Università di Pisa

Sentiment Classification with Neural Text Representation on Amazon Reviews

09/05/2025

Andrea Lepori, Gemma Ragadini, Mihnea Molnar, Pan Zhang, Loris Giunta

- Group 4 -

The problem

SENTIMENT ANALYSIS: Process of identifying, extracting, and classifying the opinions, thoughts, and impressions expressed in a text.

Interprets the **polarity** of a text

Applications: identifying customer-preferred products, managing brand reputation, conducting market research, etc.

Sarcasm, irony, informal writing, words that change meaning depending on the context.

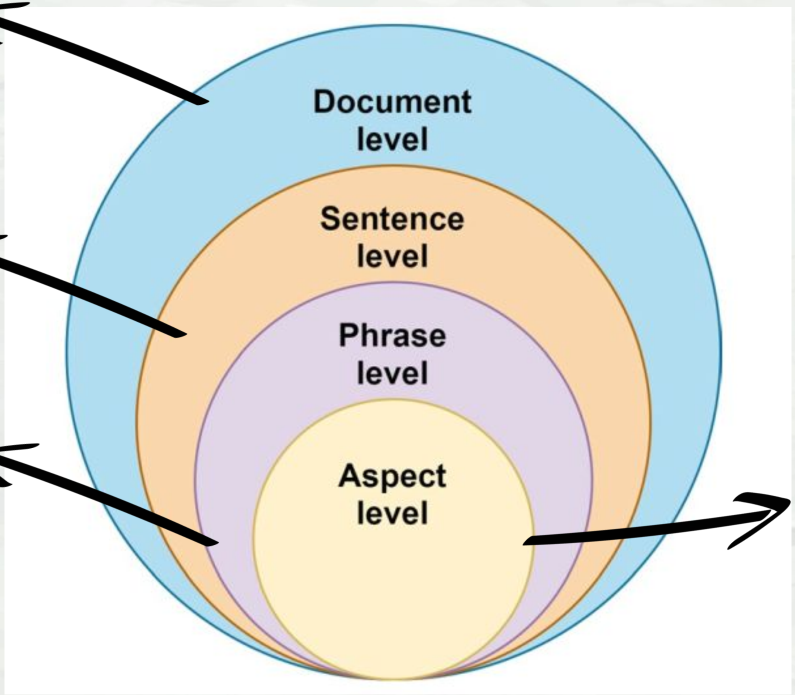
Sentiment analysis levels

Polarity assigned to the entire document.

cross-domain and cross-language analysis

"I love this restaurant!"

"The service is fast, but the food is mediocre."



"The camera is excellent, but the battery drains too quickly."

Feature Extraction

- **Terms frequency**
  Considering how often a term appears in a text by counting it with an integer value.
- **Parts of Speech tagging**
  Grammatically tagging each word in a text based on its definition and context.
- **Bag of Words (BoW)**
  representing each sentence as a vector based on the frequency of words from a fixed vocabulary, ignoring the syntactic structure and word order.
- **Word Embedding**

Word Embedding

Representation of words in a fixed-dimensional vector space.

Word2Vec uses the weights learned by a binary classifier (based on the probability that one word appears near another) as embeddings for the words.

SGNS predicts the context given a central word

CBOW predicts the central word given the contex

FastText extends Word2Vec by incorporating n-grams (again SGNS and CBOW)

Global Vectors (GloVe) does not predict words, but builds a co-occurrence matrix and learns embeddings by optimizing weights to represent semantic relationships.

Feature Selection

A characteristic can be insignificant, significant, or redundant. Feature selection chooses the most relevant features in order to eliminate the others, improve accuracy, and reduce noise.

Filter approach selects features without using machine learning, based on statistical metrics from the training data (low computational cost).

Wrapper approach is based on machine learning algorithms (more computationally expensive, but capable of finding the optimal feature set).

Embedded approach selects features during model training, integrated within the algorithm.

Hybrid approach combines filter and wrapper approaches

Methodology

Three mainly used approaches for Sentiment Analysis:

- **Lexicon Based Approach**
- **Machine Learning Approach and Neural Networks**
- **Hybrid Approach**

Researchers are continuously trying to figure out better ways to accomplish the task with **better accuracy** and **lower computational cost**.

Lexicon Based Approach - I

«**Lexicons** are the collection of tokens where each token is assigned with a **predefined score** which indicates the **neutral**, **positive** and **negative** nature of the text» (*Kiritchenko* et al. 2014).

The **score** can be assigned:

- *Simple way*: **–1** (**negative**), **0** (**neutral**), **+1** (**positive**) "*...good...*, +1", "*....excellent...*, +1"
- *Use a range*: [**–1** (**higly negative**), **+1** (**highly positive**)] – "*...good...*, 0.6", "*....excellent...*, 0.9"

PIPELINE:

TOKENIZATION → **SCORE** → **TEXT OVERVIEW**

TOKENIZATION: **Divide te text into tokens**

SCORE: **Assign a score to every single token**

TEXT OVERVIEW: **Sum the scores of every single token to determine the overall polarity of the text**

Lexicon Based Approach - II

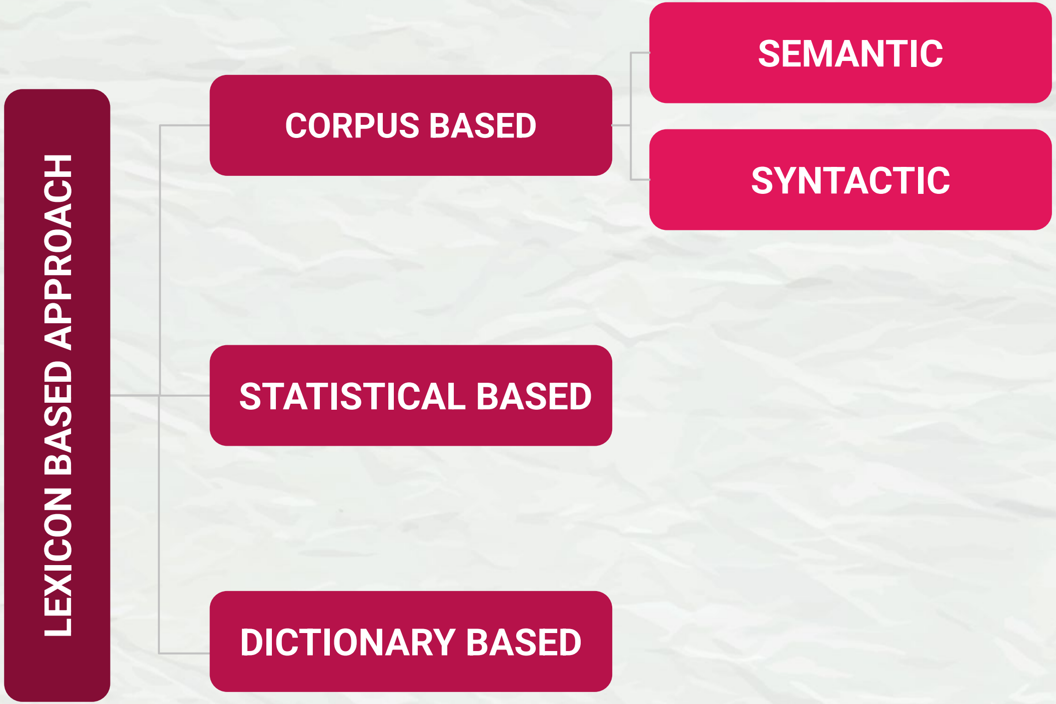
- **PROS**
  - It’s considered an **unsupervised method**
  - Not requiring any training data

- **CONS**
  - Highly domain dependent

- Word: *huge*
  - Phrase 1: “The queue for the movie was **huge**”
  - Phrase 2: “There was a **huge** lag in the network”

… Need to build **domain-specific sentiment lexicons**.

Lexicon Based Approach - III



- **Corpus Based:**
  This method expands an initial set of sentiment words by analyzing large text corpora using **syntactic and/or semantic patterns**. it considers how new words appear in context using grammatical structures to determin their sentiment orientation. Need graph models and labeled data (eg. POS-tagging).

- **Statistical Approach:**
  Uses **frequency** and **co-occurrence** patterns to **infer sentiment.** If a word frequently appears in positive or negative texts, it's assumed to share the same sentiment.

- **Dictionary Approach:**
  Uses a manually crafted dictionary of words with predefined sentiment labels (eg. positive, negative, neutral) or scores.

Machine Learning Approach - I

Machine Learning Algorithms can be used to categorize sentiments.
There are two primary ML approaches to sentiment analysis:

- **Supervised method:**
  - The algorithms **need to be trained** on a training set (**labeled data**).
  - After the training the algorithm can be applied to our data.
  - The model is able to predict a class label for a given instance (unseen)
- **Unsupervised method:**
  - Uses lexicons, knowledge bases, and ontologies, datasets, without labeled data.
  - All this data has been selected before specifically for sentiment analysis.
- **Keypoints:**
  - S.A. can be done using both methods but supervised methods are more commonly used due to their accurate results.
  - Sentiment classification treated as a text classification problem.
  - Utilizes syntactic and linguistic features.
  - Models predict sentiment labels for unseen data.

Machine Learning Approach - II

TECHNIQUE	ADVANTAGE	DISADVANTAGE
SVM - Support Vector Machine	• Most popular algorithm for SA • Gives good accuracy for large datasets	• Fine tuning could be difficult • Long time training for large datasets
LR - Logistic Regression	• Simplest model to implement and training	• Accuracy can be low for complex dataset
DT - Decision Tree	• Easy to construct • Less time for training • No big dataset required for training	• Model is more prone to overfit • Highly domain-oriented

Other strategies could be **Naive Bayes**, **Maximum Entropy** and so on…

Hybrid Approach

- Hybrid approach combines **Machine Learning techniques** and **Lexicon-Based methods** in sentiment analysis tasks to **leverage the strengths of both**.
    - **Lexicon-based methods** help capture semantic and linguistic nuances, and can reduce feature complexity.
    - **ML models** learn to generalize and detect patterns beyond the predefined vocabulary.
- **Emerging area of research** in the field of Natural Language Processing.

HYBRID APPROACH

BENEFITS

- **Higher accuracy** than using either methods alone.
- **Reduce features dimensionality** while maintaining **performance**.

DISADVANTAGES

- **More complex to implement.**
- its e**ffectiveness heavily depends** on the quality of the training data.

Neural network

Neural networks learn complex, non-linear relationships between text and sentiment.

- ANN and SVM: ANN generally outperform SVM in sentiment analysis tasks, except in cases involving limited or imbalanced data.

- RNN and its variants(LSTM, GRU, Bi-LSTM): The advantage of RNN is that it use previous information, but it suffer from vanishing and exploding gradient descent; RNN variants like LSTM, GRU, and Bi-LSTM address these limitations and are widely used in sentiment analysis.

- Attention mechanism and Transformers (e.g., BERT and GPT): Attention mechanism are being introduced, Transformers leverage attention mechanisms to efficiently capture global contextual information in parallel, enabling state-of-the-art performance in sentiment analysis. Transformers-baesd models like BERT and GPT are already trained on a massive corpus. Weights can be fine-tuned using the training dataset to get accurate results.

Other approaches

- **Aspect based sentiment analysis (ABSA)**
- **Transfer learning**

Aspect based sentiment analysis (ABSA)

Aspect based sentiment analysis is defined as the process of identifying sentiment aspects in a text by focusing on specific aspects, sentiment polarity, and opinion terms related to the text. ABSA is composed by three critical phases.

1. **Aspect detection**
2. **Polarity or sentiment categorization**
3. **Aggregation**

Challenges: Aspect-level sentiment analysis is challenging due to implicit aspect detection, precise polarity classification, and the need for advanced models like LSTM and BERT, with cultural and contextual factors adding complexity.

Transfer learning

Transfer learning is one of the advances techniques in AI, where a pre-trained model can use its acquired knowledge to transfer to a new model. The new model directly uses the previously learned features without needing any explicit training data. Training data may be used to fine-tune the model to a new task.

- Pretrained models like BERT, RoBERTa, and GPT-2 provide contextual embeddings.
- Methods include feature extraction and fine-tuning.
- Feature extraction: use pretrained embeddings as input to a classifier.
- Fine-tuning: retrain the entire model on sentiment-specific data.

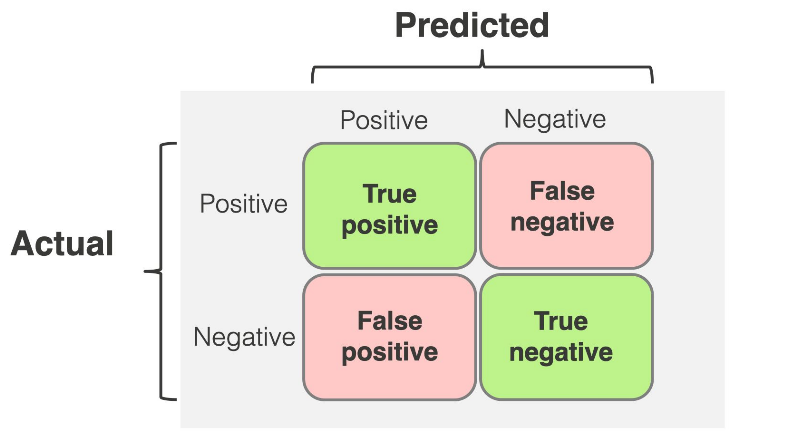
Multimodal sentiment analysis (MSA)

Multimodal sentiment analysis(MSA) adds a new level to standard text-based sentiment analysis by incorporating additional modalities such as audio and visual data.

Several studies have attempted to discern sentiment analysis in social multimedia using a variety of multimodal inputs, including visual, audio, and textual data. Such as a video portraying a person discussing a product or a movie. Typically, spoken transcripts are examined separately from face and voice expressions, and the results of unimodal, text-based sentiment analysis are combined in post to create a "MSA" system. MSA is a rapidly expanding area of study. A key area of opportunity in this subject is to enhance the mechanism of multimodal fusion. Unlike traditional emotional learning tasks that require the use of single modalities (text, speech), multimodal learning makes use of many sources of information, including language (text/transcripts/ASR), audio/acoustic, and visual modalities.

Performance evaluation parameter

Confusion Matrix:



$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

- most commonly used metric
- good metric for a **balanced** dataset

$$Precision = \frac{TP}{TP + FP}$$

Out of **predicted positives**, how many were correct?

$$Recall = \frac{TP}{TP + FN}$$

Out of **actual positives**, how many were caught?

$$Specificity = \frac{TN}{TN+FP}$$

Out of all the **actual negatives**, how many did the model correctly label as negative?

$$F_1 = \frac{2TP}{2TP+FP+FN}$$

– **harmonic mean of Recall and Precision**
– most used metric after accuracy

Applications of sentiment analysis

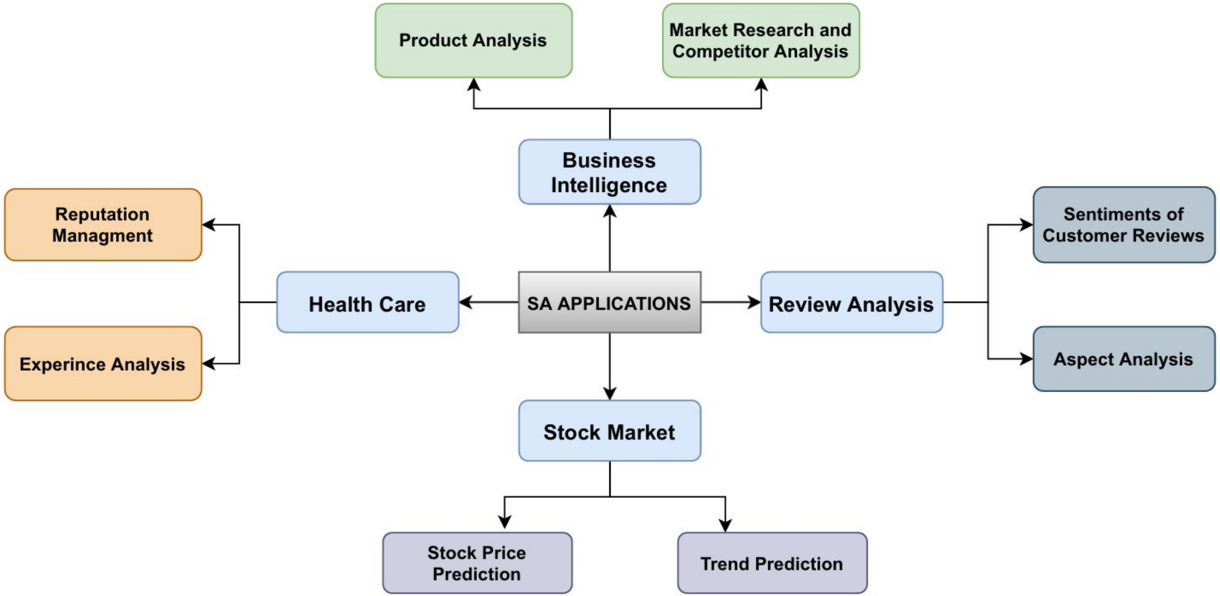


Fig. 8 Applications of sentiment analysis

A double advantage:

*"**Businesses** can use the results of Sentiment Analysis to **make product enhancements**, **examine consumer feedback**, or **develop a new marketing plan**…**to examine customers perceptions of products or services**"*

*"These studies, however, are not limited to product producers; **consumers** may use them to **review items** and **make more informed decisions**."*

*" … **Aspect-based sentiment analysis** done on hotels and restaurants will help i**dentify the aspect** with the most positive reviews and negative reviews, on which Hotels can work and make it better… **The aspect-based method will enable companies to extract the most important aspects of client feedback and service**."*

Structures of sentiment reviews

Structured sentiments

- These are **formal** and **well-organized** reviews
- They are more **focused on formal problems** such as **books or research**
- The authors are **professionals** capable of writing **thoughts** or **observations** concerning **scientific** or **factual concerns**

"The theoretical foundation of the book is weak, although the experimental setup is impressive."

Semi-Structured Sentiments

- These are **somewhat organized**, but **not fully formal**.
- Often seen in user reviews where people **list Pros and Cons separately**
- Usually **short sentences** or **phrases**

"Pros:
- Fast processor
- Good battery life
Cons:
- Overheats
- Poor camera in low light"

Unstructured Sentiments

- **Free-form text**, **informal writing**, very common in social media, blogs, or general product reviews
- The sentiment is **mixed**, **scattered**, or **embedded inside unrelated content**
- **Explicit features**: directly mentioned → "The screen is amazing." (feature = *screen*)
- **Implicit features**: not mentioned by name but implied → "It is extremely pricey." (feature = *price*)

"I bought this phone last week. The screen is amazing, but the battery drains too fast. I wish it came with a charger too. Not impressed overall."

Challenges in sentiment analysis (I)

Sarcasm To deal with **ambiguos situations** and **irony** are key difficulties in sentiment analysis

- *"Oh great, another update that makes the app even slower. Just what I needed!"*

A **sarcastic remark** about an object is intended to communicate a **negative sentiment**; yet, **conventional sentiment analysis algorithms frequently miss this meaning.**
Various method has been proposed for detecting sarcasm in language, however the problem is far from resolved as **comedy is very culturally particular**, and it is challenging for a machine to understand unique (and frequently fairly detailed) cultural allusions.

Informal style of writing is the biggest challenge to all NLP tasks as people are very casual about writing reviews or texts, they tend to use:

- **acronyms, emojis, shortcuts**

Acronyms can be handled if they are universal. There are a lot of regional acronyms which change and grow day by day.

Grammatical errors are very common in informal texts and can be handled up to a certain extent. The accuracy of sentiment analysis and NLP tasks may be improved if these errors can be corrected

Challenges in sentiment analysis (II)

Computational cost To get better accuracy, we need to increase the training data size and make the model more complex, which will exponentially increase the computational cost of the model for training; high-end GPU may be required to train a model with a huge corpus. Especially neural networks and attention models have shown that they are computationally costly

Availability of data The training data of one domain may not be applicable and valuable to other domains. For instance, a model trained on a hotel review dataset is not helpful in predicting sentiments of a stock or mutual fund dataset and viceversa

Adaptations of language consider English language, which is widely spoken worldwide, but it is seen that many English varieties are spoken worldwide based on the regions like Indian, American, British, etc. Lots of words are used differently depending upon the region there are used or different spellings for the same word

Phrases containing degree adverbs and intensifiers Consider review r1= "*The food is barely good*" and r2= "*the food is really good*". r1 is considered neutral or slightly positive, whereas r2 is considered to be highly positive. The adverbs 'barely' and 'really' decide the extent of positiveness and the word 'good'. Similarly, intensifiers (like "*very*" or "*too*") also quantify the sentiment of the sentences

Mixed Code Data Code-mixing is the employment of vocabulary and grammar from different languages in same sentence. This phenomenon is becoming increasingly common as communication between groups of people who speak different languages grows. The lack of a formal grammar for code-mixed phrases makes it challenging to identify compositional semantics which is critical if we are using e.g. Machine Learning based techniques

Project Overview

Goal: Classify sentiment in Amazon product reviews.

Task: Compare neural models using **contextual vs static embeddings.**

Focus: Binary sentiment classification (positive vs negative) as a starting point.

The Chosen Dataset

Title: *"Amazon Review Full Score Dataset"*

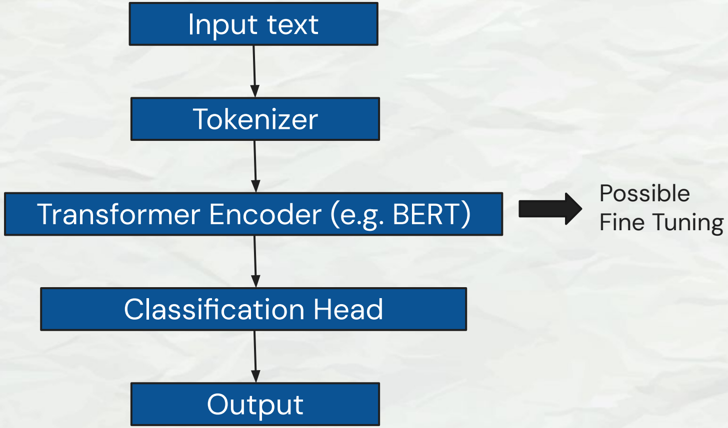
Content:

- Dataset: 3M training + 650K testing samples
- It includes: review title, review text and rating (1-5)
- Available at: https://drive.google.com/file/d/0Bz8a_Dbh9QhbZVhsUnRWRDhETzA/view?usp=share_link&resourcekey=0-Rp0ynafmZGZ5MfIGmvwLGg

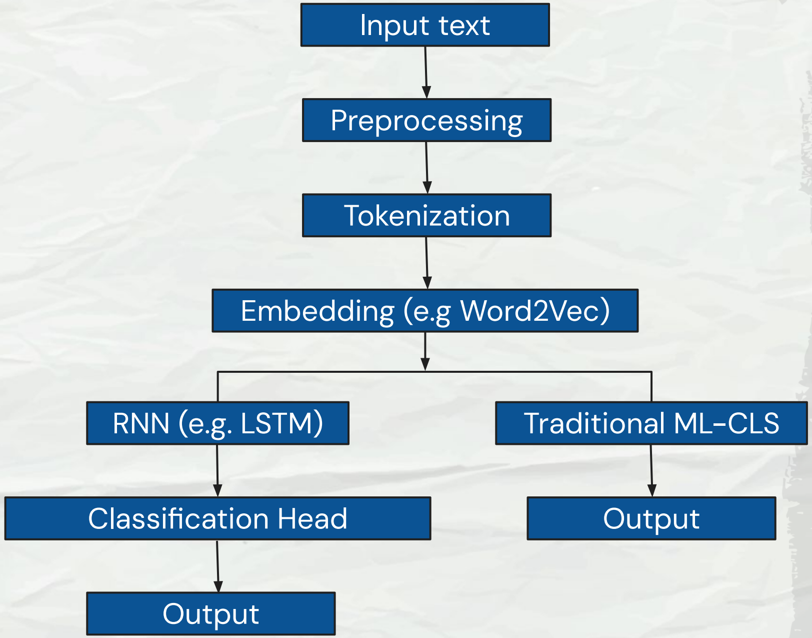
	label	title	text
0	3	more like funchuck	Gave this to my dad for a gag gift after direc...
1	5	Inspiring	I hope a lot of people hear this cd. We need m...
2	5	The best soundtrack ever to anything.	I'm reading a lot of reviews saying that this ...
3	4	Chrono Cross OST	The music of Yasunori Misuda is without questi...
4	5	Too good to be true	Probably the greatest soundtrack in history! U...

Proposed Approaches

Contextual Embeddings (Transformers-Based)



Static Embeddings



Parameter-Efficient Fine Tuning

Rather than updating all model parameters during fine-tuning, **PEFT** techniques could allow us to adapt large language models by training only a small subset of parameters. This significantly reduces resource usage and speeds up training.

Some potentially exploitable techniques that we can explore further :

- **Adapters**: insert small trainable layers between transformer layers, keeping the rest of the model frozen.
- **LoRA**: Injects low-rank matrices into attention layers to efficiently capture task-specific changes.
- **Prompt Tuning**: Learns task-specific prompts (tokens) while keeping the entire model unchanged.
- **BitFit:** Only fine-tunes bias terms in the model, achieving good performance with minimal updates.

Task design and Evaluation metrics

⭐ Rating	Mapped Sentiment
1-2	Negative 😣
3	Neutral 😐
4-5	Positive 😊

⭐ Rating	Mapped Sentiment
1-2	Negative 😣
4-5	Positive 😊

Evaluation strategy:

- **Accuracy**: Measures overall correct predictions. In case of binary class. dataset is balanced
- **Macro-averaged F1 score**: Balances performance across all the classes
- **Confusion matrix**: To analyze where misclassifications happen.

Project Status and Next Steps

- Completed the initial phase focused on research and design.
- Defined our sentiment classification setup and model comparison plan.
- Evaluated fine-tuning strategies including PEFT techniques to make it more feasible.
- Set up shared Github repository for team collaboration.
- Next: start implementation and experimentation.
- Intensive computations will be performed on Google Colab using cloud GPUs.

State of the Art Reference

Our literature review and problem framing are based on the following work:

A survey on sentiment analysis methods, applications, and challenges

Mayur Wankhade, Annavarapu Chandra Sekhara Rao, Chaitanya Kulkarni

Published in: Artificial Intelligence Review (07 February 2022)
https://doi.org/10.1007/s10462-022-10144-1

This work provided us with a structured overview of traditional and modern sentiment analysis approaches and helped us identify key research challenges and trends.

Thanks for your attention!

09/05/2025

Andrea Lepori, Gemma Ragadini, Mihnea Molnar, Pan Zhang, Loris Giunta

- Group 4 -