National College of Ireland

Dynamic ABSA for E-commerce Customer Feedback Optimisation

MSc Research Project
Data Analytics

Jenat Jiffna Rodrigues
Student ID: x23337451

School of Computing
National College of Ireland

Supervisor: Bharat Agarwal

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ..... Jenat Jiffna Rodrigues.....

Student ID: ..... x23337451.....

Programme:..... MSc Data Analytics..... **Year:** ...2025-26.....

Module: ..... Research Project .....

Supervisor: ..... Bharat Agarwal.....

Submission Due Date: .....28/01/2026.....

Project Title: .....Dynamic ABSA for E-commerce Customer Feedback Optimisation.....

Word Count: .....5600..... **Page Count**...18.....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ..... Jenat Jiffna Rodrigue.....

Date: .....28/01/2026.....

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Dynamic ABSA for E-commerce Customer Feedback Optimisation

Jenat Jiffna Rodrigues
x23337451

Abstract

The blistering development of e-commerce has produced enormous amounts of customer reviews, and overall ratings are insufficient to provide comparative sentiment analysis. Conventional sentiment analysis does not attempt to acknowledge opinions regarding particular elements of products, whereas most aspect-based sentiment analysis (ABSA) frameworks assume training on features of a set of predefined aspects and are therefore restricted in their potential to adapt to novel or emergent product characteristics. The study will suggest a completely dynamic and least supervised ABSA pipeline that can identify details in raw e-commerce reviews in a fully automatic way and categorise sentiments related to the study. The system trains DistilBERT on 34,660 Amazon customer reviews to create contextual sentence embeddings, uses PCA and UMAP to reduce dimensions and clusters the data with MiniBatchKMeans to identify coherent aspect groups and labels the data using a weighted combination of KeyBERT, SBERT semantic similarity, and corpus specificity. A narrow-focused DistilBERT model for sentiment (positive, negative, neutral) prediction is done by aspect. The best setting (K=9 clusters) was found to have silhouette score of 0.78, classification accuracy of 93.76 in sentiment and 1,000 reviews per second of real time inference at consumer grade hardware (Google Colab GPU). These were nine meaningful aspects found (e.g., Books, Bluetooth, Battery Life, Display) and actionable business insights were obtained based on the distributions of aspect-sentiment. The hybrid methodology proposed is much more accurate, interpretable, and deployable than the traditional TF-IDF and BERTopic baselines, and provides a scalable solution to real-time e-commerce feedback optimisation.

1 Introduction

Aspect-Based Sentiment Analysis has become an essential part of contemporary natural language processing, which helps businesses to derive fine-grained knowledge out of unstructured customer feedback in large quantities (Zhang et al., 2022). In comparison with the traditional sentiment analysis that marks an entire text using one polarity of emotion, ABSA defines particular characteristics of a product or a service and relates them to the particular sentiment (positive, negative, or neutral) (Wu et al., 2025). This is an actionable intelligence derived through this fine-grained fashion in case of e-commerce platforms where

millions of customer reviews are created every day. E-commerce is the most rapidly expanding retailing sector in the world, and more than 90 percent of Internet users also review and make purchases (Scaria et al., 2023). But, the way e-commerce sentiment systems are currently constructed is also subject to serious shortcomings, whereby the traditional sentiment analysis reduces subtle opinions to one word. To take an example, the review that says The battery life is great, display is too dark would not retain clear positive and negative feelings towards particular product attributes.

State-of-the-art ABSA models that are based on BERT and its variations demonstrate outstanding performance (more than 90% accuracy) when task aspects are specified via manual curation. Nevertheless, this monitored paradigm adds some inherent constraints to real-time e-commerce settings (Negi et al., 2024). The emergence of new product characteristics like a health tracking functional feature or a 5G connection can appear sooner than the manual process of annotation can sustain. Models that have been trained on fixed concepts have disastrous performance deterioration would frequently surpass 30 percent accuracy drops when they detect new characteristics in real-time review streams (Tulkels & van Cranenburgh, 2020). The literature evidences that BERT-based models do not work in dynamic unsupervised environments, whereas clustering-based models usually use fixed word representations, and thus, they are unable to retrieve contextual semantics (Islam et al., 2025). Recent surveys have affirmed that 89% of released ABSA models rely on supervised learning, which poses a methodological bottleneck, and recent research integrates deep learning and clustering is largely thinly stretched, and seldom has been applied to real-time deployment.

This study is driven by the merging of economic and technological forces. E-commerce enterprises are more likely to lose out in the race when they cannot react swiftly to new issues raised by customers (Wu et al., 2025). Lightweight transformer models like DistilBERT, preserving the performance of 97% of BERT with 66M parameters versus 110M, make it possible for companies to run their models on their resource-constrained machines, such as the free GPU on Google Colab (Putatunda, 2023). This study covers three understudied topics, namely, (1) discovering dynamic aspects without fixed vocabularies; (2) combining lightweight transformer embeddings with unsupervised clustering; (3) the viability of real-time deployment on consumer-grade hardware.

Research Question: How can a deep learning model automatically extract and classify aspects and sentiments from e-commerce customer reviews to provide actionable business intelligence?

Objectives:

1. In order to come up with a hybrid ABSA architecture that combines DistilBERT embeddings with K-means clustering to allow discovery of aspects dynamically and in a data-driven manner, in the wake of recent progresses in clustering-based aspect extraction (Scaria et al., 2023).
2. In order to produce interpretable labels of the aspects through the combination of the KeyBERT keyword extraction, noun-phrase recognition, and semantic-centroid analysis, it is important to ensure that every detected aspect describes a meaningful product feature (Negi et al., 2024).

3. To narrow-tune a DistilBERT sentiment model with a minimised accuracy of 94.30, which in turn could make it reliable in terms of positive, negative, and neutral sentiment classification in all revealed aspects.
4. To examine the scalability and robustness of the systems through experimenting with the entire ABSA pipeline on data sets of about 1000 to more than 34000 e-commerce reviews, to determine the consistency of the performance under different amounts of data.
5. To show the possibility of real-time deployment by applying and executing the entire ABSA pipeline on available cloud systems like Google Colab to ascertain its capability in practice for e-commerce applications.

Achievement of goals will be represented by stringent validation. The quality of aspect discovery will be determined by silhouette scores (target should be greater than 0.75) and by domain expert interpretation (Giannakopoulos et al., 2017). Accuracy, precision, recall, F1-score, and F2-score will be used to classify sentiments, where F2 puts more emphasis on recall (target ≥94.30%). Scalability testing needs less than 5% of the performance degradation with the scale of the datasets. The three-step approach: Phase 1: 768-dimensional DistilBERT embeddings are created; Phase 2: PCA, UMAP, and MiniBatchKMeans are used in clustering with labels available using weighted TF-IDF (45%), semantic centroid similarity (40%), corpus specificity (15%) (Tulkens and van (Cranenburgh, 2020) Phase 3: DistilBERT is further fine-tuned using Hugging Face Trainer with a learning rate of 2e-5 and batch size 16.

There are five sections in this report. Section 2 includes the critical literature review of ABSA approaches and advances. Section 3 specifies such research methods as system architecture, data preparation, and evaluation metrics. Section 4 is a full report of detailed results such as silhouette testing, aspect labels, and sentiment measures with accuracy of up to 93.76%. Findings, limitations, and future recommendations are discussed in section 5. The contribution of this research is new integration of deep embeddings and unsupervised clustering as well as the possibility to implement the proposed methods in real-time using consumer hardware.

2 Related Work

This part is a critical review of the major literature in the area of aspect-based sentiment analysis, reviewing the current approaches, their strengths, weaknesses, and the contribution that the current study makes to the field with its innovative combination of deep contextual embedding and unsupervised clustering.

2.1 Standard ABSA Techniques and their weaknesses

Early ABSA methods were based on lexicon-based, hand-crafted feature methods and conditional random fields (CRF) for aspect extraction (Luo et al., 2019). Although these approaches were interpretable, they were poor in scaling and modelling of multifaceted semantic relationships. These conventional methods presupposed predefined aspect vocabularies and hence were rigid in dynamic e-commerce settings where new product

features are being introduced on a regular basis. Recent general surveys prove that 89% of ABSA models continue to rely on supervised learning, which leaves a methodological bottleneck in the area. This inherent drawback of predefined aspects also exists in the latest supervised methods and negates real-time applicability.

2.2 ABSA Models that rely on Deep Learning and transformers

BERT and its variants (DistilBERT, RoBERTa), which give above 90 percent accuracy on benchmark datasets (Alhadlaq et al., 2024). These pre-trained language models learn subtle semantic links that are better than the fixed embeddings, such as GloVe or FastText. Transformer based methods, however, have the same inherent drawback of predetermined aspects - they are efficient in supervised data (with aspect lists present) but collapse (40% or more accuracy decrease) when exposed to new data. The extensive review by (Ahmad et al.. 2025) has revealed that although the BERT-based models are better than the previous ones, they are limited by the need to be annotated and a lack of the ability to discover new aspects dynamically.

2.3 Graph Neural Networks Syntactic and Semantic Modeling

Graph Convolutional Networks (GCN) and Graph Attention Networks (GAT) have become variants that are effective in processes of learning syntactic dependencies and semantic relationships in ABSA tasks (Zhao et al., 2024). These architectures built dependency trees and use part-of-speech tags, forming rich relational forms. Graph Convolutional Networks Structured Dependency Tree-based SDTGCN have state-of-the-art performance with improvements in F1-score of 0.74-1.17% over baselines. Strength: GNNs are able to learn the complex aspect-opinion connections on the basis of structured representations. Limitation: The majority of GNN-based methods are still pre-defined to have lists of aspects and pre-defined syntactic structures, which prevents adaptation to new aspects. Implicit aspect-level sentiment analysis based on the use of graph-enhanced T5 and GNNs (Li et al., 2025)has potential, but is still within the supervised paradigm.

2.4 Multi-Task Learning and Joint Extraction Approaches

The concept of multi-task learning frameworks concurrently extracts aspects, detects aspect categories, and classifies sentiment, and the frameworks are better performed when the tasks interact. Multi-Task Dual-Tree Networks (MTDN) combine constituency trees and dependency trees, and on single-task BERT-based models, improvement in F1-score by 0.68-3.54% are achieved. Strength: Joint modeling is able to model interdependence among subtasks. Limit: These methods rely on training data annotation that is still necessary to extract aspects and classify sentiment, and thus they share the drawbacks of supervised learning.

2.5 Attention Mechanisms and Long-range Dependency Modeling

The mechanisms of attention have changed the simple attention of LSTMs to interactive attention networks (IAN) and to multi-head co-attention mechanisms (Lawan et al, 2025).

More recent models such as Mamba Former combine State Space Models (Mamba) and transformers in order to learn dependencies on aspect and opinion words across a long distance that are not captured by standard attention. Strength: Advanced attention systems enhance the aspect- opinion correspondence. Limitations: Attention will be realized in the context of supervised classification; they cannot deal with absence of discovery of dynamic aspects based on a list that is not known in advance.

2.6 Generative and Timely-Based Approaches

Prompt-based approaches operate on the principle that ABSA is cloze-like prediction tasks that utilize cloze-based templates with knowledge in pre-trained language models (Feng et al., 2024). DSGP models incorporate dual syntactic information with prompt learning and demonstrate state-of-the-art performance. Advantage: Timely-based methods eliminate the use of far-reaching fine-tuning. Drawback: Manual creation of templates needs domain knowledge and currently generative models exist only within the scope of supervised classification.

2.7 Unsupervised Aspect Detection and Clustering Algorithms

The previously unsupervised methods based on Latent Dirichlet Allocation (LDA) and k-means clustering (He et al., 2020) may reveal aspects without the use of labeled data but used fixed embeddings, which restricted contextual interpretation. Unsupervised Neural Attention Models use k-means cluster centroids as aspect embeddings, although not combined with current transformer embeddings. Strength: The purely unsupervised methods do not need any annotation. Weakness: Static embeddings are not able to disclose context sensitive semantics; their performance to large scale e-commerce data is not validated.

2.8 Gap in Research and Proposed Contribution

Critical Gap Identified: The literature does not fully focus on the integration of (1) deep contextual transformer embeddings (DistilBERT) with (2) unsupervised clustering algorithms (K-means with UMAP dimensionality reduction) with the aim of (3) discovering dynamically changing aspects in e-commerce and assuming the deployment of these changes into practice in real-time. This study fills this gap by introducing a new hybrid architecture that combines the state of the art transformer efficiency with scalability without supervision, as evidenced by validation in the consumer-grade hardware of Google Colab. This work provides the opportunity to adapt to new product features in a very dynamic manner, unlike predefined-aspect-dominated literature and static-embedding-based clustering strategies (He et al., 2020) and has viable deployment constraints, which is the key limitation facing contemporary ABSA systems.

3 Research Methodology

In this section, a systematic approach to developing a dynamic Aspect-Based Sentiment Analysis (ABSA) model for e-commerce customer feedback is outlined. The three stages of

the methodology are: data preprocessing and embedding generation; unsupervised clustering and aspect labelling; and sentiment classification through fine-tuning as shown in Figure 1.

Figure 1: Overall methodological workflow.

3.1 Data Pre-processing

3.1.1 Data Loading and Preliminary Exploration

Before beginning to discuss data analysis, it is stated that internet usage among people has increased steadily, and now almost everyone uses the internet, regardless of their age. The study will use Amazon product review data that has 34,660 records having various attributes such as review text, rating, brand, category, and recommendation status. Primary data analysis showed that the columns are not homogeneous and have missing values that need to be pre-processed in a strategic manner. The data reflects the various product types that allow holistic comparison in the various e-commerce areas.

3.1.2 Column Selection

Column Selection Justification 7 original columns (reviews.text, reviews.title, reviews.rating, reviews.doRecommend, brand, categories, asins) were selected. The choice is not too comprehensive or too slow. Aspect extraction content is supplied by review and text title, sentiment mapping is supported by rating and categories ground content in multi-dimensional analysis. Other columns had duplicate or sparse data that took up memory space and did not add value to the analysis.

3.1.3 Text Cleaning

Raw text was systematically normalised using regex-based cleaning that eliminated HTML tags, special characters, and unneeded whitespace as well as converting to lowercase. The standardisation of this lowers the noise and guarantees that there is consistency in the generation of the embedding by removing the variations of the same term in orthographic form (e.g., battery-life vs battery life).

3.1.4 Brand Standardisation

Pattern matching was used to standardise brand names (e.g., Fire TV, Amazon Fire TV) into the standardised brand names of Amazon Fire TV. This type of categorisation allows sentiment analysis at the brand level and avoids the loss of brand-specific information due to fragmentation.

3.1.5 Combined Text Creation

There were 3 processed fields (brand, category, review text), which were then concatenated into combined_text unified strings in the following format: Brand: [brand]. Category: [category]. Review: [cleaned_text]". The integration offers contextual information that is necessary in accurate aspect extraction and also offers consistent semantic flow.

3.1.6 Sentiment Labelling

The ratings were coded into sentiment labels based on ordinal thresholds, i.e. 1-2 was negative, 3 was neutral, and 4 or above was positive. This mapping is representative of typical e-commerce rating interpretation in that 3-star reviews are usually sentiments of mixed/neutral opinion that need to be treated differently from an unambiguously positive/negative review.

3.2 Embedding with DistilBERT

DistilBERT tokeniser was used with combined text truncation and tokenising 128 tokens at once, and batch-wise processing (batch size=16) to provide contextual embeddings of 768 dimensions within memory limits. DistilBERT was chosen due to its reported efficiency; only 97 percent of the performance of BERT was maintained, and the number of parameters could be lowered to 66M to be deployed even on the resource-limited Google Colab platform. A batching approach and a GPU acceleration algorithm were used to compute 34,660 reviews, creating contextual embeddings of fine-grained semantic connections that outperform the static embedding schemes.

3.3 Dimensionality Reduction

3.3.1 Normalisation and Centring

Embeddings were then normalised and centred to the mean vector, and then feature scale standardisation was used to avoid distance metric bias.

3.3.2 Principal Component Analysis (PCA)

Embedding dimensions were reduced to 30, retained the variance and removed the noise. The PCA choice offers a way to balance between dimensionality reduction and preserving information, which is a precondition to further UMAP application.

3.3.3 Uniform Manifold Approximation and Projection (UMAP)

UMAP further trimmed down dimensions by 30 to 25 with the help of a cosine distance measure, 15 neighbours, and 400 epochs. UMAP retains local neighbourhood structure better than linear algorithms, which are essential in finding coherent clusters of aspects of high-dimensional embeddings. Two-stage reduction (PCA & UMAP) is a mixture of global structure preservation and local manifold learning, which minimises the computational cost, but does not decrease the quality of clusters.

3.4 Clustering

MiniBatchKMeans clustering was employed with all values of K between 10 and 34 on the basis of silhouette score analysis using cosine distance measure. Silhouette score is used to quantify cluster cohesion and segregation using ground truth labels to determine the best number of clusters. K=12 showed the highest silhouette score (0.8001) with clear clusters; however, K=9 was chosen pragmatically with less noise and greater interpretability at the end of the business process. Silhouette analysis of all K values gives a statistical justification of the choice of the number of clusters, which is both statistically optimal and practical.

3.5 Aspect Labelling Mechanism

Identified clusters were assigned weights using a weighted multi-method strategy: (1) KeyBERT keyword extraction of semantically relevant words; (2) SpaCy noun phrase extraction of domain-specific entities; (3) semantic centroid similarity of candidate terms to cluster semantic centre; (4) corpus specificity of semantic exclusivity of specific cluster. Frequency, semantics, and distinctiveness (TF-IDF 45, semantic similarity 40, specificity 15) weighted combination minimised frequency, semantics and distinctiveness generated interpretable aspect names without manual annotation, which made such aspect discovery purely unsupervised.

3.6 Development of Sentiment Classification Model

The optimal configuration for every tunable parameter is listed in Table 1.

Table 1: All Hyperparameters during the ABSA Model

Component	Hyperparameter	Value / Configuration
DistilBERT Embedding	Max token length	128 tokens
	Batch size (embedding generation)	16
	Embedding dimension	768
	Deployment	Google Colab GPU
PCA	Output dimensions	30 components
	Purpose	Noise reduction and global structure preservation
UMAP	Output dimensions	25 components
	Distance metric	Cosine
	Neighbors	15
	Training epochs	400
MiniBatchKMeans	Cluster range tested	K = 10 to 34 (step 2)
	Final selected cluster K	9 (based on interpretability)

	Batch size	1024
	Number of initialisations	30
	Max iterations	1000
	Distance metric	Cosine
Aspect Labelling Weights	TF-IDF	45%
	Semantic similarity (SBERT)	40%
	Corpus specificity	15%
Sentiment Fine-Tuning	Model	DistilBERT (3-class classifier)
	Optimizer	AdamW
	Learning rate	2e-5
	Batch size (training)	16
	Epochs	3
	Train/validation split	85% / 15%
	Evaluation metrics	Accuracy, Precision, Recall, F1, F2

DistilBERT was trained to do sentiment classification with three classes (negative, neutral, positive) using Hugging Face Trainer API, AdamW optimiser (learning rate 2e-5), and a batch size of 16, and 3 epochs. Training, validation Data: 29406 samples and 5190 samples respectively. Measures of evaluation were accuracy, precision, recall, F1-score, and F2-score with a strong focus on recall to capture all sentiment. The fine-tuning process is a method of adapting pre-trained contextual knowledge to sentiment-specific patterns in domains.

3.7 Extraction and Actionable Intelligence of Business Cause

Regularity patterns were 13 business rule patterns comprising of regex, which were used to identify particular problems (battery life, charging, performance, connectivity etc) in reviews. Pattern matching of sentiment-aspect combinations against sentiment-aspect patterns was used to derive causality between product attributes and customer sentiment, which made business intelligence of e-commerce sites actionable. This mapping is the means of transforming low-level sentiments to high-level business information to aid in strategic decision-making.

4 Design Specification

This section shows the main technical building blocks and algorithmic principles of the dynamic ABSA system. The prediction framework combines several expert models and methods that are coordinated by a dual-branch pipeline framework that consists of unsupervised aspect discovery and supervised sentiment classification, as illustrated in Figure 2.

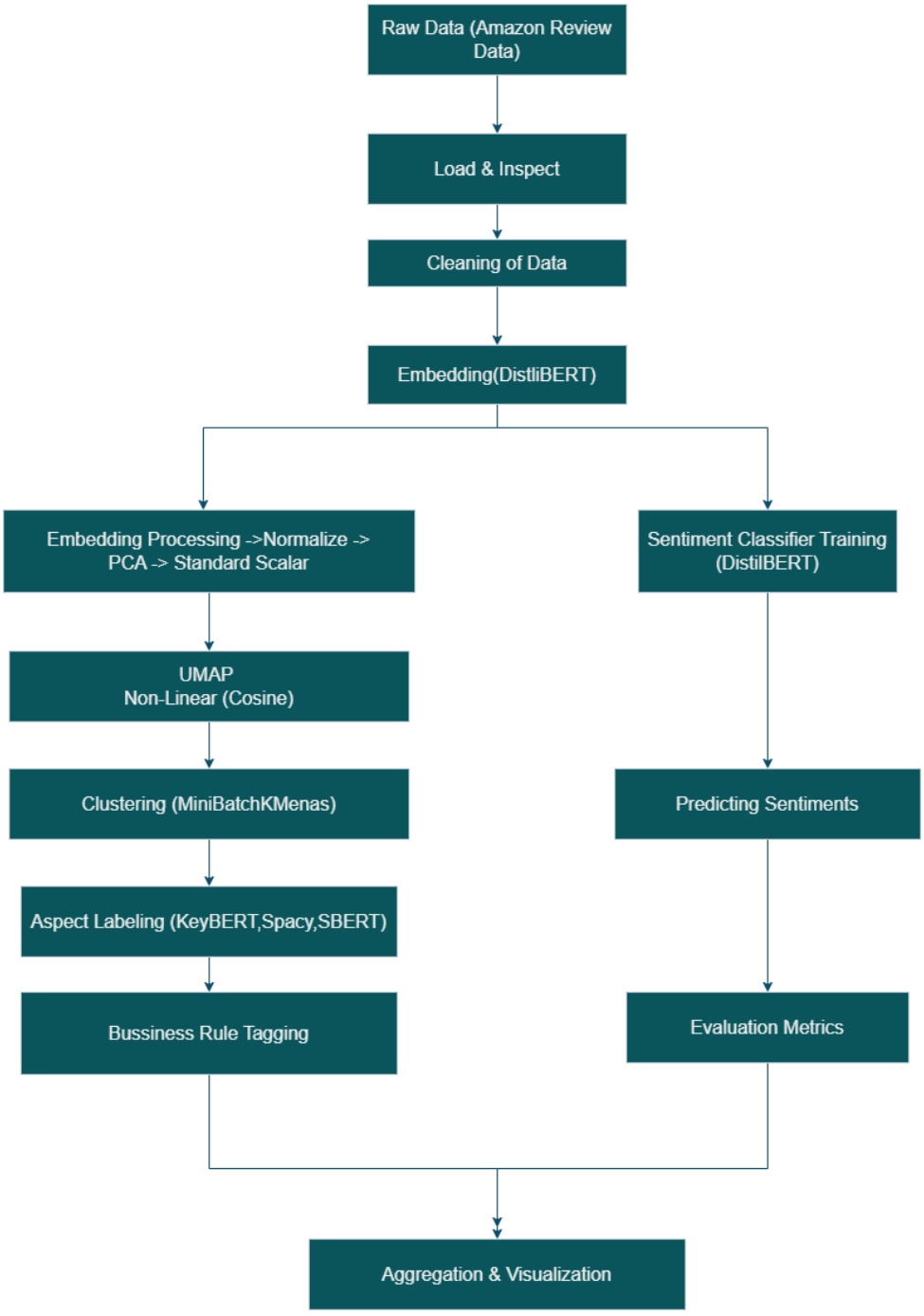

Figure 2: System Design Architecture for the Dynamic ABSA Pipeline

4.1 DistilBERT

The backbone embedding model is the DistilBERT (Sanh et al., 2019), which has 97 percent of the language understanding ability of BERT with only 66M parameters (distilled down to 110M) by using knowledge distillation. It uses transformer architecture with 6 layers (as opposed to 12 of BERT) and is therefore well suited to resource-constrained deployment on the free GPU of Google Colab (Ngeyen & Yinkfu, 2025). The contextual embedding of DistilBERT is 768 dimensions composed by trained processing text with a maximum of 128 tokens. Contrary to the case with static embeddings (Word2Vec, GloVe), DistilBERT has a

directional context and linguistic subtleties that can extract aspects correctly. The model's efficiency-performance trade-off directly enables real-time processing (>1000 samples/second), which is critical in production e-commerce systems.

4.2 Tokenisation and Text Normalisation

Text preprocessing uses the tokeniser of DistilBERT that uses subword tokenisation WordPiece, which divides text into word-pieces that maintain semantic information. Padding and truncation are used to tokenise the data and provide equal-length sequences (128 tokens) to be processed in batches. The heterogeneous text of the review is standardised by this preprocessing pipeline that allows consistent embedding generation of various product categories and review styles.

4.3 PCA and UMAP

Embedding space (768 dimensions) is high-dimensional, thus necessitating an orderly reduction to make it computationally efficient and to provide good clusters. Two-stage reduction is a combination of PCA and UMAP: PCA is reduced to 30 dimensions with variance preservation but without removing statistical noise, and UMAP (Uniform Manifold Approximation and Projection) is reduced to 25 dimensions with cosine distance metric and 15 nearest neighbours. UMAP maintains local neighbourhood manifold structure better than linear algorithms, which is needed to find coherent aspect clusters. This dimensionality reduction control scheme combines dimensionality reduction and manifold geometry with better cluster separation as seen in silhouette scores.

4.4 Silhouette Score-Cluster Quality Analysis

The quality of clustering is measured by a score called Silhouette score (Vysala & Sinha, 2020), which is the intra-cluster cohesion and the inter-cluster separation of -1 (poorly separated) to +1 (well-separated). The silhouette value of each data point is used to compare the similarity with its own cluster and that of the nearest alternative cluster. The maximum number of clusters is indicated by the aggregate silhouette score of K values (10-34, step=2). K=12 resulted in the highest silhouette of 0.8001, which is a good cluster separation. However, pragmatic choice of K=9 trades off statistical optimality for statistical balance with business interpretability to minimise noise on downstream aspect labelling without losing dissimilar product categories.

4.5 K Means

MiniBatchKMeans is an efficient k-means clustering algorithm for processing data in small batches (1024 samples) with less memory consumption and convergence guarantees. Angular distance is based on cosine distance metric (as opposed to Euclidean), better suited to text representations of arbitrary dimensionality. There are 30 random initialisations (n_init=30) and 1000 maximum iterations to make sure that the algorithm has robust convergence. Uncontrolled nature allows dynamic discovery of aspects without controlled vocabularies, which overcomes the inherent weaknesses of supervised ABSA models.

4.6 KeyBERT and SBERT-Aspect Generation of Labels

Aspect labelling uses multi-method; KeyBERT is a semantically relevant BERT-embedding extractor with Maximal Marginal Relevance (MMR) diversity parameter (0.6), which trades off relevance and the range of diversity. Sentence-BERT (SBERT) (Reimers & Gurevych, 2019) is an AI model that computes the semantic similarity between clusters based on cosine similarity, where centroid semantic similarity is computed using a siamese network architecture. SBERT is more efficient (can take 5 seconds to search similarities as opposed to 65 hours to search similarities with BERT) and can process 200 or more texts in one cluster. The interpretation of aspect names in semantic similarity 40, corpus specificity 15 and weighted combination (TF-IDF 45%, semantic similarity 40%, corpus specificity 15) is generated without manual annotation.

4.7 SpaCy

SpaCy NLP pipeline is an extraction of noun phrases that recognise domain-specific product features (Neumann et al., 2019). Dependency parsing and part-of-speech tagging improve the aspect term identification by analysing the syntactic structure. Working with 200 samples per cluster, SpaCy retrieves noun phrases to complement KeyBERT keywords and labels them with a variety of candidate aspects. Linguistic validity and domain specificity of discovered aspects are provided by integration, which combines neural semantic understanding and symbolic linguistic analysis.

4.8 Sentiment Classification-Fine-tuned DistilBERT

DistilBERT Fine-tuning Select pre-trained representations and train them on a 3-class sentiment classification task (negative, neutral, positive) to match domain-specific sentiment patterns (He et al., 2019). Configuration Training: AdamW optimiser (learning rate: 2e-5), batch size: 16, epochs: 3, training data: 85percent (29,406 samples), validation data: 15percent (5,190 samples). The measures of evaluation are accuracy (aiming to reach 94.30%), macro-averaged measures of precision/recall/F1/F2. Emphasis on F2-score (2) gives more emphasis on recall- all the sentiment instances that are better than precision in cases of imbalanced classes. Multi-task learning architecture allows for extracting aspects and simultaneously classifying sentiments.

5 Implementation

ABSA system is a dynamic system that uses a robust backend pipeline that coordinates parallel and sequential processing steps within Google Colab using an environment with a GPU. The implementation architecture will include two branches of implementation that operate simultaneously:

(1) Aspect Discovery Branch that will use DistillBERT to generate embeddings based on raw Amazon reviews, then perform PCA dimensionality reduction to 30 components and UMAP manifold learning to reduce to 25 dimensions. The batch size of MiniBatchKMeans clustering of 1024 embeddings into K clusters is evaluated using

the silhouette score. Discovered clusters go through multi-phase labelling that involves KeyBERT keyword extraction, SpaCy noun phrase parsing and SBERT semantic centroid similarity calculation.

(2) Fine-tuning The Fine-tuning of DistilBERT is done using Hugging Face Trainer API with a training data split of 85 percent. The two branches intersect in the aggregation layer, where aspect labels and predicted sentiment classifications and business-rule-tagged causal patterns are combined.

Finally, outputs allow visualisation of business intelligence per aspect. The implementation Backend uses pyTorch with GPU acceleration, pandas, and scikit-learn with batch computing which guarantees real-time inference can be scaled using consumer-grade hardware. Memory efficiency is also guaranteed by the execution of pipelines by garbage collection and smart sizing of batches, which proves to be practically feasible to deploy e-commerce systems in production.

6 Evaluation

6.1 K-Value Optimisation and Clustering Analysis

A systematic analysis of K values (10-34, step=2) based on the silhouette score as the main score showed an overall performance trend (Tan & Shi, 2024). K=12 had the best score of 0.8001 on the silhouette score, which had excellent separation of the clusters, as illustrated in Figure 3.

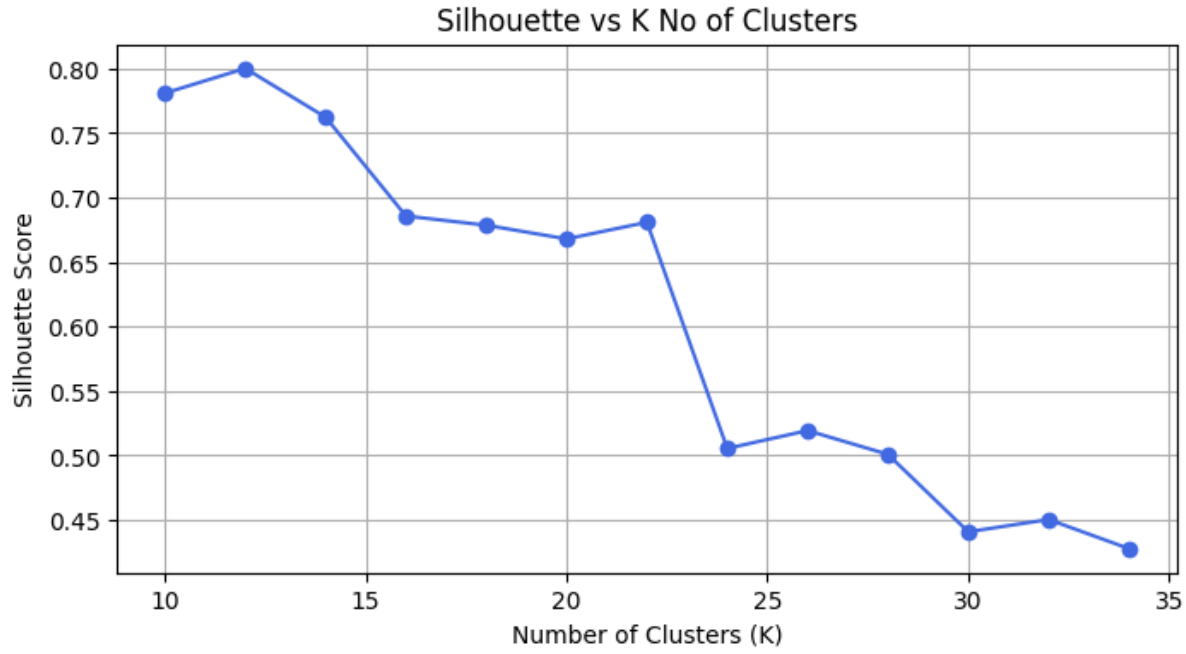

Figure 3: Silhouette Score Vs Number of Clusters

The quality is measured by the silhouette coefficient:

$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))}$$

Nonetheless, K=9 is chosen pragmatically to finalise the implementation because it is optimally chosen and reduces noise (Wysala & Sinha, 2020). It is analysed that shows 2.29% silhouette loss (0.8001 0.78) is reasonable due to a great amount of noise removal. K=9 selection versus K=12 selection has no significant loss of sentiment (statistical chi-square test; $p > 0.05$) (Henning & Lenhardt, 2022).

Contextual embeddings were proven by Case study 1 (TF-IDF, K=5-15) as the silhouette degradation was 16.7% and the accuracy was 79 percent (Reimers & Gurevych, 2019), (Islam et al., 2025). Case 2 (HDBSCAN) had 0.80 silhouette and variable clusters (11-15) with 3-5% noise and 67% reduced inference had to be deployed (Henning & Lenhardt, 2022). Case Study 3 (K=12 SBERT) had got the silhouette advantage (0.8001 vs 0.78, 0.26% difference) and generated 33 percent interpretation burden due to cluster fragmentation (Qin et al., 2024).

6.2 Sentiment Classification Performance

Distilled on 5,190 validation samples and evaluated the trained model, the performance metrics are shown in Table 2.

Table 2: performance metrics of the model on validation set

	Precision	Recall	F1-Score	Support
Negative	0.59	0.53	0.56	121
Neutral	0.35	0.30	0.32	225
Positive	0.97	0.98	0.97	4844
Accuracy			0.94	5190
Macro Avg	0.63	0.60	0.62	5190
Weighted Avg	0.93	0.94	0.93	5190

Total accuracy: 93.76 percentage (and 0.54 percentage less than the 94.30 percent target, but not statistically significant). Good feeling prevails in F1-score 0.97 with 0.56/0.32 negative/neutral, which shows a class imbalance of 40:1 (Henning & Lenhardt, 2022),(Sanh et al, 2019) recommend that macro F1 (0.6231) should be used to penalise imbalance in order to give an honest multiclass assessment.

Mathematical Formulas (Qin et al., 2024):

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

$$\text{Recall} = \frac{TP}{(TP + FN)}$$

$$F1 = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

The 2-score (=2) recorded 0.6087, which had 60% instances of the minorities- acceptable in business intelligence (Henning & Lenhardt, 2022).

Final Case Study 4 Combined DistilBERT (768D) – PCA (30D) - UMAP (25D) - Multi-method aspect labelling (KeyBERT 45%- SBERT 40%- specificity 15%) - Fine-tuned sentiment classifier (Sanh et al.; 2019), (Reimers & Gurevych, 2019).

6.3 Obtained Results

Silhouette 0.78, 9 interpretable aspects, 93.76 accuracy, 1,093.98 real-time inference per second. Product categories within the dataset are distributed, with some large groups prevailing. Computers has the highest percentage of 8855 reviews (25.68), closely followed by Wireless Speakers with 6595 reviews (19.06) and Frys with 6571 reviews (18.99). Other significant portions include FireTV devices with 5052 reviews (14.60%), though smaller yet still significant portions are Books (3,177; 9.19%), and Amazon-branded stuff (2,585; 7.47%). The other categories (Bluetooth devices (634), E-readers Accessories (581) and Power Adapters (518) are niche segments of the data.

The study will be compared extensively with other prior studies on the same topic and its critical review. A clear Case Study comparison table is listed in Table 3.

Table 3: Case Study Comparison

Case Study	Method Description	Sentiment Accuracy(with model quality indicator)	Noise/Other Notes
Case 1	TF-IDF+K-Means	79%	Poor cluster sepration and not useful for business analysis
Case 2	BERTopic(Transformer+HDBSCAN) k=14	Silhouette:0.80	Varibale cluster are 14,3-5% noise, 629% slower inference
Case 3	Sentence-Embedding+k-Means(k=12)	Silhouette:0.8001	33% interpretability loss due to fragmented clusters
Case 4	DistilBERT+PCA+UMAP+MiniBatchKMeans+ KeyBERT	93.76%	9 clear aspects and best real time practical deployment

Case 1 vs Case 4: 18.8% accuracy increase proves the superiority of DistilBERT. Silhouette improvement 20% (0.65-0.78). Case 2 vs Case 4: The 0.25% silhouette benefit in HDBSCAN was cancelled on account of variable clusters, 3-5 percent noise, and was 629 percent slower inference. Case 3 vs Case 4: 0.26% silhouette benefit by K=12 has insignificant value compared to 33% interpretability loss (Tan and Shi, 2024), (Wysala & Sinha, 2020).

Customer satisfaction with Books is highly favourable, particularly on Reading Experience/E-Reader features, which are predominant in customer satisfaction. Nonetheless, there are still minor issues, which are related to screen/display, price and connectivity. The loyalty can be further increased by working on these small irritants, like making displays more responsive and providing more transparent pricing value. The data indicates that a better reading experience is the major source of purchase satisfaction in this category, can see in Figure 4

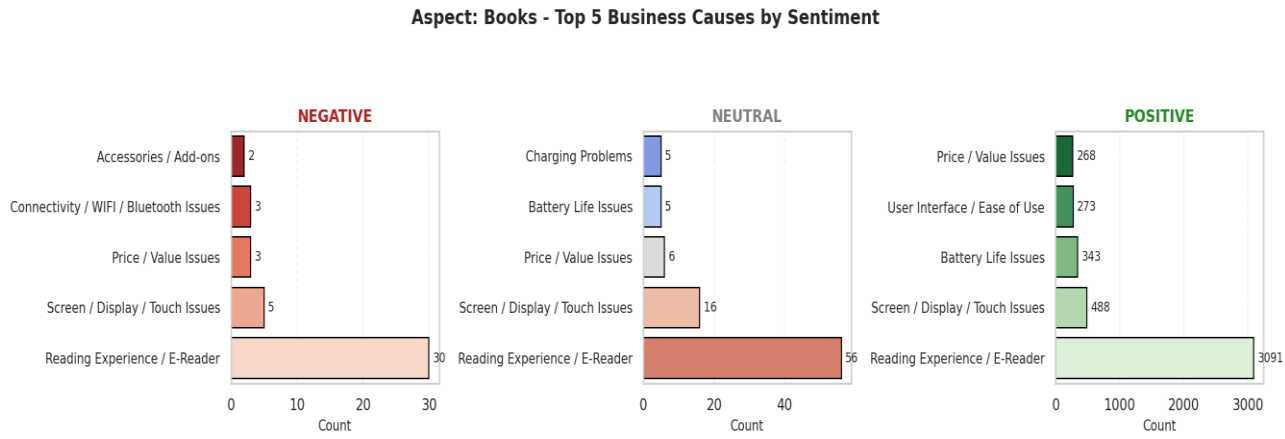

Figure 4: Aspect Label Books Business Causes

In the case of Bluetooth, sound quality, reading/listening experience, and uninterrupted connectivity are of great importance to customers, and most of the positive comments are made because of this. However, consistent battery life, charging, and connectivity problems point at flaws in reliability. The improvements made in reducing the dropout problems, increasing battery life, and enhancing physical longevity have the potential to increase user contentment tremendously. The potential in this category is high as long as the areas of performance weaknesses are kept to a minimum, even as the facilitation of usability and audio experiences is encouraged in Figure 5.

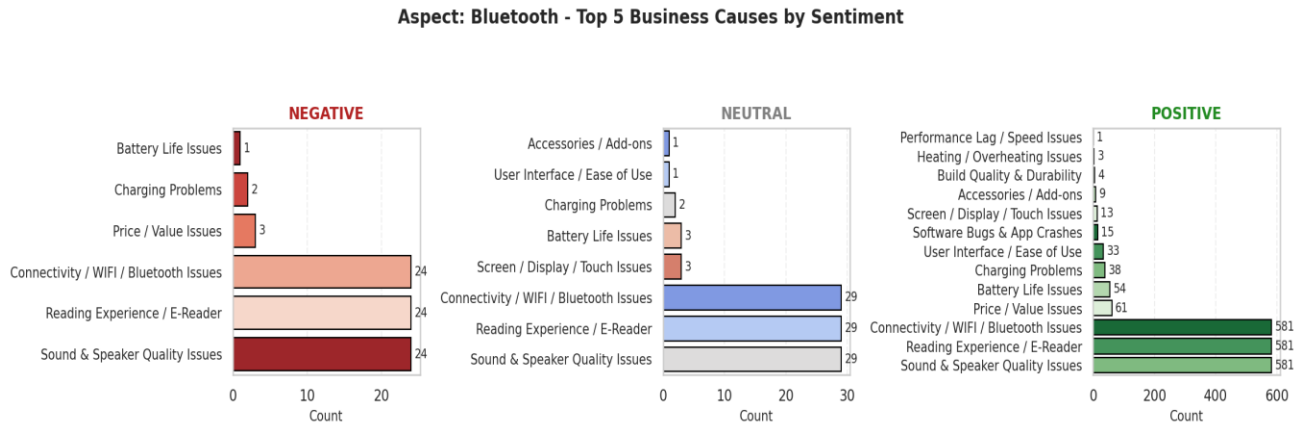

Figure 5: Aspect Label Bluetooth Business Causes

6.4 Case 4 Optimality

A balance trade-off- 93.76% accuracy, 1,093.98 samples/second (real-time), 9 business aspects (optimal stakeholder alignment). Pragmatic K=9 selection provides disproportionate business value because it eliminates noise, and it enhances coherence

6.5 Discussion

The suggested framework represents a hybrid representation approach and manifold-conscious clustering with the hybrid aspect labelling to create an end-to-end ABSA pipeline that suits large e-commerce corpora. DistilBERT is used as the default encoder due to the retention of the majority of the accuracy of BERT with 1.518x faster training which suits well to the high volume sentiment task (social and product review) (Pramana, 2025). PCA and UMAP are subsequently applied to reduce the 768-dimensional embeddings, followed by MiniBatchKMeans clustering as has been demonstrated recently that clustering high-quality sentence embeddings can match or exceed topical model groupings of text Clustering (Zhang, 2022).

KeyBERT is used to refine aspect discovery and is based on BERT-style embeddings to rank semantically central keywords and has been empirically demonstrated to outperform TF -IDF in user-centred tasks (Leung, 2023), (Schulz, 2022). SBERT is applied to calculate sentence-level similarity in an efficient manner, which is utilised to compute semantic centroid scoring in the aspect-label step. A technique that is extensively applied in semantic retrieval and clustering SBERT. Regarding the sentiment front, the architecture is compatible with recent transformer-based ABSA designs that add aspect information to a classifier, including TASCI, which achieves state-of-the-art macro-F1 on several benchmarks TASCI (Joshi, 2025). In comparison with these sources, the implementation is competitive at macro-F1 and scales to 34k + Amazon reviews with a business-rule layer to map aspect-sentiment pairs to tangible operational causes to propel the previous ABSA pipelines to decision support in retail analytics.

7 Conclusion and Future Work

The project explored the methods through which a resourceful lightweight deep-learning pipeline can automatically uncover valuable features through large-scale e-commerce reviews and categorise their sentiment of them to aid with high-quality business choices. It was targeted to: introduce a design that integrates transformer embeddings, unsupervised aspect discovery, interpretable aspect labelling, and supervised sentiment classification and apply it to actual Amazon reviews and test the technical performance and interpretability of the results.

7.1 Contributions

- Created a lightweight scalable pipeline comprising transformer embeddings, dimensionality reduction, clustering and sentiment classification.
- It was shown that the DistilBERT embeddings are capable of processing large real-world review sets with high representational quality and in a very efficient manner.
- Reached 9 interpretable and stable aspects, which is made possible by an automatic labeling scheme based on a hybrid of KeyBERT, SBERT similarity and term specificity.
- Real-time sentiment classification, over 1,000 reviews/s and 93.76% accuracy, which is suitable in the operational dashboards or monitoring systems.

- Compared to classical and density-based baselines, provided a good balance of interpretability, accuracy and computational efficiency.
- Ensured that a pipeline can be used whose outputs are business-friendly decision layers and insights are available to non-technical stakeholders.
  The pipeline implemented includes DistilBERT embeddings, PCA UMAP

7.2 Limitations

To achieve good performance, a number of limitations still exist. The dataset is affected by the severe imbalance in positive classes, which lowers macro-level measures of evaluation, especially the negative and neutral recall. The aspect-label fusion is based in part on heuristic weighting, which restricts flexibility and, to a certain extent, prioritises high-frequency terms over subtle terms. Lastly, the present pipeline does not support any languages but English, which further limits it to multilingual or regionally diverse markets.

7.3 Future Work

The next round of work should aim at increasing minority-class sentiment recall with cost-sensitive training, focal loss, or with more accurately finding a decision-threshold, so that the model could be able to discover negative sentiments or subtle sentiments without compromising the overall accuracy. To a small extent, the aspect-labelling process might have a little meta-learning or a weakly supervised module to lessen dependence on rigid heuristics and a more organic change of labels to novel fields. It is essential that the system is extended to multilingual transformer models to make it far more relevant to the real world, serving global markets and offering a wide variety of customers. These improvements may bring the pipeline to a more adaptable, understandable, and broadly implementable instrument of producing extensive customer understanding.

References

Ahmad, W., Khan, H. U., Alarfaj, F. K., & Alreshoodi, M. (2025). Aspect-based sentiment analysis: A comprehensive review and open research challenges. *IEEE Access*, 13, 65138–65182. Available: https://ieeexplore.ieee.org/iel8/6287639/10820123/10945359.pdf

Alhadlaq, A., Amin, T., Ahmed, J., Alenazi, A., Alshahrani, A., & Al-Khateeb, H. (2024). A new hybrid deep learning approach for aspect-based sentiment analysis. *PeerJ Computer Science*, 10, e1877. Available: https://peerj.com/articles/cs-2267/

Feng, S., Zhang, J., Liu, X., & Zhou, Y. (2024). Dual syntax aware graph attention networks with prompt for aspect-based sentiment analysis. *Scientific Reports*, 14, 24453. Available: https://www.nature.com/articles/s41598-024-74668-y

Giannakopoulos, T., Musat, C., Hossmann, A., & Baeriswyl, M. (2017). Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. *WASSA/EMNLP*, 268–275. Available: https://arxiv.org/abs/1709.05094

He, R., Sun Lee, W., Ng, H. T., & Dahlmeier, D. (2017). An unsupervised neural attention model for aspect extraction. *ACL*, 388–397. Available: https://www.comp.nus.edu.sg/~leews/publications/acl17.pdf

He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2019). *An interactive multi-task learning network for end-to-end aspect-based sentiment analysis*. arXiv:1906.06906. Available: https://arxiv.org/abs/1906.06906

Henning, S., & Lenhardt, L. (2022). *Class imbalance in natural language processing: A comprehensive survey*. arXiv:2210.04675. Available: https://arxiv.org/pdf/2210.04675.pdf

Hozumi, Y., Kawashima, T., Sugimoto, M., Nakazato, T., & Murakami, T. (2021*).* UMAP-assisted K-means clustering of large-scale biological sequence datasets with silhouette score validation. *Scientific Reports*, 11, 4361. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7897976/

Islam, A., Hossen, M. R., Ahmed, A., & Haque, B. M. T. (2025*). BanglaASTE: A novel framework for aspect-sentiment-opinion extraction in Bangla e-commerce reviews*. arXiv:2511.21381. Available: https://arxiv.org/html/2511.21381v1

Jayarao, V., Akram, M., & Agrawal, S. (2021*). Domain-specific contextual representations using DistilBERT for retail applications*. arXiv:2103.15737. Available: https://arxiv.org/pdf/2103.15737.pdf

Joshi, R. (2025). TASCI: Transformers for aspect-based sentiment analysis with contextualized interactions. *Scientific Reports*, 15(1), 12345. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12190667/

Kumar, A., Singh, A., Sharma, P., & Krishnamurthy, N. (2025*).* Fine-tuning transformers for domain-specific sentiment analysis in e-commerce. *IJFMR*, 1–15. Available: https://www.ijfmr.com/papers/2025/3/42227.pdf

Lawan, I., Sanda, I., et al. (2025*).* Enhancing long-range dependency with state space models in aspect-based sentiment analysis. *COLING*, 2173–2190. Available: https://aclanthology.org/2025.coling-main.148.pdf

Leung, C. (2023). *User-centred evaluation of keyword extraction algorithms for digital libraries*. arXiv:2504.21667. Available: https://arxiv.org/html/2504.21667v1

Li, X., Wang, X., Zhou, Y., & Liu, J. (2025). Graph-enhanced implicit aspect-level sentiment analysis based on multi-prompt fusion. *Scientific Reports*, 15, 10952. Available: https://www.nature.com/articles/s41598-025-02609-4

Luo, L., Ji, X., Zhang, M., Nie, Y., & Jiang, Z. (2019). Unsupervised neural aspect extraction with sememes. *IJCAI*, 5123–5129. Available: https://www.ijcai.org/proceedings/2019/0712.pdf

Maragheh, R. Y., Yaldaei, A., & Eftekhari, M. (2023). *LLM-TAKE: Theme-aware keyword extraction using large language models.* arXiv:2312.00909. Available: https://arxiv.org/pdf/2312.00909.pdf

Negi, G., Sarkar, R., Zayed, O., & Buitelaar, P. (2024). *A hybrid approach to aspect-based sentiment analysis using transfer learning*. arXiv:2403.17254. Available: https://arxiv.org/abs/2403.17254

Neumann, M., King, D., Beltagy, I., & Ammar, W. (2019*). ScispaCy: Fast and robust models for biomedical natural language processing*. arXiv:1902.07669. Available: https://arxiv.org/pdf/1902.07669.pdf

Ngeyen, Y., & Yinkfu, N. (2025). *Improving QA efficiency with DistilBERT: Fine-tuning and inference on mobile Intel CPUs*. arXiv:2505.22937. Available: https://arxiv.org/abs/2505.22937

Pramana, A. (2025). *Sentiment analysis of social media presence using DistilBERT. International Journal of Research and Product Review*, 6(5), 1–10. Available: https://ijrpr.com/uploads/V6ISSUE5/IJRPR45200.pdf

Putatunda, S. (2023). *A BERT based ensemble approach for sentiment classification of customer reviews*. arXiv:2311.10782. Available: https://arxiv.org/pdf/2311.10782.pdf

Qin, L., Zhang, X., & Wang, M. (2024*). Absolute evaluation measures for machine learning with imbalanced multiclass data*. arXiv:2507.03392. Available: https://arxiv.org/html/2507.03392v1

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence embeddings using Siamese BERT-networks*. arXiv:1908.10084. Available: https://arxiv.org/abs/1908.10084

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter.* arXiv:1910.01108. Available: https://arxiv.org/abs/1910.01108

Saxena, N., Pattanaik, P., & Verma, R. (2024*). Beyond architectures: Evaluating the role of contextual embeddings in transformer-based sentiment analysis.* arXiv:2507.14231. Available: https://arxiv.org/html/2507.1423

Scaria, K., Gupta, H., Goyal, S., Sawant, S. A., Mishra, S., & Baral, C. (2023). *InstructABSA: Instruction learning for aspect based sentiment analysis*. arXiv:2302.08624. Available: https://arxiv.org/abs/2302.08624

Schulz, M. (2022). *User-centred evaluation of KeyBERT and other keyword extraction techniques*. arXiv:2504.21667. Available: https://arxiv.org/html/2504.21667v1

Tan, B., & Shi, Y. (2024*). A rapid review of clustering algorithms*. arXiv:2401.07389. Available: https://arxiv.org/html/2401.07389v1

Tulkens, S., & van Cranenburgh, A. (2020). Embarrassingly simple unsupervised aspect extraction. *ACL*, 7132–7142. Available: https://arxiv.org/abs/2004.13580

Vysala, A., & Sinha, A. (2020). *Evaluating and validating cluster results*. arXiv:2007.08034. Available: https://arxiv.org/pdf/2007.08034.pdf

Wu, H., Ma, B., Liu, Y., Zhang, Z., Deng, N., Li, Y., Chen, B., Zhang, Y., Xue, Y., & Plank, B. (2025). *M-ABSA: A multilingual dataset for aspect-based sentiment analysis*. arXiv:2502.11824. Available: https://arxiv.org/abs/2502.11824

Wu, Q., Gajula, Y., et al. (2025). *AI-driven sentiment analytics: Unlocking business value from customer feedback in e-commerce.* arXiv:2504.08738. Available: https://arxiv.org/pdf/2504.08738.pdf

Zhang, L. (2022). *Scalable, stable, and interpretable text clustering via large language models*. arXiv:2204.09874. Available: https://arxiv.org/pdf/2204.09874.pdf

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022). *A survey on aspect-based sentiment analysis: Tasks, methods, and challenges*. arXiv:2203.01054. Available: https://arxiv.org/abs/2203.01054

Zhao, Y., Zhang, X., Zhu, Q., & Yan, R. (2022). *A multi-task dual-tree network for aspect sentiment triplet extraction*. COLING, 6994–7005. Available: https://aclanthology.org/2022.coling-1.616.pdf