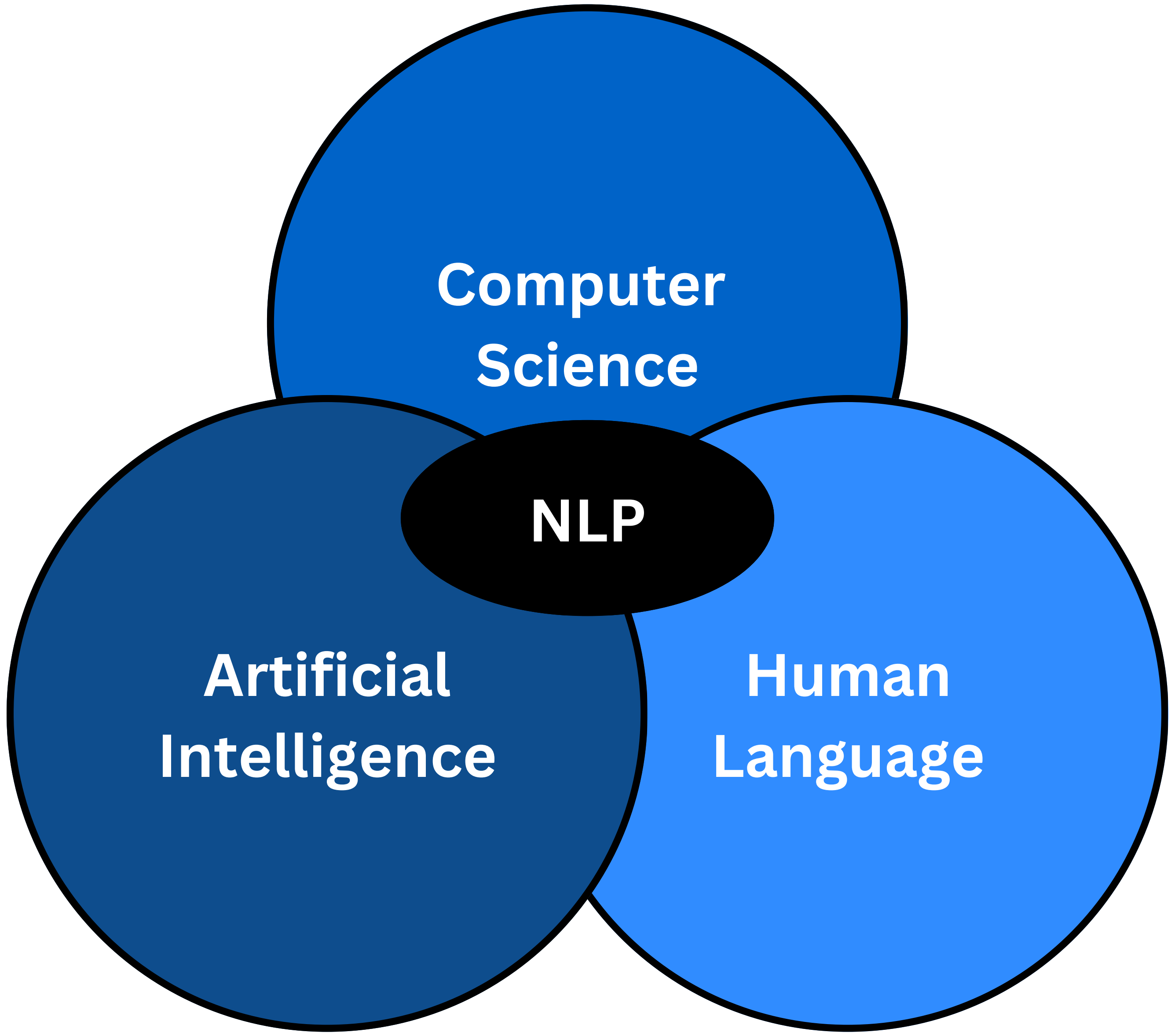
Project Proposal

# Domain-Specific Sentiment Analysis in Textual Feedback: AI-Powered Aspect Detection and Sentiment Analysis for Textual Feedback and Reviews


Isuranda M.A.K. EG/2019/3607
Jayakody D.S. EG/2019/3609
Malkith A.V.A. EG/2019/3664

# Introduction

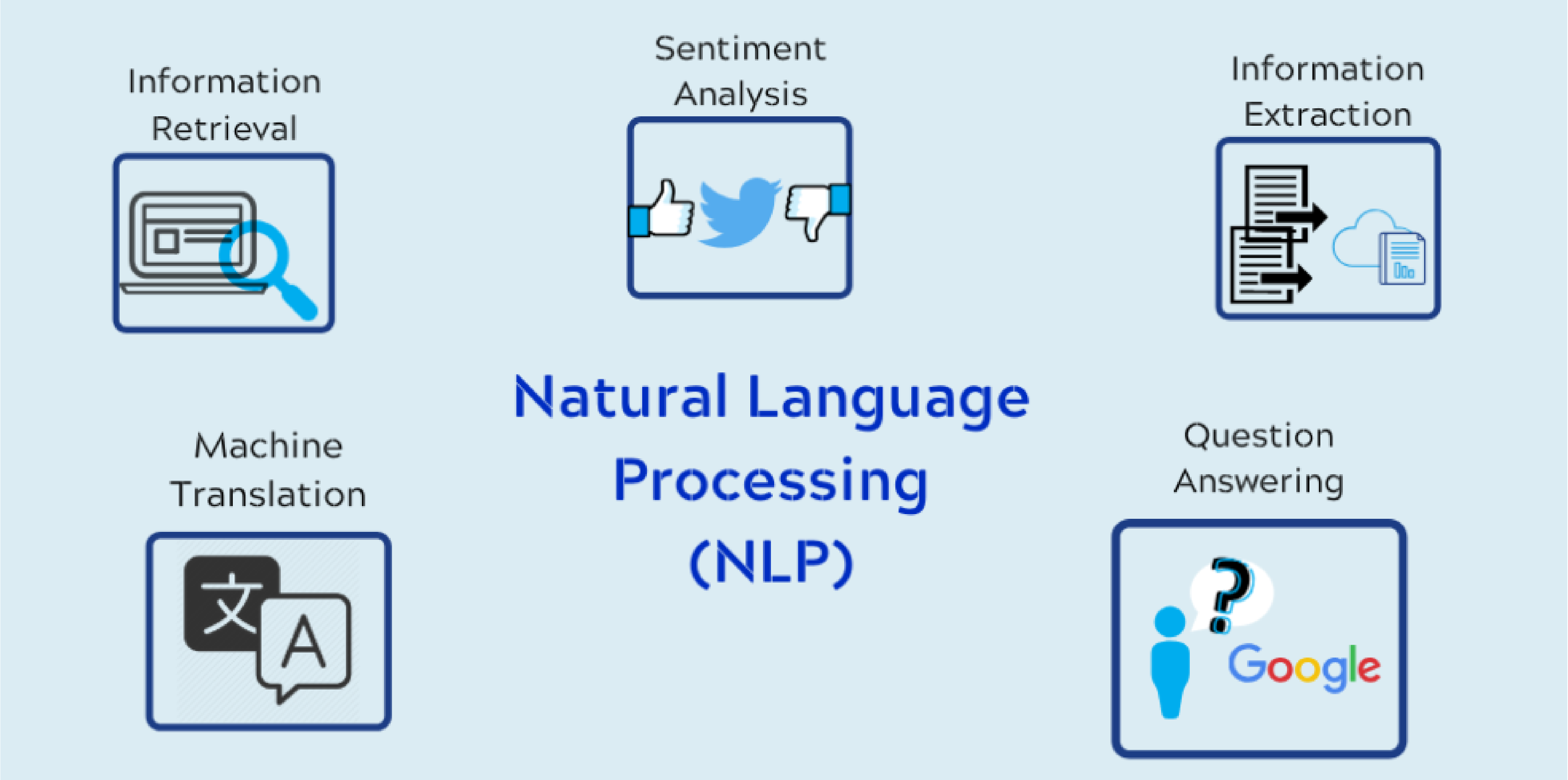
## What is NLP




Natural language processing (NLP) refers to the branch of computer science and  the branch of Artificial Intelligence concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

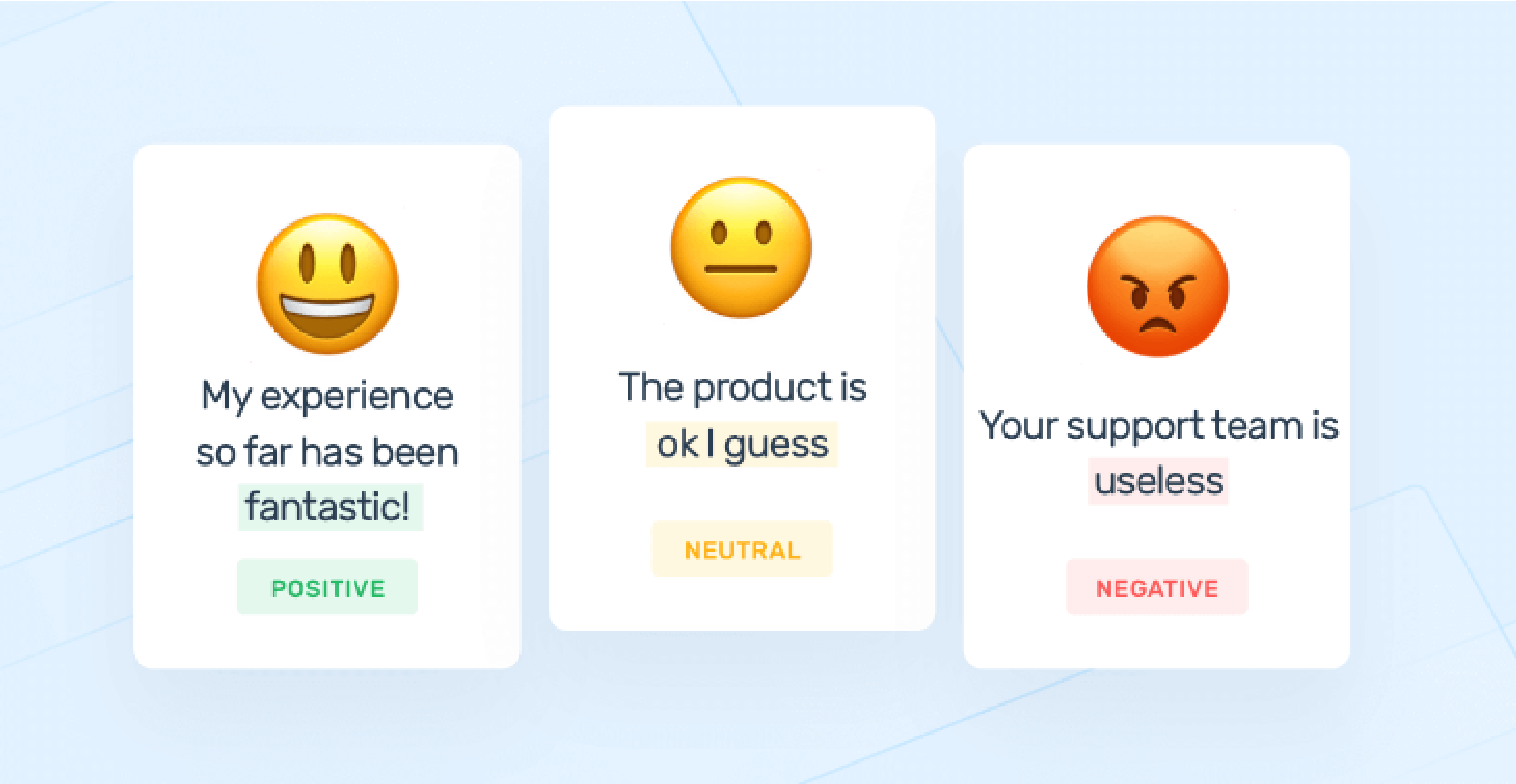
# Applications of NLP

- Language Translation
- Search Results
- Text Analytics
- Chatbots
- Information retrieval
- Sentiment analysis

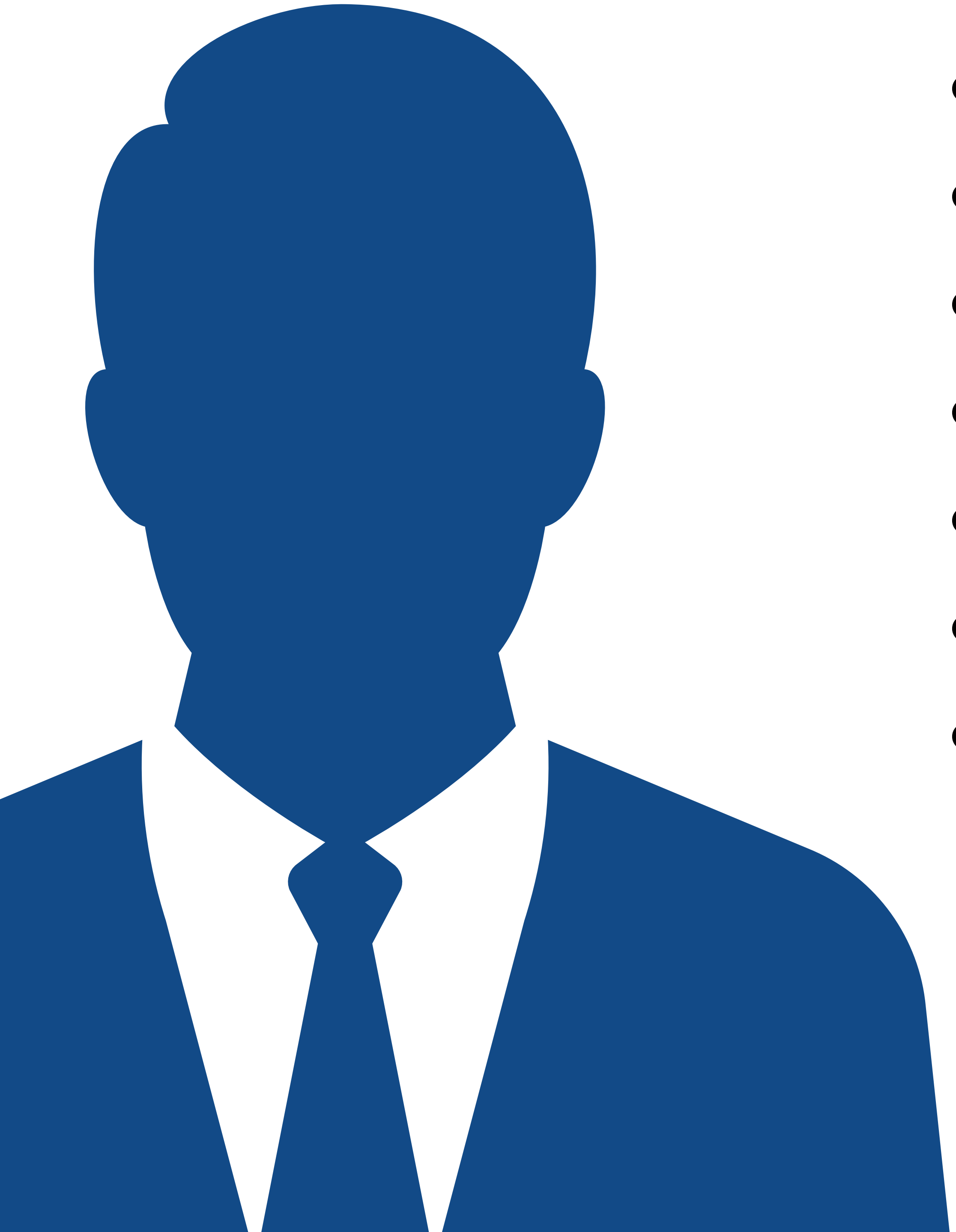


# What is Sentiment Analysis?

Sentiment analysis is a machine learning and NLP technique used to determine the overall sentiment expressed in a piece of text , such as a customer review , social media post or email



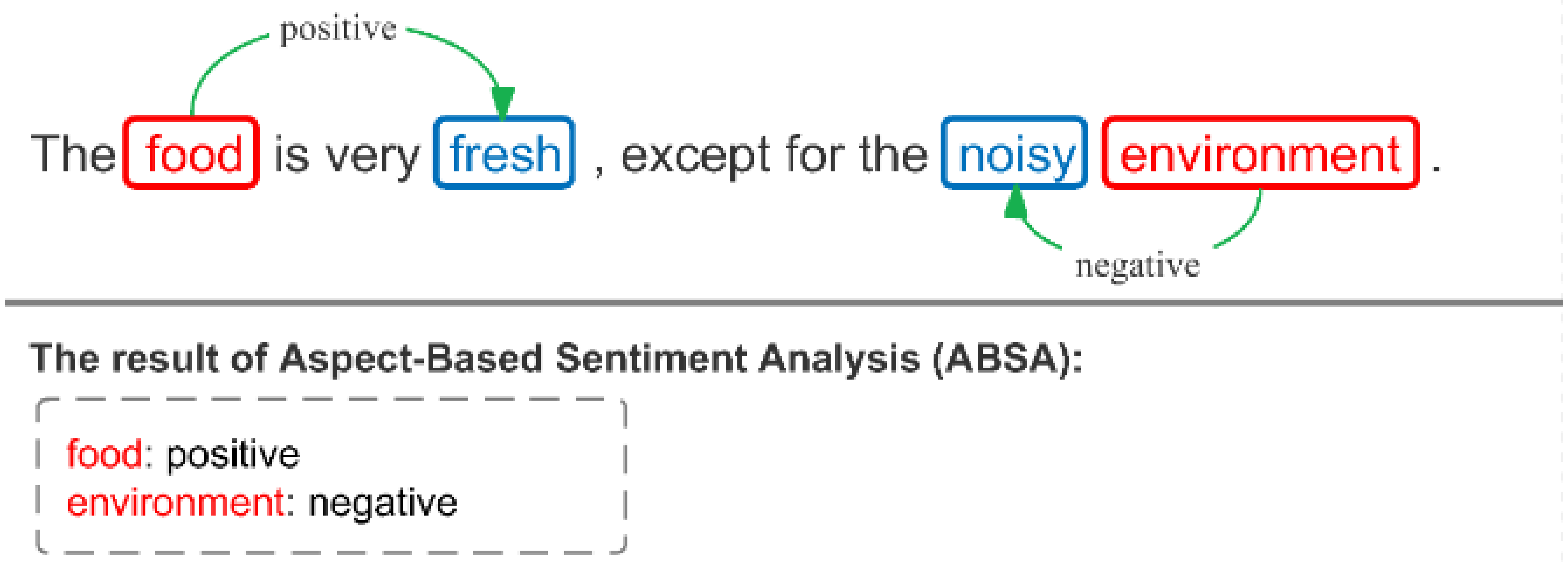
# Benefits of Sentiment Analysis



- Allows business to listen to their customers
- Recognizing business’s consumer base
- Understand customer needs
- Developing and accessing a marketing campaign
- Improving customer service
- Crisis management
- Increasing revenue from sales
- Can maintain business’s brand reputation and loyalty

# What is Aspect Based Sentiment Analysis ?

Aspect-based sentiment analysis(ABSA): A text analysis technique that divides the text data and defines its sentiment based on its aspects. It analyzes consumer feedback data by correlating emotions to different aspects of a product or service



# Project Objectives

- Investigate and assess the key differences, advantages, and disadvantages of various sentiment analysis techniques
- Develop a model that can achieve a high accuracy in recognizing aspects from sentences related to a specific domain within textual reviews.
- Build a machine learning model with a high accuracy to identify the polarity of sentences in sentiment analysis
- Develop a user-friendly interface that allows for the intuitive display of final results

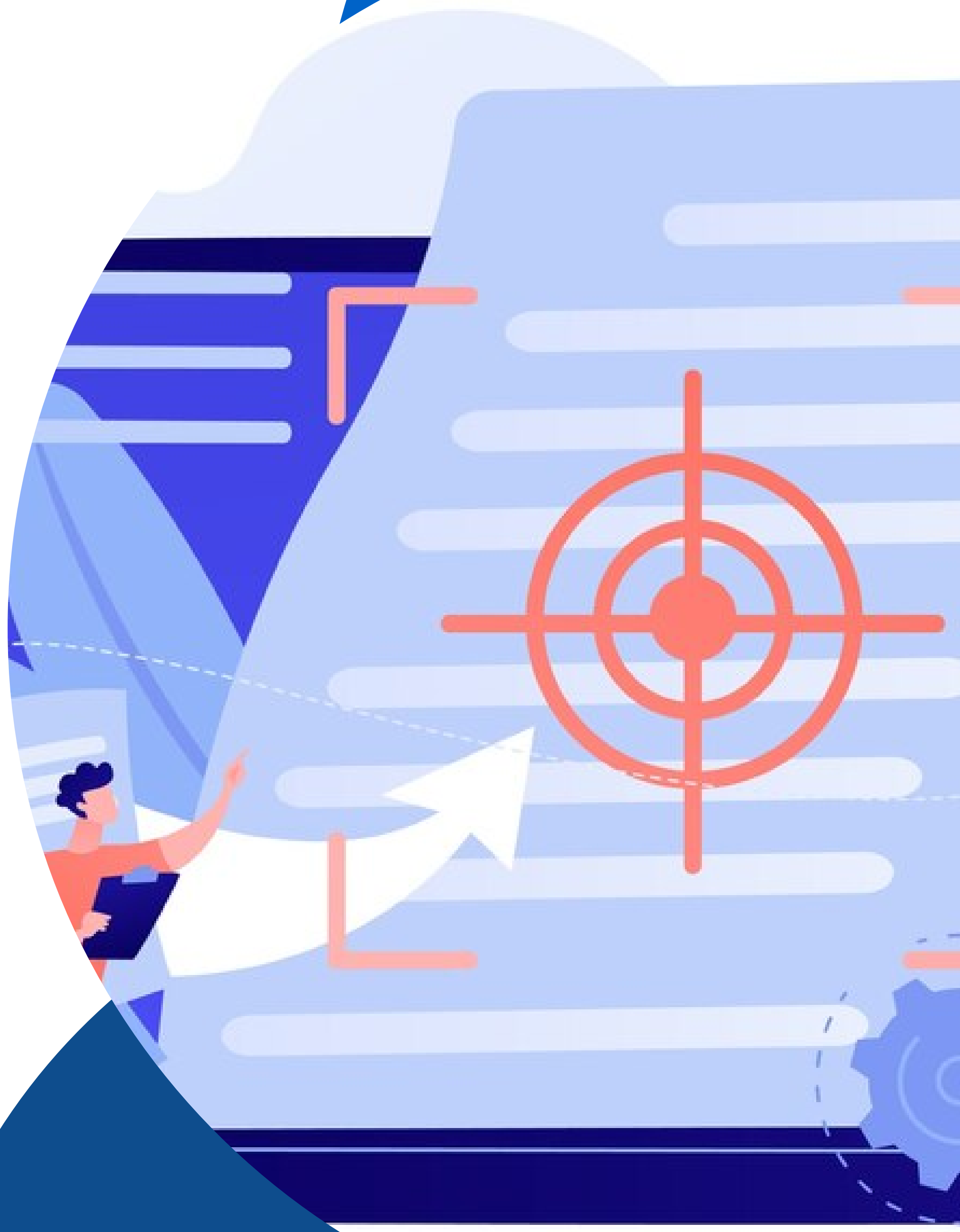
# Project Scope

- The aspects for a certain given domain are decided by the client.

- The language scope of our project will be limited to English.

- Each review under consideration will pertain to a specific domain, ensuring that aspects and sentiments analyzed are relevant and confined to that particular domain.

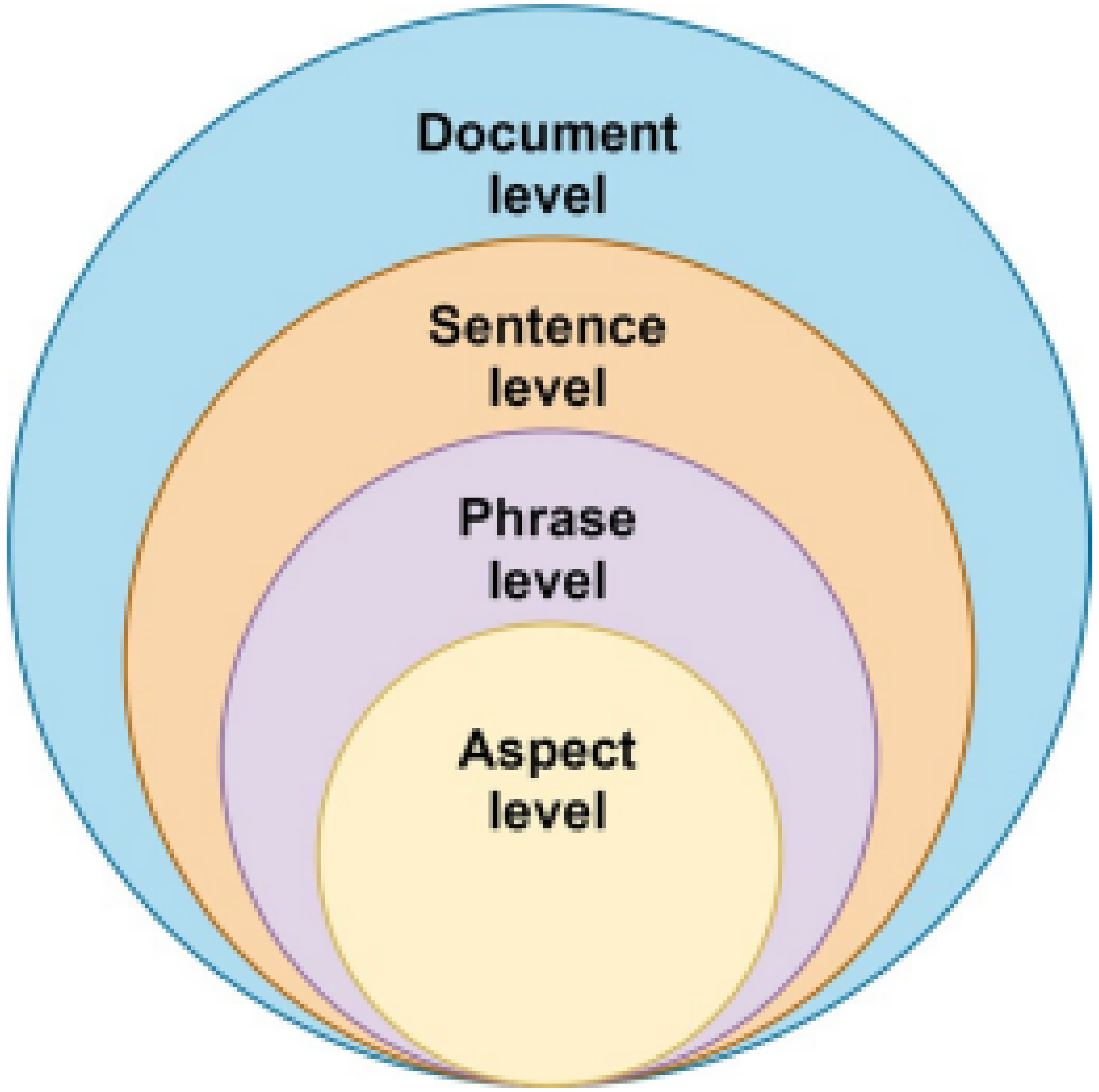
- The sentiment intensity for a certain aspect of a domain will be categorized as positive, neutral and negative.



# Levels of Sentiment Analysis

**Document-level**
Document-level sentiment analysis involves evaluating the sentiment of an entire document and assigning a single polarity to the entire content.

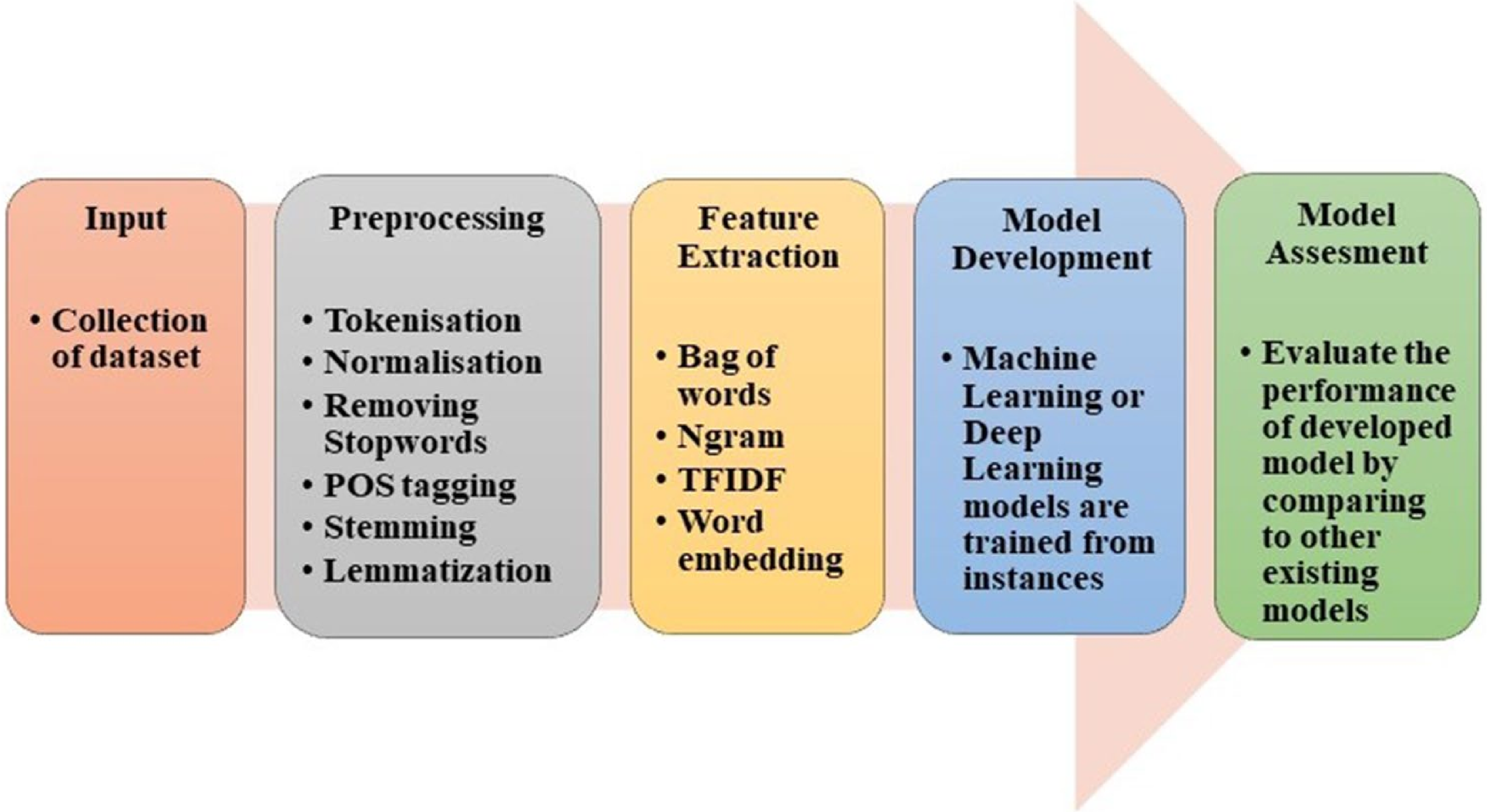
**Phrase level**
Sentiment analysis can also be conducted by identifying opinion words at the phrase level and subsequently classifying them.




**Sentence-level**
In sentence-level analysis, every sentence is examined to determine its polarity, resulting in a corresponding sentiment classification.

**Aspect level**
Sentiment analysis is performed at the aspect level. [1]

# Process of Sentiment Analysis



# Preprocessing

Preprocessing in sentiment analysis refers to the steps taken to clean and prepare text data before it is analyzed for sentiment.

- **Tokenization:** Tokenization commonly involves splitting text into words or subwords.
- **Normalization:** Normalization in the context of natural language processing refers to the process of standardizing text data to a common format or representation. This typically involves converting text to lowercase, removing punctuation, special characters, and so on.
- **Removing stopwords:** Removing stop words in NLP means getting rid of common words like "and," "the," "is," and so on. Because they usually don't add much meaning to the text. [2]

# Preprocessing

- **Part-of-Speech tagging:** This is a process in NLP where words in a text are assigned a specific part of speech (e.g., noun, verb, adjective, adverb) based on their context and grammatical relationships within a sentence.
- **Stemming:** Stemming is a process where words are reduced to their root or base form, called the "stem."
- **Lemmatization:** Words are transformed to their base or dictionary form in lemmatization. [2]

# Feature Extraction

Feature extraction in sentiment analysis involves identifying and selecting relevant aspects or characteristics of text data that are indicative of sentiment.

- **Bag of Words:** BoW model is a basic technique for representing text data as a collection of words, disregarding grammar and word order. In this model, a document is represented as a "bag" (multiset) of its constituent words, along with their frequency of occurrence in the document.
- **N-gram:** N-grams are commonly used to capture the local context and dependencies between words in a text. [3]
- **TFIDF:** TF-IDF stands for "Term Frequency-Inverse Document Frequency." It's a statistical measure to evaluate the importance of a word in a document relative to a collection of documents. [4]

# Feature Extraction

## Word Embedding

Word embedding is a technique used in NLP to represent words as dense, continuous vectors in a high-dimensional space, where each dimension represents a different aspect or feature of the word's meaning. Word embeddings are typically learned from large amounts of text data using some techniques.

- **Word2Vec:** word2vec is a 2-layer neural network that is used for vectorizing the tokens.
- **GloVe (Global Vectors):** Glove is used to generate dense vector representations of words in a high-dimensional space.
- **FastText:** FastText uses a linear classifer to train the model, which is very fast in training the model. [5]

# Model Development

Model development can be done using lexicon-based, machine learning, deep learning, and transfer models.

**Lexicon-based Models:**

- Use pre-made word lists to decide if a text is positive, negative, or neutral.
- Lexicon-based models are simple and interpretable but may lack context sensitivity and struggle with sarcasm or nuanced language. [6]

**Machine Learning Models:**

- Train computers to recognize sentiment patterns from labeled datasets.
- Common machine learning algorithms used for sentiment analysis include Support Vector Machines (SVM), Naive Bayes, Logistic Regression, and Random Forests. [1]

# Model Development

**Deep Learning Models:**

- Deep learning models, particularly neural networks, have shown remarkable performance in sentiment analysis tasks.
- Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks are commonly used architectures for sentiment analysis.
- Deep learning models can automatically learn feature representations from raw text data, enabling them to capture complex semantic relationships and contextual information. [7]

**Transfer Learning Models:**

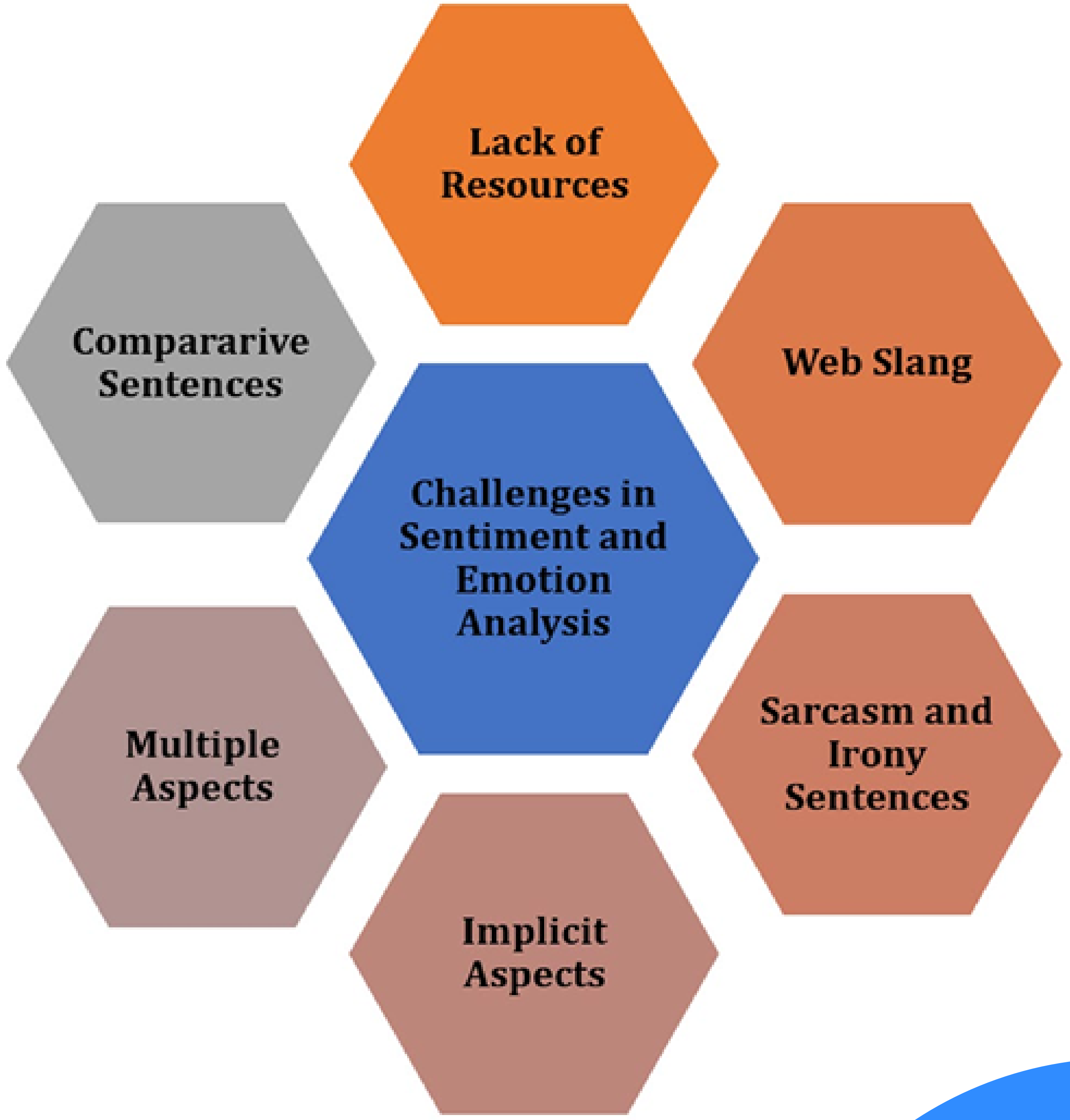
- Start with models trained on lots of data and adjust them to understand sentiment.
- Models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) can be fine-tuned on specific sentiment analysis tasks with smaller datasets, leading to improved performance.
- Benefit from previous learning and need less data to perform well. [1]

# Applications of Sentiment Analysis

1. **Business analysis**
    a. Product reviews
    b. Market research and competitor analysis
2. **Healthcare and medical domain**
    a. Reputation management
    b. Review analysis
    c. Customer reviews
3. **Review analysis**
4. **Stock market**
5. **Social media monitoring**

# Challenges in Sentiment Analysis

1. **Lack of Resources**
2. **Web Slang**
3. **Sarcasm and Irony Sentences**
4. **Implicit Aspects**
5. **Multiple Aspects**
6. **Comparative Sentences**



# Methodology

**1) Extensive research on aspect based sentimental analysis**

Conducting a thorough **literature review and research exploration** on aspect based sentiment analysis forms the foundational stage of this project. This involves a c**omprehensive examination of existing methodologies, algorithms, and models in the field.**

The aim is to understand the **current state-of-the art techniques**, **identify gaps in the literature**, and **determine the most suitable approaches for the domain-specific sentiment analysis of textual feedback.**

# Methodology

**2) Data Collection**

This includes **collecting a diverse set of textual feedback and reviews in English** to ensure the model's applicability across various linguistic contexts. The dataset should be representative of the domain under consideration, encompassing various aspects and sentiments expressed by users.

**Emojot (pvt) Ltd has agreed to hand us a real world public data set which they have been using in their platforms.**

# Methodology

**3) Data Preprocessing:**

Since data from social media platforms are **highly unstructured**, pre-processing plays a crucial role in enhancing the quality of sentiment analysis.

Techniques such as **tokenization, normalisation, stop word removal, POS tagging, stemming, and lemmatization** are employed.

These techniques aim to structure the data and eliminate irrelevant information, preparing it for further analysis.

# Methodology

**4) Aspect Based Sentimental Detection:**

Aspect-Based Sentiment Detection is hoped to be implemented using **state-of-the-art Transformer-based models**. Conventional neural ABSA models, while effective, have faced limitations in capturing complex sentiment dependencies due to context-independent word embeddings.

To address this, we hope to integrate pre-trained language models (PLMs) like **BERT and RoBERTa**, which have demonstrated significant improvements across various NLP tasks. We hope to employ the transformer architecture for its ability to capture long-range dependencies and contextual information effectively.

# Methodology

**5) Model Training:**

Training the Aspect-Based Sentiment Detection model involves **fine-tuning the pre-trained transformer model on the specific ABSA task.**

The **pre-trained transformer model's parameters are hoped to be fine-tuned using the collected and preprocessed dataset**. This process helps the model learn domain specific patterns and sentiment dependencies present in the textual feedback.

The utilization of transformer models ensures that the sentiment analysis model can capture intricate relationships between aspects and sentiments, providing a clear solution for Emojot (pvt) Ltd's domain.

# Methodology

**6) Model Evaluation:**

To assess the effectiveness of the t**rained models, rigorous evaluation metrics will be employed.**

These metrics include but are not limited to **accuracy, precision, recall, and F1 score.** The evaluation will be conducted using a **separate test dataset that was not seen by the models during the training phase.**

This step is crucial in determining the model's generalization capabilities and its performance on unseen data.

# Methodology

**7) Deploying and demonstration**

Once the model is trained and evaluated satisfactorily, the next step is to deploy the **sentiment analysis model for real-world applications**.

This involves integrating the model into **Emojot (pvt) Ltd's platforms**, allowing it to analyze and categorize textual feedback in real-time.

A **demonstration will be prepared** to showcase the model's functionality, emphasizing its accuracy and efficiency in providing aspect-based sentiment analysis results.

# References

[1] Wankhade, M., Rao, A.C.S. and Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), pp.5731-5780.
[2] Nandwani, P. and Verma, R., 2021. A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1), p.81.
[3] Razon, A. and Barnden, J., 2015, September. A new approach to automated text readability classification based on concept indexing with integrated part-of-speech n-gram features. In Proceedings of the international conference recent advances in natural language processing (pp. 521-528).
[4] Liu, Y., Wan, Y. and Su, X., 2019. Identifying individual expectations in service recovery through natural language processing and machine learning. Expert Systems with Applications, 131, pp.288-298.
[5] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5, pp.135-146.
[6] Z. Yan-Yan, Q. Bing, and L. Ting, “Integrating intra-and inter-document evi dences for improving sentence sentiment classification,” Acta Automatica Sinica, vol. 36, no. 10, pp. 1417–1425, 2010.
[7] Bandara, K., Bergmeir, C. and Smyl, S., 2020. Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert systems with applications, 140, p.112896.

# Thank You

For Your Attention