Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Automated Detection and Filtering Off-Topic Comments on YouTube

Natasha Safdar
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.

Dr. Ahmad Khan Shazli
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan. Email: ahmad.khan.fsd@superior.edu.pk

Shahzad Nazir
Baba Guru Nanak University, Nanakana Sahib, Punjab, Pakistan

ABSTRACT

Popularity of YouTube has drawn different advertisers and content sharing users to advertise their content and videos, but the individuals will do off-topic comments ("we love Prophet (PBUH), we love Allah") which are completely off-topic from that content which was uploaded by influencer. In this paper, we had implemented a novel approach for automatically detect & filtered off-topic comments of YouTube videos based on natural language processing methods. In this work, a specific dataset of comments on YouTube has been collected and preprocessed using feature extraction through Bag-of-Words. Machine learning model LLM (Large Language Model) has been trained and tested to obtain various perspectives in their outputs while classifying comments as on-topic or off-topic. We had cull and categorized the off-topic comments into various categories of both sentiment and sentence types that assist YouTubers in discovering appropriate comments for expanding their audience in this paper. Previous work has dealt with sentiment analysis (spam, positive and negative) or classification of sub-categories within the same sentence categories (e.g., question types) on a text corpus. These have limited extension to non-traditional text corpus such as YouTube comments. We tackle this problem of text extraction and classification from YouTube comments with popular statistical metrics and machine learning models. We assessed each statistical metric versus the machine learning model by cross validation and scoring. The results showed that our method incorporating standard methods works well on the classification task, justifying its viability in helping content creators boost viewer engagement on their channel.

Keywords: YouTube, Machine Learning Model, Irrelevant, Off-Topic, Comments, Contents Creators

INTRODUCTION

Over the last few years, YouTube has become incredibly popular among creators. Many creators put their video content on this site. These videos receive loads of views

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

and comments. These creators, commonly referred to as YouTubers, must constantly work on improving the quality and quantity of their contents. To achieve this, they must gather feedback from their audience through the comments box. This feedback lets them understand the influence of their creations. In addition to improving audience engagement, feedback also provides information on the aspects of the content that need improvement. But not all YouTubers have time to read all the comments on each video. Actually, they need to read all the comments to gain a complete understanding of the public interest in their content. The solution for this disturbance is discussed in our work. Our solution is to retrieve all the comments of a video and classify them into two categories: relevant & off-topic. These categories allow You Tubers to concentrate on only those comments which are of their choice. There have been several researches in sentiment analysis like Twitter sentiment analysis Apoorv Agarwal et al., YouTube polarity trend analysis Amar Krishnaet al., sentiment analysis of user comment on YouTube Hanif Bhuiyan et al., etc. But no significant amount of research has been done in sentiment analysis by classifying a sentence according to its type. We have dealt with this problem from the angle of YouTube comments. One of the easiest methods of solving the issue is to label the comments purely on lexicon , the off-topic comments can be recognized from keywords ("we love Prophet (PBUH), we love Allah,"). Additionally, this approach does poorly if a comment contains both categories. These kinds of comments are better categorized by properly extracting features from the corpus of text and employing supervised machine learning methods in the research of Ammar Ismael Kadhim et al. etc. Neural networks are a possible solution, but they're hard to tune and not easily explainable Chunting Zhou et al.. Explain ability is particularly necessary for comments that belong to more than one category to know the reasons behind a resulting category being chosen.

Our strategy is to derive features from preprocessed data and utilize the features to train established supervised learning algorithms. The supervised learning algorithms can then be fine-tuned in order to achieve the optimal results. Because the model's performance relies on the text corpus, we choose several popular fine-tuned algorithms for this purpose and monitor their performance. We try our YouTube comments dataset with five various fined tuned classification models based on two different feature extraction techniques in order to get the best outcome for every model. The performance of the models are computed in terms of cross validation score and F1 score. While our method is simple, the outcome is efficient allowing the creators to see their feedback of interest without a hassle.

The rest of the paper is structured as below: Section 2 discusses the related work on text feature extraction, sentiment analysis of YouTube comments, and sentence type classification. Section 3 provides our methodology. Likewise, Section 4 contains the results of the methodology and Section 5 concludes the paper with some comments and future improvements.

RELATED WORK

Different research has been conducted on feature extraction from text. Poche et al. in their thesis have discussed the commenting habits of viewers on coding tutorial videos. With SVM and Naive Bayes models, the comments gathered were categorized

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

into two wide categories: Content Concerns and Miscellaneous. But comment classification into multiple sentence types is a more intricate task; a fine-grained classification of "Content Concerns" was done in our work with better accuracy.

Taboada et al. used lexicon based techniques for sentiment analysis. In this research work, a word dictionary was developed and words were ranked manually into a range of -5 to +5 (negative to positive sentiment). Our supervised learning method substitutes the manual effort done by them in their classification step, giving a more accurate result. Our time consuming word scaling process was also made automatic by our machine learning method. Our work categorizes text on the basis of sentence type as well as sentiment. Siersdorfer et al. utilized linear support vector machine with a thesaurus to achieve the level of negativity and positivity of each word from comments. The accuracy of this method is “0.72”. Even without resorting to any thesauruses, we have managed to improve the accuracy of our model for sentiment analysis.

Similarly, Krishna et al. employed Naïve Bayes classification to perform polarity (sentiment) analysis over the IMDB dataset. Their method utilizes the category of every comment while selecting features, which implies it is unable to determine the polarity of a new comment since its features cannot be determined without the knowledge of its category. Our method never employs category while selecting features. In a different manner, the opinion on comments is utilized to assist in the location of the target video on YouTube by the search text using Bhuiyan et al..

This method was limited to classifying the comments as positive and negative.

Maas et al. presented a sophisticated idea by demonstrating the similarity between the words in their rational meaning along with sentimental sense. The semantic component was trained using an unsupervised model, while the sentiment component was trained using a supervised model (SVM). While their model provided the text polarity, they did not test the sentence type. Based on the traditional NLP tool, we could classify text according to both sentiment and sentence type.

On classification by sentence types, Madden et al. presented a classification scheme for YouTube comments. They suggested that individuals' commenting behavior varies across groups; some comment for promotion, some for information provision, and others for fun. Madden's effort has yielded 10 overall categories and 58 subcategories for such comments. They have only presented a schema for the task of classification. We tried to apply some of their classes into the classification problem. Cheung et al. also classified questions for Webclopedia data. Their task was restricted to classifying various types of question.

CNN models variations employed were CNN-rand, CNN-static, CNN-non-static and CNN multichannel. Hassan et al. introduced a recurrent neural network based model Long Short-Term Memory (LSTM) for sentiment analysis, on the opposite side. Both these methods yielded good results, but they haven't ventured to explore the classification of various types of sentences (such as imperative, question, corrective, and sentimental). They are also not beyond sentiment analysis or question classification. Our work addresses much more generic sentence classification.

In the majority of the current work, the imperative class sentences have not been studied appropriately. Khoo et al. [10] conducted experiments on various models for 14 sentence classes (including imperative sentence classes such as request, instruction

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

and suggestions). Naive Bayes, Decision Tree, and Support Vector Machine were the models experimented with. Support Vector Machine dominated all other models and did not experience significant impact from feature selection. They have selected the standard response emails in their work only because the sentences are well structured and there are not many grammatical mistakes. This makes the classification process easier. Our task comprises a huge number of unstructured sentences with enormous grammatical mistakes. Table. 1 is a comparative study of our work and the earlier works.

Table 1: Summary of the related work

Study	Categories	Methodology
Poche et al. [15]	Content Concerns, Miscellaneous	SVM, Naive Bayes
Taboada et al. [18]	Setiments	Lexicon
Siersdorfer et al. [16]	Sentiments	Linear SVM + thesaurus
Krishna et al. [12]	Sentiments	Naïve Bayes
Bhuiyan et al. [4]	Sentiments	Lexicon
Maas et al. [13]	Sentiment + Semantic	Unsupervised + Supervised models
Madden et al. [14]	Numerous	Manual
Cheung et al. [5]	Question types	Naïve Bayes, Decision Tree
Hassan et al. [7]	Sentiments	Recurrent NN
Khoo et al. [10]	different sentence types	Naive Bayes, Decision Tree, SVM

METHODOLOGY

In our method, we utilized python programming language as in (Rishabh Ahuja,2019) to carry out the experiment. We obtained comments from YouTube videos through the YouTube Data API, which are then preprocessed to eliminate the noise. The features are then extracted from the comments and input into the classification models. The models are finally hyper tuned to deliver the optimal outcome.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

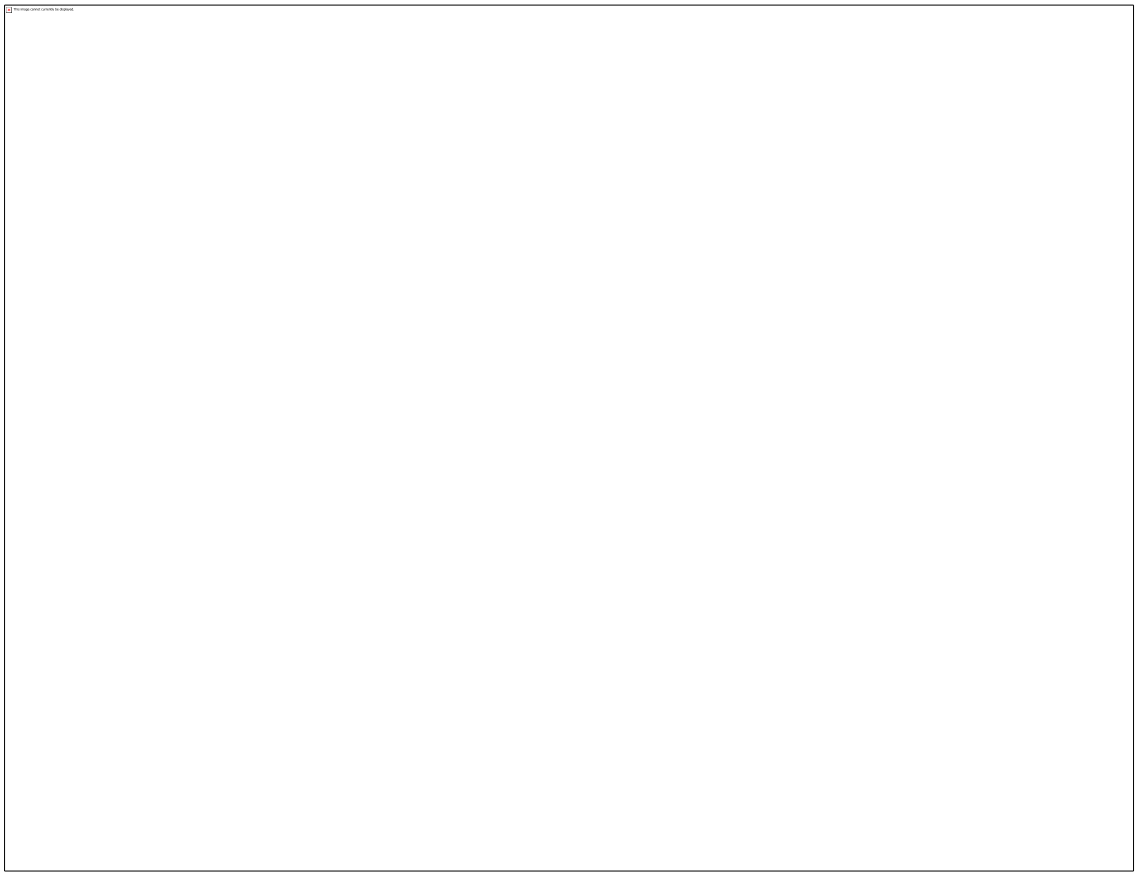


Figure 1: Process of extracting and training model

Data Collection

We constructed our dataset through YouTube comment scraping. We utilize YouTube Data API for authenticating and retrieving the comments of videos (Douji Yasmina, 2016). The API is first utilized for authentication, and then obtained credentials are passed to the comment extractor. The comment extractor scrapes comments from the comment section by scrolling all the comments and loading them dynamically. Fig. 1 illustrates the scraping process.

The data is a collection of comments randomly selected from various random videos. We used news, vlogs and songs videos in our experiments because they have a lot of diversity in comments. We then manually label the comments into 2 different classes: Relevant & off-topic. These classes are defined based on general needs of the YouTubers. Note that further categories can be established if or as needed. Table. 2 shows different classes and the content of that class. The classes are explained in more detail next.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about



Figure 2: The process of scraping

Relevant tells that the viewers perceived the content as worthy and that the content created a positive impact on them. Off-Topic provides the content that is not relate to topic and also it contains some Islamic words ("we love Prophet (PBUH), we love Allah,").

Table 2: Classes of comments with content type

Class	Content
Relevant	Which are relevant to videos
Off-Topic	Which contain some Islamic words Such as("we love Prophet PBUH, Allah is best etc")

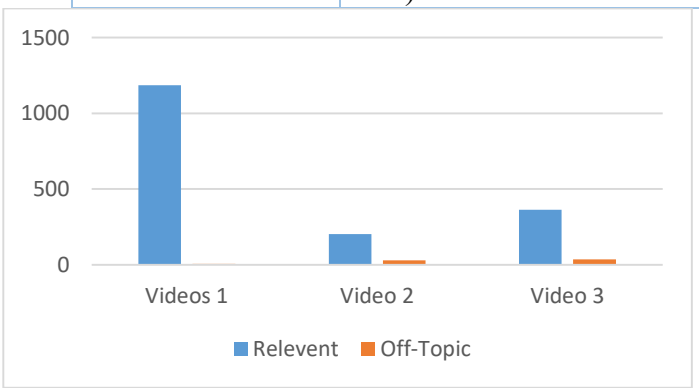


Figure 3: Number of comments in graphical form of vlog

Here, we have taken it as an overall overview of sentence as answering the question is more crucial to raise chances of the viewer to come back and remain interested in the channel's content. Fig. 2 indicates the graphical representation of the number of comments in each class.

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Data Pre-processing

It is necessary to sanitized the data and put them in proper form to enhance classification. The step of pre-processing the data addresses the following issues that make the classification task cumbersome:

Non-standard language: The text employed in the comments area is not always Standard English. Slangs and wrong form of words are present in comments, so it becomes challenging to drive features from them.

Spelling mistakes: Spellings are seldom checked by commenters because of the relaxed environment. These spelling mistakes have to be corrected; if not, the words would be adding unwanted attributes to the classification algorithm, lowering the overall classifier accuracy and affecting the efficiency of the method. For instance, words "plz" and "please" hold the same semantic meaning when used in casual writing, but without giving any attention to the wrong spelling, they would be considered as two distinct words by the classifier.

Unformatted texts: They are comments that have computer codes written in them. These are not helpful in precision of feature extraction; they merely impose unnecessary burden on the feature matrix.

Spurious comments: Not every comment posted was on the video or pertaining to the channel. A lot of viewers post comments to promote their products or simply to indicate their presence. The comments are of no use to the content providers and introduce unnecessary overhead only.

Since there are a lot of types in casual writing, we had corrected the types by using the autocorrect-library of Python. We have used Lemmatization to look at the words morphologically and bring together the similar words in one group. In addition, there are some common words that do not contribute much to the meaning in the sentence like "is", "are", and "it". We have eliminated them from the corpus. Table. 3 indicates the stopwords which were not taken into account in each category.

Conversion of text into features

After pre-processing has been done, the pre-processed text is then filtered to ensure that only the required features are left in the corpus. This reduces the burden to the classifiers and also boosts the performance of chosen models [19].

Table 3: Stopwords ignored in each class

Class	Ignored Stopwords
Positive	NA
Negative	"allah", "muhammad", "prophet", "islam","quran","hadith", "sunnah", "اللّٰہ", "الله", "رسول", "محمد", "قرآن", "حدیث", "اسلام", "مسلمان", "نیکی", "گناہ", "عبادت",

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Model Choice and Hyper parameter Adjustment

We choose Large Learning Model(LLM) because it is specialy used in Artificial Intelligence (AI) to understand and generate human like language. It is used for various purpose such as customer support, content generation language translation etc. LLM also used for Text classification, Sentiment analysis, Entity extraction, so we also gave this model first priority for our research.

RESULTS

Experimental Setup

All experiments were coded in Python programming language and run through Google Colab. The "pandas" library was utilized for data importing. Pre-processing was done by regex functionality of the "re" library. Stopwords were eliminated by the "nltk" library. Likewise, spell correctors utilized the "autocorrect" library of Python. The classification and feature selection process employed the "sklearnmodule". To train, 80% of comments were randomly taken from the dataset, and 20% was reserved for testing. The experiments were conducted on Google Colab with 16.29 GB RAM and 2.20GHz Intel(R) Xeon(R) CPU with GPU disabled.

Result Analysis

Firstly we make CSV file than we make manual labelling for preprocessing of data after that we make feature selection and the model we apply is LLM model and produce a result. For result analysis we take some videos of different channel from youtube and represent it graphically which predict relevant and off-topic comments.

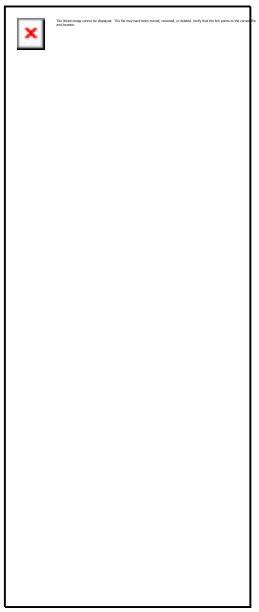


Figure 4: The complete classification

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about



Figure 5: Graphical Analysis of Vlogs Video (1)

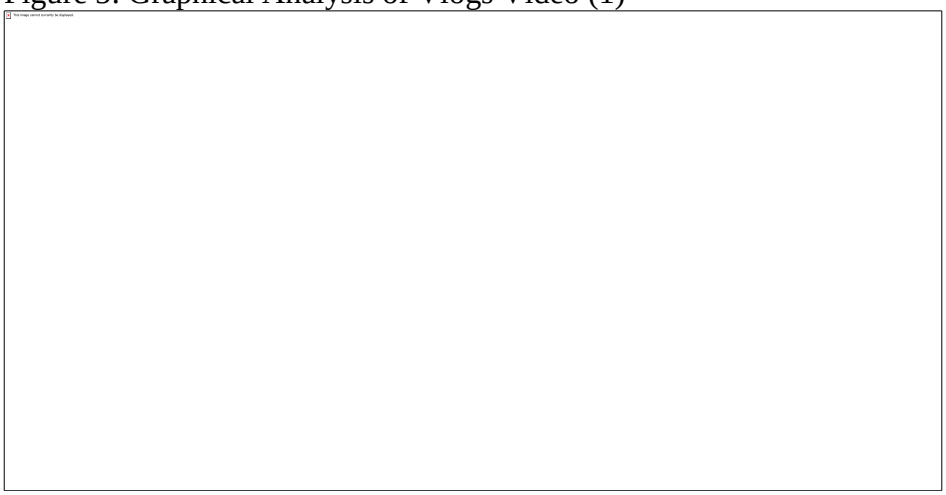


Figure 6: Graphical Analysis of Vlogs Video (2)

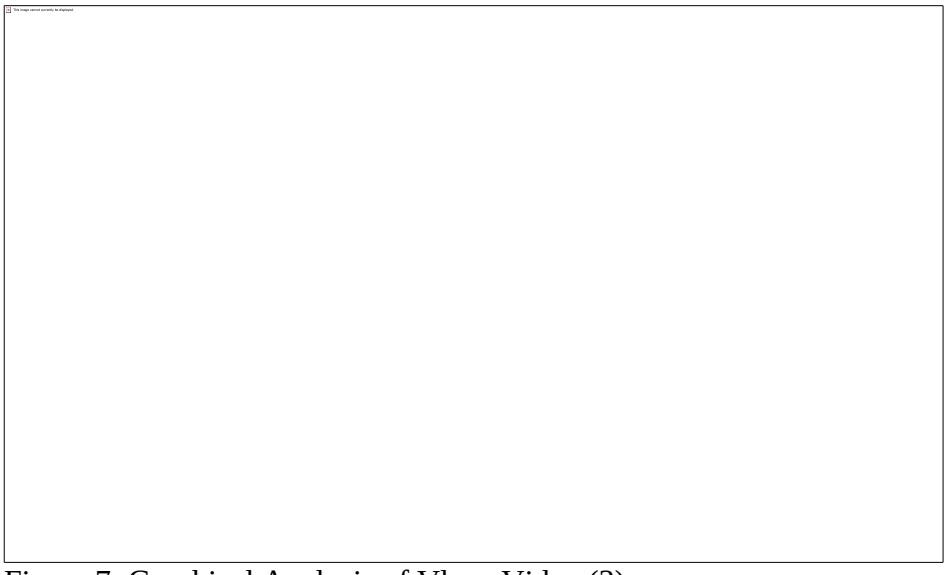


Figure 7: Graphical Analysis of Vlogs Video (3)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

Explanation:
Here above, I take a video 3 link of vlog from YouTube. The Green bar represent off-topic comments and red bar represent relevant videos.



Figure 8: Graphical Analysis of News Video (1)

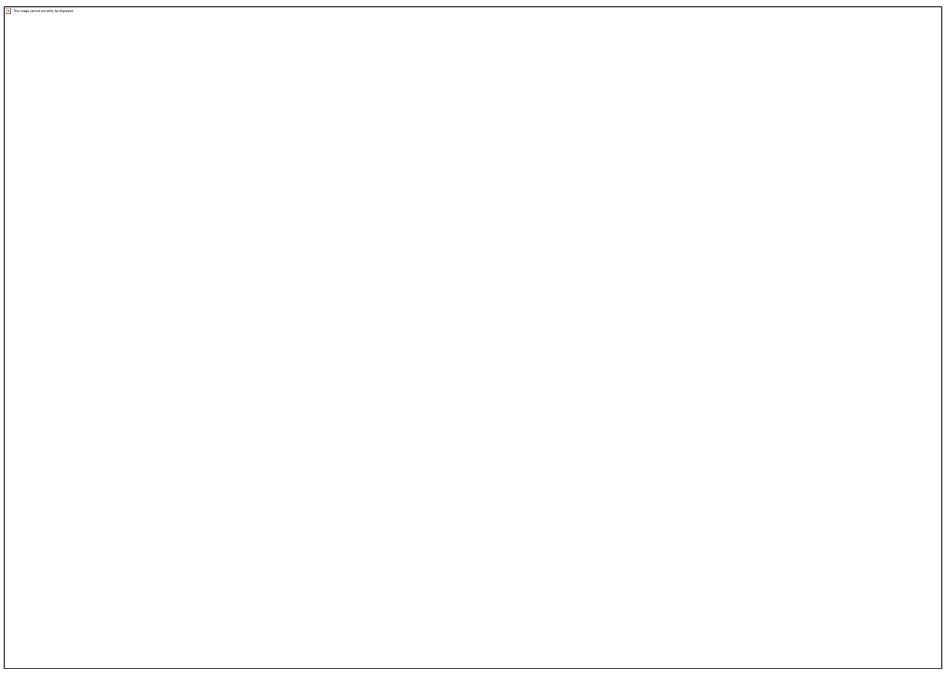


Figure 8: Graphical Analysis of News Video (2)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about



Figure 4: Graphical Analysis of News Video (3)

Explanation:
Here above, I take 3 video link of News from YouTube. The Green bar represent off-topic comments and red bar represent relevant videos.

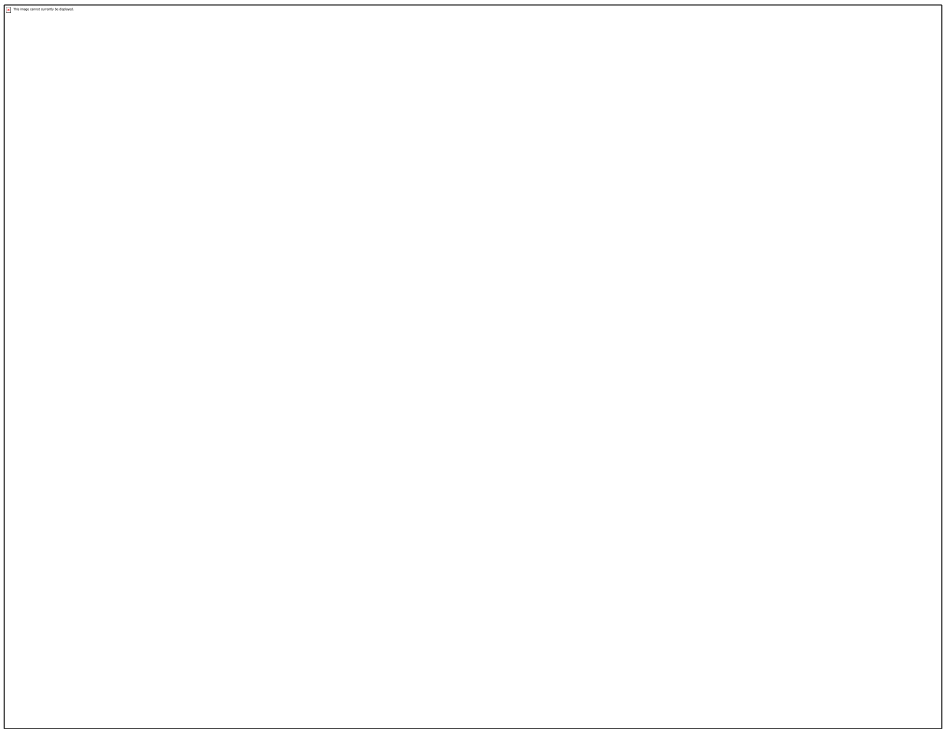


Figure 5: Graphical Analysis of Songs Video (1)

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about



Figure 5: Graphical Analysis of Songs Video (2)



Figure 5: Graphical Analysis of Songs Video

Explanation:

Here above, I take a video link of song from YouTube. The Green bar represent off-topic comments and red bar represent relevant videos.

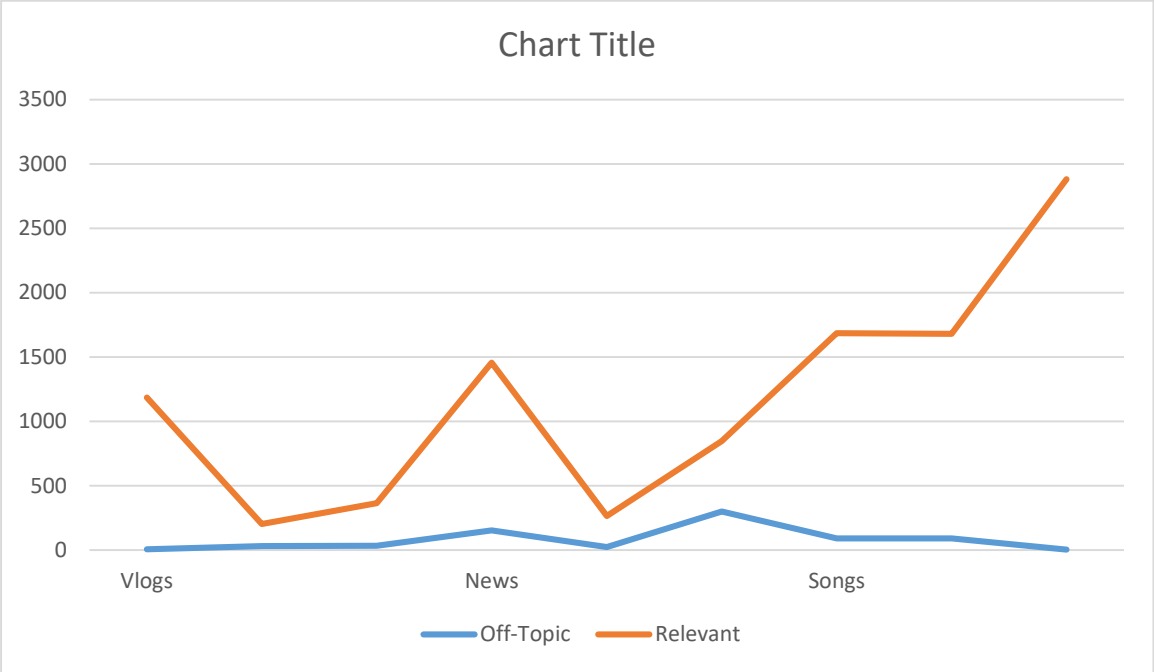
LLM (Large Language Model): Designed for natural language processing (NLP) tasks, LLMs are typically based on transformer architectures and are trained on vast amounts of text data. They're used for tasks like language translation, text generation,

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189

http://amresearchreview.com/index.php/Journal/about

and question answering.



Conclusion

With the proper classification of comments into their categories, a YouTuber can easily view corresponding comments. The corresponding comments reveal the public opinions regarding the videos whereas off-topic comments will be filtered once detected which include some Islamic terms like (Who love prophet PBUH, etc).This can assist a YouTuber to never have to scroll through hundreds of comments and filter them manually for every.

By the proper classification of the comments into their respective categories, a YouTuber can readily view each category of comment.Other categories help the YouTuber to clearly view questions raised about the video and the suggestions made about how to better the content. This can save the YouTuber from scrolling through hundreds of comments and filtering them manually for a given video. While earlier researchers have used either sentiment analysis or sentence classification of a niche, we have used both the aspectsExperiments revealed that best cross validation. The scores were achieved by Logistic Regression.

In future, the classes and sub-classes can be expanded in order to cover a more inclusive comment classification. Similarly, the class models and overall feature selection method can be enhanced for comments that lie in multiple classes. In the future, we will also extend this research by conducting a comparative study with Explainable Neural Networks (xNNs) on comments classification.

REFERENCES

Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment Analysis of Twitter Data. In Proceedings of the Workshop on

Annual Methodological Archive Research Review
http://amresearchreview.com/index.php/Journal/about
Volume 3, Issue 11 (2025)

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Languages in Social Media (Portland, Oregon) (LSM '11). Association for Computational Linguistics, USA, 30–38.

Rishabh Ahuja, Arun Solanki, and Anand Nayyar. 2019. Movie Recommender System Using K-Means Clustering AND K-Nearest Neighbor. In 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence). 263–268. https://doi.org/10.1109/CONFLUENCE.2019.8776969

Khin Zezawar Aung and Nyein Nyein Myo. 2017. Sentiment analysis of students' comment using lexicon based approach. In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). 149–154. https://doi.org/ 10.1109/ICIS.2017.7959985

Hanif Bhuiyan, Rajon Bardhan, and Dr. MD Rashedul Islam. 2017. Retrieving YouTube video by sentiment analysis on user comment. 474–478. https://doi.org/10.1109/ICSIPA.2017.8120658

Zhalaing Cheung, Khanh Linh Phan, Ashesh Mahidadia, and Achim Hoffmann. 2005. Feature Extraction for Learning to Classify Questions. In AI 2004: Advances in Artificial Intelligence, Geoffrey I. Webb and Xinghuo Yu (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1069–1075.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural Language Processing (Almost) from Scratch. J. Mach. Learn. Res. 12, null (Nov. 2011), 2493–2537.

A. Hassan and A. Mahmood. 2017. Deep learning for sentence classification. In 2017 IEEE Long Island Systems, Applications and Technology Conference (LISAT). 1–5.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political Ideology Detection Using Recursive Neural Networks. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Baltimore, Maryland, 1113–1122. https://doi.org/10.3115/v1/P14-1105

Ammar Ismael Kadhim. 2019. Survey on supervised machine learning techniques for automatic text classification. Artificial Intelligence Review 52, 1 (2019), 273–292.

Anthony Khoo, Yuval Marom, and David Albrecht. 2006. Experiments with Sentence Classification. In Proceedings of the Australasian Language Technology Workshop 2006. Sydney, Australia, 18–25. https://www.aclweb.org/anthology/ U06-1005

Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Doha, Qatar, 1746–1751. https://doi.org/10.3115/v1/D14-1181

Amar Krishna, Joseph Zambreno, and Sandeep Krishnan. 2013. Polarity Trend Analysis of Public Sentiment on YouTube. In Proceedings of the 19th International Conference on Management of Data (Ahmedabad, India) (COMAD '13). Computer Society of India, Mumbai, Maharashtra, IND, 125–128.

Online ISSN 3007-3197
Print ISSN 3007-3189
http://amresearchreview.com/index.php/Journal/about

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning Word Vectors for Sentiment Analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Portland, Oregon, USA, 142–150. https://www.aclweb.org/anthology/P11-1015

Amy Madden, Ian Ruthven, and David McMenemy. 2013. A classification scheme for content analyses of YouTube video comments. Journal of Documentation 69,

5 (2013), 693–714. https://doi.org/10.1108/JD-06-2012-0078

Elizabeth Heidi Poche. 2017. Analyzing User Comments On YouTube Coding Tutorial Videos. Master of Science in Computer Science. Louisiana State University and Agricultural and Mechanical College, Baton Rouge, United States.

Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. 2010. How Useful Are Your Comments? Analyzing and Predicting Youtube Comments

and Comment Ratings. In Proceedings of the 19th International Conference on World Wide Web (Raleigh, North Carolina, USA) (WWW '10). Association for

Computing Machinery, New York, NY, USA, 891–900. https://doi.org/10.1145/ 1772690.1772781

Richard Socher, Jeffrey Pennington, Eric H. Huang, Andrew Y. Ng, and Christopher D. Manning. 2011. Semi-Supervised Recursive Autoencoders for Predicting Sentiment Distributions. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Edinburgh, Scotland, UK., 151–161. https://www.aclweb.org/anthology/D11-1014

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede.

2011. Lexicon-Based Methods for Sentiment Analysis. Comput. Linguist. 37, 2 (June 2011), 267–307. https://doi.org/10.1162/COLI_a_00049

Yiming Yang and Jan O. Pedersen. 1997. A Comparative Study on Feature Selection

in Text Categorization. In Proceedings of the Fourteenth International Conference on Machine Learning (ICML '97). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 412–420.

Douiji Yasmina, Mousannif Hajar, and Al Moatassime Hassan. 2016. Using

YouTube Comments for Text-based Emotion Recognition. Procedia Computer Science 83 (2016), 292 – 299. https://doi.org/10.1016/j.procs.2016.04.128 The 7th International Conference on Ambient Systems, Networks and Technologies (ANT

2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.

Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. 2015. A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630 (2015).