Faculty of Engineering and Natural Sciences

Software Engineering Department

ASPECT – BASED SENTIMENT ANALYSIS FOR TURKISH

SE 401 Project Report

170706003 Furkan ÖCALAN

170706007 Mustafa Tayyip BAYRAM

Project Advisor: Assist. Prof. Dr. Volkan TUNALI

Table of Contents

List of Figures

List of Abbreviations

I. Introduction

Aspect-Based Sentiment Analysis (ABSA) is a text analysis technique that categorizes data by aspect and identifies the sentiment attributed to each one. ABSA can be used to analyze customer feedback by associating specific sentiments with different aspects of a product or service.

I. Purpose and Importance of the Project

The goal of the ABSA for Turkish is to use scientific methods to assess the emotions and thoughts of the target audience in various sectors and to create new business plans based on the information gathered. Because of its ease of accessibility, this application will be constructed in a web application style, with an interface that the user can readily comprehend and interpret by using various visualizations of the analysis results within this web application.

The first studies of Sentiment Analysis (SA) 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 [1]. Because of the extensive usage of social media and online shopping, individuals are no longer passive users who consume the content offered to them. They have become active individuals who can produce and share content themselves. In this way, the 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 amounts of data with natural human skills is either impossible or highly inefficient to arrive at a valuable conclusion for these purposes.

II. The Innovation (originality) of the Project

Project aims to obtain high performance analysis results with modern Natural Language Processing (NLP) 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. There are some methods for Turkish. Although the success rates are low to making business plans from the output. Because these solutions mostly rely on Machine Learning algorithms and standard SA. In order to obtain high-performance analysis results, modern NLP techniques based on Deep Learning (DL) 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. Our purpose is to increase the success rate and provide more useful outputs to customers.

III. Technology Areas Related to the Project

The project should be classified as a sub-branch of some fields. The Natural Language Processing field is shown here. It is required to become familiar with the following areas in order to comprehend the general application.

    a.   Artificial Intelligence (AI),

    b.   Deep Learning (DL),

    c.   Neural Networks (NN),

    d.   Natural Language Processing (NLP),

        i.   Named Entity Recognition

    e.   Web Development,

    f.   Database Management Systems (DBMS)

a) *Artificial Intelligence (AI):*

It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable [2].

There are numerous, real-world applications of AI systems today. Some of the most common examples: speech recognition, customer service, computer vision, recommendation engines and automated stock trading.
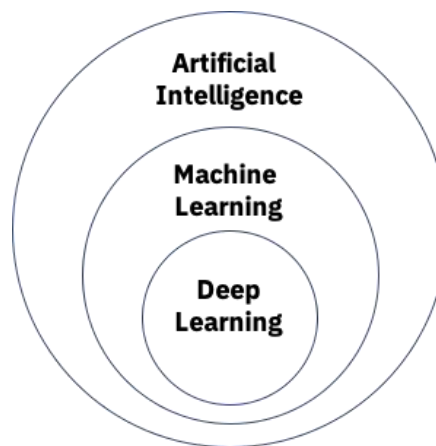
Figure 1. AI Schema

b) *Deep Learning (DL):*

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain albeit far from matching its ability allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning algorithms are incredibly complex, and there are different types of neural networks to address specific problems or datasets [3]. For example,

1.2 Convolutional neural networks (CNNs),

1.3 Recurrent neural networks (RNNs)

Finally, deep learning has applications in healthcare, customer service, financial services, law   enforcement, and many other fields.

c) *Neural Networks*

Neural networks, also known as Artificial Neural Networks (ANNs) or Simulated Neural Networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms [4]. The main task of these neural networks is to relay information. Structure of basic neural network model showed in Figure 5.
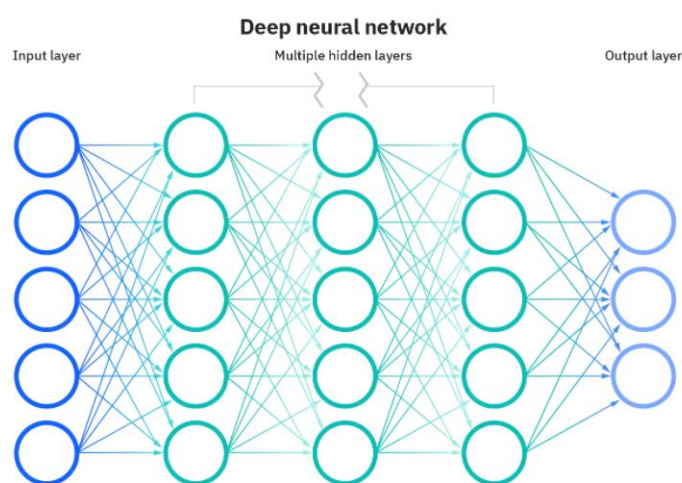
Figure 2. Neural Network

d) *Natural Language Processing (NLP):*

NLP is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

Challenges in NLP frequently involve speech recognition, natural language understanding, and natural language generation [5].

Named Entity Recognition (NER):

NER (also known as named entity identification, entity chunking and entity extraction) is a subtask of information extraction that seeks to locate and classify named entities mentioned in

unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc. [6].

There are some NER platforms that could be used such as

- GATE: General Architecture for Text Engineering is a tool used for all sorts of NLP tasks, including information extraction in many languages.

- OpenNLP: Apache OpenNLP library is a Machine Learning (ML) tool for the natural language text. It supports the almost all the NLP tasks, such as language detection, chunking, parsing, etc.

- SpaCy: It's an open-source library that used in advanced NLP tasks such as information extraction, pre-processing text for DL.

e) Web Development:

Web development, also known as website development, refers to the tasks associated with creating, building, and maintaining websites and web applications that run online on a browser [7].

f) Database Management System:

A database management system (DBMS) is a software package designed to define, manipulate, retrieve, and manage data in a database. A DBMS generally manipulates the data itself, the data format, field names, record structure and file structure. It also defines rules to validate and manipulate this data.

Database management systems are set up on specific data handling concepts, as the practice of administrating a database evolves. The earliest databases only handled individual single pieces of specially formatted data. Today's more evolved systems can handle different kinds of less formatted data and tie them together in more elaborate ways [8].

IV. Detailed Description of the Project

Project name is called 'Aspect-Based Sentiment Analysis for Turkish '. To understand the meaning and goal of the ABSA, one must become familiar with NLP and SA.

Nowadays, NLP is a field of computer science that should be considered and has a great potential of benefit for people. In more detail, it's a field of Artificial Intelligence (AI) that focuses on teaching computers to interpret texts and spoken language in the same way that people do. NLP unites computational linguistics rule-based modeling of human language with statistical, machine learning, and Deep Learning (DL) models. When these technologies are used together, computers can process human language in the form of text or audio data and understand its full meaning, along with the intent and emotion of the speaker or author.

SA is a text analysis technique based on NLP and aims to determine the class that the given text wants to express emotionally. The first studies of SA are referred to as sentimental polarity which aims to classify given text as positive, negative, and neutral. In the literature, all these studies work under the information retrieval [2]. Through to become widely used of social media and e-commerce, humans started becoming individuals who can produce and share content. Thus, analyzes will be able to provide positive change in the field of such as trade, politics, social events. In addition, estimations of strategic

plans which can be made by using these analyzes will provide positive impact in peoples live. Humans usually cares comments which made by other people because they might represent a valuable thing. It's impossible or extremely inefficient to analyze large amounts of data with human skills to achieve at a valuable conclusion for these purposes.

SA is divided into three levels: document, sentence, and phrase/aspect level. Document-level and sentence-level sentiments cannot provide sufficient information that is important for decision making. This type of analysis is very effective way to examine people's opinion but not enough. Such information can be obtained by ABSA. To give an example, when the customer gives feedback about restaurant, one usually makes a comment on some aspects of the restaurant. This comment doesn't mean that customer completely liked or disliked. While the customer's overall opinion of the restaurant may be positive or negative, this person can indicate both positive and negative idea about different aspects of the restaurant in one review. From the comment "Dessert was good, but the waiter's attention and service were bad", positive conclusions can be drawn about the desserts, but negative conclusions about the waiter. This is where ABSA makes difference from SA. The main purpose of ABSA is to extract and summarize reviews which made by people about some areas. As seen from Figure 1, from the example sentence, firstly, "Dessert" and "waiter" expressions are determined as the targets to be found as a result of the analysis, and "Food" and "Person" are their categories. Then, by performing a SA for each of the targets, ABSA is concluded.

Figure 3. Example Sentence

Customer feedback plays an important role in making new improvements in the product which might result in the enhancement of its sales. Customer reviews can be considered as a wealth of insights that let companies know what they're doing right and wrong. Sentiment analysis is an NLP technique that helps classify the reviews provided by the customers and states whether it is positive, negative, or neutral. This gives a clear idea about the customer satisfaction ratio concerning the three class. But companies, more than ratio, are interested in knowing the reason behind negative or neutral reviews. Such reviews can help them to understand the loopholes in their product or service and in turn help them to solve them.

SA is not capable to give such insights to the service providers and hence cannot be used to extract the cause of customer dissatisfaction [9]. The main problems that exist in the current techniques are inability to perform well in different domains, inadequate accuracy and performance in sentiment analysis based on insufficient labeled data, incapability to deal with complex sentences that require more than sentiment words and simple analyzing.

The application will be developed in NodeJS to enable the creation of a web-based application. Model works will be conducted with Python application and will be connected to a ExpressJS server. Data flow will be provided through services. MySQL will be used as a DBMS with necessary ORM's (SQLAlchemy – Python) (NodeJS - Sequalize) to facilitate interaction.

In various business areas, companies will be the target customers. The public end users will not be targeted as a customer.

V.   Detailed Problem Requirements

There are two types of requirements: functional and non-functional.

Functional Requirements:

These are the requirements that the end user expresses as essential features that the system should provide. As part of the contract, all these functionalities must be included into the system.

Non-Functional Requirements:

According to the project contract, these are the quality restrictions that the system must meet. The importance of these aspects, as well as the amount to which they are implemented, vary by project.

Functional Requirements
1.   The user must be authorized when they log in.
1.1 User must be able to sign up with specifying their companies if user is not registered.
a.   User shall be able to renew user's password if user forgets the password.
2.   User must be uploading the reviews file in required formats for new analyze.
2.1 The system gives an error message if the file does not match the format.
2.2 User shall be able to select current industries of the company for new analyze.
3.   The system must display and save analysis results.
1.   The system shall be able to tag outcome as positive, negative, or neutral.
2.   The system must be visualized with tree architecture in accordance with the analysis results. From the general analysis result, the specific analysis result will be accessible through visualizations.
4.   User must be able to examine previous analyze results.
4.1 User can export previous analyzes.
4.2 User shall be able to import previous analyzes in the analyze result format of the system.

5.	User can edit company's and user's information.
Non-Functional Requirements	
1.	Named Entity Recognition success rate (F score) must be greater than 80%.
2.	Sentiment Analysis success rate (F score) must be greater than 85%.
3.	Web systems loading time must be less than 5 seconds.
4.	Web system must have responsive design.

Software Quality Attributes:

- Availability: Because this application will be critical to business planning, we will have a goal of four nines (99.50%) availability.

- Correctness: The application should never allow anyone to read messages or discussions not intended for that person.

- Maintainability: The application should use continuous integration so that features and bug fixes can be deployed quickly without downtime.

- Usability: The interface should be easy to learn without a tutorial and allow users to accomplish their goals without errors.

II. High Level Design

High-level design (HLD) explains the architecture that would be used to develop a system. It can be classified in two heading, namely system design and database design. Although database design is a subset of system design, it is a topic worth delving into more.

➔ System Design

System design is classified under two main headings, namely web application architecture and platform requirements. Each heading explained below.

1. *Web Application Architecture*

The project is a web application, and it will be client/server architecture. The term client/server refers to a logical separation of duties in which the client is the application (also known as the front-end) and the server is the database management system (also known as the back end). The functionalities will be available when a user is entered in. Web application system architecture:

- Front-end software: HTML, CSS, JavaScript, PugJS
- Back-end software: NodeJS (ExpressJS), Python
- Database Management System: MySQL
    - ORM's: SQLAlchemy, Sequalize
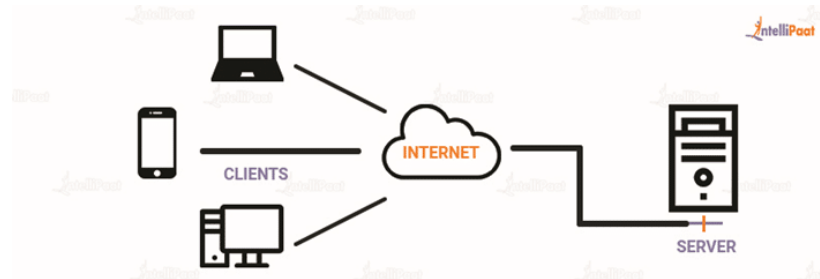
Figure 4. Client-Server Architecture
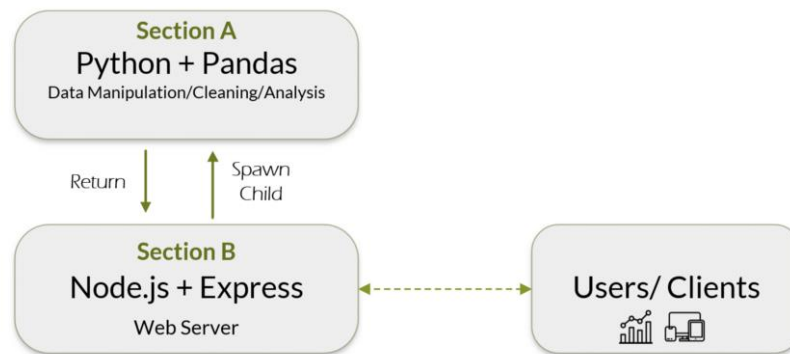
Figure 5. System Architecture

2. Platform Requirements:

To access the web project, HTML and JavaScript must be supportable by the browser. This should be done in accordance with Figure 4's browser compatibility.

o Modern web browsers with JavaScript enabled in it.

Internet Explorer	Microsoft Edge	Mozilla Firefox	Chrome	Opera	Safari
8 +	Latest	Latest	22 +	17 +	12 +

Figure 6. JavaScript availability in modern browser versions.

➔ Database Design

The backend of the project will be a SQL-based relational database.

➤ Database Management System: MySQL

MySQL has been chosen because it is the world's most popular open-source database is that it provides comprehensive support for every application development need. Besides, it has simplicity.

> *Object Relational Model (ORM)*

The most significant reason to use an ORM is to be able to have a sophisticated, object-oriented business model while still being able to store it and run efficient queries against a relational database efficiently.

1. SQLAlchemy: It provides a full suite of well-known enterprise-level persistence patterns, designed for efficient and high-performing database access, adapted into a simple and Pythonic domain language [10].
2. Sequalize: Sequelize is a promise based Node.js ORM for Postgres, MySQL, MariaDB, SQLite and Microsoft SQL Server [11].

> *Conceptual Database Design (Entity Relationship Diagram)*
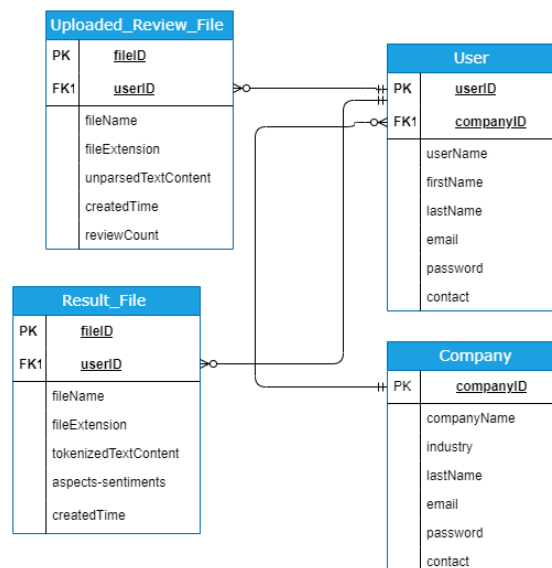
Figure 7. Entity Relationship Conceptual Database Schema

a. Company: Registry that will make business plans according to the analysis. Companies must have representatives and they are called customers. Company's working field must gather and store for future analysis. ABSA is changeable sector of the company such as food, health, tourism, etc.

b. User: Representatives of the company. Also, customer's contacts will be stored.

c. Uploaded_Review_File: Before performing an analysis, the user uploads this file to the system. This file had better to stored for showing where the analyzes came from and to access later.

d. Result_File: The file that is going to be created once the analysis is finished. This file must be stored to allow for subsequent analysis examination. Keeping files rather than graphics will be more efficient.

The database schema elements will be change in the future. As more improvements are variable and experimental, new tables will be added later.

III.   Methods and Techniques

Crisp-DM

CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, is an industry-proven way to guide the data mining efforts [12].

1. As a **methodology**, it includes descriptions of the typical phases of a project, the tasks involved with each phase, and an explanation of the relationships between these tasks.
2. As a **process model**, CRISP-DM provides an overview of the data mining life cycle.

This methodology will be used as a roadmap.
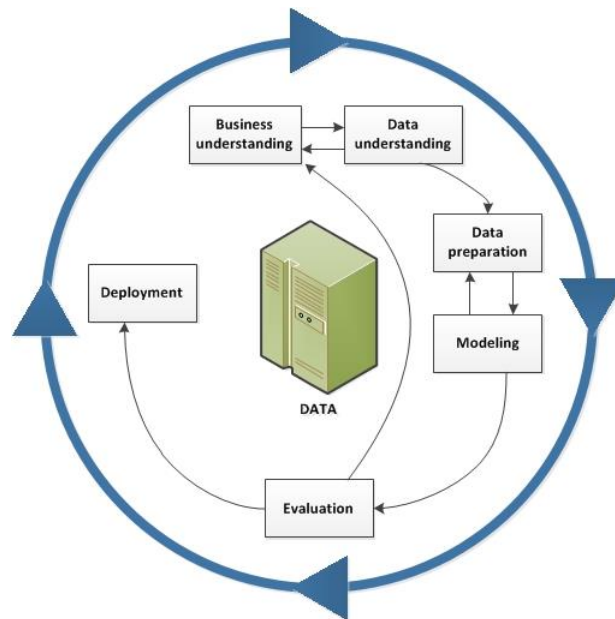
Figure 8. Data Mining Life Cycle

➢ Data Collection

Basically, data will be collected from Turkish datasets in the literature and reviews from plenty types of e-commerce web sites about product and services. The data target to be obtained must be at least 5,000 comments from different business areas (restaurant, hotel, etc.).

➢ Data Pre-processing

In order to make the obtained data more suitable for the model, it must go through the analysis, then cleaning and editing processes. In this way, it is aimed not to give data that does not need the model to be used. This process consists of such as dividing the text into sentences, separating the sentences

into words, removing stop words and punctuation, etc. All these operations will be performed using Python language and libraries.

1.   Tokenization:

Tokenization is the process of converting text into tokens before transforming it into vectors. It is also easier to filter out unnecessary tokens. For example, a document into paragraphs or sentences into words. In this case tokenizing partitions the reviews into words.

2.   Normalization:

Words which look different due to casing or written another way but are the same in meaning need to be process correctly. Normalization processes ensure that these words are treated equally. For example, changing numbers to their word equivalents or converting the casing of all the text.

'100' → 'one hundred'

'Apple' → 'apple'

3.   Lower Casing:

Converting character to the same case so the same words are recognized as the same. In this case letters can be converted to lowercase for conforming to the standard.

4.   Removing:

Some pieces are the most commonly occurs which are not relevant in the context of the data and do not contribute any deeper meaning to the phrase. Such as numbers, punctuations, stop words, URL and HTML tags should be removed.

5.   Stemming:

Stemming is the process of reducing inflected (or sometimes derived) words to their word stem, base, or root form.

6.   Lemmatization:

Lemmatization is similar to stemming in reducing inflected words to their word stem but differs in the way that it makes sure the root word (also called as lemma) belongs to the language.

➢ Data Labeling

Data Labeling refers to the process of adding tags or labels to raw data such as images, videos, text, and audio. These tags form a representation of what class of objects the data belongs to and helps a DL model learn to identify that class of objects when encountered in data without a tag [13].

More accurately labeled coupled with a larger quantity of labeled data creates more useful deep learning models, as the resulting machine learning model bases their decisions on all the labeled data [14].

Open-source Label Studio and Doccano tools will be used for data labeling.

1.   Label Studio:

Label Studio is an open-source data labeling tool for labeling and exploring multiple types of data. It provides to perform different types of labeling with many different data formats [15].

2.   Doccano:

Doccano is an open-source text annotation tool that was created with humans in mind. It has text categorization, sequence labeling, and sequence to sequence annotation characteristics [16].

➢ Training Model

A model that will allow larger analyzes and predictions from a small sample will be created and this model will be trained with the arranged data. Process of training model consists of 2 stages. the target-categories will be determined using the Named Entity Recognition (NER) model, and then the polarization of the detected target-categories will be extracted. Depending on performance of model data adequacy will be checked. If it's inadequate when model performance checked, number of data will be increased. All these operations will be performed using Python language and libraries. Pretrained language models such as BERT, DistilBERT and ELECTRA, which are thought to increase the success rate, and deep learning techniques such as CNN, RNN, and LSTM will be used. Generally, stages of model training and test follow same way and are outlined in Figure 8.
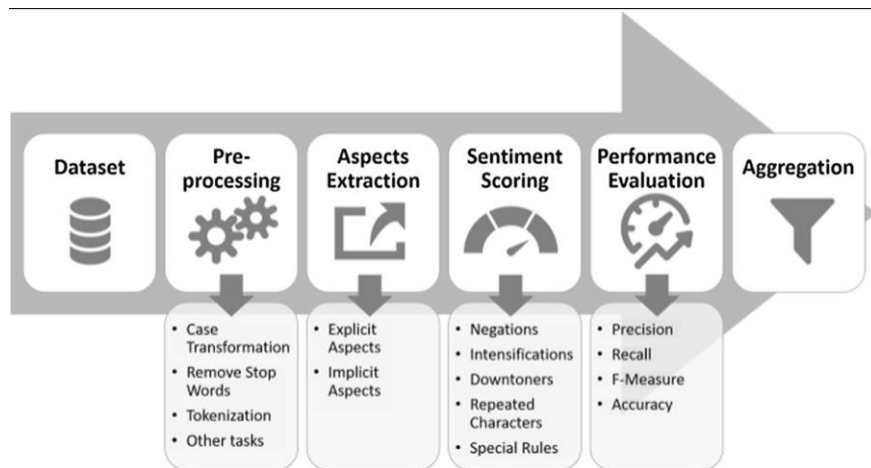

Figure 9. ABSA Steps

a) Pretrained Language Models

PLMs (Pretrained Language Models) are huge neural networks used for a range of NLP tasks. They work on a pretrain-finetune model, in which models are pretrained on a huge text corpus before being finetuned on a downstream task. The following are the most common models that will be used in the project.

1.   BERT (Bidirectional Encoder Representations from Transformers):

In 2018, Google announced this model. As understood from its name, that model analyzes text from both right and left side. In this way, semantics and relationship between words are analyzed better. This model is developed to solve plenty types of problems such as Chatbot, text classification, etc. In addition, this development process took around 4 days with 16 TPU for bert_large (used 2.5M row data) and with 4 TPU for bert_base (2,5B row data). Lastly, to solve and specified model in the problems, extra layers must be added.

1. DistilBERT:

   DistilBERT is a compact, fast, inexpensive, and light Transformer model that has been trained using BERT base. It has 40% less parameters than BERT-base-uncased, and it runs 60% quicker while keeping over 95% of BERT's performance on the GLUE language understanding benchmark [17].

1. ELECTRA:

   It is a method for self-supervised language representation learning. It can be used to pre-train transformer networks using relatively little compute. ELECTRA models are trained to distinguish "real" input tokens vs "fake" input tokens generated by another neural network, similar to the discriminator of a GAN [18] .

b) Techniques In Training Stage

Although there are current Turkish language solutions, these solutions typically give conventional SA services based on classic Machine Learning algorithms, and the outputs of this approach do not provide sufficient and meaningful information to product and service manufacturers. Beyond traditional sentiment analysis, this project will feature ABSA, as well as the usage of big and complicated language models based on Deep Learning and Natural Language Processing technologies. As a result, consumers with a high success rate will receive more beneficial and valuable outputs. The techniques to be used are listed below.

1. Convolutional Neural Networks (CNN):

   Neural networks are a set of algorithms designed to recognize patterns. These patterns are numbers contained in vectors that are translated from real-world data such as images, sound, text, or time series. A convolutional neural network is a neural network that applies convolutional layers to local features [19].

1. Recurrent Neural Networks (RNN):

   A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data. These deep learning algorithms are commonly used for ordinal or temporal problems, such as language translation, NLP, speech recognition, and image captioning; they are incorporated into popular applications such as Siri, voice search, and Google Translate. Like feedforward and CNNs, RNNs utilize training data to learn. They are distinguished by their "memory" as they take information from prior inputs to influence the current input and output. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of RNNs depend on the prior elements within the sequence. While future events would also be helpful in determining the output of a given sequence, unidirectional RNNs cannot account for these events in their predictions [20].

Recurrent Neural Network vs. Feedforward Neural Network
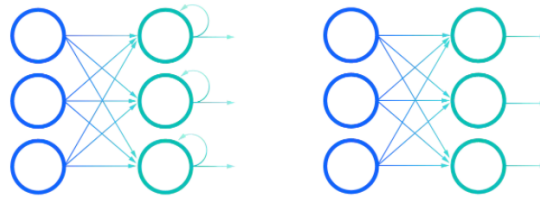
Comparison of Recurrent Neural Networks (on the left) and
Feedforward Neural Networks (on the right)

Figure 10. RNN

1. **Long Short-Term Memory (LSTM):**

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. It can process not only single data points (such as images), but also entire sequences of data (such as speech or video). For example, LSTM is applicable to tasks such as unsegmented, connected handwriting recognition, speech recognition and anomaly detection in network traffic or IDSs (intrusion detection systems) [21] .

➢ **Web Application Development**

Web application will be developed using NodeJS which is a framework of JavaScript. The main purpose is to develop a web application that can provide ease of use and accessibility, providing a result of analyzes using a variety of visualization techniques.

➢ **Establish Connection the Model and Web Application**

To ensure that the analyzes and predictions of the model are used in the Web Application appropriately. At this stage, communication with the Web Application will be provided by using the necessary Python services and web sockets.

IV. Detailed System Architecture

➢ **Input/Output Diagrams**

In systems analysis and software engineering, the input–process–output (IPO) model is a widely used approach for characterizing the structure of an information processing program or another process.
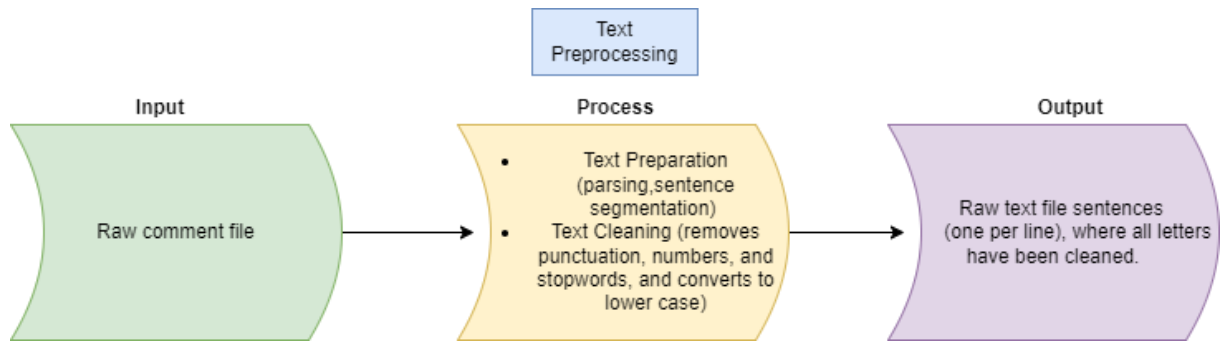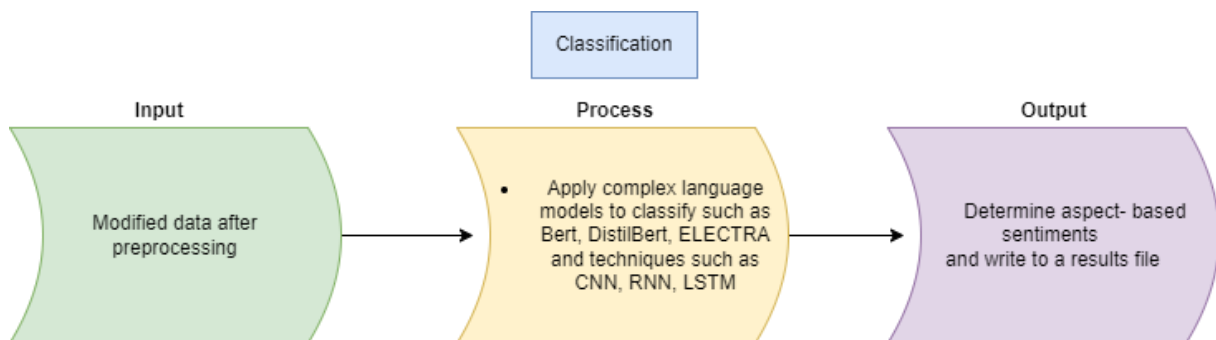
Text
Preprocessing

Input	Process	Output
Raw comment file	• Text Preparation (parsing, sentence segmentation) • Text Cleaning (removes punctuation, numbers, and stopwords, and converts to lower case)	Raw text file sentences (one per line), where all letters have been cleaned.

Figure 11. Text Processing

Classification

Input	Process	Output
Modified data after preprocessing	• Apply complex language models to classify such as Bert, DistilBert, ELECTRA and techniques such as CNN, RNN, LSTM	Determine aspect- based sentiments and write to a results file

Figure 12. Classification

Visualizing

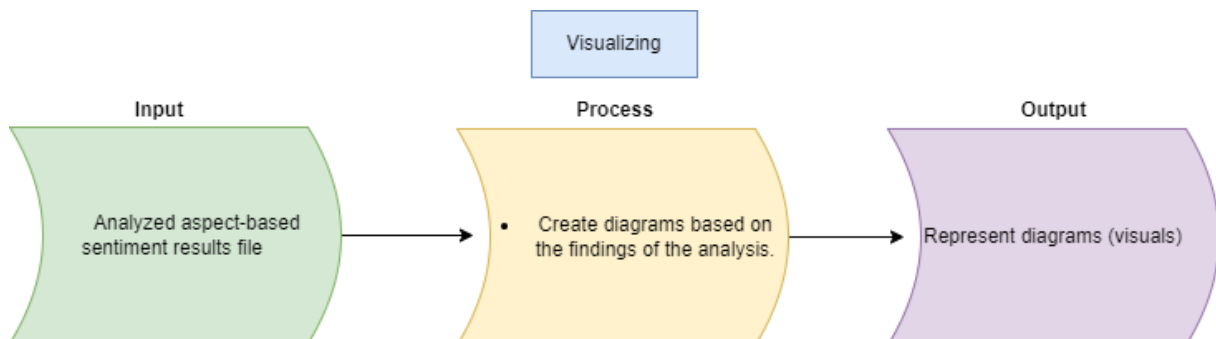| Input	Process	Output
Analyzed aspect-based sentiment results file	• Create diagrams based on the findings of the analysis.	Represent diagrams (visuals)

Figure 13. Visualizing

➢ Definitions of Functions and APIs

➢ *Functions*
  ➢ User Login/Registration (Authentication/Authorization)
        The process of confirming a person's identification is known as authentication. A web application can be interacted with in a variety of ways. Using user levels, access to specific

actions or pages can restrict. The process of regulating user access by assigning roles and privileges is known as authorization.

The user will be able to register and log in as an individual on behalf of the customer-company authorized company. The authorization and authentication phases for the company are critical in this regard.

➢ File Uploading for Analyze

Users should be able to upload a comment-review file that they want to analyze in the system.

➢ Showing Visualizations in the Dashboard

On the dashboard, the system displays the analysis results. As a result, users can study each analysis performed by the system from the reviews file in great depth.

➢ Dashboard for Previous Analysis Results

From the dashboard, users should be able to examine the results of prior analyses. As a result, they will be able to compare and investigate in depth. This section's goal is to make future analysis more straightforward.

➢ *APIs:*

An application programming interface, or API, enables companies to open up their applications' data and functionality to external third-party developers, business partners, and internal departments within their companies. This allows services and products to communicate with each other and leverage each other's data and functionality through a documented interface. Developers don't need to know how an API is implemented; they simply use the interface to communicate with other products and services. API use has surged over the past decade, to the degree that many of the most popular web applications today would not be possible without APIs [22].

➢ python-bridge API

Node.js to Python bridge to ensure communication.

➢ RESTful API

Representational state transfer (REST) defines a set of standards for web services. An API is an interface that software programs use to communicate with each other. Therefore, a RESTful API is an API that conforms to the REST architectural style and constraints. REST systems are stateless, scalable, cacheable, and have a uniform interface [23]. For sending information to the client, it will need to use.

➢ Class Diagram

The UML Class diagram is a graphical notation used to construct and visualize object-oriented   systems. A class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's:

➢ classes,

➢ their attributes,

➢ operations (or methods),
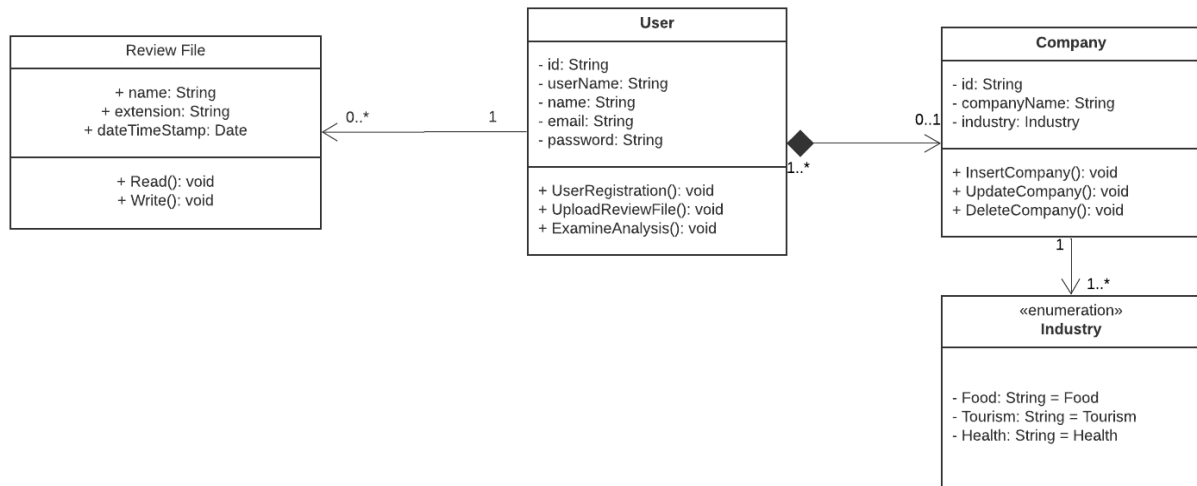
➢ and the relationships among objects [24].

Figure 14. Class Diagram

Class diagrams explained above. In order to explain detailly:

e. User Class:

> User class has attributes which is related to person such as username, name, email, password. This user has some methods that can be used on web application such as UserRegistration (register to the web site), UploadReviewFile (uploading a review file to analyze and see the results in visual format) and ExamineAnalysis (checking previous results and current results after analyzed finished from user panel).

f. Company Class:

> Company class has attributes which represents the company such as companyName, industry. Gathering the company's industry will be necessary for analysis. Analyzes will be conducted by industry-specific structured model. It should be used as an enumaration to begin with, as there is no way to train models for every industry. This company has some methods which can initialize itself such as InsertCompany (to register company to the system), UpdateCompany (update the name or industry of the company) and DeleteCompany (delete the company from the system).

g. Review File Class:

> Review File class represents the file which needs to be uploaded by user. It has some attributes such as name, extension of the file, dateTimeStamp (to define when it is uploaded

to the system). In addition, it has attributes for the system such as read and write file. Because file needs to be read and written by the system.
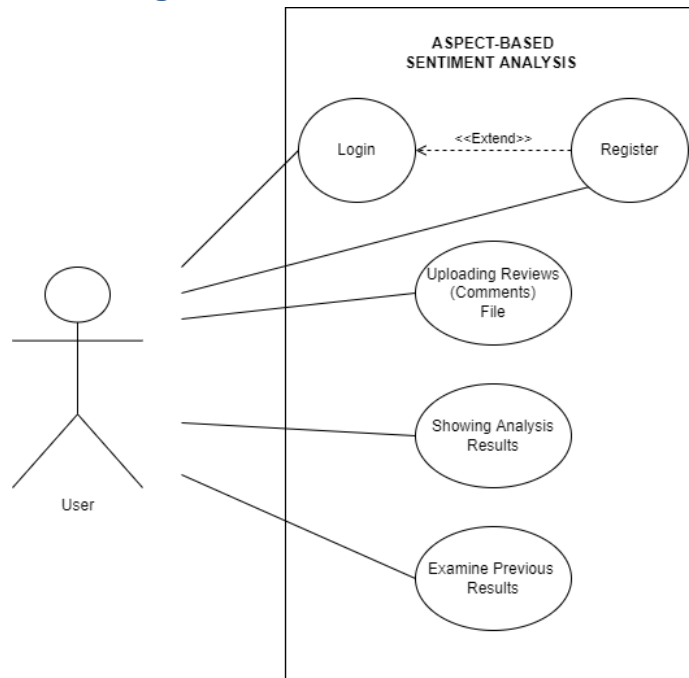
➢ Use Case Diagram

Figure 15. Use Case Diagram

➢ Use Case Definition

Use Case Name	Login	
Description	Login case describes how the user logs into a system. If the user exists, then directs to main page and have an access. Otherwise, user doesn't exist, or wrong password would be some problems.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must exist in the database.	
Post-Condition	The user will be logged in	
Main Scenario	Serial No	Steps
	1	Enter username and password

	2	Login information will be validated
	3	User will be redirected to the main menu
Extension	2a	Invalid username Error message will be shown
	2b	Invalid password Error message will be shown
	2c	Forgot password A password reset page will be sent to the user's email.

Use Case Name	Register	
Description	If no one else has registered with the same email address, register use case will add the user's information to the database.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must not exist in the database.	
Post-Condition	The user will be registered in	
Main Scenario	**Serial No**	**Steps**
	1	Enter username, name, email, password, company, and department under the company.
	2	User information will be added to the database.
	3	User will be redirected to the login page.
Extension	2a	If email or username already exists. Error message will be shown.

Use Case Name	Uploading Reviews (Comments) File	
Description	The user uploads the review-comment file to be analyzed. Initial step to start the analysis.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must be logged in. File type will be suitable for standards.	
Post-Condition	The file will be sent for analysis.	
Main Scenario	Serial No	Steps
	1	The user clicks upload button and select the review file in requested formats.
	2	After file is approved by the user, analysis process begins.
Extension	2a	Uploading invalid file. Error message will be shown
	2b	If the file is not approved by the user, process will not begin for the file.

Use Case Name	Showing Analysis Results
Description	Defines showing analysis results by visualization techniques such as graphics, tables, etc. User can examine easily what system performed from the file.
Actors	User

Pre-Condition	The device must be connected to the internet. The user must be logged in. The user must be uploaded reviews file.	
Post-Condition	The user will be examined results	
Main Scenario	**Serial No**	**Steps**
	1	User opens the main menu
	2	User uploads the reviews file
	3	Result visualizations-data are shown to the user and saved.
Extension	2a	Uploading invalid file Error message will be shown
	2b	Uploading file in different format Error message will be shown

Use Case Name	Examine Previous Results from Dashboard	
Description	Defines checking previous results from user's records if user has already done some analysis. That presents a comparison between previous results and new one.	
Actors	User	
Pre-Condition	The device must be connected to the internet. The user must be logged in. The user must be done analysis before.	
Post-Condition	The user will be checked results	
Main Scenario	**Serial No**	**Steps**
	1	User opens the dashboard
	2	User checks the results and make comparison
Extension	2a	Hasn't done analysis yet Show message about not existing analysis

➢ Data Flow

A Data Flow Diagram (DFD) is a classic visual representation of how data moves through a system. A tidy and clear DFD can graphically display the correct amount of system need. It can be done manually, automatically, or in a combination of the two.

It depicts how data enters and exits the system, as well as what changes the data and where it is stored.

A DFD's goal is to demonstrate the breadth and bounds of a system. It can be used as a tool for communication between a system analyst and everyone involved in the order that serves as a beginning point for system change. A data flow graph or bubble chart is another name for the DFD.

Below, detailly explained the data flow of how to analyze data (Figure 16) and data flow of the system (Figure 17). In addition, analyze data flow is put into the system data flow as a blue process.
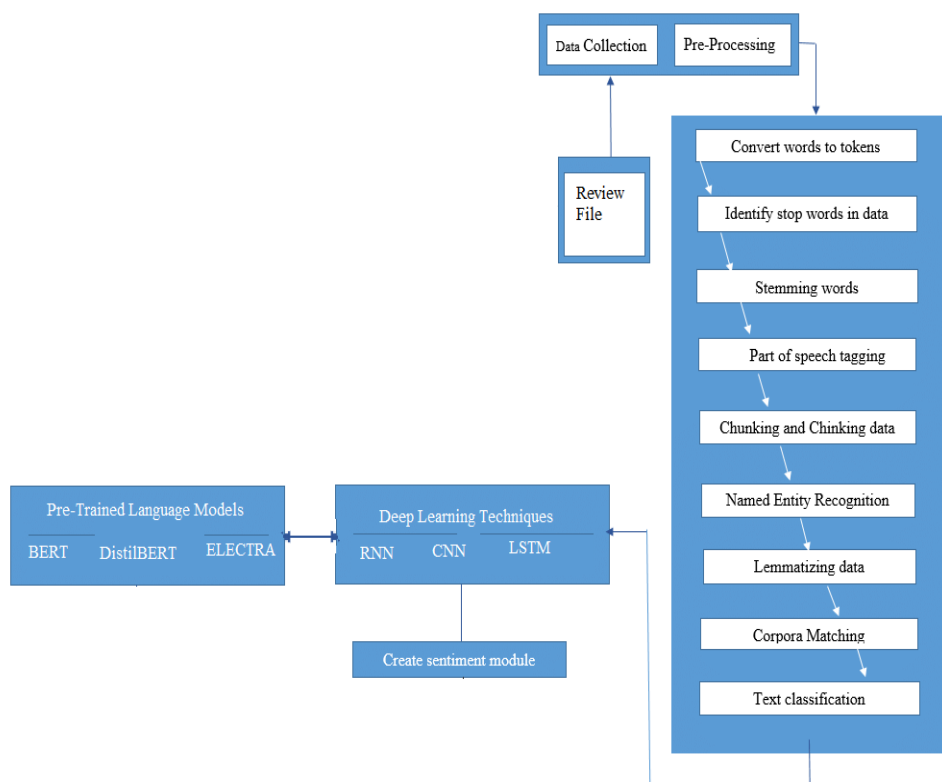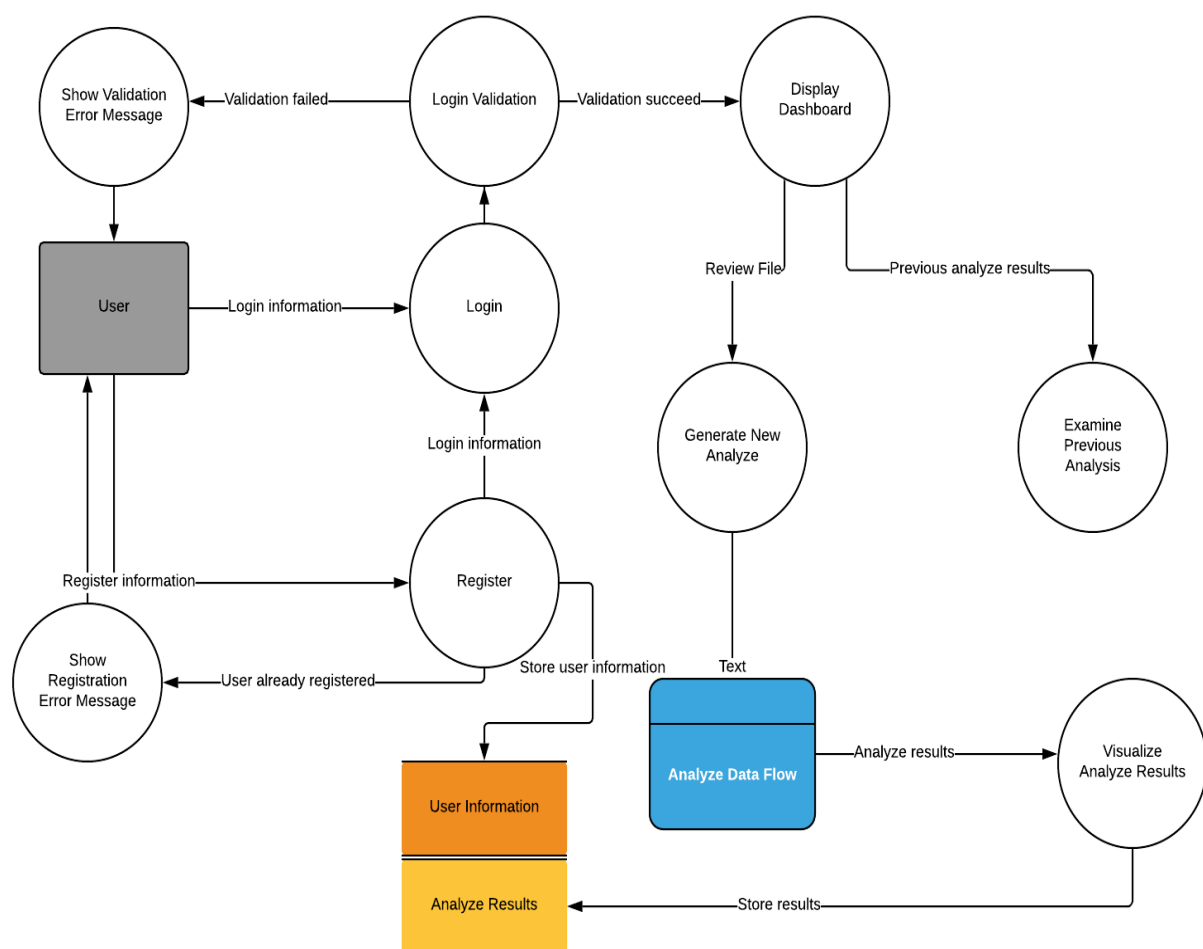

Figure 16. Analyze Data Flow
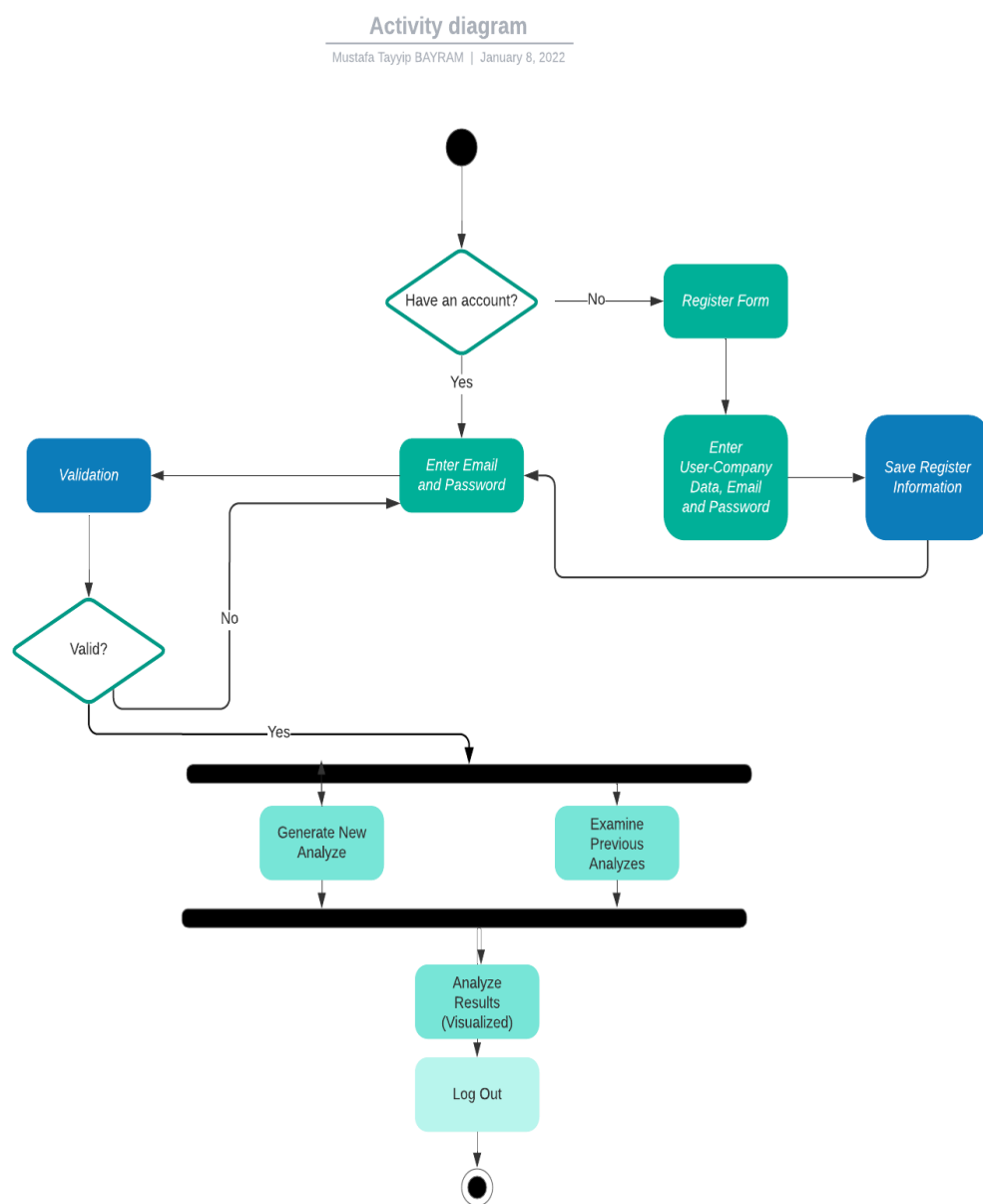
Figure 17. System Data Flow Diagram

> ➢ Activity Diagram

Figure 18. Activity Diagram
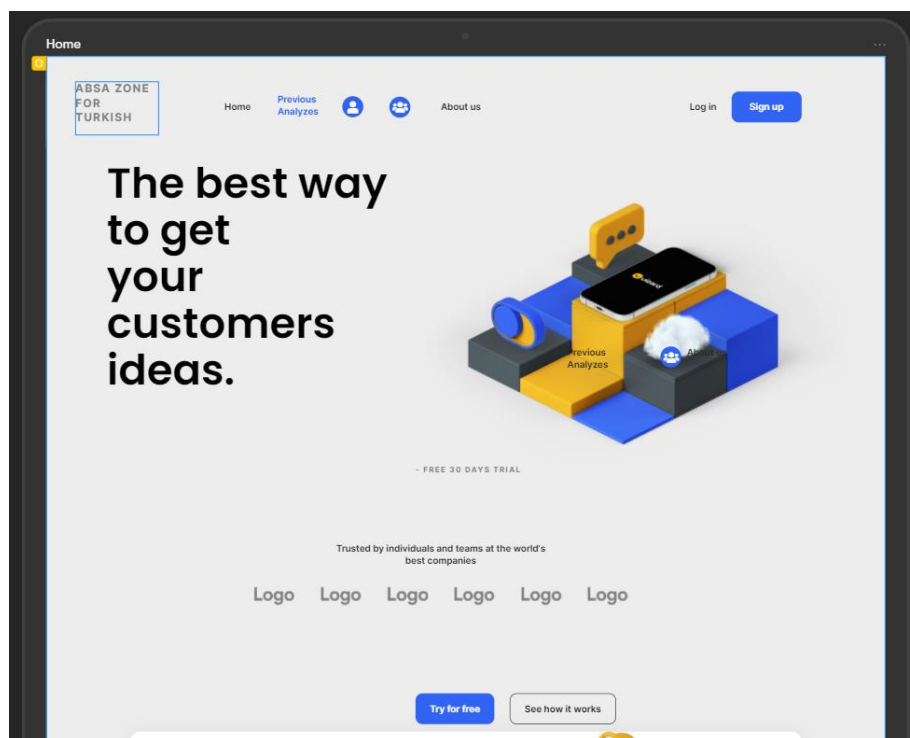
➢ User Interface Design
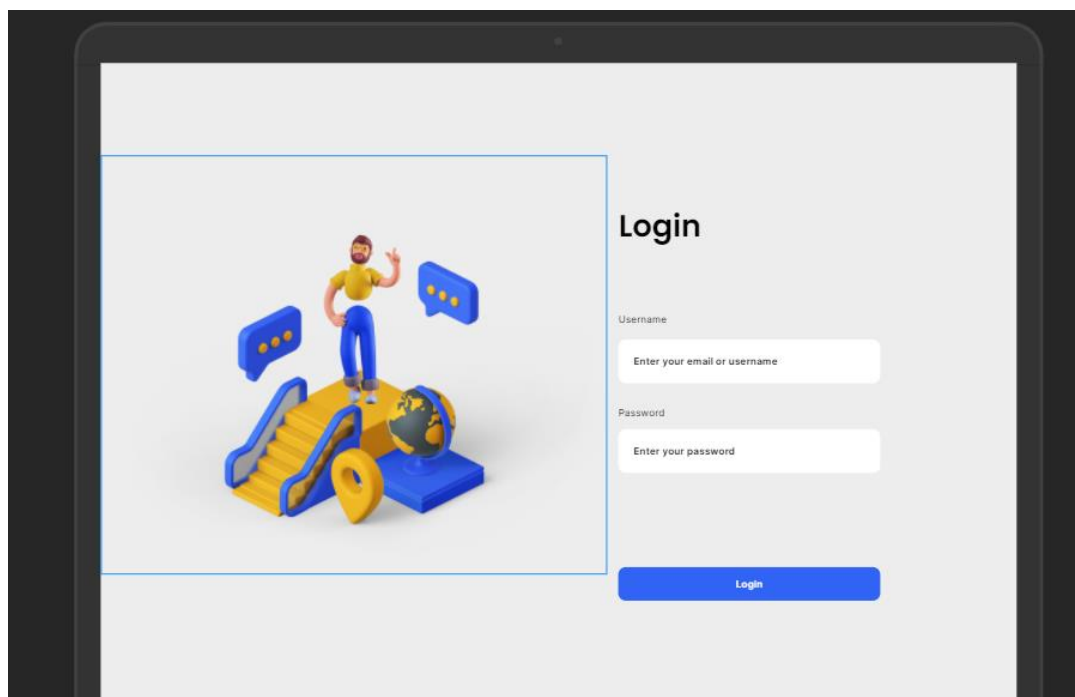

Figure 19. Main Dashboard
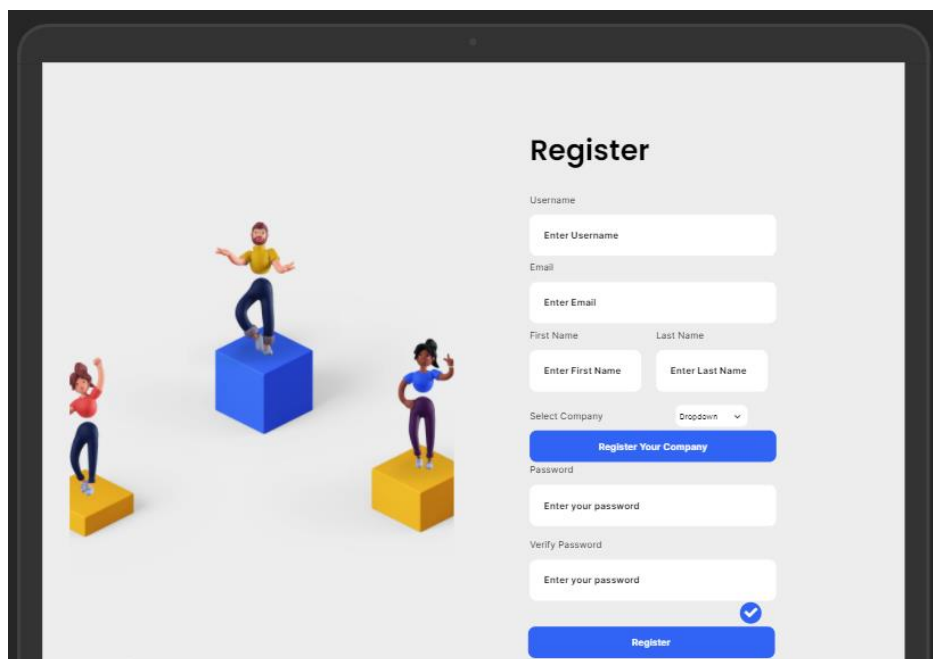
Figure 20. Login Page

Figure 21. Register
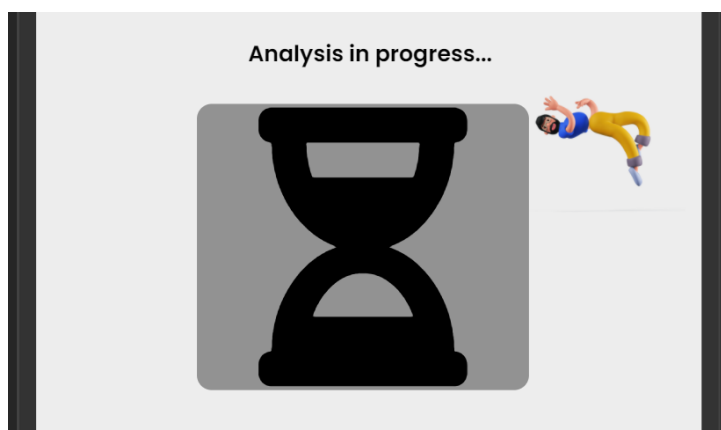
Figure 22. Uploading File
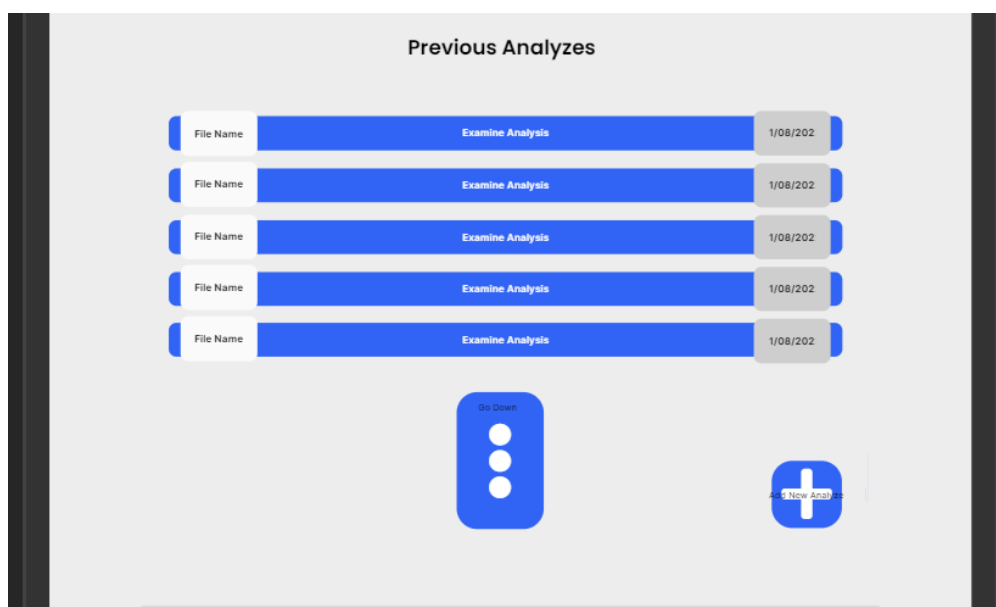

Figure 23. Analyze Progress

Figure 24. Previous Analyzes

Figure 25. Analyze Results (Microsoft PowerBI)

References

[1]   Ş. E. ŞEKER, «Yönetim Bilişim Sistemleri Ansiklopedisi Cilt 3, Sayı 3 Eylül 2016,» [Çevrimiçi]. Available: https://ybsansiklopedi.com/wp-content/uploads/2016/09/aralik20163-3kapak.pdf#page=19. [Erişildi: 18 18 2021].

[2]   J. McCarthy, "WHAT IS ARTIFICIAL INTELLIGENCE?," Computer Science Department, [Online]. Available: https://homes.di.unimi.it/borghese/Teaching/AdvancedIntelligentSystems/Old/IntelligentSyste

ms_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_what isai.pdf. [Accessed 26 12 2021].

[3]   IBM, "What is Deep Learning?," [Online]. Available: https://www.ibm.com/cloud/learn/deep-learning. [Accessed 26 12 2021].

[4]   IBM, "What are neural networks?," [Online]. Available: https://www.ibm.com/cloud/learn/neural-networks. [Accessed 26 12 2021].

[5]   Wikipedia, "Natural language processing," [Online]. Available: https://en.wikipedia.org/wiki/Natural_language_processing. [Accessed 28 12 2021].

[6]   Wikipedia, "Named-entity recognition," [Online]. Available: https://en.wikipedia.org/wiki/Named-entity_recognition. [Accessed 26 12 2021].

[7]   Brainstation, "What Is Web Development?," [Online]. Available: https://brainstation.io/career-guides/what-is-web-development. [Accessed 28 12 2021].

[8]   techopedia, "Database Management System (DBMS)," [Online]. Available: https://www.techopedia.com/definition/24361/database-management-systems-dbms. [Accessed 28 12 2021].

[9]   P. Shahnawaz and Astya, "Sentiment analysis: Approaches and open issues," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017, pp. 154-158.

[10]   "SQLAlchemy," [Online]. Available: https://www.sqlalchemy.org/. [Accessed 28 12 2021].

[11]   "Sequalize," [Online]. Available: https://sequelize.org/. [Accessed 28 12 2021].

[12]   IBM, "CRISP-DM Help Overview," [Online]. Available: https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview. [Accessed 26 12 2021].

[13]   H. Bandyopadhyay, "What is Data Labeling and How to Do It Efficiently," [Online]. Available: https://www.v7labs.com/blog/data-labeling-guide#data-labeling. [Accessed 26 12 2021].

[14]   Labelbox, "Data Labeling 101," [Online]. Available: https://labelbox.com/data-labeling-overview. [Accessed 26 12 2021].

[15]   "Get started with Label Studio," Label Studio, [Online]. Available: https://labelstud.io/guide/#What-is-Label-Studio. [Accessed 26 12 2021].

[16]   "Doccano," [Online]. Available: https://doccano.github.io/doccano/. [Accessed 26 12 2021].

[17]   "DistilBERT," Hugging Face, [Online]. Available: https://huggingface.co/docs/transformers/model_doc/distilbert. [Accessed 26 12 2021].

[18]   K. C. a. M.-T. L. a. Q. V. L. a. C. D. Manning, "ELEKTRA Pre-training Text Encoders as Discriminators Rather Than Generators," [Online]. Available: https://openreview.net/pdf?id=r1xMH1BtvB. [Accessed 26 12 2021].

[19]   K. Tran, "Towards Data Science | Convolutional Neural Network in Natural Language Processing," [Online]. Available: https://towardsdatascience.com/convolutional-neural-network-in-natural-language-processing-96d67f91275c. [Accessed 26 12 2021].

[20]   IBM, "Recurrent Neural Networks," [Online]. Available: https://www.ibm.com/cloud/learn/recurrent-neural-networks. [Accessed 26 12 2021].

[21]   Wikipedia, "Long short-term memory," [Online]. Available: https://en.wikipedia.org/wiki/Long_short-term_memory. [Accessed 26 12 2021].

[22]   IBM, "What is an application programming interface (API)?," [Online]. Available: https://www.ibm.com/topics/api. [Accessed 3 1 2022].

[23]   T. Rascia, "Node.js, Express.js, and PostgreSQL: CRUD REST API example," [Online]. Available: https://blog.logrocket.com/nodejs-expressjs-postgresql-crud-rest-api-example/#whatisarestfulapi. [Accessed 2 1 2022].

[24]   V. Paradigm, "UML Class Diagram Tutorial," [Online]. Available: https://www.visual-paradigm.com/guide/uml-unified-modeling-language/uml-class-diagram-tutorial/. [Accessed 3 1 2022].

[25]   "Arxiv," [Online]. Available: https://arxiv.org/ftp/arxiv/papers/2006/2006.04611.pdf. [Accessed 10 2021].

[26]   "Research Gate," [Online]. Available: https://www.researchgate.net/publication/335421201_Aspect Based_Sentiment_Analysis_Methods_in_Recent_Years. [Accessed 10 2021].

[27]   "TowardsDS," [Online]. Available: https://towardsdatascience.com/aspect-based-sentiment-analysis-using-spacy-textblob-4c8de3e0d2b9. [Accessed 10 2021].

[28]   "MonkeyLearn," [Online]. Available: https://monkeylearn.com/blog/aspect-based-sentiment-analysis/. [Accessed 10 2021].

[29]   "Trust Pilot," [Online]. Available: https://www.trustpilot.com/about. [Accessed 10 2021].

[30]   "Medium," [Online]. Available: https://intellica-ai.medium.com/aspect-based-sentiment-analysis-everything-you-wanted-to-know-1be41572e238. [Accessed 10 2021].

[31]   IBM, "CRISP-DM Help Overview," [Online]. Available: https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview.