SMU Data Science Review

Volume 6
Number 2 *Summer 2022*

Article 6

2022

Predicting Twitch.tv Donations using Sentiment Analysis

Alexander J. Gilbert
Southern Methodist University, alexgilbert84@gmail.com

Jason Herbaugh
Southern Methodist University, jherbaugh@mail.smu.edu

Feby Cheruvathoor
Southern Methodist University, fcheruvathoor@mail.smu.edu

Ben Williams
bentropy@gmail.com

Alex Tozzo
actozzo@gmail.com

1 Introduction

Video game streaming can be defined as the process where content creators record themselves playing video games while other gaming enthusiasts communicate with those creators in real time. This concept grew as a natural extension of earlier websites, such as justin.TV, where users would broadcast user generated live video content. As the popularity of streaming and video games has continued to grow over time, video game streaming has become an incredibly popular platform for community engagement around the world.

Each of the biggest video gaming streaming platforms is backed by a major company. YouTube Gaming is backed by Google, whereas Microsoft partners with Facebook. For our paper, we will be looking at Twitch.TV—owned by Amazon—the successor to Justin.TV and which was rebranded Twitch Interactive in 2014.

Video game streamers broadcast themselves playing games to highlight their skills and foster a community. Viewers of the streaming platforms watch different games to get a preview of what a game looks like or to follow their favorite personalities as the streamers play games. As Twitch has grown, streamers have been looking to make money by monetizing their broadcasts using ads and earning viewer donations and subscriptions. As the streamer games, viewers of the channel can donate "tips" to the streamers. Viewers can also subscribe to a channel, as doing so allows them preferred access to exclusive features such as VIP chat and channel-specific rewards provided by the streamer. Streamers set a streaming schedule ahead of time so that followers know when they are online and can maximize their audience.

As popular streamers gain momentum, communities tend to form around these strong personalities. In such communities, it is common for "gamer" vocabulary to become prevalent in the everyday lexicon. From game-to-game or even stream-to-stream communities will develop their own manner of communicating using slack, game references, and emojis. Cabeza-Ramirez et al. (2020) [3] detailed what factors influence users to subscribe to game streaming platforms like Twitch. In their study, they explored the various motivational elements related to the use of the streaming platform. Their hypothesis centered on user social and demographic aspects and how they found the game streaming platform quite different than simply watching someone play a game. User social commitments, interests, charity, and numerous other factors lead the user to connect to a specific game community. Our research is focused on finding the correlation between the sentiments of chat room discussions and user donations, which is explained in the sections below.

A study by Eligio and Kaschak [6] have examined how the lexical representation of gaming-related words have different associations between veteran gamers and community members when compared to non-gamers. Video games have always had their own subculture and lexicon, including memes, emojis, and nuances around language.

Looking at the comprehensive approach of how language and sentiment analysis drives people to donate to a channel in conjunction with the other content that streamers also curate, we have aimed to identify the patterns that drive donations based upon a myriad of factors, including number of viewers, gender of streamer, and genre of gameplay. A study conducted by Yoganathan et al. (2021) [25] broadly covered the characteristics of the donation process and intentions behind viewers donating to streamers they like or their channels of interest. With this thought in mind, our research intends to track this intention of the viewer by analyzing sentiment and using it to predict when a donation will occur.

The research questions are as follows: Does the overall sentiment of a chat room lead to the audience donating to the streamer? Does the audience donate based upon certain criteria or opinions of the content? What is the role of the game in conjunction with sentiment?

We studied these questions while uncovering the sentiment of the room and how this relates to donations. Our contributions are as follows:

- Baseline scoring for predicting the occurrence of streaming donations.
- Support that the gender of the streamer and genre of games are weak factors with regard to predicting donations.
- Evidence that collective sentiment of a viewer base is useful for predicting the occurrence of donations.
- Examples of boosting methods such as LightGBM and XGBoost being useful in improving prediction models for this use-case.

2 Literature Review

A limited number of studies have examined Twitch. Of the studies, the papers can be classified into a handful of categories limited to understanding audience participation in chat, investigating emojis used in Twitch chat, and exploring tips and donations on Twitch. The following section captures previous research work done.

2.1 Prior Analysis

In a study by Chen et al. (2021) [4], text mining was utilized to understand how real-time comments by Twitch.tv viewers affected the other viewers of a channel. After acquiring chat log data for dozens of different games and identifying features using support vector machines, Chen et al. identified three main factors that can influence the viewership of streams: “game instruction,” “game content,” and “game characters.” The study focused on viewership, specifically without examining donations. Several feature selection techniques were utilized: least absolute shrinkage and selection operator (LASSO), support vector machines – recursive feature elimination (SVM-RFE), and a chi-square test. We have utilized each method in analyzing our dataset.

Jia et al. (2016) [10] examined replays and live streams between streamers and their audience to see what drove activity. The study examined both the characteristics of streamer and audience, such as the popularity of the streamer, activity level of streamer, and spectator interactions with replays. The authors showed that creators often emphasized their individual skills, but spectators appreciated team play tactics.

Constant interaction with viewers can take an emotional toll on steamers, which can influence their performance and, ultimately, the interaction between the streamer and viewer. Woodcock and Johnson (2019) [23] noted the workload of streaming and its impact on the livelihood of streamers. It was not uncommon for a streamer to present as a "character" to reduce the emotional energy expended while streaming. That character also enabled the streamer to connect with their audience in interesting ways that could form a bond, leading to the individual goals of the streamer. With knowledge of how sentiment aligns with monetary rewards, a streamer may be more effective at driving viewer behavior while in character rather than out of character.

Another study by Xu et al. (2021) [24] showed that the success of video game streaming depended on the active participation of the viewers because this created economic, hedonic, and social values. A study by Hilvert-Bruce et al. (2018) [8] utilized ordinal linear regression analysis to identify the six motivations that drive live-stream engagement: social interaction, sense of community, meeting new people, entertainment, information seeking, and a lack of external support in real life.

Additional research has been performed to analyze what elements affect a viewer's disposition toward commenting during a stream. Wang and Li (2020) [22] showed that the viewership size of a channel, streamer gender, gifting toward the streamer, and duration of a stream all affected the commentary. According to their results, females engage audiences at a rate higher than male streamers while gifts contribute to greater overall engagement. Streams that go on too long were shown to have a negative impact on viewer participation. Their study used the Chinese streaming platform of Huajiao, which has similarities to Twitch.tv, and they expanded their study to include not only video games, but also talent shows, which are popular on the platform. The concept of interactional ritual chains, a theory describing how interactions can increase or deplete emotional energy depending on how they are interpreted by a group, played a significant role in Wang's study. In the analysis, it was suggested that streamers can influence their communities by manipulating the use of symbols to influence commentary. An interesting area of research in this space would be to analyze the effect on donations with regard to viewers who have a propensity to avoid participation in chat. Does chat-based engagement drive donations, or are silent viewers just as likely to donate?

Another study by Song et al. (2021) [21] presented a deep learning model that evaluated "epic" moments according to audiences' collective evaluation. Their study showed that they could identify epic moments from user-generated contexts covering a variety of contexts, such as victory, funny, awkward, or embarrassing. Another study by Recketenwald [20] took a micro-level look at both communication and play in conjunction with each other, introducing the concept of "pivoting," which is like the "roar of the stadium." This idea of momentum towards a particular sentiment in chat is central to the topic addressed in our research, as will be expressed when talking about a "rolling sentiment window" below.

The neologisms of Twitch users cannot be ignored when performing a semantic analysis of Twitch chatlogs. According to Dolin et al. (2021) [5], in 2021, there were 8.09 million emotes used on the Twitch platform. Twitch users can subscribe to channels to use special emotes or have the ability to use Twitch default emotes. Table 1 shows the top emotes used on April 15, 2022 from the TwitchEmotes.com tracking website. It should be kept in mind that the large majority of those emotes were unique to conversations on Twitch, so longstanding semantic models lacked the capacity to effectively analyze Twitch chat. In their study, Dolin et al. [5] developed a framework using word embedding and k-NN to enhance the then-contemporary models. The result was an unsupervised framework called LOOVE (learn out of vocabulary emotions), which could more effectively characterize Twitch emotes and outperform the best previous framework's accuracy by 7.36%.

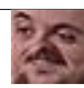
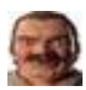
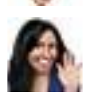
Rank	Emotes	Code	Usage
1		forsenE	1,727,137
2		homyLol	1,114,181
3		chipsaKEK	1,045,561
4		TriHard	529,446
5		HeyGuys	485,207
6		Kreygasm	437,714
7		LUL	421,595

Table 1: Emotes trending for the day - April 15, 2022. Twitchemotes.com

Another study conducted by Eligio and Kaschak [6] detailed the lexical representation of gaming-related words and the effects on the video game experience. Here, a veteran gamer who had played video games for a long time was more likely to associate game-related meanings in a word association task. However, in our study, we were not keen to identify the gamers' experience and age, which was not available from the Twitch dataset. This is an interesting area for research as there are potentially complex interactions based on age to be seen. Typically, older people are more likely to have disposable income for tipping, but there could be some generational or demographic influence on the culture around tipping that might encourage a younger audience to tip more.

Ravid et al. (2015) [19] performed a survey of different tasks, approaches, and applications regarding opinion mining and sentiment analysis. The study showed that the invention of Web 2.0 and increased usage of social media platforms led to electronic word of mouth (eWOM) statements becoming more prevalent, enabling customers to share their point of view. This concept links back to the study by Wang and Li in that the opportunity for one to share their point of view more-easily is likely to lead to increased interaction with a community and raising the question of whether they are more likely to donate based on this interaction.

Social media has been a big factor in contributing studies looking into sentiment analysis tasks, but most prior studies have mainly focused on Twitter. For example, the study "Multi-class Twitter sentiment classification with emojis." [15] used a method of extracting emojis from chat logs and processing them separately. The approach adopted by Li. et al was to find the sentiment score of the text component of a sentence and the emoji segment of a sentence individually before blending the scores to arrive at the final sentiment score. Because there are no sentimental polarity scores provided by Twitch, prior analysis has not explored sentimental classification using emotes. Kim et al. (2022) [12] analyzed the use of emotes in toxic chat as an attempt to process the mix of text chat and emotes. Barbieri et al. (2017) [2] attempted to predict the emotes in the chat by understanding the intended meaning of the message

After examining numerous studies conducted in the live streaming gamer space, there were only a handful using sentiment analysis. Hence, we built sentiment analysis models exclusively to find the correlation between donations and the overall sentiment score while donations were made in chat rooms. In addition to sentiment analysis, we layered in additional context such as gender of streamers, games being played, and stream size to improve the model's predictive capability.

2.2 Hypothesis

Our hypothesis is that it is possible to develop a machine learning model to predict the occurrence of donations to Twitch streamers during their live streams based on real-time chat data by their viewers, with a focus on using viewer sentiment analysis as a feature in the dataset.

3 Experiment

This section discusses the methodology involved in the research, including data collection from Twitch chat logs, data cleansing and preparation using the NLTK library, performing an exploratory data analysis on the collected data to understand the raw data, developing the data pipeline, and building machine learning models before finally performing the model evaluations.

3.1 Methodology

The main goal of the project was to use chat log data such as viewer data, streamer data, stream metadata, and chat messages processed for sentiment to predict the occurrence of donations.

First, we collected the chat data and prepared the texts for sentiment analysis. Individual chats from these logs were lemmatized and processed by the lexical analyzer and given a polarity score based on the language in the room during the time of the donation. The different lexical models used in our research processed the chat logs separately, and model benchmarking was thereafter carried out.

3.2 Data Collection and Data Processing

The first step in building the pipeline involved retrieving chat logs from the Twitch platform.

The “TwitchDownloader” command-line interface (CLI), which was developed by Lay295 (GitHub persona), was used to download VOD streams from the Twitch VOD portal. The application was customized to download chat logs from select streams and then convert each log into .txt or JSON format for processing. The objective was to retrieve the key attributes, such as user details, chat messages, emotes details, and other metadata information. In addition, gamer demographic information such as username, subscription details, gender, and streaming duration was prepared while retrieving the chat logs from the Twitch streaming VOD portal.

Table 2: Example of several columns of data provided by TwitchDownloader Chat logs that were collected across five different genres of games to obtain a broad sample of streamers. Our targeted games were among the most popular games on the Twitch platform. We chose the following games as the target of our inquiry: Hearthstone, FIFA 2022, Apex Legends, Minecraft, and Dota 2.

streamer.id	Time Stamp	Message Body	Game
64073463	2015-06-13T02:53:37.708196Z	vixClap	Minecraft
64073463	2021-07-12T19:52:39.036878Z	KEKW	Minecraft
64073463	2014-12-02T17:40:22.73143Z	booooo	Minecraft
64073463	2018-04-08T18:17:20.276222Z	SlowClap	Minecraft
64073463	2015-06-13T02:53:37.708196Z	Cool under pressure OMEGALUL	Minecraft

Utilizing the TwitchDownloader CLI, a Python script was developed to scrape the chat log data. Twitch allows VODs to be kept on the platform for a maximum of 14 days for regular streamers, which means all data of interest in a study must be collected within that timeframe. The streams were selected based on the viewer count of the individual stream (between 450 and 3,300) and the length of the stream (at least 2 hours).

The viewer count requirement of 450 to 3,300 was based on the fact that there are enough streamers able to achieve this count to enable our data collection and the streamers who achieve this viewer count are capable of generating enough donations to earn an income comparable to many full-time jobs in the United States.

Table 3: Counts of comments made per stream and number of streams per Gender. The count represents comments made on each gender's stream, not count of comments made by each gender.

Gender	Total Comments	Stream Count
Female	35971	111
Male	64482	189

Two-hour streams were the minimum stream length because we chose to analyze one hour of stream chat, starting at the top of the second hour. Many streams go on air without any content or gaming for the first hour, which would have prevented us from gaining valuable data from chat.

A total of 300 English-speaking VODs were collected, each from a unique, random streamer. The streams were evenly split between the five games chosen for the study. Those games, Hearthstone, FIFA 2022, Apex Legends, Minecraft, and DOTA 2 were selected because they are representative of the popular genres found on Twitch and were in the Top 25 most popular games on Twitch at the time of the study.

Table 4: Number of comments and streams per Game.

Game	Total Comments	Stream Count
Apex Legends	19285	60
Dota 2	14759	60
Fifa 22	19998	60
Hearthstone	13380	60
Minecraft	33031	60

3.3 Exploratory Data Analysis

Once the full dataset was collected, basic summary statistics were calculated. Table 3 shows that most of the comments were made on channels with male streamers. The average, however, is closely aligned. When looking at the comments for each game, there is a clear variance in the data. Although each game is represented by 60 streams, Minecraft has a high plurality of comments relating to it.

The words frequently used in this sample of chat data were the lingo most often used by those in the gaming sphere. “KEKW” is a term made famous on the Twitch platform itself and is derived from a translation of “LOL” (laugh aloud) between opposing factions in World of Warcraft. “Pog,” meaning “play of the game,” while not originating from the game Overwatch, was made famous thanks to being featured at the end of every Overwatch match. The term “GG” often refers to “good game” or “get good.”

Figure 1: Wordcloud of highly occurring words in dataset

Currently available, public lexical knowledgebases such as NLTK (which was used for this study) will have difficulty appropriately representing the words and phrases found in Twitch chat. This is a limitation of the study but developing a new knowledgebase would take an incredible amount of effort and was out of scope of this paper.

To obtain a better understanding of the basics of the data, an analysis of the metadata was performed. During initial data ingestion, we captured several data points from the stream. We recorded the gender of the streamer, the average user count of the stream, and the game that the streamers were playing. Streams over two hours in length and only in English were considered.

The games were chosen strategically because they represent a cross-section of the most popular genres in video games and streaming. Hearthstone is a digital collectible card game, whereas Minecraft is an open world sandbox game. FIFA 22 is a football simulation game, and Dota 2 is one of the most popular multiplayer online battle arena (MOBA) video games. Finally, Apex Legends is a first-person shooter mixed in with a battle royale video game.

It was found that the user count in each stream was fairly close, regardless of

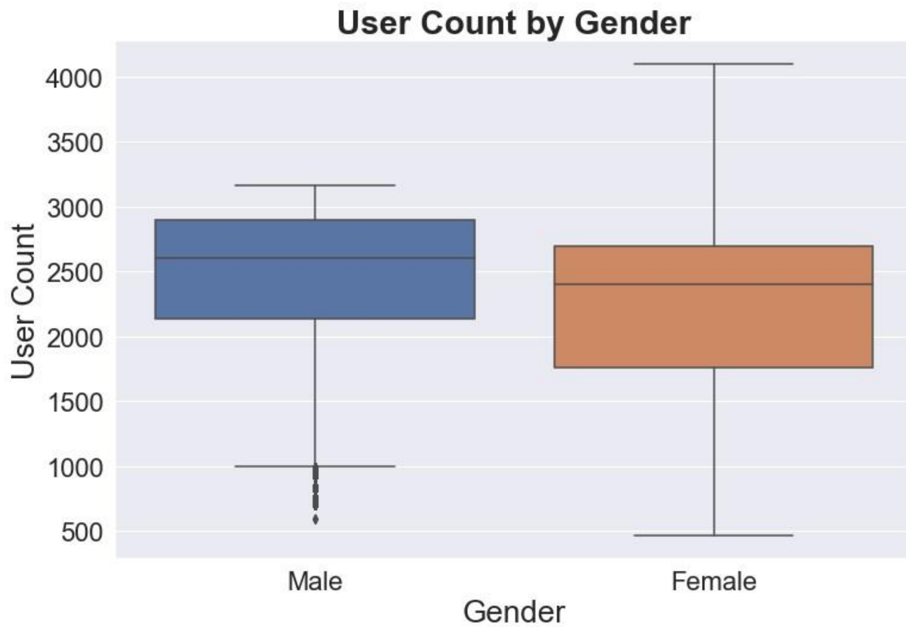


Figure 2: The user counts of viewers in all streams is fairly even.

whether the streamer was male or female. Male streamers had slightly higher user counts at 2,550 versus female streamers at 2,450 (Figure 2). The user count per game was inconsistent, so this was considered during the analysis. If possible, future research should be performed with datasets of similar size for each game (Figure 3).

Hearthstone streamers had the greatest range of viewers while Minecraft streamers enjoyed higher viewers and a greater visibility when streaming. This is most likely due to Minecraft being one the most popular games across several video game platforms.

Although, at a high level, the user counts were similar regardless of gender, when each game was considered separately, there was a clear distinction of user count based on gender (Figure 4). It would be extremely difficult to find a more evenly distributed user count for analysis based on these games because there has been a historical preference for viewership for certain genres of games based on gender.

Examining the stream count per game grouped by gender (Figure 5), females were underrepresented in both FIFA 22 and Hearthstone. Every attempt was made to collect equal samples of both genders. Apex Legends and Dota 2 had the most equal representation among the male and female streamers.

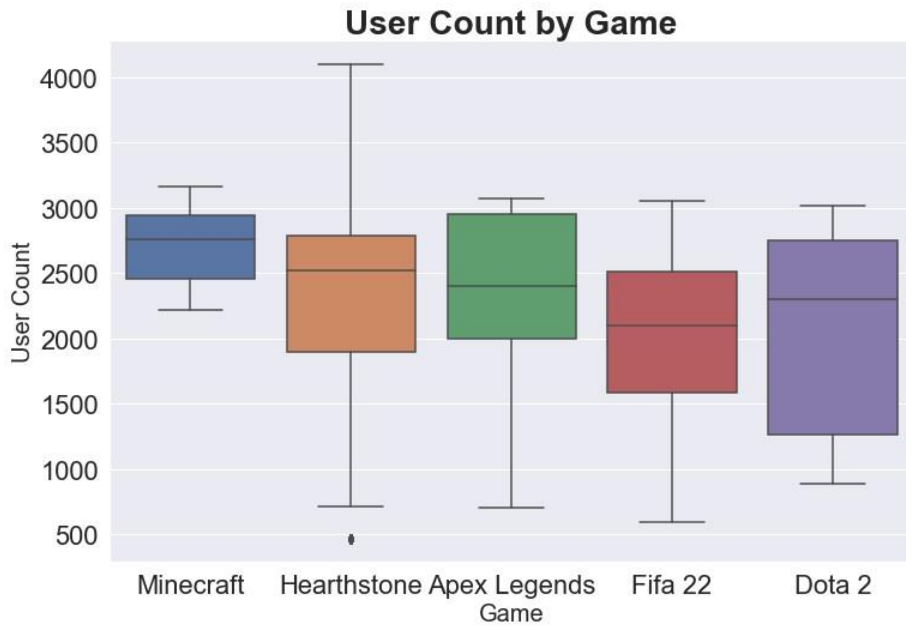


Figure 3: There is some imbalance in the user counts by game.

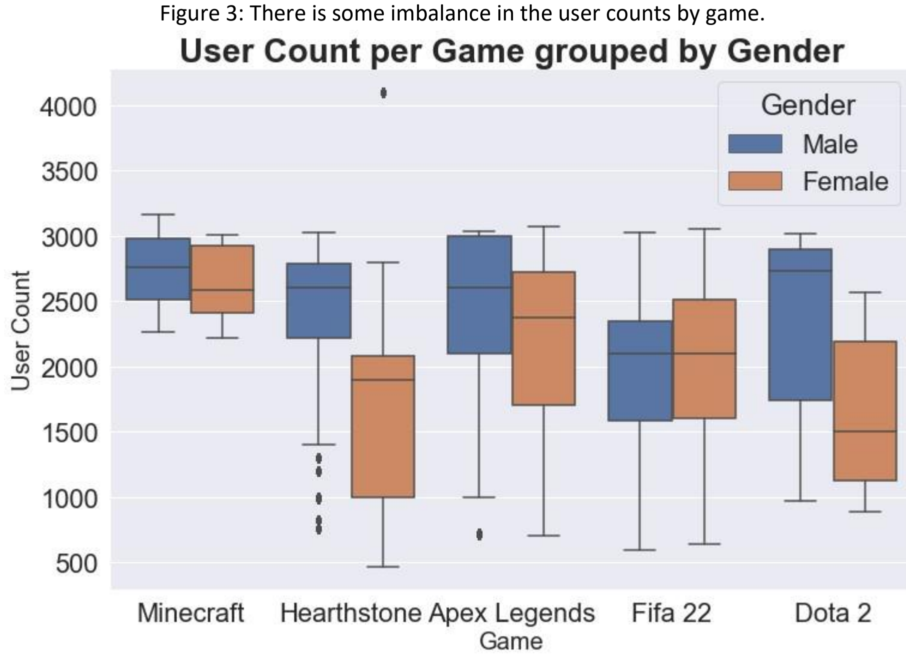


Figure 4: The data set was not evenly split between genders for all games.

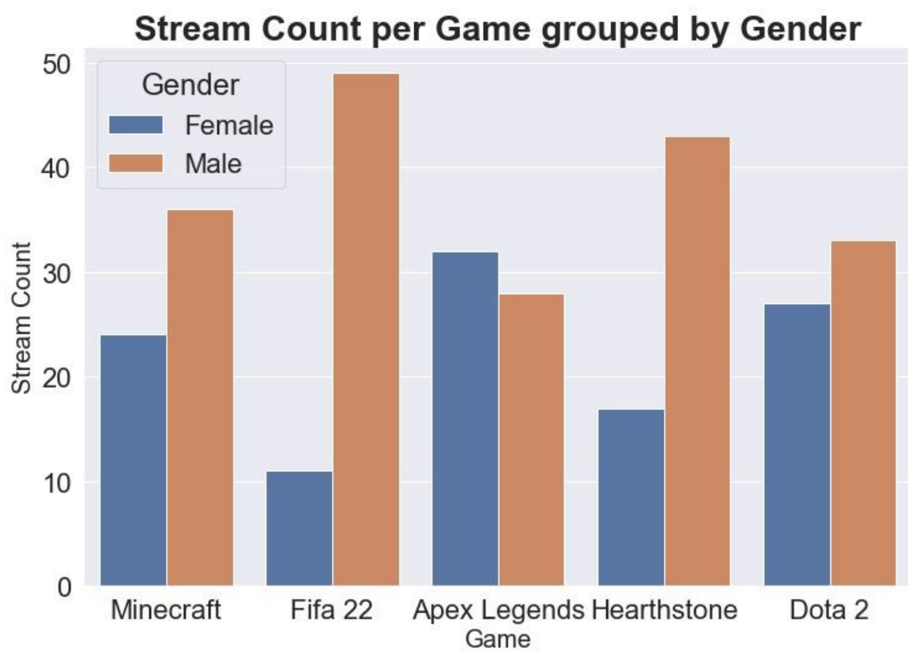


Figure 5: An even split of stream counts per gender was attempted but was not possible based on the selection criteria.

3.4 NLP Pipeline

After the data was gathered, examined, and cleaned, the NLP pipeline as shown in Figure 6, was applied using the NLTK package. The first step in the NLP pipeline was to break down the body of the chat messages into easily digestible tokens. Next, text cleansing was applied to remove noise in the data, such as special characters, punctuation, and whitespace. Chat messages that contained emotes were ingested from the TwitchDownloader and translated to their text representation. All text was standardized so that the text was lower case. Words were normalized to the root normal form, stemmed of the prefixes and suffixes, and lemmatized. At this stage of tokenization, 100,453 individual chat messages were processed, and 375,073 tokens were generated.

During our study, three distinct pretrained sentiment analysis frameworks were applied to the chat messages so that the sentiment could be calculated.

Valance aware dictionary for sentiment reasoning (VADER) is a simple rule-based model for general sentiment analysis. VADER uses a list of lexical words that are labeled as positive or negative according to their semantic orientation. As seen in a paper by Hutto et al. (2014) [9], the accuracy of the VADER model is 96% and generalized more favorably across multiple contexts.

The TextBlob sentiment analyzer returns two properties for a given input sentence. TextBlob measures both polarity and subjectivity. Polarity is a measure of sentiment between -1 and 1, while subjectivity is a measure of opinion that ranges from 0 to 1. The general reported accuracy for sentiment analysis is 77.2% [18]. For our purposes, only polarity was considered. As seen in a paper by Hazarika et al. (2020) [7], TextBlob has been used across Twitter to analyze sentiment across machine learning classifiers and was found to be efficient and accurate.

Finally, we used a DistilBERT model within Flair. This is a pretrained model based on the BERT architecture. Flair was first proposed as a state-of-the-art NLP framework in the paper by Akbik et al. (2019) [1].

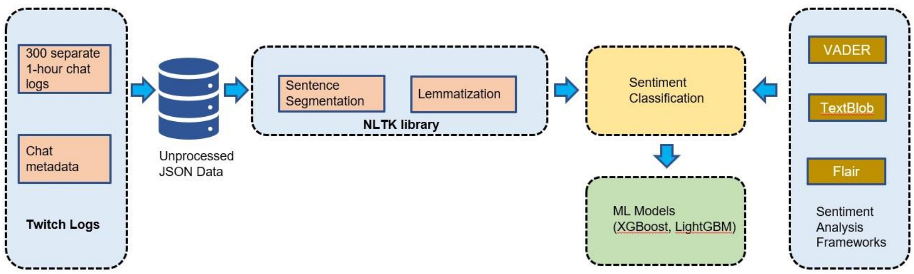


Figure 6: Data Pipeline used for this study.

3.5 Rolling Window Framework

After creating a feature for each VADER, TextBlob, and Flair to store the sentiment values for each individual message, a novel approach for creating a 10-message rolling sentiment value was applied. This rolling sentiment value was calculated by averaging the sentiment scores for each sentiment analysis model and assigning that average as the value.

Table 5: How sentiment scores compare with rolling scores.

Mean of previous 10 VADER sentiment scores & RollVader
Mean of previous 10 Flair sentiment scores & RollFlair
Mean of previous 10 TextBlob sentiment scores & RollBlob

3.6 Model Building and Evaluation

Once the NLP pipeline was resolved and sentiment scores applied, the data were in the final format for model evaluation. The goal of analysis was to identify the model that most effectively predicted whether a particular record would be associated with a subscription or gift. The scoring methods considered accuracy, precision, and recall. The components of F-1 Score, precision, and recall were particularly important because of the large imbalance in the dataset. It would be possible for a model to achieve relatively high accuracy by predicting only “Not-Subscribed,” but the model could not be considered good because it would never classify “Subscribed” records correctly.

After our analysis and pipeline of the data had been built, binary classification models were evaluated for accuracy, precision, and recall scores. Models were evaluated with grid search cross-validation so that the best hyperparameters were chosen. Each set of parameters included a balanced class weight property to ensure that the smaller target class was equally represented.

After finishing the stochastic gradient descent (SGD), decision tree, and boosted models, additional ensembled models were considered. In a paper by Lochter et al. (2016) [17], ensembled models were proposed to perform opinion detection in short text messages because they improved performance in most text categorizations. A separate paper by Korovkinas et al. [13] employed an SVM and Naive-Bayes classification ensemble method for sentiment analysis. Finally, Kazmaier et al. (2022) [11] showcased the power of ensemble learning in sentiment analysis and proposed median performance improvements up to 5.53% over individual models.

4 Results

As seen in table 6, SGD based classifiers performed rather poorly. The SVM model had an accuracy of 78.92%, precision of 5.40%, and recall of 73.05%. The SGD model performed slightly worse, with an accuracy of 93.70%, precision of 7.92% and recall of 27.37%.

Table 6: Comparison of accuracy & F-1 scores for each model.

Name	Accuracy	Precision	Recall
Random Forest	99.12%	92.34%	49.59%
XGBoost	99.08%	78.79%	58.85%
SVM	78.92%	5.40%	73.05%
SGD	93.70%	7.92%	27.37%
LightGBM	99.05%	77.99%	57.61%

Counting our evaluation of the results, a random forest algorithm was performed as the middle of the pack classifier. The model produced a high accuracy of 99.12%, with a precision of 92.34% and recall of 49.59%. It is to be expected that methods building off random forest such as LightGBM and XGBoost would show improvements over the base model, which was observed in the results we obtained.

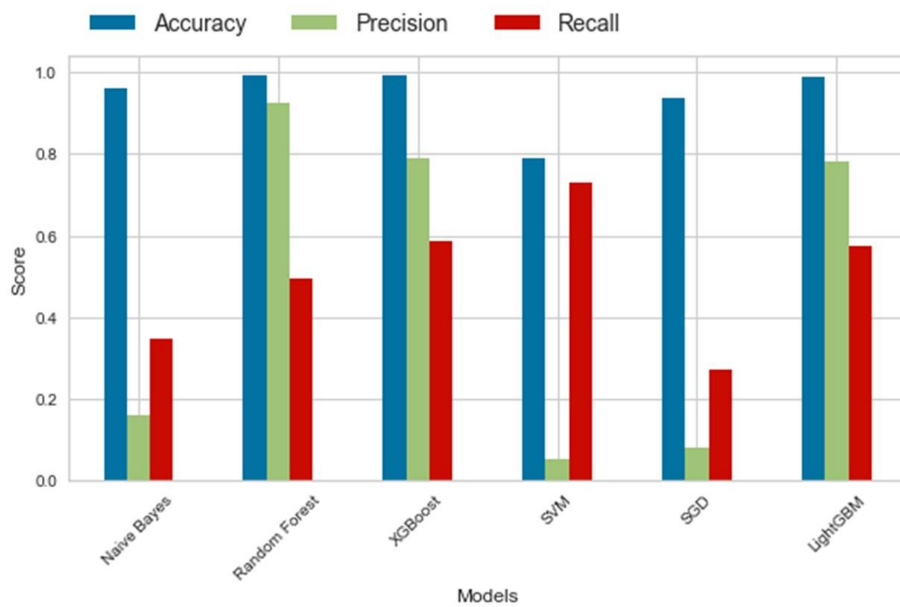


Figure 7: Visual representation of accuracy & F-1 scores.

The analysis found the most effective models to be the Gradient Boosted Decision trees, which had the highest accuracy, precision, and recall. Because there were no real differences in the results between LightGBM and XGBoost, LightGBM was chosen as the preferred model because of the speed at which the algorithm performs. The LightGBM model had an accuracy of 99.05%, precision of 77.99%, and recall of 57.61% and completed approximately 6.5 times as fast as XGBoost. As seen in Figure 8, the most important features were the Flair sentiment, followed by user count and the rolling sentiment scores for Text Blob, Flair, and VADER.

5 Discussion

The analysis provided evidence that sentiment analysis can be effectively utilized to predict the occurrence of donations to a streamer with an excellent balance of accuracy, recall, and precision. Considering the rarity of donations in chat, a high rate of true positive predictions without making false negative predictions is what makes this effort valuable. The analysis also showed that the novel "rolling window" method was a useful way to interpret chat sentiment data relative to time. In fact, the rolling averages were some of the more important features for making good predictions. To improve the usefulness of this concept, future work could benefit from identifying some threshold for a rolling sentiment after which the next message would represent a donation as opposed to the current model which successfully predicts if the last message represents a donation.

When conceptualizing this study, our intuition was that both the type of game and the gender of the streamer would have a significant impact on the ability to predict donations. As seen in Figure 8, sentiment was a much more important element in making this determination. That is not to say there is not a significant impact of gender or game in a user's propensity to donate. It may still be the case that users are more likely to donate to women rather than men or that Minecraft viewers are more likely to donate than FIFA 22 viewers, but if either of those things is the case the impact of those facts is overshadowed by sentiment for the purposes of making predictions.

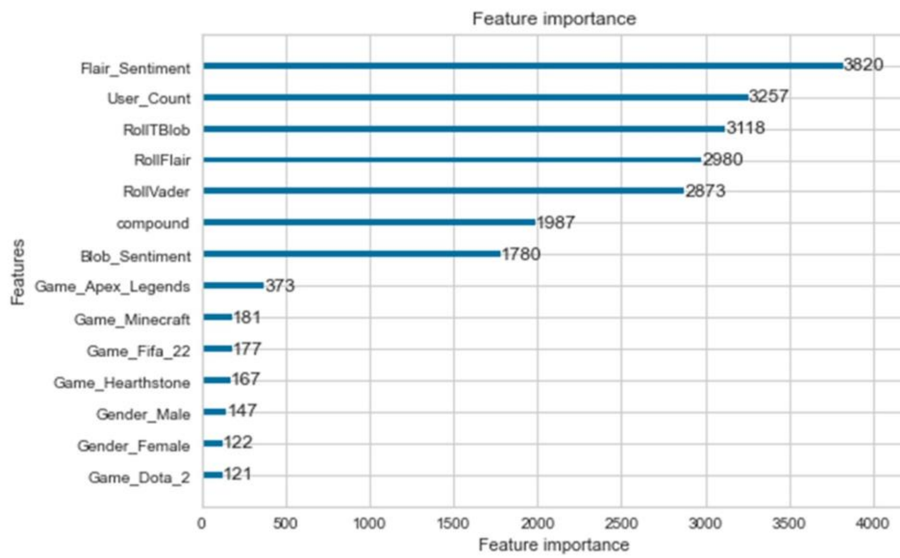


Figure 8: Important features defined by LightGBM.

The study was based on historical data rather than collected experimentally, so it would be an overreach to make strong suggestions for streamers about how they should behave based on the current study. The long-term effects of maintaining any sentiment were not considered, nor was any information about viewers outside of the context of each stream was known. It is possible that there are social dynamics at play that far outweigh the effects of real-time sentiment that should not be ignored when interpreting the data.

We have developed a baseline method for predicting donations based on sentiment that can be built upon by streamers, game developers, social platforms, and brands to improve their understanding of users to drive monetization that will encourage the growth of platforms. As a baseline, this method should be evaluated and improved to increase the predictive capabilities for future models.

5.1 Limitations & Future Work

While our study achieved useful results, there are many opportunities to expand on the research to improve our collective understanding of viewer sentiment.

Although emotes have been acknowledged as an important element of communication on Twitch, this study could be greatly improved in this area. At the outset of the data collection and processing phase, it became abundantly clear that the enormous and ever-evolving dimensionality of communication through emotes was worthy of years' worth of study in itself. A study performed by Liu et al. (2021) [16] found that emojis are effective for improving the accuracy of sentiment analysis algorithms. Adding in an emoji sentiment lexicon such as the "Emoji Sentiment Ranking," as seen in Novak et al. (2015) [14], may also augment the dataset. We are hopeful that other researchers are successful in creating libraries similar to VADER, Flair, and TextBlob in a manner that gives proper attention to Twitch emotes. When that day comes, integrating those findings into research on sentiment analysis such as ours could amplify the findings to a significant degree.

Our study took the approach of analyzing the general sentiment of the chat room without paying attention to metadata about the users, but a focus on the individual viewers could also add new dimensions this research. Understanding demographic information, history about viewer interactions, attributing sentiment scores to previous chat, and watching donation history could all be valuable avenues for exploration. Of course, these opportunities present new ethical challenges, but with proper attention to those concerns, the effort could be worthwhile.

The viewer counts of each chat log analyzed were a limiting factor as well. We looked at streams from a few hundred to a few thousand viewers, but Twitch offers a platform for those who have a dozen or less to tens of thousands of viewers. By expanding the scope of stream sizes, it may be found that different attributes influence the prediction of donations in different ways.

Finally, tools that enable a user to make real-time predictions could dramatically change the interactions between streamers and viewers. Analyzing historical data is helpful in an academic environment but has limited usefulness to a streamer in the moment that they are able to make use of the data. Providing tooling for them to gain actionable insights in the moment is where real value can be captured outside the academic realm.

5.2 Ethics

The current study contains information about users on the Twitch platform. Therefore, every avenue to protect the user's anonymity was utilized. Personal identifying data included in the metadata provided by TwitchDownloader that was not critical to the study was removed from the dataset and any required information was anonymized, so it was unknown to the researchers. A large problem on the Twitch platform is doxing, which is an act of revealing private information online. By scrubbing the data ahead of the study, this minimized the ability for doxing to occur.

As the area of study evolves, close consideration should be given to the idea of targeted content in general. It is generally accepted that many people are against the idea of their identity being tracked for the purpose of serving advertisements, even when that personal information is publicly available. While Twitch is a public platform, following from this perceived invasion of privacy some users may not appreciate the idea that their data will be analyzed and used by a streamer to influence the content being created. Giving viewers an opportunity to opt in or opt out of their data being used in this way may be a reasonable answer to this problem of privacy.

Communication analysis tools like the ones developed for this study are potentially powerful devices. As these tools become more widely available to the streamer community, there is a question about the professionalism to be expected by users. Typically, streamers are amateur gaming enthusiasts trying to connect with a community and entertain them. Even professional streamers are most often solo performers without many resources at their disposal. When offering analysis tools to this userbase, the ethical considerations should be put on full display to ensure the streamers are being thoughtful about how they are using the data and insights available to them.

Another aspect of the Twitch platform is that community guidelines that are followed when streaming during video game sessions. Illegal conduct, such as breaking the law, violent behavior, hateful conduct, or unauthorized sharing of private information, is against Twitch's term of service. All efforts were made to verify that our dataset was clean of this information.

6 Conclusion

Sentiment analysis can be useful for predicting donations on Twitch. Our model was able to predict when someone donated to the stream over 57% of the time. Neither gender nor genre of the game seemed to have an impact on the donation in the chat room.

The best model was the one with that combined TextBlob, VADER, and Flair sentiment. The best model also included both boosted frameworks (XGBoost and LightGBM).

Acknowledgements

We want to thank our advisors, Ben Williams and Alex Tozzo: Ben, who helped us understand the novelty of the research we set out to perform, and Alex, who guided us in making proper decisions about the models and our findings. We appreciate the opportunity to perform research in an area we are passionate about, given to us by Southern Methodist University. Thank you to our wives who put up with our late nights and thank you to our kids who had to wait their turns to watch Twitch and play Fortnite.

References

[1] Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). Flair: An easy-to-use framework for state-of-the-art nlp. *Proceedings of the 2019 Conference of the North*. https://doi.org/10.18653/v1/n194010

[2] Barbieri, F., Espinosa-Anke, L., & Ballesteros, M. (2017). Towards the understanding of gaming audiences by modeling twitch emotes. Retrieved July 4, 2022, from https://aclanthology.org/W17-4402.pdf

[3] Cabeza-Ram´ırez, L. J., Sanchez-Ca´ nizares, S. M., & Fuentes-Garc˜ ´ıa, F. J. (2020). Motivations for the use of video game streaming platforms: The moderating effect of sex, age and self-perception of level as a player. *International Journal of Environmental Research and Public Health*, *17*, 7019. https://doi.org/10.3390/ijerph17197019

[4] Chen, W.-K., Chen, L.-S., & Pan, Y.-T. (2021). A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. *Applied Soft Computing*, *111*, 107704. https://doi.org/10.1016/j.asoc.2021.107704

[5] Dolin, P., d'Hauthuille, L., & Vattani, A. (2021). Feelsgoodman: Inferring semantics of twitch neologisms. *arXiv:2108.08411 [cs]*. Retrieved July 4, 2022, from https://arxiv.org/abs/2108.08411

[6] Eligio, R. B., & Kaschak, M. P. (2020). Gaming experience affects the interpretation of ambiguous words (S. Triberti, Ed.). *PLOS ONE*, *15*, e0243512. https://doi.org/10.1371/journal.pone.0243512

[7] Hazarika, D., Konwar, G., Deb, S., & Bora, D. J. (2020). Sentiment analysis on twitter by using textblob for natural language processing. *Annals of Computer Science and Information Systems*. https://doi.org/10.15439/ 2020km20

[8] Hilvert-Bruce, Z., Neill, J. T., Sjoblom, M., & Hamari, J. (2018). Social mo-¨ tivations of live-streaming viewer engagement on twitch. *Computers in Human Behavior*, *84*, 58–67. https://doi.org/10.1016/j.chb.2018.02.013

[9] Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[10] Jia, A. L., Shen, S., Epema, D. H. J., & Iosup, A. (2016). When game becomes life. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *12*, 1–24. https://doi.org/10.1145/2957750

[11] Kazmaier, J., & van Vuuren, J. H. (2022). The power of ensemble learning in sentiment analysis. *Expert Systems with Applications*, *187*, 115819. https://doi.org/10.1016/j.eswa.2021.115819

[12] Kim, J., Wohn, D. Y., & Cha, M. (2022). Understanding and identifying the use of emotes in toxic chat on twitch. *Online Social Networks and Media*, *27*, 100180. https://doi.org/10.1016/j.osnem.2021.100180

[13] Korovkinas, K., & Danenas, P. (2017). Svm and na˙ ¨ıve bayes classification ensemble method for sentiment analysis. *Baltic Journal of Modern Computing*, *5*. https://doi.org/10.22364/bjmc.2017.5.4.06

[14] Kralj Novak, P., Smailovic, J., Sluban, B., & Mozeti´ c, I. (2015). Sentimentˇ of emojis (M. Perc, Ed.). *PLOS ONE*, *10*, e0144296. https://doi.org/10. 1371/journal.pone.0144296

[15] Li, M., Ch'ng, E., Chong, A. Y. L., & See, S. (2018). Multi-class twitter sentiment classification with emojis. *Industrial Management & Data Systems*, *118*, 1804–1820. https://doi.org/10.1108/imds-12-2017-0582

[16] Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., & Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, *2*, 246–252. https://doi.org/10.1016/j.jnlssr.2021. 10.003

[17] Lochter, J. V., Zanetti, R. F., Reller, D., & Almeida, T. A. (2016). Short text opinion detection using ensemble of classifiers and semantic indexing. *Expert Systems with Applications*, *62*, 243–249. https://doi.org/10.1016/j. eswa.2016.06.025

[18] Mas Diyasa, I. G. S., Marini Mandenni, N. M. I., Fachrurrozi, M. I., Pradika, S. I., Nur Manab, K. R., & Sasmita, N. R. (2021). Twitter sentiment analysis as an evaluation and service base on python textblob. *IOP Conference Series: Materials Science and Engineering*, *1125*, 012034. https://doi.org/ 10.1088/1757-899x/1125/1/012034

[19] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, *89*, 14–46. https://doi.org/10.1016/j.knosys.2015.06.015

[20] Recktenwald, D. (2017). Toward a transcription and analysis of live streaming on twitch. *Journal of Pragmatics*, *115*, 68–81. https://doi.org/10.1016/ j.pragma.2017.01.013

[21] Song, H., Park, K., & Cha, M. (2021). Finding epic moments in live content through deep learning on collective decisions. *EPJ Data Science*, *10*. https://doi.org/10.1140/epjds/s13688-021-00295-6

[22] Wang, M., & Li, D. (2020). What motivates audience comments on live streaming platforms? (C. S. Park, Ed.). *PLOS ONE*, *15*, e0231255. https: //doi.org/10.1371/journal.pone.0231255

[23] Woodcock, J., & Johnson, M. R. (2019). The affective labor and performance of live streaming on twitch.tv. *Television & New Media*, *20*, 813– 823. https://doi.org/10.1177/1527476419851077

[24] Xu, X.-Y., Luo, X. (, Wu, K., & Zhao, W. (2021). Exploring viewer participation in online video game streaming: A mixed-methods approach. *International Journal of Information Management*, *58*, 102297. https://doi. org/10.1016/j.ijinfomgt.2020.102297

[25] Yoganathan, V., Osburg, V.-S., & Stevens, C. J. (2021). Freedom and giving in game streams: A foucauldian exploration of tips and donations on twitch. *Psychology & Marketing*, *38*. https://doi.org/10.1002/mar.21483