Real-time Aspect Sentiment Quad Predictions for Vietnamese Youtube Comments

Hao Phuc Nguyen[1,2], Tai Huu Le[1,2], Luong Ngoc Nguyen[1,2], and Hop Trong Do[1,2]

[1] Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City, Vietnam
[2] Vietnam National University, Ho Chi Minh City, Vietnam
{21522407, 21522562, 21522311}@gm.uit.edu.vn
{hopdt@uit.edu.vn}

Abstract. Nowadays, there are lots of game shows on the media and many people become famous. This causes the problem that many people fake famous, or famous in a negative way. To solve this problem NLP is the best way, NLP can extract the aspect of the entity, and can show the sentiment of words about the entity. In last 5 years, ABSA has been a trend of NLP because it can solve many problems in the media and other aspects of life and lately more approach of ABSA has been create and especially Aspect Sentiment Quad Prediction (ASQP), with ability to predict 4 elements (aspect category, aspect term, opinion term, sentiment polarity), it appears as an best approach in ABSA for media problem.

Keywords: Deep Learning · NLP · ASBA · ASQP · Apache Spark · Apache Kafka · Spark Streaming

1 Introduction

The surge in reality shows has generated a vast amount of social media data, which holds significant value for various stakeholders. Producers can utilize this data to enhance their shows, while viewers can rely on it to decide whether a show is worth watching. Additionally, this data often contains controversies related to various aspects and entities, making it a fertile ground for aspect-based sentiment analysis (ABSA).

Sentiment analysis (SA) classifies data as Positive, Negative, or Neutral, providing users with valuable insights. However, traditional sentiment analysis cannot delve into the specifics of particular aspects, components, or functionalities of an entity. To address this limitation, researchers have introduced aspect-based sentiment analysis (ABSA), which allows for a more detailed examination of an entity's attributes.

ABSA is a fine-grained opinion mining technique that focuses on analyzing sentiment at the aspect level. It involves four key elements: **aspect category** (identifying and categorizing the relevant aspects of the entity), **aspect term** (terms explicitly or implicitly mentioned in the text), **opinion term** (expressions of opinion related to the aspect), and **sentiment polarity** (indicating the sentiment class).

Recently, ABSA has garnered considerable attention due to its technical advancements. Despite its popularity, most ABSA models only perform partial predictions rather than providing a comprehensive aspect-level sentiment analysis. To overcome this limitation, researchers have developed a new approach called Aspect Sentiment Quad Prediction (ASQP). This approach aims to predict all four elements—aspect category, aspect term, opinion term, and sentiment polarity—simultaneously for a given opinionated sentence.

In this paper:
Input: a comment from Rap Viet video or King of Rap video on YouTube.
Output: a quad including the aspect category, aspect term, opinion term, and sentiment polarity.

Nowadays, Vietnam hosts numerous reality shows that often spark controversies involving various entities such as participants, show content, music, and images....Most of these shows lack a comprehensive rating system to evaluate different aspects. We believe that implementing Aspect Sentiment Quad Prediction (ASQP) could significantly enhance show quality. This approach would provide producers with valuable insights for improvement and offer viewers an objective perspective on the show, the candidates, the music, the examiner...

2 Related Works

Aspect-Based Sentiment Analysis (ABSA) was first introduced as a SemEval task in 2014 (SE-ABSA14), providing benchmark datasets of English reviews and a common evaluation framework Potinki et al. (2014) [10]. These datasets included aspect terms (e.g., "hard disk," "pizza") and their polarity for laptop and restaurant reviews, as well as broader aspect categories (e.g., FOOD) with their polarity specifically for the restaurant domain. The task was continued in SemEval 2015 (SE-ABSA15) to encourage more comprehensive research by introducing a new ABSA framework. This framework required that all identified components of expressed opinions (aspects, opinion target expressions, and sentiment polarities) adhere to a set of guidelines and be interlinked within tuples. In this new framework, an aspect category is defined as a combination of an entity type (e.g., LAPTOP, KEYBOARD, CUSTOMER SUPPORT, RESTAURANT, FOOD) and an attribute type (e.g., USABILITY, QUALITY, PRICE), thereby clarifying the distinction between entities and the specific facets being evaluated Potinki et al. (2015) [9]. The SemEval-2016 task-5 (SE-ABSA16) further extended the SE-ABSA15 dataset to include new domains such as Hotels, Consumer Electronics, Telecom, Museums, and other languages Potinki et al.

(2016) [8].

In recent years, numerous researchers have published studies on aspect-based sentiment analysis (ABSA) with promising results (Nazir et al. (2022) [5]; Do et al. (2019) [1]). However, most of these studies do not predict all quads in a single step. In 2020, Peng et al. (2020) [7] proposed the aspect sentiment triplet extraction (ASTE) task, which garnered significant attention (Xu et al. (2020) [14]; Huang et al. (2021) [2]). Building on this, Zhang et al. (2021) [15] introduced Aspect Sentiment Quad Prediction in 2021, marking a new advancement in natural language processing (NLP).

In Vietnam, many universities and researchers have contributed to the field of aspect-based sentiment analysis (ABSA) by publishing various datasets and studies. Notable examples include the Vietnamese ABSA corpus on smartphone reviews by Mai & Le (2018) [4] and the SA-VLSP2018 dataset, which focuses on hotel and restaurant reviews, published by Nguyen et al. (2018) [6]. A related dataset was created by Nguyen et al. (2019) for sentiment classification and aspect extraction, building on earlier work. Additionally, Van Thin et al. (2021) [13] developed a sentence-level dataset on the same topic, annotated with high inter-annotator agreement. The UIT-ViSFD benchmark dataset, created by Luc Phan et al. (2021) [3], evaluates ABSA for mobile e-commerce, while Thanh et al. (2021) introduced the UIT ViSD4SA dataset, a benchmark Vietnamese smartphone feedback dataset featuring span detection for ABSA. Most recently, Tran et al. (2022) [12] applied ABSA to a dataset on e-commerce beauty product reviews. Despite these efforts, there remains a shortage of research on new approaches to ABSA Thin & Nguyen (2023) [11], and even more so on Aspect Sentiment Quad Prediction (ASQP).

In addition, in 2024, Duy and his colleagues made significant efforts to create a comprehensive and detailed dataset on Rap Viet. This not only represents an important achievement for them but also serves as a solid foundation for us in our research and development of our own dataset. Thanks to the dedication and hard work of Duy and his associates, we can continue to build upon and expand what they have established, contributing to the holistic and sustainable development of this field.

3 Dataset Description

3.1 Source and Crawling Method

Our dataset was built by collecting comment data from 2 popular rap TV shows in Vietnam: Rap Viet and King of Rap on YouTube. The data collection was done using the Youtube API and the Selenium library, includes various details that provide an in-depth look into these popular Vietnamese rap competition shows. The dataset encompasses information about comments of people for the

Aspect Category	**Definition**
Character - Outlook	Comments about the person outlook
Character - Voice	Comments about the person voice (mostly is show candidate)
Character - Overall	Comments describe other aspects of the person (style, character. . .)
Show	Comments that concern to the show
Song	Comments refer to the music, the melody, the song
Examiner	Comments about the examiner and coach that appear in the show
Others	Spam comments

Table 1: The description of each column

Candidates, examnier or show, stage,... Our dataset includes 8405 comments crawled from 2 playlist, the first one including 118 videos about Rap Viet in season 3 and another one including 14 videos about King of Rap. About sentiment polarity, we design it with 3 labels: Positive, Negative, and Neutral. For the aspect category detection, we design this subtask with 6 aspects including Character - Overall, Character - Voice, Character - Outlook, Examiner, Show and Song. We also have an aspect named Others to detect spam or comments that do not have any aspect or sentiment. Their definitions are shown in Table [1]. We split the dataset into 4 species: train set with 7000 comments, validation set with 500 comments, test set with 500 comments, and stream set with 405 comments for simulating data streaming with Spark and Kafka

3.2 Annotation Process

First of all, we define a basic annotation guideline. For training, we randomly take 100 comments to annotate, then we calculate Cohen's Kappa score for those annotated comments. For labels that do not have enough 3 agreements, we gather and discuss a new rule to update our guidelines. We trained four rounds to obtain Cohen's Kappa score higher than 85% before performing data annotation independently.

After the training phase, the rest of the data was split into 3 parts and each member supervison 1 part following the rules in our annotation guideline. Those comments in the training phase are instruction for ChatGPT to help us in annotation process. Comments was give to ChatGPT to extract 4 elements of quads and ours member acts as supervisor and controller. Of course, after using chatGPT to annotate, we will check again to ensure accuracy.

3.3 Data Preprocessing

For data preprocessing, we have applied many techniques:

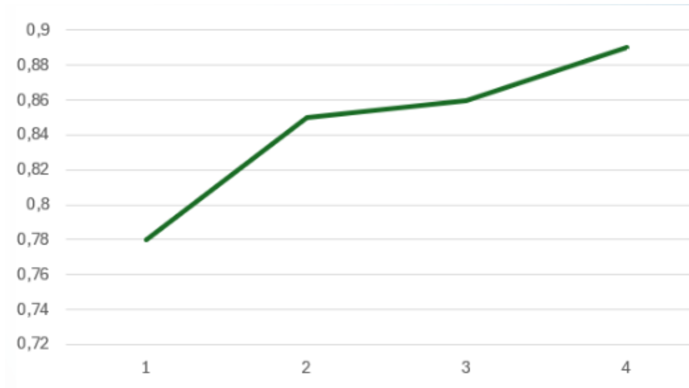

Fig. 1: Cohen - Kappa Score of 4 training phrase

– Remove HTML
– Convert unicode, normalize acronyms, teencode, emoji
– Remove unnecessary, duplication characters

Padding is used with a maximum length of 128, which is chosen based on the word length distribution chart of the dataset shown in Figure [2].

3.4 Statistics

Our dataset contains 8405 comments, including 6 sentiment aspects except "Others", with each sentiment aspect having three sentiment labels: Neural, Positive, and Neg ative. As you can see in Table [2], there are 4 aspects that receive a high num ber of comments compared to other aspects: "Character - Overall" with 3492 comments classified, "Others" (2338), "Song" (1770) and "Show" (1587). These 4 aspects have accounted for more than 87% of the dataset, which mean that our dataset is imbalance. In terms of sentiment distribution, Positive labels dominate across all aspects (accounting for average 70%), showing more interest of viewers to rappers and shows. This also shows that there is an imbalance in the number of labels in each aspect, which can affect the performance of deep learning models. Overall, this dataset provides a valuable foundation for sentiment analysis, with further exploration recommended to uncover deeper insights into specific aspects and sentiments.

4 Model

4.1 Approach

We use the Paraphrase Generation method presented in Aspect Sentiment Quad Predictions as a Paraphrase Generation paper ([15]). With an input comment,

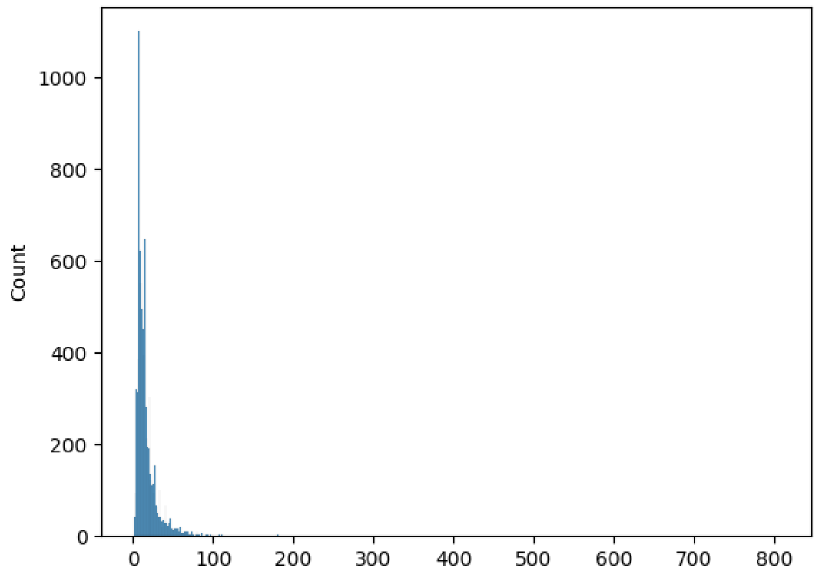

Fig. 2: Word length distribution of the dataset

	Character - Overall	Character - Voice	Character - Outlook	Examiner	Show	Song	Others
Positive	2357	518	100	135	937	638	0
Negative	688	212	35	95	453	612	0
Neutral	447	71	23	69	197	520	2338
Total	3492	801	158	399	1587	1770	2338

Table 2: Aspect category and Sentiment polarity distribution in the dataset

aspect sentiment quad predictions (ASQP) aim to predict aspect category, aspect term, opinion term, and sentiment polarity. This method paraphrases the input comment to neglect unnecessary details and highlight the major sentiment elements:

Aspect category là **Sentiment polarity** bởi vì **Aspect term** là **Opinion term**.

Two examples of paraphrasing the input comments for training using the Para phrase Generation Method are presented in Figure [3]. After training the model with these paraphrased sentences, it will perform the prediction as the form above, which can be used to extract the sentiment quads. The details of this process are shown in Figure [4].

Input – 1	Rap Việt hay hơn nhiều đúng không mọi người
Label – 1	(c, a, o, p) (Show, Rap Việt, hay hơn nhiều, POS)
Target – 1	Show là tuyệt bởi vì Rap Việt là hay hơn nhiều
Input – 2	Đêm chung kết chất lượng hơn bên RV một ngàn lần
Label – 2	(c, a, o, p) (Show, Đêm chung kết, chất lượng hơn bên RV, POS)
Target – 2	Show là tuyệt bởi vì Đêm chung kết là chất lượng hơn bên RV

Fig. 3: Two examples for Paraphrase Generation Method

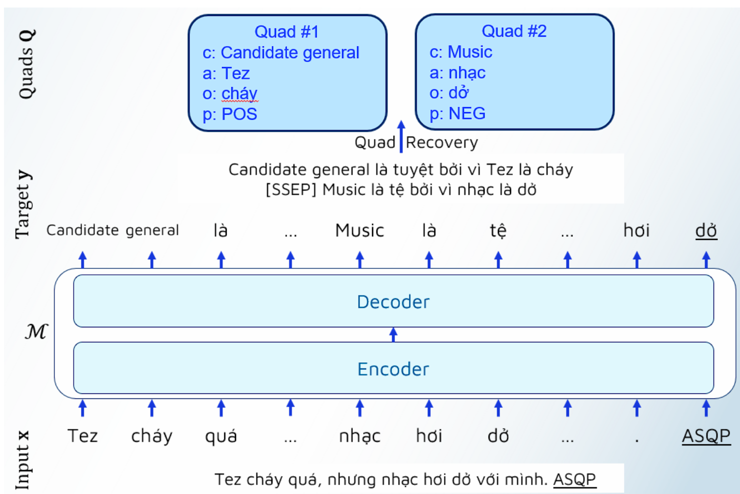

Fig. 4: Paraphrase Generation model framework overview

4.2 Models

Generating text is the task of generating new text given another text. These models can, for example, fill in incomplete text or paraphrase. The process involves generating coherent and meaningful text that resembles natural human communication. Text generation has gained significant importance in various fields, including natural language processing, content creation, customer service, and coding assistance. LLMs, or Large Language Models, are the key component behind text generation. In a nutshell, they consist of large pretrained transformer models trained to predict the next word (or, more precisely, token) given some input text.

For Vietnamese text generation, we use three models: ViT5, BARTPho and ByT5

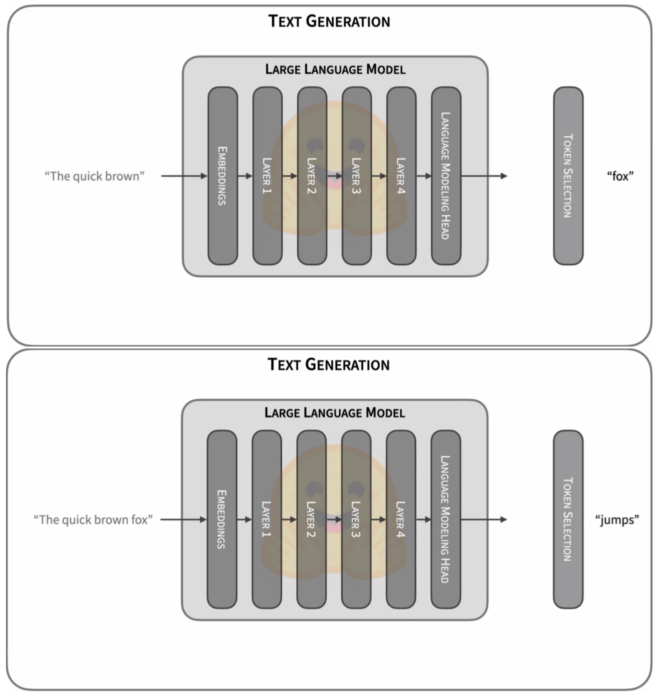

Fig. 5: Example of Text Generation by Large Language Model

ViT5: T5 is an encoder-decoder model pre-trained on a multi-task mixture of unsupervised and supervised tasks for which each task is converted into a text-to-text format. T5 works well on a variety of tasks out-of-the-box by prepending a different prefix to the input corresponding to each task.

ViT5 is a pretrained sequence-to-sequence Transformer model for the Vietnamese language. It is a pretrained Transformer-based encoder-decoder model for the Viet namese language. With T5-style self-supervised pretraining, ViT5 is trained on a large corpus of high-quality and diverse Vietnamese texts.

BARTPho: BART is a denoising auto encoder for pretraining sequence-to-sequence models. It uses a standard Transformer-based neural machine translation architecture which, despite its simplicity, can be seen as generalizing BERT (due to the bidirectional encoder), GPT (with the left-to-right decoder), and many other more recent pretraining schemes.

BARTPho is a large-scale monolingual sequence-to-sequence model pre-trained for Vietnamese. BARTpho uses the "large" architecture and the pre-training scheme of the sequence-to-sequence denoising autoencoder BART, thus it is especially suitable for generative NLP tasks.

ByT5: ByT5 is a tokenizer-free extension of the mT5 model. Instead of using a subword vocabulary like most other pretrained language models (BERT, XLM-R, T5, GPT-3), ByT5 model operates directly on UTF-8 bytes, removing the need for any text preprocessing. Beyond the reduction in system complexity, parameter-matched ByT5 models are competitive with mT5 across a range of tasks, and outperform mT5 on tasks that involve noisy text or are sensitive to spelling and pronunciation.

ByT5 represents an important step towards more universal and flexible NLP models by adopting a token-free approach. By processing text at the byte level, it simplifies the handling of diverse text inputs and reduces some of the complexities associated with traditional tokenization methods.

5 Experiments and Analysis

5.1 Experiment settings

Because the method used in this paper uses a text generation model to paraphrase the input comments, we will apply 3 model to see models' ability to learn and generate text. Adam with 3e-4 learning rate and 1e-8 epsilon is used as Optimizer. We use batch size of 16 and 20 epochs with an Early Stopping function.

5.2 Evaluation metrics

A sentiment quad prediction is counted as correct if and only if all the predicted elements are exactly the same as the gold labels ([[15]]). F1-score macro, which is the harmonic mean of the precision and recall, is used because of the imbalance of aspect and sentiment in the dataset. It thus symmetrically represents both precision and recall in one metric and gives us a more general view of the models' performance.

5.3 Results and Discussion

Our model was successful in identifying and extracting important aspects from the input text. We have divided the results into 3 tables with different evaluation aspects using the F1-score index.

In the overall experimental results shown in Table [3], we evaluate the performance of ViT5, BARTpho and byT5 pre-trained models. byT5 has a best performer precision (0.2790), recall (0.2626), f1-score (0.2706) follow by ViT5 with precision (0.2710), recall (0.2496), f1-score (0.2598) and BARTpho stand last with precision (0.2722), recall (0.2232), f1-score (0.2453).In overall, this result this lower than the dataset before because there are less spam comment than the dataset before. The observed results show that byT5 has a better score than other model about 3% in all 2 evaluation metrics (recall and f1-score) prove

	Precision	Recall	F1-macro
ViT5	0.2710	0.2496	0.2598
BARTpho	0.2722	0.2232	0.2453
byT5	0.2790	0.2626	0.2706

Table 3: The overall experimental results

ability to learn and generate text better than other 2 models and it seems to be a suitable model for the data. In precision the results show that 3 models have mostly equal results.ability to learn and generate text than other 2 models and it seems to be a suitable model for the data.

	ViT5	BARTpho	byT5
Character - Outlook	0.44	0.44	0.40
Character - Voice	0.61	0.50	0.57
Character - Overall	0.69	0.65	0.71
Show	0.71	0.60	0.66
Song	0.72	0.60	0.70
Examiner	0.45	0.42	0.61
Others	0.65	0.61	0.68

Table 4: Aspect Category Detection F1-macro

Opposed to the overall result, Aspect Category Detection table [4] shows that with the new data, the model has better results with all aspects having a f1-score higher than 0.44. This is a big improvement if compared to previous projects with 0 in aspects with less quantity (Character - Outlook, Examiner). With Aspect Category Detection, ViT5 shows that it has a better ability to extract aspects with best results in aspects like 'Character - Outlook', 'Character - Voice', 'Show', 'Song'. And with 3 other aspects 'Character - Overall' , 'Examiner', 'Others' byT5 has a better result. Overall, all 3 models have positive results with aspects that have large quantities 'Character - Overall' with 0.69, 'Show' with 0.71 and 'Song' with 0.72. With small quantity aspect ViT5 has better performance in Character - Outlook with 0.45 and most impressive is byT5 with Examiner with 0.61 f1-score in spite of just having 399 comments in the dataset.

For Sentiment Classification Table [5], all 3 models has a positive result from 0.45 to 0.8 f1-score. ViT5 has a best result in Negative (0.52) and Positive (0.8). For Neutral byT5 has a little better than ViT5 (0.71 with 0.7 for ViT5). With Sentiment has large quantity, models have a good result from 0.66 to 0.8 but

	ViT5	BARTpho	byT5
Negative	0.52	0.45	0.50
Positive	0.80	0.67	0.79
Neutral	0.70	0.66	0.71

Table 5: Sentiment Classification F1-macro

with Negative which has small quantity the model has quite bad result, this seem because the quantity of Negative is not enough for models to learn.

5.4 Error analysis

With bad result in prediction quad but good result in Aspect Category Detection and Sentiment Classification seem like models has a big problem in extract opinion term. This because the model can extract all of opinion term, some has more words, some has less words and some have a exactly Aspect Category, Aspect Term, Sentiment Polarity but wrong about Opinion Term. Besite that there any some types of errors: cannot detect the aspect/opinion term, misclassify the sentiment, cannot detect the aspect category, and detect the wrong aspect category, see Table [6].

With the first comment of Table [6], model was predict wrong aspect from 'Show' to 'Others', this make other elements of the quad turn wrong because we have stipulated if aspect is 'Others' others elements must be NULL and neutral. For the second comment, model was misclassify the sentiment, with the opinion term is 'quá tệ và nhạt' the sentiment should be negative but model has predict it is a positive. The third comment, it has a entity is 'Chị kiều' but model can not detect it in to Aspect Term. With the fourth comment, model detect wrong opinion term 'đang rap cái gì' instead of 'Không hiểu'.

6 Spark Streaming

Spark Streaming is a real-time data processing framework within the Apache Spark ecosystem, designed for seamless integration with batch processing. It enables high-throughput, fault-tolerant stream processing, allowing users to process live data streams in micro-batches.

Spark Streaming can provide fault recovery and scalable, parallel processing. With its flexibility and compatibility with various data sources, Spark Streaming is a powerful tool for handling continuous data streams in a distributed and resilient manner.

Our system is divided into two parts: real-time data collecting and real-time analysis. Its architecture is presented in the figure below:

Comments	Predictions	True Labels
Chung kết king of the rap chất hơn rap viet	Others NULL neutral NULL	Show Chung kết king of the rap positive Chất
Chương trình sẽ hay hơn khi để một MC khác dẫn chương trình quá tệ và nhạt	Character - Overall MC positive quá tệ và nhạt	Character - Overall MC negative quá tệ và nhạt
Chị kiều keo ghê	Character - Overall NULL positive keo ghê	Character - Overall Chị kiều positive keo ghê
Không hiểu bà Lona đang rap cái gì luôn ấy được vào vòng trong t cũng chịu	Character - Overall Lona negative đang rap cái gì	Character - Overall Lona negative Không hiểu

Table 6: Some incorrect prediction examples of ViT5

6.1 Real-time Data Collecting

For the first part, we're tapping into the YouTube API to gather data. This valuable information is then sent off to a Kafka producer, creating a smooth flow for data streaming.

Once in Kafka, our consumer component steps in to grab the data. This data journey continues as it makes its way into Spark Streaming DataFrames. This setup allows us to process the data in small, continuous chunks, enabling us to keep an eye on sentiment changes in real time.

6.2 Real-time Analysis

In this part, we receive the data from the TCP socket and apply the data pre-processing layer that we mentioned before. Result of this process is presented in Figure [7]. Then, we deploy the ViT5 pre-trained model which is trained offline. After getting the results, we categorize them into for each Vietnamese Rap contestant from aspect terms that are predicted, we create statistical charts and deploy a website application using Streamlit based on those aspect categories

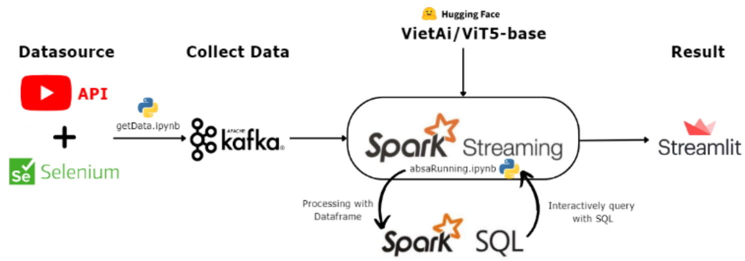

Fig. 6: The end-to-end system's architecture using Spark Streaming

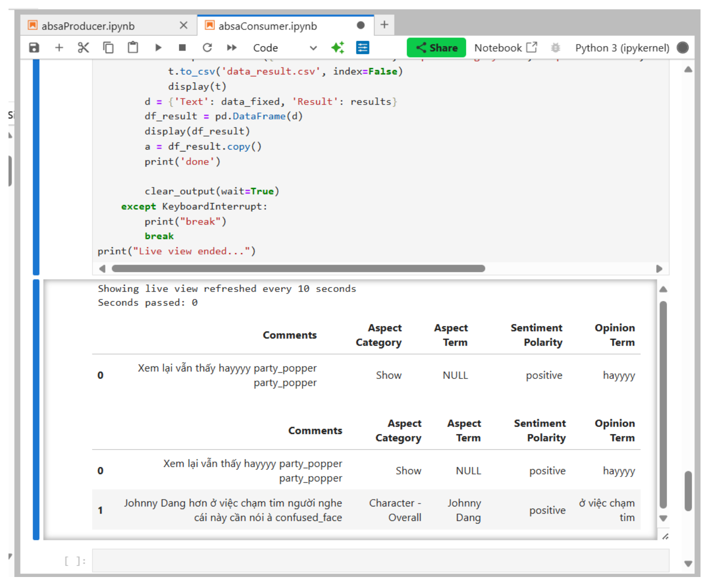

Fig. 7: Consumer results

shown in Figure [8]. This website will help users catch up with the overall statistics of Rap Viet candidates by knowing their aspect category and sentiment polarity that are commented on livesream YouTube videos by audiences.

7 Conclusion

In this paper, we have presented an enhanced dataset based on the original Rap Viet dataset, combining data from Vietnam's two most popular game shows. This expanded and more comprehensive dataset is designed for aspect-based sentiment analysis (ABSA), targeting quad predictions: aspect category, aspect term, opinion term, and sentiment polarity. We applied the Paraphrase Generation method to predict quads, incorporating all four elements simultaneously.

Fig. 8: Real time candidate statistics Streamlit demo

Our deployment of three models ViT5, BARTPho and byT5 to our dataset and received an acceptable result..

Looking ahead, we plan to continue expanding our dataset to ensure greater balance and inclusivity, extending our focus to additional shows and potentially other fields such as education and social media influencers (KOLs). Our ultimate goal is to develop a system capable of automatically collecting, preprocessing, analyzing comments, and visualizing results. Such a system would be immensely beneficial, providing producers with insights to improve their shows and giving viewers a comprehensive overview of contestants. Additionally, we aim to further research Aspect Sentiment Quad Prediction (ASQP), explore multimodal ABSA, and experiment with models like LLAMA to create a robust Social Listening system.

References

1. Do, H.H., Prasad, P.W., Maag, A., Alsadoon, A.: Deep learning for aspect-based sentiment analysis: a comparative review. Expert systems with applications **118**, 272–299 (2019)

2. Huang, L., Wang, P., Li, S., Liu, T., Zhang, X., Cheng, Z., Yin, D., Wang, H.: First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction. arXiv preprint arXiv:2102.08549 (2021)
3. Luc Phan, L., Huynh Pham, P., Thi-Thanh Nguyen, K., Khai Huynh, S., Thi Nguyen, T., Thanh Nguyen, L., Van Huynh, T., Van Nguyen, K.: Sa2sl: From aspect-based sentiment analysis to social listening system for business intelligence. In: Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14. pp. 647–658. Springer (2021)
4. Mai, L., Le, H.B.: Aspect-based sentiment analysis of vietnamese texts with deep learning. In: Asian Conference on Intelligent Information and Database Systems (2018), `https://api.semanticscholar.org/CorpusID:3750939`
5. Nazir, A., Rao, Y., Wu, L., Sun, L.: Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing **13**(2), 845–863 (2022). `https://doi.org/10.1109/TAFFC.2020.2970399`
6. Nguyen, H.T., Nguyen, H.V., Ngo, Q.T., Vu, L.X., Tran, V.M., Ngo, B.X., Le, C.A.: Vlsp shared task: sentiment analysis. Journal of Computer Science and Cybernetics **34**(4), 295–310 (2018)
7. Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., Si, L.: Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 8600–8607 (2020)
8. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S.M., Eryiğit, G.: SemEval-2016 task 5: Aspect based sentiment analysis. In: Bethard, S., Carpuat, M., Cer, D., Jurgens, D., Nakov, P., Zesch, T. (eds.) Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). pp. 19–30. Association for Computational Linguistics, San Diego, California (Jun 2016). `https://doi.org/10.18653/v1/S16-1002`, `https://aclanthology.org/S16-1002`
9. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: SemEval-2015 task 12: Aspect based sentiment analysis. In: Nakov, P., Zesch, T., Cer, D., Jurgens, D. (eds.) Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 486–495. Association for Computational Linguistics, Denver, Colorado (Jun 2015). `https://doi.org/10.18653/v1/S15-2082`, `https://aclanthology.org/S15-2082`
10. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: SemEval-2014 task 4: Aspect based sentiment analysis. In: Nakov, P., Zesch, T. (eds.) Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 27–35. Association for Computational Linguistics, Dublin, Ireland (Aug 2014). `https://doi.org/10.3115/v1/S14-2004`, `https://aclanthology.org/S14-2004`
11. Thin, D., Nguyen, N.: Aspect-category based sentiment analysis with unified sequence-to-sequence transfer transformers. VNU Journal of Science: Computer Science and Communication Engineering **39**(2) (2023). `https://doi.org/10.25073/2588-1086/vnucsce.662`, `//jcsce.vnu.edu.vn/index.php/jcsce/article/view/662`
12. Tran, Q.L., Le, P.T.D., Do, T.H.: Aspect-based sentiment analysis for Vietnamese reviews about beauty product on E-commerce websites. In: Dita, S., Trillanes, A., Lucas, R.I. (eds.) Proceedings of the 36th Pacific Asia Conference on Language, In-

formation and Computation. pp. 767–776. Association for Computational Linguistics, Manila, Philippines (Oct 2022), https://aclanthology.org/2022.paclic-1.84
13. Van Thin, D., Nguyen, N.L.T., Truong, T.M., Le, L.S., Vo, D.T.: Two new large corpora for vietnamese aspect-based sentiment analysis at sentence level. ACM Trans. Asian Low-Resour. Lang. Inf. Process. **20**(4) (may 2021). https://doi.org/10.1145/3446678, https://doi.org/10.1145/3446678
14. Xu, L., Li, H., Lu, W., Bing, L.: Position-aware tagging for aspect sentiment triplet extraction. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 2339–2349. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.183, https://aclanthology.org/2020.emnlp-main.183
15. Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., Lam, W.: Aspect sentiment quad prediction as paraphrase generation. arXiv preprint arXiv:2110.00796 (2021)