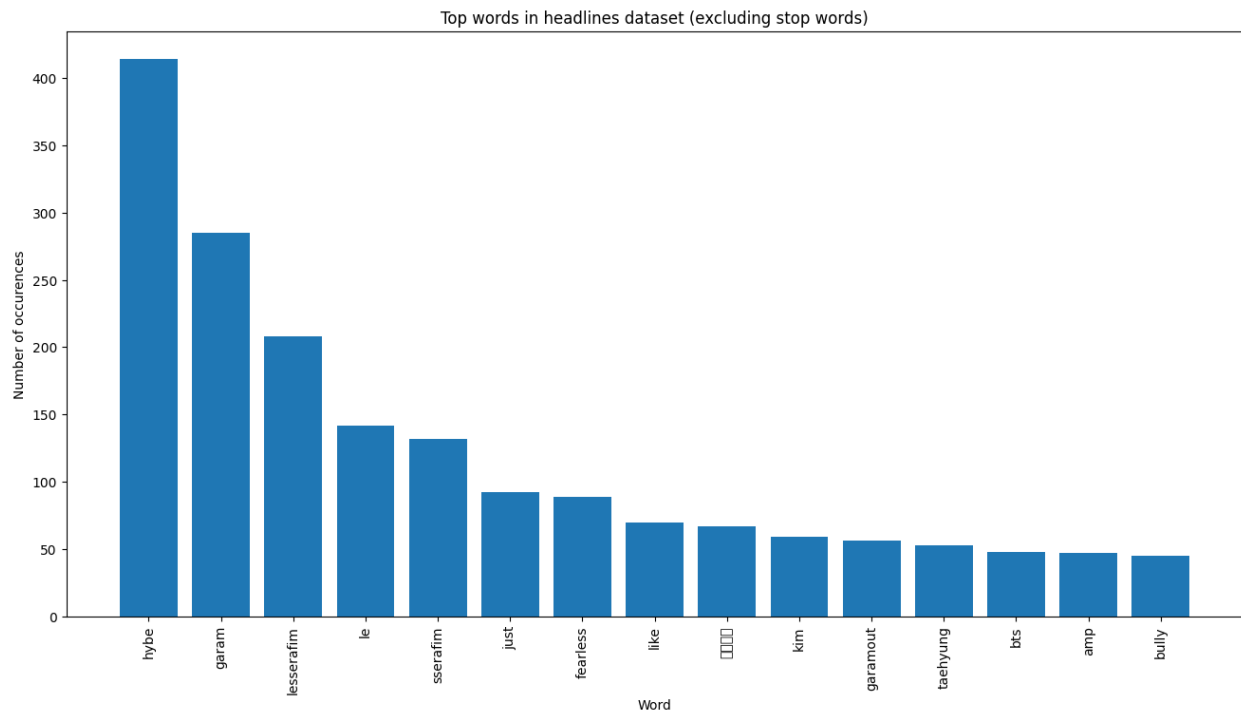


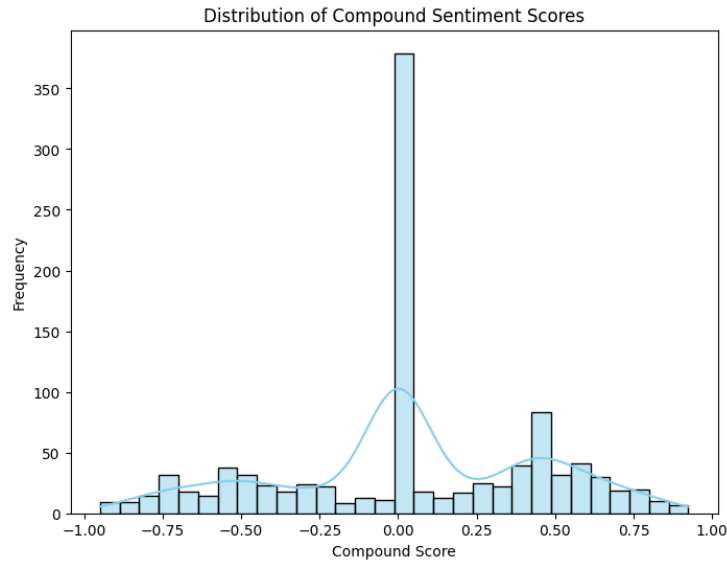
Exploratory Data Analysis

The dataset used in the study are tweets from May 26 - June 2, 2022 related to the Kpop girl group, Le Sserrafim, and their company, Hybe. The dataset did not contain any null values. Moreover, the histogram below shows the top fifteen frequently used words from the dataset excluding english stop words. The word 'hybe' was used more than 400 times, obtaining the highest number of occurrences. This is followed by 'garam' with more than 250 occurrences and 'le sserafim' with around 200 occurrences.

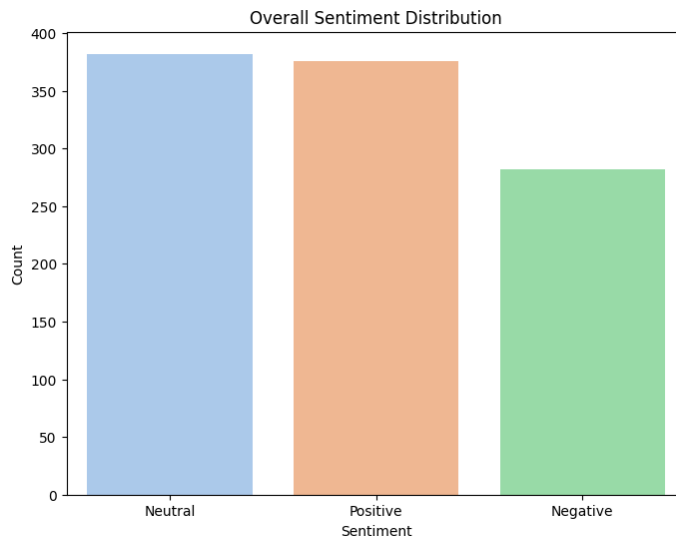


Sentiment Analysis

Sentiment analysis is an application of natural language processing that classifies texts into three groups namely positive, neutral, and negative (Mergim and Jaumin 21). Valence Aware Dictionary for Sentiment Reasoning (VADER) is the method used for sentiment analysis in the project. The VADER algorithm calculates the sentiment score for each text in the lemmatized column of the dataset. The figure below features the distribution of the compound sentiment scores which is the basis for the overall sentiment of each document. The distribution is skewed to the left signifying that the lump of the data has zero to positive compound scores. In addition, it shows that the majority of the documents in the dataset have a compound score of 0.0, followed by a score of 0.50.



The bar graph below shows the count of the overall sentiment distribution for each document in the dataset. It shows that most documents are labeled as neutral sentiments, indicating a compound score of between -0.05 and 0.05. Positive sentiments come in second with documents having a compound score of greater than 0.05. Negative sentiments come in last with documents having a compound score of less than 0.05.



Topic Modeling

Topic modeling is a technique that uncovers the thematic structure of a collection of documents through grouping themes as topics (Kherwa and Bansal 1-2). Latent Dirichlet Allocation (LDA) algorithm is employed for topic modeling. The word clouds below feature texts belonging to a particular theme followed by the interpretation of the topics summarized in the table.

Topic 0

think show postage song
sserafim ve
music already set even girl today
lesserafim love
big group date get hybe
kpop know make leader

Topic 1

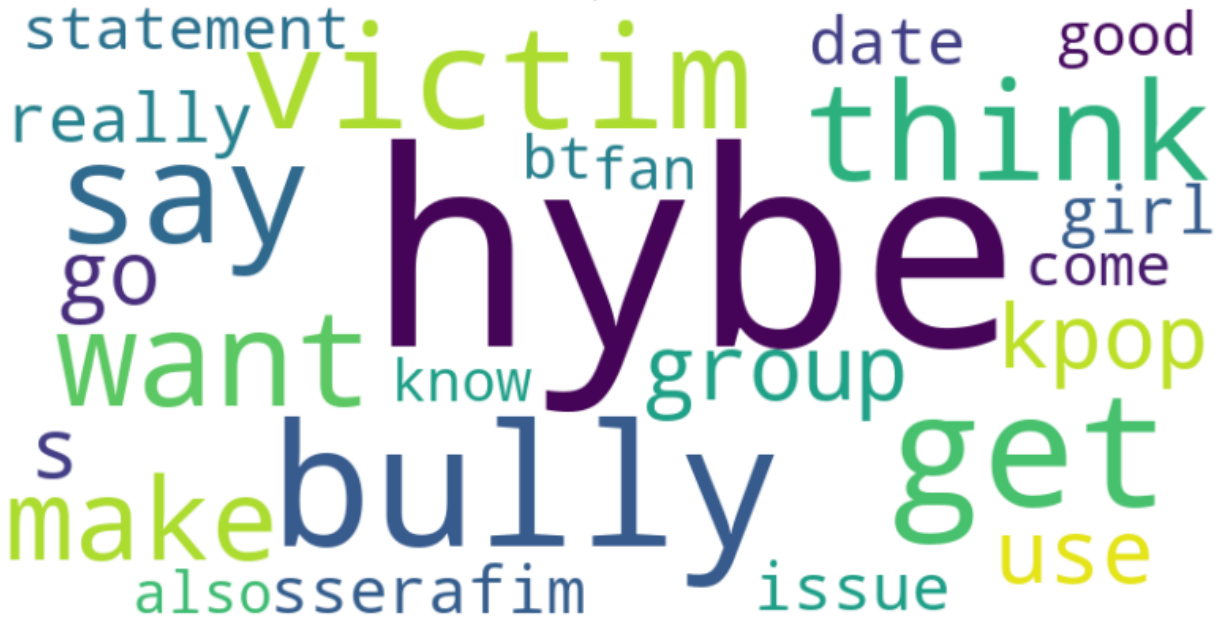
even lesserafim gar amout
date king go
know literally hybe day protect
let good s people
say chart artist person
new ve make lesserafim m

Topic 2



A word cloud for Topic 2 featuring various terms related to the K-pop group LESSERAFIM. The words are arranged in a circular pattern with different colors. The most prominent words are 'hybe', 'lesserafim', and 'fearless'. Other visible words include 'week', 'chart', 'people', 'music', 'garam', 'issue', 'peak', 'bully', 'cover', 'go', 'get', 'want', 'new', 'stream', 'imlesserafim', 'know', 've', 'take', 'let', and '□□□□'.

Topic 3



A word cloud for Topic 3 featuring various terms related to the K-pop group LESSERAFIM. The words are arranged in a circular pattern with different colors. The most prominent words are 'hybe', 'bully', 'victim', 'think', 'say', 'want', 'get', 'make', 'group', 'kpop', 'use', 'girl', 'come', 'date', 'good', 'statement', 'really', 'go', 'know', 'bt fan', 'also', 'lesserafim', 'issue', and 's'.

[illegible]

Topic	Texts	Interpretation
Topic 0	music, think, show, sserafim. Lesserafim, big, kpop, know, make, leader, girl, postage, song, hybe, leader, love, gidle, today, already, set	Popularity of Le Sserafim as a kpop girl group under Hybe
Topic 1	hybe. people, garam out, know, make, sserafim, good, chart, literally, protect, artist, person, go	Hybe's protection on their artists
Topic 2	lesserafim, hybe, bully, issue, garam, peak, fearless, go, get, cover, music, chart, week, new, take know	Promoting Le Sserafim's new song called 'Fearless' to top the charts
Topic 3	H=hybe, victim, bully, statement, kpop, think, say, go, kpop, girl, get, issue, use	Bullying issue concerning Le Sserafim
Topic 4	Sserafim, hybe, case, garamout, lesserafim, girl, go, fake, garambully, take, vote, say, show, wear, fearless,	Accusations on Garam as a bully and Hybe filing a case

	know, see, need, new, group, girl	
--	-----------------------------------	--

Conclusion

The project explored sentiments of tweets related to Le Sserafim collected from May 26 to June 2, 2022. Majority of the tweets evoke a neutral sentiment having a polarity score of 0.0. The study of (Grishechko 79) explains that neutral tweets signifies a fact-based approach in communication with a lack of strong emotional associations Furthermore, five word cloud figures were generated representing five potential topics from the analysis. The results showed that the tweets were heavy around the alleged bullying issue of a group's member named Kim Garam. The multiple overlapping words in different themes suggests the need to refine topic models to produce cleaner key themes which are more interpretable.

References

- Grishechko, Elizaveta G. "The Linguistic Landscape of "Controversial": Sentiment and Theme Distribution Insights." *GEMA Online Journal of Language Studies/GEMA*, vol. 24, no. 1, 28 Feb. 2024, pp. 79–97, <https://doi.org/10.17576/gema-2024-2401-05>.
- Kherwa, Pooja, and Poonam Bansal. "Topic Modeling: A Comprehensive Review." *ICST Transactions on Scalable Information Systems*, 2018, pp. 1-2. *ResearchGate*, Topic Modeling: A Comprehensive Review.
- Mergim, Hoti H., and Ajdari Jaumin. "SENTIMENT ANALYSIS USING THE VADER MODEL FOR ASSESSING COMPANY SERVICES BASED ON POSTS ON SOCIAL MEDIA." *sciendo*, vol. 18, no. 2, 2023, p. 21. *ResearchGate*, https://www.researchgate.net/publication/376973160_Sentiment_Analysis_Using_the_Vader_Model_for_Assessing_Company_Services_Based_on_Posts_on_Social_Media.