ChattyTicket: Classifying Emotion and Toxicity in Valorant Chats using Multi-Task Learning with Bi-LSTM Algorithm and BERT

Johndel N. Encabo
College of Science and Computer Studies
De La Salle University - Dasmariñas
EJN2008@dlsud.edu.ph

Maria Cassandra B. Vitug
College of Science and Computer Studies
De La Salle University - Dasmariñas
VMB0604@dlsud.edu.ph

Tita R. Herradura
College of Science and Computer Studies
De La Salle University - Dasmariñas
trherradura@dlsud.edu.ph

ABSTRACT

This study investigates communication dynamics within the popular multiplayer game Valorant, particularly focusing on the challenges posed by toxic conversations despite its potential for fostering connections, especially during the pandemic when entertainment serves as a form of leisure. To address this issue, we propose a novel multi-task learning architecture that integrates Bi-LSTM (Bidirectional Long-Short Term Memory network) and BERT (Bidirectional Encoder Representations from Transformers) models. This architecture was chosen for its superior performance classification problem and a BERT pre-trained model which can provide additional features as a backbone of the model.

Experimentally, our classifier achieves an accuracy of 91.81 percent in toxicity prediction and 86.74 percent in emotion prediction, shedding light on prevalent emotions such as anger and instances of cyberbullying. The insights garnered from these results hold the potential to cultivate healthier gaming communities. We introduce the ChattyTicket API for chat text evaluation, alongside a web application that has garnered positive user feedback, featuring its usability and interactivity. Continuous revisions guided by constructive criticism, are aimed at enhancing the platform and improving user experience. These revisions address challenges such as discerning sarcasm and fostering a more positive gaming environment.

KEYWORDS

Toxicity, emotion, Bi-LSTM, BERT backbone, toxicity classification, emotion classification, multi-task learning architecture, deep learning, game chat analysis

1 INTRODUCTION

Amidst the pandemic, individuals sought solace in their homes, turning to various forms of entertainment, including video games, which became a significant leisure activity, especially among the youth. In a time marked by social isolation, multiplayer games like Valorant emerged as a refuge, offering a platform for forging connections, shared experiences, and camaraderie. The interactive nature and immersive storytelling of these games not only entertain but also serve as therapeutic outlets, contributing to their widespread appeal as means of social interaction during the pandemic [11, 12, 19].

Valorant, a prominent first-person shooter game, gained traction during this period, emphasizing the crucial role of effective communication in its team-based gameplay. While features like text and voice chat enrich the gaming experience by facilitating strategic collaboration, the competitive and anonymous nature of these platforms often leads to toxic interactions among players. This toxicity, characterized by verbal abuse, harassment, and bullying, has been pervasive, affecting not only the gaming experience but also the mental well-being of players [1, 10, 14, 22, 26].

Communication in chat, whether through text or voice, often carries with it a range of emotions. In competitive games like Valorant, players frequently express their reactions to their teammates, which can manifest as anger, sadness, happiness, disgust, or fear, depending on the circumstances. These reactions can significantly impact the outcome of the game, either positively or negatively as they can influence the emotional state of others. As Davidson et al (1994) [6] noted, universal emotions are characterized by distinct signals, physiologies, and timelines, with variations in onset, duration, and decline. Typically, emotions do not endure beyond an hour, and if they persist for an extended period without interruption, they are more likely to be classified as a mood [6]. This understanding allows researchers to examine the emotional content of text and its effect on gameplay, providing insights into player experience.

The prevalence of gaming toxicity has spurred researchers to explore deep learning algorithms, such as Bi-LSTM, for analyzing the emotional and toxic content of in-game chat. Studies have shown promising results, with algorithms like Bi-LSTM demonstrating high accuracy in detecting abusive or toxic language having a high classification accuracy compared to other existing cyberbullying detection algorithms. Bi-LSTM is to detect abusive or toxic content in any messaging application. [18] By employing multi-task learning architectures, researchers aim to further enhance performance and provide developers and community managers with effective tools to identify and mitigate toxic behavior within online gaming communities.

This study employs the Bi-LSTM deep learning algorithm to classify emotion and toxicity in Valorant text chat, aiming to enhance the gaming experience for all players. The research focuses on developing web applications to gather data, aiming to improve the online gaming experience by identifying emotions and tackling toxic behavior in text chat. It introduces ChattyTicket, a system that utilizes a Bi-LSTM algorithm for classifying emotion and toxicity in Valorant chats, aiming to improve the gaming experience.

The research targets players within the South-East Asia server, the desired population sample comprises one thousand Valorant players from this region. The dataset gathered only comprises of Valorant text chat logs, it excludes communication forms beyond text chat like voice chat and external messaging platforms.

The study's findings are anticipated to benefit the Valorant community by offering insights into common toxic language and prevalent emotions during gameplay. Additionally, developers can utilize the data to refine in-game censorship mechanisms and possibly initiate warning and banning systems to filter out the players who are toxic. Future researchers can explore alternative algorithms for improved accuracy in toxicity detection and emotion classification. Leveraging the Bi-LSTM algorithm, developers can create practical applications for monitoring and managing in-game communication, enhancing player experience, and promoting a healthier gaming community.

2 RELATED WORKS

Online gaming has witnessed a surge in popularity during the pandemic, facilitating interactions and shared experiences among millions worldwide. However, this rise has coincided with an increase in toxic behavior, including harassment and hate speech, posing challenges to players' mental well-being [4]. Shen et al. (2020) suggest that toxic behavior in online games is often rationalized and perpetuated by players, with exposure in prior games increasing the likelihood of future toxic acts, particularly among experienced players [25]. Tyler (2020) notes that certain competitive games like Counter-Strike: Global Offensive and League of Legends are particularly prone to toxicity, with varying degrees of enforcement and consequences for toxic behavior. To address this issue, there is a need for effective toxicity detection models, such as the one proposed by the researchers in this study [10, 20, 21].

Asgher et al. (2022) highlight the challenge of accurately analyzing text emotions and propose a Deep Learning approach, specifically Bi-LSTM, for improved emotion detection. Their study focuses on enhancing emotion classification accuracy, albeit with limitations regarding word embedding and language scope [2]. Lee et al. (2023) suggests the use of transformer transfer learning for emotion annotation, offering a faster alternative to manual annotation, albeit with potential limitations in capturing subtle social emotions [11]. Meanwhile, the researchers in this study adopt a Modified Discrete Emotions model, adding a "neutral" category to Dr. Ekman's well-established model, which identifies six basic emotions [6]. This model facilitates the categorization of sentiments and enhances understanding of associated emotions, thereby providing insights into the emotional states expressed in textual data."

Toxicity in online gaming can manifest in various forms, including verbal toxicity, grieving behavior, and cheating. Dzigurski (2022) further delineates sub-themes within these categories, underscoring the multifaceted nature of toxic behavior in gaming environments in his study of Toxicity in the game World of Tanks: A participant observation ethnography, thematic analysis, content analysis and autoethnography (Dissertation), this category can be seen on Table 1 [8]. Understanding these nuances is crucial for developing effective strategies to mitigate toxicity and foster a positive gaming experience for all users [1, 8, 14]. Table 1 comprises themes and sub-themes representing toxic behaviors observed in the game of World of Tanks. These classifications were from battle chats and were tallied across 120 games, totaling 383 occurrences across all classified chats [8].

Table 1: Occurrence of themes/sub-themes in battle chats

No	Theme/sub-theme	Occurrence	%	No. of battles	%
1	Gamesplaining	92	24	37	31
2	Ableism	89	23	53	44
3	Male preserve	60	15.7	36	30
4	Sarcasm	53	14	29	24
5	Positive	27	7	16	13.3
6	Blaming others	18	4.7	12	10
7	RNG complaints	13	3.4	4	3.3
8	Sexism	10	2.6	10	8.3
9	SPG complaints	5	1.3	2	1.7
10	MM complaints	5	1.3	5	4
11	Ageism	4	1	3	2.5
12	EBR complaints	3	0.8	1	0.8
13	Cyberbullying	/	/	3	2.5
14	Game complaints	2	0.5	2	1.7
15	Racism	1	0.26	1	0.8
16	Map Complaints	1	0.26	1	0.8

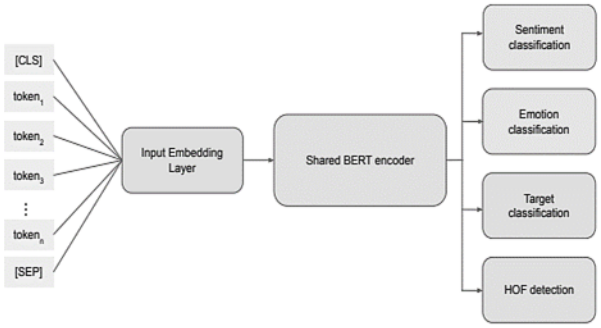

Figure 1: Vanilla Tree-like Architecture in with Sentiment, Emotion, Target Detection

Multi-Task Learning (MTL) has emerged as a powerful approach in machine learning, enabling the simultaneous training of multiple tasks to capture generalized and complementary knowledge from specific tasks [3]. Huang et al. (2022) used MTL in their study on abuse and emotion classification, utilizing linguistic context and pretrained language models like BERT to enhance the algorithm's representational capability. They employed different decoders and a cross-attention component, achieving superior performance compared to other methods [13].

Researchers often explore various architectures of MTL to optimize model performance. The focus of this study lies in Parallel Architectures, particularly the Vanilla Tree-like Architecture and its variants, which allow tasks to run in parallel, sharing certain layers for efficiency [3]. Plaza-del-Arco et al. (2021) applied Tree-Like Architecture with a Shared Bert Encoder in their study on hate speech and offensive language recognition, leveraging transfer learning to enhance model performance [7], their model architecture shown on the Figure 1.

Moreover, Parallel Feature-fusion is another MTL architecture that actively combines features from different tasks to create task-specific representations [3]. While Supervision at Different Feature Levels also shows promise, its applicability to this study may require further exploration. Overall, understanding and implementing different MTL architectures offers researchers opportunities to enhance model efficiency and performance across various NLP tasks.

Cruz and Cheng's (2022) research focus on advancing natural language processing technology for Filipino speakers, with the aim of developing resources and models for the Filipino language to enhance access to crucial services. By conducting a literature review, they identify areas for improvement and suggest further research directions, laying the groundwork for enhancing technology accessibility among Filipino-speaking communities [5]. Similarly, Molina et al. (2021) contribute to this goal by creating Tagalog language models using multi-source data and the BERT pre-training technique, addressing challenges in developing language models for resource-constrained languages [16].

Alampay et al. (2020) delve into the connection between cyberbullying/cybervictimization and empathy among adolescent Filipinos, underscoring the importance of understanding this relationship for intervention and prevention programs [24]. Pujante (2021) investigates the use and impacts of "trash talk" in gaming environments, shedding light on social dynamics, and emphasizing the need for promoting a positive gaming culture [17]. Ferrer et al. (2021) tackle the identification of profanity in the Filipino language, aiming to develop efficient methods for monitoring and controlling online communication to foster a civil and safe online environment [9].

Ong's (2022) project, together with collaborators, focuses on building a context-aware digital lexicon for Tagalog and English to address challenges in lacking local resources and continuously updated datasets [21]. Sagum et al. (2019) contributes to this effort by creating a Filipino WordNet using semi-supervised learning, further enriching language resources for Tagalog. These studies collectively contribute to advancing natural language processing technology for Filipino languages and fostering a safer and more inclusive online environment for Filipino speakers [23].

3 METHODOLOGIES

3.1 Equipment

This study utilized the following equipment: (1) personal computers and (2) internet. The website served as the platform for generating output, using Django to create the API connecting to the model deployed on an AWS server. Next.js was utilized to develop the interface featuring an input box for entering words/sentences, deployed on a serverless instance of Vercel.

3.2 Data Collection

The aim of data collection was to acquire authentic text chat data from the online game Valorant. Convenience sampling was employed, distributing the survey to public Facebook groups, Discord servers, group chats, and other online communities frequented by Valorant players aged 18 and above.

There was a total of 789 participants, all were asked to complete a survey questionnaire detailing their Valorant gameplay experience, including interactions within the in-game chat. This information could be submitted by typing text into the provided input box or by attaching screenshots of their game chat box. Only in-game chat data was collected, excluding any other personal information like their full name. The survey questionnaire was integrated into a website developed by the researchers. Google Forms was also used as an alternative way to collect data on gameplay experience and text chat screenshots. Demographic information such as gender, country, age, and in-game details such as rank were also collected for further analysis. The dataset gathered consists of 4 columns: username, text/chat, toxicity label and emotion label.

A post-evaluation survey, distributed through Microsoft Forms, was sent to individuals who utilized the web application, inviting them to share feedback on their experience with the website. This feedback serves to guide future improvements to the application. To protect sensitive information like unfiltered datasets containing usernames, researchers pledged to delete screenshots and archive the dataset post the removal of any identifiable participant details. This archival and deletion process was implemented after the utilization of data for training and result acquisition.

3.3 Data Processing

The course of the system's data processing started in gathering the data from the survey answered by the participants. The screenshots of the in-game chat box were included, then extracted using OCR and manual extraction. Take note that OCR was used just for extraction and not included in the model made, and it does not affect the performance of it. We excluded the number of games played due to the absence of information given by the participants.

In the context of emotion classification for Valorant chats, Ekman's renowned model was applied, which delineates six primary emotions universally expressed and recognized across cultures: happiness, sadness, anger, fear, surprise, and disgust [6]. In this paper, a modified discrete emotion model was used, introducing an additional category, "neutral," to account for instances where game chats exhibit a lack of strong emotional intensity or are used to express information to teammates.

The proponents utilized Dzigurski's comprehensive classification framework from 2022, which encompasses various toxic behaviors prevalent in Valorant chats, as referenced in Table 1. To simplify the categorization process and enhance the coherence of the analysis, similar forms of toxicity were grouped as seen on Table 2. Ableism, Male Preserve, Ageism, Racism, and Sexism were labeled "Multiple Discrimination". Toxicity related to in-game experiences, including RNG Complaints, MM Complaints, EBR Complaints, Map Complaints, and Game complaints, was placed within the category of "Gameplay Experience Complaints". Lastly, the "Gamesplaining" category was created by merging instances from both Gamesplaining and Blaming Others [1, 14]. The proponentsomitted the other toxicity classifications as they were not relevant to the game under study.

3.3.1 *Data Labelling Validation.*

To ensure the accuracy and reliability of emotion classification, the proponents engaged the expertise of a board-certified psychometrician. Collaborative consultations were conducted with the psychometrician for suiTable labels

Table 2: Game Toxicity Classifications

Toxicity	Combined Classifications
Sexism **Ableism** **Male Preserve** **Ageism** **Racism**	Multiple Discrimination
RNG complaints **Map complaints** **MM complaints** **EBR complaints** **Game complaints**	Gameplay Experience Complaints
Sarcasm	Sarcasm
Cyberbullying	Cyberbullying
Blaming others	Blaming others
Gamesplaining	Gamesplaining
Not Toxic	Not Toxic

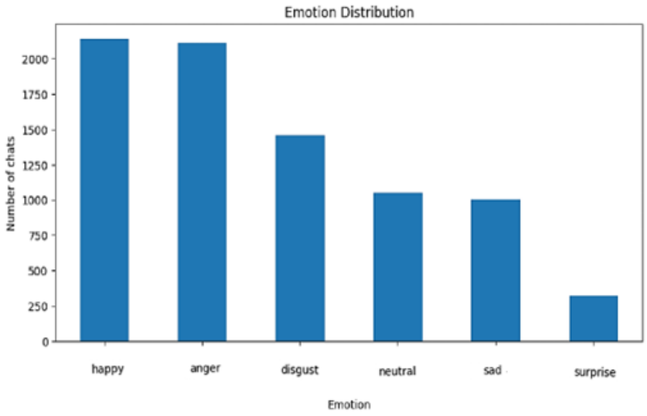

Figure 2: Emotion Distribution After Data Labelling

for the dataset, specifically utilizing Ekman's six basic emotions and a neutral category.

The labeled dataset then underwent rigorous validation process in subsequent sessions with the psychometrician. The psychometrician provided official endorsement by signing the validation form and issuing a certificate upon successful validation. This action affirmed the robustness of the emotion classification framework employed in this study. After labelling the emotions and toxicity, uneven distribution of them was discovered as shown on Figure 2. This indicates augmentation of the dataset was required.

3.3.2 Data Augmentation. To address an unbalanced dataset characterized by a low number of categories in emotion, proponents employed oversampling techniques. This process aimed to augment the representation of underrepresented categories, as illustrated in Figure 2 of the emotion classification results. Oversampling was applied to the toxicity class to achieve balance between the two classes. Subsequently, undersampling was conducted on the toxicity class to mitigate the disproportionately high number of non-toxic instances, as depicted in Figure 3. These efforts yielded a dataset comprising 8,274 rows, up from the original 8,086.

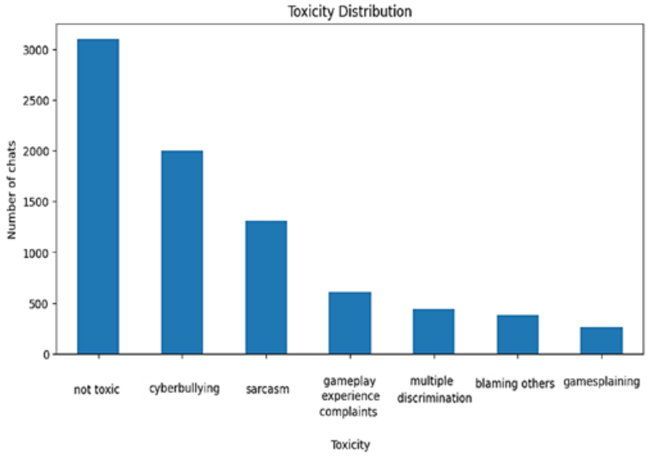

Figure 3: Toxicity Distribution After Data Labelling

3.3.3 Data Preparation. Preparing the chat data was the next step, which utilized some of the NLP techniques including the removal of duplicates and removing some of special characters that might unrelated such as underscores, quotation marks, etc., however question mark and exclamation point was not removed due to its context in the game like it can be a disbelief in the in-game chat.

After the cleaning and augmenting, the dependent variables such as the emotions and toxicities undergone one-hot encoded to make categorical data possible to work with in the model. This one-hot encoded makes the categories into a finite set of label values, this consists of 1 and 0s, the position of 1 in the set determines the category. Using the Scikit-learn library for one-hot encoding facilitated the process, providing the flexibility to transform the encoded data back to its categorical values when needed. The text inputs were prepared for the model using the BERT Tokenizer. The completed processed dataset was split into two parts: 80% for training and 20% for testing.

3.4 Model Architecture

The model consists of 4 major layers, first was first input layer which accepts the tokenize version of the chat data more specific is the input ids, then the label data in a one-hot encoded version. The second layer was the pre-trained BERT Layer, which is used for feature extraction which gives additional feature as the backbone of the model, receives data from input layer then produces a pooler output that is used in the next layer, the Classifier layer. This was split into two tasks connected by a trunk, the toxicity classification and emotion classification, using Bi-LSTM as the classifier for each task, this helps to retain information without duplicating the context using its memory [15].

Lastly, the output layer or the dense layer that accepts an input from the Bi-LSTM layers because there's two task two output layer was made, first was for emotion that have a unit of 6, second was for toxicity which have a unit of 7. Both output layers have an activation

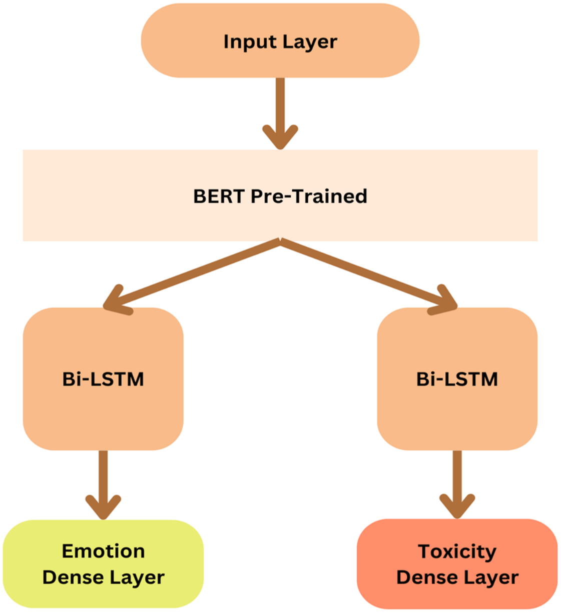

Figure 4: Multitask Classification Model

function SoftMax. The optimizer used was Adam with a learning rate of 1e-4 with a learning decay of 0.01. Then a categorical cross entropy loss function for both toxicity and emotion were used as par multiple categories was used. The metrics used were discussed in Section 3.6.

3.4.1 *Regularization Techniques.*

To prevent overfitting, the model employed dropout regularization with a rate of 0.6 up to 0.7. Additionally, task specific L1 and L2 regularization were applied. For the Emotion task, L1 and L2 coefficients were 0.0045 and 0.03 respectively, while for the Toxicity task, they were 0.006 and 0.03. These regularization techniques were calibrated manually during experimentation to optimize performance.

3.4.2 *Hyperparameters Tuning .*

The following hyperparameters included: epoch which start from 30 until 20, then batch size from 32 to 70, also number of LSTM layers that have a value of 40 down from 50 after calibrations. Learning decay also applied which discussed on Section 3.4. The hyperparameter tuning done was a manual trial and error to find better results in the model. It was inefficient, due to time constraint resorting to this method has been chosen.

3.5 Model Implementation

The model was implemented using the python libraries such as keras for deep learning packages based on Tensorflow for the pre-built layers for the model like the Bi-LSTM, Input and Dense Layer, also the metrics is already built in because of Tensorflow, then Scikit-learn for preprocessing and metrics computation of the data like the one-hot encoder for the emotion and toxicity and the confusion matrix. For the pre-trained BERT, transformers library from HuggingFace was used, also includes the BERT tokenizer used for the tokenization of the text. Two varieties of BERT were used: BERT-base-uncased is for general use and BERT-base-multilingual-cased for multi-language that can provide features from different languages.

3.6 Model Evaluation and Validation

In the model evaluation and validation phase, our primary objective is to ensure the accuracy and reliability of our trained model. Using manual analysis using metrics was used during experimentation, when found one of the metrics was low or indicate a possible overfitting then adjusting of the hyperparameter was done, from dropout, L1 and L2 regularization, batch size and number of LSTM layers. This process somehow helps mitigate overfitting and provides a robust assessment of the model's generalization ability. To quantify the model's performance, key metrics were used including precision, recall, F1 score, and accuracy. These metrics collectively offer insights into the model's predictive capabilities, balancing between correctness and completeness of predictions across the different classes.

Moreover, visual aid such as confusion matrices to gain a deeper understanding of the model's performance characteristics, identifying potential areas for improvement and assessing it discriminative power effectively. Through these comprehensive evaluation strategies, the aim is to ensure that our model not only learns accurately but also generalizes well to unseen data, thereby enhancing its utility and reliability in practical applications. After getting these results, another model was created with different BERT, a BERT with multilanguage capability.

4 RESULTS AND DISCUSSION

4.1 Model Performance

After conducting several experiments, the proponents compared the performance of two models using different BERT variants, this was shown on Table 3. The BERT Base model achieved an accuracy of 86.74%, a recall of 82.23%, and a precision of 91.17%. Its AUC is 98.85%, and the F1-Score is 63.50%. The BERT Multilingual model showed slightly better performance, with an accuracy of 87.70%, a recall of 84%, and a precision of 91.23%. Its AUC is 98.67%, and the F1-Score is 60.46%.

The model training several hyperparameters were used, but the best hyperparameters that were found, also been used in the model were 50 epochs, a batch size of 70, learning rate in Adam optimizer of 0.0001, then a LSTM (Long Short-Term Memory) layer of 0.75, also an L1 and L2 regularization was applied for both emotion and toxicity tasks with a value of 0.0045 and 0.03.

The utilization of Bi-LSTM as the classifier, supported by BERT from Transformers, demonstrates strong performance in identifying both toxicity and emotion especially the BERT Multilingual. However, toxicity exhibits superior performance compared to emotion. This is attributed to certain words or sentences lacking sufficient context to accurately determine the conveyed emotion; for instance, "GG" may signify happiness or sadness depending on the in-game situation. The performance disparity is evident in the confusion matrices depicted in Figures 3 and 4, which illustrate the classifier's predictive capabilities across various categories. Furthermore, the multitask architecture efficiently deployed yields dual outputs, enhancing overall output efficiency.

Table 3: Model's Performance Score

Model BERT	Task	Accuracy %	Recall%	Precision%	AUC%	F1-Score
With Base	**Toxicity**	**91.81**	**90.38**	**92.90**	**99.6**	**81.28**
	Emotion	**86.74**	**82.23**	**91.17**	**98.85**	**62.83**
With Multilingual	**Toxicity**	**93.88**	**92.85**	**94.77**	**99.63**	**77.96**
	Emotion	**87.70**	**84**	**91.23**	**98.67**	**60.46**

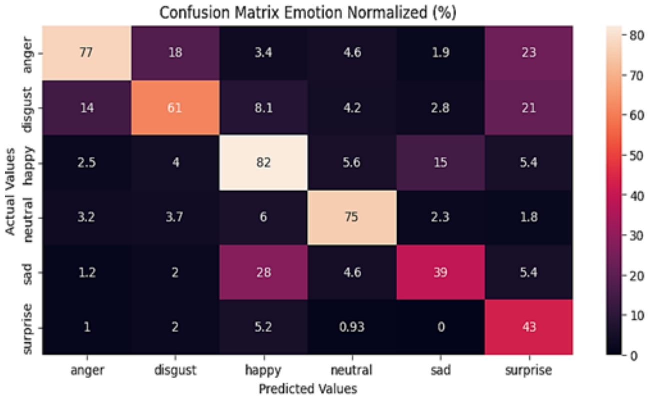

Figure 5: Confusion Matrix Emotion Normalized

Two confusion matrices were created to demonstrate the model with BERT Multilingual has capability in determining the toxicity and emotion of the chat in each category. These matrices were normalized for clarity and ease of interpretation. Figure 5 illustrates the normalized confusion matrix for emotion classification, depicted as a heatmap, where brighter colors denote higher values and darker colors indicate lower values. Figure 5 visually compares actual predictions with predicted values for each emotion classification, revealing successful identification rates. Specifically, the accuracy rates are 77% for anger, 81% for disgust, 82% for happy, 75% for neutral, 39% for sad, and 43% for surprise. However, 28% of instances where actual sadness is present are misclassified as happy due to the ambiguity of certain expressions like "GG", which could signify happiness or sadness depending on the context.

Figure 6 illustrates the confusion matrix for toxicity classification, presented as a heatmap. Brighter colors signify higher values, while darker colors represent lower values. The toxicity classification demonstrates effective prediction of actual values, with accuracy rates of 81% for blaming others, 88% for cyberbullying, 78% for gameplay experience complaints, 91% for gamesplaining, 82% for multiple discrimination, 60% for sarcasm, and 89% for not toxic. However, it is worth noting that there is a 33% chance of misclassifying actual non-toxic instances as sarcasm, likely due to the nuanced nature of sarcasm detection and the contextual complexity involved in distinguishing between genuine non-toxic communication and sarcastic remarks. Overall, the model performs well, achieving a high accuracy rate of 81.28% in predicting the actual classification.

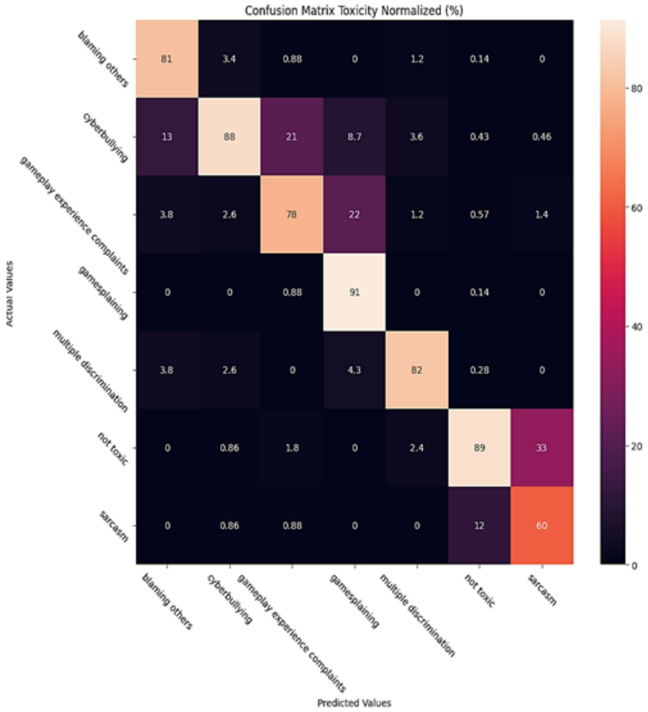

Figure 6: Confusion Matrix Toxicity Normalized

4.2 Evaluation Discussion

ChattyTicket's evaluation reveals diverse user preferences and positive reception for its functionality, usability, performance, interactivity, and UI design. While most users are satisfied, valuable feedback suggests enhancements for future versions, including improved analytics display, more visual content, additional emotions in the model, and enhanced accuracy.

5 CONCLUSION AND FUTURE WORKS

5.1 Conclusion

In summary, the multi-task learning architecture, which integrates Bi-LSTM classifiers for toxicity and emotion with a BERT backbone, demonstrates significant effectiveness in predicting multiple classes for both tasks. The model achieved an accuracy of 93.88% in classifying toxicity and 87.7% in classifying emotion, highlighting the refined performance of the model with bert base multilingual. Despite accuracy improvements, challenges persist, particularly in discerning sarcasm within the 'not toxic' category due to contextual nuances.

Moreover, the development of an accessible API has facilitated the evaluation of chat text for both emotion and toxicity, showcasing accurate and efficient performance. The examination of datasets underscores the prevalence of toxic chats in gaming environments like Valorant, encompassing cyberbullying, sarcasm, and discrimination. The evaluation of the web application revealed diverse user preferences and garnered high praise for its utility, usability, performance, and interactivity. Constructive criticism has guided subsequent revisions, ensuring continuous enhancement to meet customer expectations and elevate the overall user experience.

In conclusion, the multi-task learning architecture, in conjunction with an accessible API, demonstrates promising results in predicting toxicity and emotion in chat text. The exploration of diverse datasets and user feedback has provided valuable insights into the prevalence of toxic behaviour in gaming environments and the effectiveness of the web application in meeting user needs. Constructive criticism has been instrumental in refining the platform, ensuring its ongoing alignment with user expectations, and enhancing the overall user experience.

5.2 Future Works

Future research should aim to focus on observing players during Valorant gameplay to provide more accurate labeling and contextualization of emotions, either through in-game observation or post-game interviews. To achieve this, researchers are advised to allocate at least a month or longer for data gathering, given the typical game duration of around 45 minutes. Extending the data collection period enables the accumulation of a more comprehensive dataset. Furthermore, researchers should consider incorporating additional classifiers such as Naïve Bayes or Logistic Regression to compare with Bi-LSTM performance and explore alternative architectures for multi-task learning beyond the Vanilla Tree-Like Structure to assess potential performance improvements.

REFERENCES

[1] Jenna Abrera, Maria Laureta, Kent Miculob, Fatima Valencia, Justin David Pineda, Jayvee Cabardo, and Lorena Rabago. 2019. Harassment Exposure Model Using Sentiment Analysis on Facebook Pages. In *2019 International Conference on Information Management and Technology*. 175–179.

[2] Muhammad Zubair Asghar, Adidah Lajis, Muhammad Mansoor Alam, Mohd Khairil Rahmat, Haidawati Mohamad Nasir, Hussain Ahmad, Mabrook S. Al-Rakhami, Atif Al-Amri, and Fahad R. Albogamy. 2022. A Deep Neural Network Model for the Detection and Classification of Emotions from Textual Content. *Complexity* 2022, 1 (2022), 8221121. https://doi.org/10.1155/2022/8221121 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1155/2022/8221121

[3] Shijie Chen, Yu Zhang, and Qiang Yang. 2024. Multi-Task Learning in Natural Language Processing: An Overview. arXiv:2109.09138 [cs.AI] https://arxiv.org/abs/2109.09138

[4] Debating Communities and Networks XI. 2020. "Shat up u noob" – The primary causes of toxicity in online gaming communities. https://networkconference.netstudies.org/2020Curtin/2020/05/13/shat-up-u-noob-the-primary-causes-of-toxicity-in-online-gaming-communities/

[5] Jan Christian Blaise Cruz and Charibeth Cheng. 2021. Improving Large-scale Language Models and Resources for Filipino. arXiv:2111.06053 [cs.CL] https://arxiv.org/abs/2111.06053

[6] Richard Davidson, Paul Ekman, Nico Frijda, Harold Goldsmith, Jerome Kagan, Richard Lazarus, Jaak Panksepp, David Watson, and Lee Clark. 1994. How are emotions distinguished from moods, temperament, and other related affective constructs? *The nature of emotion: Fundamental questions. Series in affective science* (01 1994).

[7] Flor Miriam Plaza del Arco, Sercan Halat, Sebastian Padó, and Roman Klinger. 2022. Multi-Task Learning with Sentiment, Emotion, and Target Detection to Recognize Hate Speech and Offensive Language. arXiv:2109.10255 [cs.CL] https://arxiv.org/abs/2109.10255

[8] Sasa Dzigurski. 2022. *Toxicity in the game World of Tanks: A participant observation ethnography, thematic analysis, content analysis and autoethnography.* Master's thesis. Uppsala University, Department of Game Design.

[9] Beverly P. Ferrer, Crystelle T. Tomilas, Lorenz P. Mallare, Benedict D. Pineda. Jr., Ana F. De Guzman, Famela N. Siapno, Kathleen B. Payang, Michael B. Yu, Allen D. Bolaños, and Zareena D. Lee. 2021. A machine learning model for the profanity detection in the Filipino language. *Journal of Engineering Science and Technology Special* special issue (10 2021), 37–46.

[10] Enrico Gandolfi and Richard Ferdig. 2021. Sharing dark sides on game service platforms: Disruptive behaviors and toxicity in DOTA2 through a platform lens. *Convergence: The International Journal of Research into New Media Technologies* 28 (07 2021), 135485652110288. https://doi.org/10.1177/13548565211028809

[11] Vicpher Garnada. 2020. ONLINE GAMING ADDICTION AND ACADEMIC ATTITUDES: THE CASE OF COLLEGE STUDENTS IN THE PHILIPPINES. *INTERNATIONAL REVIEW OF HUMANITIES AND SCIENTIFIC RESEARCH* (03 2020), 417–426. https://www.researchgate.net/profile/Vicpher-Garnada/publication/339687790_ONLINE_GAMING_ADDICTION_AND_ACADEMIC_ATTITUDES_THE_CASE_OF_COLLEGE_STUDENTS_IN_THE_PHILIPPINES/links/5e5fc5a4a6fdccbeba1c5c61/ONLINE-GAMING-ADDICTION-AND-ACADEMIC-ATTITUDES-THE-CASE-OF-COLLEGE-STU pdf

[12] Karen Gasper, Lauren A. Spencer, and Danfei Hu. 2019. Does Neutral Affect Exist? How Challenging Three Beliefs About Neutral Affect Can Advance Affective Research. *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.02476

[13] Yucheng Huang, Rui Song, Fausto Giunchiglia, and Hao Xu. 2022. A Multitask Learning Framework for Abuse Detection and Emotion Classification. *Algorithms* 15, 4 (2022). https://doi.org/10.3390/a15040116

[14] Sidney V. Irwin, Anjum Naweed, and Michele Lastella. 2023. The AACTT of Trash Talk: Identifying Forms of Trash Talk in Esports Using Behavior Specification. *Journal of Electronic Gaming and Esports* 1, 1 (2023), jege.2022–0024. https://doi.org/10.1123/jege.2022-0024

[15] Beakcheol Jang, Myeonghwi Kim, Gaspard Harerimana, Sang-ug Kang, and Jong Wook Kim. 2020. Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. *Applied Sciences* 10, 17 (2020). https://doi.org/10.3390/app10175841

[16] Shengyi Jiang, Yingwen Fu, Xiaotian Lin, and Nankai Lin. 2021. *Pre-trained Language Models for Tagalog with Multi-source Data.* Springer Nature Switzerland AG 2021, 210–223. https://doi.org/10.1007/978-3-030-88480-2_17

[17] Nicolas T. Pujante, Jr. 2021. Speech for Fun, Fury, and Freedom: Exploring Trash Talk in Gaming Stations. *Asian Journal of Language, Literature and Culture Studies* 4, 1 (January 2021), 1–11. http://science.sdpublishers.org/id/eprint/217/

[18] Pranav Kompally, Sibi Chakkaravarthy Sethuraman, Steven Walczak, Samuel Johnson, and Meenalosini Vimal Cruz. 2021. MaLang: A Decentralized Deep Learning Approach for Detecting Abusive Textual Content. *Applied Sciences* 11, 18 (2021). https://doi.org/10.3390/app11188701

[19] Ryan Labana, Jehan Hadjisaid, Adrian Imperial, Kyeth Jumawid, Marc Jayson, M Lupague, and Daniel Malicdem. 2020. Online Game Addiction and the Level of Depression Among Adolescents in Manila, Philippines. *Central Asian Journal of Global Health* 9 (12 2020). https://doi.org/10.5195/cajgh.2020.369

[20] Sanghyub John Lee, JongYoon Lim, Leo Paas, and Ho Seok Ahn. 2023. Transformer transfer learning emotion detection model: synchronizing socially agreed and self-reported emotions in big data. *Neural Comput. Appl.* 35, 15 (jan 2023), 10945–10956. https://doi.org/10.1007/s00521-023-08276-8

[21] E. Ong. 2022. *Building a rich digital lexicon for low-resource Philippine languages. FilWordNet.* http://filwordnet.dlsu.edu.ph/in-focus/digital-lexicon-philippine-languages-overview/

[22] W. Opinion. 2020. *Toxicity in Gaming Is Dangerous. Here's How to Stand Up to It.* https://www.wired.com/story/toxicity-in-gaming-is-dangerous-heres-how-to-stand-up-to-it/

[23] Ria Sagum, Aldrin Ramos, and Monique Llanes. 2019. FICOBU: Filipino WordNet Construction Using Decision Tree and Language Modeling. *International Journal of Machine Learning and Computing* 9 (02 2019), 103–107. https://doi.org/10.18178/ijmlc.2019.9.1.772

[24] Lee Jungmin Shannen Tadena, Kim Shin-Jeong. 2021. Empathy, cyberbullying, and cybervictimization among Filipino adolescents. *Child Health Nurs Res* 27, 1 (2021), 65–74. https://doi.org/10.4094/chnr.2021.27.1.65 arXiv:http://e-chnr.org/journal/view.php?number=1730

[25] Cuihua Shen, Qiusi Sun, Taeyoung Kim, Grace Wolff, Rabindra Ratan, and Dmitri Williams. 2020. Viral vitriol: Predictors and contagion of online toxicity in World of Tanks. *Computers in Human Behavior* 108 (2020), 106343. https://doi.org/10.1016/j.chb.2020.106343

[26] Zicheng Zhu, Renwen Zhang, and Yuren Qin. 2022. Toxicity and prosocial behaviors in massively multiplayer online games: The role of mutual dependence, power, and passion. *Journal of Computer-Mediated Communication* 27, 6 (09 2022), zmac017. https://doi.org/10.1093/jcmc/zmac017 arXiv:https://academic.oup.com/jcmc/article-pdf/27/6/zmac017/45870068/zmac017.pdf