iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IX **Month of publication:** September 2026

DOI: https://doi.org/10.22214/ijraset.2026.84938

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IX Sep 2026- Available at www.ijraset.com

AI-Driven Service Analytics for Business Decision Support Using Explainable AI

Yeduvakula Niharika[1], G. Ramesh Naidu[2]
[#]*Department of Computer Science and Engineering, Sanketika Vidya Parishad Engineering College, Visakhapatnam, Andhra Pradesh, India*

Abstract: ***Online reviews hold detailed but unstructured evidence about how customers experience a business, yet managers seldom have the time or tooling to convert this text into prioritised and trustworthy actions. This paper presents an AI-driven service analytics framework that transforms raw customer reviews into aspect-level insights and explains the reasoning behind every sentiment judgement. Reviews are encoded with a compact sentence transformer (all-MiniLM-L6-v2) and the resulting dense vectors are kept in a lightweight SQLite repository. For each business aspect, namely food, service, ambience and pricing, a retrieval stage selects the most semantically relevant reviews through cosine similarity, after which a DistilBERT classifier fine-tuned on SST-2 estimates the positive-class probability of the aspect-bearing sentences. The averaged scores are mapped to positive, mixed or negative labels and linked to a rule base of managerial recommendations that is ordered so the weakest aspect is addressed first. To make every output auditable, SHAP partition explanations attribute each prediction to individual tokens. On a balanced sample of restaurant reviews drawn from the UCI Sentiment Labelled Sentences corpus, the framework scored service and ambience at 0.75, food at 0.66 and pricing at 0.50, identifying pricing as the first improvement priority. The SHAP study of a contrastive review showed that the conjunction "but" and negative descriptors counterbalance a strongly positive opening word, demonstrating how explanations reveal the model's internal reasoning. The entire pipeline runs locally with open-source models and no paid APIs, which makes it practical for small and medium enterprises.***

Keywords: Explainable Artificial Intelligence, SHAP, Aspect-Based Sentiment Analysis, Retrieval-Augmented Analytics, Sentence Transformers, DistilBERT, Business Decision Support, Customer Review Mining.

I. INTRODUCTION

Customer feedback has moved from comment cards and survey forms to open, public platforms where anyone can describe a visit in their own words. For a service business such as a restaurant, these short texts are a continuous stream of evidence about what is working and what is not. A single sentence may praise the cooking, complain about the waiting time and question the bill, all at once. Reading such feedback manually is feasible for a handful of entries, but it becomes impractical as volume grows, and star ratings alone compress several distinct experiences into one number that cannot tell a manager where to act.

Natural Language Processing (NLP) offers an obvious remedy. Sentiment classifiers built on transformer architectures now label review polarity with high reliability [1], [2]. However, a document-level polarity label answers only whether a customer was satisfied, not why. Business decisions need finer granularity: an owner has to know whether dissatisfaction arises from food quality, staff behaviour, the physical environment or perceived value for money. This requirement motivates aspect-based sentiment analysis, in which opinions are resolved against specific facets of the service [3], [4]. A second barrier is trust. Deep neural classifiers are frequently criticised as opaque, and a manager is understandably reluctant to reallocate budget or retrain staff because a model returned a number without justification. Explainable Artificial Intelligence (XAI) addresses this gap by attributing a model's output to the input features that drove it [5], [6]. Among the available methods, SHapley Additive exPlanations (SHAP) is attractive because it rests on cooperative game theory and yields additive, locally faithful attributions [7]. This work combines these threads into a single decision-support pipeline. Rather than classifying every review blindly, the framework first retrieves the reviews that are semantically closest to each business aspect, drawing on the retrieval idea popularised by retrieval-augmented generation [8], [9]. It then scores only the relevant evidence, converts the scores into categorical judgements and concrete recommendations, and supplies token-level explanations for inspection. The main contributions of the paper are as follows:

1) *Aspect-Oriented Retrieval:* a dense-embedding retrieval layer that isolates evidence for each business aspect from a shared review store without requiring manually labelled aspect terms.
2) *Actionable Decision Layer:* a transparent scoring and thresholding scheme that maps aspect sentiment to managerial recommendations and ranks aspects by urgency.

3) *Built-in Explainability:* integration of SHAP partition explanations so that each sentiment estimate can be traced to the words responsible for it.
4) *Low-Cost Deployment:* a fully local implementation based on open-source models and an embedded database, suitable for organisations without access to commercial AI services.

The remainder of the paper is organised as follows. Section II reviews related research. Section III describes the proposed methodology. Section IV outlines the implementation environment. Section V reports and discusses the results, and Section VI concludes with directions for future work.

II. RELATED WORK

A. Review Mining and Sentiment Analysis

Early work on customer review mining extracted frequently mentioned product features and summarised the opinions attached to them using lexical rules [4]. Pang and Lee [10] consolidated the field of opinion mining, describing the progression from lexicon-driven methods to supervised machine learning. Recursive neural models trained on the Stanford Sentiment Treebank showed that sentiment is compositional and can be modelled at the phrase level [11]. Zhang et al. [12] later surveyed how convolutional, recurrent and attention-based deep networks overtook classical feature engineering for polarity detection.

B. Transformer Language Models

The self-attention mechanism introduced in the Transformer [13] made it possible to pre-train deep bidirectional encoders such as BERT [1], which reached leading results on many classification benchmarks after light fine-tuning. Because full-size BERT is expensive to deploy, DistilBERT [2] applied knowledge distillation to retain most of its accuracy with roughly 40% fewer parameters and faster inference. Open libraries such as Hugging Face Transformers [14] have since made these models accessible to practitioners through uniform interfaces.

C. Aspect-Based Sentiment Analysis

The SemEval-2014 shared task formalised aspect-based sentiment analysis for restaurant and laptop reviews, defining sub-tasks for aspect term extraction, aspect category detection and polarity assignment [3]. Most systems in this line require annotated aspect terms or categories for training. In contrast, the present work treats aspects as natural-language descriptors and relies on semantic similarity to locate evidence, which removes the need for aspect-level labels.

D. Dense Retrieval and Retrieval Augmentation

Sentence-BERT [15] adapted BERT with siamese and triplet objectives so that semantically similar sentences lie close together under cosine distance. MiniLM [16] compressed such encoders through deep self-attention distillation, and derived checkpoints such as all-MiniLM-L6-v2 produce 384-dimensional embeddings at low cost. Dense passage retrieval [9] demonstrated that learned embeddings can outperform sparse term matching for locating relevant text, and retrieval-augmented generation [8] showed the value of grounding a model's output in retrieved evidence. The proposed framework borrows the retrieval step of this paradigm and applies it to analytics rather than text generation.

E. Explainable Artificial Intelligence

Surveys by Adadi and Berrada [5] and Barredo Arrieta et al. [6] classify XAI methods by scope, model dependence and presentation, and argue that explanation is essential for responsible deployment of AI in decision-making. LIME [17] fits local surrogate models around individual predictions, while SHAP [7] unifies several attribution methods under Shapley values and guarantees properties such as local accuracy and consistency. SHAP has been widely adopted for tabular models; applying it to transformer-based text classifiers for business reporting, as done here, remains comparatively less explored.

F. Research Gap

Existing studies usually address sentiment accuracy, aspect extraction or explainability in isolation. Few works combine semantic retrieval, aspect-level scoring, managerial recommendation and token-level explanation in one lightweight system that a small business could run without specialised infrastructure. The framework described in the next section is designed to close this gap.

III. PROPOSED METHODOLOGY

A. System Architecture

The architecture of the proposed framework is shown in Fig. 1. It is organised as a sequential pipeline of seven stages: review acquisition, preprocessing, sentence embedding, vector storage, aspect-oriented retrieval, sentiment scoring and decision support. A SHAP explainer operates alongside the sentiment model and passes its attributions to the decision-support layer, so that explanations accompany the final recommendations.

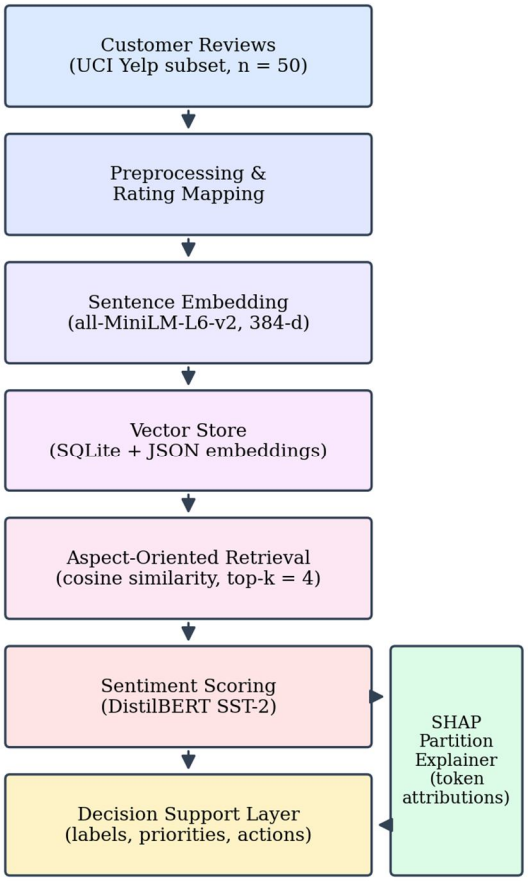


Fig. 1 Architecture of the proposed explainable service analytics framework

B. Dataset and Preprocessing

The experiments use the Yelp portion of the UCI Sentiment Labelled Sentences dataset, which was introduced by Kotzias et al. [18]. This subset contains 1,000 restaurant review sentences, each tagged as positive (1) or negative (0). To align the data with a star-rating view familiar to business users, positive labels were mapped to a rating of 4.5 and negative labels to 1.5, and every record was assigned to a single business entity. A balanced working set of 50 reviews, 25 positive and 25 negative, was drawn with a fixed random seed of 42 and shuffled to ensure reproducibility. Balancing prevents the aggregate scores from being biased simply by the majority class.

C. Semantic Embedding and Storage

Each review is converted into a 384-dimensional vector with the all-MiniLM-L6-v2 sentence transformer [15], [16]. The vectors are serialised as JSON and stored with the business name, review text and rating in an SQLite table. SQLite was chosen because it is serverless, requires no configuration and is bundled with Python, which keeps the deployment footprint small while still supporting structured queries over the review store.

D. Aspect-Oriented Retrieval

Four service aspects are considered, each represented by a descriptor string containing characteristic vocabulary, as listed in Table I. The descriptor is embedded with the same encoder to obtain a query vector q. For every stored review vector d_j the cosine similarity is computed as

$$\text{sim}(\mathbf{q}, \mathbf{d}_j) = \frac{\mathbf{q} \cdot \mathbf{d}_j}{\|\mathbf{q}\| \, \|\mathbf{d}_j\|} \tag{1}$$

Reviews are ranked by (1) and the k = 4 most similar entries are retained as evidence for the aspect. Operating on meaning rather than exact keywords allows the system to connect, for example, a complaint about a "bland" dish with the food aspect even when the word "food" does not appear.

TABLE I
ASPECT DESCRIPTORS USED FOR RETRIEVAL

Aspect	Descriptor Vocabulary
Food	food, quality, taste, menu, dishes, flavor, meal, cuisine, ingredients
Service	staff, waiter, service, speed, helpful, rude, friendly, wait
Ambience	ambience, atmosphere, decor, noisy, cozy, environment, clean, dirty
Pricing	price, expensive, cheap, overpriced, value, cost, money, worth

E. Aspect Sentiment Scoring

Retrieved reviews are split into sentence fragments at terminal punctuation, and fragments longer than eight characters that contain at least one descriptor term are kept as aspect-bearing units. If no fragment qualifies, the complete retrieved reviews are used instead. Each unit u is passed to a DistilBERT model fine-tuned on the SST-2 benchmark [2], [11], which returns the probability P(pos | u) that the unit expresses positive sentiment. The aspect score S_a is the mean over the set of units U_a:

$$S_a = \frac{1}{|U_a|} \sum_{u \in U_a} P(\text{pos} \mid u) \tag{2}$$

The score is then converted into a categorical label L_a using two thresholds. An aspect is labelled positive when $S_a \geq 0.65$, negative when $S_a \leq 0.40$, and mixed otherwise. The intermediate band deliberately absorbs uncertain or contradictory evidence so that managers are not pushed towards decisive action on ambiguous signals.

F. Decision Support Layer

A rule base associates every combination of aspect and label with a concise managerial recommendation, giving twelve possible actions in total. Positive labels trigger advice to preserve and promote a strength, mixed labels suggest standardisation or targeted improvement, and negative labels call for urgent corrective measures. Aspects are finally sorted by ascending score, so that the report opens with the area most in need of attention. Keeping this layer rule-based makes the link between evidence and advice fully transparent and easy for domain experts to edit.

G. Explainability with SHAP

SHAP explains a prediction f(x) by distributing it among the input features according to their Shapley values [7]. For a feature i drawn from the full feature set F, the attribution is

$$\phi_i = \sum_{T \subseteq F \setminus \{i\}} \frac{|T|! \, (|F| - |T| - 1)!}{|F|!} [f(T \cup \{i\}) - f(T)] \tag{3}$$

where T ranges over subsets of features that exclude i. The attributions satisfy the additive property

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \tag{4}$$

in which φ_0 is the expected model output and M is the number of input tokens. Exact evaluation of (3) is exponential in M, so the framework uses the SHAP Partition explainer, which exploits a hierarchical grouping of tokens to approximate Shapley values efficiently for text. A positive φ_i indicates that a token pushes the prediction towards the positive class, and a negative value indicates the opposite.

IV. IMPLEMENTATION

The framework was implemented in Python and executed on Google Colaboratory using CPU resources. All models are open-source and are downloaded once from the Hugging Face model hub, after which the pipeline runs without any external API calls. Table II summarises the main software components and their roles.

TABLE II
IMPLEMENTATION COMPONENTS

Component	Tool and Role
Language	Python 3 on Google Colaboratory
Embedding	sentence-transformers, all-MiniLM-L6-v2 (384-d vectors)
Storage	SQLite, reviews table with JSON-encoded embeddings
Retrieval	NumPy cosine similarity, top-k = 4
Sentiment	Hugging Face Transformers, DistilBERT fine-tuned on SST-2
Explanation	SHAP Partition explainer on the sentiment pipeline
Reporting	pandas and Matplotlib for tables and charts

The workflow proceeds in six steps: installing dependencies, loading and balancing the dataset, embedding and storing reviews, retrieving evidence per aspect, scoring sentiment and generating recommendations, and finally computing SHAP explanations and visual summaries. Because the embedding model has only about 22 million parameters and DistilBERT about 66 million, both fit comfortably in commodity memory.

V. RESULTS AND DISCUSSION

A. Retrieval Behaviour

A preliminary retrieval test with the query “food quality taste dishes” returned four reviews with cosine similarities between 0.46 and 0.54. All four concerned the taste, texture or selection of dishes, for instance a remark that the food lacked flavour and texture, confirming that the embedding space groups reviews by topic even when phrasing varies. All four were also negative reviews, a reminder that retrieval selects by topic and not by polarity, which is precisely the separation of concerns intended by the design.

B. Aspect-Level Sentiment

Table III reports the aspect scores, labels, priority ranks and recommended actions produced by the framework, and Fig. 2 visualises the scores against the two decision thresholds. Service and ambience obtained the highest score of 0.75, food reached 0.66, and pricing scored 0.50. Three aspects were therefore labelled positive and one mixed, while no aspect fell into the negative band.

TABLE III
ASPECT SENTIMENT AND RECOMMENDED ACTIONS

Aspect	Score	Label	Rank	Recommended Action
Pricing	0.50	Mixed	1	Introduce combo deals and loyalty discounts to improve value perception
Food	0.66	Positive	2	Maintain food quality as the key asset for retention
Service	0.75	Positive	3	Reward staff and sustain current service standards
Ambience	0.75	Positive	4	Continue maintaining the welcoming atmosphere

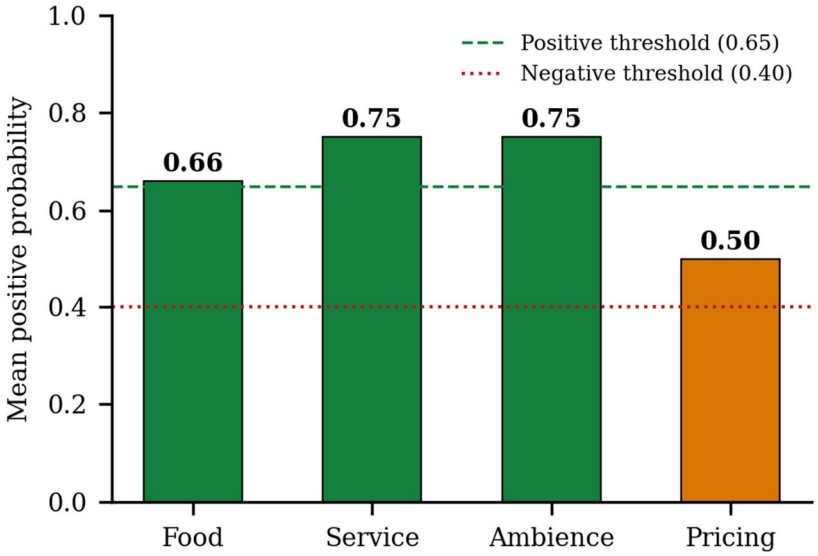


Fig. 2 Aspect-level sentiment scores with decision thresholds

From a decision-support viewpoint, the most useful outcome is the ranking. Pricing is the only aspect that failed to reach the positive threshold, so the report places it first and proposes value-oriented interventions. Food lies only marginally above the positive threshold, which suggests that it is a strength but not a secure one; a manager reading the report would reasonably monitor it closely. Service and ambience appear to be stable assets. This prioritised view converts fifty unstructured sentences into four clear statements that can be discussed in a management meeting.

C. Explainability Analysis

To examine how the classifier reaches its conclusions, SHAP was applied to the constructed review “Amazing food but terrible slow service and overpriced menu”, which deliberately mixes praise and criticism. Fig. 3 shows the eight tokens with the largest absolute attributions towards the positive class.

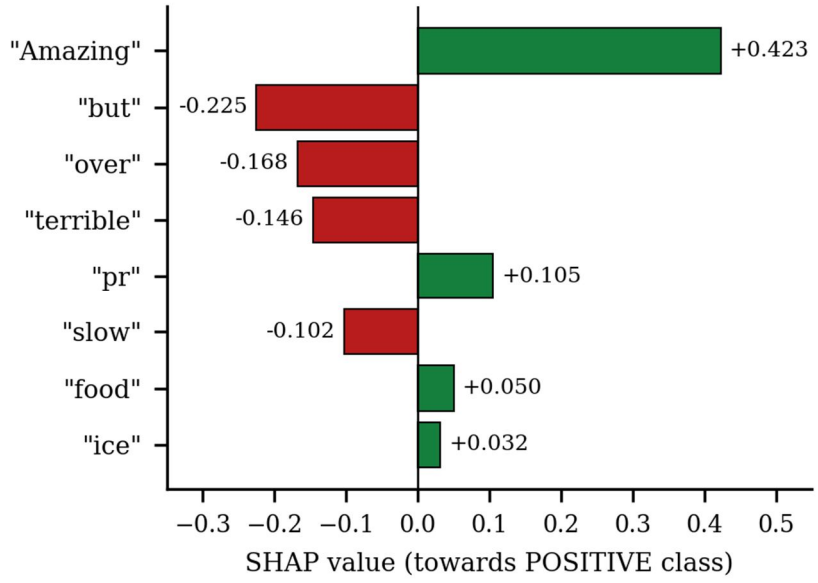


Fig. 3 SHAP token attributions for a mixed-sentiment review

The word “Amazing” contributed the strongest positive push (+0.423), and “food” added a small positive effect (+0.050). Against this, the contrastive conjunction “but” produced the largest negative contribution (−0.225), followed by “over” (−0.168), “terrible” (−0.146) and “slow” (−0.102). The prominence of “but” shows that the model has learned the discourse role of contrast: the conjunction signals that the positive opening is about to be qualified. The explanation also exposes a subtler effect of sub-word tokenisation. The word “overpriced” is split into “over”, “pr” and “ice”; the first fragment carries a negative attribution, while the remaining fragments receive small positive values (+0.105 and +0.032). Such detail would remain hidden without an explanation method and highlights where a manager should interpret individual attributions with care.

For business users, the value of these explanations lies in verification. A manager can see that a negative service judgement rests on words like “terrible” and “slow” rather than on incidental vocabulary, which builds confidence in the recommendation. Conversely, when attributions appear unreasonable, the explanation provides a concrete basis for challenging or correcting the model.

D. Comparison with Conventional Approaches

A conventional document-level classifier would assign a single polarity to the review above and hide the contrast between food and service. Keyword counting would capture aspect mentions but ignore negation and context. The proposed framework differs from both by combining semantic retrieval, which identifies relevant evidence without exact keyword matches, with contextual transformer scoring and post-hoc explanation. It also produces direct recommendations, whereas most sentiment tools stop at labels and leave interpretation to the user.

E. Limitations

Several limitations should be acknowledged. The evaluation uses a small sample of 50 short sentences from one domain, so the scores illustrate the behaviour of the pipeline rather than establish statistical generalisation. The dataset provides only document-level labels, which means aspect-level predictions could not be validated against a ground truth; quantitative measures such as precision and recall at the aspect level are therefore left for future work. Aspect scores depend on the wording of the descriptors and the value of k, and a broader descriptor can retrieve a different evidence set from a narrower test query. The SST-2 classifier was trained on movie reviews and may not capture every nuance of restaurant language. Finally, SHAP explanation required roughly 50 seconds for a single sentence on a CPU runtime, which is acceptable for reporting but would need acceleration for real-time use.

VI. CONCLUSION AND FUTURE SCOPE

This paper presented an explainable service analytics framework that converts customer reviews into aspect-level sentiment, prioritised recommendations and token-level explanations. By pairing dense semantic retrieval with a distilled transformer classifier, the framework isolates evidence for food, service, ambience and pricing without aspect annotations. A transparent thresholding and rule layer turns scores into actions, and SHAP attributions allow each judgement to be inspected. In the case study, the framework identified pricing as the principal area for improvement while confirming service and ambience as strengths, and its explanations revealed how contrastive words and sub-word fragments shape the model’s output. Because it relies only on open-source components and an embedded database, the solution can be adopted by small and medium enterprises at minimal cost.

Future work will extend the evaluation to larger, multi-business review collections with aspect-level annotations, such as the SemEval restaurant benchmarks, to measure accuracy quantitatively. Domain-adapted sentiment models, automatic discovery of new aspects through clustering, temporal tracking of sentiment trends and an interactive dashboard for managers are further planned enhancements. Integrating a generative language model to draft context-specific recommendations from the retrieved evidence, while retaining SHAP-based verification, is another promising direction.

VII. ACKNOWLEDGMENT

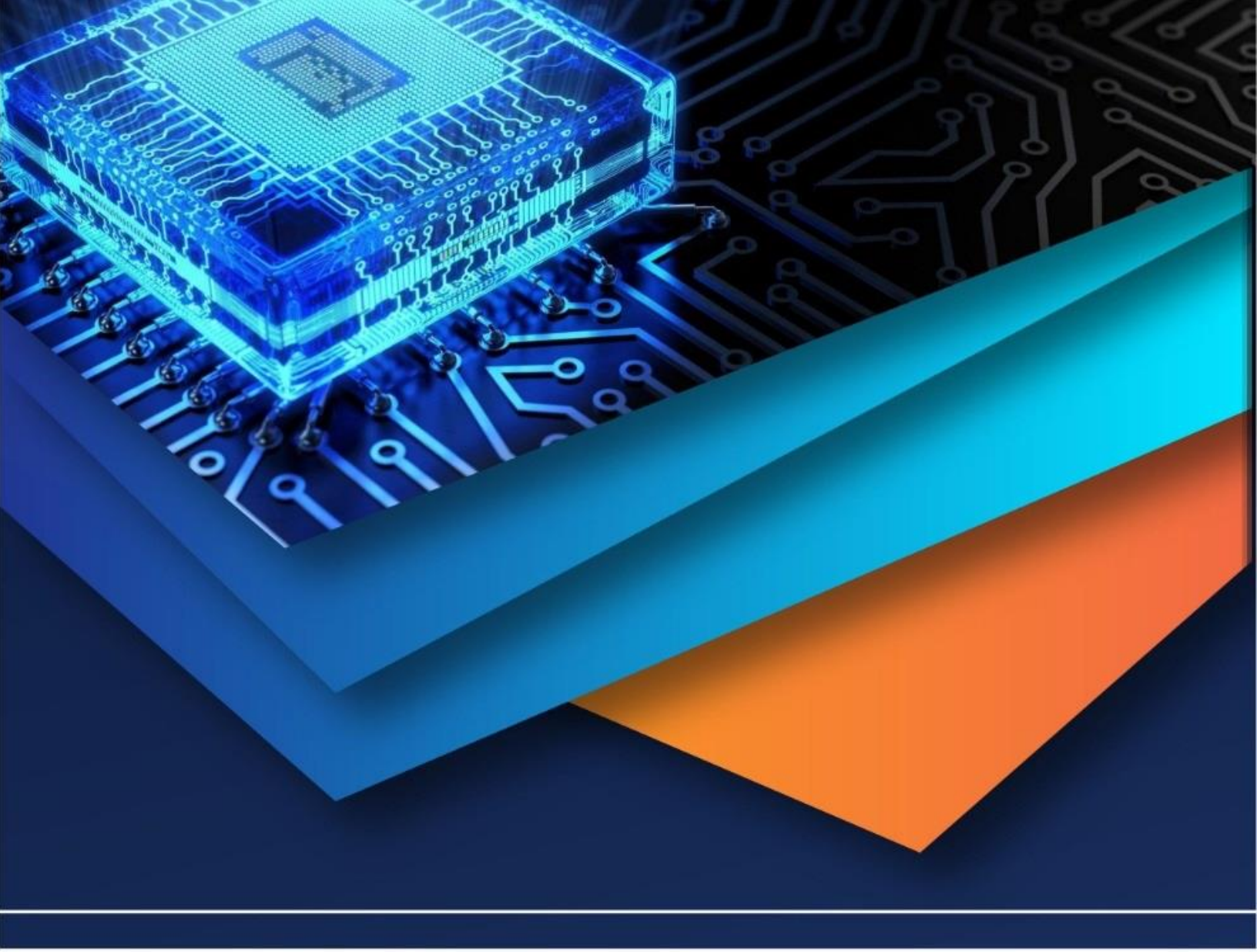
The authors thank the Management, the Principal and the Head of the Department of Computer Science and Engineering, Sanketika Vidya Parishad Engineering College, Visakhapatnam, for their encouragement and for providing the computing facilities and academic environment that made this research possible.

REFERENCES

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[2] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter,” arXiv preprint arXiv:1910.01108, 2019.

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect based sentiment analysis," in Proc. 8th Int. Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 27–35.
[4] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2004, pp. 168–177.
[5] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," IEEE Access, vol. 6, pp. 52138–52160, 2018.
[6] A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Information Fusion, vol. 58, pp. 82–115, 2020.
[7] [7] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 4765–4774.
[8] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474.
[9] V. Karpukhin et al., "Dense passage retrieval for open-domain question answering," in Proc. Conf. Empirical Methods in Natural Language Processing, 2020, pp. 6769–6781.
[10] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
[11] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in Proc. Conf. Empirical Methods in Natural Language Processing, 2013, pp. 1631–1642.
[12] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," WIREs Data Mining and Knowledge Discovery, vol. 8, no. 4, e1253, 2018.
[13] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998–6008.
[14] T. Wolf et al., "Transformers: State-of-the-art natural language processing," in Proc. Conf. Empirical Methods in Natural Language Processing: System Demonstrations, 2020, pp. 38–45.
[15] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.
[16] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 5776–5788.
[17] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[18] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, "From group to individual labels using deep features," in Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2015, pp. 597–606.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY