Exploring Aspect-Based Sentiment Classification in Coursera Reviews

# Abstract:

Aspect-based sentiment classification (ABSC) is a burgeoning area within natural language processing (NLP) aimed at identifying sentiment specific to targeted entities or features within text. This study contributes to the field by proposing novel methodologies and techniques for enhanced ABSC accuracy. We focus on the unique context of Coursera, utilizing a web-scraped dataset of course reviews to analyse sentiments towards key aspects: instructor performance and course content. Our research addresses specific challenges and opportunities presented by Coursera reviews, advancing not only general sentiment analysis but also offering valuable insights for researchers, practitioners, and developers of educational platform sentiment analysis systems.

# Keywords:

Aspect-based sentiment classification, NLP, Coursera, educational reviews, instructor performance, course content, web scraping

# Introduction:

ABSC delves deeper than traditional sentiment analysis, which classifies overall document sentiment, by pinpointing sentiment expressed towards specific aspects. This fine-grained approach is crucial for understanding nuanced opinions in various domains, including education.

This research focuses on Coursera as a case study, leveraging a dataset of scraped reviews to explore ABSC in the context of educational experiences. We focus on two critical aspects: instructor performance and course content, as these significantly influence student perception.

# Dataset:

## 1.Data Scraping:

A systematic web scraping process was conducted on the Coursera platform utilizing the Requests library. The data collection focused on a representative sample of 1000 freely accessible courses. The outcome of this effort resulted in the acquisition of approximately 478,775 individual reviews from diverse users.

## 2.Data Pre-processing before labelling:

The pre-processing procedures encompass the cleansing, normalization, and transformation of textual data, aiming to optimize its appropriateness for subsequent labelling stages. These specific pre-processing steps consist of:

### 2.1-Mention Removal:

This step eliminates mentions and URLs from the review to focus solely on the textual content.

### 2.2-Case Normalization:

The text is converted to lowercase to eliminate case sensitivity and ensure consistency.

### 2.3-Repeated Letter Handling:

This step identifies and removes consecutive repetitions of letters to improve the readability of the review.

### 2.4-Number Merging:

Numerical values are merged to represent them as single entities, reducing redundancy and improving data consistency.

### 2.5-Lemmatization:

Lemmatization employs linguistic knowledge to reduce words to their base forms, considering grammatical context and capturing the most accurate meaning.

### 2.6-Aspect-Specific Term Merging:

Specific terms related to the 'teacher' and 'course' aspects are merged to create more coherent and meaningful representation of the review's sentiment towards these aspects.

### 2.7-Correct Spelling:

## 3. Data Labelling Based on SenticNet5:

The sentiment lexicon, SenticNet5, was downloaded, and the concepts were extracted from the sentic.txt file. The polarity intensity and polarity value were utilized to represent sentiment scores, which were integrated into a dictionary. In this dictionary, words from the lexicon served as keys, and their corresponding polarity values were stored as entries.

For sentiment analysis, two predefined aspects were established: Aspect 1, consisting of words associated with educators (e.g., "teacher"), and Aspect 2, encompassing terms related to educational components (e.g., "course"). These predefined aspects served as the basis for labelling the data during the sentiment analysis process.

To prepare the reviews for analysis, spacy sentence detection was employed to split them into individual sentences. Subsequently, sentences were segregated into two lists based on their relevance to the predefined aspects.

Tokenization was performed on each sentence, breaking it down into individual words. Sentiment scores were assigned to these words using the predefined dictionary. The average sentiment score for each sentence was then calculated by accumulating individual word scores and dividing by the total number of words, excluding sentences containing words not present in the dictionary.

The final sentiment score for each aspect was obtained by aggregating the average sentiment scores of all relevant sentences in the corresponding list. This process facilitated a comprehensive evaluation of the overall sentiment related to each aspect.

To classify reviews, two labels were assigned based on the average sentiment value for each aspect, employing three threshold values: positive threshold (greater than 0.3), negative threshold (smaller than 0), and neutral for all other cases.

Reviews lacking alignment with any predefined aspect were labelled as 'neutral' and included in the pre-processed data.

## 4. Data Pre-processing after labelling:

The pre-processing steps involve cleaning, normalizing, and transforming the text to enhance its suitability for downstream sentiment analysis tasks. The specific steps include:

### 4.1-Punctuation Removal:

Punctuation marks are eliminated to simplify the text and focus on the core linguistic components.

### 4.2-Stop Word Removal:

Common stop words, such as "a," "the," and "an," are eliminated to reduce redundancy and focus on more meaningful content.

### 4.3-Single-Letter Word Removal:

Single-letter words, which often convey little meaning, are removed to reduce noise and improve the quality of the text.

## 5. Data Down-Sampling:

Down-sampling to Mitigate Class Imbalance for Aspect-Based Sentiment Analysis

### 5.1-Data Filtering:

5.1.1-Data points were initially filtered based on their respective aspect (instructor or course) and sentiment label (negative, neutral, or positive).

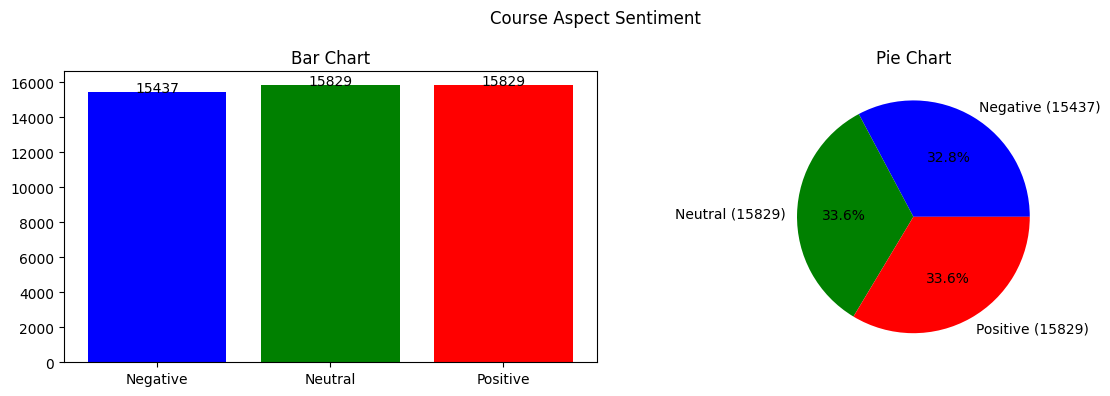
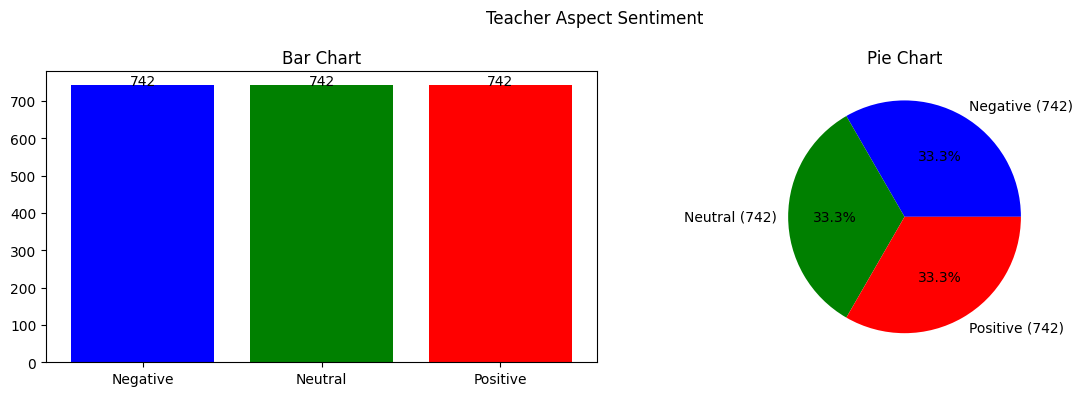
5.1.2-Duplicates were removed within each filtered subset to ensure data integrity.

### 5.2-Balanced Dataset Creation:

5.2.1- For each aspect, random subsets of 742 data points (instructor aspect) or 15,829 data points (course aspect) were drawn from each sentiment class.

5.2.2- These subsets were then concatenated, creating balanced datasets for both aspects.

5.2.3- Remaining duplicates were eliminated to maintain dataset consistency.



# Methodology:

## 1. Word Embeddings and Positional Encoding:

We leverage pre-trained Glove vectors to initialize the representation of individual words in a review, denoted as S.

For each word within S, we extract its corresponding Glove vector, , forming a sequence . Each is a dense vector of dimension , where n represents the word's position in the sequence and dimx signifies the embedding dimension.

Additionally, we incorporate relative positional information using separate positional encodings. For each word in the sequence, a corresponding positional encoding is obtained, forming another sequence Similar to word embeddings, each is a n x diml vector, where denotes the dimension of the positional encoding.

## 2. Construction of Bi-LSTM:

The resultant output from the Bidirectional Long Short-Term Memory (Bi-LSTM) model, represented as is formed through the concatenation of word embeddings X and positional encodings L during the input phase. Each input vector wi within HS possesses a dimensionality of , where dimw is equivalent to the sum of . This amalgamated representation adeptly encompasses both the semantic significance of individual words and their positional relevance within the review. Consequently, the Bi-LSTM can proficiently capture sequential dependencies and contextual information.

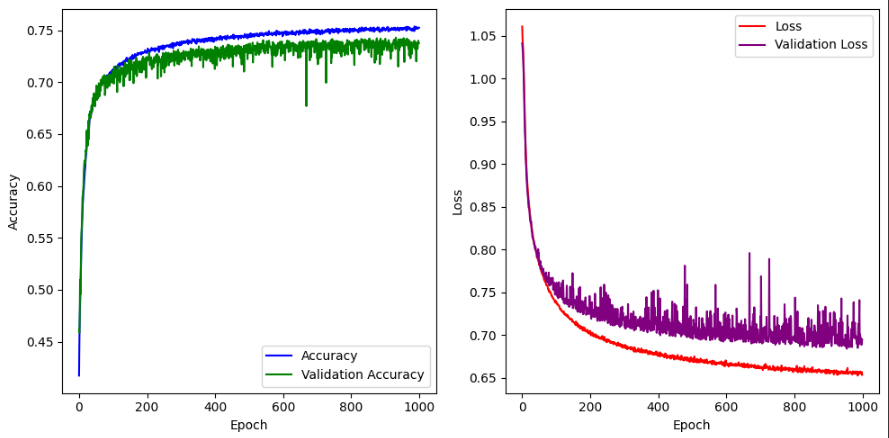
## 3. Attention-over-attention (AOA) module:

This module calculates the attention weights for the sentence based on the interaction between the sentence and the aspect representations. The AOA module uses two attention mechanisms:

* Target-to-sentence attention: This attention mechanism focuses on the parts of the sentence that are most relevant to the aspect.
* Sentence-to-target attention: This attention mechanism focuses on the most informative words in the aspect itself.

## 4.1. Baseline:

Our classification architecture employs a single fully connected layer. This layer receives its input from the Attention-over-attention (AOA) module and maps it to one of three distinct classes.



|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.4648257725180 | 0.4435382685069 | 0.4539325842696 |  |
| Neutral | 0.4577259475218 | 0.5266849649283 | 0.4897901304594 | 0.51836734693 |
| Positive | **0.64958123953** | **0.58175817581** | **0.61380183602** |  |

***Table 1.***

***For teacher aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.5037593984962 | 0.4718309859154 | 0.4872727272727 |  |
| Neutral | 0.46875 | **0.7** | **0.56149732620** | 0.49082568807 |
| Positive | **0.531645569** | 0.2916666666666 | 0.3766816143497 |  |

***Table 2.***

***For course aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.4630457201787 | 0.4422193040052 | 0.4523929471032 |  |
| Neutral | 0.4570301493378 | 0.5183764781080 | 0.4857741838873 | 0.51964972234 |
| Positive | **0.65278733654** | **0.59485732204** | **0.62247744052** |  |

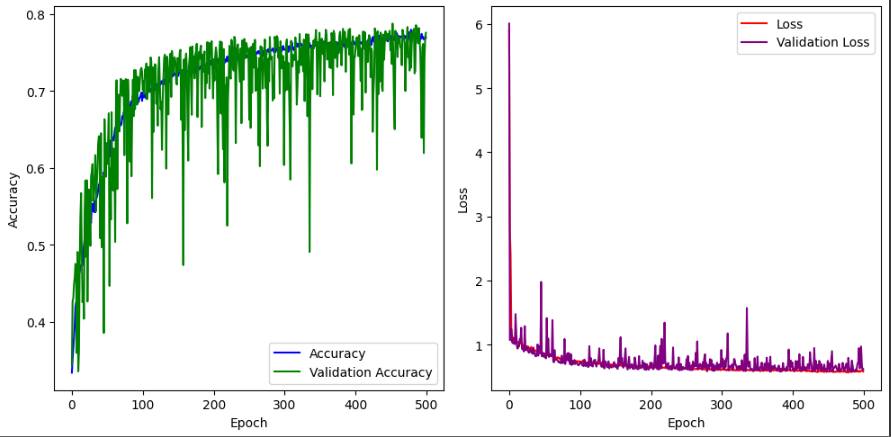
***Table 3.***

## 4.2. Bert features:

Our approach utilized the (BERT) model. Data was fed into the BERT model, and features were extracted from the hidden layer. Subsequently, these features were concatenated with features derived from the Attention-over-attention (AOA) module. Three individual training experiments were conducted, the details and results of which will be elaborated upon further.

### 4.2.1. Experiment 1:

The architecture of the model comprises four internal layers, in addition to an input layer and an output layer. The input layer consists of 400,000 neurons, corresponding to the feature dimensions of the input data. Each neuron in the internal layers implements a non-linear activation function, responsible for transforming the input received from the previous layer. The final output layer contains three neurons, corresponding to the three distinct output classes the model is designed to predict.



|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7348269994327 | 0.8127352572145 | 0.7718200774501 |  |
| Neutral | **0.81378242353** | 0.6410491003354 | 0.7171613783691 | 0.76897959183 |
| Positive | 0.7702519642373 | **0.85298529852** | **0.80951025056** |  |

**Table 4.**

***For teacher aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.6568047337278 | **0.78169014084** | **0.71382636655** |  |
| Neutral | 0.5691489361702 | 0.7133333333333 | 0.6331360946745 | 0.623853211009 |
| Positive | **0.68354430379** | 0.375 | 0.484304932735 |  |

***Table 5.***

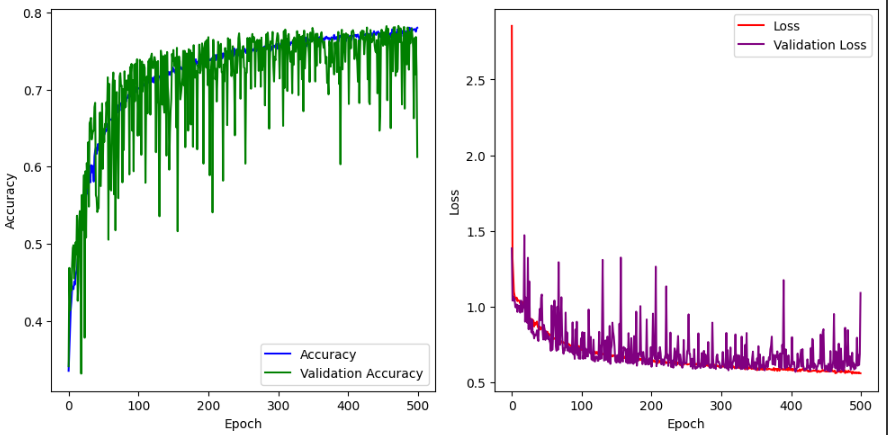
***For course aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7387548406315 | 0.8141825344714 | 0.7746368889583 |  |
| Neutral | **0.83298538622** | 0.6375838926174 | 0.7223026792179 | 0.77573686458 |
| Positive | 0.7721483942414 | **0.87456883035** | **0.82017350389** |  |

***Table 6.***

### 4.2.2. Experiment 2:

The architecture of the model comprises six internal layers, in addition to an input layer and an output layer. The input layer consists of 400,000 neurons, corresponding to the feature dimensions of the input data. Each neuron in the internal layers implements a non-linear activation function, responsible for transforming the input received from the previous layer. The final output layer contains three neurons, corresponding to the three distinct output classes the model is designed to predict.



|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7332589285714 | 0.8243412797992 | 0.7761370348493 |  |
| Neutral | **0.80910142691** | 0.6398292162244 | 0.7145776566757 | 0.7692857142 |
| Positive | 0.7764283742754 | **0.84398439843** | **0.80879815986** |  |

***Table 7.***

***For teacher aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | **0.76928571428** | **0.76760563380** | **0.70096463022** |  |
| Neutral | 0.560439560439 | 0.68 | 0.6144578313253 | 0.62155963302 |
| Positive | 0.7058823529411 | 0.4166666666666 | 0.5240174672489 |  |

***Table 8.***

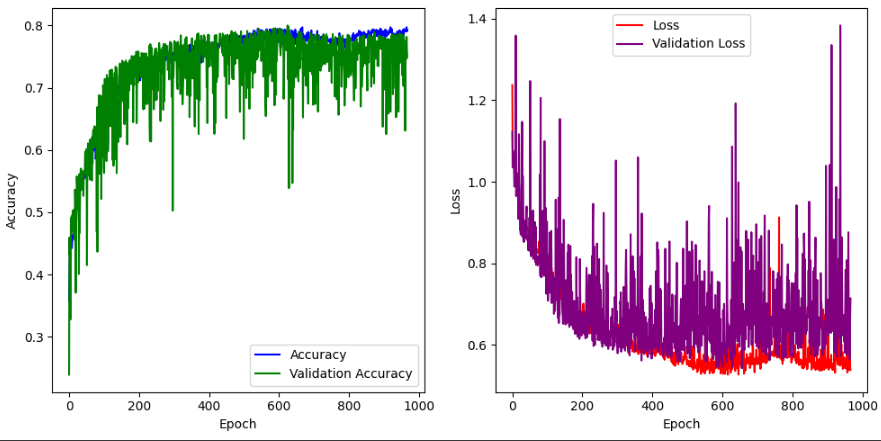
***For course aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7376281112737 | 0.8269862114248 | 0.7797554558117 |  |
| Neutral | **0.82787225217** | 0.6379034835410 | 0.7205776173285 | 0.77616403246 |
| Positive | 0.7781232334652 | **0.86328002508** | **0.81849264159** |  |

***Table 9.***

### 4.2.3. Experiment 3:

The architecture of the model comprises eight internal layers, in addition to an input layer and an output layer. The input layer consists of 400,000 neurons, corresponding to the feature dimensions of the input data. Each neuron in the internal layers implements a non-linear activation function, responsible for transforming the input received from the previous layer. The final output layer contains three neurons, corresponding to the three distinct output classes the model is designed to predict.



|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7736938031591 | 0.7989335006273 | 0.7861111111111 |  |
| Neutral | **0.7935264054514** | 0.7102775236352 | 0.7495976826520 | 0.78510204081 |
| Positive | 0.7886929750909 | **0.84548454845** | **0.81610194034** |  |

***Table 10.***

***For teacher aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | **0.78510204081** | 0.7676056338028 | **0.75694444444** |  |
| Neutral | 0.5416666666666 | **0.78** | 0.6393442622950 | 0.63990825688 |
| Positive | 0.7162162162162 | 0.3680555555555 | 0.4862385321100 |  |

***Table 11.***

***For course aspect:***

|  | precision | recall | F1 | accuracy |
| --- | --- | --- | --- | --- |
| Negative | 0.7749523204068 | 0.8003939592908 | 0.7874677002583 |  |
| Neutral | **0.81353438764** | 0.7069351230425 | 0.7564979480164 | 0.79186245194 |
| Positive | 0.7902257787939 | **0.86704296017** | **0.82685406698** |  |

***Table 12.***

# results:

|  | **Metric** | **Baseline** | **EXP1** | **EXP2** | **EXP3** |
| --- | --- | --- | --- | --- | --- |
|  | precision | 0.4648257725 | 0.7348269994 | 0.7332589285 | **0.773693803** |
| **Negative** | recall | 0.4435382685 | 0.8127352572 | **0.824341279** | 0.7989335006 |
|  | F1 | 0.4539325842 | 0.7718200774 | 0.7761370348 | **0.7861111111** |
|  | precision | 0.4577259475 | **0.813782423** | 0.8091014269 | 0.7935264054 |
| **Neutral** | recall | 0.5266849649 | 0.6410491003 | 0.6398292162 | **0.710277523** |
|  | F1 | 0.4897901304 | 0.7171613783 | 0.7145776566 | **0.749597682** |
|  | precision | 0.6495812395 | 0.7702519642 | 0.7764283742 | **0.788692975** |
| **Positive** | recall | 0.5817581758 | **0.852985298** | 0.8439843984 | 0.8454845484 |
|  | F1 | 0.6138018360 | 0.8095102505 | 0.8087981598 | **0.816101940** |
|  | Accuracy | 0.518367346 | 0.768979591 | 0.768979591 | **0.785102040** |

***Table 13.***

# Conclusion:

The AOA-LSTM model proposed in this paper achieves state-of-the-art performance on aspect-level sentiment classification tasks, outperforming several existing methods. This demonstrates the effectiveness of the attention-over-attention (AOA) module in capturing the interaction between aspects and sentences, which is crucial for accurately predicting the sentiment of an aspect within a sentence. Overall, the AOA-LSTM model is a promising approach for aspect-level sentiment classification with potential applications in various domains.

# References:

[1] B. Huang, Y. Ou, and K. M. Carley, “Aspect level sentiment classification with attention-over-attention neural networks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2018, pp. 197–206. doi: 10.1007/978-3-319-93372-6\_22.