

Date of version 23/11/2025

Digital Object Identifier

# Aspect-Based Sentiment Analysis for Turkish E-Commerce Reviews Using BERTurk

Ahmet Furkan Faideci<sup>1</sup>, Hozan Baydu<sup>2</sup>, and Serkan Eren, <sup>3</sup>,

<sup>1,2,3</sup> Artificial Intelligence Graduate Program, Department of Computer Engineering, Faculty of Engineering, Gebze Technical University, Kocaeli, Türkiye

a.faideci2025@gtu.edu.tr, h.baydu2024@gtu.edu.tr, s.erens2018@gtu.edu.tr

This work was conducted as part of the CSE 684: Natural Language Processing course in the Master's Degree Program at Gebze Technical University.

**ABSTRACT** This project presents a Turkish Aspect-Based Sentiment Analysis (ABSA) framework developed on a large-scale e-commerce review dataset collected from the Hepsiburada platform. Although ABSA has been well explored in English, existing Turkish sentiment models remain limited in their ability to handle multi-aspect feedback. A total of 54,416 customer reviews were gathered from 507 products across 10 major categories. From this large-scale corpus, an unsupervised aspect discovery method using Sentence-BERT and K-Means clustering was employed to identify seven latent aspect categories, including *Price/Performance*, *Delivery*, and *Quality*. Subsequently, a Gold Standard subset of 2,000 reviews was expertly annotated to train the proposed system. The model architecture is based on fine-tuning *BERTurk* using a Multi-Task Learning (MTL) framework. Unlike traditional multi-label approaches, the model utilizes a shared encoder with seven independent classification heads to simultaneously predict the specific sentiment polarity (Positive, Negative, or Neutral/Absent) for each aspect. The system produces structured outputs such as "Price": Positive", "Quality": Negative" for a single review. The overall goal is to contribute a robust, reproducible ABSA pipeline and high-quality benchmark resources for Turkish e-commerce analytics.

**INDEX TERMS** Aspect-Based Sentiment Analysis, Turkish NLP, BERTurk, Multi-Label Classification, E-Commerce Reviews , Hepsiburada

## I. INTRODUCTION

THE increasing use of e-commerce platforms has led to an exponential growth in user generated product reviews, which serve as a critical source of information for mining customer preferences and operationalizing feedback [1]. These reviews contain granular opinions on diverse product attributes, such as quality, logistics, pricing, and usability, making them indispensable for Natural Language Processing (NLP) tasks, including sentiment analysis and opinion mining. In Türkiye, Hepsiburada exemplifies this scale as a leading marketplace; as of late 2024, the platform reported a Net Promoter Score (NPS) of 74 and served over 12 million active customers, driven by initiatives such as its Premium membership program which reached 3.7 million users [2].

While traditional sentiment analysis typically assigns a single scalar polarity (positive, negative, or neutral) to an entire document, such document level approaches fail to capture the complexity of customer feedback where conflicting opinions coexist (e.g., "fast shipping" but "poor quality"). Aspect-Based Sentiment Analysis (ABSA) addresses this limitation

by identifying specific aspect terms and determining their associated sentiment polarities. ABSA has been rigorously standardized in English, largely driven by benchmark datasets from SemEval shared tasks [3]. Recent surveys [4], [5] classify these objectives into "Single ABSA" tasks (e.g., aspect extraction) and "Compound ABSA" tasks (e.g., joint aspect-sentiment extraction), highlighting the necessity of the latter for actionable business intelligence.

In contrast, ABSA research for the Turkish language remains under resourced. Existing studies have predominantly focused on document level sentiment classification using classical machine learning or lexicon based methods [6]. While the landscape has improved with the release of large-scale benchmarks like TrSAV1 [7], which provides 150,000 labeled reviews, comprehensive resources for fine-grained token level aspect annotation remain scarce. Furthermore, the agglutinative morphology of Turkish poses unique challenges for aspect extraction, creating a clear need for deep learning methodologies capable of capturing long distance syntactic dependencies.

To address this gap, this study proposes a transformer-based ABSA framework tailored for Turkish e-commerce text. The methodology leverages a dataset of 54,416 reviews collected from Hepsiburada across 10 major product categories, ensuring broad domain coverage from electronics to cosmetics. A detailed account of the dataset construction and annotation process is presented in Section III.

Given the proven efficacy of pre-trained language models, the proposed approach employs BERTurk [8] as the backbone encoder. Recent empirical studies [9] demonstrate that transformer-based pipelines significantly outperform traditional architectures in identifying aspect categories within Turkish multi-domain reviews. Building on this foundation, this research aims to design a robust methodology for extracting aspect-sentiment pairs, contributing both a structured dataset and a reproducible ABSA pipeline to the Turkish NLP community.

The remaining sections of this paper describe related work (Section II), the dataset characteristics (Section III), the proposed methodology (Section IV), and the preliminary findings (Section V).

## II. RELATED WORK

Sentiment analysis, also known as opinion mining, has established itself as a fundamental task in Natural Language Processing (NLP), enabling the automated extraction of subjective information from user-generated content [1]. While traditional approaches successfully categorize text at the document or sentence level, they often lack the granularity required to disentangle conflicting opinions regarding different attributes of a single entity (e.g., a product with a "great screen" but "poor battery").

### A. EVOLUTION OF ASPECT-BASED SENTIMENT ANALYSIS

To address the limitations of coarse-grained analysis, Aspect-Based Sentiment Analysis (ABSA) focuses on identifying specific aspect terms and resolving their associated sentiment polarities. Early research relied heavily on feature engineering and static sentiment lexicons, but the field has since shifted toward deep learning architectures. As detailed in recent comprehensive surveys by Zhang et al. [4] and Hua et al. [5], modern ABSA is categorized into "Single ABSA" tasks (e.g., Aspect Term Extraction) and "Compound ABSA" tasks (e.g., Aspect-Sentiment Triplet Extraction). The latter, which solves extraction and classification jointly, is currently considered the state-of-the-art for minimizing error propagation.

### B. TURKISH SENTIMENT ANALYSIS LANDSCAPE

Unlike English, which benefits from massive annotated corpora and established benchmarks like the SemEval Shared Tasks [3], research on Turkish sentiment analysis faces significant resource constraints. Early studies on Turkish predominantly utilized morphological analysis combined with classical machine learning algorithms such as Naïve Bayes

and Support Vector Machines (SVM) to predict document-level polarity [6].

A major bottleneck has been the scarcity of large-scale, publicly available datasets. While the SemEval-2016 Task 5 provided a gold-standard dataset for Turkish ABSA, it is limited to the restaurant domain and contains fewer than 2,000 sentences, restricting its utility for training deep neural networks from scratch. More recently, Aydoğan and Kocaman introduced TrSAV1 [7], a massive benchmark comprising 150,000 e-commerce reviews. However, TrSAV1 is annotated only for document-level sentiment (Positive/Neutral/Negative), leaving a gap for large-scale, token-level aspect annotations required for fine-grained ABSA.

### C. TRANSFORMER-BASED APPROACHES IN TURKISH

The advent of Transformer architectures has revolutionized Turkish NLP. The release of BERTurk [8], a BERT model pre-trained specifically on a large Turkish corpus, addressed many challenges posed by the language's agglutinative morphology. Recent studies have begun to leverage these models for ABSA. For instance, Bilgin and Turan [9] demonstrated that fine-tuning BERTurk on sequential classification tasks significantly outperforms traditional lexicon-based methods in extracting aspect categories. Similarly, other recent works have explored pipeline approaches that first extract targets and then classify sentiments [10].

Despite these advancements, most existing Turkish ABSA frameworks either rely on small, single-domain datasets (like restaurants) or treat aspect extraction and sentiment classification as isolated pipeline steps. There remains a distinct lack of a unified, multi-domain framework trained on a large-scale corpus that can handle the compound task of extracting aspect-sentiment pairs simultaneously. This study aims to bridge this gap by introducing a broad-coverage dataset and a transformer based methodology tailored for the Turkish e-commerce context.

## III. DATASET CONSTRUCTION

This study establishes a large-scale, multi-domain corpus of Turkish e-commerce product reviews sourced from Hepsiburada, one of the dominant online marketplaces in Türkiye. As of late 2024, Hepsiburada hosts over 200 million Stock Keeping Units (SKUs) and serves approximately 12 million active customers [2], making it a representative source for analyzing Turkish consumer sentiment.

The primary objective of the dataset construction was to mitigate the domain-dependency and polarity imbalance often observed in raw web-scraped corpora. The final curated dataset consists of 54,416 user-generated reviews collected from 507 products across 10 major categories, specifically structured to support Aspect-Based Sentiment Analysis (ABSA) and cross-domain generalization.

### A. CATEGORY SELECTION

To ensure broad semantic coverage, the data collection process targeted eleven high-volume product categories. Re-

views were successfully harvested from the following ten categories, selected to represent varying levels of subjectivity and aspect complexity (e.g., "technical specs" in Electronics vs. "aesthetic preferences" in Fashion):

- Home Improvement
- Fashion (Clothing & Footwear)
- Sports & Fitness Equipment
- Cosmetics
- Supermarket
- Books
- Kitchenware
- Small Home Appliances
- Auto Accessories
- Mother, Baby & Toys

*Note on Exclusions:* The **Furniture** category was excluded from the final dataset. During the extraction phase, significant inconsistencies were observed in the Document Object Model (DOM) structure of furniture product pages—likely due to dynamic content loading for variable product attributes (dimensions, fabrics)—which compromised the reliability of the automated scraper.

### B. PRODUCT SAMPLING AND POLARITY BALANCING

To minimize selection bias, a stratified sampling approach was employed based on the platform's "Best Sellers" ranking algorithm. For each category, the top 60 products were targeted by scanning the first three results pages. Out of the initial pool of 516 targeted products, 507 were successfully processed.

A critical challenge in e-commerce sentiment analysis is the "J-shaped" distribution of user ratings, where reviews are overwhelmingly skewed toward positive (5-star) values [11]. To address this, we implemented a quota-based sampling strategy:

- We filtered reviews by star rating (1 to 5).
- A cap of 100 reviews per rating level per product was applied.
- This ensured a balanced distribution (uniform polarity), allowing the model to learn from negative and neutral feedback effectively, which is often underrepresented in random sampling.

### C. DATA PREPROCESSING

Following extraction, the raw text data underwent a standardized preprocessing pipeline to ensure quality and consistency (see Algorithm 1). **Crucially, unlike traditional pipelines that lowercase all text, we adopted a case-preserving strategy** to ensure compatibility with the *cased* BERT model used in the methodology, which relies on capitalization for entity recognition (e.g., distinguishing special brand names).

The cleaning steps involved:

- 1) **Deduplication:** Duplicate entries were removed based on the review text to prevent data leakage between train/test splits.

- 2) **Null Handling:** Rows with missing or empty review texts were dropped.

---

### Algorithm 1 Data Construction and Preprocessing Pipeline

```

Require:  $U_{categories}$ : List of Hepsiburada category URLs
Require:  $Q_{star}$ : Quota per star rating (e.g., 100)
Ensure:  $D_{final}$ : Balanced and preprocessed dataset
1:  $D_{raw} \leftarrow \emptyset$ 
2: for all category  $\in U_{categories}$  do
3:    $P_{list} \leftarrow \text{ScrapeBestSellers}(category, \text{depth} = 3 \text{ pages})$ 
4:   for all product  $\in P_{list}$  do
5:     for star  $\leftarrow 1$  to 5 do
6:        $R_{batch} \leftarrow \text{GetReviews}(product, star)$ 
7:        $\triangleright$  Apply quota to balance sentiment distribution
8:       if Count( $R_{batch}$ )  $> Q_{star}$  then
9:          $R_{batch} \leftarrow \text{RandomSample}(R_{batch}, Q_{star})$ 
10:      end if
11:       $D_{raw} \leftarrow D_{raw} \cup R_{batch}$ 
12:    end for
13:  end for
14: end for
15: Preprocessing Phase:
16: for all review  $\in D_{raw}$  do
17:    $t \leftarrow \text{review.text}$ 
18:    $t \leftarrow \text{RemoveDuplicates}(t)$ 
19:    $\triangleright$  Preserve Turkish characters and Case for BERTTurk
20:   if Length( $t$ )  $< 2$  sentence then
21:     Discard review
22:   else
23:      $D_{final}.\text{add}(t)$ 
24:   end if
25: end for
return  $D_{final}$ 

```

---

### D. FINAL DATASET CHARACTERISTICS

The post-processed corpus contains **54,416** reviews. Unlike single-domain datasets (e.g., movie reviews), this corpus spans diverse semantic fields—from the durability of "Auto Accessories" to the texture of "Cosmetics". From this large-scale pool, a subset of 5,000 reviews was used for unsupervised aspect discovery, and a Gold Standard subset of 2,000 reviews was manually annotated for the supervised training phase described in the next section.

**TABLE 1. Summary of Final Dataset Statistics**

Metric	Count
Total Reviews	54,416
Unique Products	507
Categories	10
Discovery Subset	5,000
Annotated Training Subset	2,000

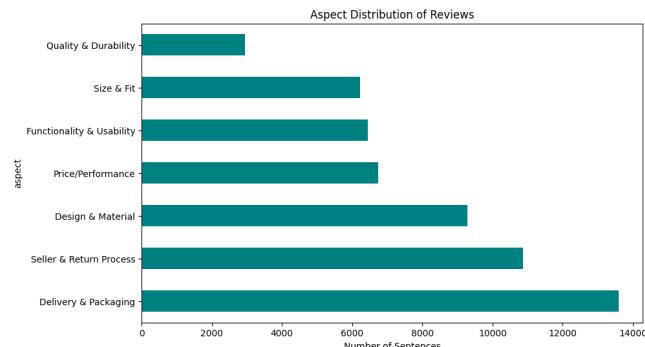
## IV. METHODOLOGY

The proposed methodology adopts a hybrid approach that combines unsupervised clustering for aspect discovery with a supervised Multi-Task Learning (MTL) framework for sentiment prediction. The system is built upon *BERTurk* (dbmdz/bert-base-turkish-cased), a state-of-the-art language model optimized for Turkish, fine-tuned to simultaneously predict sentiment polarities for seven distinct product aspects.

### A. UNSUPERVISED ASPECT EXTRACTION AND CLUSTERING

Data Preprocessing and Granularity for this intermediate phase of the study, a subset of 5,000 raw customer reviews was utilized to establish the aspect extraction pipeline. A critical initial step in the methodology was sentence segmentation. Since a single consumer review often contains multiple distinct opinions (e.g., "The delivery was fast, but the screen quality is poor"), analyzing the review as a whole introduces noise into the clustering process. Therefore, we utilized the NLTK library to tokenize reviews into individual sentences, resulting in a more granular dataset where each data point represents a single semantic unit. Semantic Representation and Dimensional Reduction To capture the semantic meaning of the sentences beyond mere keyword matching, we employed the Sentence-BERT (SBERT) architecture. Specifically, the paraphrase-multilingual-MiniLM-L12-v2 model was selected for its robustness in handling the Turkish language. This model mapped each sentence into a high-dimensional vector space. Subsequently, Principal Component Analysis (PCA) was applied to reduce these embeddings to 50 dimensions. This dimensional reduction step was essential to mitigate the "curse of dimensionality" and enhance the performance of the subsequent K-Means clustering algorithm by filtering out noise while retaining variance. Unsupervised Clustering and Keyword Extraction The core of the aspect discovery process relied on the K-Means clustering algorithm, initialized with  $k = 10$  clusters to discover latent topics within the corpus. To interpret the semantic focus of each cluster, we utilized a class-based TF-IDF approach. A custom stopword list was rigorously curated to exclude general sentiment adjectives (e.g., "good," "bad," "beautiful") and generic verbs. This ensured that the highest-ranking keywords for each cluster represented specific aspects (e.g., "cargo," "price," "fabric," "battery") rather than the sentiment associated with them. Aspect Mapping and Distribution (Figure 1) Following the clustering process, the 10 generated clusters were manually inspected and mapped to 7 distinct aspect categories to create a coherent taxonomy. For instance, clusters sharing semantic proximity, such as those related to "shipping" and "packaging," were merged into a single label.

As illustrated in Figure 1, the distribution of these aspects reveals that "Seller & Return Process" (Satıcı ve İade Süreci) and "Delivery & Packaging" (Teslimat ve Paketleme) are the most frequently discussed topics in this datasets. This unsupervised approach successfully allowed for the automated



**FIGURE 1. Aspect Distribution.**

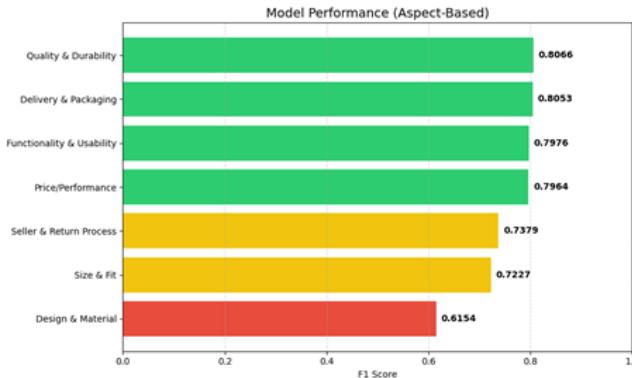
categorization of unlabeled text into specific domains such as "Price/Performance," "Design & Material," and "Quality & Durability" without the need for manual training data.

### B. SUPERVISED LEARNING AND MULTI-HEAD MODEL ARCHITECTURE IMPLEMENTATION

Datasets Annotation and Sentiment Schema following the unsupervised clustering phase, a specific subset of 2,000 customer reviews was selected to establish a "Gold Standard" datasets for the intermediate training phase. To enable Aspect-Based Sentiment Analysis (ABSA), each review within this datasets was manually annotated according to the seven aspect categories previously identified (e.g., Price/Performance, Delivery, Quality). A granular sentiment polarity schema was applied for each aspect using a numerical scale: -1 (Negative), 0 (Neutral or Not Mentioned), and 1 (Positive). This multi-label annotation approach ensures that a single review can carry different sentiment polarities for different aspects simultaneously (e.g., positive for delivery but negative for product quality).

**Model Architecture:** Multi-Head BERT To process this complex label structure, we designed a Multi-Head BERT architecture. We utilized the dbmdz/bert-base-turkish-cased (*BERTurk*) pre-trained model as the shared encoder to generate contextual word embedding suitable for the Turkish language. Unlike standard classification models that output a single prediction, our architecture features seven independent classification heads (linear layers) attached to the pooled output of the BERT encoder. This design allows the model to compute a separate loss and generate a distinct sentiment prediction for each aspect concurrently during a single forward pass, significantly increasing computational efficiency.

**Performance Evaluation** (Figure 2) The datasets was split into training (80%), validation (10%), and test (10%) sets. The model achieved a General Macro-F1 score of 0.7546, which indicates robust performance for an intermediate dataset size of 2,000 samples. As illustrated in Figure 2, the model demonstrated superior performance in objective categories such as "Quality & Durability" (F1: 0.8066) and "Delivery & Packaging" (F1: 0.8053), where the terminology used by customers is often explicit and unambiguous. Con-



**FIGURE 2.** Aspect Based Model Performance.

versely, the "Design & Material" category yielded a lower score (F1: 0.6154). This disparity suggests that comments regarding design are inherently more subjective and context-dependent, making them harder for the model to generalize with a limited datasets. These results validate the efficacy of the Multi-Head BERT approach for Turkish ABSA tasks.

#### Algorithm 2 Unsupervised Aspect Category Discovery

**Require:**  $D_{raw}$ : Raw review corpus  
**Require:**  $K$ : Number of clusters (e.g., 10)  
**Require:**  $S_T$ : paraphrase-multilingual-MiniLM-L12-v2  
**Ensure:**  $A_{categories}$ : Identified aspect mapping

- 1: **Step 1: Sentence Segmentation**
- 2:  $S_{all} \leftarrow \emptyset$
- 3: **for all**  $review \in D_{raw}$  **do**
- 4:      $sentences \leftarrow \text{NLTK.tokenize}(review)$
- 5:     **for all**  $s \in sentences$  **do**
- 6:         **if** Length( $s$ ) > 2 words **then**
- 7:              $S_{all}.\text{append}(s)$
- 8:         **end if**
- 9:     **end for**
- 10: **end for**
- 11: **Step 2: Embedding Generation**
- 12:  $Model_{st} \leftarrow \text{SentenceTransformer}(S_T)$
- 13:  $V_{emb} \leftarrow Model_{st}.\text{encode}(S_{all})$   $\triangleright$  Generate dense vectors
- 14: **Step 3: Dimensionality Reduction & Clustering**
- 15:  $V_{reduced} \leftarrow \text{PCA}(V_{emb}, \text{components} = 50)$
- 16:  $Clusters \leftarrow \text{KMeans}(V_{reduced}, n\_clusters = K)$
- 17: **Step 4: Keyword Extraction (Labeling)**
- 18:  $A_{categories} \leftarrow \emptyset$
- 19: **for**  $k \leftarrow 0$  to  $K - 1$  **do**
- 20:      $Docs_k \leftarrow \{s \mid s \in S_{all} \text{ where } Clusters.\text{predict}(s) = k\}$
- 21:      $Keywords_k \leftarrow \text{TF-IDF}(Docs_k, \text{top\_n} = 8)$
- 22:      $AspectName \leftarrow \text{ManualMap}(Keywords_k)$   $\triangleright$  e.g., 'Kargo'  $\rightarrow$  'Delivery'
- 23:      $A_{categories}.\text{add}(k \rightarrow AspectName)$
- 24: **end for**
- 25:     **return**  $A_{categories}$

#### Algorithm 3 Multi-Task Aspect-Based Sentiment Analysis Training

**Require:**  $D_{train}$ : Dataset with 7 aspect labels per review  
**Require:**  $M_{bert}$ : Pre-trained bert-base-turkish-cased  
**Require:**  $A$ : Set of 7 Aspects (Price, Delivery, Quality, etc.)  
**Require:**  $C$ : Classes {0: Neg, 1: Neu, 2: Pos}

- 1: **Architecture Setup:**
- 2:  $Encoder \leftarrow M_{bert}$
- 3: **for**  $i \leftarrow 1$  to  $|A|$  **do**
- 4:      $Head_i \leftarrow \text{Linear}(768 \rightarrow 3)$   $\triangleright$  One classifier per aspect
- 5: **end for**
- 6:  $Optimizer \leftarrow \text{AdamW}(\text{lr} = 2e^{-5})$
- 7: **Training Loop:**
- 8: **for**  $epoch \leftarrow 1$  to  $E$  **do**
- 9:     **for all**  $(text, Y_{true}) \in D_{train}$  **do**
- 10:          $Tokens \leftarrow \text{Tokenizer}(text, \text{max\_len} = 128)$
- 11:          $H_{last} \leftarrow Encoder(Tokens).\text{last\_hidden\_state}$
- 12:          $v_{cls} \leftarrow \text{Dropout}(H_{last}[:, 0, :])$   $\triangleright$  Pool token
- 13:          $L_{total} \leftarrow 0$
- 14:         **for**  $i \leftarrow 1$  to  $|A|$  **do**
- 15:              $logits_i \leftarrow Head_i(v_{cls})$   $\triangleright$  Shape:  $logits_i \leftarrow Head_i(v_{cls})$
- 16:              $loss_i \leftarrow \text{CrossEntropy}(logits_i, Y_{true}[:, i])$
- 17:              $L_{total} \leftarrow L_{total} + loss_i$
- 18:         **end for**
- 19:          $L_{total}.\text{backward}()$
- 20:          $Optimizer.\text{step}()$
- 21:     **end for**
- 22:     **end for**
- 23: **Inference Function:**
- 24: **Function Predict( $text$ )**
- 25:      $v_{cls} \leftarrow \text{Encode}(text)$
- 26:      $Predictions \leftarrow \{\}$
- 27:     **for**  $i \leftarrow 1$  to  $|A|$  **do**
- 28:          $class\_id \leftarrow \text{ArgMax}(Head_i(v_{cls}))$
- 29:         **if**  $class\_id \neq 1$  **then**  $\triangleright$  Ignore Neutral/None
- 30:              $Predictions[A_i] \leftarrow C[class\_id]$
- 31:         **end if**
- 32: **end for** **return**  $Predictions$
- 33: **End Function**

The architecture is defined as follows:

- **Shared Encoder Layer:** The input sequence  $X = \{x_1, x_2, \dots, x_n\}$  is fed into the pre-trained BERT encoder (dbmdz/bert-base-turkish-cased). This shared layer captures the global semantic context and produces contextualized embeddings  $H \in \mathbb{R}^{n \times d}$ , where  $d = 768$ .
- **Pooling:** The final hidden state of the special classification token is extracted ( $h_{CLS} \in \mathbb{R}^{768}$ ) to serve as the aggregate representation of the entire review. A dropout layer ( $p = 0.1$ ) is applied to  $h_{CLS}$  to prevent overfitting.
- **Multi-Head Classification Layer:** The pooled representation  $h_{CLS}$  is simultaneously fed into  $K = 7$  independent classification heads (linear layers), where  $K$  corresponds to the number of discovered aspect categories

(e.g., Delivery, Price, Quality). Each head  $k$  projects the shared embedding to an aspect-specific logit vector  $z_k \in \mathbb{R}^3$ .

- **Output & Activation:** Each head is responsible for a 3-class classification task. The final prediction for aspect  $k$  is obtained by applying the **argmax** operation over the logits:

$$\hat{y}_k = \text{argmax}(W_k h_{CLS} + b_k) \quad (1)$$

where  $\hat{y}_k \in \text{Table2}$ . This allows the model to disentangle conflicting sentiments (e.g., Positive for *Delivery* but Negative for *Quality*) within the same sentence.

**TABLE 2.**

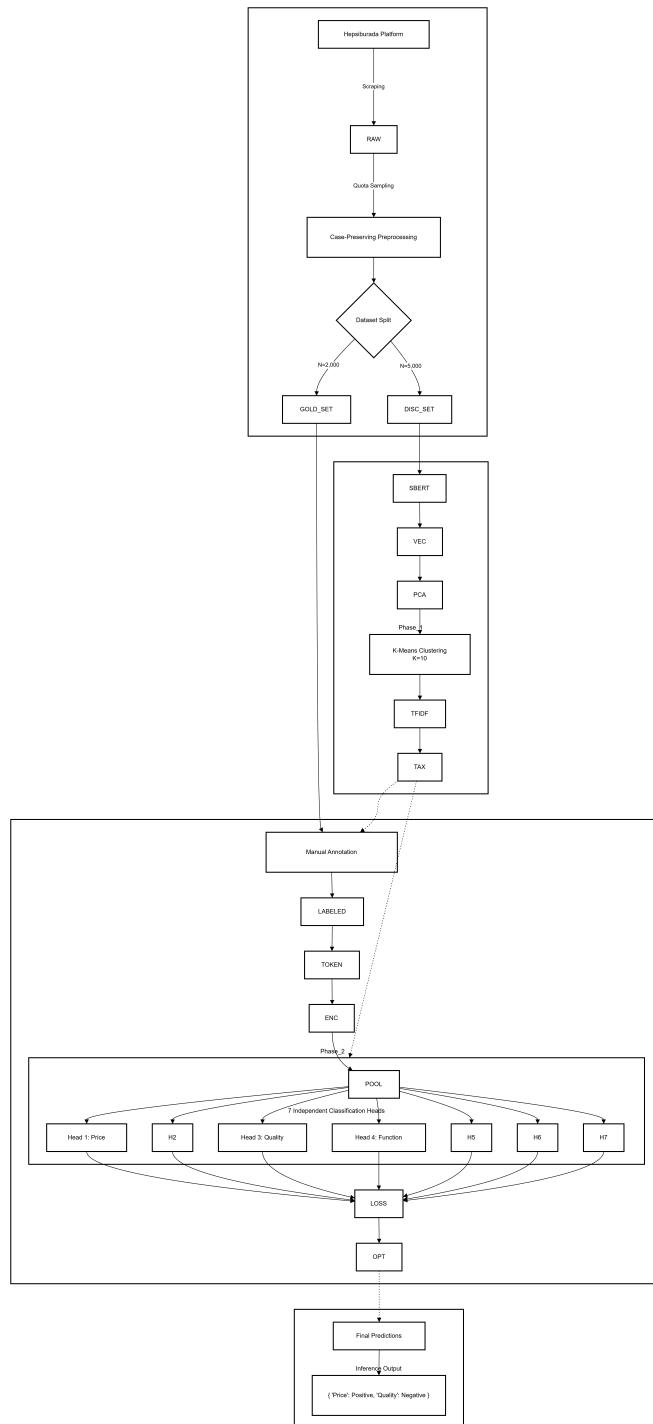
Sentiment	Value
Negative	0
Neutral/Absent	1
Positive	2

## V. CONCLUSION

This study has established a foundational framework for Aspect-Based Sentiment Analysis (ABSA) in the Turkish e-commerce domain, addressing the critical scarcity of multi-domain resources. By constructing a large-scale corpus of 54,416 reviews and distilling a **Gold Standard** subset for supervised training, we successfully pioneered a data-driven approach to aspect taxonomy. Utilizing unsupervised clustering, we identified seven latent aspect categories—ranging from *Price/Performance* to *Seller Communication*—without relying on rigid, pre-defined ontologies.

The proposed **Multi-Task Learning (MTL)** architecture, built upon the *BERTurk* encoder, has demonstrated the viability of simultaneous sentiment prediction, achieving a General Macro-F1 score of **0.7546**. This confirms that a shared encoder with independent classification heads can effectively disentangle conflicting sentiments (e.g., positive delivery vs. negative quality) within a single text, capturing granular nuances that traditional document-level models overlook.

However, this research represents the initial phase of a broader investigation. While the current results are promising, challenges remain in generalizing to highly subjective categories like *Design*. Future phases of this project will focus on three key areas: (1) expanding the annotated corpus to include implicit aspect expressions, (2) benchmarking the current BERT-based pipeline against generative Large Language Models (LLMs) to enhance reasoning capabilities in subjective contexts, and (3) refining the joint loss function to better mitigate the class imbalance between the dominant “Neutral/Absent” class and active sentiment labels. Ultimately, these continuing efforts aim to evolve the current prototype into a fully deployable, a for actionable customer insight generation in the Turkish market.



**FIGURE 3.** Proposed *BERTurk*-based \*\*Multi-Task\*\* ABSA architecture with joint aspect-sentiment detection.

**APPENDIX. A**

The complete implementation and experimental resources are provided in the project's GitHub Repository . Collected Data are stored in the Kaggle .

**REFERENCES**

- [1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.
- [2] D-MARKET Electronic Services & Trading, "Hepsiburada announces fourth quarter and full year 2024 financial results," <https://investors.hepsiburada.com>, 2025, accessed: Nov 2025.
- [3] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 2016, pp. 19–30.
- [4] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, "A survey on aspect-based sentiment analysis: Tasks, methods, and challenges," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11 019–11 038, 2023.
- [5] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," *Artificial Intelligence Review*, vol. 57, no. 11, pp. 1–50, 2024.
- [6] M. Demircan and O. Tuna, "Developing turkish sentiment analysis models using various machine learning techniques," *Machine Learning with Applications*, vol. 6, p. 100163, 2021.
- [7] M. Aydoğan and V. Kocaman, "Trsavl1: A new benchmark dataset for classifying user reviews on turkish e-commerce websites," *Journal of Information Science*, vol. 49, no. 6, pp. 1711–1725, 2023, first published online in 2022.
- [8] S. Schweter, "Berturk: Bert models for turkish," Zenodo, 2020, doi:10.5281/zenodo.3770924.
- [9] M. Bilgin and M. Turan, "Aspect-based sentiment analysis for turkish reviews using token and sequential classification methods," *International Journal of Computer and Information Technology*, vol. 14, no. 3, pp. 90–99, 2025.
- [10] S. Özkan *et al.*, "Goal-based sentiment analysis on turkish smartphone reviews," in *International Conference on Artificial Intelligence and Data Processing (IDAP)*, 2022.
- [11] N. Hu, P. A. Pavlou, and J. Zhang, "Overcoming the j-shaped distribution of product reviews," *Communications of the ACM*, vol. 52, no. 10, pp. 144–147, 2009.
- [12] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.
- [13] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations (ICLR)*, 2019.
- [14] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [15] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 2019.
- [16] T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38–45.



**AHMET FURKAN FAIDEKİ** received his B.Sc. degree in Electrical Engineering from Istanbul Technical University, Türkiye, in 2021. He is currently pursuing an M.Sc. degree in Artificial Intelligence at Gebze Technical University. He works as a Senior Embedded Software Engineer at FEV Türkiye, where he focuses on automotive software development, model-based design, and intelligent automation for embedded systems. His professional interests include artificial intelligence, deep learning, and intelligent control systems in automotive applications.



**HOZAN BAYDU** received his B.Sc. degree in Electrical Engineering from Yıldız Technical University, İstanbul, Türkiye, in 2022, and is currently pursuing an M.Sc. degree in Artificial Intelligence at Gebze Technical University. He worked as a Machine Learning Engineer at ADATECH, where he focuses on model development, data processing, and optimization for intelligent systems. His research interests include natural language processing, deep learning, and applied artificial intelligence.



**SERKAN EREN** received his B.Sc. degree in Electronics Engineering from Gebze Technical University, Türkiye, in 2023, and is currently pursuing an M.Sc. degree in Artificial Intelligence at the same institution. He works as an Automotive Software Engineer at FEV Türkiye, contributing to Python-based automation, DevOps pipelines, and embedded software integration. His professional interests include deep learning, software engineering for intelligent systems, and the development of data-driven applications for automation and natural language processing.