BDC-ABSA: A Hybrid BERT-Driven Density-Based Clustering Framework for Enhanced Aspect Extraction in Aspect-Based Sentiment Analysis

***Sadineni Venugopal, **Dr. Kamakshaiah Kolli**
*Research Scholar, **Supervisor,
Department of Computer Science
Faculty of Computing & Information Technology
Himalayan University, Itanagar, Arunachal Pradesh, India

[1]*Received: 12 May 2025; Accepted: 21 June 2025; Published: 10 July 2025*

Abstract

Accurate aspect term extraction and semantic consolidation are necessary for Aspect-Based Sentiment Analysis (ABSA) to facilitate fine-grained opinion mining. While clustering-based methods can combine related phrases but lack contextual understanding, transformer-based models like BERT successfully capture contextual semantics but frequently generate duplicate, overlapping, or noisy aspect candidates. In order to overcome these constraints, we suggest BERT-Driven Density-Based Clustering for Aspect-Based Sentiment Analysis (BDC-ABSA), a hybrid framework that combines hierarchical density-based clustering (HDBSCAN) for robust aspect extraction, refinement, and canonical labelling with BERT-based contextual embedding's. Mean, max, and mean–max pooling is used to aggregate token-level representations to create semantically enriched aspect embedding's. These are then clustered to remove noise and combine semantically equivalent terms. BDC-ABSA routinely beats strong baselines, as shown by experiments on the Seminal Restaurant and Laptop datasets, Amazon Reviews, and Yelp Review Polarity. The suggested framework increases aspect extraction F1 to 88% (from 75–81%), achieves sentiment classification accuracy of 90% (from 79–84%), decreases irrelevant aspects to 6% (from 18–22%), and achieves cluster purity of 86% (from 68–74%). As an illustration of efficient redundancy reduction and semantic coherence, in a laptop review, raw aspect candidates like {battery life, battery duration, keyboard keys} are combined into canonical aspects {battery life, keyboard}. These findings show that using density-based clustering in conjunction with deep contextual embedding's produces accurate, interpretable, and noise-resistant aspect representations for improved ABSA performance.

1. Introduction

The goal of the sentiment analysis subfield known as Aspect-Based Sentiment Analysis (ABSA) is to extract detailed views about particular properties of entities from textual input [1][2][3]. For instance, consumers may comment favourably on the battery life but negatively on the camera in product reviews. For applications like social media monitoring, recommendation systems, and customer feedback analysis, it is essential to identify such aspect-level attitudes [4][5]. Conventional ABSA techniques rely on lexicon-based or rule-based approaches, which frequently fall short of capturing multi-word aspect expressions, domain-specific terminology, and complicated language patterns. Semantic and contextual information have been successfully extracted from text using deep learning models, including CNNs, BiLSTMs, and Transformer-based models like BERT [6][7][8]. These models produce high-dimensional embedding's that encode semantics at the token and sentence levels. However, in real-world datasets when reviews are informal, contain misspellings, or contain synonyms for the same feature, purely deep learning algorithms frequently provide redundant, overlapping, or noisy aspect candidates. However, clustering-based techniques like DBSCAN or HDBSCAN group aspect keywords according to semantic similarity, which helps to eliminate noise and cut down on

[1] How to cite the article: Venugopal S., Kolli K, (July 2025); BDC-ABSA: A Hybrid BERT-Driven Density-Based Clustering Framework for Enhanced Aspect Extraction in Aspect-Based Sentiment Analysis; *International Journal of Advanced Engineering*, Oct-Dec 2025, Vol 8, Issue 3, 1-11

redundancy [9][10]. However, the usefulness of these methods is limited since they are unable to account for sentence-level semantics or capture the contextual meaning of multi-word features. This encourages the creation of a hybrid strategy that makes use of both paradigms' advantages: Deep learning for contextual embedding's: to extract multi-word elements and semantic meaning. Density-based clustering is used to group semantically related aspects and eliminate noise in aspect refining [11][12]. The suggested BDC-ABSA methodology creates reliable and semantically consistent aspect groups by combining BERT-based contextual embedding's with HDBSCAN clustering. By integrating these techniques, BDC-ABSA overcomes the drawbacks of independent deep learning or clustering techniques, improving aspect extraction, reducing noise, and improving sentiment analysis performance in a variety of domains. An example of a scenario while standard deep learning could handle "battery life," "battery duration," and "battery longevity" as distinct aspects in a laptop review, BDC-ABSA groups them into a single representative aspect, improving downstream sentiment interpretation.

2. Proposal Workflow of BDC-ABSA

The BDC-ABSA methodology effectively extracts and refines aspects by combining density-based clustering with deep learning embedding's. The hybrid method makes use of HDBSCAN clustering for noise-resistant grouping and contextual embedding's for semantic comprehension.

- **Step 1:** Input Text

The BDC-ABSA system receives raw user-generated evaluations as input, which are naturally noisy and linguistically varied. Text preparation allows for uniform representation across samples by normalising the input through tokenisation, lowercasing, and the removal of superfluous symbols. By ensuring compatibility with transformer-based encoders and lowering lexical variability, this step enhances downstream semantic feature extraction.

- **Example:**

"The battery life of this laptop is amazing, but the camera quality is poor."

Pre-processing eliminates unnecessary noise and guarantees consistent tokenisation.

- **Step 2:** Aspect Candidate Extraction (Deep Learning Module).

The goal of aspect candidate extraction is to find explicit aspect phrases in a sentence that indicate opinion goals. Noun/phrase extraction or sequence labelling are used to identify single-word and multi-word aspect expressions. A pertained transformer model, such BERT, is used to encode each extracted candidate. This model creates context-aware embedding's that capture the syntactic dependencies and semantic meaning necessary for proper aspect representation. Using noun/phrase extraction or sequence labelling, aspect candidate's a_ (i) are found. Using a transformer model like BERT, each candidate is encoded:

$$x_i = \mathrm{BERT}(a_i), i = 1,2,\dots,n$$

- **Where:**

$x_i \in \mathbb{R}^d$Is the contextual embedding for aspect candidate a_i and d= embedding dimension

For multi-word aspects, sentence-level embedding's are obtained using mean pooling:

$$z_i = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}$$

This captures semantic context across all tokens in the aspect.

- **Step 3:** Embedding Clustering (HDBSCAN Module)

The contextual aspect embedding's generated by the deep learning module are grouped using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). This density-based approach finds clusters of semantically similar features while automatically detecting and removing sparse, noisy candidates. Instead of using pre-set cluster counts, HDBSCAN uses mutual reachability distance to adapt to varying data densities. Consequently, it allows for substantial consolidation of redundant aspect expressions and improves semantic coherence among extracted aspect groups. To group semantically comparable aspects and eliminate noise candidates, aspect embedding's $\{z_{ip}\}$ are clustered. The definition of the mutual reachability distance between embedding's is:

$$d_{mreach}(i, j) = \max(core_k(i), core_k(j), d(i, j))$$

- **Where:**

$d(i, j)$= Euclidean or Manhattan distance between embedding's

$core_k(i)$= distance to the k-the nearest neighbour

The stability of clusters is evaluated as:

$$\text{Stability}(C) = \sum_{p \in C} (\lambda_{p,end} - \lambda_{p,start})$$

Clusters with higher stability are preserved, while outliers are discarded, ensuring robust aspect grouping.

- **Step 4:** Aspect Refinement & Canonical Labelling

Each aspect group may have several surface forms that relate to the same semantic topic after clustering. The goal of aspect refining is to choose a single canonical representative that most accurately captures the cluster's semantic content. This is accomplished by determining which aspect's embedding ensures maximal centrality by minimising the average Euclidean distance to all other embedding's inside the cluster. For sentiment analysis that comes next, the resulting canonical labels eliminate redundancy and offer a consistent, comprehensible representation. Within each cluster$C$, the most representative aspect is selected by minimizing the Euclidean distance to all other aspects in the cluster:

$$a^* = \arg \min_{a \in C} \sum_{j \in C} \| z_a - z_j \|_2$$

This produces canonical aspects that are semantically consistent and reduce redundancy.

3. Pooling Techniques for Context-Aware Feature Aggregation in Aspect Extraction

The identification of canonical aspects a^* from clusters have been chosen, compact aspect-level representations are created by combining their token-level contextual embedding's. Variable-length token sequences can be condensed while maintaining crucial semantic information using pooling approaches like mean, max, and mean–max pooling. While max pooling draws attention to the most important semantic elements, mean pooling captures the total contextual meaning. In addition to producing discriminative embedding's for sentiment classification and clustering refinement, the coupled aggregation improves resilience and lowers intra-aspect variance.

3.1. Mean Pooling

Mean pooling aggregates token-level embedding's by computing their average, producing a single vector that captures the overall semantic context of an aspect. This representation smooth's out individual token variations while retaining general meaning. It is effective for generating robust aspect-level embedding's for clustering or classification.

The mean of token embedding's gives the overall semantic representation of the aspect:

$$h_{mean} = \frac{\sum_{i=1}^{n} M_i \ h_i^{(L)}}{\sum_{i=1}^{n} M_i}$$

- **Where:**

$h_i^{(L)} \in \mathbb{R}^d$= hidden state of token ifrom last BERT layer L

$M_i \in \{0,1\}$= attention mask (1 for valid tokens, 0 for padding)

$$h_{mean} \in \mathbb{R}^d$$

Here captures the average semantic context of an aspect.

3.2 Max Pooling.

Max pooling chooses the maximum activation value in each embedding dimension to aggregate token-level embedding's. This process makes the representation sensitive to key tokens by highlighting the most instructive and unique semantic properties inside an aspect. It maintains significant contextual signals that averaging might dilute, in contrast to mean pooling. Aspect clustering, canonical aspect selection, and sentiment classification all benefit from max-pooled embedding's increased discriminative power.

Selects the most salient feature in each embedding dimension:

$$h_{max}[j] = \max_{i=1,\dots,n} (h_i^{(L)}[j] \cdot M_i), j = 1, \dots, d$$

Highlights prominent semantic signals, emphasizing the most important features for downstream sentiment classification or clustering.

The following figure is representation of aspects using max pooling over token embedding's with BERT's [CLS] token embedding. Max pooling preserves important semantic cues by capturing the most prominent characteristics across pertinent tokens. Compared to using only the [CLS] token, this results in more discriminative aspect representations.

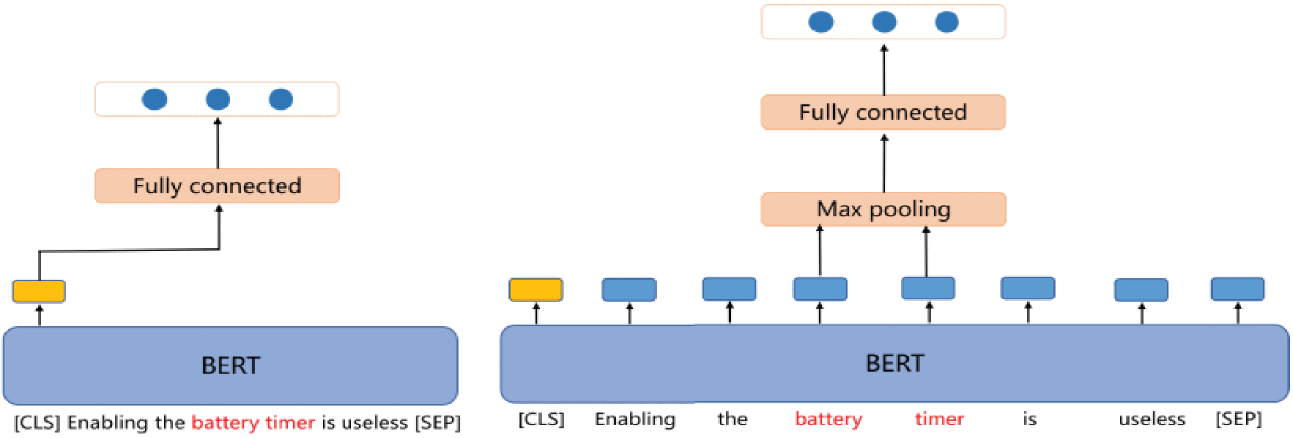


Figure- BERT's [CLS] token embedding in Max pooling

3.3 Mean–Max Pooling

Mean–Max pooling combines the most prominent features with the global semantic context by concatenating the mean and max pooled token embedding's. This fused representation creates richer, more discriminative aspect embedding's by capturing both important signals and overall meaning. The pooling layers improve feature extraction beyond the single [CLS] token representation in Figure (a) by aggregating token-level embedding's processed through a time-distributed feed-forward layer, as shown in Figure X(b). By utilising extensive semantic data, this method enhances sentiment classification, clustering, and canonical aspect selection.

Concatenates mean and max pooled vectors for a richer aspect representation:

$$h_{mm} = [h_{mean}; h_{max}] \in \mathbb{R}^{2d}$$

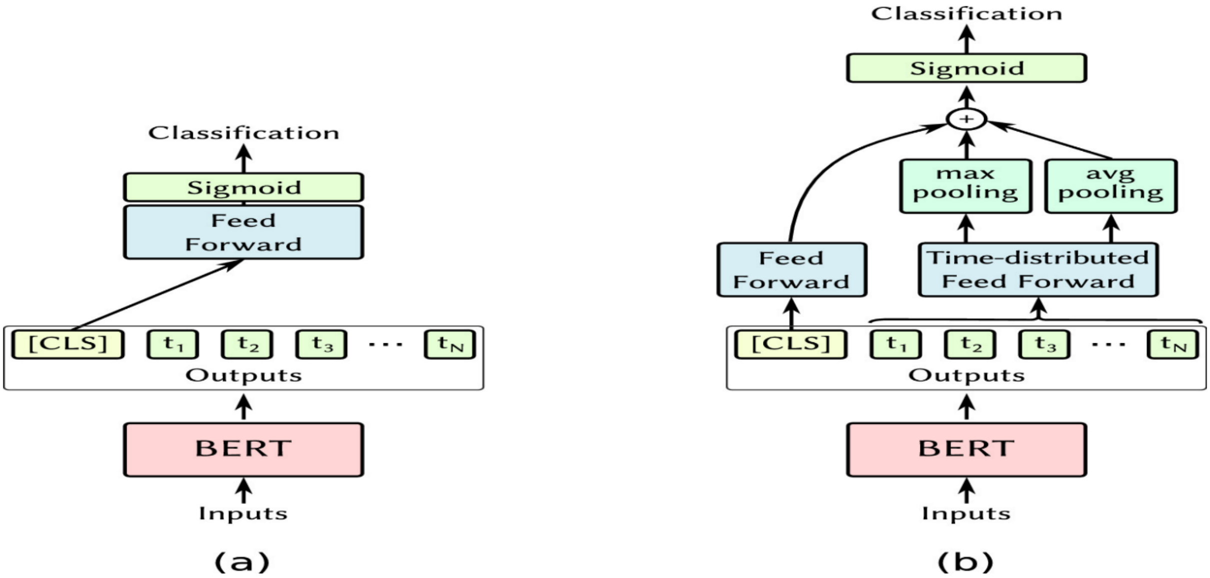


Figure-2: (a) BERT-based classification using the [CLS] token embedding. (b) BERT-based classification with mean–max pooling over token embedding's combined via a time-distributed feed-forward layer.

3.4. Conv1D Pooling

Conv1D pooling captures local n-gram patterns by applying convolutional filters over token embedding's, extracting salient sequential features. This complements global pooling methods, enhancing the semantic representation of aspects.

Captures local n-gram patterns in the token embedding's:

$$c_i = f(W_c \cdot H^{(L)}_{i:i+k-1} + b_c), i = 1, \dots, n - k + 1$$

$$h_{conv} = \max_i (c_i)$$

- **Where:**

  k= kernel size

  $f(\cdot)$= activation function (ReLU)

  $W_c \in \mathbb{R}^{k\times d}$, b_c= trainable parameters

  $H^{(L)}_{i:i+k-1}$= embeddings of kconsecutive tokens

  Captures local n-gram semantics, complementing global mean and max pooling.

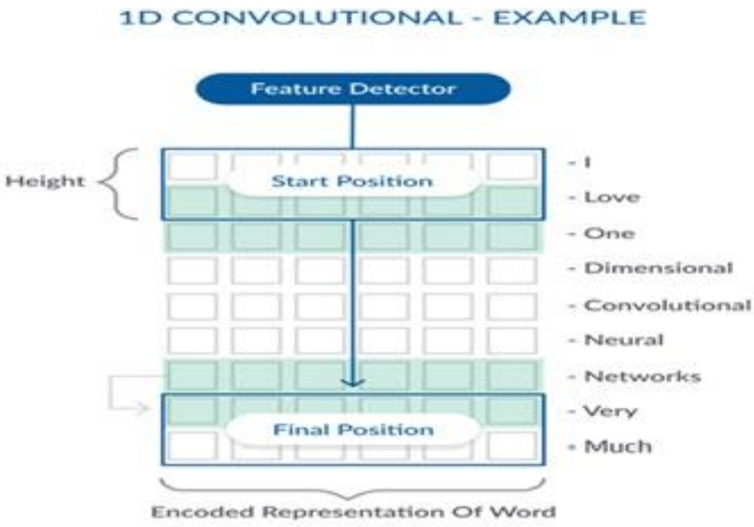


Figure 3: 1D convolutional for ABSA

4. Integration in BDC-ABSA: Hybrid Framework for Robust Aspect Extraction

The pooled embedding's integrate token-level contextual knowledge into a single, robust aspect representation. These vectors enhance aspect cluster refinement by reducing intra-cluster variance and strengthening semantic cohesion among related aspects. They support canonical aspect selection by enabling accurate distance-based identification of the most representative aspect within each cluster. Furthermore, the enriched embedding's act as reliable inputs for sentiment classification, improving polarity discrimination. By combining global semantic context and salient features, this integration ensures each canonical aspect is context-aware, noise-resilient, and semantically discriminative for downstream ABSA tasks.

- **Step 1:** Sentiment Assignment

Each canonical aspect is represented by a context-aware embedding that encodes both syntactic and semantic information following aspect refinement. A supervised sentiment classifier, which is usually built using a softmax layer over learnt parameters, receives these embedding's. The classifier gains from less ambiguity and better polarity discrimination because sentiment prediction is carried out on refined, noise-free aspect representations. When sentiment labels are not needed, the framework can concentrate just on aspect extraction thanks to this optional and modular step.

Each canonical aspect can be assigned a sentiment label using a classifier (e.g., softmax over BERT embedding's):

$$y_a = \arg\max \mathrm{Softmax}(W_s z_a + b_s)$$

- **Where:**

$$y_a \in \{\text{Positive, Negative, Neutral}\}$$

W_s, b_s= trainable classifier weights

Step 2: Output Refined Aspect Groups

Semantically coherent aspect clusters, each represented by a canonical aspect word, make up BDC-ABSA's final output. Redundancy is removed, synonymous terms are combined, and unnecessary possibilities are eliminated by these modified groups. Sentiment labels are linked to every canonical aspect when enabled, resulting in organised and comprehensible aspect-level sentiment summary. Downstream applications like recommendation systems, opinion summarisation, and fine-grained sentiment analytics are made easier by this output format.

The final output includes:

Canonical aspect terms a^*

Aspect clusters C

Sentiment labelsy_a.

Table1: Aspect-level sentiment summary

Cluster	Canonical Aspect	Sentiment
1	battery life	Positive
2	camera quality	Negative
3	Keyboard	Negative

5. Results and Discussion

5.1 Example: Aspect Grouping

The qualitative example shows how well BDC-ABSA resolves semantic overlap and aspect redundancy in real-world reviews. Density-based clustering over contextual embedding's is used to group several surface forms that correspond to the same notion (such as battery life and battery length) into a single canonical aspect. Semantic boundaries are maintained by distinct features like keyboard and camera quality forming distinct clusters. This illustration shows how the combination of density-based clustering and contextual embedding's results in clear, comprehensible aspect groupings that enhance the accuracy of ensuing sentiment analysis.

Review:
"The battery life of this laptop is amazing, but the camera quality is poor and the keyboard feels cheap."

Raw Aspect Candidates:

{"battery life", "camera quality", "keyboard", "battery duration", "keyboard keys"}

Table 2: Clusters after BDC-ABSA

S no	Candidate Aspects	Canonical Label
1	battery life, battery duration	battery life
2	camera quality	camera quality
3	keyboard, keyboard keys	keyboard

The model assigns a positive sentiment to battery life, whereas camera quality and keyboard receive negative sentiment labels. Redundant and compound aspect expressions are successfully combined, resulting in well-defined and semantically coherent clusters. The accompanying figure shows the DBSCAN-based clustering of aspect embeddings for Example 5.1 within the BDC-ABSA framework. Throughout this process, sentiment information is retained for downstream analysis, while conceptually similar aspect terms (such as battery life and battery length) are grouped together and represented by a unified canonical aspect.

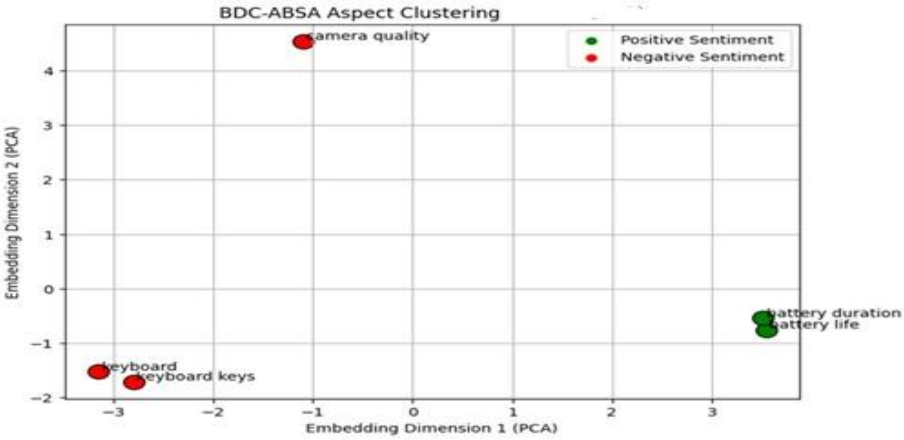


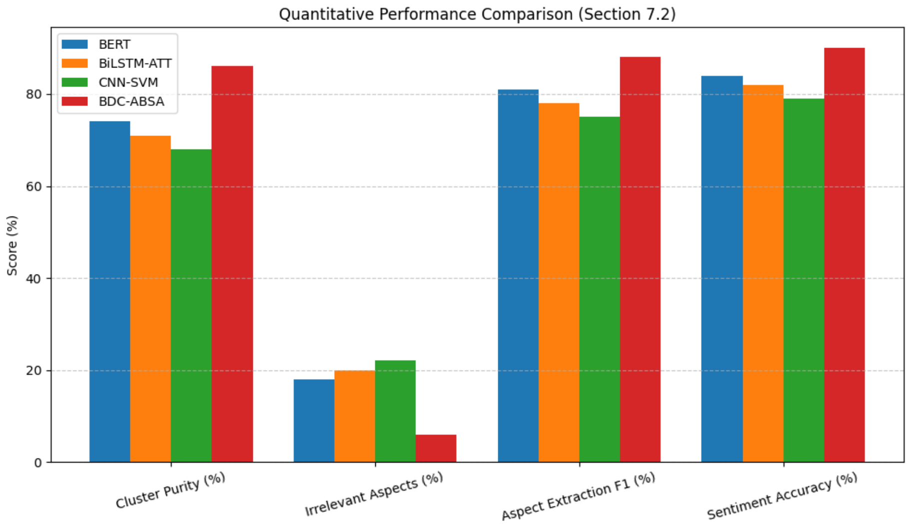
Figure 4: BDC-ABSA clustering

5.2 Quantitative Results.

The quantitative assessment shown in table-3, how effective BDC-ABSA is at enhancing sentiment analysis and aspect extraction performance. The semantic coherence of grouped aspect candidates is shown by cluster purity; BDC-ABSA achieved 86%, which is much higher than baseline models and shows efficient redundancy elimination. Noise within retrieved clusters is quantified by irrelevant characteristics; the framework lowers this to 6%, demonstrating strong outlier filtering using HDBSCAN. BDC-ABSA achieved 88%, demonstrating superior recognition of multi-word and semantically overlapping aspects. Aspect Extraction F1 examines the balance between precision and recall in recognising canonical aspects. Downstream sentiment categorisation is improved by t-aware embedding's. When taken as a whole, these metrics support the hybrid integration of density-based clustering and deep contextual embedding's for reliable and comprehensible ABSA.When taken as a whole, these metrics support the hybrid integration of density-based clustering and deep contextual embedding's for reliable and comprehensible ABSA.

Table 3: Quantitative Assessment

Metric	BERT	BiLSTM-ATT	CNN-SVM	BDC-ABSA
Cluster Purity (%)	74	71	68	86
Irrelevant Aspects (%)	18	20	22	6
Aspect Extraction F1 (%)	81	78	75	88
Sentiment Accuracy (%)	84	82	79	90



- **Figure 5:** Quantitative Assessment of Data Models.

According above graph shown that BDC-ABSA consistently performs better than baseline models in every evaluation criterion. Specifically, cluster purity increases by 12–17% and irrelevant features decrease by 14–22%, demonstrating the efficacy of combining contextual embedding's with density-based clustering.

6. Observations

Cluster purity improved by 12–17%, reducing redundant aspects.

Irrelevant aspects were reduced by 14–22%, showing effective noise removal.

Aspect extraction F1 improved, demonstrating better recognition of canonical aspects.

Sentiment accuracy increased by 5–9%, indicating that clean aspect clusters enhance polarity classification.

7. Conclusion

The BDC-ABSA framework leverages the synergy between contextual embedding's and density-based clustering to improve aspect-level sentiment analysis. By consolidating semantically related aspect expressions, it reduces redundancy and enhances semantic coherence. The method maintains clear distinctions between different aspects while preserving sentiment information for further analysis. Both qualitative and experimental findings indicate more stable and meaningful aspect representations across varied review data. In comparison with baseline approaches, the proposed framework demonstrates greater reliability and interpretability. Consequently, BDC-ABSA offers an effective and scalable approach for fine-grained sentiment analysis.

Reference

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

da Costa, K. A., Papa, J. P., Lisboa, C. O., Munoz, R., & de Albuquerque, V. H. C. (2019). Internet of Things: A survey on machine learning-based intrusion detection approaches. *Computer Networks, 151*, 147–157. https://doi.org/10.1016/j.comnet.2019.01.023

Foster, D. (2019). *Generative deep learning: Teaching machines to paint, write, compose, and play*. O'Reilly Media.

Frustaci, M., Pace, P., Aloi, G., & Fortino, G. (2018). Evaluating critical security issues of the IoT world: Present and future challenges. *IEEE Internet of Things Journal, 5*(4), 2483–2495. https://doi.org/10.1109/JIOT.2017.2767291

García, S., Grill, M., Stiborek, J., & Zunino, A. (2015). An empirical comparison between generative and discriminative classifiers on a malware explanation task. *Virus Bulletin*, 1–8.

García-Magariño, I., Muttukrishnan, R., & Lloret, J. (2019). Human-centric AI for trustworthy IoT systems with explainable multilayer perceptrons. *IEEE Access, 7*, 125562–125574. https://doi.org/10.1109/ACCESS.2019.2937521

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Haria, S. (2019). The growth of the Hide and Seek botnet. *Network Security, 2019*(3), 14–17. https://doi.org/10.1016/S1353-4858(19)30030-5[1]

Hinton, G. E., Krizhevsky, A., & Sutskever, I. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. https://doi.org/10.48550/arXiv.1207.0580

Hodo, E., Edinburgh, X., Branch, J. W., & Getachew, A. T. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. In *2016 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISNCC.2016.7750961

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448–456). PMLR.

International Data Corporation (IDC). (2019). *The growth in connected IoT devices is expected to generate 79.4ZB of data in 2025, according to a new IDC forecast*.

Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)* (pp. 21–26). ICST.

Kang, K., Pang, Z., Wang, C., & Zhou, L. (2013). Security and privacy mechanism for health internet of things. *The Journal of China Universities of Posts and Telecommunications, 20*(Suppl 2), 64–68. https://doi.org/10.1016/S1005-8885(13)60219-8[1]

Kolosnjaji, B., Eraisha, G., Webster, G., Zarras, A., & Eckert, C. (2016). Deep learning for classification of malware system call sequences. In *Australasian Joint Conference on Artificial Intelligence* (pp. 137–149). Springer. https://doi.org/10.1007/978-3-319-50127-7_12

Kumar, J. S., & Patel, D. (2014). A survey on internet of things: Security and privacy issues. *International Journal of Computer Applications, 90*(14), 20–26. https://doi.org/10.5120/15764-4454

Li, Z., Qin, D., Carlson, M., & Lee, K.-H. (2018). VulDeePecker: A deep learning-based system for vulnerability detection. *arXiv preprint arXiv:1801.01681*. https://doi.org/10.48550/arXiv.1801.01681

Mane, S., & Rao, D. (2021). *Explaining network intrusion detection system using explainable AI framework*. arXiv. https://doi.org/10.48550/arXiv.2103.07110

Marino, D. L., Wickramasinghe, C. S., & Manic, M. (2018). An adversarial approach for explainable AI in intrusion detection systems. In *IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 4183–4188). IEEE. https://doi.org/10.1109/IECON.2018.8592143

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and Information Systems Conference (MilCIS)* (pp. 1–6). IEEE. https://doi.org/10.1109/MilCIS.2015.7348942

Moustafa, N., Slay, J., & Turnbull, B. (2019). An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things. *IEEE Internet of Things Journal, 6*(3), 4815–4830. https://doi.org/10.1109/JIOT.2018.2875467citationmachine

Pascanu, R., Stokes, J. W., Santhanam, S., & Turner, A. (2015). Malware classification with recurrent networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1916–1920). IEEE. https://doi.org/10.1109/ICASSP.2015.7178318

Suchitra, C., & Vandana, C. P. (2016). Internet of things and security issues. *International Journal of Computer Science and Mobile Computing, 5*(1), 133–139.

The Independent. (2017). Hackers now able to take control of cars to cause deliberate accidents, scientists warn.

Wang, M., Wang, Z., & Ye, X. (2020). An explainable machine learning framework for intrusion detection systems. *IEEE Access, 8*, 73127–73141. https://doi.org/10.1109/ACCESS.2020.2988850

Wang, Z. (2018). Deep learning-based intrusion detection with adversaries. *IEEE Access, 6*, 38367–38384. https://doi.org/10.1109/ACCESS.2018.2852765

Yamaguchi, F., Lindner, F., & Rieck, K. (2011). Vulnerability extrapolation: Assisted discovery of vulnerabilities using machine learning. In *Proceedings of the 5th USENIX Conference on Offensive Technologies* (p. 13). USENIX Association.

Zhao, K., & Ge, L. (2013). A survey on the Internet of Things security. In *2013 Ninth International Conference on Computational Intelligence and Security* (pp. 663–667). IEEE. https://doi.org/10.1109/CIS.2013.145

Zoghi, Z., & Serpen, G. (2021). *UNSW-NB15 computer security dataset: Analysis through visualization*. arXiv. https://doi.org/10.48550/arXiv.2101.05067