Medicon Engineering Themes
Volume 3 Issue 4 October 2022
Research Article

Identity Access Management in Metaverse

Manasi Chhibber[1]*, Sarthak Agarwal[1]* and Amit Dubey[2]

[1]B. Tech AI, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India

[2]National Security Expert, Cyber Forensics Head: Red Teaming, Tech Mahindra, India

***Corresponding Author:** Manasi Chhibber and Sarthak Agarwal, B. Tech AI, Amity School of Engineering & Technology, Amity University, Noida, Uttar Pradesh, India.

Received: August 26, 2022; **Published:** September 22, 2022

Abstract

Metaverse is a virtual apace where people can enter and interact with other users in a computer-generated surroundings. It is an integrated system of 3-dimensional virtual worlds. These are preferably accessed via a VR headset. Users' eye movements, their voice commands and feedback controllers are mainly used to navigate around the metaverse. In this way they are able to fully immerse themselves in the virtual world and simulate their presence. With the advent of this new technology, challenges are bound to occur. Most importantly challenges related to security of users need to be dealt with. Every person entering and interacting in the metaverse must be a verified user so that identity frauds do not happen, hence keeping the users safe from malicious attacks. To deal with this issue, we have come up with two features, first, an avatar facial tracking based and second, a chat based. By using NLP and Deep Learning, we were able to 73.77% uniquely identity a person in a WhatsApp chat group with other people. For the facial recognition-based security feature, we have just laid out our theoretical approach. In future, we're planning to make use of other features as well such as speech pattern recognition, body positioning retrieval, retinal movement tracking, etc. for allowing only authorized users to enter and interact in the metaverse.

Keywords: metaverse; avatar security; Natural Language Processing; Deep Learning

Introduction

In his 1992 science fiction book Snow Crash, author Neal Stephenson originally introduced the term "metaverse." Stephenson described the metaverse as a vast digital universe that could coexist with the physical world in which we all live. The metaverse is a 3D representation of the Internet and general computing. The majority of interactions back when these two technologies (the internet and computing) originally arose were text-based. Then they gradually shifted to a media-based model (photos, videos, livestreams). 3D user interfaces and experiences are the next level up. Second, the metaverse may be seen to always be contained within a computer and the internet if we consider a mobile phone to be like carrying a computer about in our pocket. The metaverse will grow to be much more widely dispersed, democratic, fluid, and diversified as it grows. Experts predict that the metaverse will advance virtual reality by enabling users to float into the virtual world and perform tasks like hosting events and even getting married using digital avatars.

Metaverse are prevalent in the gaming world, but nowadays concerts are being held in the metaverse by musicians and entertainment companies as well. Top teams like Manchester City are developing virtual stadiums so that spectators can watch games and, presumably, buy virtual goods. The sports business is following suit. Online education and governmental services will likely present the metaverse with its most far-reaching prospects.

Uses of the metaverse

The metaverse is hailed as a key role in boosting the digital economy due to its high value projection. The metaverse will boost the global economy's main growth engine, the digital economy. Although the metaverse is already being predicted as the future of gaming, fashion, and even parties, experts contend that its best-case application will probably be in the field of education. A fast buck may be made by investing in virtual nations that don't exist in reality, but the metaverse will only be really valuable when it is put to use in ways that improve people's lives.

Is metaverse ready?

Not right now. Interoperability is the largest barrier to Ball and Zuckerberg's metaverse becoming a reality. Interoperability is essential if the metaverse is to advance the internet to its next level, yet the barriers are so great that they appear insurmountable. There are technological difficulties, such as transferring assets across graphics engines and rendering them accurately on a dizzying array of hardware combinations. There are also legal and financial difficulties, such as getting several companies to agree not to fence off their gardens and getting around intellectual property rights. It's a lot more difficult than, say, deciding on a standard for hypertext links.

Beyond that, people need to be convinced that they want this. The technology we use to access these worlds must be at least as portable and pleasant to use as a smartphone, or it will appear to be a step back from the mobile internet it is meant to replace. Additionally, one must wonder how strong the desire to spend time in such a virtual environment truly is despite the fact that the science-fiction attractiveness of it may appear clear on the surface.

Metaverse is a growing platform on the internet where people interact with each other with the help of their avatars but there are some security threats in the metaverse.

Security Vulnerabilities

Along with a wide range of benefits and applications of metaverse, there are certain things that could possess a challenge in smooth and secure functioning of the technology. Here are a few typical security issues that could arise in metaverse realms:

- *Moderation difficulties*. In the majority of the metaverses, there is no access to aid or support. For instance, nonfungible token theft may render a user helpless.
- *Users' vulnerabilities*. Heavy-duty devices with plenty of software and memory include VR and AR headsets. They are also easy prey for both intentional and accidental hacking. Furthermore, location fraud and gadget manipulation allow criminals to join the metaverse, assume users' identities, and wreak havoc.
- *Communications between users*. Because the goal of the metaverse experience is to facilitate user-to-user contact, the foundation of these connections is trust and exchange of goods and services. A one evil guy can inflict a lot of harm. Scale-based moderation is essential and must be addressed.
- *Privacy*. There are no laws governing the metaverse, and the requirement for data collecting for a genuinely customized immersive experience necessitates privacy breach. Most of the time, users are unaware of the amount of data they are supplying. Additionally, because virtual experiences are borderless, unlike other legislation like GDPR that include obligations for provincial sovereignty, the platform owner and the property owners are ultimately responsible for maintaining privacy.
- *Authentication*. It might be difficult to verify that someone is who they claim to be. How can you be certain that the person you are interacting with is who they say they are? Consider telemedicine as an example. How can a patient tell whether the person they speak to is a health care provider? How can a landowner verify a doctor's qualifications before letting them to practice?
- *Compromising of the access point.* As a headset is generally used to enter the VR metaverse, a breach of the headset endpoint might result in total control of the user's avatar.

Literature Review

Although very little research has been done in the field of the metaverse, but there are some studies that have made use of behavioural analysis to detect compromised accounts on social media.

Egele et al. performed statical anomaly detection to detect compromised profiles by detecting a drastic change in the behavior based on user information present in Facebook and Twitter data. Viswanath and other researchers used P.C.A. on user's Facebook history to detect changes in behavior patterns of users [20]. Vandam and other researchers studied about selected characteristics of Twitter accounts, such as the frequency of hashtags or mentions in tweets for detecting compromised accounts on Twitter [20]. Karimi et al. [20] used LSTM networks to learn to identify chronological irregularities by capturing chronological relationships inside user accounts. Seyler et al. [20] used semantic text analysis to detect compromised social media accounts with the help of uni-gram language models and KL-divergence measures. Almozaini et al [22] created a framework for the detection of silent hacking of users' accounts on social media using Deep learning models to detect behavioral deviations of users and prove the fallouts of classical machine learning algorithms for silent hack detection.

Methodology

We have worked on two types of security features for metaverses. One is Avatar Facial Mapping based and another one is Chat-based authentication feature. The working of the facial mapping-based security feature has only been theoretically explored whereas, the chat-based security feature was built as a PoC with the help of WhatsApp chats.

Avatar Facial Mapping based Security Feature

A solution is required to secure & verify the identity of the Avatar to prevent them from getting duplicated and being used for various malicious purposes that may harm the reputation of the person in the metaverse. To keep a track of uniqueness of all the users, we can use IP address, MAC address and other digital addresses but they can also be spoofed easily in the current digital world. So, this approach aims at maintaining the uniqueness and preventing this forgery by tracking the facial expressions of the Avatar. These are exclusive as they are just a copy of a real person's expressions that are being imitated by another virtual figure. Hence, for avatar facial mapping, we can proceed in two ways:

Corrective Approach

Initially we will make an Avatar and use it to give a speech (10-30 sec). The same avatar will be used by 5 different people for the same speech then we will be capturing the facial expressions of an avatar while it speaks. After we create this dataset of facial expressions vs words for each person, we will test whether this data is sufficient for differentiating or not. Here either we can either map facial expressions to each word in the speech eliminating the problem of speaking rate that can vary person to person or we can take speech data extrapolate or reduce speech of person to a fixed time like 10 sec also to eliminate the problem of speaking rate and then capture the expressions. Therefore, we will be tracking facial expressions of an Avatar while it is in use.

Preventive Approach

Another thing we can do is to train an agent/model while making an avatar i.e., we ask the user to record 5-6 video samples helping us in collection of data regarding his natural facial expressions as well his accent of speech, speaking rate etc. After we get this data, we train our model and track the behavior regularly. As soon as we see some unusual activity, we verify it with some additional collected data of the user and if something is found malicious, we block that avatar. In the end giving a blue tick against the avatar after it has been used for some time for authenticity. We can keep the labelled data set of the user's facial expressions and voice w.r.t different sentiments. By this, we can identify whether the user has spoken the same sentence in a different mood or is it an unauthorized person.

Given below are the steps in the methodology we idealized:

1. We are going to use avatar maker UMA plugin in Unity to capture our facial expression and put it inside our avatar.

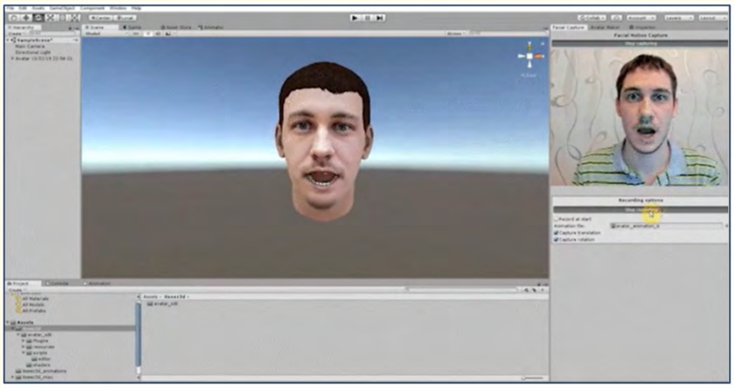


Figure 1: Making Avatars using Unity.

2. We are going to use Mediapipe library to mark 50 landmarks on the avatar's face generated by unity so that we can capture its facial movements while speaking.

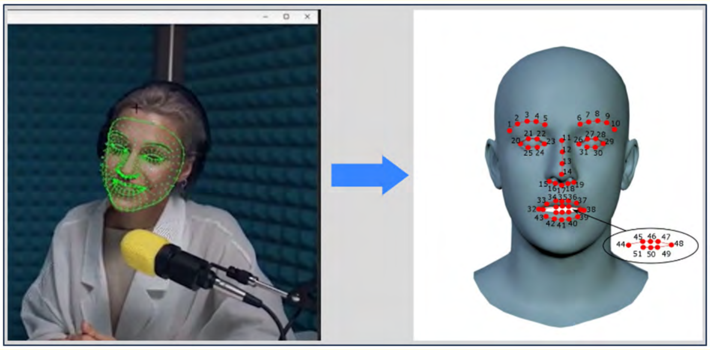


Figure 2: Mapping Landmarks from Face Mesh.

3. We will convert the 3D coordinates of each of these landmarks into relative distances with respect to the central point of all these face landmarks at certain time frames, so that our computations can be simplified.

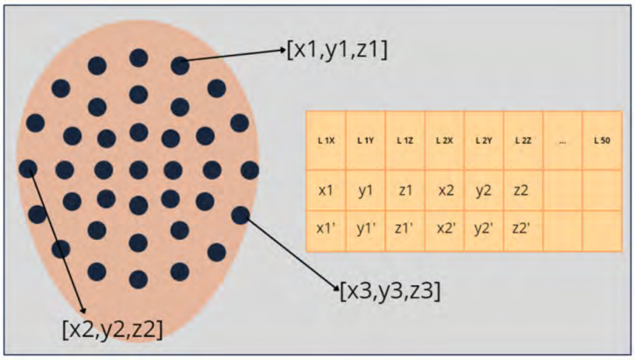


Figure 3: Recording 3D coordinates of all the Landmarks.

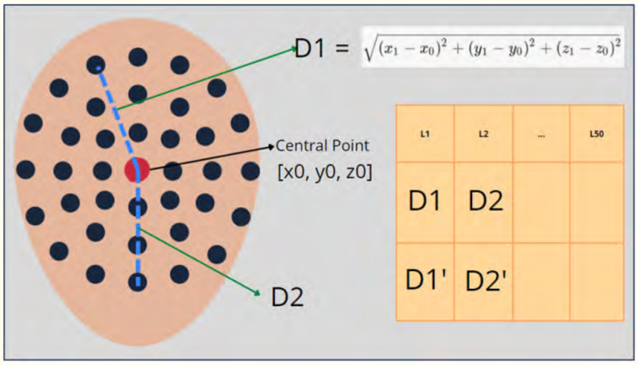


Figure 4: Converting 3D coordinates to Relative Distances.

4. We will use PCA to apply dimensionality reduction on our dataset to optimize the model training. After some computations the final dataset would be ready.

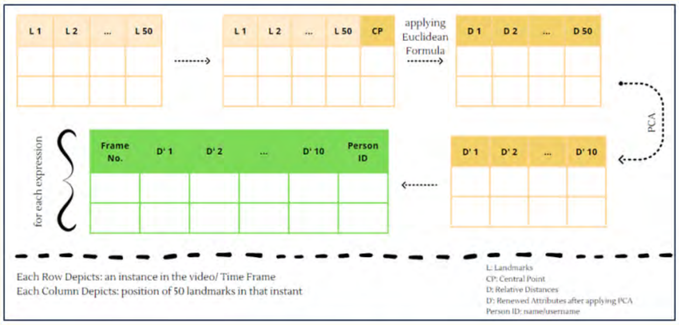


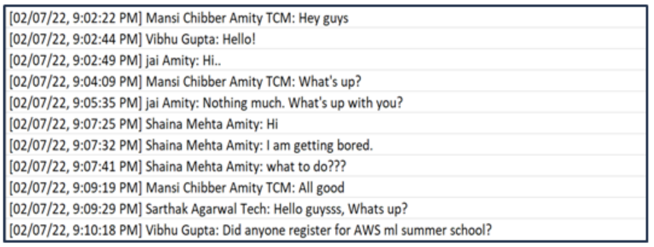
Figure 5: Process to prepare the Dataset.

5. Finally, we'll use some ML models or Neural Networks to train on the data and predict whether a user in the metaverse in authentic or not.

In this way, we will be able to distinguish and authenticate metaverse users even if they are speaking the same sentence or expressing the same emotion.

Chat based Security Feature

1. The dataset was prepared first. In this project, we created the dataset using WhatsApp Group Chat messages. These messages were exported in the form of Microsoft Excel Workbook. It had a total of 988 records. In the chat, total 5 people participated, and the aim was to test whether an ML model can identify a person named 'Vibhu' or not.



```
[02/07/22, 9:02:22 PM] Mansi Chibber Amity TCM: Hey guys
[02/07/22, 9:02:44 PM] Vibhu Gupta: Hello!
[02/07/22, 9:02:49 PM] jai Amity: Hi..
[02/07/22, 9:04:09 PM] Mansi Chibber Amity TCM: What's up?
[02/07/22, 9:05:35 PM] jai Amity: Nothing much. What's up with you?
[02/07/22, 9:07:25 PM] Shaina Mehta Amity: Hi
[02/07/22, 9:07:32 PM] Shaina Mehta Amity: I am getting bored.
[02/07/22, 9:07:41 PM] Shaina Mehta Amity: what to do???
[02/07/22, 9:09:19 PM] Mansi Chibber Amity TCM: All good
[02/07/22, 9:09:29 PM] Sarthak Agarwal Tech: Hello guysss, Whats up?
[02/07/22, 9:10:18 PM] Vibhu Gupta: Did anyone register for AWS ml summer school?
```

Figure 6: Collected Messages.

2. The implementation of this project was done using Python programming language in Google Colab. This project runs on the Google Cloud Platform.
3. After loading the dataset, standard preprocessing steps were done. The dataset was refined, null values were removed, label encoding was done, etc.
4. The NLP pipeline was setup. All the steps under natural language processing were performed. In the end, we got a dataset of clean texts only.
5. Vectorization was performed and the dataset was split for training and testing.
6. First, a Naïve Bayes model was trained and tested for accuracy. Later, a Neural Network model was prepared to test its functioning in comparison to the classic ML model.

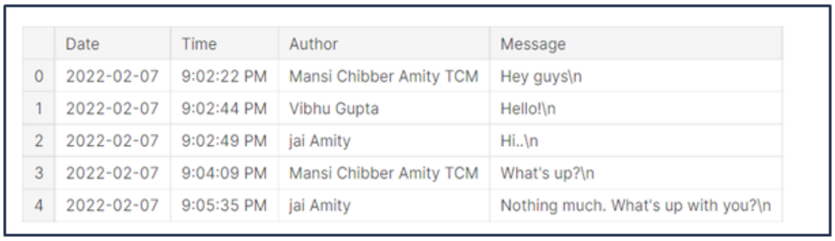
In this way, we were able to accurately assess the performance of both of these models and 73.77% correctly predict whether the messages were from 'Vibhu' or an imposter.

Experimentation and Analysis

The step-by-step making and implementation of the chat-based security feature has been explained below:

Data Pre-Processing and EDA

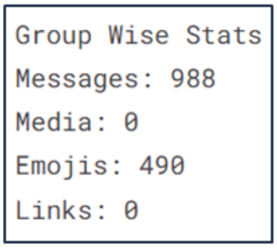
1. Loaded the dataset into a Pandas data frame.



	Date	Time	Author	Message
0	2022-02-07	9:02:22 PM	Mansi Chibber Amity TCM	Hey guys\n
1	2022-02-07	9:02:44 PM	Vibhu Gupta	Hello!\n
2	2022-02-07	9:02:49 PM	jai Amity	Hi..\n
3	2022-02-07	9:04:09 PM	Mansi Chibber Amity TCM	What's up?\n
4	2022-02-07	9:05:35 PM	jai Amity	Nothing much. What's up with you?\n

Figure 7: Loading the Dataset.

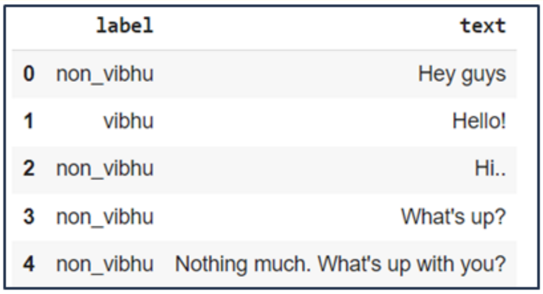
2. Removed the unnecessary punctuations, time stamps, URL links, and unnecessary messages from the data.



```
Group Wise Stats
Messages: 988
Media: 0
Emojis: 490
Links: 0
```

Figure 8: Overall Statistics.

3. Classified the data into the target person ('Vibhu') and the Non-Target person ('Non-Vibhu') by performing label encoding and checked imbalance in between the two classes.



	label	text
0	non_vibhu	Hey guys
1	vibhu	Hello!
2	non_vibhu	Hi..
3	non_vibhu	What's up?
4	non_vibhu	Nothing much. What's up with you?

Figure 9: Dataset after Label Encoding.

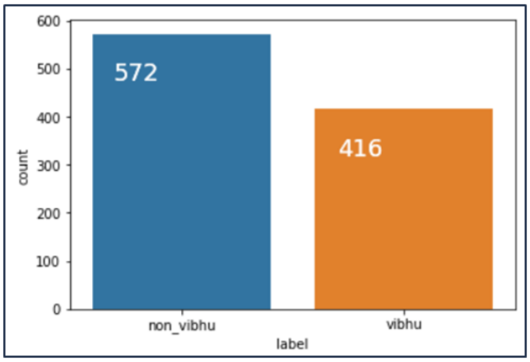
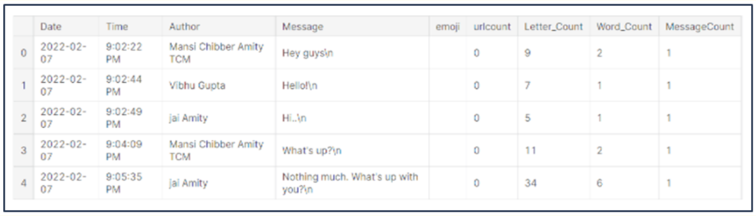


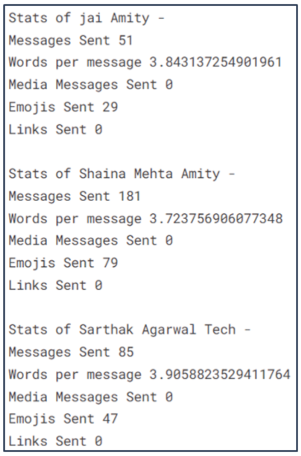
Figure 10: Visualizing the imbalance in the Target Classes.

4. Analyzed the dataset by plotting various graphs.



	Date	Time	Author	Message	emoji	urlcount	Letter_Count	Word_Count	MessageCount
0	2022-02-07	9:02:22 PM	Mansi Chibber Amity TCM	Hey guys\n		0	9	2	1
1	2022-02-07	9:02:44 PM	Vibhu Gupta	Hello!\n		0	7	1	1
2	2022-02-07	9:02:49 PM	jai Amity	Hi..\n		0	5	1	1
3	2022-02-07	9:04:09 PM	Mansi Chibber Amity TCM	What's up?\n		0	11	2	1
4	2022-02-07	9:05:35 PM	jai Amity	Nothing much. What's up with you?\n		0	34	6	1

Figure 11: Exploring the contents of the Dataset.



```
Stats of jai Amity -
Messages Sent 51
Words per message 3.843137254901961
Media Messages Sent 0
Emojis Sent 29
Links Sent 0

Stats of Shaina Mehta Amity -
Messages Sent 181
Words per message 3.723756906077348
Media Messages Sent 0
Emojis Sent 79
Links Sent 0

Stats of Sarthak Agarwal Tech -
Messages Sent 85
Words per message 3.9058823529411764
Media Messages Sent 0
Emojis Sent 47
Links Sent 0
```

Figure 12: Individual User Stats.

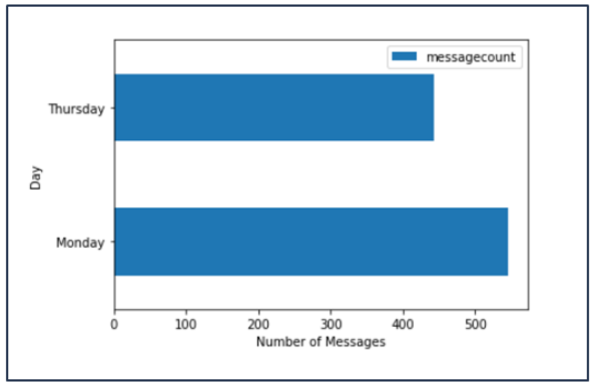


Figure 13: Bar Graph denoting the no. of Messages sent each day.

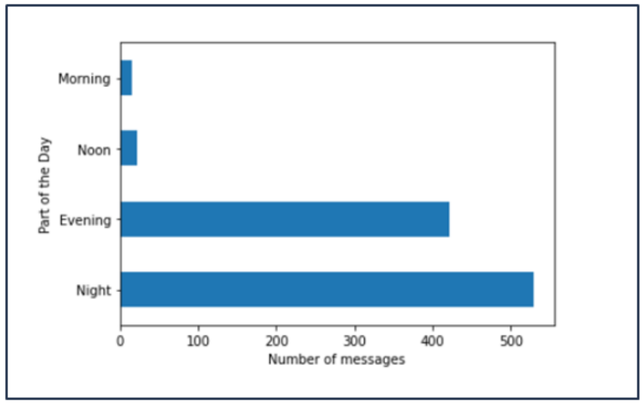


Figure 14: Bar Graph denoting Message Count during various hours of the day.

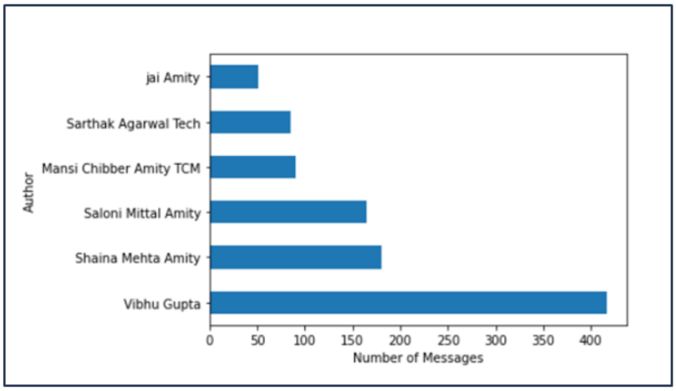


Figure 15: Bar Graph denoting the no. of Messages sent by each person.

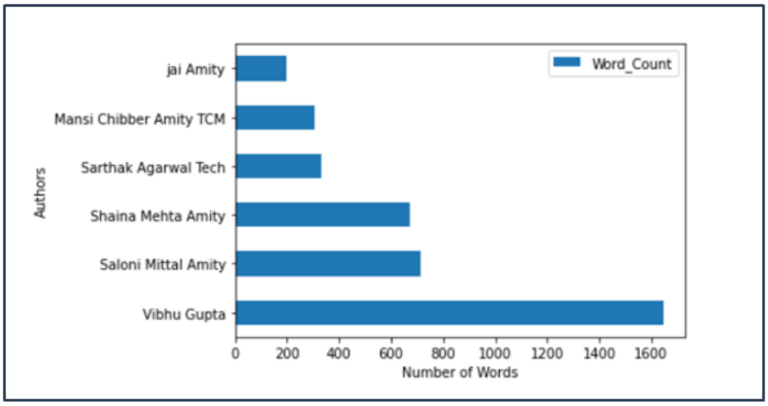


Figure 16: Bar Graph denoting Total Word Count of messages sent by each person.

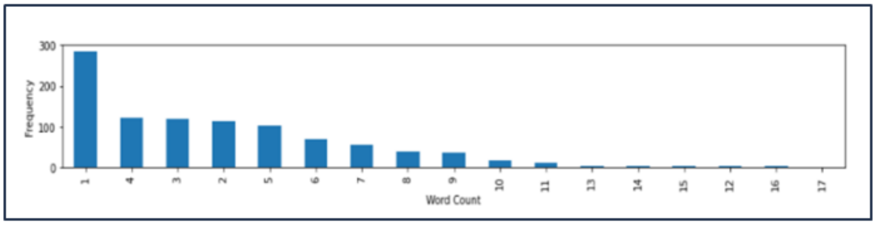


Figure 17: Bar Graph denoting Most Common Word Count in a Message indicating that people prefer Short Messages.

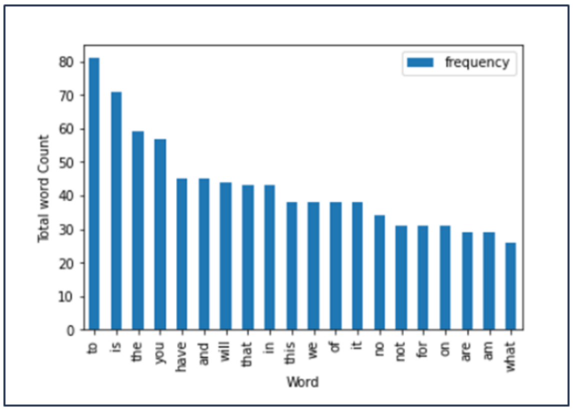


Figure 18: Bar Graph denoting Most Frequently used Words while Chatting.

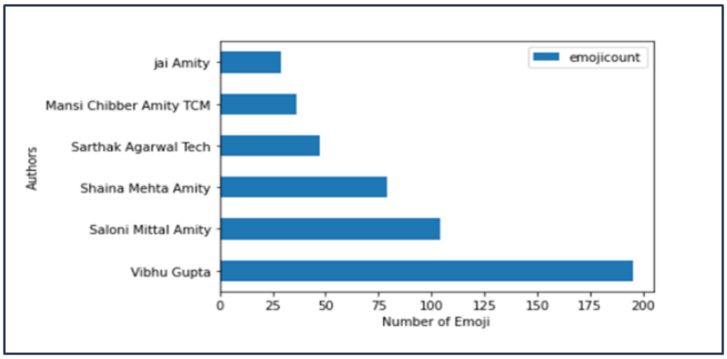
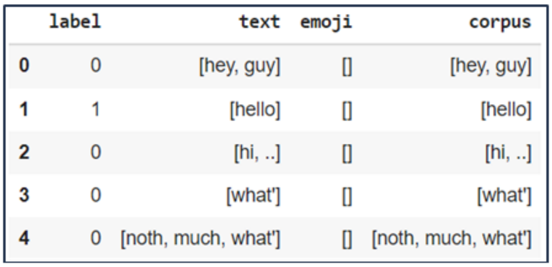


Figure 19: Bar Graph denoting no. of Emoji sent by each user.

Setting up the NLP Pipeline

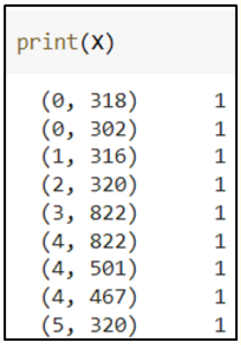
1. A function was drafted to remove stop words, tokenize the texts and stem all the words in between them. All the texts were looped through, and this function was called to clean the messages.



	label	text	emoji	corpus
0	0	[hey, guy]	[]	[hey, guy]
1	1	[hello]	[]	[hello]
2	0	[hi, ..]	[]	[hi, ..]
3	0	[what']	[]	[what']
4	0	[noth, much, what']	[]	[noth, much, what']

Figure 20: Cleaned Messages.

2. All these words were converted into vectors so that they could be understood and computed by the mathematical model.



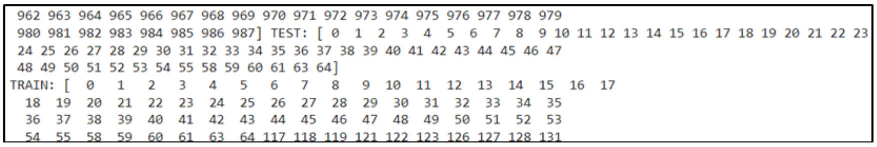
```
print(X)

  (0, 318)	1
  (0, 302)	1
  (1, 316)	1
  (2, 320)	1
  (3, 822)	1
  (4, 822)	1
  (4, 501)	1
  (4, 467)	1
  (5, 320)	1
```

Figure 21: Training Set after Vectorization.

Training Models and Checking Accuracy

1. Stratified K-fold method was used to randomly sample the dataset over and over again and select the best possible training and testing sets.



```
 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979
 980 981 982 983 984 985 986 987] TEST: [ 0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
 48 49 50 51 52 53 54 55 58 59 60 61 63 64]
TRAIN: [   0    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17
   18   19   20   21   22   23   24   25   26   27   28   29   30   31   32   33   34   35
   36   37   38   39   40   41   42   43   44   45   46   47   48   49   50   51   52   53
   54   55   58   59   60   61   63   64  117  118  119  121  122  123  126  127  128  131
```

Figure 22: Training and Testing Sets.

2. First, a Naïve Bayes model was trained. It was able to provide us 58.95% accurate results.

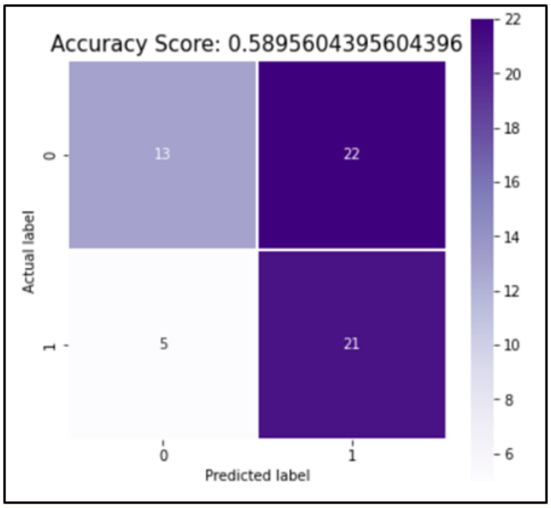
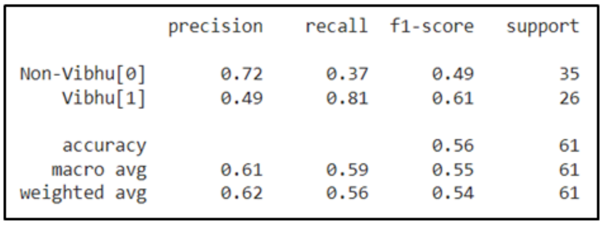


Figure 23: Confusion Matrix after training using Naïve Bayes Model.



```
              precision    recall  f1-score   support

 Non-Vibhu[0]       0.72      0.37      0.49        35
     Vibhu[1]       0.49      0.81      0.61        26

     accuracy                           0.56        61
    macro avg       0.61      0.59      0.55        61
 weighted avg       0.62      0.56      0.54        61
```

Figure 24: Classification Report after training using Naïve Bayes Model.

3. Undersampling and oversampling were performed to vanish out the imbalance in the target classes. Undersampling helped us boost the accuracy to 62.80%, whereas after oversampling, the results remained the same.

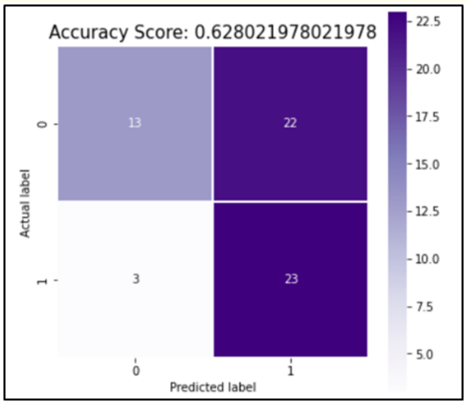
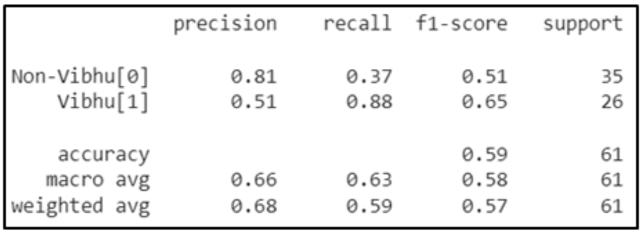


Figure 25: Confusion Matrix after Undersampling.



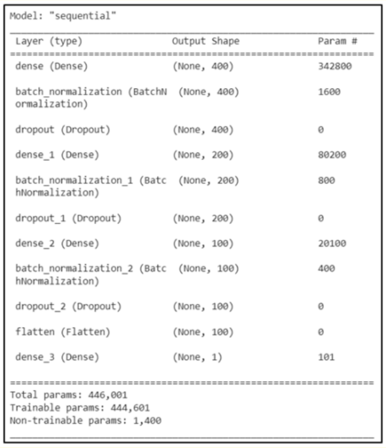
```
              precision    recall  f1-score   support

Non-Vibhu[0]       0.81      0.37      0.51        35
    Vibhu[1]       0.51      0.88      0.65        26

    accuracy                           0.59        61
   macro avg       0.66      0.63      0.58        61
weighted avg       0.68      0.59      0.57        61
```

Figure 26: Classification Report after Undersampling.

4. Second, a sequential Neural Network architecture was prepared for training the data on. It was able to provide 73.77% accurate results.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 400)               342800

 batch_normalization (BatchN  (None, 400)              1600
 ormalization)

 dropout (Dropout)           (None, 400)               0

 dense_1 (Dense)             (None, 200)               80200

 batch_normalization_1 (Batc  (None, 200)              800
 hNormalization)

 dropout_1 (Dropout)         (None, 200)               0

 dense_2 (Dense)             (None, 100)               20100

 batch_normalization_2 (Batc  (None, 100)              400
 hNormalization)

 dropout_2 (Dropout)         (None, 100)               0

 flatten (Flatten)           (None, 100)               0

 dense_3 (Dense)             (None, 1)                 101

=================================================================
Total params: 446,001
Trainable params: 444,601
Non-trainable params: 1,400
_________________________________________________________________
```

Figure 27: Neural Network Model Architecture.

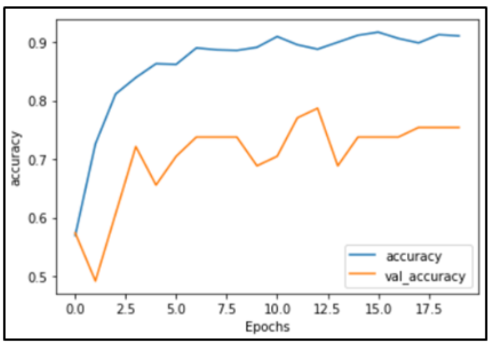


Figure 28: Accuracy Trend Line.

Results and Discussion

In this section, the comparative study of models is presented on the basis of their accuracy. In this project, the following two classification models were selected for identifying the forgery of metaverse avatars:

1. Gaussian Naïve Bayes.
2. Neural Networks.

For Gaussian Naïve Bayes, the model was trained and accuracy that could be obtained using it was predicted. For Neural Networks architecture, the model was trained on 20 epochs having the batch size of 10 each. Adam optimizer was used, and Binary Cross-Entropy loss was taken as the loss function. The accuracy comparison has been listed below.

Model	***ROC Accuracy (%)***
Gaussian Naïve Bayes	58.95604395604396
Convolutional Neural Networks	73.77049326896667

Table 1: Accuracy of Different Models.

Overall, it can be seen that Neural Networks gave better accuracy than Gaussian Naïve Bayes, so it is the best model selected for the purpose. Hence, the final accuracy of the selected model will be 73.77%. This model will be used for further development of security features in the metaverse.

Conclusion and Future Scope

This project deals with the detection of forgery of metaverse avatars using behavioral analysis of text data using machine learning and deep learning algorithms. The comparative study between classical machine learning algorithms and Artificial Neural Networks has been presented. It was found that Neural Networks gave better accuracy than classical machine learning algorithms of about 73.77049326896667 percent. This model will be useful for preventing the forgery of metaverse avatars in the future.

Since it is ongoing research, more data collection is required for more accurate behavioral analysis to get better results, and deployment of the model is also required. Moreover, research on forgery detection on metaverse avatars based on voice data and facial expressions of the avatar is still going on. Researchers are trying their best to make the experience of the metaverse for the users secure and safe.

References

1. What is the metaverse, and what can we do there?.
2. What Is the Metaverse? An Explanation for People Who Don’t Get It.
3. The metaverse, explained while boardrooms chase a VR internet no one asked for, video game metaverses are all around us.
4. Top metaverse cybersecurity challenges to consider.
5. Neural Networks.

Volume 3 Issue 4 October 2022