Analysis of an Emotional Contagion Detection Algorithm Based on Sentiment Evaluation

by

Arleen Kaur Arora

Bachelor's in Computer Engineering, University of Mumbai, 2021

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Computer Science

In the Graduate Academic Unit of Computer Science

Supervisor(s):	Luigi Benedicenti, Ph.D, Computer Science
Examining Board:	Paul Cook, Ph.D, Computer Science, Chair
	Francis Palma, Ph.D, Computer Science
	Shivam Saxena, Ph.D, Electrical and Computer Engineering

This thesis is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

December, 2023

Abstract

Emotional contagion, or the transfer of emotions between individuals, is a well-studied phenomenon in social psychology. In recent years, researchers have become interested in understanding how emotional contagion can impact team dynamics in software development, particularly on GitHub where communication takes place primarily online. In this study, we propose an approach to detect emotional contagion in collaborative software development platform using sentiment analysis tool available in Mathematica. This approach is based on a previously published framework. We collected data from several GitHub repositories and analyzed the emotional content of developers' comments to identify patterns of emotional contagion. Our findings suggest that emotional contagion exists and the algorithm defined can be used on various data sets. Our method is applied on five GitHub repositories, however the method is general and can be reused for experimental validation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Emotional contagion is defined as a process wherein one person's experience in a group is transmitted to others in the same group [1]. The thesis is based on the framework proposed in previous work [2]. The framework contains a method to detect whether or not emotional contagion exists in a collaborative software development where such collaboration happens through online means, and more specifically GitHub. The framework used for implementing it requires no face to face human interaction and the GitHub repository used was Opencv [3]. The cited research is based on a framework which visually allows one to assess trends in a plot of the integrated dataset. This framework thus allows one to measure emotional contagion based on qualitative analysis.

The framework uses a pattern recognition method [2]. Consistent sequences determine whether emotional contagion exists or not. The method was computationally slow, leading the researchers to opt for an alternative approach. This approach splits the events into short sequences and each of them is a emotional contagion event. This method can be applied to bigger data-sets to check whether the pattern is discerned or not [2].

Software development is less than a century old and methods for its production have evolved more over that time [4]. Like craftspeople in the 1950s, skilled programmers had a thorough knowledge and comprehension of their field. The software was created iteratively, with errors in the code being repaired until the user was happy. The code-and-fix method survived because software was not that complex and there was no more efficient technique for developing software. However, the code-and-fix approach did not last long [5]. This laissez-faire strategy gave way to more structured approaches as software usage increased and firms came to rely on computers for daily operations. By the mid-sixties, management wanted software development to be a managed and controlled process much like other industrial activities [6]. Therefore, to accomplish Taylorism [22] and implement the waterfall model [7], developers turned to a more than fifty-year-old paradigm, called "Scientific Management" [8]. As applied to software development, Scientific Management led to the development of factory-like concepts. R. W. Bemer of General Electric was among its earliest proponents [9]. By the late 1960s, the term 'software factory' was in popular use and became associated with computer-aided tools, management-control systems, modularization, and reusability [10]. Attempts were made to introduce statistical control in software engineering [11]. Efficiency of software development processes were measured through the use of control charts. Models such as CMMI (Capability Maturity Model Integrity) gained popularity for defining and improving software development processes [11]. Although, introduced in 2000s, agile manufacturing and agile software development have roots that can be traced to both Lean and Agile manufacturing paradigms introduced in the 1970s and 1990s respectively [4].

The advent of computers has created new software development procedures and processes have changed over time [11]. These techniques have been updated to reflect the most recent advancements in computer hardware, software development tools, and organisational administration of software development teams. With this development, new software development techniques have emerged as a result of worldwide private and public software development initiatives. Today, software development is one of the most potent, essential, and necessary fields. Software Development Life Cycle (SDLC) is all about the minimization of failure and the maximization of product quality. To make the development work in a step by step procedure and precisely, SDLC came into existence. The SDLC defines the framework that includes different activities and tasks to be carried out during the software development process [12]. There are different types of models like the waterfall model, V shaped model, evolutionary prototyping model, spiral model, iterative and incremental model, and agile model. To ensure a project's success, it may be necessary to select the appropriate SDLC model based on the particular problems and needs of the project. Every model has benefits and drawbacks. Software development is divided into a set of activities that allow any software development company to control the software product easily. Software development life cycle models complete the software development process step by step. When the process is powerful, the result will be strong as well, and the project will succeed [12].

Large software systems are created with the collaboration of geographically dispersed developers using global software engineering. In this environment, a large number of software engineers must collaborate while navigating difficulties brought on by geographic, temporal, cultural, and linguistic variety [13]. Effective team collaboration is required for software projects to be successful across the full development lifecycle. Working together might be challenging, yet it is proven that collaborative development produces software that is better than what a single developer could produce [9]. In order to continue offering value to clients, software development teams should adhere to follow best practises to be a dependable counterpart and active contributor.

A famous example of open collaboration is open source software (OSS), which is what happens when individuals work together to create something in order to accomplish a common objective. It provides the framework for team members to contribute their knowledge, skills, and expertise in order to create new products. GitHub is a web-based code hosting service that uses the Git distributed version control system. Furthermore, GitHub is collaborative and available for free. As of June 2022, GitHub reported having over 83 million developers and more than 200 million repositories, including at least 28 million public repositories [14]. GitHub has become an essential tool in technology areas that demand collaboration, such as globally distributed software development. After one opens a pull request in a repository, collaborators or team members can comment on the comparison of files between the two specified branches, or leave general comments on the project as a whole.

Scholars have investigated how human factors impact the creation of software products [15]. These factors include, for instance, personality [16], cognitive capacity [17], and experience [18]. Especially in agile development, affect has drawn numerous scholars recently to highlight the positive and negative factors that can either increase or inhibit positive communication among software team members. The Oxford Dictionary of Psychology defines affect as “Emotion or subjectively experienced feeling such as happiness, sadness, fear, or anger” [19]. This can lead to emotional contagion spreading from one person to another, or across larger groups.

As revealed by Domagalski (1999), the word ‘contagion’ has derived from a Latin word ‘contagio’ that means ‘from touch’ [20]. Historically, emotional contagion has been directly related to the feelings of empathy and sympathy [21], stemming from the German term Einfühlung [22]. Emotional contagion is also a phenomenon that people experience the same emotion by "catching" other's expressed emotion during their communication [23].

Software developers that experience negative emotional contagion tend to be dissatisfied, which has an adverse influence on the software company for which they work [24]. It seems that the study of emotional contagion can help the software building process. This study should help the quality of software products produced. The aim of our research is to find out ways to detect emotional contagion in an Open Source Software environment. Our research would help in future to find ways to reduce the impact of emotional contagion while developing software. Also, we will investigate whether the proposed method would be suitable for a wide range of projects. Hence, the contributions of this thesis can be listed as:

1. Determining if an existing algorithm can detect emotional contagion in large GitHub Repositories. Detecting episodes of emotional contagion in both short and long comments. Detecting episodes of emotional contagion in AND, OR sequences created for determining consistency of sentiment in both short and long comment sequences.

2. Determining the robustness of the algorithm used to detect emotional contagion by computing correlations between short and long comments and AND and OR sequences.

Our objective is to answer the following research questions:

1. Can an existing algorithm be applied to large GitHub repositories to detect emotional contagion?

- Finding presence of emotional contagion episodes in both short and long comments.

- Determining whether the sentiment is consistent throughout short and long comments.

2. Is the algorithm used to detect emotional contagion robust?

- Computing correlations between the data sets used.

The remainder of this thesis is structured as follows. Chapter 2 will discuss previous work. In Chapter 3, we describe the algorithm used to detect emotional contagion. Chapter 4 presents the results of our analysis. Chapter 5 includes discussions and ethical considerations of data used in our study. We also mention the limitations of this research. Finally, in Chapter 6, we summarise the contributions of this thesis, and we briefly conclude our work and discuss potential future research directions.

Chapter 2

Related Works

Several studies have explored the detection of emotional contagion in open source collaborative software environments using various approaches. This chapter summarizes previous research on detecting emotional contagion, different mechanisms and approaches, including the impact of emotional contagion. We also review previous research on collaborative software development and it's effects.

2.1 Emotional Contagion Overview

Emotional contagion is defined as "the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person and, consequently, to converge emotional" [23]. One widely known argument on emotional contagion is that emotional contagion among people is composed of a two-stage process. The first stage is described as "people automatically and continuously mimic and synchronize their movements with the facial expression," and the second stage is described as "emotional experience is affected, moment to moment, by the activation of and feedback from facial mimicry" [25]. Emotional contagion can be triggered by facial expressions, indirect human interactions, and/or by observing other people's behaviour in direct and indirect interactions. Furthermore, emotional contagion can be triggered physiologically or neurologically by synchronizing with the emotional state of others during human interactions. Emotional Contagion classified as positive, negative, neutral in professional work can affect creativity, group rapport, user focus, and job satisfaction [26].

By studying emotional contagion, we can identify strategies for managing emotions and promoting positive emotional experiences that can lead to improved team performance and well-being. Also, understanding emotional contagion can help individuals and organizations to better manage emotions and promote positive emotional experiences.

2.2 Collaborative Software Development

In this section, we focus upon collaborative software environment and the presence of emotional contagion in collaborative software developments. In collaborative software development, team members often work on complex projects under tight deadlines, which can create stress and pressure. This can increase the likelihood of emotional contagion occurring within the team. Therefore, it is important to study as emotional contagion might exist in collaborative software development because it is a natural part of human interaction and can be exacerbated by the high-pressure, deadline-driven nature of software development projects and the challenges of communicating and collaborating in a virtual environment.

Software development is a highly collaborative activity in which developers engage in collaborative tasks and interact with shared artefacts. The effectiveness of collaboration determines whether many small- to large-scale software development businesses succeed or fail. [1]. Therefore, one of the key concerns in project management is how to increase collaboration efficiency. Efficiency of cooperation is influenced by a variety of factors, including task definition, team trust, and technological usability. Anger, disgust, fear, joy, sadness, and surprise are examples of basic emotions which can influence job satisfaction, group cohesion, user focus, and creativity in the workplace [27]. In the collaborative relationships in open source software (OSS) projects, where the work is volunteer-driven, hence developers happiness during the development process is therefore paramount [28]. Happiness here is used as a colloquial term which means to research in organizational behavior and psychology [29]. To identify the key factors behind collaborators sentiment relations and understand how these relations interact with collaborative relationships is important for managing Open Source Software projects [26].

The success of software development largely depends on developers collaboration efficiency and many factors influence the formation of successful collaborative relationships. Work by researchers suggests that human-related issues, such as rapport and transactive memory, are important for collaborative work [26]. The abilities needed by an effective team are recognised as shared intention, sharing of objectives, plans, and knowledge of the environment, as well as understanding of roles and duties and team awareness. Generally, collaborative relationships are formed when two people work together to accomplish common goals. In GitHub, issue reports are used by team members to ask for advice, and express and share opinions related to software maintenance and evolution [30]. Collaboration between two developers is defined as the issue resolution process they both participate in. The posting of comments under an issue by two developers indicates a collaborative relationship. According to the research, encouraging positive sentiment linking among collaborators can increase their sense of closeness, and doing so in software development will improve the collaboration ecosystem. It is also proposed that, in order to improve collaboration willingness and effectiveness, negative sentiment impacts are more likely to be reduced than enhanced by adjusting and reassigning collaborators based on network features as well as elements impacting sentiment consistency. The results also indicate that developers of different positions in the collaboration network tend to have fewer collaborations in practice, which may reduce sentiment consistency accordingly. This inspires us that we should promote the collaborations among developers of different positions in the collaboration network. More specifically, to encourage positive developers of high degree (i.e., of central position) to cooperate with negative developers of low degree (i.e., of peripheral position) can bring more gains to projects. These findings tell us when we are going to coordinate the OSS projects, we should take the consistency of developers sentiments into account in order to promote their collaborations in a task. As a measure of implementations, the consistency of developer sentiments can be monitored so that we can take appropriate measures to regulate the organization of the development process. Consistency of developers sentiment also encourages us to incorporate these new features into the future monitoring tools [26].

2.3 Presence of Emotional Contagion

This section describes various instances were emotional contagion was found. Evidence of presence of emotional contagion in live streams was detected by researchers. Researchers used Vader which is a sentiment analysis tool to detect emotional contagion in live stream conversations. The sentiment of subsequent chat messages is correlated with previous chat messages from viewers and oral messages from the videos. Though the effects differ depending on the sort of live stream, in general, viewer chat is a better indicator of ensuing chat messages than oral ones in live videos. Although the findings suggested that emotional contagion is present during live streaming, they concluded that alternate theories such verbal mimicking cannot be completely ruled out [27].

Other results suggest that, selfie-based emotional contagion over the social network was confirmed throughout the entire evaluation. The authors developed a phone application named “SmileWave” in order to detect emotional contagion. It was found that there was a weak correlation between smile degree of posted selfie photograph and increase in smile degree when users view other users posted photographs. A possibility that the effect on emotional contagion may differ depending on the degree of facial expression was determined. On the other hand, it was also found that viewing five photographs with highly smile degree has a possibly significant impact on the smile degree of user’s selfie when taking photographs. Although, a previous work argued that emotional contagion is instantaneous, there is a possibility that the momentary effect of emotional contagion may accumulate continuously and cause longer-term emotional contagion as observed in their evaluation [23]. This required more careful investigation because it was not confirmed in this experiment that all five smiley selfie photographs have an influence; instead it might be the influence of only the photograph just before the photograph was taken. Therefore, we concluded that selfie-based emotional contagion occurs over the social network. To the best of their knowledge, extensive evaluation is the first work that deeply investigated and revealed the effects of selfie-based emotional contagion over the social network [21].

With data from millions of Facebook users, a study showed that rainfall directly influences the emotional content of their status messages, and it also affects the status messages of friends in other cities who are not experiencing rainfall. For every one person affected directly, rainfall alters the emotional expression of about one to two other people, suggesting that online social networks may magnify the intensity of global emotional synchrony. The authors developed an instrumental variables regression, a technique pioneered in economics, to detect emotional contagion on the data from Facebook [31].

Additional findings while exploring hybrid open source state that, where developers and other stakeholders with various backgrounds exist, on applying sentiment analysis, it aims to understand if the community members behave differently, depending on their background in the company that owns the community IPR (Intellectual property rights), being an independent contributor, or representing a collaborating organization.

Sentiment analysis classifier and appropriate statistical tools were used here to extract and classify the information on emotional contagion episodes. Given the large number of developer's contributions, the analysis is statistically significant, within the limits highlighted in the section on threats to validity. This paper provides two novel contributions: the first is to detect emotional contagion episodes in a large hybrid development project. The second is to determine the origin of such emotional contagion episodes and analyse the difference between the proportion of positive, neutral, and negative contributions from internal developers and external developers [32]. The analysis shows that external contributors in general seem to have more positive comments than the internal ones. This trend is visible in terms of proportion of contributions per contributor category as well as in contagion sequences clustered by pull requests.

Sentiment analysis uses natural language processing, text analysis and computational techniques to automate the extraction or classification of sentiments from texts [33]. Researchers analyze commits and pull requests on GitHub and find that more negative emotions are expressed in security-related discussions [34]. Java projects are found to attract more negative comments while projects with more distributed teams attract more positive comments [35]. A study investigating commit logs on GitHub finds that Tuesday's comments have the most negative sentiments [36].

In a previous thesis, researchers used two approaches were employed to test hypotheses related to emotion detection in commit messages. The first approach involved using the Classify function in Mathematica, which employs supervised machine learning algorithms to classify commit messages. The second approach utilized Mathematica's built-in classifier called Sentiment to gather sentiment from the text. The results indicated that the trained classifier performed differently compared to the built-in sentiment classifier, particularly in classifying messages as neutral, positive, or negative. The research highlighted the number of messages classified under each sentiment category using both methods and discusses the effectiveness and limitations of each approach [37].

The motivation behind investigating the impact of emotional contagion in collaborative software development is to better understand how emotions can affect team members and the team as a whole. By studying emotional contagion, we can identify potential positive and negative outcomes of emotional contagion in software development teams, such as increased creativity, improved collaboration, decreased motivation, and decreased job satisfaction. Understanding emotional contagion can also help software development teams to develop strategies for managing emotions and improving team dynamics.

Chapter 3

Approach

This thesis describes the adoption of a previously designed algorithm to detect the presence of emotional contagion in the comment section of various open source GitHub repositories [38]. We describe the in-built sentiment classifier used in Mathematica (the tool adopted for our analysis) and explain its step-by-step implementation and apply it to five data sets. Furthermore, we explain how we used AND and OR operators on the comment sequences and computed correlations among the data sets to determine to robustness of the algorithm used.

To enable the replication of this approach by other researchers, we will outline the steps involved in this study.

Step 1. Retrieve data from the repository.

Step 2. Import the file in Mathematica after making the required changes.

Step 3. Apply in-built Sentiment Classifier to classify Positive, Negative and Neutral comments.

Step 4. Determine if episodes of contagion exist.

Step 5. Create AND, OR functions using short and long comments.

Step 6. Determine if episodes of contagion exists on the obtained output.

Step 7. Compute correlations between short and long comments and AND and OR.

3.1 Sentiment Classifier

The Sentiment classifier is one of the standard built-in classifiers in Mathematica [39]. The classification is based on Supervised Machine Learning algorithm. This classifier attempts to infer the sentiment that a snippet of text conveys. The input text should typically be one or a few sentences. This classifier assumes the text conveys only one sentiment. The probabilities reflect the belief in these sentiments, not the proportion of sentiments. The current version only works for the English language. The sentiments are classified into four classes:- 1. Positive 2. Negative 3. Neutral 4. Indeterminate. For example: Classify [“Sentiment”,{“I am so sad”, “My phone broke again”,”I love this movie”}] Output: {Negative, Negative, Positive}

To apply the sentiment classifier we used GitHub repositories as datasets. To investigate the algorithm we used five GitHub repositories namely TensorFlow [40], Keras [41], Pytorch [42], React Native [43] and Pillow [44]. To retrieve the complete data set of interest for evaluation, we used the following line generating command: git log – pretty=format:'tˆ3%antˆ3,tˆ3%adtˆ3, tˆ3%s tˆ3, tˆ3%btˆ3' > ./GitlogFull-

Dataset.csv

Due potential presence of ”stray” double quotes, which can cause confusion for the import mechanism in Mathematica, 3-character terminator tˆ3 is used. We loaded this generated file in Text Editor and replaced all the double quotation marks with single quotation marks, then finally replaced the multi-character terminator tˆ3 with double quotation marks. After loading the entire data set in Mathematica we converted the dates into actual Mathematica dates as the date format in the log file is not directly recognized by Mathematica. We then applied the sentiment analysis classifier. To ensure accurate computation of cumulative and integral functions, we removed the indeterminate results from the classifier's output. Indeterminate threshold was set to zero, which was done by specifying the IndeterminateThreshold option in the classifier. We used the MapAt function as shown in Fig. 3.1, which was native to Mathematica and also allowed to apply a specified function to a column of the database which we needed. We used MapAt command for both columns i.e 3,4. Column 3 contained short comments in the project and column 4 contained long comments. Short comments are commit messages summarizing changes briefly and are typically concise, often a single line. Long comments provide more detailed information about the modifications, reasons, or context behind the commits and are multiple lines long.

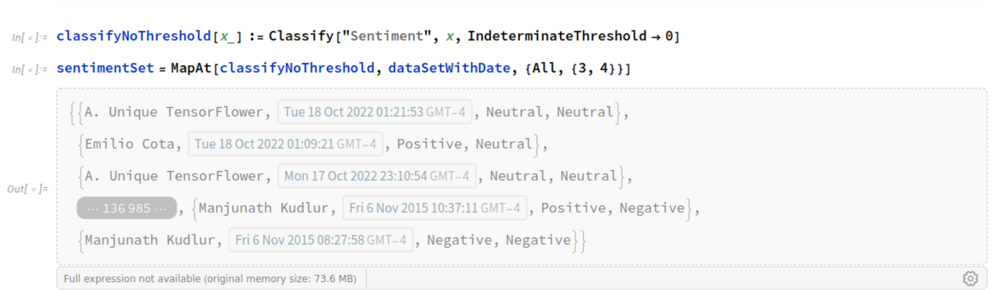


Figure 3.1: Application of MapAt function

In order to get positive, negative and neutral trends easily in form of cumulative charts for the entire data set, we replaced negative sentiment with -1, positive sentiment with +1, and neutral sentiment with 0 as shown in Fig. 3.2. We then plotted the short comment and long comment lists as sequences.

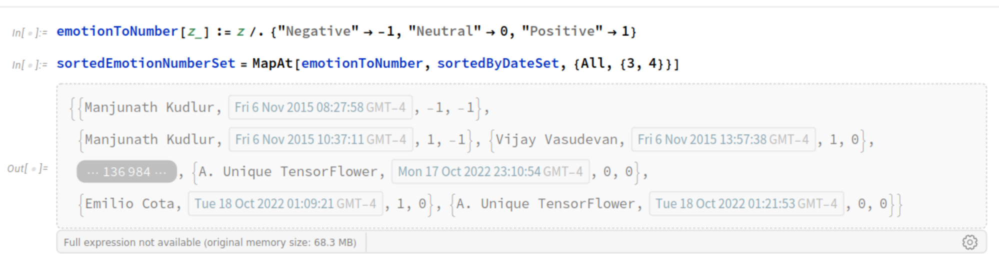


Figure 3.2: Assigning values to sentiment polarities

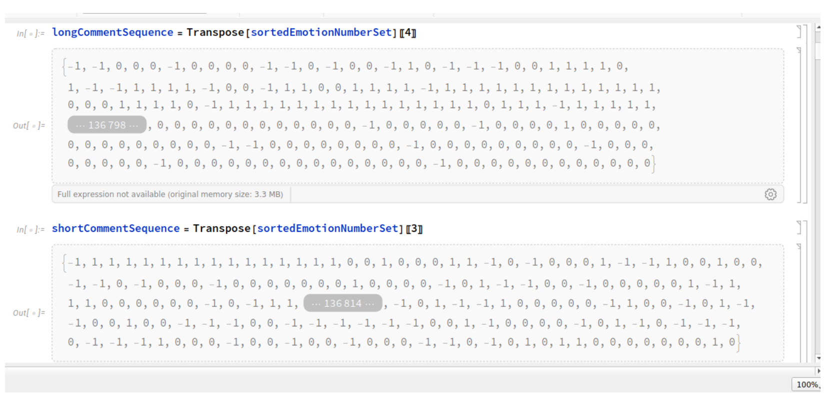


Figure 3.3: Long Comment Sequence and Short Comment Sequence

In order to convert the lists to a numerical format which would make it easier to draw graphs, an interpolating function was created. To do this, it was required to extract columns which was possible by transposition as shown in Fig. 3.6. After plotting these sequences we could detect how sequences were but did not get a temporal view of the accumulation of positive, negative or neutral sentiments. Creating an interpolating function and applying transposition on short and long comment sequences followed by verifying the results obtained from using cumulative function was the fastest way to achieve the desired result.

Finally, to detect emotional contagion, it was necessary to match the pattern for positive, negative and neutral values. To achieve this, we counted sequences and to display the count of sequences we plotted sequences as a histogram. To list the different occurrences we used DeleteDuplicates and created a table of frequencies by counting how many times a pattern occurs in the list. We repeated all the above steps for positive, negative and neutral contagion.

3.2 Using Logical "AND" Operator on Comment Sequence

To further analyze the algorithm and determine whether the sentiment is consistent throughout short and long comments, we added a AND function and computed it on both long and short comment sequences as shown in Fig. 3.4. The outcomes derived from the AND function would indicate whether or not the sentiment conveyed in both short and long comment sequences is consistent throughout. After running the data set through a sentiment classifier, we build a function that performs an AND operation on both long and short comment sequences. The function implements a logic where the output is "1" if the long and short comment sequences convey the same sentiment that we were looking for, and "0" otherwise. We created a constant array of '1', '0' or '-1' depending on the sentiment that we were looking for. MapThread in Mathematica applies the function to corresponding pairs of elements and hence we used it to apply our function to the entire dataset as shown in Fig. 3.5. The resulting list obtained after applying this function was then used to find contagion sequences.

```
In[ ]:= ValueOr = -1
Out[ ]= -1

In[ ]:= SentimentAnd[longCommentSequence_, shortCommentSequence_, ValueOr_] :=
         If[longCommentSequence == ValueOr && shortCommentSequence == ValueOr, 1, 0]
```

Figure 3.4: Logical AND Function

```
In[ ]:= longAndShort = MapThread[SentimentAnd, {longCommentSequence, shortCommentSequence, x}]
```

Figure 3.5: Application of MapThread

3.3 Using Logical "OR" Operator on Comment Sequence

Logical OR was performed with the motive to find out whether or not either of the comment sequence, i.e. short and long, has the sentiment (positive, negative or neutral respectively) that we are looking for. After applying sentiment classifier on the data set, we created a function to perform OR operation on long and short comment sequences as shown in Fig. 3.6. The function utilizes a logical construct such that in the event if either the long or the short comment sequence conveyed the sentiment that we were looking for, the resulting output is assigned a value of "1"; conversely, the output is assigned a value of "0". The subsequent steps were identical to those taken when implementing logical AND.

```
In[ ]:= ValueOr = -1
Out[ ]= -1

In[ ]:= SentimentOr[longCommentSequence_, shortCommentSequence_, ValueOr_] :=
         If[longCommentSequence == ValueOr || shortCommentSequence == ValueOr, 1, 0]
```

Figure 3.6: Logical OR Function

```
In[ ]:= longOrShort = MapThread[SentimentOr, {longCommentSequence, shortCommentSequence, x}]
```

Figure 3.7: Application of OR Function on Long and Short Comments

3.4 Correlations

To find out whether or not there is a connection between short comments, long comments, AND of short and long comments, OR of short and long comments; we computed correlations as shown in Fig. 3.8. Finding correlation values can be helpful for identifying patterns of emotional contagion in large data sets. While detecting emotional contagion, we used correlation to identify whether or not the emotional state of one data set is related to the emotional state of another data set. We found correlations to determine the consistency of sentiment expressed in comment sequences across data sets. If the results show high value of correlations, we can infer that there is robustness between the data sets. If the values are relatively low, we can say that there is distinctness between the data sets used. Spearman Correlation Coefficient was used here. For this test we used Spearman Correlation Coefficient as it allows the x, y values to be continuous or ordinal and approximate normal distributions for x, y values are not required. We used the in-built Correlation function in Mathematica to get results.

```
In[•]:= Correlation[longAndShortPos, longOrShortPos] // N
Out[•]= 0.200065
```

Figure 3.8: Correlation

Fig. 3.8 shows that the values for x, y to find correlation are longAndShortPos i.e the list obtained when AND operator was applied to long and short comment sequence for positive contagion sequence for a data set, longOrShortPos i.e the list obtained when OR operator was applied to long and short comment sequence for positive contagion sequence for a data set [44].

We chose five data sets for our analysis, namely TensorFlow, Keras, PyTorch, Pillow, React-Native. Our motive was to choose different size of data sets to analyze the versatility of the algorithm. Each data set used has its own characteristics. Table 4.1 describes some properties of the data sets used.

Data set	Contributors	Year	Features
TensorFlow	3380	2016	90 authors have pushed 276 commits to master and 286 commits to all branches.
Keras	1127	2017	6 authors have pushed 7 commits to master and 11 commits to all branches. On master, 38 files have changed and there have been 785 additions and 662 deletions.
Pytorch	2761	2016	Excluding merges,158 authors have pushed 290 commits to main and 1,517 commits to all branches.
React Native	2485	2015	Excluding merges, 42 authors have pushed 131 commits to main and 155 commits to all branches.
Pillow	379	2015	Excluding merges, 1 author has pushed 11 commits to main and 11 commits to all branches

Table 3.1: Data sets and their characteristics

Chapter 4

Results

In this section we present the results we obtained from applying the process described in chapter 3 on five data sets, namely Tensorflow, Keras, Pillow, Pytorch and React Native. At first, we considered results obtained by computing histograms from contagion sequences obtained from all data sets. On each data set the proposed algorithm was applied on short comment sequence and long comment sequence. The analysis were expanded by detecting emotional contagion after implementing AND and OR logical operations. We also present the correlation values found for each data set.

After performing sentiment analysis, it was then possible to create an affect sequences. The resulting sequences were integrated and plotted. Tables following also show the affect sequence charts wherein, we show the first 500 data points in the affect sequence and highlight a few instances of emotional contagion found in it. The X axis is time, and the Y axis is the emotional accumulation level.

4.1 Tensorflow

Results when tested with TensorFlow dataset are as follows: (Refer A.1 for sequences)

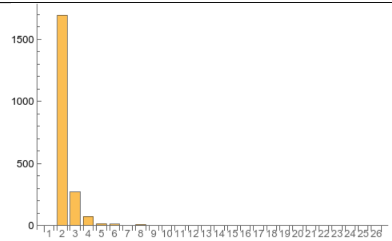
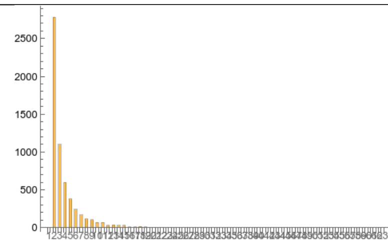
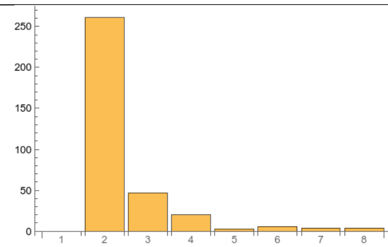
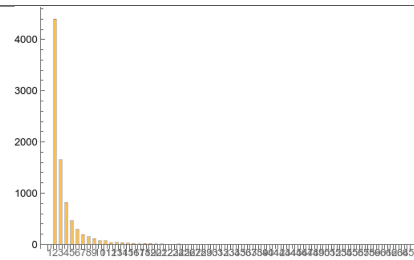
Positive Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.1: Tensorflow- Positive Sentiment

4.1.a. We can observe that majority of the comment sequences are of length 2, which occurred 1694 times. The number of comment sequences decreases as the length of the sequence increases, with only one occurrence of sequences having a length of 9 and 26. The maximum sequence length observed is 26.

4.1.b. Majority of the comment sequences are of length 2, which occurred 2780 times. The number of comment sequences decreases as the length of the sequence increases, with only one occurrence of sequences having a length of 27, 29, 31, 33, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, and 54. The maximum sequence length observed is 63.

4.1.c. Majority of comments are of length 2 with a count of 261. Lengths 3 and 4 have a count of 47 and 20 respectively.

4.1.d. The most frequent comment sequence length was 2, with 4,397 comments, followed by a comment sequence length of 3, with 1,650 comments.

Category	Correlation
Short-Long Comments	0.029714
AND-OR	0.245897

Table 4.2: TensorFlow- Positive Sentiment

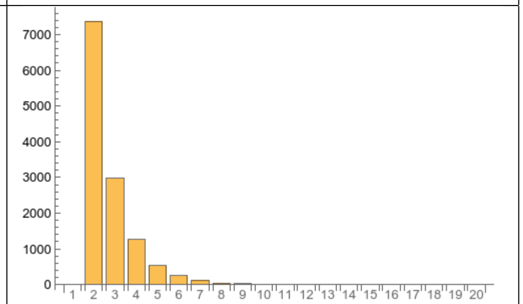
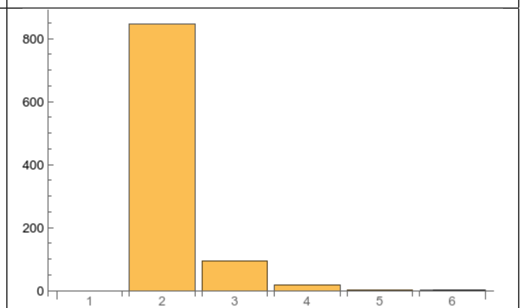
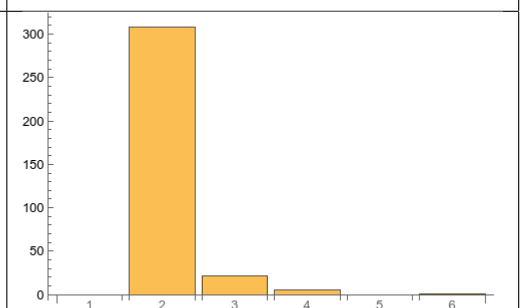
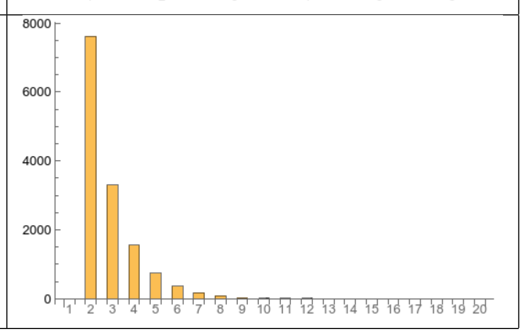
Negative Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.3: Tensorflow- Negative Sentiment

4.3.a. Majority of the negative contagion comment sequences are of length 2, which occurred 7367 times. The number of comment sequences decreases as the length of the sequence increases, with only one occurrence of sequences having a length of 15, 16, 17, and 20. The maximum sequence length observed is 20.

4.3.b. The majority of comments in the sequence have a small negative sentiment lengths. This is indicated by the fact that the pairs 1,0, 3,95, 4,18, 5,3, and 6,2 all have lengths of 95 or less, while only one comment has a higher length of 846. Comment 2 appears to be an outlier in terms of negative sentiment. It has a length of 846, which is significantly higher than all the other comments in the sequence.

4.3.c. There are 308 occurrences of negative contagion with a sequence length of 2. There are 21 occurrences of negative contagion with a sequence length of 3. There are 5 occurrences of negative contagion with a sequence length of 4 and 1 occurrences of sequence length 6. There are no occurrences of negative contagion with a sequence length of 5. We see that negative contagion tends to diminish or become less in comment sequences longer than four comments.

4.3.d. The most frequent comment sequence length was 2, with 7625 negative sentiment comments, followed by a comment sequence length of 3, with 3302 negative sentiment comments. As the comment sequence length increases, the frequency of negative sentiment comments decreases, with only 1 negative sentiment comment found in comment sequences of length 16 and 17.

Category	Correlation
Short-Long Comment Sequence	0.0297147
AND-OR	0.224137

Table 4.4: TensorFlow- Negative Sentiment

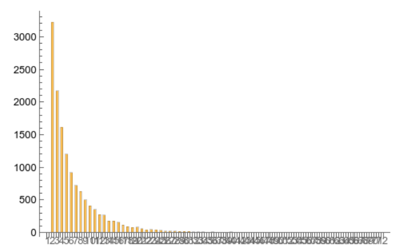
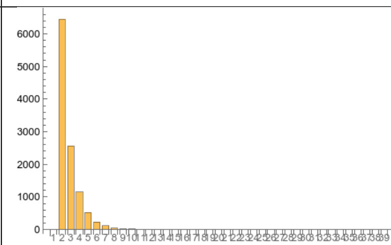
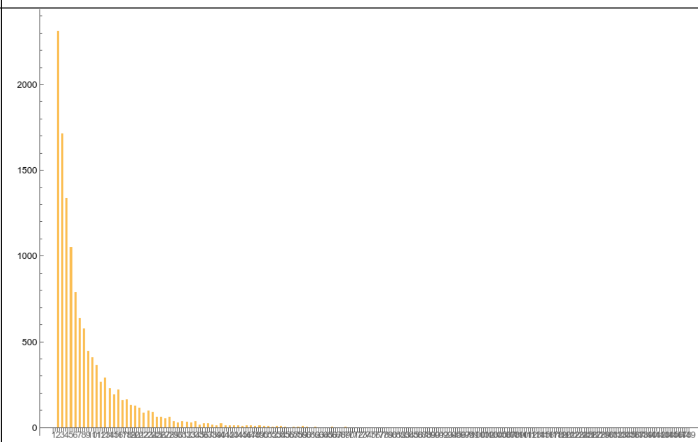
Neutral Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.5: Tensorflow- Neutral Sentiment

4.5.a. Comments 2 and 3 have significantly higher neutral sentiment lengths than the other comments in the sequence.

4.5.b. The emotional contagion frequency length decreases as the comment sequence progresses. There is a sharp drop in emotional contagion length from the first to the second comment. The emotional contagion length gradually decreases from the second to the twelfth comment. From the thirteenth to the sixteenth comment, there is a slight increase in the emotional contagion length. The emotional contagion length decreases again from the seventeenth to the thirty-fifth comment. There is a slight increase in the emotional contagion score from the thirty-sixth to the fortieth comment, followed by a decrease in the subsequent comments.

4.5.c. The majority of comments with a neutral sentiment are two or three words long, as indicated by the high counts in 2, 6443 and 3, 2570. As the length of the comments increases, the count decreases. There are some neutral comments that are longer in length, as indicated by the non-zero counts for comments that are 18, 25, and 39 words long. However, the number of comments with a neutral sentiment decreases significantly for comments that are longer than 20 words.

4.5.d. The output of applying the OR operator on the comment sequences shows that the comments in the data set have varying lengths, with a significant number of comments having a length between 2 and 10.

Category	Correlation
Short-Long Comment Sequence	0.0297147
AND-OR	0.322815

Table 4.6: TensorFlow- Neutral Sentiment

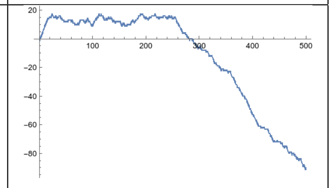
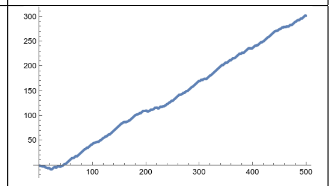
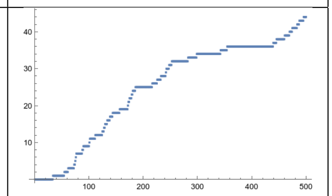
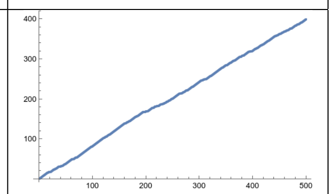
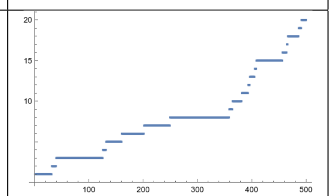
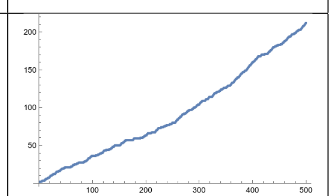
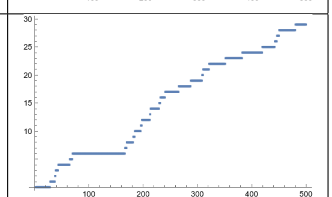
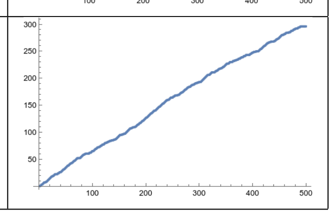
Category	Affect Chart
Short Comments	
Long Comments	
AND Positive	
OR Positive	
AND Negative	
OR Negative	
AND Neutral	
OR Neutral	

Table 4.7: TensorFlow- Affect Charts

4.2 React-Native

Results when tested with React-Native data set are as follows: (Refer A.2 for sequences)

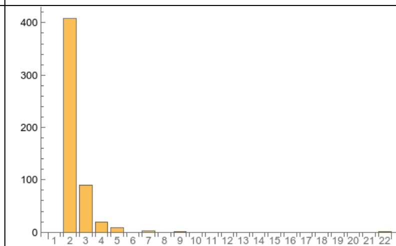
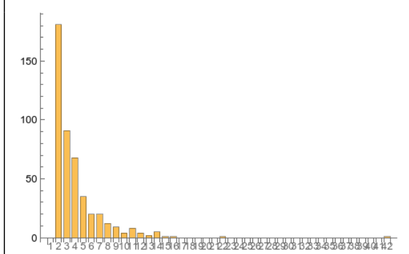
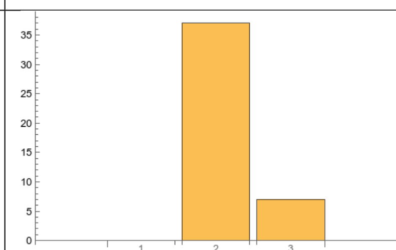
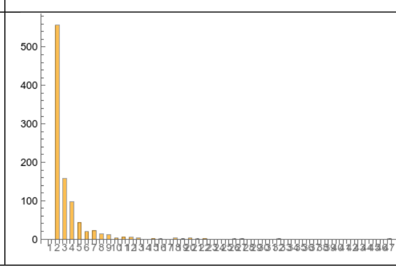
Positive Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.8: React Native- Positive Sentiment

4.8.a. Most occurrences of positive emotional contagion happen in sequences of length 2. There are also a significant number of occurrences in sequences of length 3. The occurrences decrease as the sequence length increases. There is a small number of occurrences in sequences of length 4 and 5.Beyond length 5, the occurrences drop significantly. No occurrences for sequences of length 1, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, and 21 are found. There are 2 occurrences in a sequence of length 7 and one occurrence in a sequence of length 9.

4.8.b. The most frequent occurrences of positive emotional contagion in long comment sequences are observed in sequences of length 2. There is a consistent number of occurrences in sequences of length 3, 4, 5, 6, and 7. The number of occurrences starts to decrease gradually from length 8 onwards. The data shows no occurrences for sequences from length 17 to 41. There are isolated occurrences in sequences of length 10, 11, 12, 13, 14, 15, 16, 22, and 42.

4.8.c. The contagion sequences obtained after applying the AND operation are relatively shorter, with sequence lengths ranging from 1 to 3. The majority of positive contagion sequences have lengths of 2 and 3, with 37 occurrences and 7 occurrences, respectively.

4.8.d. The most common length for positive contagion sequences is 2, with 557 occurrences. Neutral contagion sequences with lengths of 3, 4, 5, and 6 also have significant occurrences. As the length of the sequences increases beyond 6, the number of occurrences decreases.

Category	Correlation
Short-Long Comment Sequence	0.102389
AND-OR	0.227417

Table 4.9: React-Native- Positive Sentiment

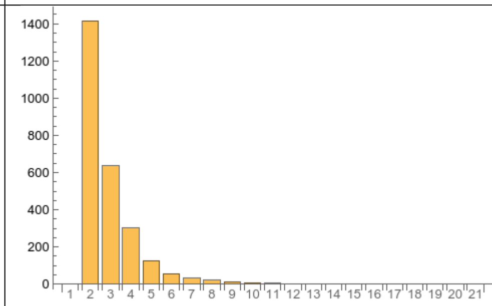
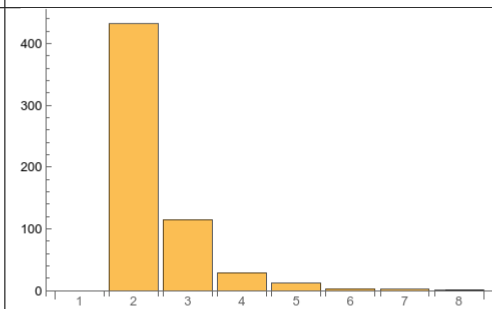
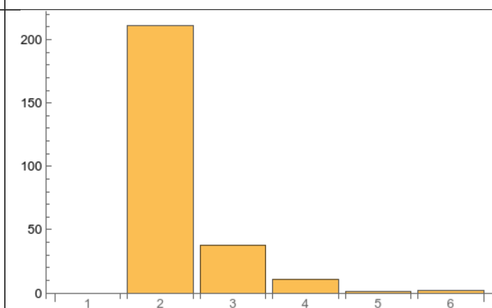
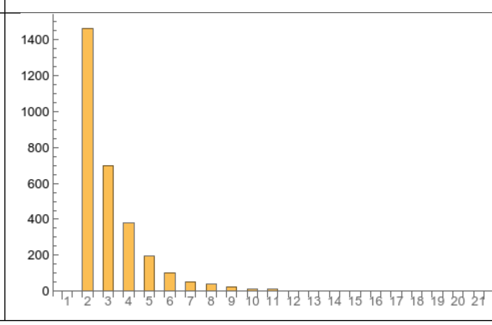
Negative Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.10: React-Native- Negative Sentiment

4.10.a.The most frequent occurrences of negative emotional contagion in short comment sequences are observed in sequences of length 2. There is a significant number of occurrences in sequences of length 3. The number of occurrences decreases as the sequence length increases from 4 to 10. Beyond length 10, the occurrences of negative emotional contagion become infrequent. There are only a few occurrences in sequences of length 11, 13, 15, and 21. The data shows no occurrences for sequences of length 1, 12, 14, 16, 17, 18, 19, and 20 except for one at 21.

4.10.b. The most frequent occurrences of negative emotional contagion in long comment sequences are observed in sequences of length 2. There are relatively fewer occurrences in sequences of length 3. The occurrences further decrease as the sequence length increases. The data shows minimal occurrences in sequences of length 4 to 8.

4.10.c. The contagion sequences obtained after applying the AND operation exhibit a range of lengths, from 2 to 6. The most common lengths for negative contagion sequences are 2 and 3, with 211 occurrences and 38 occurrences, respectively. There are also less occurrences of negative contagion sequences with lengths of 4, 5, and 6.

4.10.d. The most common length for negative contagion sequences is 2, with 1463 occurrences. Neutral contagion sequences with lengths of 3, 4, 5, and 6 also have significant occurrences. There are a few occurrences of neutral contagion sequences with lengths of 13, 14, 15, and 18.

Category	Correlation
Short-Long Comment Sequence	0.102389
AND-OR	0.312419

Table 4.11: React-Native- Negative Sentiment

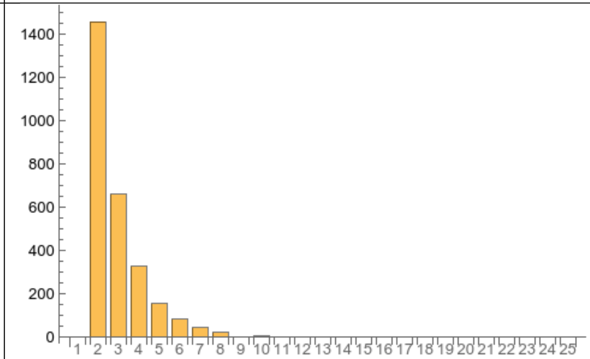
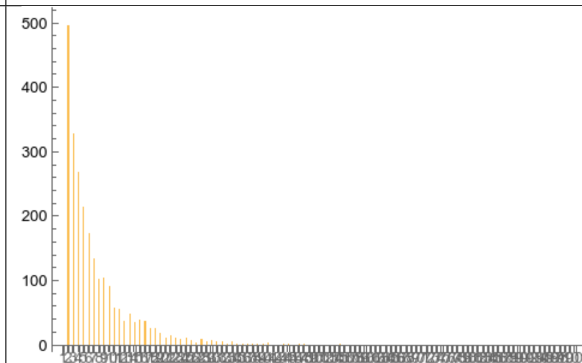
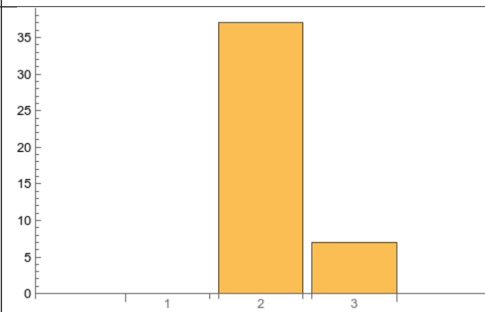
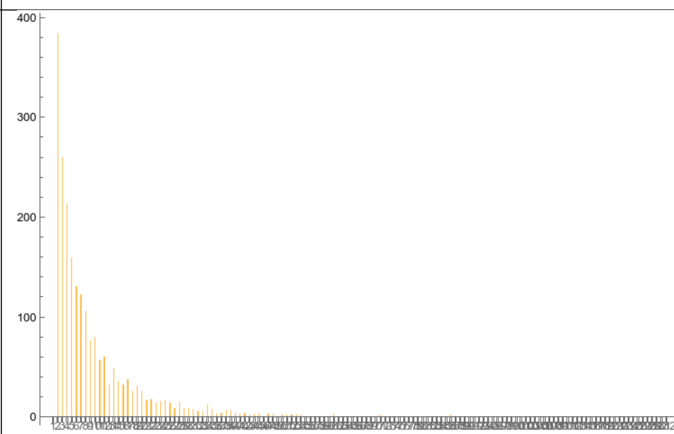
Neutral Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.12: React-Native- Neutral Sentiment

4.12.a. The most frequent occurrences of neutral emotional contagion in short comment sequences are observed in sequences of length 2. There is a significant number of occurrences in sequences of length 3. The number of occurrences decreases as the sequence length increases from 4 to 9. There are rare occurrences in sequences of length 10 to 15. The data shows no occurrences for sequences of length 1, 16, 17, 18, 19, 20, 21, 22, 23, and 24. There is one occurrence in a sequence of length 25.

4.12.b. Neutral emotional contagion is observed in a wide range of sequence lengths in long comment sequences, ranging from 2 to 101. The number of occurrences per length varies, with a trend of decreasing occurrences as the sequence length increases. The most frequent occurrences of neutral contagion are observed in sequences of lengths 2 to 10. There are less occurrences of neutral contagion in longer comment sequences, including sequences of lengths 11 to 101. The data shows no occurrences for sequences of length 1 and for sequence lengths 46, 49, 51, 52, 54, 56 to 59, 61 to 63, 66 to 76, 78 to 87, and 89 to 98. There are a few isolated occurrences in sequences of lengths 99 and 101.

4.12.c. The most common length for neutral contagion sequences is 2, with 37 occurrences. There are also occurrences of neutral contagion sequences with a length of 3.

4.12.d. The contagion sequences obtained after applying the OR operation exhibit a wide range of lengths, from 2 to 132. Neutral contagion sequences with lengths of 2, 3, 5, 6, and 7 also have significant occurrences. As the length of the sequences increases beyond 7, the number of occurrences decreases.

Category	Correlation
Short-Long Comment Sequence	0.102389
AND-OR	0.301003

Table 4.13: React-Native- Neutral Sentiment

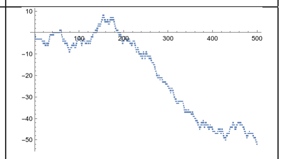
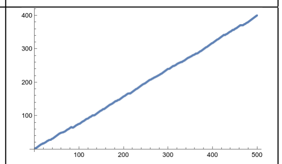
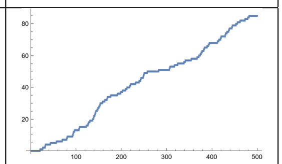
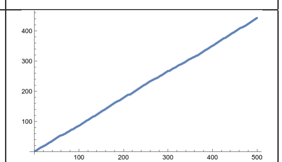
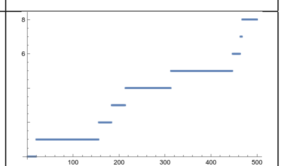
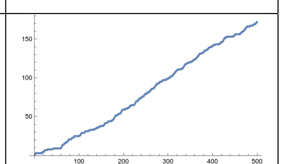
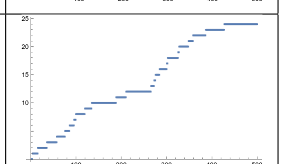
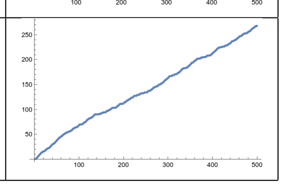
Category	Affect Chart
Short Comments	
Long Comments	
AND Positive	
OR Positive	
AND Negative	
OR Negative	
AND Neutral	
OR Neutral	

Table 4.14: React-Native- Affect Charts

4.3 Keras

Results when tested with Keras dataset are as follows: (Refer A.3 for sequences)

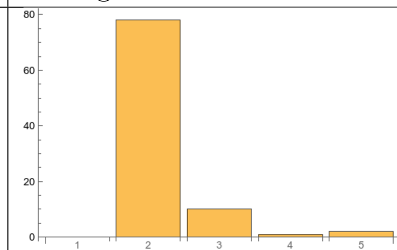
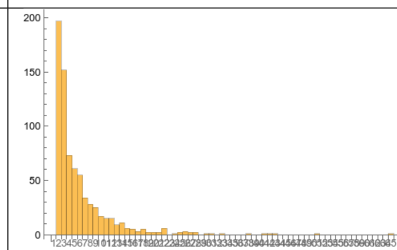
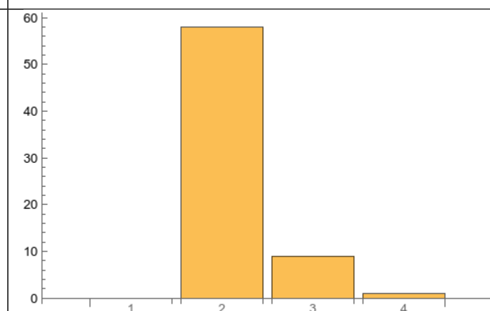
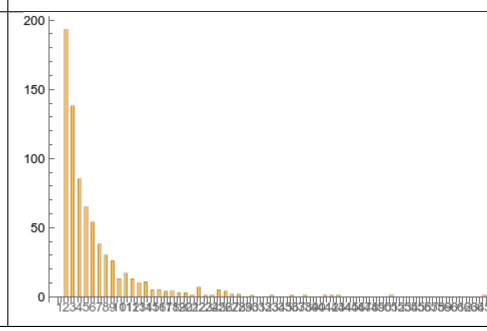
Positive Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.15: Keras- Positive Sentiment

4.15.a. Positive contagion was observed in 78 instances of length 2 sequences. Positive contagion occurred 10 times in length 3 sequences. Positive contagion was observed only once in length 4 sequences and twice in length 5 sequences.

4.15.b. Positive emotional contagion is observed in a range of sequence lengths in long comment sequences, spanning from 2 to 65. There is a decreasing trend in the number of occurrences as the sequence length increases. The most frequent occurrences of positive contagion are observed in sequences of lengths 2 to 10. Also, there are significant occurrences of positive contagion in sequences of lengths 11 to 20. There are no occurrences for sequences of length 1 and for sequence lengths 23, 29, 32 to 36, 38 to 44, 46 to 50, 52 to 56, 58 to 61, 63, and 65. Finally, there are a few isolated occurrences in sequences of lengths 21, 22, 24, 25, 26, 27, 28, 30, 31, 33, 37, 41, 42, 43, 51, and 65.

4.15.c. The most common length for positive contagion sequences is 2, with 58 occurrences. Positive contagion sequences with lengths of 3 and 4 have fewer occurrences compared to length 2. The occurrence of positive contagion decreases as the length of the sequences increases, with only one occurrence for sequences of length 4.

4.15.d. The contagion sequences obtained after applying the OR operation have lengths ranging from 2 to 65. The most common length for contagion sequences is 2, with 193 occurrences. The occurrences gradually decrease as the length of the sequences increases.

Category	Correlation
Short-Long Comment Sequence	0.089631
AND-OR	0.20924

Table 4.16: Keras- Positive Sentiment

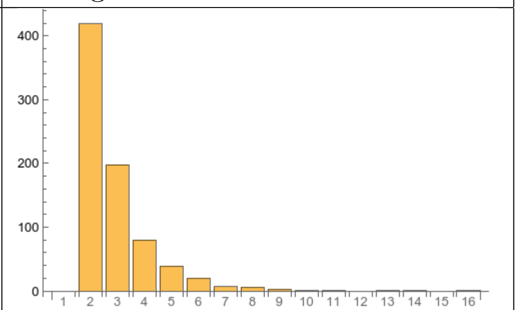
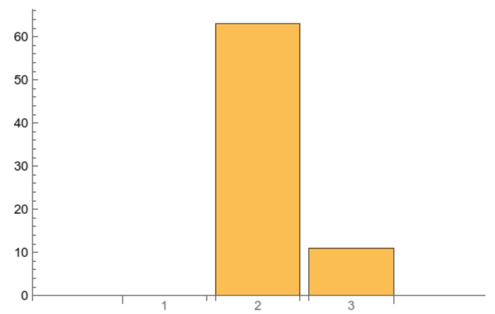
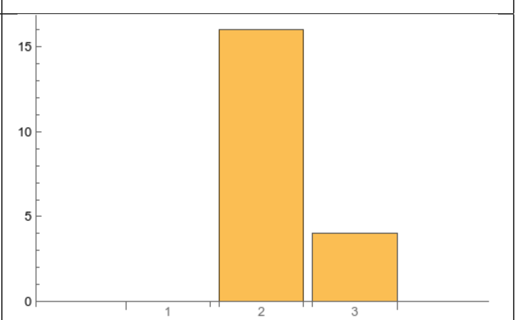
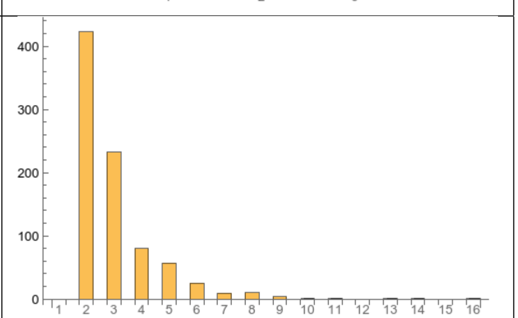
Negative Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.17: Keras- Negative Sentiment

4.17.a. Negative emotional contagion is observed in a range of sequence lengths in short comment sequences, spanning from 2 to 16. The most frequent occurrences of negative contagion are observed in sequences of lengths 2 and 3. Beyond sequence length 5, the occurrences of negative contagion become less frequent. There are isolated occurrences in sequences of lengths 6, 7, 8, 9, 10, 11, 13, and 14.

4.17.b. Negative emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 3. The number of occurrences per length is relatively low compared to short comment sequences. The most frequent occurrences of negative contagion are observed in sequences of length 2.

4.17.c. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 3. The most common length for neutral contagion sequences is 2, with 16 occurrences. Neutral contagion sequences with a length of 3 also have occurrences, with 4 instances.

4.17.d. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 16. The most common length for neutral contagion sequences is 2, with 423 occurrences. 3, 4, 5, 6 have 232, 80, 10, 4 occurrences respectively. Beyond 6 as the length increases occurrences decrease.

Category	Correlation
Short-Long Comment Sequence	0.089631
AND-OR	0.234048

Table 4.18: Keras- Negative Sentiment

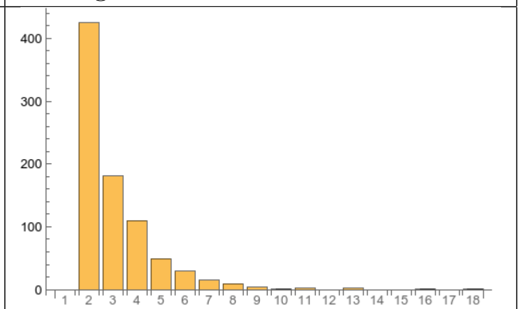
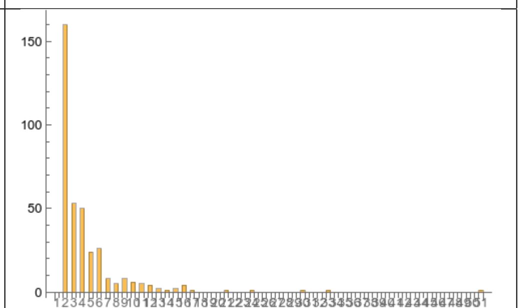
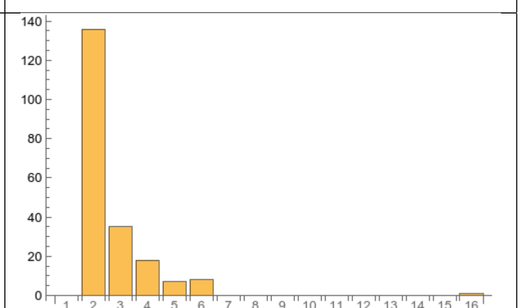
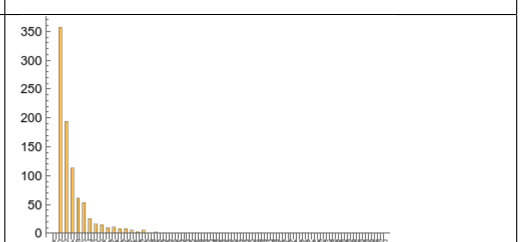
Neutral Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.19: Keras- Neutral Sentiment

4.19.a.Neutral emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2 to 18. The number of occurrences per length varies, with the highest number of occurrences observed in sequences of length 2, followed by sequences of length 3 and 4. As the sequence length increases beyond 4, the number of occurrences per length decreases. The occurrences for sequences of length 14, 15, 17, and 18 are minimal.

4.19.b. Neutral emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 51. The number of occurrences per length varies, with the highest number of occurrences observed in sequences of length 2, followed by sequences of length 3 and 4. As the sequence length increases beyond 4, the number of occurrences per length decreases. The occurrences for sequences of length 18, 19, 20, 22, 23, 25, 26, 27, 28, 29, 30, 32, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and 50 are all zero, indicating that neutral emotional contagion is absent in sequences of these lengths.

4.19.c. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 16. The most common length for neutral contagion sequences is 2, with 136 occurrences. Neutral contagion sequences with lengths of 3, 4, 5, and 6 also have significant occurrences, with 35, 18, 7, and 8 respectively. There are no occurrences of neutral contagion sequences with lengths beyond 6, and a single occurrence at 16.

4.19.d. The contagion sequences obtained after applying the OR operation have lengths ranging from 2 to 56. The most common length for neutral contagion sequences is 2, with 357 occurrences. Neutral contagion sequences with lengths up to 56 are observed, although with decreasing frequencies.

Category	Correlation
Short-Long Comment Sequence	0.089631
AND-OR	0.334704

Table 4.20: Keras- Neutral Sentiment

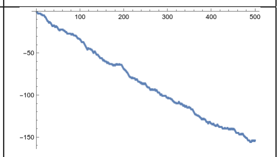
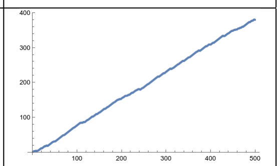
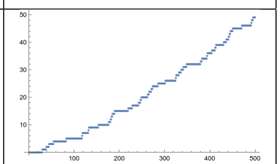
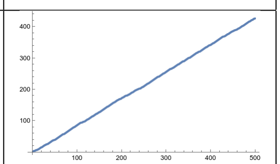
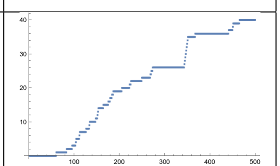
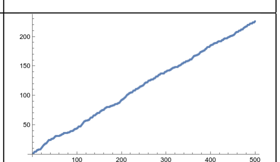
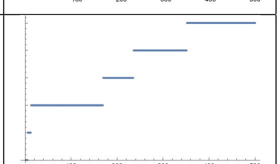
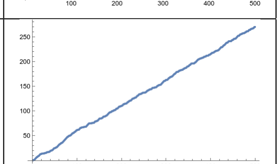
Category	Affect Chart
Short Comments	
Long Comments	
AND Positive	
OR Positive	
AND Negative	
OR Negative	
AND Neutral	
OR Neutral	

Table 4.21: Keras- Affect Charts

4.4 Pytorch

Results when tested with Pytorch dataset are as follows: (Refer A.4 for sequences)

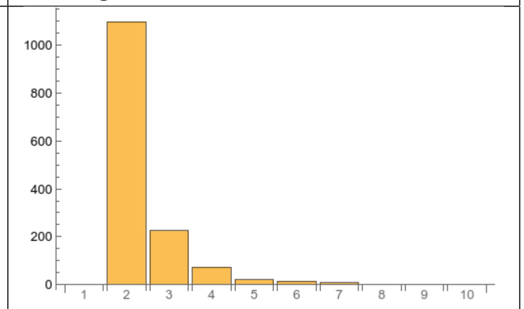
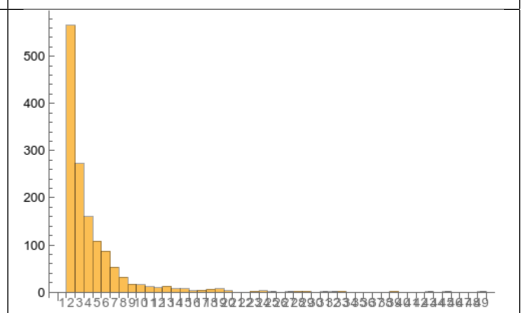
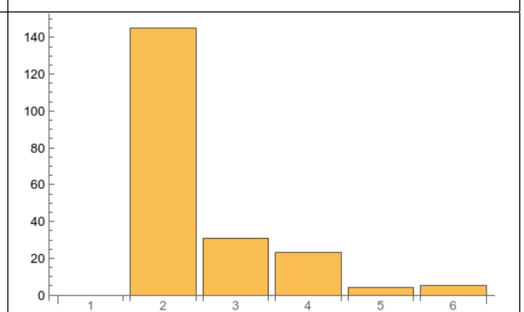
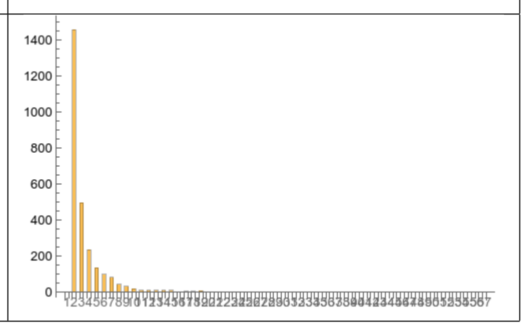
Positive Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.22: Pytorch- Positive Sentiment

4.22.a. Positive emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2 to 10. Highest number of occurrences observed in sequences of length 2, followed by sequences of length 3 and 4. As the sequence length increases beyond 4, the number of occurrences per length decreases, and there are no occurrences for sequences of length 9.

4.22.b. Positive emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 49. The number of occurrences per length varies, with the highest number of occurrences observed in sequences of length 2, followed by sequences of length 3 and 4. Occurrences are observed in sequences of length 5 to 10, with relatively consistent frequencies. There are some rare occurrences in sequences with lengths beyond 10.

4.22.c. The contagion sequences obtained after applying the AND operation have lengths ranging from 2-6. The most common length for positive contagion sequences is 2, with 145 occurrences. Further as the length increases, frequency decreases.

4.22.d. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 57. The most common length for negative contagion sequences is 2, with 1458 occurrences. Negative contagion sequences with lengths beyond 2 are less frequent, with 498 occurrences for length 3, 237 occurrences for length 4, and decreasing frequencies for longer sequences.

Category	Correlation
Short-Long Comment Sequence	0.0248618
AND-OR	0.252579

Table 4.23: Pytorch- Positive Sentiment

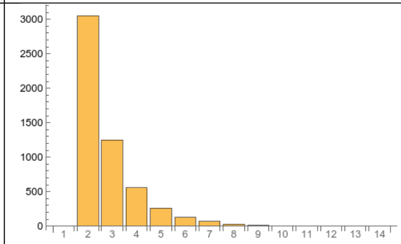
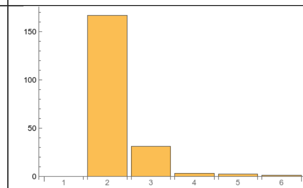
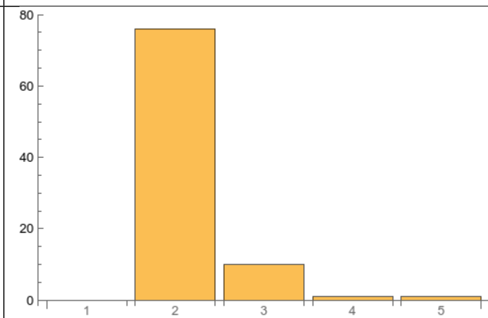
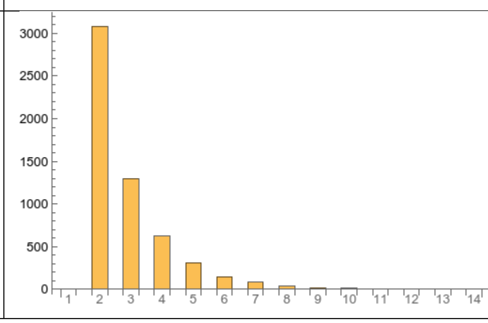
Negative Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.24: Pytorch- Negative Sentiment

4.24.a. Negative emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2 to 14. The number of occurrences per length varies, with the highest number of occurrences observed in sequences of length 2, followed by sequences of length 3 and 4. The number of occurrences gradually decreases as the sequence length increases from 5 to 14.

4.24.b. Negative emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 6. The number of occurrences per length decreases as the sequence length increases. The highest number of occurrences is observed in sequences of length 2, followed by sequences of length 3. As the sequence length increases beyond 3, the number of occurrences decreases significantly. There are only a few occurrences observed in sequences of length 4 to 6.

4.24.c. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 5. The most common length for negative contagion sequences is 2, with 76 occurrences. Negative contagion sequences with lengths beyond 2 are less frequent, with 10 occurrences for length 3 and 1 occurrence each for lengths 4 and 5.

4.24.d. The contagion sequences obtained after applying the OR operation have lengths ranging from 2 to 14. The most common length for positive contagion sequences is 2, with 3082 occurrences. Contagion sequences with lengths beyond 2 are less frequent, with 1300 occurrences for length 3, 621 occurrences for length 4, and decreasing frequencies for longer sequences.

Category	Correlation
Short-Long Comment Sequence	0.0248618
AND-OR	0.183197

Table 4.25: Pytorch- Negative Sentiment

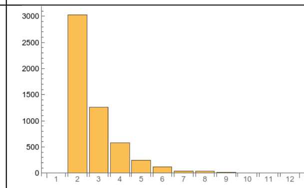
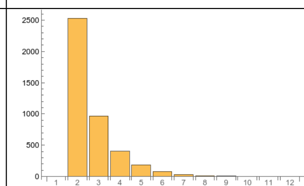
Neutral Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.26: Pytorch- Neutral Sentiment

4.26.a. Neutral emotional contagion is observed in small comment sequences, with sequence lengths ranging from 2 to 12. The highest number of occurrences is observed in sequences of length 2, followed by sequences of length 3 and 4. While the number of occurrences decreases as the sequence length increases, there is a slight increase in occurrences for sequences of length 7 and 8. However, the occurrences decrease again for longer sequences i.e length 9 to 12.

4.26.b. Neutral emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 401. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3 and 4. The occurrences become less frequent as the sequence length increases beyond 8, with rare occurrences observed for longer sequences.

4.26.c. The contagion sequences obtained after applying the AND operation have lengths ranging from 2 to 12. The most common length for positive contagion sequences is 2, with 2529 occurrences. Positive contagion sequences with lengths beyond 2 are less frequent, with 970 occurrences for length 3, 404 occurrences for length 4, 186 occurrences for length 5, and decreasing frequencies for longer sequences.

4.26.d. The contagion sequences obtained after applying the OR operation have lengths ranging from 2 to 401. The most common length for positive contagion sequences is 2, with 664 occurrences. Positive contagion sequences with lengths beyond 2 are less frequent, with decreasing frequencies as the length increases. However, there are still notable occurrences for longer sequences, such as 375 occurrences for length 3, 268 occurrences for length 4, and so on.

Category	Correlation
Short-Long Comment Sequence	0.0248618
AND-OR	0.264544

Table 4.27: Pytorch- Neutral Sentiment

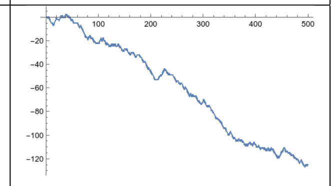
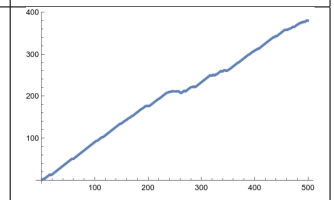
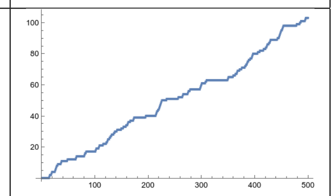
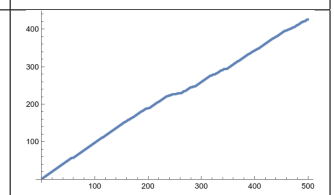
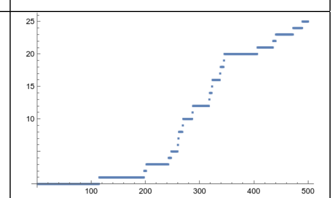
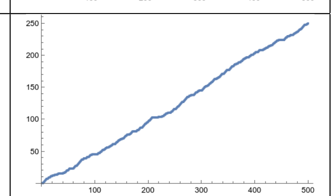
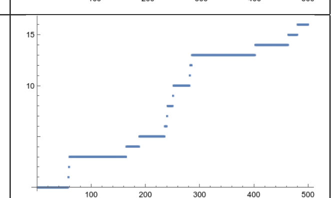
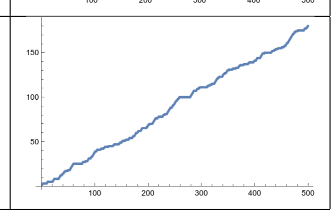
Category	Affect Chart
Short Comments	
Long Comments	
AND Positive	
OR Positive	
AND Negative	
OR Negative	
AND Neutral	
OR Neutral	

Table 4.28: Pytorch- Affect Charts

4.5 Pillow

Results when tested with Pillow data set are as follows: (Refer A.5 for sequences)

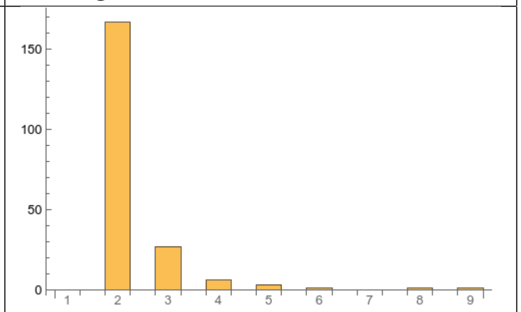
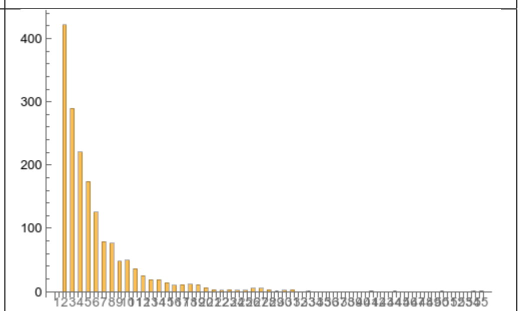
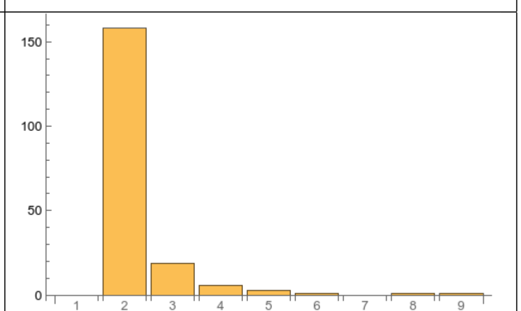
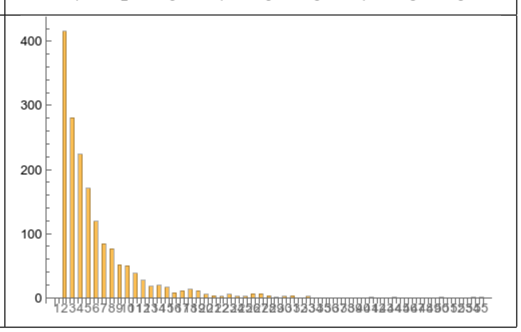
Positive Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.29: Pillow- Positive Sentiment

4.29.a. Positive emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2-9. As the sequence length increases beyond 3, the number of occurrences decreases. There are only a few occurrences for sequences of length 4, 5, 6, 8, 9. There are no occurrences for sequences of length 1 and 7.

4.29.b. Positive emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 55. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3 and 4. As the sequence length increases beyond 4, the number of occurrences starts to decrease gradually. There are no occurrences for sequences of length 1, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 51, 52, 53, and 58. Some longer sequences, such as those of length 50, 54, and 55, still have some occurrences.

4.29.c. Contagion sequences with a length of 2 are the most frequent, occurring 158 times. As the sequence length increases, the frequency of occurrence decreases. Contagion sequences of length 3 occur 19 times, while sequences of length 4 occur 6 times.There is a further decrease in frequency for longer contagion sequences. Sequences of length 5 occur 3 times, and sequences of length 6, 7, 8, and 9 occur once.

4.29.d. Contagion sequences with a length of 2 are the most frequent, occurring 416 times. Contagion sequences with a length of 3 occur 281 times. Contagion sequences with lengths ranging from 4 to 28 are also observed, with decreasing frequency as the length increases.

Category	Correlation
Short-Long Comment Sequence	0.26535
AND-OR	0.200065

Table 4.30: Pillow- Positive Sentiment

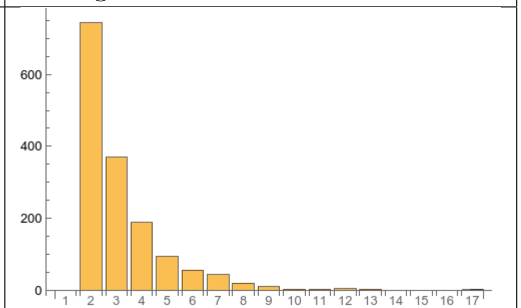
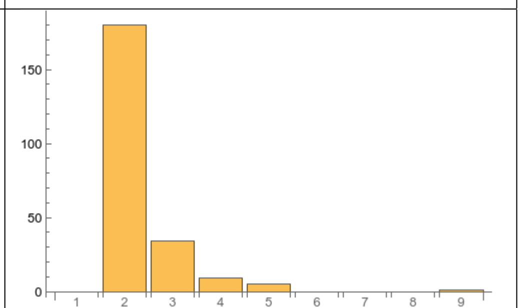
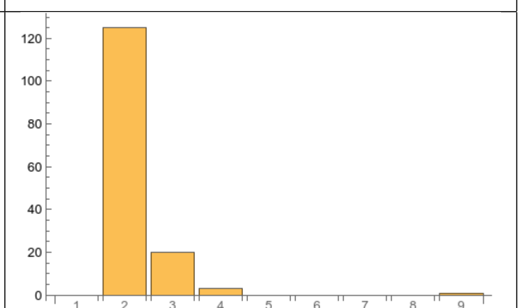
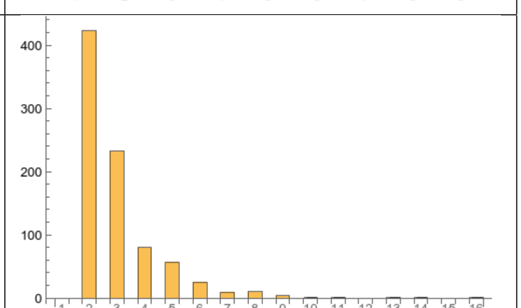
Negative Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.31: Pillow- Negative Sentiment

4.31.a. Negative emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2 to 17. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3 and 4. As the sequence length increases beyond 4, the number of occurrences starts to decrease gradually. There are no occurrences for sequences of length 1, 14, 15, 16. Some longer sequences, such as those of length 10, 11, 12, and 13, still have occurrences.

4.31.b. Negative emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 9. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3. As the sequence length increases beyond 3, the number of occurrences starts to decrease. Sequences of length 4 and 5 still have a few occurrences, but no occurrences are observed for sequences of length 6, 7, and 8. There are no occurrences for sequences of length 1 and 6 to 8. Negative emotional contagion is still observed in longer sequences, with a single occurrence in sequences of length 9.

4.31.c. Contagion sequences with a length of 2 are the most frequent, occurring 125 times. As the sequence length increases, the frequency of occurrence decreases. Contagion sequences of length 3 occur 20 times, while sequences of length 4 occur 3 times. No contagion sequences with a length of 5, 6, 7, or 8 are observed. No contagion sequences with a length of 1 are observed.

4.31.d. Contagion sequences with a length of 2 are the most frequent, occurring 423 times. Contagion sequences with a length of 3 occur 232 times. Contagion sequences with lengths ranging from 4 to 6 are also observed, although with decreasing frequency as the length increases. Contagion sequences with lengths 7 and above occur less frequently, with single occurrences for lengths 10, 11, 13, 14, and 16.

Category	Correlation
Short-Long Comment Sequence	0.26535
AND-OR	0.309103

Table 4.32: Pillow- Negative Sentiment

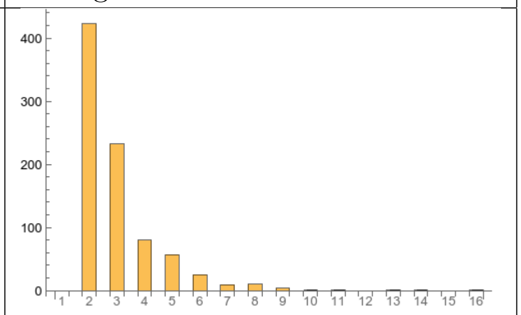
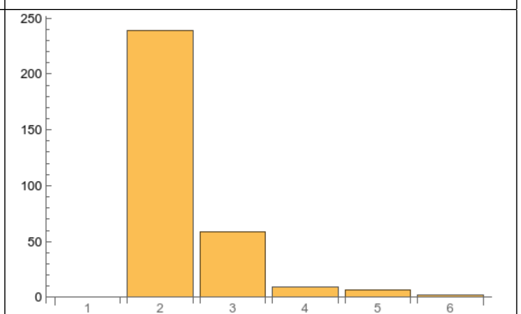
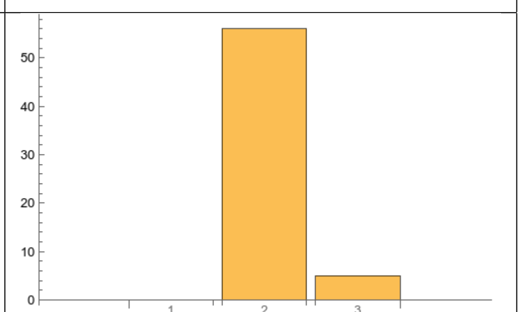
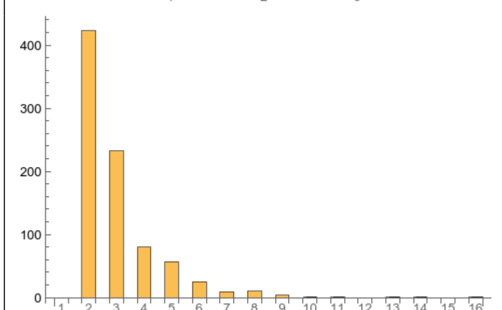
Neutral Sentiment	Histogram
a.Short Comment Sequence	
b.Long Comment Sequence	
c.AND	
d.OR	

Table 4.33: Pillow- Neutral Sentiment

4.33.a. Neutral emotional contagion is observed in short comment sequences, with sequence lengths ranging from 2 to 20. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3. As the sequence length increases beyond 3, the number of occurrences starts to decrease. Sequences of length 4 and 5 still have a significant number of occurrences, but the frequency decreases for longer sequence lengths. There are no occurrences for sequences of length 1, 13, 14, 15, 16, 17, 18, and 19. Neutral emotional contagion is still observed in longer sequences, with a few occurrences in sequences of length 6 to 12 and a single occurrence of length 20.

4.33.b. Neutral emotional contagion is observed in long comment sequences, with sequence lengths ranging from 2 to 6. The highest number of occurrences is observed for sequences of length 2, followed by sequences of length 3. As the sequence length increases beyond 3, the number of occurrences becomes less frequent. Sequences of length 4 and 5 still have a few occurrences, but the frequency decreases further. There are no occurrences for sequences of length 1. Neutral emotional contagion is observed in longer sequences, with a few occurrences in sequences of length 4, 5, and 6 within the given long comment sequences.

4.33.c. Contagion sequences with a length of 2 are the most frequent, occurring 56 times. Contagion sequences with a length of 3 occur 5 times.

4.33.d. Contagion sequences with a length of 2 are the most frequent, occurring 786 times. Contagion sequences with lengths of 3 and 4 are also relatively common, occurring 407 and 222 times, respectively. Contagion sequences with lengths ranging from 5 to 9 are observed, although with decreasing frequency as the length increases.

Category	Correlation
Short-Long Comment Sequence	0.265351
AND-OR	0.211636

Table 4.34: Pillow- Neutral Sentiment

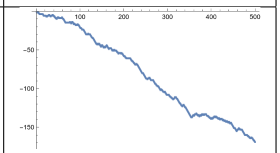
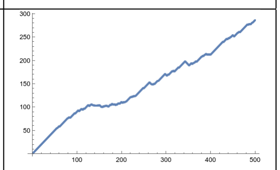
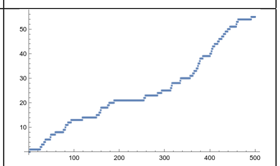
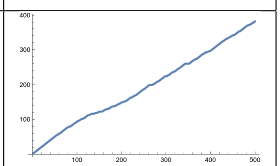
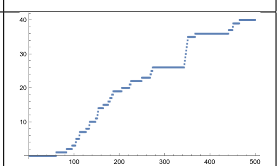
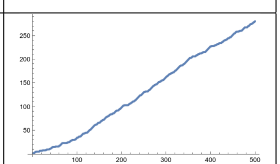
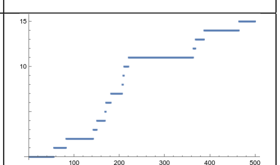
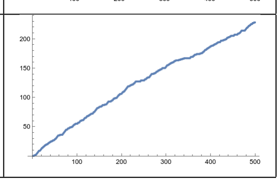
Category	Affect Chart
Short Comments	
Long Comments	
AND Positive	
OR Positive	
AND Negative	
OR Negative	
AND Neutral	
OR Neutral	

Table 4.35: Pillow - Affect Charts

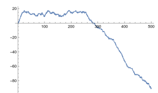
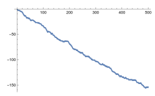
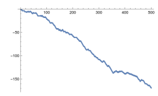
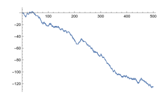
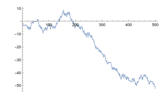
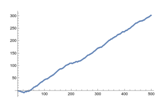
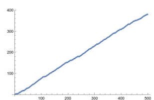
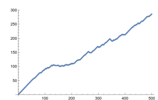
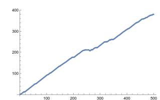
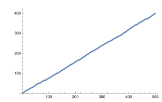
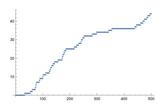
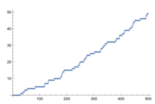
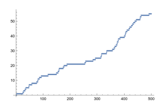
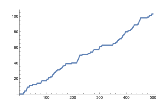
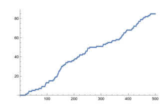
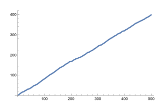
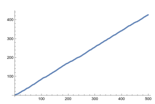
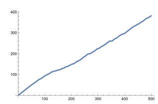
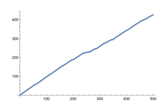
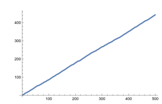
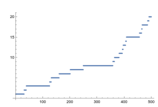
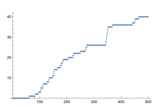
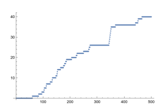
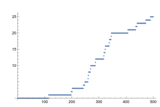
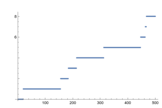
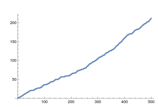
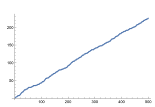
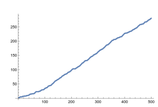
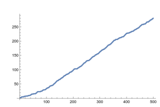
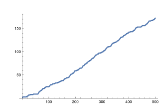
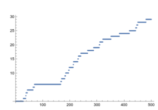
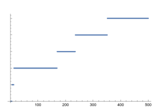
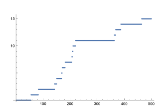
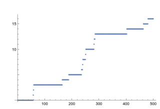
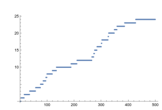
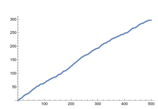
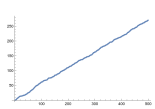
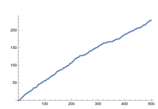
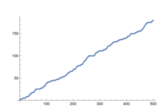
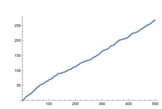
Category	Tensorflow	Keras	Pillow	Pytorch	React-Native
Short Comment					
Long Comment					
AND Positive					
OR Positive					
AND Negative					
OR Negative					
AND Neutral					
OR Neutral					

Table 4.36: Affect Charts of all data sets used

Chapter 5

Discussion

In this chapter, we discuss the findings and implications of our research on detecting emotional contagion in collaborative software development. Additionally, we will highlight the limitations of our study and suggest avenues for future research.

Our study aimed to investigate the presence and impact of emotional contagion in open source software systems. By analyzing comments on five GitHub repositories and applying sentiment analysis, we were able to identify emotional expressions. The findings of our study revealed several important insights. We were able to detect emotional contagion on a wide variety of data sets which suggests that the methodology used is versatile and can be reused on a large number of data sets.

The consistency of sentiment classification across different lengths of comment sequences is a crucial aspect of our research findings. It indicates that the sentiment classifier employed in our study is reliable and robust, capable of accurately capturing and classifying emotions in various contexts and communication styles within the open software collaborations. The results obtained on implementing AND Operator suggested that the sentiment conveyed in both long and short comment sequences were consistent. It showed that the sentiment classifier used is reliable and consistent. The reliability of the sentiment classifier is of great significance for future applications and research in the field as it can provide accurate insights and help businesses make informed decisions.

The implementation of the logical OR operator in our study yielded noteworthy insights regarding the presence of sentiment across different lengths of comment sequences i.e short and long. Specifically, our findings indicated that if we were searching for a particular sentiment, such as positive, negative, or neutral, we could identify instances of that sentiment in either short or long comment sequences. This finding holds significant implications for sentiment analysis within the context of open software collaborations. It suggests that the occurrence of a specific sentiment is not confined to a particular comment length. Instead, sentiment expression can exist in both short and long comment sequences, highlighting the flexibility and variability in how individuals convey their emotions within collaborative software development. The ability to identify the desired sentiment in both short and long comment sequences expands the practical applications of sentiment analysis in open software collaborations.

Looking at the correlation values provided, we can make some comments on the connect between short comments, long comments, AND of short and long comments , OR of short and long comments. Firstly, we can observe that the positive sentiment category has a positive correlation with all five GitHub repositories, indicating that as the positive sentiment increases, so does the mention of these repositories. The strength of the correlation varies from 0.20 for Pillow to 0.25 for PyTorch, but all correlations are positive. Secondly, we can observe that the negative sentiment category has a negative correlation with TensorFlow, Keras, and PyTorch, indicating that as the negative sentiment increases, the mention of these data sets decreases. The strength of these correlations is relatively weak, ranging from -0.18 for PyTorch to -0.23 for Keras. However, there is a positive correlation between the negative sentiment and Pillow and React Native, indicating that as the negative sentiment increases, so does the mention of these two data sets. The strength of these correlations is also relatively weak, ranging from 0.20 for Pillow to 0.31 for React Native. Finally, we can observe that the neutral sentiment category has a positive correlation with Keras and React Native, indicating that as the neutral sentiment increases, so does the mention of these data sets. The strength of these correlations is moderate, ranging from 0.30 for React Native to 0.33 for Keras. However, there is a weak positive correlation between the neutral sentiment and TensorFlow and PyTorch, and a weak negative correlation with Pillow.

Data set	Sentiment	Long-Short Comment	AND-OR
Tensorflow	Positive	0.029	0.245
	Negative	0.029	0.224
	Neutral	0.029	0.322
React Native	Positive	0.102	0.227
	Negative	0.102	0.312
	Neutral	0.102	0.301
Keras	Positive	0.089	0.209
	Negative	0.089	0.234
	Neutral	0.089	0.334
Pytorch	Positive	0.024	0.252
	Negative	0.024	0.183
	Neutral	0.024	0.264
Pillow	Positive	0.265	0.200
	Negative	0.265	0.309
	Neutral	0.265	0.211

Table 5.1: Correlations

Hence, based on these correlation values, we can observe some similarities and differences between the data sets. While all data sets have a positive correlation with the positive sentiment category, the direction of the correlation for the negative and neutral sentiment categories varies across the GitHub repositories. Additionally, the strength of the correlations also varies between the GitHub repositories and sentiment categories. Therefore, while there may be some similarities between the data sets (such as the positive correlation between the positive sentiment category and all data sets), the variations in the correlation values suggest that the data sets are distinct from each other.

5.1 Ethics of using comments from Github

Ethical use of information is becoming increasingly important, especially since generative AI tools have been available to the public. Such tools require very large amount of data to train, but it is unclear if such data includes copyrighted material; and it is even less clear if the responses produced by such tools can be considered original or derivatives. In our study, we chose Open Source Software because of its inherently open nature: contributions and elaborations, including derivative work, are encouraged. We also chose a public platform, GitHub, which makes every user agree that public material will be made available to all on the Internet. Additionally, the Fair Dealing Exception of the Copyright Act allows one to make use of copyrighted materials for research, as long as the use of such material is considered "fair," which has been defined in court, and as far as we understand our work fits under the definition. Table 5.2 shows the licenses of the five databases we consulted. These licenses are all more permissive than the Fair Dealing Exception, and in most cases even allow commercial use of the software. This provides solid ethical grounds for our work at the time of writing. It is notable, however, that because of the popularity of generative AI tools and their use of all Internet resources for training, a number of court cases are currently being heard that might have unpredictable results in the future. Thus, ethical considerations are evolving and it will be important to adjust research policies accordingly in the future.

Repository	Licensing details
Tensorflow	Apache License 2.0- A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code. Permissions : Commercial use, Modification, Distribution, Patent use, Private use.
Keras	Apache License 2.0- A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code. Permissions : Commercial use, Modification, Distribution, Patent use, Private use.
Pytorch	Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met: 1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. 2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution. 3. Neither the names of Facebook, Deepmind Technologies, NYU, NEC Laboratories America and IDIAP Research Institute nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.
React Native	MIT License- A short and simple permissive license with conditions only requiring preservation of copyright and license notices. Licensed works, modifications, and larger works may be distributed under different terms and without source code. Permissions-Commercial use, Modification, Distribution, Private use.

Table 5.2: Github repositories and their licenses

Repository	Licensing details
Pillow	Permission to use, copy, modify and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appears in all copies, and that both that copyright notice and this permission notice appearing supporting documentation , and that the name of Secret Labs AB or the author not be used in advertising or publicity pertaining to distribution of the software without specific, written prior permission.

Table 5.3: Pillow repository and it's license

5.2 Limitations

Emotional contagion can occur in different contexts, such as in face-to-face interactions, through social media, or in virtual environments. The contextual differences can affect the nature of emotional contagion, making it difficult to develop universal measures for detecting it. While the method we study here is potentially applicable to a wide variety of GitHub repositories, it has some limitations.

Firstly, the comments received on GitHub repositories were classified into three categories - positive, negative, and neutral - using the in-built sentiment classifier in Mathematica. However, emotions can be classified into a much wider range than just these three categories. For instance, people can experience emotions such as happiness, anger, sadness, disgust, surprise, and fear, among others. The in-built sentiment classifier in Mathematica, with its fixed set of sentiment categories (Positive, Negative, Neutral, Uncertain), may not be able to accurately capture the full spectrum of emotions expressed in the comments.

Secondly, the in-built sentiment classifier in Mathematica might not always accurately identify the sentiment of a piece of text. This can be attributed to the fact that sentiment analysis is a complex task and involves understanding the nuances of language. For instance, the sentiment classifier may encounter difficulties in discerning the technical and non-technical forms of language that influence the sentiment of the text. The application of a more focused sentiment analysis tool could potentially alleviate this limitation.

To address these limitations, future research could involve developing a more nuanced sentiment analysis framework that can capture a wider range of emotions and better account for the nuances of language. This could involve training sentiment classifiers on a more diverse range of sentiment categories and incorporating more advanced natural language processing techniques that can better handle figurative language and other nuances in language. Additionally, collecting a larger and more diverse data set of comments in different languages like French, Chinese, Hindi etc and from different sources could help improve the generalizability of the sentiment analysis results.

Language barriers can pose a challenge when analyzing text data for sentiment analysis. The in-built sentiment classifier in Mathematica is a pre-trained model that has been primarily trained on English language text. As a result, the accuracy of the sentiment classifier may be lower for text in other languages that it was not trained on. This means that if this sentiment analysis framework is used on text data in other languages, the results may not be as accurate as they would be for English language text.

To overcome language barriers in sentiment analysis, it might be useful to use a sentiment analysis framework that has been specifically trained on text data in the language(s) of interest. This could typically involve locating or generating a data set of text data in the desired language(s) and training a sentiment classifier tailored for that language(s). Alternatively, a more general sentiment analysis framework trained on multiple languages can be utilized to mitigate the effects of language barriers.

Emotional contagion research typically takes place in controlled laboratory settings where researchers can manipulate certain variables and control for other factors that may influence emotional contagion. However, in open software collaborations, emotional contagion can be influenced by a variety of contextual factors that are beyond the control of researchers. For example, project deadlines, technical challenges, or social dynamics can all impact emotional contagion dynamics in open software collaborations. This means that the results obtained from emotional contagion research in open software collaborations may not be directly comparable to those obtained in laboratory settings [25].

Furthermore, our research has focused on analyzing emotional contagion in open software collaborations using GitHub as a platform. However, using an open software collaboration platform like GitHub has its own set of disadvantages. In particular, communication often takes place online, which can impact emotional contagion dynamics as some participants may choose to remain anonymous, which can also impact emotional contagion dynamics. The anonymity present in open software collaborations may have impacted the validity of our results.

In addition to contextual factors, emotional contagion can also be influenced by gender and diversity factors. For example, the proportion of women or under-represented groups in a collaboration, or the ways in which emotions are expressed differently across cultures can all impact emotional contagion dynamics. However, our research does not address these gender and diversity constraints.

The size and structure of a collaboration may impact the spread of emotional contagion. For example, in larger collaborations like TensorFlow repository comments, emotions may spread more slowly or be less impactful, while in smaller collaborations like Keras, emotional contagion may be more intense. Our research does not discuss the significance of size and structure of collaborations.

Chapter 6

Conclusion

Detecting emotional contagion in open source development is crucial to improve collaboration, communication, code quality, and the health and well-being of team members. Emotional contagion can negatively impact team productivity, lead to conflicts, and affect the quality of the code produced [45]. By detecting emotional contagion early on, teams can prevent negative consequences and promote a positive work environment. We studied communication patterns and emotional expressions in five open source software development projects by conducting analysis. We used an algorithm which uses sentiment analysis to measure emotional contagion and identify the emotional influences in GitHub repository. We used an in-built sentiment classifier in Mathematica to detect positive, negative and neutral sentiments. We used Git repository commands, to retrieve information on all modifications stored in the repository. The information we gathered includes the change log and commits notes, which we processed for compatibility with Mathematica, the tool we use for analysis. We performed sentiment analysis on five data sets namely TensorFlow, Keras, Pytorch, Pillow and React Native using Mathematica's sentiment analysis tool.

The main research question of this study was *Can an existing algorithm be applied to large GitHub repositories to detect emotional contagion?* We determined the presence of episodes of emotional contagion using quantitative analysis method, which we propose as an improvement to our previous visual analysis method. We performed separate analysis on both short and long comment sequences. We found occurrences of emotional contagion in the comments of all repositories. We also incorporated a method to create a list by applying AND logic and OR logic on long and short comment sequence. We then performed analysis on those lists to compare the results obtained. AND logic adheres to the following theory that whenever a positive occurrence occurs in both a long and a short comment sequence, we count it as positive, or 1. When there are same occurrences in both short and long comment sequence, i.e negative or neutral, we count it as positive or 1. When the long and short comment values do not match, we assign it a value of 0. In using logical OR operator, if either sequence (long or short) contains the sentiment we are seeking, we could count it as 1. If not, we are uninterested and count it as 0. For instance, the output of the OR logical operator for positive sentiment is 1 if the long comment sequence has positive sentiment (1) and the short comment sequence has negative emotion (-1). Further, we tried to count sequences of the sentiment that we are investigating for. For example, while looking for positive contagion we say :- SequenceCount[shortCommentSequence, 1, 1, 1]. The results showed that there was connection between the sentiments in both long and short comments. The second research question that we answer in this study *is the algorithm used to detect emotional contagion robust?* Algorithm is robust which was inferred from observing the correlation values which were not very high. The contributions of this thesis can be clearly listed as follows:

1. Algorithm to detect emotional contagion in Open Source Software Environment can be applied to large Github Repositories.

- Episodes of emotional contagion detected in both short and long comments.

- Episodes of emotional contagion detected in AND, OR sequences created for determining consistency of sentiment in both short and long comment sequences.

2. Algortithm used to detect emotional contagion is robust which was inferred from observing the correlation values.

For future work, different lines of research can be taken. We have only produced results for AND and OR logics; XOR, NOT, and NAND logic operators have not been investigated. It would be interesting to know the extent of emotional contagion present when these logic operations are applied on long and short comment sequence. XOR could interpret the presence or absence of certain sentiment, NOT could reverse the sentiment polarity and NAND would have opposite results than what AND achieved. We also only take into account asynchronous exchanges across a single point of contact, which is a GitHub repository. It will be intriguing if the investigated algorithm can be applied to Gitlab, Bitbucket and similar platforms. Sentiment analysis simply reveals the kind of emotion (Positive, Neutral, or Negative) present in a text, not its intensity. Our reliance on a general-purpose sentiment analysis tool limits the method's efficacy because, in order to make this repeatable, we must make sure that the same classifier is applied to each data set.

Bibliography

[1] Carolina Herrando and Efthymios Constantinides. Emotional contagion: a brief overview and future directions. *Frontiers in psychology*, 12:712606, 2021.

[2] Luigi Benedicenti. Detecting emotional contagion in collaborative software development.

[3] "opencv/opencv: Open source computer vision library.". *Github*, Last Date accessed: 11 June 2023.

[4] Ashish Kakar and Akshay Kakar. A brief history of software development and manufacturing. 2020.

[5] Barry W. Boehm. A spiral model of software development and enhancement. *Computer*, 21(5):61–72, 1988.

[6] Gisle Hannemyr. Technology and pleasure: Considering hacking constructive. *First Monday*, 1999.

[7] "what is sdlc waterfall model?" software testing help, www.softwaretestinghelp.com/what-is-sdlc-waterfall-model/. Last Accessed: 11 June 2023.

[8] Daniel A Wren. The centennial of frederick w. taylor's the principles of scientific management: A retrospective commentary. *Journal of Business & Management*, 17(1), 2011.

[9] RW Bemer. Position papers for panel discussion: The economics of program production. information processing 68, 1969.

[10] Michael A Cusumano et al. The software factory: origins and popularity in japan. 1988.

[11] Woonghee Tim Huh. Software process improvement: operations perspectives. In *PICMET'01. Portland International Conference on Management of Engineering and Technology. Proceedings Vol. 1: Book of Summaries (IEEE Cat. No. 01CH37199)*, volume 1, pages 428–429. IEEE, 2001.

[12] M Vivek Bhatnagar and M Ahmad Rather. A comparative study of software development life cycle models. *International Journal of Application or Innovation in Engineering & Management (I JA I EM)*, 4(10), 2015.

[13] Paolo Tell, David Raffo, Liguo Huang, Igor Steinmacher, Ricardo Britto, Eray Tüzün, and Paul Clarke. Summary of the 1st icssp-icgse joint event. *ACM SIGSOFT Software Engineering Notes*, 45(4):31–34, 2020.

[14] "build software better, together." github, github.com/about. *Github*, Last Accessed: 11 June 2023.

[15] Abdulaziz Alhubaishy and Luigi Benedicenti. Toward a model of emotion influences on agile decision making. In *2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, pages 48–51. IEEE, 2017.

[16] Clark Briggs and Patrick Little. Impacts of organizational culture and personality traits on decision-making in technical organizations. *Systems Engineering*, 11(1):15–26, 2008.

[17] Murray R Barrick, Greg L Stewart, Mitchell J Neubert, and Michael K Mount. Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology*, 83(3):377, 1998.

[18] Bjørnar Tessem. An empirical study of decision making, participation, and empowerment in norwegian software development organisations. In *Agile Processes in Software Engineering and Extreme Programming: 12th International Conference, XP 2011, Madrid, Spain, May 10-13, 2011. Proceedings 12*, pages 253–265. Springer, 2011.

[19] Andrew M. Colman. "affect", https://shorturl.at/acc17. Last Accessed: 8 August 2023.

[20] V Vijayalakshmi and Sanghamitra Bhattacharyya. Emotional contagion and its relevance to individual behavior and organizational processes: A position paper. *Journal of Business and Psychology*, 27:363–374, 2012.

[21] Wataru Sasaki, Yuuki Nishiyama, Tadashi Okoshi, and Jin Nakazawa. Investigating the occurrence of selfie-based emotional contagion over social network. *Social Network Analysis and Mining*, 11(1):8, 2021.

[22] Waltraut J Stein. *On the problem of empathy: The collected works of Edith Stein Sister Teresa Bendicta of the cross discalced carmelite volume three.* Springer Science & Business Media, 2012.

[23] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. Emotional contagion. *Current directions in psychological science*, 2(3):96–100, 1993.

[24] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. Unhappy developers: Bad for themselves, bad for process, and bad for software product. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*, pages 362–364. IEEE, 2017.

[25] Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[26] Lingjia Li, Jian Cao, and David Lo. Sentiment analysis over collaborative relationships in open source software projects. 2020.

[27] Jiajing Guo and Susan R Fussell. A preliminary study of emotional contagion in live streaming. In *CONFERENCE Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, pages 263–268, 2020.

[28] Karim R Lakhani and Eric Von Hippel. How open source software works: "free" user-to-user assistance. *Produktentwicklung mit virtuellen Communities: Kundenwünsche erfahren und Innovationen realisieren*, pages 303–339, 2004.

[29] Barbara L Fredrickson. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, 56(3):218, 2001.

[30] Edward Loper and Steven Bird. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*, 2002.

[31] Lorenzo Coviello, Yunkyu Sohn, Adam DI Kramer, Cameron Marlow, Massimo Franceschetti, Nicholas A Christakis, and James H Fowler. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315, 2014.

[32] Luigi Benedicenti, Tommi Mikkonen, and Jurka Rahikkala. Analysing the sentiments in a hybrid floss community based on commits.

[33] Bing Liu et al. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010):627–666, 2010.

[34] Daniel Pletea, Bogdan Vasilescu, and Alexander Serebrenik. Security and emotion: sentiment analysis of security discussions on github. In *Proceedings of the 11th working conference on mining software repositories*, pages 348–351, 2014.

[35] Emitza Guzman, David Azócar, and Yang Li. Sentiment analysis of commit comments in github: an empirical study. In *Proceedings of the 11th working conference on mining software repositories*, pages 352–355, 2014.

[36] EMHL Ekanayake and TD Weerasinghe. Impact of emotional contagion on work performance of employees. *Kelaniya Journal of Human Resource Management*, 14(2), 2019.

[37] Rashmi Dhakad and Luigi Benedicenti. Detecting emotional contagion in oss projects. In *The 7th World Congress on Electrical Engineering and Computer Systems and Science*, 2021.

[38] Luigi Benedicenti. Emotional contagion in open software collaborations. In *Open Source Systems: 16th IFIP WG 2.13 International Conference, OSS 2020, Innopolis, Russia, May 12–14, 2020, Proceedings 16*, pages 47–54. Springer, 2020.

[39] Wolfram. https://www.wolfram.com/mathematica/. Last Accessed: 1 December 2023.

[40] "an open source machine learning framework for everyone." github, github.com/tensorflow/tensorflow. *Github*, Last Accessed: 11 June 2023.

[41] "deep learning for humans." github, github.com/keras-team/keras. *Github*, Last Accessed: 11 June 2023.

[42] "tensors and dynamic neural networks in python with strong gpu acceleration." github, github.com/pytorch/pytorch. *Github*, Last Accessed: 11 June 2023.

[43] "a framework for building native applications using react." github, github.com/facebook/react-native. *Github*, Last Accessed: 11 June 2023.

[44] "python imaging library (fork)." github, github.com/python-pillow/pillow. *Github*, Last Accessed: 11 June 2023.

[45] Daniel Graziotin, Fabian Fagerholm, Xiaofeng Wang, and Pekka Abrahamsson. Consequences of unhappiness while developing software. In *2017 IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion)*, pages 42–47. IEEE, 2017.

Appendix A

Results

A.1 Tensorflow

Positive Sentiment - Short Comment Sequence

{1, 0}, {2, 1694}, {3, 270}, {4, 73}, {5, 16}, {6, 12}, {7, 3}, {8, 9}, {9, 1}, {10, 0}, {11, 0}, {12,0}, {13, 0}, {14, 0}, {15, 1}, {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 0}, {22, 0}, {23,0}, {24, 0}, {25, 0}, {26,1}

Long Comment Sequence

{1, 0}, {2, 2780}, {3, 1104}, {4, 598}, {5, 381}, {6, 243}, {7, 168}, {8, 117}, {9, 107}, {10, 68}, {11, 67}, {12, 32}, {13, 35}, {14, 32}, {15, 29}, {16, 11}, {17, 8}, {18, 14}, {19, 9}, {20, 5}, {21,5}, {22, 6}, {23, 4}, {24, 7}, {25, 3}, {26, 3}, {27, 1}, {28, 4}, {29, 1}, {30, 1}, {31, 0}, {32,3}, {33, 1}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 1}, {42, 1}, {43,0}, {44, 0}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 0}, {52, 0}, {53, 0}, {54,0}, {55, 0}, {56, 0}, {57, 0}, {58, 0}, {59, 0}, {60, 0}, {61, 0}, {62, 0}, {63, 1}

AND

{1, 0}, {2, 261}, {3, 47}, {4, 20}, {5, 3}, {6, 6}, {7, 4}, {8, 4}

OR

{1, 0}, {2, 4397}, {3, 1650}, {4, 815}, {5, 466}, {6, 289}, {7, 191}, {8, 155}, {9, 116}, {10, 74}, {11, 73}, {12, 32}, {13, 41}, {14, 33}, {15, 29}, {16, 19}, {17, 11}, {18, 17}, {19, 15}, {20, 8}, {21, 8}, {22, 6}, {23, 7}, {24, 9}, {25, 4}, {26, 6}, {27, 1}, {28, 6}, {29, 1}, {30, 1}, {31, 0}, {32,4}, {33, 1}, {34, 1}, {35, 1}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 1}, {42, 1}, {43,0}, {44, 0}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 0}, {52, 0}, {53, 0}, {54,0}, {55, 0}, {56, 0}, {57, 0}, {58, 0}, {59, 0}, {60, 0}, {61, 0}, {62, 0}, {63, 0}, {64, 0}, {65,1}

A.1.1 Negative Sentiment

Short Comment Sequence

{1, 0}, {2, 7367}, {3, 2975}, {4, 1259}, {5,536}, {6, 252}, {7, 123}, {8, 46}, {9, 25}, {10, 8}, {11, 5}, {12, 3}, {13, 2}, {14, 2}, {15, 1}, {16, 1}, {17, 1}, {18, 0}, {19, 0}, {20, 1}

Long Comment Sequence

{1, 0}, {2, 846}, {3, 95}, {4, 18}, {5, 3}, {6, 2}

AND

{1, 0}, {2, 308}, {3, 21}, {4, 5}, {5, 0}, {6, 1}

OR

{1, 0}, {2, 7625}, {3, 3302}, {4, 1553}, {5, 747}, {6, 369}, {7, 171}, {8, 83}, {9, 31}, {10, 14}, {11, 10}, {12, 11}, {13, 3}, {14, 3}, {15, 2}, {16, 1}, {17, 1}, {18, 0}, {19, 0}, {20, 1}

A.1.2 Neutral Sentiment

Short Comment Sequence

{1, 0}, {2, 8095}, {3, 3944}, {4, 2101}, {5, 1082}, {6, 576}, {7, 325}, {8, 172}, {9, 88}, {10, 52}, {11, 36}, {12, 20}, {13, 5}, {14,7}, {15, 7}, {16, 4}, {17, 2}, {18, 3}, {19, 0}, {20, 1}, {21, 0}, {22, 0}, {23, 0}, {24, 0}, {25, 0}, {26, 0}, {27, 0}, {28, 0}, {29, 0}, {30, 0}, {31, 0}, {32, 1}, {33,0}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 1}, {42, 0}, {43, 0}, {44,0}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 0}, {52, 0}, {53, 0}, {54, 0}, {55,0}, {56, 0}, {57, 0}, {58, 0}, {59, 0}, {60, 0}, {61, 0}, {62, 0}, {63, 0}, {64, 0}, {65, 0}, {66,0}, {67, 0}, {68, 0}, {69, 0}, {70, 0}, {71, 0}, {72, 0}, {73, 0}, {74, 0}, {75, 0}, {76, 0}, {77,0}, {78, 0}, {79, 0}, {80, 0}, {81, 0}, {82, 0}, {83, 0}, {84, 0}, {85, 0}, {86, 0}, {87, 0}, {88,0}, {89, 0}, {90, 0}, {91, 0}, {92, 0}, {93, 0}, {94, 0}, {95, 0}, {96, 0}, {97, 0}, {98, 0}, {99,0}, {100, 0}, {101, 0}, {102, 0}, {103, 0}, {104, 0}, {105,0}, {106, 0}, {107, 0}, {108, 1}

Long Comment Sequence

{1, 0}, {2, 3219}, {3, 2174}, {4, 1616}, {5, 1202}, {6, 918}, {7, 725}, {8, 631}, {9, 501}, {10,412}, {11, 352}, {12, 277}, {13, 267}, {14, 178}, {15, 175}, {16, 157}, {17, 118}, {18, 93}, {19, 80}, {20, 84}, {21, 59}, {22, 43}, {23, 47}, {24, 40}, {25, 38}, {26, 21}, {27, 26}, {28, 25}, {29, 16}, {30, 18}, {31, 18}, {32, 14}, {33, 12}, {34, 12}, {35, 7}, {36, 14}, {37, 6}, {38, 5}, {39,8}, {40, 9}, {41, 6}, {42, 3}, {43, 2}, {44, 3}, {45, 2}, {46, 4}, {47, 0}, {48, 3}, {49, 1}, {50,1}, {51, 2}, {52, 1}, {53, 1}, {54, 1}, {55, 0}, {56, 0}, {57, 0}, {58, 1}, {59, 1}, {60, 1}, {61,0}, {62, 1}, {63, 0}, {64, 0}, {65, 2}, {66, 2}, {67, 0}, {68, 0}, {69, 0}, {70, 0}, {71, 0}, {72,1}

AND

{1, 0}, {2, 6443}, {3, 2570}, {4, 1158}, {5, 525}, {6, 230}, {7, 119}, {8, 44}, {9, 25}, {10, 18}, {11, 12}, {12, 7}, {13, 1}, {14, 3}, {15, 1}, {16, 2}, {17, 0}, {18, 2}, {19, 0}, {20, 0}, {21, 0}, {22, 0}, {23, 0}, {24, 0}, {25, 1}, {26, 0}, {27, 0}, {28, 0}, {29, 0}, {30, 0}, {31, 0}, {32, 0}, {33,0}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 1}

OR

{1, 0}, {2, 2314}, {3, 1716}, {4, 1341}, {5, 1054}, {6, 792}, {7, 638},
{8, 579}, {9, 447}, {10,409}, {11, 363}, {12, 267}, {13, 292}, {14, 229}, {15, 193},

{16, 223}, {17, 160}, {18, 163}, {19, 133}, {20, 128}, {21, 114}, {22, 87}, {23, 97}, {24, 91}, {25, 64}, {26, 63},

{27, 56}, {28,62}, {29, 38},

{30, 30}, {31, 39}, {32, 33}, {33, 29}, {34, 38}, {35, 19},{36, 26}, {37, 27}, {38, 16}, {39, 15},

{40, 24}, {41, 13}, {42, 13}, {43, 13}, {44, 15}, {45, 9}, {46, 13}, {47, 12}, {48, 7}, {49, 12}, {50, 9}, {51, 8}, {52, 4}, {53, 7}, {54, 7}, {55, 4}, {56, 1}, {57, 4}, {58, 4}, {59, 8}, {60, 6}, {61, 2}, {62, 3}, {63, 2}, {64, 0}, {65, 2}, {66, 3}, {67, 0}, {68, 1}, {69, 3}, {70,0}, {71, 1}, {72, 0}, {73, 1}, {74, 1}, {75, 1}, {76, 0}, {77, 0}, {78, 0}, {79, 0}, {80, 2}, {81,0}, {82, 1}, {83, 0}, {84, 1},{85, 2}, {86, 1}, {87, 0}, {88, 0}, {89, 0}, {90, 0}, {91, 1}, {92,1}, {93, 0}, {94, 0}, {95, 0}, {96, 0},{97, 0}, {98, 0}, {99, 0}, {100, 0}, {101, 0}, {102, 0}, {103, 0}, {104, 0}, {105, 0}, {106, 0}, {107, 0}, {108, 1}, {109, 0}, {110, 0}, {111, 1}, {112, 0}, {113, 0}, {114, 0}, {115, 0}, {116, 0}, {117, 1}, {118, 0}, {119, 0}, {120, 0}, {121, 0}, {122, 0}, {123, 0}, {124, 0}, {125, 0}, {126, 0}, {127, 0}, {128, 0}, {129,0}, {130, 0}, {131, 0}, {132, 0}, {133, 0}, {134, 0}, {135, 0}, {136, 0},{137, 0}, {138, 0}, {139, 0}, {140, 0}, {141, 0}, {142, 0}, {143, 0}, {144, 0}, {145, 0},{146, 0}, {147, 0}, {148, 1}, {149, 1}

A.2 React-Native

Positive Sentiment - Short Comment Sequence

{1, 0}, {2, 408}, {3, 90}, {4, 20}, {5, 8}, {6, 0}, {7, 2}, {8, 0}, {9, 1}, {10, 0}, {11, 0}, {12, 0}, {13, 0}, {14, 0}, {15, 0}, {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 0}, {22, 1}

Long Comment Sequence

{1, 0}, {2, 181}, {3, 91}, {4, 68}, {5, 35}, {6, 20}, {7, 20}, {8, 12}, {9, 9}, {10, 4}, {11, 8}, {12,4}, {13, 2}, {14, 5}, {15, 1}, {16, 1}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 0}, {22, 1}, {23,0}, {24, 0}, {25, 0}, {26, 0}, {27, 0}, {28, 0}, {29, 0}, {30, 0}, {31, 0}, {32, 0}, {33, 0}, {34,0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 0}, {42, 1}

AND

{1, 0}, {2, 37}, {3, 7}

OR

{1, 0}, {2, 557}, {3, 158}, {4, 97}, {5,44}, {6, 20}, {7, 23}, {8, 13}, {9,11}, {10, 3}, {11, 6}, {12, 5}, {13, 3}, {14,0}, {15, 2}, {16, 1}, {17, 0}, {18, 3}, {19,1}, {20, 3}, {21, 1}, {22, 2}, {23,0}, {24, 0}, {25, 0}, {26, 1}, {27, 1}, {28,0}, {29, 0}, {30, 0}, {31, 0}, {32, 1}, {33, 0}, {34,0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39,0}, {40, 0}, {41, 0}, {42,0}, {43, 0}, {44, 0}, {45,0}, {46, 0}, {47,1}

A.2.1 Negative Sentiment

Short Comment Sequence

{1, 0}, {2, 1415}, {3, 637}, {4, 302}, {5, 126}, {6, 55}, {7, 33}, {8, 20}, {9, 13}, {10, 7}, {11, 5}, {12, 0}, {13, 2}, {14, 0}, {15, 2}, {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 1}

Long Comment Sequence

{1, 0}, {2, 432}, {3, 115}, {4, 29}, {5, 12}, {6, 3}, {7, 2}, {8, 1}

AND

{1, 0}, {2, 211}, {3, 38}, {4, 11}, {5, 1}, {6, 2}

OR

{1, 0}, {2, 1463}, {3, 699}, {4, 380}, {5, 194}, {6, 102}, {7, 53}, {8, 39}, {9, 26}, {10, 12}, {11,11}, {12, 0}, {13, 3}, {14, 1}, {15, 3}, {16, 0}, {17, 0}, {18, 2}, {19, 1}, {20, 0}, {21, 1}

A.2.2 Neutral Sentiment

Short Comment Sequence {1, 0}, {2, 1454}, {3, 660}, {4, 328}, {5, 154}, {6, 83}, {7, 44}, {8, 22}, {9, 3},{10, 5},{11, 3}, {12, 2},{13, 1}, {14, 1}, {15, 2}, {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 0}, {21, 0}, {22, 0}, {23,0}, {24, 0}, {25, 1}

Long Comment Sequence

{1, 0}, {2, 497}, {3, 328}, {4, 269}, {5, 215}, {6, 174}, {7, 135}, {8, 102}, {9, 104}, {10, 91}, {11, 59}, {12, 56}, {13, 38}, {14, 49}, {15, 35}, {16, 39}, {17, 38}, {18, 27}, {19, 26}, {20, 19}, {21, 11}, {22, 15}, {23, 11}, {24, 9}, {25, 11}, {26, 8}, {27, 4}, {28, 9}, {29, 6}, {30, 8}, {31, 5}, {32, 5}, {33, 3}, {34, 5}, {35, 3}, {36, 3}, {37, 3}, {38, 2}, {39, 3}, {40, 3}, {41, 4}, {42, 1}, {43,1}, {44, 2}, {45, 3}, {46, 0}, {47, 2}, {48, 2}, {49, 0}, {50, 1}, {51, 0}, {52, 0}, {53, 1}, {54,0}, {55, 2}, {56, 0}, {57, 0}, {58, 0}, {59, 0}, {60, 1}, {61, 0}, {62, 0}, {63, 0}, {64, 1}, {65,0}, {66, 0}, {67, 0}, {68, 0}, {69, 0}, {70, 1}, {71, 1}, {72, 0}, {73, 0}, {74, 0}, {75, 0}, {76,0}, {77, 0}, {78, 0}, {79, 0}, {80, 0}, {81, 0}, {82, 0}, {83, 0}, {84, 0}, {85, 0}, {86, 0}, {87,0}, {88, 0}, {89, 0}, {90, 0}, {91, 0}, {92, 0}, {93, 0}, {94, 0}, {95, 0}, {96, 0}, {97, 0}, {98,0}, {99, 1}, {100, 0}, {101, 1}

AND

{1, 0}, {2, 37}, {3, 7}

OR

{1, 0}, {2, 384}, {3, 260}, {4, 214}, {5, 160}, {6, 131}, {7, 123}, {8, 106}, {9, 77}, {10, 80}, {11,57}, {12, 60}, {13, 32}, {14, 49}, {15, 36}, {16, 32}, {17, 38}, {18, 26}, {19, 31}, {20, 26}, {21, 17}, {22, 18}, {23, 14}, {24, 16}, {25, 17}, {26, 14}, {27, 9}, {28, 15}, {29, 9}, {30, 9}, {31,8}, {32, 6}, {33, 6}, {34, 12}, {35, 8}, {36, 3}, {37, 4}, {38, 7}, {39, 7}, {40, 4}, {41, 3}, {42,4}, {43, 2}, {44, 3}, {45, 3}, {46, 0}, {47, 3}, {48, 3}, {49, 0}, {50, 2}, {51, 2}, {52, 2}, {53,2}, {54, 2}, {55, 0}, {56, 0}, {57, 1}, {58, 0}, {59, 0}, {60, 0}, {61, 3}, {62, 1}, {63, 1}, {64,1}, {65, 1}, {66, 0}, {67, 0}, {68, 0}, {69, 0}, {70, 1}, {71, 2}, {72, 0}, {73, 0}, {74, 1}, {75,0}, {76, 1}, {77, 1}, {78, 1}, {79, 1}, {80, 0}, {81, 0}, {82, 0}, {83, 1}, {84, 1}, {85, 0}, {86,2}, {87, 0}, {88, 0}, {89, 0}, {90, 0},{91, 0}, {92, 0}, {93, 0}, {94, 0}, {95, 0}, {96, 0},{97,0}, {98, 0}, {99, 1}, {100, 0}, {101, 0}, {102, 0}, {103, 0}, {104, 0}, {105, 0},{106, 0}, {107,0}, {108, 0}, {109, 0}, {110, 0}, {111, 0}, {112, 0}, {113, 0}, {114, 1}, {115, 0}, {116, 0}, {117, 0}, {118, 0}, {119, 1}, {120, 0}, {121, 0}, {122, 0}, {123, 0}, {124, 0}, {125, 0}, {126, 0},{127, 0}, {128, 0}, {129, 0}, {130, 0}, {131, 0}, {132, 1}

A.3 Keras

Short Comment Sequence {1, 0}, {2, 78}, {3, 10}, {4, 1}, {5, 2}

Long Comment Sequence

{1, 0},{2, 197},{3, 152},{4, 73}, {5, 61},{6, 55},{7, 34},{8, 28}, {9, 25},{10, 17},{11, 15},{12, 15},{13, 9},{14, 11},{15, 6},{16, 5},{17, 3},{18, 5}, {19, 2},{20, 2},{21, 2},{22, 6},{23, 0},{24, 1},{25, 2},{26, 3},{27, 2}, {28, 2}, {29, 0}, {30,1}, {31, 1}, {32, 0}, {33, 1}, {34,0}, {35, 0}, {36, 0}, {37, 0}, {38, 1}, {39, 0}, {40, 0}, {41, 1}, {42, 1}, {43, 1}, {44, 0}, {45,0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 1}, {52, 0},{53, 0}, {54, 0}, {55, 0}, {56,0}, {57, 0}, {58, 0}, {59, 0}, {60,0},{61, 0}, {62, 0}, {63, 0},

64, 0}, {65, 1}

AND

{1, 0}, {2, 58}, {3, 9}, {4, 1}

OR

{1, 0}, {2, 193}, {3, 138}, {4, 85}, {5, 65}, {6, 54}, {7,38}, {8,30}, {9, 26}, {10, 13}, {11, 17}, {12, 13}, {13, 10}, {14, 11}, {15,5}, {16, 5}, {17, 4}, {18, 4}, {19, 3}, {20, 3}, {21, 1}, {22, 7},{23, 1}, {24, 1}, {25, 5}, {26, 4}, {27, 2}, {28, 2}, {29, 0}, {30,1}, {31, 0}, {32, 0}, {33, 1}, {34, 0}, {35, 0}, {36, 1}, {37, 0}, {38, 1}, {39, 0}, {40, 0}, {41, 1}, {42, 1}, {43, 1}, {44, 0}, {45,0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 1}, {52, 0}, {53, 0}, {54, 0}, {55, 0}, {56, 0}, {57, 0}, {58, 0}, {59, 0}, {60, 0}, {61, 0}, {62, 0}, {63, 0}, {64, 0}, {65, 1}

A.3.1 Negative Sentiment

Short Comment Sequence {1, 0}, {2, 419}, {3, 198}, {4, 80}, {5, 39}, {6, 20}, {7, 7}, {8, 6}, {9, 3}, {10, 1}, {11, 1}, {12,0}, {13, 1}, {14, 1}, {15, 0}, {16,1}

Long Comment Sequence

{1, 0}, {2, 63}, {3,11}

AND

{1, 0}, {2, 16}, {3, 4}

OR

{1, 0}, {2, 423}, {3, 232}, {4, 80}, {5, 57}, {6, 25}, {7, 9}, {8, 10}, {9, 4}, {10, 1}, {11, 1}, {12,0}, {13, 1}, {14, 1}, {15, 0}, {16, 1}

A.3.2 Neutral Sentiment

Short Comment Sequence {1, 0}, {2, 425}, {3, 181}, {4, 110}, {5, 49}, {6, 29}, {7, 15}, {8, 9}, {9, 4}, {10, 1}, {11, 2}, {12,0}, {13, 2}, {14, 0}, {15, 0}, {16, 1}, {17, 0}, {18, 1}

Long Comment Sequence

{1, 0}, {2, 160}, {3, 53}, {4, 50}, {5, 24}, {6, 26}, {7, 8}, {8, 5},{9, 8}, {10, 6}, {11, 5}, {12, 4}, {13, 2}, {14, 1}, {15, 2}, {16, 4}, {17, 1}, {18, 0}, {19, 0}, {20, 0}, {21, 1}, {22, 0}, {23, 0}, {24,1}, {25, 0}, {26, 0}, {27, 0}, {28, 0}, {29, 0}, {30, 1}, {31,0}, {32, 0}, {33, 1}, {34, 0}, {35,0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 0}, {42, 0}, {43, 0}, {44, 0}, {45, 0}, {46,0}, {47, 0}, {48, 0}, {49, 0}, {50, 0}, {51, 1}

AND

{1, 0}, {2, 136}, {3, 35}, {4, 18}, {5, 7}, {6, 8}, {7, 0}, {8, 0}, {9, 0}, {10, 0}, {11, 0}, {12, 0}, {13, 0}, {14, 0}, {15, 0}, {16, 1}

OR

{1, 0}, {2, 357}, {3, 194}, {4, 113}, {5, 61}, {6, 53}, {7, 25}, {8, 16}, {9, 15}, {10, 10}, {11, 11}, {12, 8}, {13, 8}, {14, 5}, {15, 3}, {16, 5}, {17, 0}, {18, 2}, {19, 1}, {20, 1}, {21, 1}, {22, 0}, {23,0}, {24, 1}, {25, 1}, {26, 1}, {27, 1}, {28, 1}, {29, 0}, {30, 0}, {31, 0}, {32, 1}, {33, 0}, {34,0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 0}, {42, 0}, {43, 0}, {44, 0}, {45,1}, {46, 1}, {47, 0}, {48, 0}, {49, 0}, {50,0}, {51, 1}, {52, 0}, {53, 0}, {54, 0}, {55, 0}, {56,1}

A.4 Pytorch

Positive - Short Comment Sequence {1, 0}, {2, 1097}, {3, 225}, {4, 73}, {5, 20}, {6, 14}, {7, 7}, {8, 1}, {9, 0}, {10, 1}

Long Comment Sequence

{1, 0}, {2, 566}, {3, 273}, {4, 160}, {5, 108}, {6, 87}, {7, 53}, {8, 32}, {9, 17}, {10, 16}, {11,11}, {12, 9}, {13, 13}, {14, 7}, {15, 7}, {16, 3}, {17, 4}, {18, 6}, {19, 7}, {20, 3}, {21, 0}, {22, 0}, {23, 2}, {24, 3}, {25, 1}, {26, 0}, {27, 1}, {28, 2}, {29, 2}, {30, 0}, {31, 1}, {32, 1}, {33, 2}, {34,0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 2}, {40, 0}, {41, 0}, {42, 0}, {43, 1}, {44, 0},{45,1}, {46, 0}, {47, 0}, {48, 0}, {49, 1}

AND

{1, 0}, {2, 145}, {3, 31}, {4, 23}, {5, 4}, {6, 5}

OR

{1, 0}, {2, 1458}, {3, 498}, {4, 237}, {5, 135}, {6, 101}, {7, 84}, {8, 44}, {9, 33}, {10, 20}, {11,10}, {12, 10}, {13, 12}, {14, 11}, {15, 12}, {16, 3}, {17, 5}, {18, 5}, {19, 7}, {20, 3}, {21, 1}, {22, 0}, {23, 3}, {24, 2}, {25, 1}, {26, 0}, {27, 0}, {28, 2}, {29, 2}, {30, 0}, {31, 2}, {32, 2}, {33,1}, {34, 0}, {35, 1}, {36, 1}, {37, 0}, {38, 0}, {39, 1}, {40, 0}, {41, 0}, {42, 1}, {43, 0}, {44, 0}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 1}, {50, 0}, {51, 0}, {52, 1}, {53, 0}, {54, 1}, {55, 0}, {56, 0}, {57, 1}

A.4.1 Negative Sentiment

Short Comment Sequence {1, 0}, {2, 3050}, {3, 1252}, {4, 557}, {5, 258}, {6, 128}, {7, 69}, {8, 30}, {9, 12}, {10, 6}, {11,5}, {12, 4}, {13, 0}, {14, 3}

Long Comment Sequence

{1, 0}, {2, 167}, {3, 31}, {4, 3}, {5, 2}, {6, 1}

AND

{1, 0}, {2, 76}, {3, 10}, {4, 1}, {5,1}

OR

{1, 0}, {2, 3082}, {3, 1300}, {4, 621}, {5, 313}, {6, 146}, {7, 81}, {8, 36}, {9, 19}, {10, 8}, {11,7}, {12, 6}, {13,0}, {14, 3}

A.4.2 Neutral Sentiment

Short Comment Sequence {1, 0}, {2, 3026}, {3, 1259}, {4, 577}, {5, 249}, {6, 121}, {7, 42}, {8, 35}, {9, 16}, {10, 7}, {11,3}, {12, 2}

Long Comment Sequence

{1, 0}, {2, 532}, {3, 296}, {4, 177}, {5, 100}, {6, 98}, {7, 77}, {8, 73}, {9, 67}, {10, 49}, {11,47}, {12, 30}, {13, 34}, {14, 35}, {15, 32}, {16, 32}, {17, 29}, {18, 28}, {19, 35}, {20, 20}, {21, 23}, {22, 15}, {23, 11}, {24, 28}, {25, 20}, {26, 16}, {27, 10}, {28, 14}, {29, 19}, {30, 8}, {31, 18}, {32, 13}, {33, 10}, {34, 13}, {35, 17}, {36, 10}, {37, 11}, {38, 14}, {39, 11}, {40, 14}, {41, 11}, {42, 7}, {43, 4}, {44, 10}, {45, 6}, {46, 10}, {47, 10}, {48, 8}, {49, 6}, {50, 4}, {51, 7}, {52, 3}, {53, 7}, {54, 4}, {55, 7}, {56, 5}, {57, 6}, {58, 2}, {59, 7}, {60, 2}, {61, 8}, {62, 5}, {63, 8}, {64,1}, {65, 1}, {66, 4}, {67, 7}, {68, 3}, {69, 9}, {70, 2}, {71, 2}, {72, 5}, {73, 4}, {74, 2}, {75,3}, {76, 0}, {77, 7}, {78, 5}, {79, 2}, {80, 2}, {81, 3}, {82, 1}, {83, 3}, {84, 2}, {85, 4}, {86,1}, {87, 0}, {88, 1}, {89, 4}, {90, 1}, {91, 2}, {92, 0}, {93, 3}, {94, 2}, {95, 0}, {96, 0}, {97,2} 98, 2}, {99, 0}, {100, 2}, {101, 1}, {102, 1}, {103, 1}, {104, 3}, {105, 0}, {106, 2}, {107,2}, {108, 0}, {109, 1}, {110, 1}, {111, 0}, {112, 1}, {113, 3}, {114, 1}, {115, 1}, {116, 0}, {117, 0}, {118, 0}, {119, 0}, {120, 1}, {121, 0}, {122, 0}, {123, 0}, {124, 0}, {125, 2}, {126, 0}, {127, 0}, {128, 1}, {129, 0}, {130, 1}, {131, 0}, {132, 2},{133, 0}, {134, 2}, {135, 0}, {136, 0}, {137, 0},{138, 1}, {139, 1}, {140, 0}, {141, 0}, {142, 1}, {143, 2}, {144, 1}, {145, 0}, {146, 1}, {147, 1}, {148, 0}, {149, 0}, {150, 1}, {151, 2}, {152, 0}, {153, 0}, {154, 1}, {155, 1}, {156, 0}, {157, 0}, {158, 0},{159, 1}, {160, 0}, {161, 0}, {162, 1}, {163, 0}, {164, 0}, {165, 1}, {166, 0}, {167, 1}, {168, 0}, {169, 1}, {170, 0}, {171, 0}, {172, 0},{173, 0}, {174, 0}, {175, 0}, {176, 0}, {177, 0}, {178, 0}, {179, 0},{180, 0}, {181, 0}, {182, 0}, {183, 0}, {184, 0}, {185, 0}, {186, 0}, {187, 0}, {188, 0}, {189, 0}, {190, 0}, {191, 0}, {192, 0}, {193, 0}, {194, 0}, {195, 0}, {196, 0}, {197, 0}, {198, 0},{199, 0}, {200, 0}, {201, 0}, {202, 0}, {203, 0}, {204, 0}, {205, 0}, {206, 0},{207, 0}, {208, 0}, {209, 0}, {210, 0}, {211, 0}, {212, 0}, {213, 0}, {214, 0}, {215, 0}, {216, 0}, {217, 0}, {218, 0}, {219, 0}, {220, 0}, {221, 0}, {222, 0}, {223, 0}, {224, 0}, {225, 0}, {226, 0},{227, 0}, {228, 0},{229, 0}, {230, 0}, {231, 0}, {232, 0}, {233, 0}, {234, 0}, {235, 0}, {236, 0}, {237, 0}, {238, 0}, {239, 0}, {240, 0}, {241, 0}, {242, 0},{243, 0}, {244, 0}, {245, 0}, {246, 0}, {247, 0}, {248, 0}, {249, 0}, {250, 0}, {251, 0}, {252, 0}, {253, 0}, {254, 0}, {255, 0}, {256, 0}, {257, 0}, {258, 1}, {259, 1}, {260, 0}, {261, 0}, {262, 0}, {263, 0},{264, 0}, {265, 0}, {266, 0}, {267, 0}, {268, 0}, {269, 0}, {270, 0}, {271, 0}, {272, 0}, {273, 0}, {274, 0}, {275, 0}, {276, 0}, {277, 0},{278, 0}, {279, 0}, {280, 1}, {281, 0}, {282, 0}, {283, 0}, {284, 0},{285, 0}, {286, 0}, {287, 0}, {288, 0}, {289, 0}, {290, 0}, {291, 0}, {292, 0}, {293, 1}, {294, 0}, {295, 0}, {296, 0}, {297, 0}, {298, 0},{299, 0}, {300, 0}, {301, 0}, {302, 0}, {303, 0}, {304, 0}, {305, 0},{306, 1}, {307, 0}, {308, 0}, {309, 0}, {310, 0}, {311, 0}, {312, 0}, {313, 0}, {314, 0}, {315, 0}, {316, 0}, {317, 0}, {318, 0}, {319, 0},{320, 0}, {321, 0}, {322, 0}, {323, 0}, {324, 0}, {325, 0}, {326, 0}, {327, 0}, {328, 0}, {329, 0}, {330, 0}, {331, 0}, {332, 0}, {333, 0},{334, 0}, {335, 0}, {336, 0}, {337, 0}, {338, 0}, {339, 0}, {340, 0},{341, 0}, {342, 0}, {343, 0}, {344, 0}, {345, 0}, {346, 0}, {347, 0},{348, 0}, {349, 0}, {350, 0}, {351, 0}, {352, 0}, {353, 0}, {354, 0}, {355, 0}, {356, 0}, {357, 0}, {358, 0}, {359, 0}, {360, 0}, {361, 0},{362, 0}, {363, 0}, {364, 0}, {365, 0}, {366, 0}, {367, 0}, {368, 0},{369, 0}, {370, 0}, {371, 0}, {372, 0}, {373, 0}, {374, 0}, {375, 0}, {376, 0}, {377, 0}, {378, 0}, {379, 0}, {380, 0}, {381, 0}, {382, 0},{383, 0}, {384, 0}, {385, 0}, {386, 0}, {387, 0}, {388, 0}, {389, 0},{390, 0}, {391, 0}, {392, 0}, {393, 0}, {394, 0}, {395, 0}, {396, 0}, {397, 0}, {398, 0}, {399, 0}, {400, 0}, {401, 1}

AND

{1, 0}, {2, 2529}, {3, 970}, {4, 404}, {5, 186}, {6, 79}, {7, 34}, {8, 15}, {9, 9}, {10, 3}, {11, 3}, {12, 1}

OR

{1, 0}, {2, 664}, {3, 375}, {4, 268}, {5, 165}, {6, 120}, {7, 98}, {8, 79}, {9, 61}, {10, 41}, {11,32}, {12, 26}, {13, 23}, {14, 23}, {15, 29}, {16, 27}, {17, 18}, {18, 23}, {19, 15}, {20, 19}, {21, 16},{22, 15}, {23, 14}, {24, 13}, {25, 15}, {26, 14}, {27, 7}, {28, 14}, {29, 19}, {30, 8}, {31, 18}, {32, 13}, {33, 10}, {34, 13}, {35, 17}, {36, 10}, {37, 11}, {38, 14}, {39, 11}, {40, 14}, {41, 11}, {42, 7}, {43, 4}, {44, 10}, {45, 6}, {46, 10}, {47, 10}, {48, 8}, {49, 6}, {50, 4}, {51, 7}, {52, 3}, {53, 7}, {54, 4}, {55, 7}, {56, 5}, {57, 6}, {58, 2}, {59, 7}, {60, 2}, {61, 8}, {62, 5}, {63, 8}, {64,1}, {65, 1}, {66, 4}, {67, 7}, {68, 3}, {69, 9}, {70, 2}, {71, 2}, {72, 5}, {73, 4}, {74, 2}, {75,3}, {76, 0}, {77, 7}, {78, 5}, {79, 2}, {80, 2}, {81, 3}, {82, 1}, {83, 3}, {84, 2}, {85, 4}, {86,1}, {87, 0}, {88, 1}, {89, 4}, {90, 1}, {91, 2}, {92, 0}, {93, 3}, {94, 2}, {95, 0}, {96, 0}, {97,2}, {98, 2}, {99, 0}, {100, 2}, {101, 1}, {102, 1}, {103, 1}, {104, 3}, {105, 0}, {106, 2}, {107,2}, {108, 0}, {109, 1}, {110, 1}, {111, 0}, {112, 1}, {113, 3}, {114, 1}, {115, 1}, {116, 0}, {117, 0}, {118, 0}, {119, 0}, {120, 1}, {121, 0}, {122, 0}, {123, 0}, {124, 0}, {125, 2}, {126, 0}, {127, 0}, {128, 1}, {129, 0}, {130, 1}, {131, 0}, {132, 2}, {133, 0}, {134, 2}, {135, 0}, {136, 0}, {137, 0},{138, 1}, {139, 1}, {140, 0}, {141, 0}, {142, 1}, {143, 2}, {144, 1}, {145, 0}, {146, 1}, {147, 1}, {148, 0}, {149, 0}, {150, 1}, {151, 2}, {152, 0}, {153, 0}, {154, 1}, {155, 1}, {156, 0}, {157, 0}, {158, 0},{159, 1}, {160, 0}, {161, 0}, {162, 1}, {163, 0}, {164, 0}, {165, 1}, {166, 0}, {167, 1}, {168, 0}, {169, 1}, {170, 0}, {171, 0}, {172, 0},{173, 0}, {174, 0}, {175, 0}, {176, 0}, {177, 0}, {178, 0}, {179, 0},{180, 0}, {181, 0}, {182, 0}, {183, 0}, {184, 0}, {185, 0}, {186, 0}, {187, 0}, {188, 0}, {189, 0}, {190, 0}, {191, 0}, {192, 0}, {193, 0}, {194, 0}, {195, 0}, {196, 0}, {197, 0}, {198, 0},{199, 0}, {200, 0}, {201, 0}, {202, 0}, {203, 0}, {204, 0}, {205, 0}, {206, 0},{207, 0}, {208, 0}, {209, 0}, {210, 0}, {211, 0}, {212, 0}, {213, 0}, {214, 0}, {215, 0}, {216, 0}, {217, 0}, {218, 0}, {219, 0}, {220, 0}, {221, 0}, {222, 0}, {223, 0}, {224, 0}, {225, 0}, {226, 0},{227, 0}, {228, 0},{229, 0}, {230, 0}, {231, 0}, {232, 0}, {233, 0}, {234, 0}, {235, 0}, {236, 0}, {237, 0}, {238, 0}, {239, 0}, {240, 0}, {241, 0}, {242, 0},{243, 0}, {244, 0}, {245, 0}, {246, 0}, {247, 0}, {248, 0}, {249, 0}, {250, 0}, {251, 0}, {252, 0}, {253, 0}, {254, 0}, {255, 0}, {256, 0}, {257, 0}, {258, 1}, {259, 1}, {260, 0}, {261, 0}, {262, 0}, {263, 0},{264, 0}, {265, 0}, {266, 0}, {267, 0}, {268, 0}, {269, 0}, {270, 0}, {271, 0}, {272, 0}, {273, 0}, {274, 0}, {275, 0}, {276, 0}, {277, 0},{278, 0}, {279, 0}, {280, 1}, {281, 0}, {282, 0}, {283, 0}, {284, 0},{285, 0}, {286, 0}, {287, 0}, {288, 0}, {289, 0}, {290, 0}, {291, 0}, {292, 0}, {293, 1}, {294, 0}, {295, 0}, {296, 0}, {297, 0}, {298, 0},{299, 0}, {300, 0}, {301, 0}, {302, 0}, {303, 0}, {304, 0}, {305, 0},{306, 1}, {307, 0}, {308, 0}, {309, 0}, {310, 0}, {311, 0}, {312, 0}, {313, 0}, {314, 0}, {315, 0}, {316, 0}, {317, 0}, {318, 0}, {319, 0},{320, 0}, {321, 0}, {322, 0}, {323, 0}, {324, 0}, {325, 0}, {326, 0}, {327, 0}, {328, 0}, {329, 0}, {330, 0}, {331, 0}, {332, 0}, {333, 0},{334, 0}, {335, 0}, {336, 0}, {337, 0}, {338, 0}, {339, 0}, {340, 0},{341, 0}, {342, 0}, {343, 0}, {344, 0}, {345, 0}, {346, 0}, {347, 0},{348, 0}, {349, 0}, {350, 0}, {351, 0}, {352, 0}, {353, 0}, {354, 0}, {355, 0}, {356, 0}, {357, 0}, {358, 0}, {359, 0}, {360, 0}, {361, 0},{362, 0}, {363, 0}, {364, 0}, {365, 0}, {366, 0}, {367, 0}, {368, 0},{369, 0}, {370, 0}, {371, 0}, {372, 0}, {373, 0}, {374, 0}, {375, 0}, {376, 0}, {377, 0}, {378, 0}, {379, 0}, {380, 0}, {381, 0}, {382, 0},{383, 0}, {384, 0}, {385, 0}, {386, 0}, {387, 0}, {388, 0}, {389, 0},{390, 0}, {391, 0}, {392, 0}, {393, 0}, {394, 0}, {395, 0}, {396, 0}, {397, 0}, {398, 0}, {399, 0}, {400, 0}, {401, 1}

A.5 Pillow

Positive Sentiment - Short Comment Sequence {1, 0}, {2, 167}, {3, 27}, {4, 6}, {5, 3}, {6, 1}, {7, 0}, {8, 1}, {9,1}

Long Comment Sequence

{1, 0}, {2, 422}, {3, 290}, {4, 221}, {5, 174}, {6, 126}, {7, 79}, {8, 77}, {9, 48}, {10, 50}, {11,36}, {12, 25}, {13, 19}, {14, 19}, {15, 14}, {16, 10}, {17, 11}, {18, 12}, {19, 11}, {20, 6}, {21, 3}, {22, 2}, {23, 3}, {24, 2}, {25, 2}, {26, 5}, {27, 5}, {28, 3}, {29, 1}, {30, 2}, {31, 3}, {32,0}, {33, 1}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 1}, {42, 0}, {43,0}, {44, 1}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 1}, {51, 0}, {52, 0}, {53, 0}, {54,1}, {55, 1}

AND

{1, 0}, {2, 158}, {3, 19}, {4, 6}, {5, 3}, {6, 1}, {7, 0}, {8, 1}, {9,1}

OR

{1, 0}, {2, 416}, {3, 281}, {4, 224}, {5, 171}, {6, 119}, {7, 84}, {8, 76}, {9, 51}, {10, 50}, {11,38}, {12, 27}, {13, 18}, {14, 19}, {15, 16}, {16, 8}, {17, 11}, {18, 13}, {19, 11}, {20, 5}, {21,3}, {22, 2}, {23, 5}, {24, 2}, {25, 2}, {26, 6}, {27, 6}, {28, 3}, {29, 1}, {30, 2}, {31, 3}, {32,0}, {33, 2}, {34, 0}, {35, 0}, {36, 0}, {37, 0}, {38, 0}, {39, 0}, {40, 0}, {41, 1}, {42, 0}, {43,0}, {44, 1}, {45, 0}, {46, 0}, {47, 0}, {48, 0}, {49, 0}, {50, 1}, {51, 0}, {52, 0}, {53, 0}, {54,1}, {55, 1}

A.5.1 Negative Sentiment

Short Comment Sequence {1, 0}, {2, 744}, {3, 370},{4, 188}, {5, 93}, {6, 55}, {7, 43}, {8, 19}, {9, 10}, {10, 2}, {11, 3}, {12, 4}, {13, 2}, {14, 0}, {15, 0}, {16, 0}, {17, 1}

Long Comment Sequence

{1, 0}, {2, 180}, {3, 34}, {4, 9}, {5, 5}, {6, 0}, {7, 0}, {8, 0}, {9,1}

AND

{1, 0}, {2, 125}, {3, 20}, {4, 3}, {5, 0}, {6, 0}, {7, 0}, {8, 0}, {9,1}

OR

{1, 0}, {2, 423}, {3, 232}, {4, 80}, {5, 57}, {6, 25}, {7, 9}, {8, 10}, {9, 4}, {10, 1}, {11, 1}, {12,0}, {13, 1}, {14, 1}, {15, 0}, {16, 1}

A.5.2 Neutral Sentiment

Short Comment Sequence {1, 0}, {2, 807}, {3, 333}, {4, 155}, {5, 70}, {6, 38}, {7, 14}, {8, 11}, {9, 2}, {10, 3}, {11, 4}, {12, 1}, {13, 0}, {14, 0}, {15, 0}, {16, 0}, {17, 0}, {18, 0}, {19, 0}, {20, 1}

Long Comment Sequence

{1, 0}, {2, 239}, {3, 59}, {4, 9}, {5, 7}, {6, 2}

AND

{1, 0}, {2, 56}, {3, 5}

OR

{1, 0}, {2, 786}, {3, 407}, {4, 222}, {5, 131}, {6, 73}, {7, 34}, {8, 33}, {9, 20}, {10, 8}, {11, 8}, {12, 2}, {13, 6}, {14, 1}, {15, 1}, {16, 2}, {17, 1}, {18, 0}, {19, 0}, {20, 1}

Vita

Candidate's full name: Arleen Kaur Arora
University attended : University of Mumbai, Bachelor's in Computer Engineering, 2016-2021