TED UNIVERSITY

CMPE 491

AI-driven Review Aggregator and Analyzer for Strategic Product Insights
Analysis Report

Submission Date:
19.04.2024

Team Members:

Harun UĞURLU
Murat Efe DOĞAN
Ezgi Gülce YAZICI

Supervisor:

Dr. Venera ADANOVA

Jury Members:

Prof. Dr. Müslim BOZYİĞİT
Dr. Emin KUĞU

1. Introduction

AI-driven Review Aggregator and Analyzer for Strategic Product Insights project, which seeks to create an advanced tool utilizing artificial intelligence and web scraping methods. Its primary goal is to gather and analyze product reviews from various online platforms efficiently.

This report includes details of the project's objectives, requirements, and constraints ensuring clarity. Through this project, businesses aim to gain valuable insights into consumer sentiments, enhancing their decision-making processes and strategic planning efforts.

2. Current system

As this is a new project, there is currently no existing system in place to address the project. This document serves as an analysis report, providing a thorough examination of the requirements and considerations for the proposed system. As the project progresses, the recommendations and requirements outlined in this report will be integrated into the system during the development phase.

3. Proposed system

3.1 Overview

The system is designed to harness the power of artificial intelligence and web technologies in order to aggregate and analyze online reviews about a product. The primary goal is to provide businesses with valuable insights that can guide their development and marketing processes. By periodically scraping data from various online platforms and utilizing natural language processing techniques, the system will apply sentiment analysis and feature extraction. The outcome of this will be an extensive understanding of user sentiments towards the product, highlighting the strong and weak areas for improvement.

The system will provide a user-friendly dashboard for users that displays and intuitively visualizes the analyzed data, allowing users to view the trends, sentiment scores, and common themes in the reviews. The dashboard will have the functionalities to filter insights by date range, review sources, as well as exporting the data.

The dashboard will include a section to add new products and easily switchable tabs between different product review analyses.

3.2 System Boundaries

What the System Will Do?

- **Easy Product Submission Form:** It will include a form where the user can submit a new product and select the sources for review analysis.
- **Data Collection:** The system will periodically scrape the data from the online sources of the user's choice. It will handle large volumes of data and is configured to fetch updates at regular intervals.
- **Data Processing:** Exploiting a trained LLM, the system will process the data to perform sentiment analysis (positive, negative, and neutral) and identify common features in the reviews by sentiment category.
- **Data Display and Visualization:** Insights derived from the processed data will be exhibited to the user. It will include a sentiment score, reviews categorized by sentiment, the most mentioned features in reviews by sentiment, a filtering feature, and graphs that show the change in the reviews.
- **Report Generation:** Users will have the option to generate reports based on the analyzed data that can be configured to focus on specific aspects such as specific features of the product and date ranges for offline inspection.

What the System Will Not Do

- **Real-time Analysis:** While the system updates regularly, it is not designed for real-time analysis of the reviews. The frequency of updates will be determined by the scraping schedules and data processing capabilities.
- **Unstructured Data Processing:** The system will work with structured and semi-structured data. It will not handle unstructured data like images or videos.
- **Personal Data Collection:** The system will not collect or store personal data from the product reviews. All data collection and processing will focus on publicly available data and it will adhere to privacy laws and regulations.
-

3.3 Functional Requirements

1. Sign In, Sign Up:

Users should be able to securely sign in and sign up into the system using their credentials to access its features and functionalities

2. New Product Submission Form:

The system should provide a user-friendly onboarding form for submitting new products, allowing users to specify whether the product is physical or digital.

3. Platform Selection for Feedback Collection:

Users should be able to select specific online platforms(e.g., Hepsiburada, Trendyol, Google Play Store, Steam) from which to collect product feedback, ensuring targeted data retrieval.

4. Product Reviews Dashboard:

The system should present a comprehensive dashboard displaying aggregated product reviews, sentiments, and insights for easy access and analysis.

5. Customer Reviews Viewing:

Users should have the ability to directly view customer reviews within the system, facilitating quick assessment and understanding of consumer feedback.

6. Positive and Negative Review Graphs:

The system should generate graphical representations, such as line graphs, illustrating trends in positive and negative reviews over time to visualize changes effectively.

7. Positive and Negative Review Percentage Pie Chart:
Users should be presented with pie charts depicting the percentage distribution of positive and negative reviews, providing a clear overview of sentiment distribution.

8. Liked and Disliked Features Analysis:

The system should analyze product reviews to identify the top five most liked and disliked features and display them as a list to the user, allowing users to understand consumer preferences and concerns easily.

9. Product Tabs for Dashboard Navigation:

The dashboard should include intuitive tabs or navigation options, enabling users to switch seamlessly between different product dashboards for efficient analysis.

10. Feedback and Data Collection with Web Scraping:

The system should employ web scraping techniques to collect product reviews periodically from various online sources, ensuring a comprehensive dataset for analysis.

11.AI Review Analysis:

The system should analyze collected reviews to extract insights, sentiments, and trends using the power of artificial intelligence, providing users with valuable information for decision-making.

12. Review Filtering:

Users should be able to filter the analyzed results in the dashboard using different aspects such as by date range, platform, specific feature of the product.

13. Data Export:

Users should be able to export the data into a form such as a spreadsheet for offline inspection of the data using the dashboard.

3.4 Nonfunctional Requirements

1. Performance

- Response Time: The system should respond to user queries and display results within 2 seconds under normal load conditions.
- Throughput: The system must be capable of processing at least 1000 reviews per minute during peak hours without degradation in performance.
- Scalability: The system should be scalable to accommodate an increasing amount of data and simultaneous users, especially during product launch events or marketing campaigns.

2. Reliability

- Availability: The system should be available 99.9% of the time, ensuring minimal downtime.
- Fault Tolerance: The system must have mechanisms to recover from common failures without affecting the end-user experience. This includes redundancies for data storage and automatic failover processes.
- Data Integrity: Ensure that data is accurately processed, stored, and retrieved without errors or corruption.

3. Usability

- User Interface: The dashboard must be intuitive and easy to navigate, even for users with limited technical skills. It should also be responsive to different screen sizes and devices.
- Accessibility: The system should be accessible according to international web accessibility standards, such as WCAG 2.1.
- Documentation: Provide comprehensive user documentation that includes guides on system usage, troubleshooting, and FAQs.

4. Security

- Data Security: Implement industry-standard security measures, including data encryption at rest and in transit, to protect sensitive information from unauthorized access.
- Authentication and Authorization: The system should support secure login mechanisms and ensure that users can only access data and functionalities relevant to their permissions.
- Audit Trails: Maintain detailed logs of user activities and system changes to aid in monitoring and forensic analysis.

5. Maintainability

- Modularity: The system should be designed in a modular fashion, allowing individual components to be updated independently without impacting the entire system.
- Code Quality: Adhere to coding standards that promote readability and reusability of the source code.
- Updates and Patches: The system should support easy implementation of updates and patches to address any issues and improve functionality over time.

6. Scalability

- Horizontal Scaling: The system architecture should support horizontal scaling to manage increases in load by adding more machines or instances as needed.
- Vertical Scaling: The system should also support vertical scaling to enhance the capabilities of existing machines to handle greater loads.

3.5 Pseudo requirements

Support for Emoji Analysis: Demanding detailed analysis of emojis used in reviews, including sentiment analysis and frequency tracking, which might be interesting but not critical for understanding overall product sentiment.

Real-time Customer Support Chatbot: The inclusion of a real-time chatbot for customer support directly within the platform, could add complexity and maintenance overhead without significantly enhancing the core functionality of reviewing products.
Integration with Every Payment Gateway: Expecting integration with every possible payment gateway for handling subscription fees or premium features, even if focusing on a few widely used options would suffice for the target audience.
Customizable Profile Picture and Username: Requesting the ability for users to upload and change their profile picture and username at any time.

3.6 System models

3.6.1 Scenarios

1. Scenario: New User Registration

- **Scenario:** A new user wants to access the system to submit a product to receive product reviews. They visit the platform and click on the "Sign Up" button.
- **Action:** The user fills out the registration form with their email address and creates a password. They submit the form and receive a verification email.
- **Outcome:** After verifying their email, the user successfully signs up for the system and gains access to its features.

2. Scenario: User Login

- **Scenario:** A user wants to log in to the system.
- Action: The user fills out the login form with their email address and password.
- Outcome: If the user enters their email and password correctly, the user successfully logs in to the system.

3. Scenario: User Forgets Password

- **Scenario:** A user attempts to log in to the system but realizes they have forgotten their password.
- **Action:** The user clicks on the "Forgot Password?" link displayed on the login page.
- **Outcome:** The system prompts the user to enter the email address associated with their account. After entering the email address, the user submits the form.
- **Action:** The system sends a password reset link to the user's email address.
- **Outcome:** The user checks their email and clicks on the password reset link. They are directed to a page where they can enter a new password.
- **Action:** The user enters a new password and confirms it.
- **Outcome:** The system updates the user's password, and the user is redirected to the login page. They can now successfully log in using their new password.

4. Scenario: Submitting a New Product

- **Scenario:** A user, acting as a product manager, needs to add a new product to the system for feedback collection. They log in and navigate to the "New Product Submission" form.
- **Action:** The user fills out the form, specifying the product type (physical or digital), and enters relevant details such as product name, description, and category.
- **Outcome:** Upon submission, the new product is successfully added to the system, ready for feedback collection from users.

5. Scenario: Selecting Platforms for Feedback Collection

- **Scenario:** A user, acting as a marketing analyst, wants to collect feedback for a specific product from the Hepsiburada and Trendyol platforms only.
- **Action:** The user accesses the system's platform selection feature and chooses Hepsiburada and Trendyol from the available options.
- **Outcome:** The system retrieves feedback exclusively from the selected platforms, providing targeted data for analysis.

6. Scenario: Analyzing Product Reviews

- **Scenario:** A user, acting as a product analyst, needs to review customer feedback for a newly launched product.
- **Action:** The user navigates to the "Product Reviews Dashboard" and views the aggregated reviews, sentiments, and insights displayed.
- **Outcome:** The user gains valuable insights into customer sentiment and product performance, enabling informed decision-making.

7. Scenario: Selecting Product From Tabs

- **Scenario:** The user is exploring the dashboard, which allows navigation through different products using tabs.
- **Action:** The user clicks on one of the tabs corresponding to a specific product.
- **Outcome:** The system updates the dashboard to display insights, reviews, and sentiments specific to the selected product.

8. Scenario: Filtering Dashboard Insights by Date Range

- **Scenario:** A user, a marketing analyst, needs to analyze product feedback data within a specific timeframe.
- Action: The user navigates to the dashboard section of the system.
- **Outcome:** The system displays a range of insights, including aggregated product reviews, sentiments, and trends.
- **Action:** The user locates the filtering options for a date range.
- **Outcome:** The system presents dropdown menus or date pickers for selecting the desired date range.
- **Action:** The user selects a date range spanning the past month.
- **Outcome:** The system updates the dashboard to display insights based on the selected date range, providing a more focused view of the data for the specified timeframe.

9. Scenario: Filtering Dashboard Insights by Review Sources

- **Scenario:** A user, a product manager, wants to analyze feedback data from specific review sources.

- **Action:** The user navigates to the dashboard section of the system.
- **Outcome:** The system displays a range of insights, including aggregated product reviews, sentiments, and trends.
- **Action:** The user locates the filtering options for review sources.
- **Outcome:** The system presents checkboxes or dropdowns for selecting review sources.
- **Action:** The user chooses to filter reviews from a specific review source, such as a particular e-commerce platform.
- **Outcome:** The system updates the dashboard to display insights based on the selected review source, providing a more focused view of the data from the chosen platform.

10. Scenario: Exporting Dashboard Insights

- **Scenario:** A user, an operations manager, needs to export dashboard insights for further analysis.
- **Action:** The user navigates to the dashboard section of the system.
- **Outcome:** The system displays a range of insights, including aggregated product reviews, sentiments, and trends.
- **Action:** The user decides to export the dashboard insights for further analysis.
- **Outcome:** The system offers an "Export Data" button or similar functionality.
- **Action:** The user clicks on the "Export Data" button.
- **Outcome:** The system generates a downloadable file containing the dashboard insights in a suitable format, such as CSV or Excel.

3.6.2 Use case model

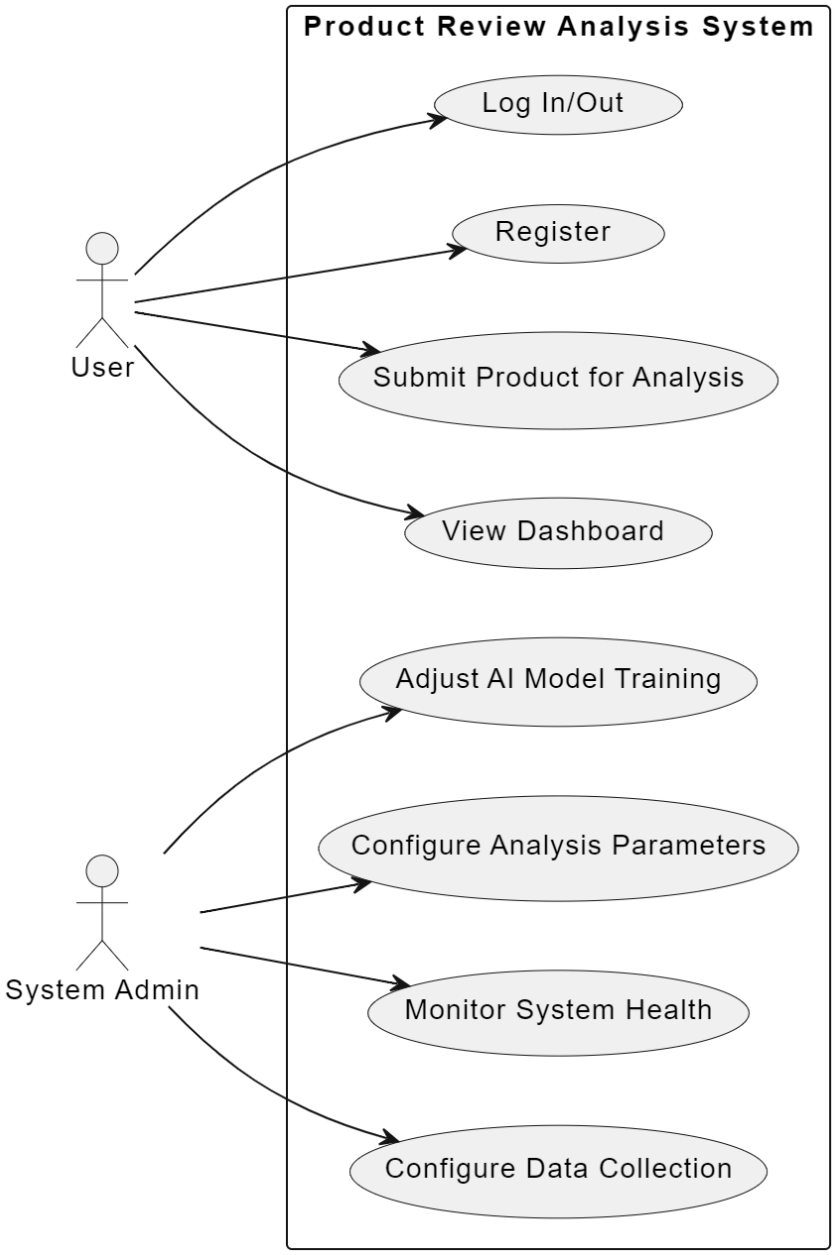

Figure 1: Use Case model of the system

- **WebScraper:** This component not shown in the use case diagram is responsible for automatically collecting data from various online platforms. It acts autonomously once configured or triggered by an administrator or scheduled task. It does not require continuous oversight or interaction from an administrator.
- **AI Model:** This component not shown in the use case diagram is an automated system that processes data, performs sentiment analysis, and extracts features from the collected data. It operates on the data that is fed into it, typically after being

triggered by a system event or administrator action. The analysis it conducts is autonomous and based on its pre-trained algorithms.

- **Log In/Out:** Users can authenticate themselves to access the system.
- **Register:** Users can create a new account to use the system.
- **Submit Product for Analysis:** Users can input products into the system for further review and analysis.
- **Configure Data Collection:** System Administrators set up and manage the parameters for the WebScraper, determining what data it should collect and how frequently.
- **View Dashboard:** Users access a visual interface displaying the results of product reviews and analysis.
- **Configure Analysis Parameters:** System Administrators set up and adjust the parameters for the AI Model's analysis, such as specifying keywords or phrases for feature extraction.
- **Monitor System Health:** System Administrators oversee the operational status of the system, including the WebScraper and AI Model, to ensure they are functioning correctly.
- **Adjust AI Model Training:** System Administrators manage the training of the AI Model, which involves providing it with new data, tweaking learning parameters, or initiating re-training cycles.

3.6.3 Object and class model

In the Object and Class Model of the AI-Driven Review Aggregator and Analyzer for Strategic Product Insights project, several key entities are identified to facilitate the system's functionality and organization:

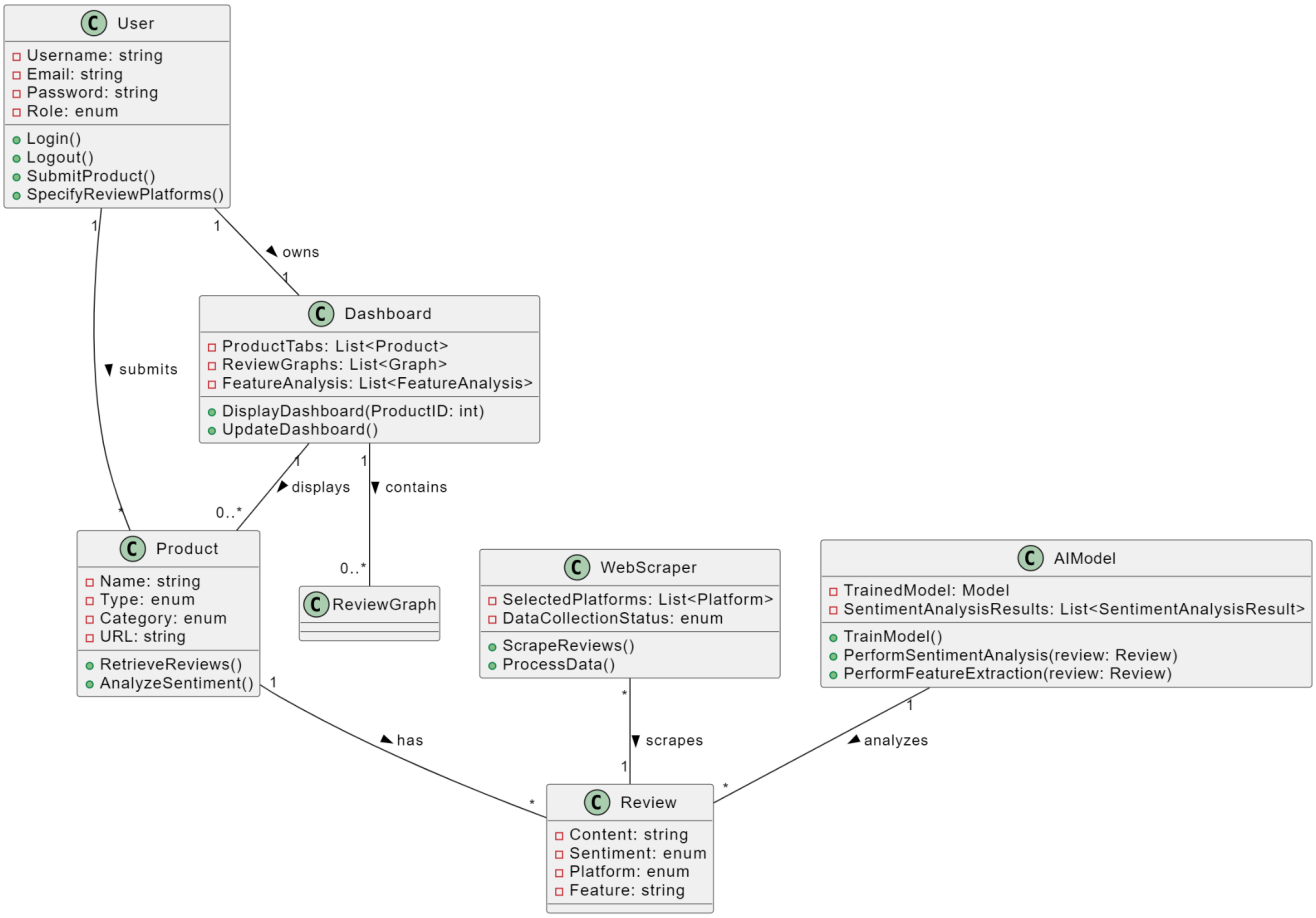

Figure 2: Class model of the system

1. **User Class:**
    - **Attributes:** Username (string), Email (string), Password (string), Role (enum)
    - **Methods:** Login(), Logout(), SubmitProduct(), SpecifyReviewPlatforms()
    - **Description:** Represents users interacting with the system. Users have distinct roles, such as administrators or regular users, and can perform actions like logging in, logging out, and submitting new products for analysis.

2. **Product Class:**
    - **Attributes:** Name (string), Type (enum), URL, Category (enum)
    - **Methods:** RetrieveReviews(), AnalyzeSentiment()
    - **Description:** Represents individual products submitted for analysis. Each product has a name, type (physical or digital), and URL linking to its online presence. Methods include retrieving reviews from online platforms and analyzing sentiment using AI models.
3. **Review Class:**
    - **Attributes:** Content (string), Sentiment (enum), Platform (enum), Feature
    - **Methods:** None
    - **Description:** Represents individual reviews collected from online platforms. Reviews contain textual content, sentiment classification (positive, negative, or neutral), and information about the platform from which they were retrieved.
4. **Dashboard Class:**
    - **Attributes:** ProductTabs (List<Product>), ReviewGraphs (List<Graph>), FeatureAnalysis (List<FeatureAnalysis>)
    - **Methods:** DisplayDashboard(ProductdID), UpdateDashboard()
    - **Description:** Represents the dashboard interface presenting analyzed insights to users. The dashboard consists of tabs for different products, graphs illustrating trends in reviews, and an analysis of liked and disliked product features. Methods include displaying and updating the dashboard content.
5. **WebScraper Class:**
    - **Attributes:** SelectedPlatforms (List<Platform>), DataCollectionStatus (enum)
    - **Methods:** ScrapeReviews(), ProcessData()
    - **Description:** Represents the web scraping component responsible for collecting product reviews from selected online platforms. The class manages the list of selected platforms for data collection and the status of the data collection process. Methods include scraping reviews from platforms and processing collected data.
6. **AIModel Class:**
    - **Attributes:** TrainedModel, SentimentAnalysisResults (List<SentimentAnalysisResult>)
    - **Methods:** TrainModel(), PerformSentimentAnalysis(), PerformFeatureExtraction()
    - **Description:** Represents the AI model used for sentiment analysis of product reviews. The class includes the trained model and stores the results of sentiment analysis. Methods include training the model and performing sentiment analysis on collected reviews.

3.6.4 Dynamic models

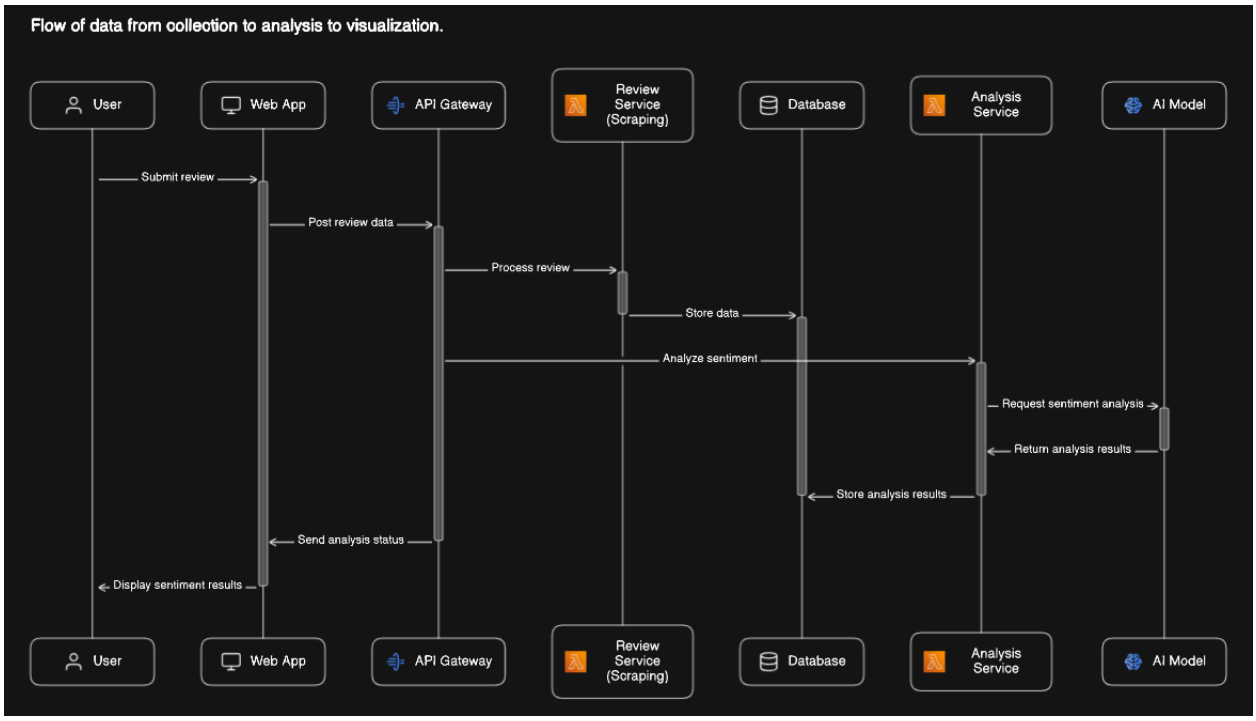

Figure 3: A sequence diagram of the system.

This sequence diagram in Figure 3, illustrates the end-to-end flow of data from the user submitting a review, the backend processing and analyzing the review, and finally, the results being displayed back to the user.

Participants:

1. **User:**
   Represents the end-user interacting with the web application.

2. **Web App:**
   The user interface where the user submits reviews and views analysis results.

3. **API Gateway:**
   Manages API requests and routes them to the appropriate backend services.

4. **Review Service (Scraping):**
   The backend service responsible for processing and scraping review data.

5. **Database:**
   Stores raw and processed review data.

6. **Analysis Service:**
   Responsible for performing sentiment analysis on the review data.

7. **AI Model:**
   The AI model used by the Analysis Service to perform sentiment analysis.

Flow of Events:

1. **Submit Review:**
   User submits a review via the Web App.

2. **Post Review Data:**
   The Web App sends the review data to the API Gateway.

3. **Process Review:**
   The API Gateway forwards the review data to the Review Service (Scraping).
   The Review Service (Scraping) processes the review data.

4. **Store Data:**
   The Review Service (Scraping) stores the processed review data in the Database.

5. **Analyze Sentiment:**
   The Review Service (Scraping) requests sentiment analysis from the Analysis Service.
   The Analysis Service requests the sentiment analysis from the AI Model.

6. **Return Analysis Results:**
   The AI Model returns the sentiment analysis results to the Analysis Service.
   The Analysis Service stores the analysis results in the Database.

7. **Send Analysis Status:**
   The Analysis Service sends the status and results of the analysis back to the Web App via the API Gateway.

8. **Display Sentiment Results:**
   The Web App displays the sentiment analysis results to the User.

3.6.5 User interface - navigational paths and screen mock-ups

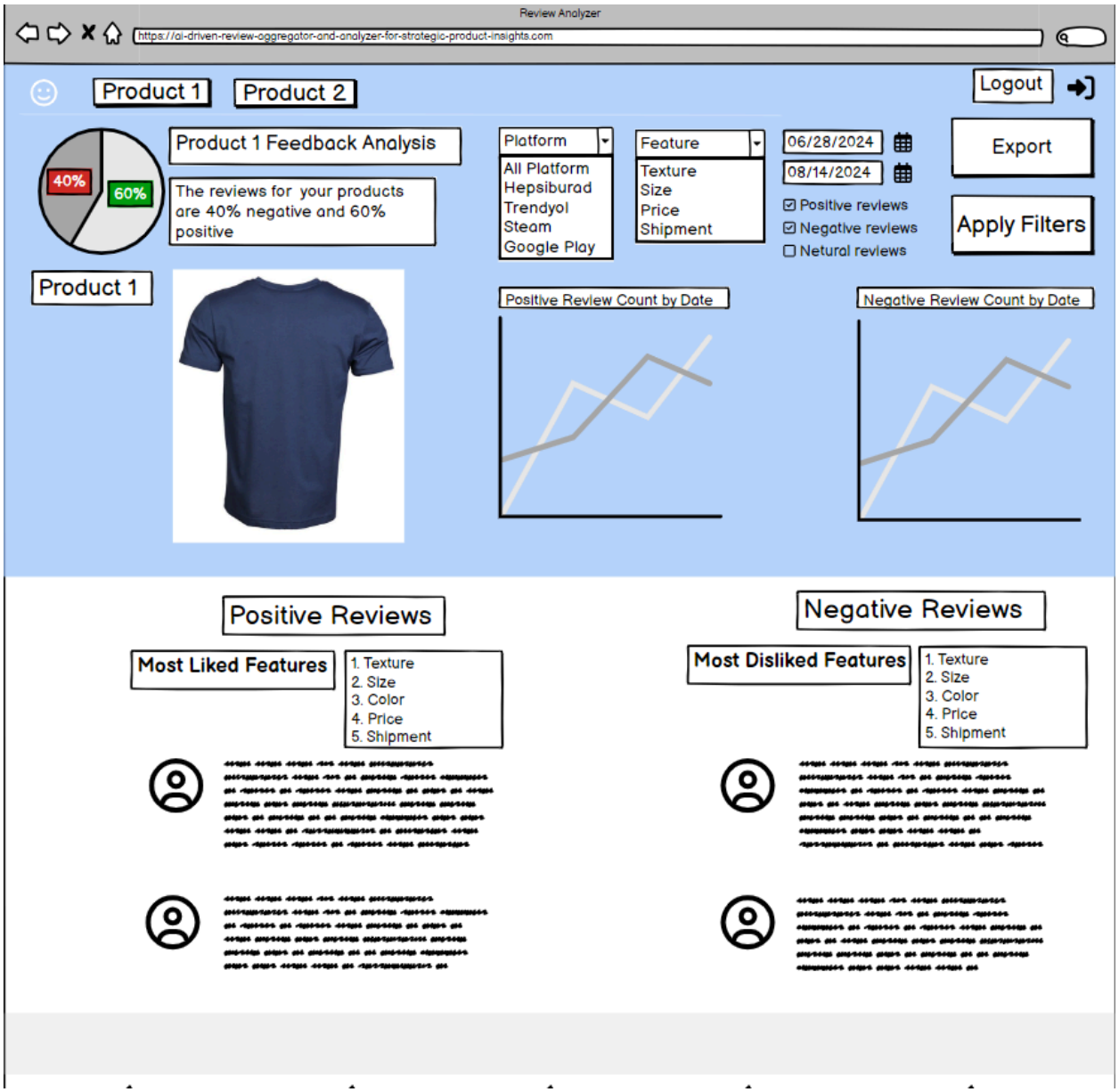

Figure 5: Screen mock-up design of the User Dashboard

The Figure 5 is a screen mockup, representing the user dashboard of the system. It showcases an overview of product feedback analysis. This interface is designed to help product owners and businesses to gain valuable insights into customer reviews, both positive and negative, across multiple platforms. Below is a detailed explanation of each section and element within the mockup:

Header Section

1. **Logout Button**
   The logout button allows users to securely log out of their accounts, positioned in the top right corner.

2. **Product Tabs**
   There are tabs labeled “Product 1”, and “Product 2”. They allow users to quickly switch between their products to view their respective reviews and analysis.

Main Analysis Area

1. **Product Feedback Analysis Summary:** In the center of the screen there is an area which displays an overall summary of the customer reviews. There is a Pie Chart, showing the percentage of the positive and negative reviews and a summary text that briefly summarizes the sentiment analysis.
2. **Product Image:** There is an image of the currently being analyzed product (e.g., an image of a T-shirt for "Product 1").

3. **Filters Panel**
   - **Platform Dropdown:** Allows users to filter reviews by specific platforms (e.g., Hespiburada, Trendyol, Steam, Google Play).
   - **Feature Dropdown:** Enables users to filter reviews based on specific features (e.g., Texture, Size, Price, Shipment).
   - **Date Range Selector:** Users can select a date range to filter reviews within specific time periods.
   - **Review Type Checkboxes:** Options to filter by review type (Positive reviews, Negative reviews, Neutral reviews).
   - **Apply Filters Button:** Applies the selected filters to update the review data displayed.
   - **Export Button:** Allows users to export the filtered data into formats such as XLS or PDF.

4. **Review Count Charts:**
   - **Positive Review Count by Date:** A line chart showing the trend of positive reviews over time.
   - **Negative Review Count by Date:** A line chart showing the trend of negative reviews over time.

Review Insights:

1. **Positive Reviews Section:**

    - **Most Liked Features:** Lists the top features that customers liked the most (e.g., Texture, Size, Color, Price, Shipment).
    - **Positive Review Examples:** Displays excerpts from positive reviews, along with user icons indicating individual reviewers.

2. **Negative Reviews Section:**

    - **Most Disliked Features:** Lists the top features that customers disliked the most (e.g., Texture, Size, Color, Price, Shipment).
    - **Negative Review Examples:** Displays excerpts from negative reviews, along with user icons indicating individual reviewers.

4. Glossary

AI[1]: Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems.
Emoji[2]: Emojis serve as a visual representation of human emotions, living beings, objects, and even certain symbols. These are in widespread use across the internet such as in text messaging, social media platforms, and any informal modes of communication.
Hepsiburada[3]: Hepsiburada.com is a Turkish e-commerce company in business since 1998.
Trendyol[4]: Trendyol is an e-commerce platform based in Turkey, with its headquarters in Istanbul.
LLM[5]: A large language model (LLM) is a deep learning algorithm that can perform a variety of natural language processing (NLP) tasks.
Structured Data[6]: Structured data — typically categorized as quantitative data — is highly organized and easily decipherable by machine learning algorithms.
Unstructured Data[7]: Unstructured data, typically categorized as qualitative data, cannot be processed and analyzed via conventional data tools and methods.
Google Play Store[8]: Google Play, also known as the Google Play Store or Play Store and formerly Android Market, is a digital distribution service operated and developed by Google.
Steam[9]: Steam is a video game digital distribution service and storefront developed by Valve Corporation.
Web Scraping[10]: Web scraping is an automatic method to obtain large amounts of data from websites.
User Interface(UI)[11]: The user interface is the point at which human users interact with a computer, website or application.
(WCAG) 2.1[12]: Web Content Accessibility Guidelines (WCAG) 2.1 defines how to make Web content more accessible to people with disabilities. Accessibility involves a wide range of

disabilities, including visual, auditory, physical, speech, cognitive, language, learning, and neurological disabilities.
FAQs[13]: A document (as on a website) that provides answers to a list of typical questions that users might ask regarding a particular subject.

5. References

[1]Rouse, M. (n.d.). AI (Artificial Intelligence). TechTarget. Retrieved from https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence

[2]Gupta, A. (2022, December 9). Emoji meanings: 25+ symbols explained with examples. 91mobiles. Retrieved from https://www.91mobiles.com/hub/emoji-meanings/#:~:text=Emojis%20serve%20as%20a%20visual,any%20informal%20modes%20of%20communication

[3]Hepsiburada. (n.d.). In Wikipedia. Retrieved from https://tr.wikipedia.org/wiki/Hepsiburada

[4]Trendyol. (n.d.). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Trendyol

[5]Elastic. (n.d.). What are large language models? Retrieved from https://www.elastic.co/what-is/large-language-models

[6]IBM. (2020, December 15). Structured vs. unstructured data: What's the difference? Retrieved from https://www.ibm.com/blog/structured-vs-unstructured-data/

[7]BM. (2020, December 15). Structured vs. unstructured data: What's the difference? Retrieved from https://www.ibm.com/blog/structured-vs-unstructured-data/

[8]Steam (service). (n.d.). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Steam_(service)

[9]Google Play. (n.d.). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Google_Play

[10]Anand, S. (n.d.). What is web scraping and how to use it? GeeksforGeeks. Retrieved from https://www.geeksforgeeks.org/what-is-web-scraping-and-how-to-use-it/

[11]Indeed Editorial Team. (2021, February 22). What is user interface? Indeed Career Guide. Retrieved from https://www.indeed.com/career-advice/career-development/user-interface

[12]Web Content Accessibility Guidelines (WCAG) 2.1. (2018, June 5). World Wide Web Consortium (W3C). Retrieved from https://www.w3.org/TR/WCAG21/#:~:text=Web%20Content%20Accessibility%20Guidelines%20

(WCAG)%202.1%20defines%20how%20to%20make,%2C%20learning%2C%20and%20neurological%20disabilities

[13]Merriam-Webster. (n.d.). FAQ. In Merriam-Webster.com dictionary. Retrieved from https://www.merriam-webster.com/dictionary/FAQ