International Journal of Science, Technology and Society
2021; 9(6): 281-288
http://www.sciencepublishinggroup.com/j/ijsts
doi: 10.11648/j.ijsts.20210906.14
ISSN: 2330-7412 (Print); ISSN: 2330-7420 (Online)

Applying Machine Learning Models to Classify Xenophobic Tweets Against Asians, With Data Analysis of Hate Crimes

Gi Joon Chang[1, *], Seoyoon Choi[2, *], Gyeongmin Han[3, *], Heuiseo Kim[4, *], Inselbag Lee[5, *]

[1]Big Heart Christian School, YongIn, South Korea

[2]Seoul International School, Seongnam, South Korea

[3]Cardigan Mountain School, Canaan, United States

[4]Palisades Park High School, Palisades Park, United States

[5]St. Mark's School, Southborough, United States

Email address:

gijoonchang@gmail.com (Gi J. Chang), choieunie@gmail.com (S. Choi), gyeongmin.andy.han@gmail.com (G. Han), love6happy1@gmail.com (H. Kim), leeinselbag@gmail.com (I. Lee)

[*]Corresponding author

To cite this article:

Gi Joon Chang, Seoyoon Choi, Gyeongmin Han, Heuiseo Kim, Inselbag Lee. Applying Machine Learning Models to Classify Xenophobic Tweets Against Asians, With Data Analysis of Hate Crimes. *International Journal of Science, Technology and Society*. Vol. 9, No. 6, 2021, pp. 281-288. doi: 10.11648/j.ijsts.20210906.14

Received: September 23, 2021; **Accepted**: November 8, 2021; **Published**: November 19, 2021

Abstract: This paper offers insight to the COVID-19 pandemic and its effect on people's attitudes towards certain minority groups, particularly Asians, Asian-Americans, and Pacific Islanders. With the Coronavirus first being identified in Wuhan, China, xenophobia, and racism towards groups pertaining to the supposed origins of the COVID-19 pandemic have been on the rise. Along with the violent physical attacks on these groups, this paper will focus on the online hate and xenophobia that Asians face due to their race, ethnicity, country of origin, and/or others. In this paper, Python is employed as the primary programming language; external libraries such as pandas, NumPy, sklearn, WordCloud, and matplotlib are imported for handling data. In analyzing the racism against Asians, keywords such as "Asian Hate," "Hate Crime" and "anti-Asian" are utilized, and the Python programming language is employed to sift through Google News articles with these keywords and identify patterns in the words' usages. Furthermore, the frequencies of the keywords' usages on online platforms such as Twitter are also analyzed in the form of comma-separated files, with patterns of usage over time before and after the COVID-19 pandemic began being identified. Randomly selected tweets are classified into five categories: anti-Asian, not anti-Asian, not English, hate against others racial groups, and support towards Asians. These tweets are classified by artificial intelligence using machine learning methods of logistic regression, support vector machine, and Naive Bayes; the artificial intelligence was taught using pre-classified data sets. Classified tweets represent the implication and relevance between the tweets and xenophobia. This classification model of xenophobia is expected to be used in social media content censoring and enhance the internet chatting etiquette. The goal of this classification model is to terminate anti-Asian hatred and lower the overall level of societal racism.

Keywords: Asian Hate, COVID-19, Xenophobia, Racism, Online Hate

1. Introduction

Racism, though an outdated concept in theory, is still widely prevalent in today's society, particularly due to the protection and anonymity that the virtual world provides individuals on online platforms. Though it is undeniable that all people of color are subject to forms of racism both online and offline, this paper will focus on Asian Americans and Pacific Islanders who are part of the group that go through racial discrimination by extracting data from various platforms [1].

After the COVID-19 pandemic began, the hatred crimes toward Asians have increased by nearly 150% in number from previous years [2]. Over 4,000 Asian Americans and Pacific Islanders have been victims of hate incidents due to physical violence against them; this statistic does not include the number of AAPI (Asian American and Pacific Islander)

individuals who died outside of these causes, simply the ones reported for having died from hate crimes [3].. For instance, misconceptions such as lab-leak theory has increased the number of people who believed that Wuhan was responsible for the pandemic even though WHO, World Health Organization, concluded the theory was "extremely unlikely" [4]. Eventually, Asians were targeted as well, which could have been a cause for this rise in Asian hate crimes (whether that be due to racist sentiments that developed during the pandemic, or racist ideas that solidified due to the pandemic being blamed, by many, on Asians) [5].

Furthermore, the press and social media have significantly influenced people's opinions and ideas regarding racial segregation [6]. Platforms such as Twitter allow individuals to post hateful words and messages while hidden behind a screen, and the spread of these sentiments only worsen the issue online as well as in real life. Moreover, with so many people searching for reliable and accurate information online, the issue of misinformation, conspiracy theories, and fake news are also fueling anti-Asian narratives [7]. Of course, with a rise in hate crimes and online hate comes a rise in the fight against these issues, and the Stop AAPI Hate movement has gained much traction in the past year [8]. However, Asian hate is still very prevalent today, and trends will be examined to analyze the effect of COVID-19 on anti-Asian sentiments (as well as the support shown towards AAPI individuals) and its extent.

This paper will discuss two main terms: xenophobia and racism. To define each term, xenophobia is the "fear and hatred of strangers or foreigners," while racism is "prejudice, discrimination, or antagonism directed at a marginalized or minority group based upon their race" [9]. Asian Americans not only face racism but also xenophobia, and the distinction between these terms should be noted for the later discussion.

To clearly visualize how COVID-19 has impacted Asian hate, research was conducted by comparing the frequencies of mentions of words such as "Asian hate" and "anti-Asian" on platforms including Twitter and Google News before and after the COVID-19 pandemic began. The resulting data was displayed in line graphs, pie charts, bar charts, and Word Clouds for comparison. Open-source libraries were imported to conduct the data analysis, such as pandas and matplotlib on Python. The following paper has been organized into three sections: results of racial discrimination online, binary classification models, and the conclusion. The first section deals with the frequency of anti-Asian keywords on Google News and twitter before and after COVID-19. Adding on, the second section deals with the binary classification models using machine learning and the trend of xenophobic content on social media. The final section concludes the significance of anti-Asian data and the classification model's future impact on anti-Asian issues.

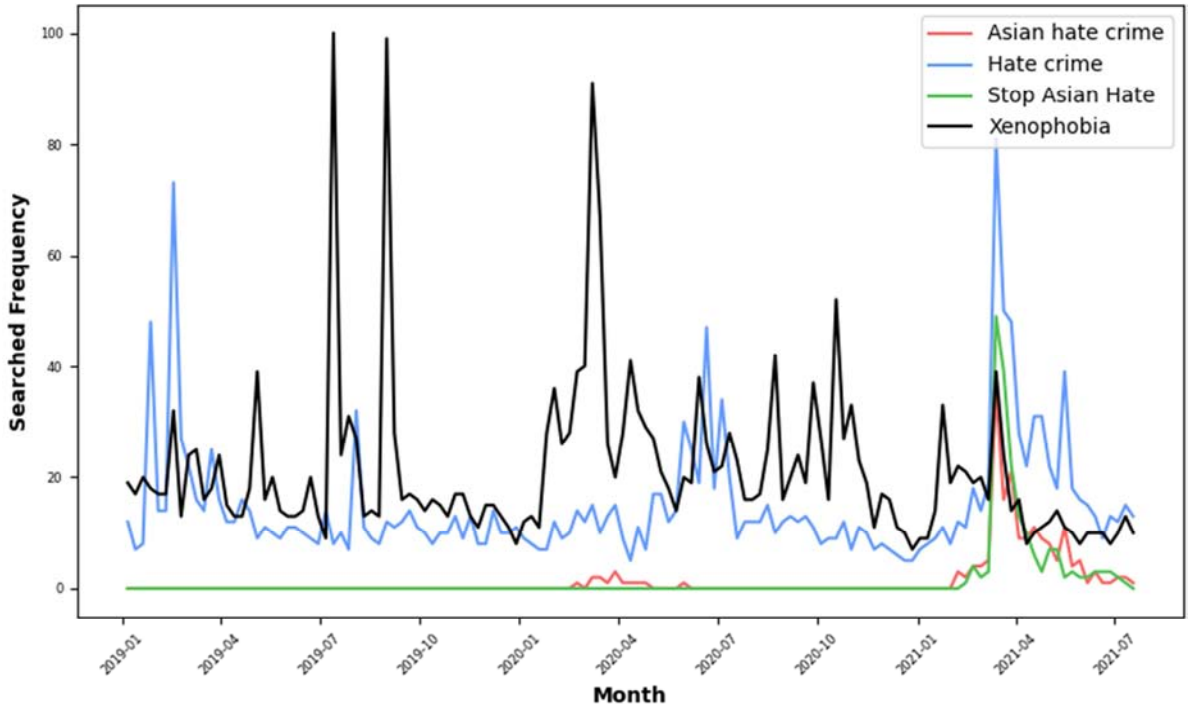

Figure 1. *Asian hate crime related keywords search frequency from 2019 until now.*

2. Results of Racial Discrimination Online

2.1. Keywords Related to Asian Hate

The keyword "Asian Hate Crime" (orange) stayed relatively low until the start of 2021 and began to rise at an unprecedented rate around March of 2021. On the other hand, "Hate Crime" (blue) had lots of ups and downs between 2019 and 2021, but it also reached its highest around March of 2021. "Xenophobia" (purple) also had many ups and downs in its search frequency, but it started to rapidly rise when the pandemic started; the search frequency is gradually decreasing, reaching its lowest

point around July of 2021. "Stop Asian Hate" (green) stayed steadily at the bottom until March of 2021. The date was basically the turning point of most of the keywords' search frequency, as this was the time when the concept of 'Corona-Delta' started to become an actual thing for people, stirring the attitudes of American people towards Asians.

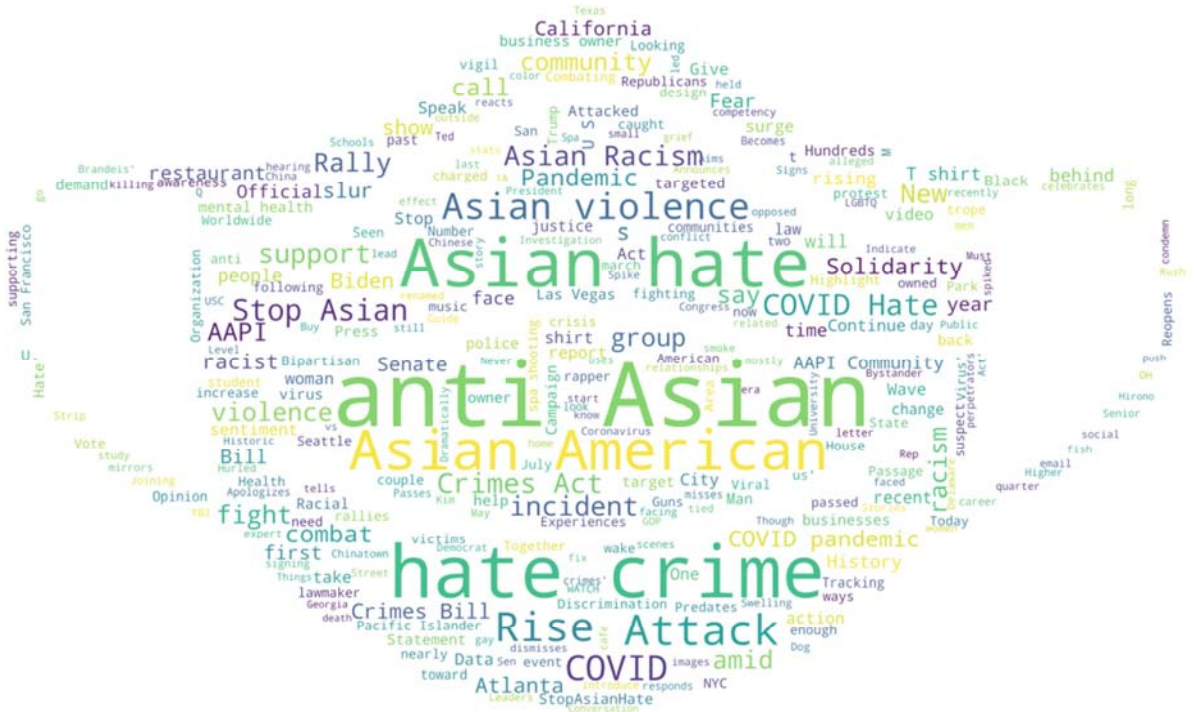

Figure 2. *Word Cloud of the 300 most frequently used words in Google News articles containing the keywords "Anti-Asian," "COVID Hate Crimes," and "Stop Asian Hate"*

The Word Cloud is in the shape of a face mask and contains the 300 most frequently used words in Google News articles containing the keywords "Anti-Asian," "COVID Hate Crimes," and "Stop Asian Hate." The more space the word takes up on the Word Cloud, the higher its frequency of use. The three largest words other than the specific keywords themselves are "Asian hate," "hate crime," and "Asian American." One thing to note is that though there are words directly related to the keywords, such as "Asian violence" and "Stop Asian Hate," there are also words such as "Asian racism" and "racist," as well as "solidarity" and "rally." This shows that this is not only an issue about violent crimes and attacks against Asians, but also a matter of the systemic racism that is still prevalent today and the continuous fight for racial equality, whether that is through physical protests or words of unity against racism and hate against Asians, Asian Americans, and Pacific Islanders in America.

2.2. Asian Hate Crime Victims Before and After COVID-19 Pandemic

In figures 3 and 4, the races are labeled as following: Asians other than labeled separately as A; African Americans as B; Chinese as C; Cambodian as D; Filipino as F; Guamanian as G; Hispanic, Latin, and Mexicans as H; American Indians and Alaskan Natives as I; Japanese as J; Korean as K; Laotian as L; other races not labeled as O; Pacific Islanders as P; Samoans as S; Hawaiian as U; Vietnamese as V; white as W; unknown as X; and Asian Indians as Z.

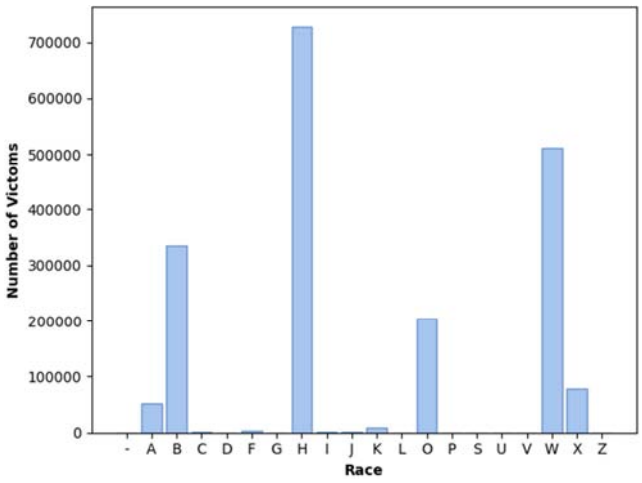

Figure 3. *Number of victims of hate crimes of different races from 2010 to 2019.*

The data from figure 3 shows that 64,791 victims were identified as Asians and additional 652 victims were Pacific Islanders. This totals about 3.407 percent among all races. However, Asian victims, including those from Pacific Islands, increased in percentage in 2020 to 3.916%, showing that Asians and Pacific Islanders have become more vulnerable to crime. 2020 was the year when the COVID-19 started spreading rapidly in the United States, with the first confirmed case in the United States dated January 21, 2020 [10].

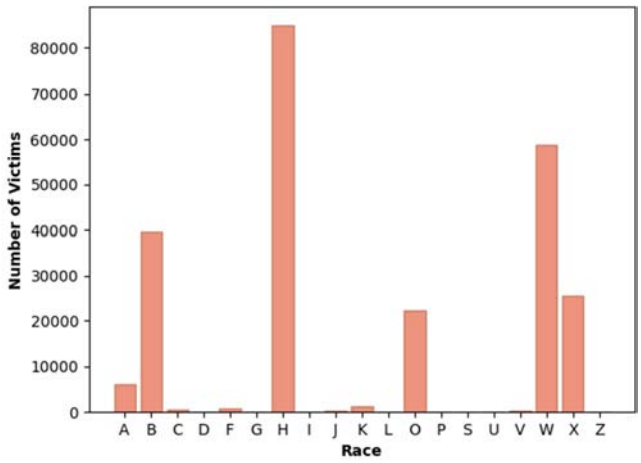

Figure 4. *Number of victims of hate crimes of different races in 2020, retrieved from the police department of the City of LA.*

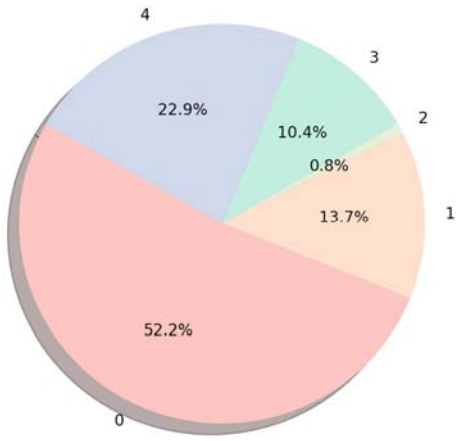

Figure 5. *Ratio of the number of tweets selected randomly with relevance to the key words "Asian hate" before the COVID-19 pandemic.*

2.3. Asian Hate Related Tweets Before COVID-19 Pandemic

With the keyword "Asian Hate," 250 Tweets from before the COVID-19 pandemic began (June 8, 2018-December 31, 2019) were analyzed based on 4 different categories: 0) if the Tweet didn't show hate towards Asians, 1) if it *did* contain hate towards Asians, 2) if it was not in English, 3) if it was racism and/or hate directed at another racial minority, 4) or if it actually *supported* Asians and advocated for the decrease of prejudice, stereotypes, and/or hate. 130, or 52.2% of the Tweets turned out not to be Asian hate, and just happened to contain those two words in it (type 0). These Tweets would generally be entirely unrelated to the discussion of race itself. An example of a Tweet classified as type 0 would be "Who's the legendary Asian midfielder? For me, Park Ji Sung... What a player. But still, I hate the club, not the player," by user *pudseposen*.

34, or 13.7% of the Tweets expressed hate towards Asians (type 1), generally through perpetuating harmful stereotypes or, on occasion, engaging in hate speech. An example of a type 1 Tweet would be "I f***ing hate doing any business with Chinese/Asian People, they wanna charge extra for EVERYTHING." by user *rosangelthebest*.

0.8%, or 2, Tweets were not in English (type 2); both were in French. For example: "sans parler du racisme dont il fait preuve," by user *recklessdragons*.

10.4%, or 26, Tweets expressed racism towards other minorities (type 3), such as African Americans. For example: "Black men are the only group of people that disrespect and tear down their women," by user *jjessiica_*.

Finally, 22.9%, or 57, Tweets supported Asians and Asian Americans and Pacific Islanders (type 4), calling out the use of outdated stereotypes and microaggressions. An example of a type 4 Tweet would be:

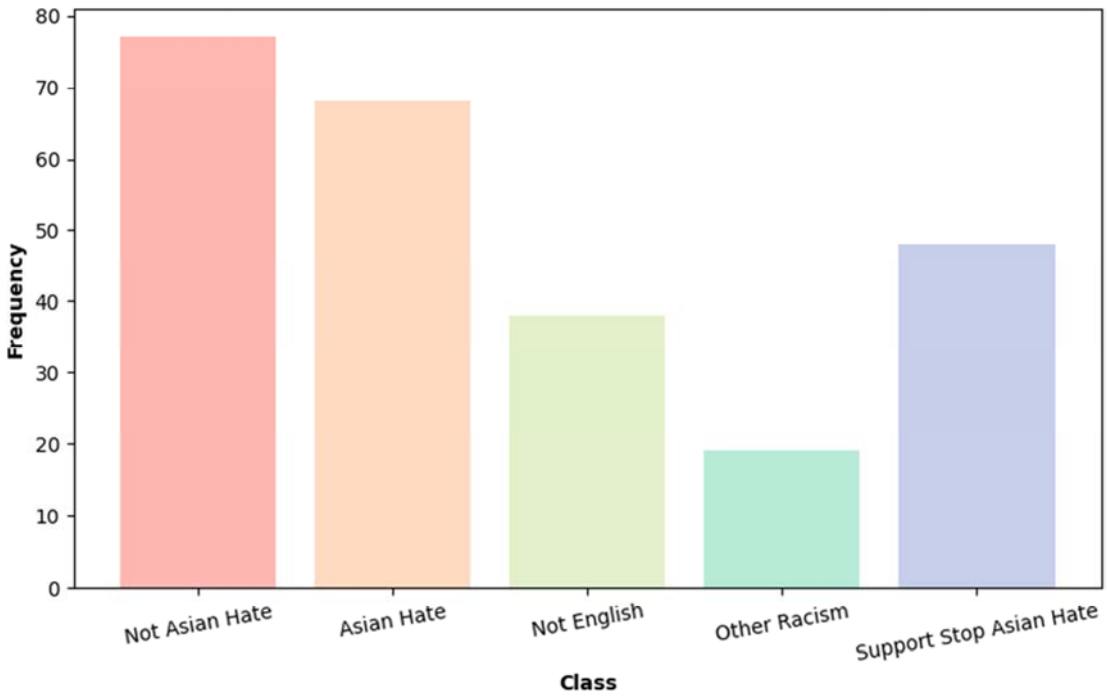

Figure 6. *The number of tweets selected randomly with relevance to the keyword "Asian hate" before the COVID-19 pandemic.*

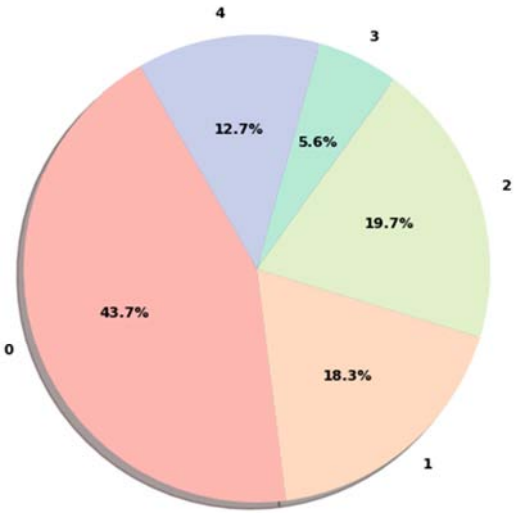

Figure 7. *Ratio of the number of randomly selected tweets with relevance to Asian hate after the COVID-19 pandemic began, represented through a pie graph. Number descriptors below: Key- 0: Not Asian Hate, 1: Asian Hate, 2: Not English, 3: Other racism, 4: Support Asians/Stop AAPI Hate.*

In the figure above (figure 7), with the keyword "Asian Hate," randomly selected 251 tweets have been collected and categorized into 5 different categories: if the tweet is not related to Asian hate, if the tweet is related to Asian Hate, if the tweet is not written in English, if the tweet is racist to other race, or if the tweet supports Asians. 29% (75) of the tweets were revealed not to be racist towards Asians, as they were simply referring to the term "Asian" for discussion of other topics, such as famous celebrities or game related conversations. At the same time, 28.3% (71) of the tweets turned out to be hostile to Asians, as they both directly and indirectly insulted Asians for common stereotypes about Asians. 19.9% (50) of the tweets weren't interpretable, as they were written in another language, and 7.2% (18) of the tweets were being racist to other races by using "Asians" as comparison. 14.7% (37) of the tweets were friendly to Asians, as they praised Asian food/culture, or the people.

2.4. Asian Hate Related Tweets After COVID-19 Pandemic

The randomly collected tweets, shared after the COVID-19 outbreak, represent the public's interest on certain topics on that day. The data was categorized into 5 different categories by their contents: not Asian hate, Asian hate, not English, other Racism, support Asians/Stop AAPI Hate. More than 30% of the collected tweets had relevance to Asian Hate, and 18.3% of the tweets collected were about Stop AAPI Hate. When summing up the number of racism related tweets, either Asian hate or not, they compose more than 35% of total tweets collected. Knowing that such a significant portion of people's opinions online contain sensitive issue related contents, Asian hate issue is a significant societal issue requiring a countermeasure.

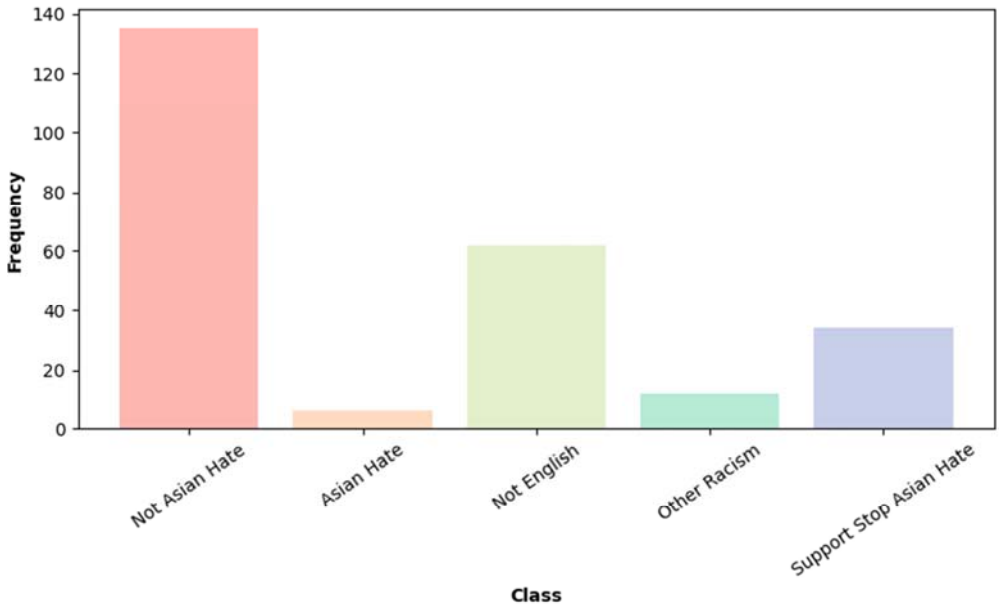

Figure 8. *The number of tweets selected randomly with relevance to the keyword "Asian Hate" after the COVID-19 pandemic.*

In figure 8, 250 tweets have been randomly selected and categorized into five categories based on their contents: Not Asian Hate, Asian Hate, Not English, Other Racism, and Support/Stop Asian Hate. 54%(135) of the tweets were not relevant to Asian Hate. 2.4%(6) of the tweets were directly relevant to Asian Hate. 41%(62) of the tweets were not English. 4.7%(12) of the tweets were pertaining to other Racism. Lastly, 13.5%(34) of the tweets were supporting the Stop of Asian Hate.

The randomly collected tweets, shared after the COVID-19 outbreak, represent the public's interest on certain topics on that day. The data was categorized into 5 different categories by their contents: not Asian hate, Asian hate, not English, other Racism, support Asians/Stop AAPI Hate. More than 30% of

the collected tweets had relevance to Asian Hate, and 18.3% of the tweets collected were about Stop AAPI Hate. When summing up the number of racism related tweets, either Asian hate or not, they compose more than 35% of total tweets collected. Knowing that such a significant portion of people's opinions online contain sensitive issue related contents, Asian hate issue is a significant societal issue requiring a countermeasure. Figures 5-8 shows that the Asian hate was a big problem in online even before COVID-19.

2.5. Correlation of Hate Crimes with Xenophobia

Table 1. *Number of Victims due to Hate Crime Categorized as Gender Identity, Age, Race, Religion, Ethnicity/National Origin, Sexual Orientation, and Disability throughout the year from 2010 to 2019.*

Categories of Hate Crimes	Number of Victims
Gender Identity	99
Age	23
Race	1399
Religion	3035
Ethnicity/National Origin	418
Sexual Orientation	989
Disability	12

The hate crimes listed above are classified by bias categories. Specific bias hate crimes such as race, religion, and ethnicity are greatly influenced by xenophobia, and most of the victims were prone to crime with these categories. Race, religion, and ethnicity make up a total 81.2 percent, with religion alone occupying 50.8 percent. From this we can conclude that xenophobia is a big problem both offline and online. To solve this problem, in the next section, we implement the classification models for xenophobia tweets.

3. Binary Classification Model for Xenophobia Tweets

In this section, we are going to introduce three machines made by ourselves that use three distinct algorithms, which later will be discussed. As smartphones became common, SNS became much more accessible than it did before. Though the increase of accessibility seemed to be beneficial for human society, it didn't turn out the way it was expected to be. One of the biggest problems has been racism in social networks, which has been a problem even before 2016. According to a survey done by a group of researchers in 2016 to randomly selected American adults about racism on Twitter, the survey revealed that more than 80% of the participants has seen a racial discrimination [11]. As time passed, even more racism had been revealed, urging people to take action. The machines that later will be discussed can be helpful tools to collect data not only from Twitter but from other SNS as well, allowing SNS servers or companies to prevent any race-hating comments or posts. To implement the classification model, we use three different classification algorithms: Logistic regression, support vector machine (SVM), and Naive Bayesian.

3.1. Logistic Regression

Logistic regression is a traditional way of handling statistics; it is also useful in machine learning. Logistic regression predicts whether something is true or false based on statistics using the logistic function [12]. For example, a logistic regression of a random person being obese may turn out to be obese, which is true, and not obese, which is false. The probabilities of becoming true or false is decided by the data collected and the input value.

3.2. Support Vector Machine

Support vector machines are another way of handling and classifying data into two types. This method functions as a way where a decision boundary is drawn using extreme data points, and the type of data is selected whether it lies on either one of the zones. The vectors of the decision boundary bounded by two vectors crossing the extreme data points are called support vectors. These support vectors are optimized through multiple steps for enhanced accuracy of classification [13].

3.3. Naive Bayes

Naive Bayes is an algorithm, which is based on Bayes algorithm, that is a very useful tool to use when dealing with thousands of data sets. It becomes even more useful mainly when predicting what class an unknown set of data would belong to. It has three branches: Gaussian Naive Bayes, Multinomial Naive Bayes, and Bernoulli Naive Bayes, in which all of them fall under the category of conditional probability. Out of all its strengths, its strongest side is that it is extremely fast and easy to understand. Additionally, it doesn't require as much training data, making it one of the most popular algorithms among people [14, 15].

Table 2. *Table of counts of classification.*

Class	Count
0 (not xenophobic)	1385
1 (xenophobic)	6
2 (not English)	28

The data in Table 2 have been collected by using the xenophobia-classifying machine (made by ourselves) to collect around 1,500 tweets that were posted on March of 2021 from twitter in which contained the keyword "COVID," then classify them into different categories. The number 0 was used to classify tweets that were neutral/positive toward foreigners, or not xenophobic. At the same time, 1 was used to classify tweets that were offensive towards other [racial] groups, often those that were insulting other countries' cultures, trends, or generally the people themselves, etc. The number 2 was used to classify tweets that were written in other languages. Despite the fact that the keyword, "COVID," didn't have any correlation with any racial group, there still were some tweets that were hostile to certain racial groups. Even more, twitter isn't the perfect reflection of how people today feel toward foreigners, meaning that there can

be a higher percentage of people that are xenophobic, which is area that people must work on.

3.4. Machine Learning Models to Classify Xenophobia Tweets

Xenophobia is the fear of strangers or foreigners. While analyzing the data from Twitter, the tweets were labeled into three classes based on their information: 0 - not xenophobia, 1 - xenophobia, 2- not English. The tweets were marked as one if they presented hate comments or criticism against different races. One example of a xenophobic remark will be a derogatory remark about a specific culture. Xenophobia is a highly severe problem in social media nowadays. Therefore, to lower the rate of xenophobic remarks in social media, particularly Twitter, classification models were created from this analyzed data. These classification models include Logistic Regression, Support Vector Machine, and Naive Bayes, and they were coded to efface any xenophobic remarks from getting posted on Twitter. These programs prevent people from posting any xenophobic comments on Twitter by determining if it is an insulting or not insulting comment.

As shown in Figure 9, both the Logistic Regression and Support Vector Machine models had accuracies of around 0.9913, while the Naive Bayes model had an accuracy of approximately 0.9970. The accuracy rate around 99.45 ±0.25% is a percentage that is safe to say that it's pretty much perfectly accurate. Even more, the models are not only capable of classifying posts uploaded in twitter, but even comments uploaded in YouTube or Instagram, or generally any SNS websites one can think of, enabling the companies, or even individual user, of the models to collect much more data from a much wider range in a shorter time than before.

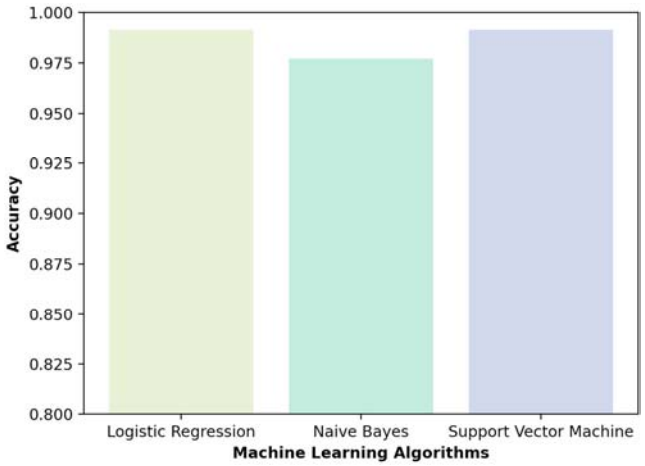

Figure 9. *The accuracy of each machine learning algorithm in classifying data for xenophobia.*

4. Conclusion

This paper analyzed how the COVID-19 pandemic impacted the racism rate towards Asians, Asian-Americans, and Pacific Islanders. With the coronavirus first being identified in Wuhan, China, xenophobia, and racism toward Asians increased significantly. This paper mainly focused on online hate towards Asians before and after COVID-19. The comments on SNS platforms, such as Twitter, with the keyword "Asian hate" or "anti-Asian" were analyzed to determine the increased rate of racism towards Asians after the pandemic. From the EDA, the paper found that the racial related crime is one of the serious problems among the different types of hate crimes. Moreover, classification models were created to lower the derogatory or xenophobic remarks prevalent in social media. Classification models included Logistic Regression, Support Vector Machina, and Naive Bayes, and these models were coded to delete any xenophobic comments from getting posted. The model has average of 90% accuracy on classifying the racial contexts. These classification models have the potential to significantly lower online hate towards any group. For the future work, we are going to implement the model that can classify the xenophobic context in various social media platform.

References

[1] Gover, A., Harper, S., & Langton, L. (2020). Anti-Asian hate crime during the COVID-19 pandemic: Exploring the reproduction of inequality. American Journal of Criminal Justice, 45 (4), 647-667.

[2] Brendan Lantz, and Marin R. Wenger. (2021, August). Are Asian Victims Less Likely to Report Hate Crime Victimization to the Police? Implications for Research and Policy in the Wake of the COVID-19 Pandemic, Crime & Delinquency (CAD).

[3] Hitman, Gadi & Harel, Dror. (2016). Hate Crimes—Methodological, Theoretical & Empirical Difficulties—A Pragmatic & Legal Overview. Journal of Cultural and Religious Studies. 4. 10.17265/2328-2177/2016.01.001..

[4] Tavernise, S., & Oppel, R. A. (2020, March 23). Spit On, Yelled At, Attacked: Chinese-Americans Fear for Their Safety. The New York Times. https://www.nytimes.com/2020/03/23/us/chinese-coronavirus-racist-attacks.html.

[5] Martin, A. (2021, July 15). Why is it so difficult to stop abuse on social media? Sky News. https://news.sky.com/story/why-is-it-so-difficult-to-stop-abuse-on-social-media-12354192.

[6] Shimizu, K. (2020, February 11). 2019-nCoV, fake news, and racism. The Lancet. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30357-3/fulltext.

[7] Afolabi, Oyeronke & Holder, Raymond. (2021). Social Media and Racism in 21 st Century America: A Case Study of Twitter. Merriam-Webster. (n.d.). Xenophobia vs. racism: Explaining the difference. Merriam-Webster.

[8] AJMC. (2021, January 2). A Timeline of COVID-19 developments in 2020. AJMC. https://www.ajmc.com/view/a-timeline-of-COVID19-developments-in-2020.

[9] Anderson, M. (2020, August 20). Social media conversations about race. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2016/08/15/social-media-conversations-about-race/.

[10] https://www.MachineLearningMastery. (2020, April 7). 4 Types of Classification Tasks in Machine Learning. Retrieved August 5, 2021, from Machine Learning Mastery website: https://machinelearningmastery.com/types-of-classification-in-machine-learning/.

[11] Rohith Gandhi. (2018, June 7). Support Vector Machine — Introduction to Machine Learning Algorithms. Retrieved August 5, 2021, from Medium website: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.

[12] Peng, Joanne & Lee, Kuk & Ingersoll, Gary. (2002). An Introduction to Logistic Regression Analysis and Reporting. Journal of Educational Research - J EDUC RES. 96. 3-14. 10.1080/00220670209598786.

[13] Kaviani, Pouria & Dhotre, Sunita. (2017). Short Survey on Naive Bayes Algorithm. International Journal of Advance Research in Computer Science and Management. 04.

[14] Wibawa, Aji & Kurniawan, Ahmad & Murti, Della & Adiperkasa, Risky Perdana & Putra, Sandika & Kurniawan, Sulton & Nugraha, Youngga. (2019). Naïve Bayes Classifier for Journal Quartile Classification. International Journal of Recent Contributions from Engineering, Science & IT (iJES).

[15] Rish, Irina. (2001). An Empirical Study of the Naïve Bayes Classifier. IJCAI 2001 Work Empir Methods Artif Intell. 3.