香港中文大學
The Chinese University of Hong Kong

Final Year Project Report 2023

Virtual Reality Application Comfort Level Rating Evaluator

Author:

Junjie XIE

Supervisor:

Prof. Michael R. LYU

LYU2201

Faculty of Engineering

Department of Computer Science and Engineering

May 12, 2023

Acknowledgement

I would like to express my gratitude to Professor Michael R. Lyu, my supervisor, and Ms. Shuqing Li, one of Michael's PhD students, for their guidance and invaluable advice in helping us throughout this final year project.

Abstract

VR applications have gradually become popular in recent years, and people are full of endless yearning and anticipation for the metaverse. However, for the special nature of VR applications compared to traditional mobile applications, the same VR applications may bring varying degrees of physiological reactions to humans, which is a major obstacle to advancing the development of the metaverse. In general, it is a time-consuming and long-term task to determine the comfort level of VR applications, and through this project, we would like to work on designing a quick evaluator tool for the comfort level rating of VR applications, obtaining a relatively accurate VR comfort level rating in a short period of time to meet the relevant needs of users, VR application developers and VR application release platform. At the same time, we also utilized information from five mainstream VR application platforms, namely Steam, Oculus Store, Oculus App Lab, SideQuest, and Viveport, to construct our own VR application information dataset to train the comfort level rating evaluator tool and facilitate subsequent research and discussion.

Contents

1 Introduction

1.1 Metaverse

The term metaverse was first presented in Neal Stephenson's 1992 cyberpunk novel Snow Crash [57]. At that time, the idea of a metaverse was first thought to be mere science fiction, but it now seems that it may become a part of our reality in the upcoming years, thanks to the rapid advancements in virtual reality (VR), augmented reality (AR), and other related technologies.

The metaverse is an online, shared, permanent, three-dimensional virtual world that enables connections between individuals from all walks of life, where users can create their own avatars, build their own virtual spaces, and interact with others in a seamless way, which breaks down traditional barriers of geography, culture, and identity. In addition, similar to how the internet links several websites together using a single browser, it would connect multiple platforms [26].

The metaverse represents a new frontier in human experience, offering a shared and permanent digital space where people can connect, create, and explore in ways that were previously impossible.

Figure 1: Metaverse: A Shared Virtual Environment [1]

1.2 Metaverse and Virtual Reality

Metaverse and VR are two different concepts, but there is a close connection between the metaverse and virtual reality. It can be said that virtual reality is the key technology and infrastructure of the metaverse.

The reason behind it is that immersing oneself in a digitally created virtual world where we can engage is a part of virtual reality, with the help of which, humans can experience and converse in a three-dimensional, computer-generated world. As the users enter or be absorbed into this virtual environment and have power over objects or can carry out a series of acts while there, which means virtual reality technology provides the necessary sensory input and interaction mechanisms for the metaverse, enabling users to participate and create their own experiences and content in the metaverse [15].

We might simply consider that it is virtual reality that enables the users to participate in the metaverse. As virtual reality technology continues to develop and improve, it will play an increasingly important role in the Metaverse and beyond.

Figure 2: Virtual Reality Headset [17]

1.3 Virtual Reality Application

VR applications refer to various applications created by applying virtual reality technology, which can be used in games, education, training, entertainment, medical care, and other fields.

At present, the status of VR applications can be said to be developing rapidly and very hot. According to a report released by the International Data Corporation (IDC), in China alone, the global VR and AR (augmented reality) market size has reached 20 billion US dollars in 2022 and is expected to reach 120 billion US dollars by 2026 [9]. And according to statistics from Statista, the number of virtual reality (VR) and augmented reality (AR) users, for 2023 would be forecast to reach over 110 million users [2].

There are some famous virtual reality application examples such as *Blade and Sorcery*, *Beat Saber*, and *Half-Life: Alyx*.

Figure 3: Game: Blade and Sorcery in 2018 [56]

Figure 4: Game: Beat Saber in 2019 [12]

Figure 5: Game: Half-Life: Alyx in 2020 [51]

1.4 Virtual Reality Application Comfort Level Rating

The same virtual reality applications would bring different experiences to people, due to the different physical acceptance capabilities of each person [23]. Thus we use a metric, called the virtual reality application comfort level rating to measure how comfortable and immersive a virtual application is, which is typically a rating system that is used to provide an indication of how likely a user is to experience discomfort or nausea while using a virtual reality application.

The virtual reality application comfort level rating ratings are usually determined based on a number of factors, such as the degree of motion and acceleration in the VR environment, the quality of the graphics and audio, and the length of the VR experience.

And Meta Corporation defines virtual reality application comfort level rating as a way to let users know the intensity of an app before they decide to try the experience [31].

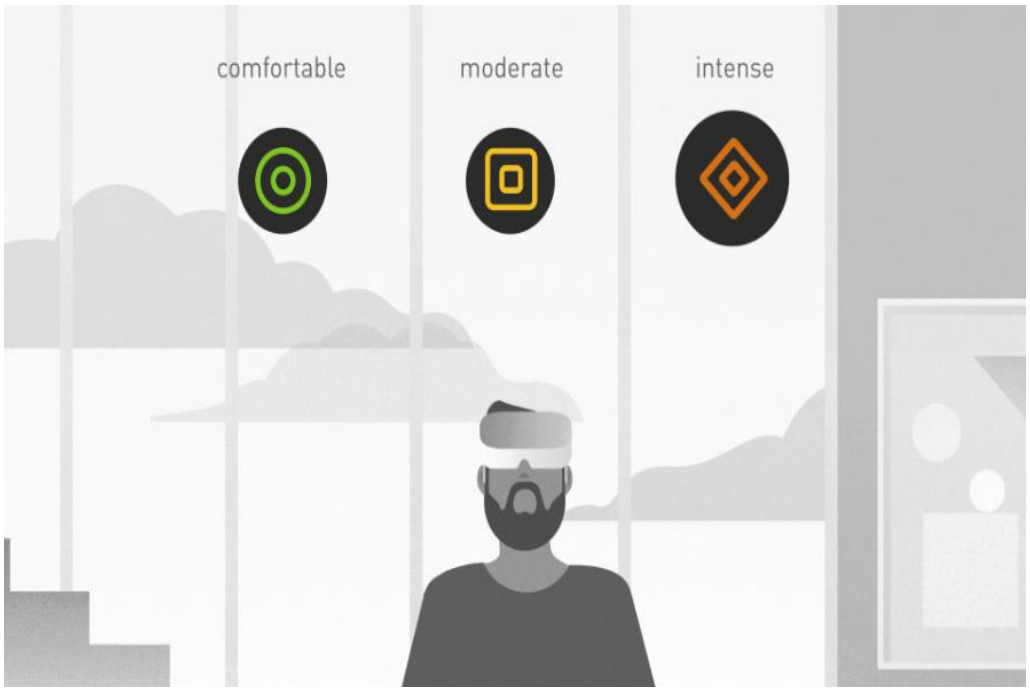

Figure 6: Comfort Level Rating [31]

Comfortable experiences are appropriate for most people, although this rating doesn't mean that an experience is going to be comfortable for everyone. These experiences generally avoid camera movement, player motion, or disorienting content and effects [31].

Moderate experiences are appropriate for many but certainly not everyone. These experiences might incorporate some camera movement, player motion, or occasionally disorienting content and effects [31].

Intense experiences aren't appropriate for most people, especially those who are new to VR. These experiences incorporate significant camera movement, player motion, or disorienting content and effects [31].

Unrated experiences may contain intense content, which may not be right for most people, especially those who are new to VR. If you experience motion sickness or other VR effects from an unrated experience, you should select another unrated experience, or try an experience in the Store that's rated for your comfort level [31].

2 Background

Now, There are five mainstream virtual reality application release platforms, namely Steam, Oculus Store, Oculus App Lab, SideQuest, and Viveport, but only Oculus Store and SideQuest actively provide information about the comfort level of their related applications, on Oculus Store, the information provider is the official Meta, but on sidequest, the provider is its related application developer.

2.1 Application Platform

2.1.1 Steam

The platform was first released in 2003 [59] and has since become one of the largest and most complete digital distribution systems for video games in the whole world, with over 130 million active users worldwide as of 2021 [6]. On the site, gamers can buy, download, discuss, upload, and share games and software.

Of course, this platform offers a dedicated section for VR games and experiences, providing users with easy access to a wide range of VR content from different developers.

Steam also offers tools for VR game development, including the SteamVR software development kit (SDK), which allows developers to create VR experiences that are compatible with different VR devices, such as the HTC Vive, Oculus Rift, and Windows Mixed Reality headsets.

Figure 7: Steam Platform Logo [45]

2.1.2 Oculus Store

At the Oculus Connect conference, Oculus VR launched the ”Oculus Platform” for developers to share their virtual reality experiences and apps. This updated Oculus Share marketplace will enable users to browse the Oculus Platform in virtual reality and download apps, games, and entertainment experiences starting in the fall of 2014 for the Samsung Gear VR created by Oculus [8].

The Oculus Platform was claimed eventually to have Rift, iOS, Android, Windows Phone, Chrome, Firefox, Safari, and Internet Explorer versions. By establishing this market, Oculus might attract an ecosystem to its mobile and PC-based VR devices. Oculus Platform might become one of the first platforms for developers to sell the VR experiences they create [8].

The Oculus store launched in 2016, along with the release of the Oculus Rift headset, and has since become a popular destination for VR enthusiasts to purchase and download games, apps, and other VR experiences [58].

The Oculus Store offers a curated selection of VR content, which is compatible with various Oculus headsets, including the Oculus Rift, Oculus Quest, and Oculus Go. With prices starting at a few dollars and going up to more expensive premium titles, the store offers a mixture of both free and paid content.

In addition to games and apps, the Oculus Store also offers 360-degree videos, immersive virtual tours, and other VR experiences. The store includes an intuitive design that makes it simple for consumers to find new application, buy it, and download it immediately to their Oculus headset.

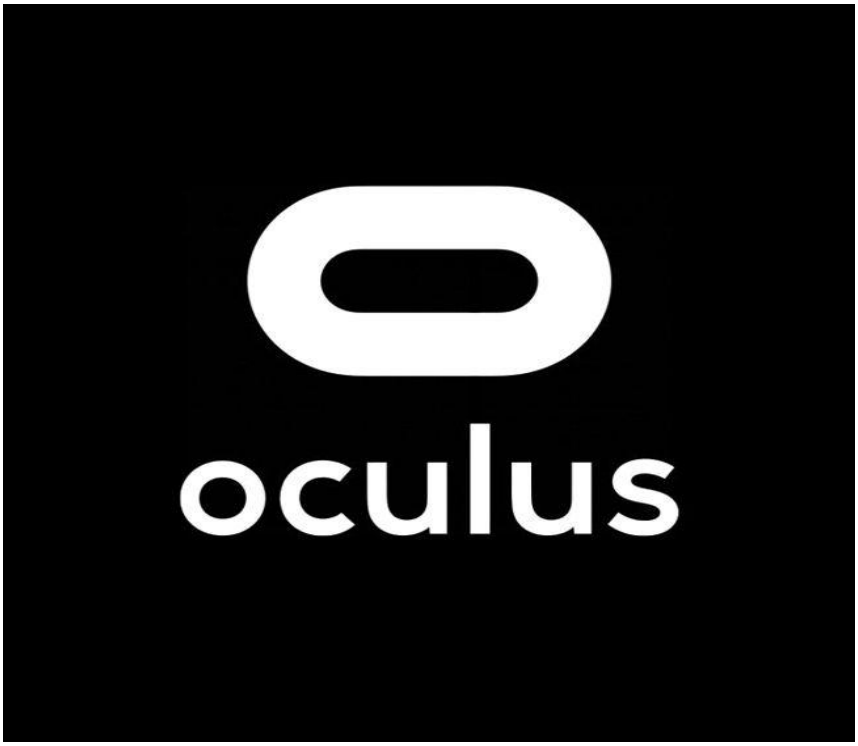

Figure 8: Oculus Store Platform Logo [53]

2.1.3 Oculus App Lab

Oculus App Lab is a new way provided by Oculus for developers to distribute apps directly to consumers safely and securely, with the help of direct links or platforms like SideQuest, without requiring store approval and without sideloading [55]. And the App Lab does not require the same level of review as the Oculus Store, allowing developers to quickly and easily release their content to the public, hence it provides an opportunity for independent developers to reach a wider audience by making their content more accessible to Oculus headset users.

App Lab makes it possible that even if the application is early in development, experimental, or aimed at a unique audience, it can still get the application directly to a community.

While the App Lab offers more flexibility for developers, it also requires them to take responsibility for the quality and safety of their content. Oculus provides guidelines for developers to follow, but ultimately it is up to them to ensure that their content meets certain standards and is safe for users to experience [25].

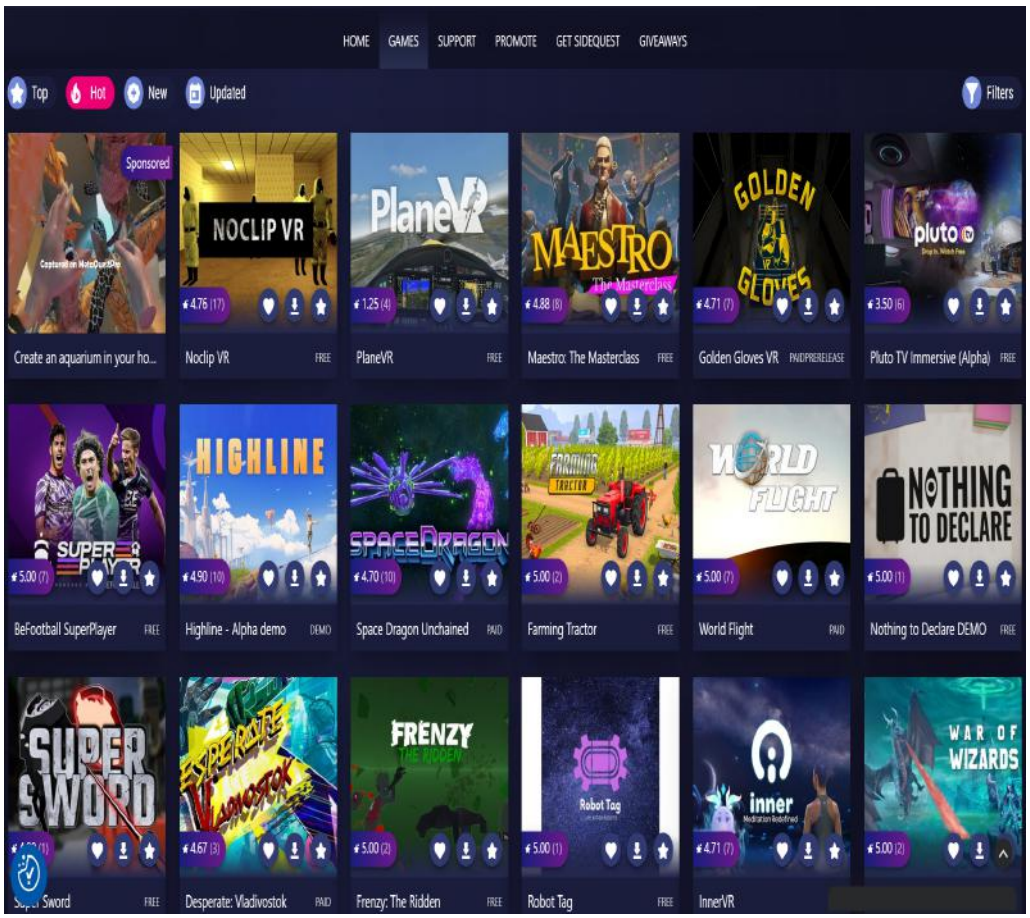

Figure 9: Oculus App Lab Screenshot

2.1.4 Viveport

Viveport is a virtual reality platform created by HTC, the same company that produces the HTC Vive VR headset. The platform is designed to provide VR users with a curated selection of high-quality VR content, including games, experiences, and other applications [16]. This platform builds on the rapid expansion and popularity of VR gaming.

The Viveport platform is different from other VR app stores in that it focuses on a subscription-based model [3], which means users pay a monthly fee to access a library of VR content. Users can also purchase individual titles on the platform if they choose. The subscription-based model allows users to access a wide range of VR content without having to purchase each title individually, which can be more cost-effective for

heavy VR users.

In addition to the subscription model, the Viveport platform also offers a developer program, which provides developers with tools and resources to create VR content for the platform [54]. Developers can publish their VR content on the Viveport platform and receive a share of the revenue generated by their titles.

With Viveport, the officials hope to further their mission of liberating human imagination from the constraints of reality by providing the most comprehensive and varied collection of VR experiences currently accessible [47].

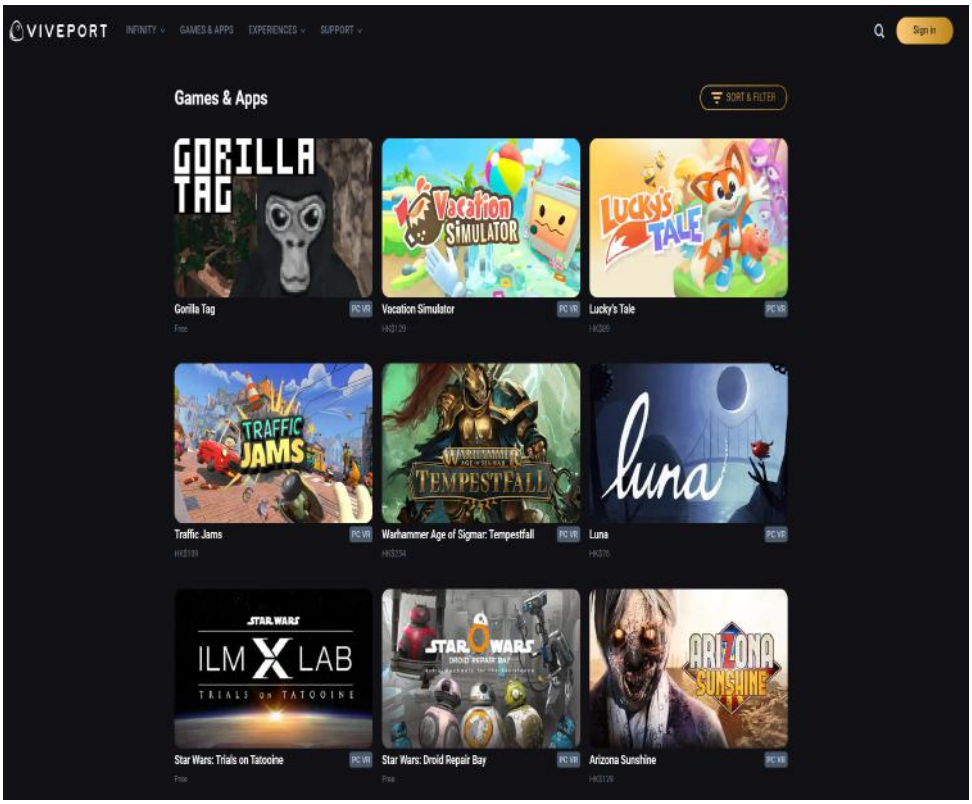

Figure 10: Viveport Platform Screenshot

2.1.5 SideQuest

SideQuest is a VR content platform for standalone headsets like Quest and Quest 2 where users can explore a vast collection of VR titles ranging from experimental games, tech demos, and game ports to full-scale games listed on the official Oculus store and App Lab [50].

As an alternative app store, SideQuest allows users to download and install VR content that may not be available on official app stores due to various reasons, such as experimental apps or games that are still in development. This user-driven platform fosters interaction and experimentation with VR apps by fostering meaningful relationships between users and developers. Because of this, SideQuest frequently has some of the most inventive and unusual titles! Games like Gorilla Tag, Pavlov, Puzzling Places, Ancient Dungeon, Gun Raiders, Quest3Doom, and QuestCraft, which have amassed hundreds of thousands of downloads, were all born on SideQuest and have gone on to become real sensations in the VR community.

Figure 11: SideQuest Platform Logo [60]

3 Motivation

Due to the special nature of virtual reality applications compared to traditional mobile applications, even the same virtual reality applications may bring varying degrees of physiological reactions to humans, which is a major obstacle to advancing the development of the metaverse. For example, if a virtual reality application contains a lot of spinning, jumping, and flashing action content, it is likely to be an application that gives the user an extreme sense of vertigo. People generally tend not to use this kind of application.

However, determining the comfort level of VR applications is a very time-consuming and long-term task, which needs to do extensive experiments on different groups of users. Nowadays, if there exists an automatic tool for the comfort level rating evaluation of virtual reality applications in the current market, it would save the virtual reality application development corporations a lot of time and money resources. Quickly and accurately rating the comfort level of a virtual reality application using automatic tools could benefit users, and developers as well as virtual reality application platforms.

3.1 User

Nowadays, if users do not have the resources to learn about the comfort level information of virtual reality applications, they can only know the comfort level of VR applications through their personal use, which might cause discomfort or severe negative physiological reactions such as dizziness or even fainting.

For users, the automatic tool for the comfort level rating evaluation of virtual reality applications could provide a more efficient and reliable way to assess the potential discomfort or negative physiological reactions they might experience before using a VR application. Hence, users can determine if the application is suitable for them or not.

By knowing the comfort level rating of a VR application beforehand, users can make informed decisions on whether to try it or avoid it. This can help reduce the risk of negative physiological reactions and increase user satisfaction with the overall virtual reality experience, which means that it could be easier for users to find applications that are more suited to their individual needs and preferences.

3.2 Developer

For developers, such an automatic tool can save time and money resources in evaluating the comfort level of their applications during the process of developing applications. Instead of conducting time-consuming and costly experiments on different groups of users, developers can use the tool to evaluate their VR applications' comfort level quickly and more efficiently. This can help them identify and fix issues related to user comfort and improve the quality of their VR applications, ultimately leading to increased user adoption and satisfaction. It could promote developers bringing their applications to market faster and more efficiently, improving their competitiveness in the market.

Besides, the automatic evaluation tools can also help developers understand what factors contribute to a comfortable experience. Subsequently, developers can make informed decisions about the design of their applications, leading to more positive user experiences, increased user satisfaction, and more widespread adoption of their applications. For example, developers can adjust or remove the corresponding elements that may cause discomfort to users.

3.3 Platform

As virtual reality application publishing platforms, with this automatic tool, they can quickly identify and remove applications that are likely to cause discomfort or negative experiences for users, which can effectively reduce the cases in that the user applies for a refund due to the physiological harm and makes negative comments about the platform for misleading application information. Thus, it could promote the overall user experience on the platform.

By providing a safer and more comfortable user experience, users are more likely to spend more time using the platform, which can lead to increased engagement and potentially increased revenue for the platform. Besides, providing more reliable and accurate information about the virtual reality applications they host could improve their reputation among users and potentially attract new users to the platform. This can help the platform become more successful and profitable in the long run and contribute to the growth of the overall virtual reality industry.

4 Dataset

4.1 Data Source

In order to be able to perform comfort rating experiments for VR applications, we need a corresponding dataset as a basis. However, there does not exist available dataset for this purpose, so we needed to prepare it on our own. We utilized the crawler tool to obtain information about the virtual reality applications from the websites of the five mainstream virtual reality application release platforms, namely Steam, Oculus Store, Oculus App Lab, SideQuest, and Viveport. This process took a lot of time and effort, as we needed to sift and sort through the data to make sure they fit our needs. In this way, we were able to obtain a sufficient amount of data to build a reliable virtual reality application comfort level rating prediction model.

4.2 Web Crawling Tool—Scrapy

Scrapy is a Python implementation of an application framework written to crawl web data and extract structured data. Scrapy is often used in a range of applications including data mining, information processing, or storing historical data. Utilizing the Scrapy framework, leveraging its set of pre-built features, we could simply implement a crawler to crawl the text, images, and videos of a given website.

Figure 12: A Web Crawling Tool: Scrapy [10]

When scanning websites and extracting structured data from their pages, Scrapy is one of the quickest high-level web crawling and web scraping frameworks available because Scrapy uses the Twisted asynchronous network framework to handle network communications, it can speed up the download speed, and we do not have to implement our own asynchronous framework.

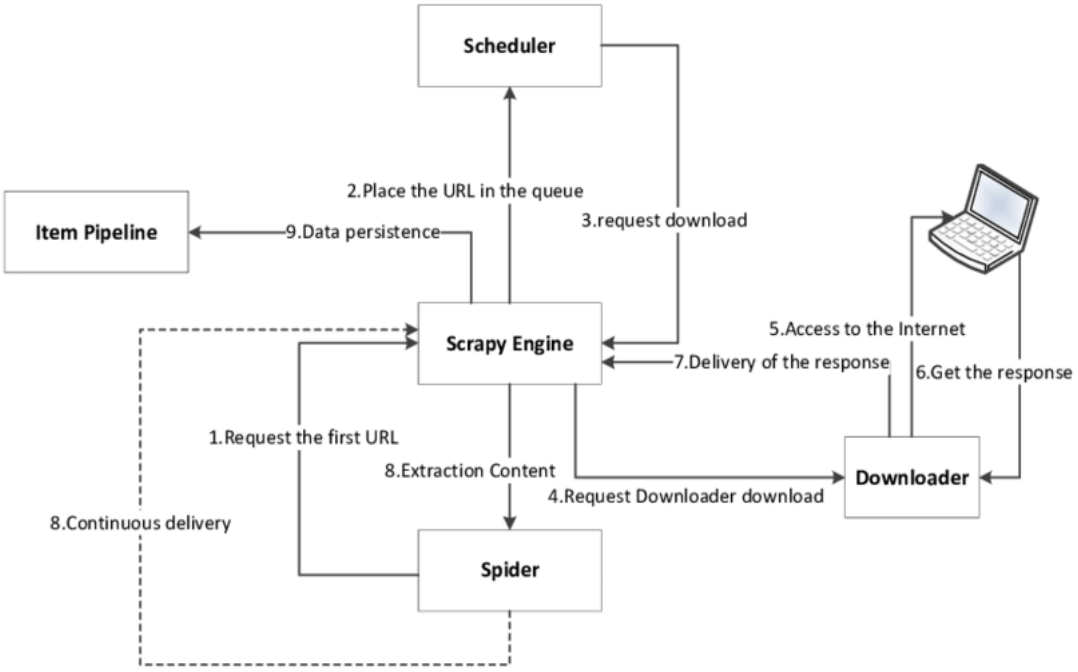

Figure 13: Overall Architecture of the Scrapy Framework [28]

4.3 Application Example on Platform Website

4.3.1 Steam example

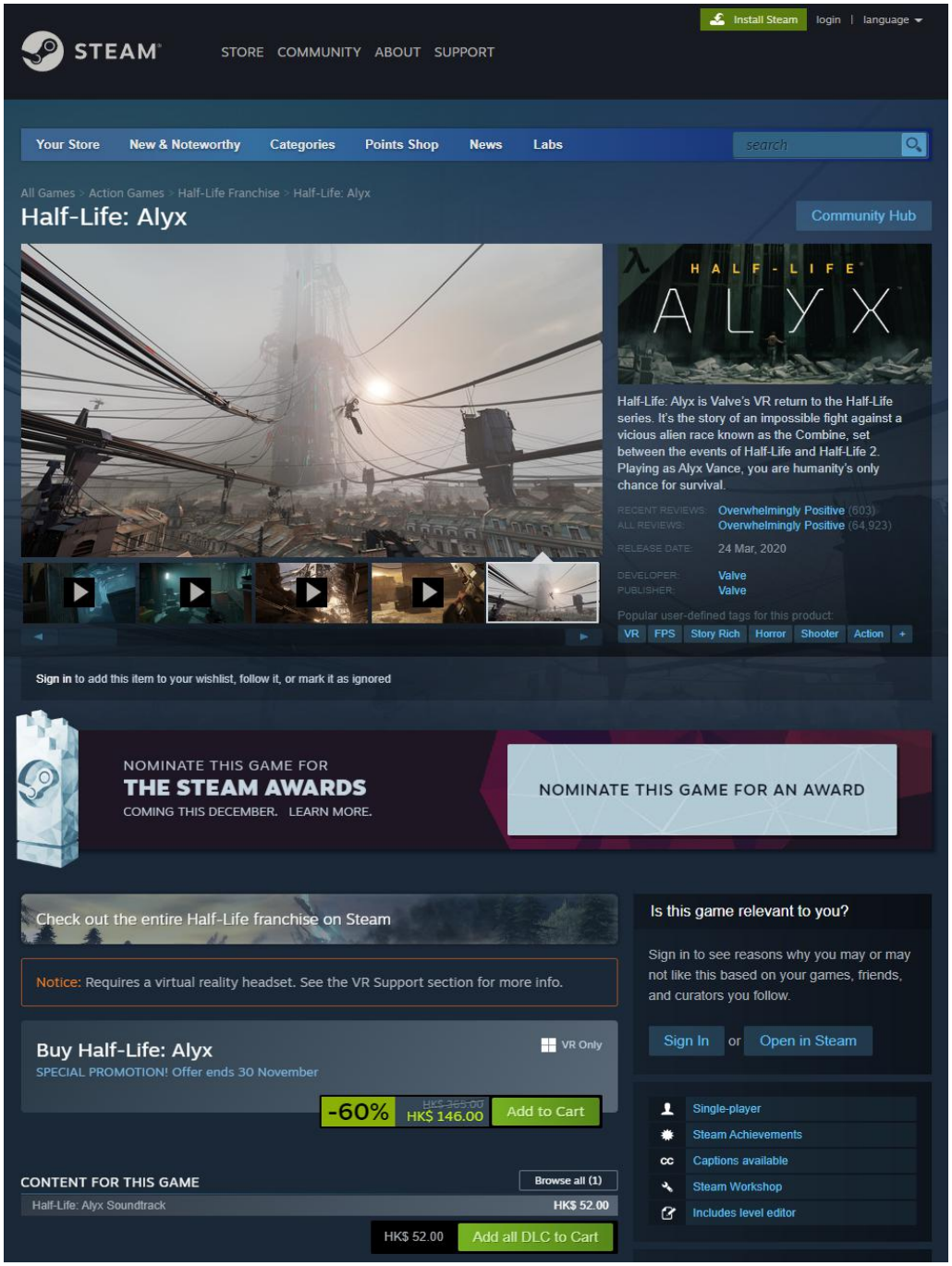

Figure 14: Half-Life: Alyx's information on steam (part1)

The Steam platform provides information about the game title, game description, game clips, etc. But users can not get information about the comfort level of that application.

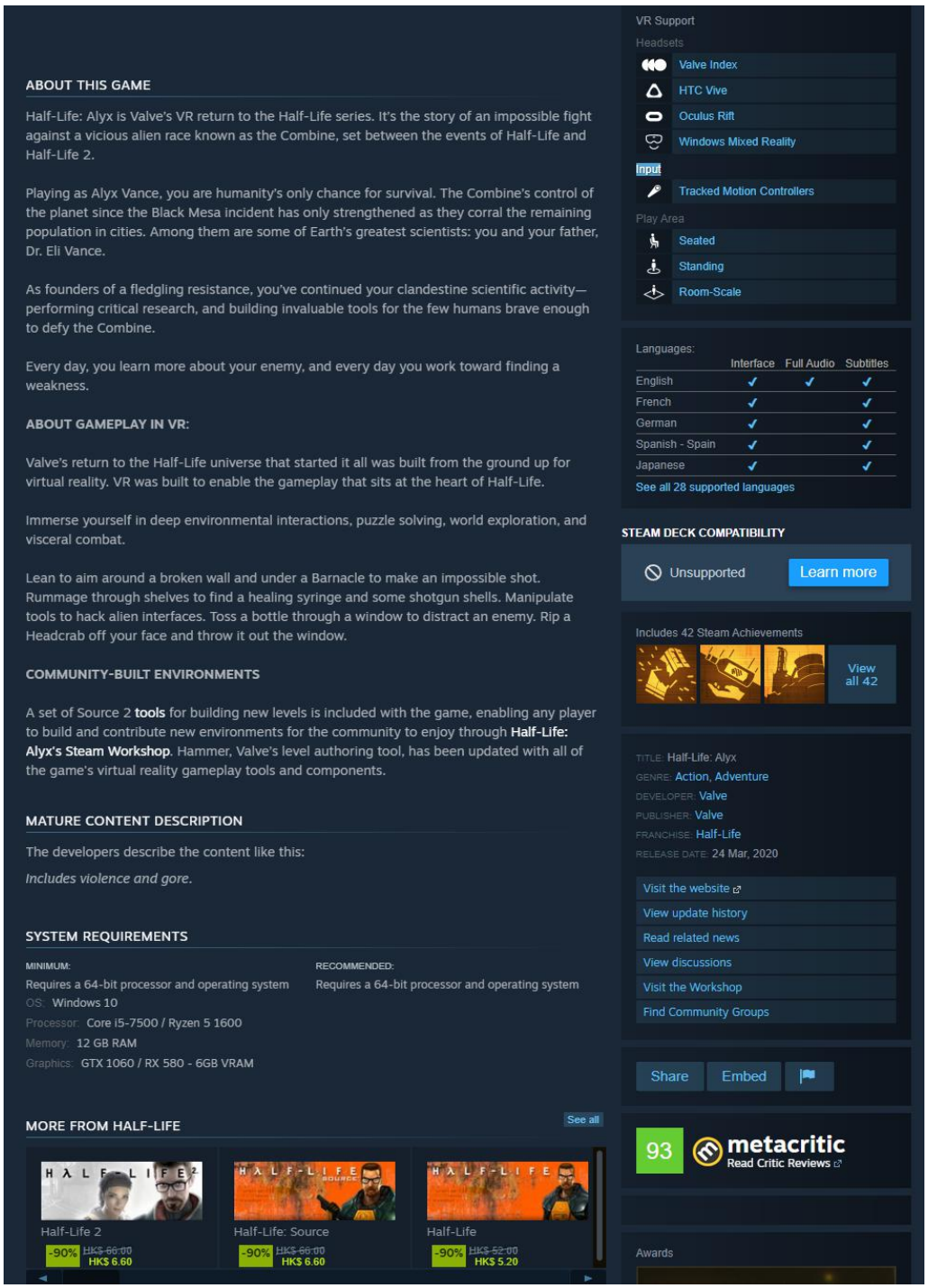

Figure 15: Half-Life: Alyx's information on steam (part2)

It also provides information about system requirements, supported VR headsets, etc.

4.3.2 Oculus Store Example

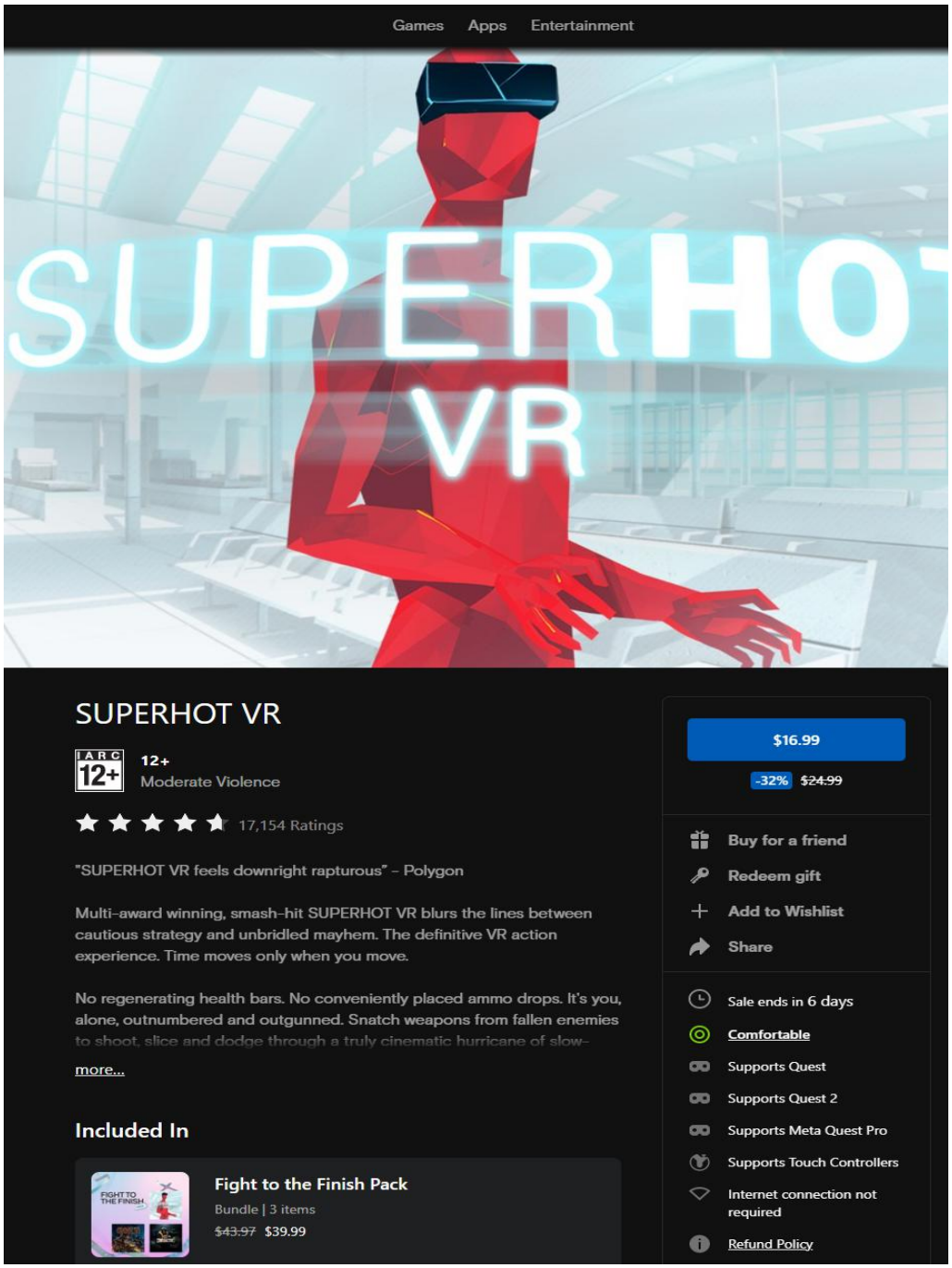

Figure 16: SUPERSHOT VR's information on oculus store (part1)

It provides information about the game title, game description, game pictures, etc. And users can get information about the comfort level for that application at the right bar of this page.

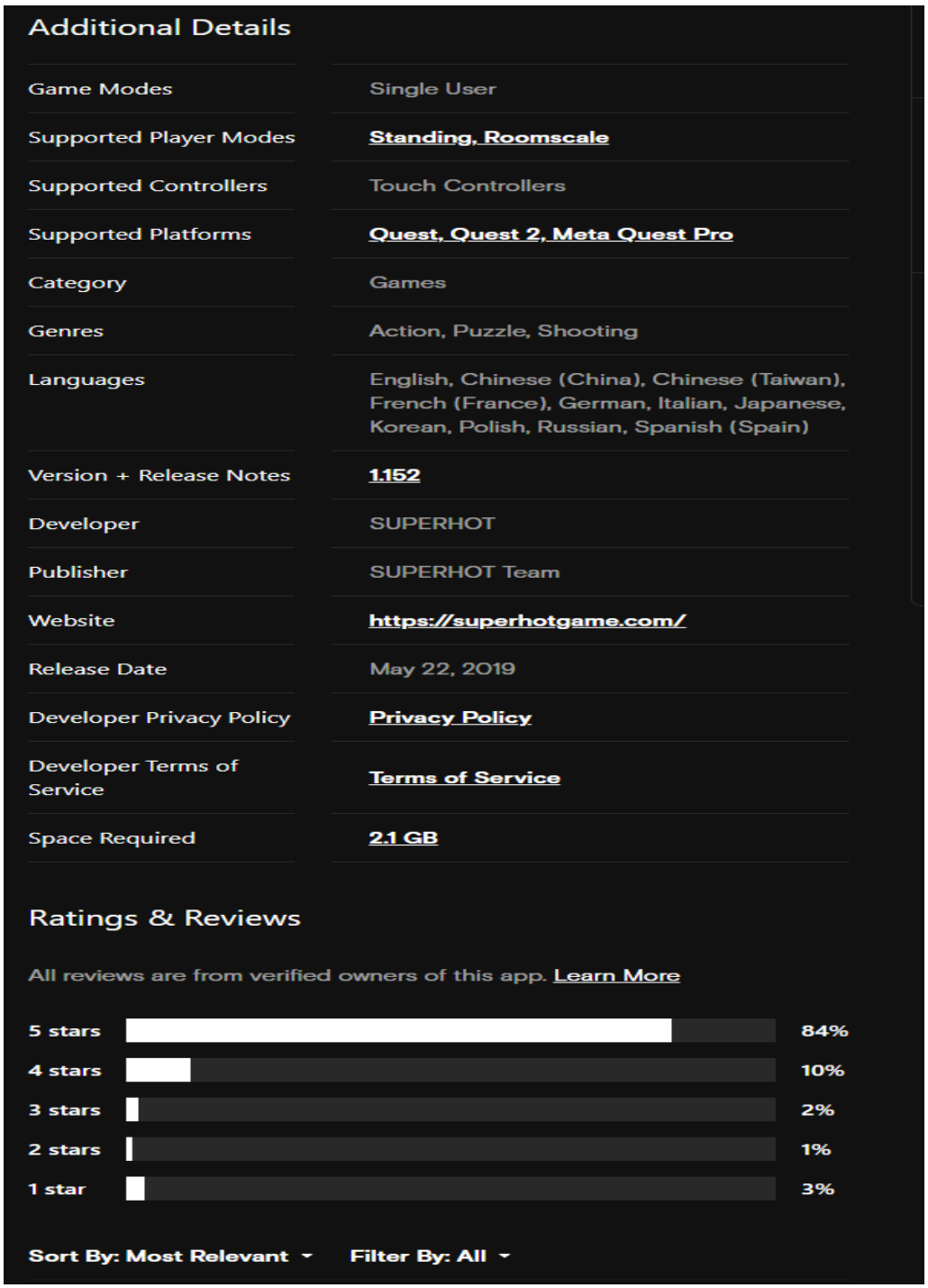

Figure 17: SUPERSHOT VR's information on oculus store (part2)

It also provides information about the game developer, languages, reviews, etc.

4.4 Construct Datasets

Since the layout of the information on the web pages of each platform and the data response content or format is different, we need to use different dealing approaches when crawling the web pages of different platforms, of which there are two main ways.

4.4.1 Dataset construction Time

For the sake of data consistency, we crawled steam, oculus store, oculus app lab, sidequest and viveport on the same day, October 5, 2022, for all their related VR application data, such as application description, application screenshots, application videos, user reviews, etc.

4.4.2 Extra Data from HTML Response

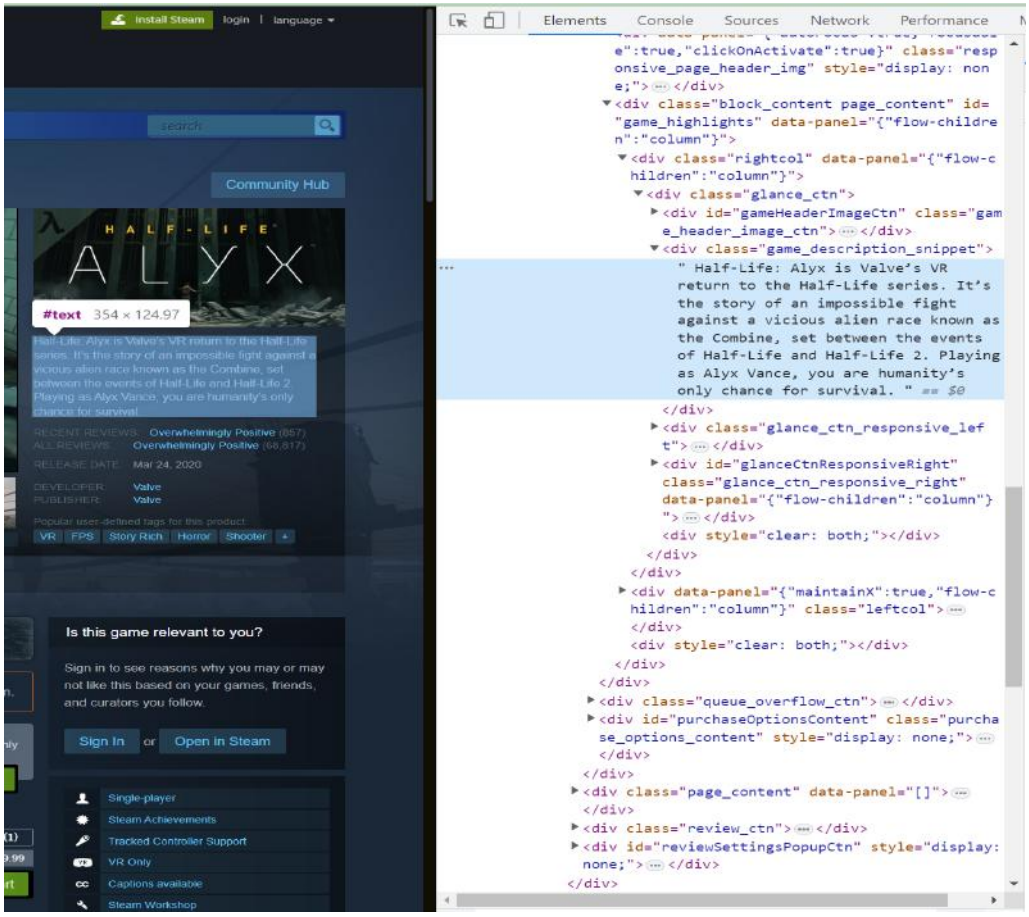

Figure 18: Steam Page Structure (partial)

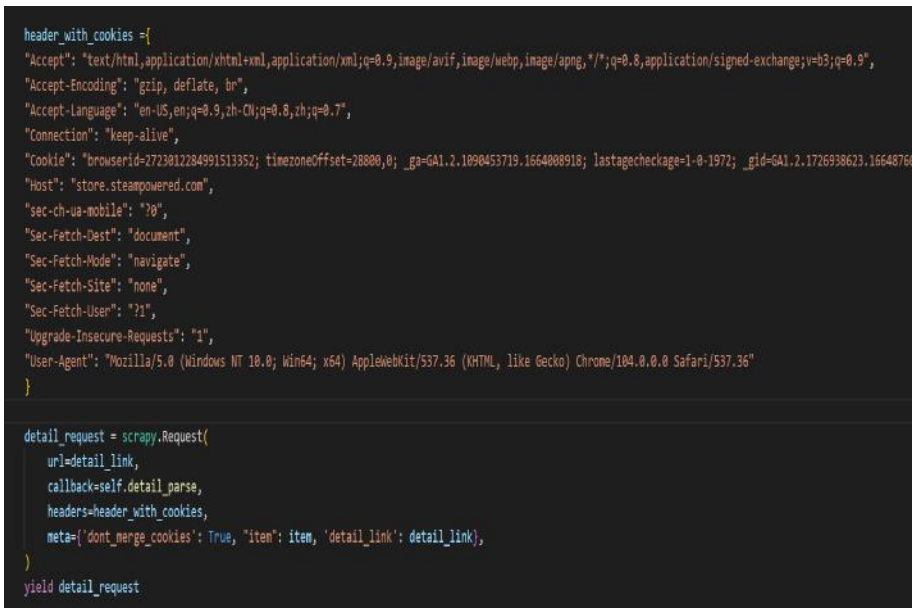

Figure 19: Steam crawler codes request part (partial)

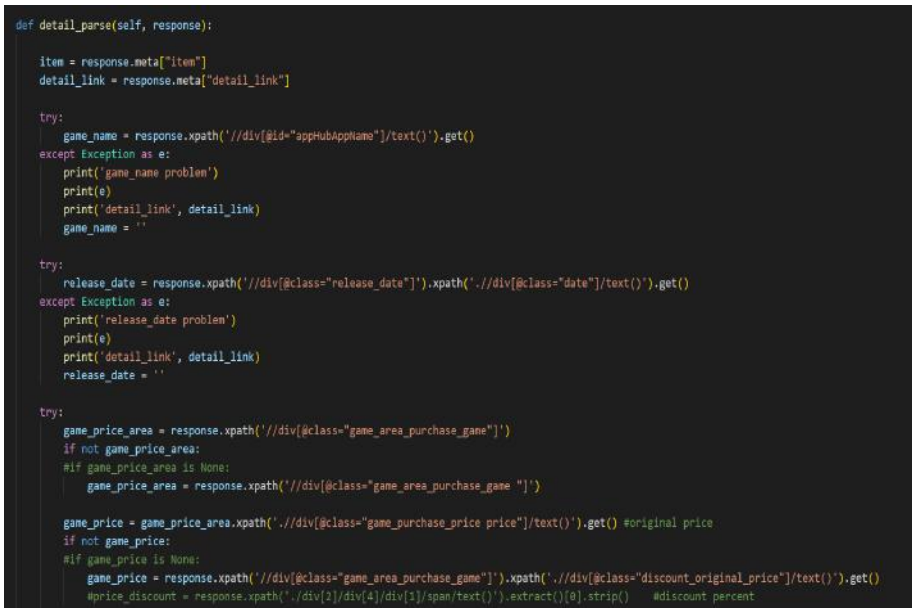

Figure 20: Steam crawler codes parse part (partial)

For example, when we build the Steam virtual reality application dataset, we directly get all related field information utilizing the Document Object Model (DOM) tree.

4.4.3 Extra data from graphql JSON response

For example, when we build the viveport virtual reality application dataset, we directly make requests to `https://www.viveport.com/graphql` and get its JSON response. Subsequently, we manipulate its response and extract related information about applications.

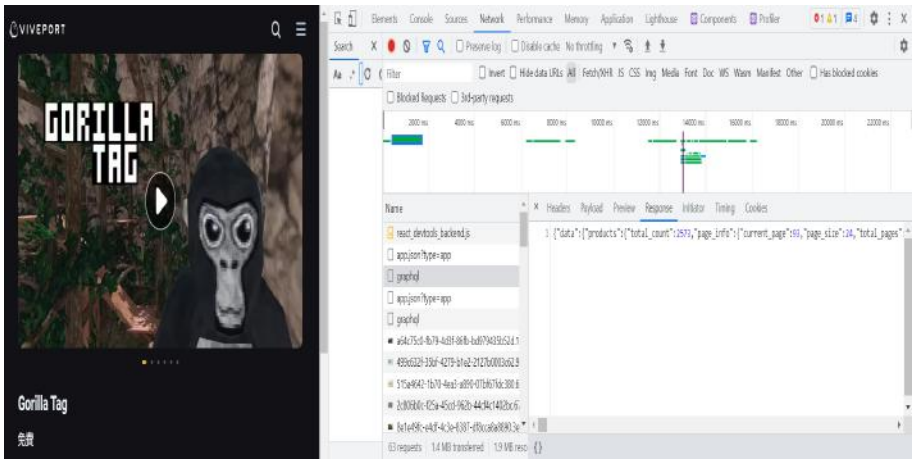

Figure 21: Viveport platform server response (partial)

```
detail_request1 = scrapy.Request(url=url,
    callback=self.detail_parse,
    body=json.dumps(base_post_request_data),
    method='POST',
    headers = {
        "accept": "*/*",
        "accept-encoding": "gzip, deflate, br",
        "accept-language": "en-US,en;q=0.9",
        "authorization": "authorization",
        "content-type": "application/json",
        "cookie": "csdata&id=a0eeb891fc4406798b819daee9adbb3cd1f81967cea679ed68602cee2d2fd917; _ga=GA1.2.1324714856.1656609238; OptanonAlertBoxClosed=2022-06
        "origin": "https://www.viveport.com",
        "referer": "https://www.viveport.com/filter-page/app",
        "sec-ch-ua-mobile": "?0",
        "sec-fetch-dest": "empty",
        "sec-fetch-mode": "cors",
        "sec-fetch-site": "same-origin",
        "user-agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/104.0.0.0 Safari/537.36",
    },
    meta={'dont_merge_cookies': True, "page_number": i}) ### must need
```

Figure 22: Viveport crawler request part (partial)

```
def review_parse(self,response):

    #print("enter to review parse")
    sample = response.meta["sample"]
    game_detail_request_url = response.meta["detail_link"]
    game_sku = response.meta["game_sku"]
    current_review_page = response.meta["current_page"] + 1
    #print(sample['game_name'], "review current_review_page = ", current_review_page)

    review_page_number = 1
    if response.meta["current_page"] != 0:
        review_page_number = response.meta["review_page_number"]

    #if response.meta["current_page"] != 0:
        sample = response.meta["sample"]
        data = response.text
        data = data.replace('\\"', "'")
        data = data.replace("\/", "/")
        a_dict = json.loads(data)

        page_reviews = a_dict['data']['products']['items'][0]['reviews']['items']
        sample["reviews"].extend(page_reviews)
        review_page_number = a_dict['data']['products']['items'][0]['reviews']['page_info']['total_pages']
    if current_review_page > review_page_number:
        yield sample

    base_get_review_post_request= {
        "operationName": "getProduct",
        "variables": {
            "sku": "05144741-96a1-42dc-b636-5a4e00c88be2",
            "reviewFilter": {},
            "reviewSortCondition": {
            "date": "DESC"
            },
            "reviewPageSize": 10,
            "reviewCurrentPage": 1
        },
```

Figure 23: Viveport crawler parse part (partial)

4.4.4 Sample Data

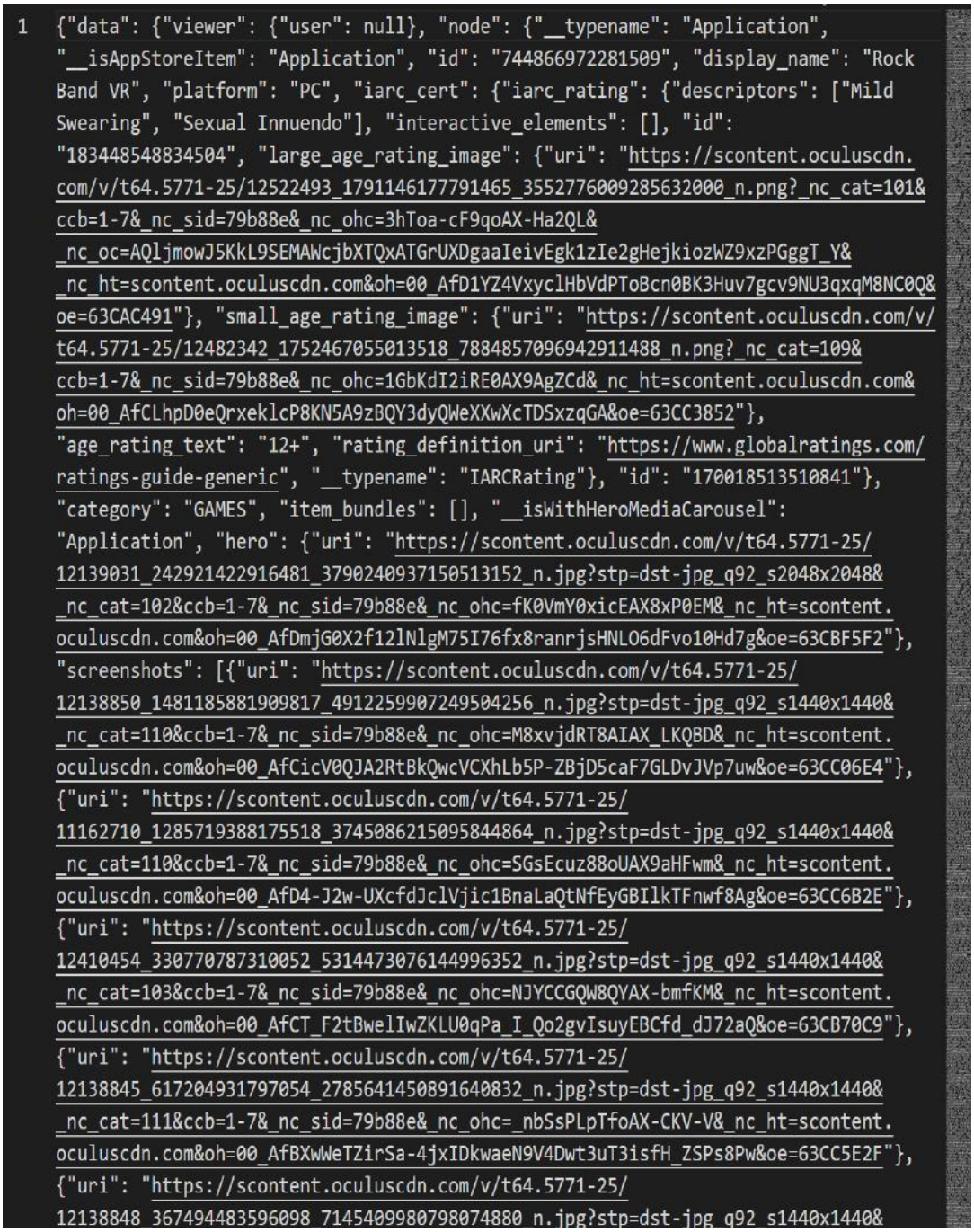

Figure 24: One Oculus Dataset Sample (partial)

4.4.5 Dataset Quantity

Dataset List	
Platform	Quantity
Steam	4428
Oculus Store	3965
SideQuest	3680
VIVEPORT	2225
Oculus Lab	1375

Since only the Oculus Store dataset contains the comfort rating for the application published by the Oculus Meta Official, for the subsequent experiments and their analyses, we concentrate on the dataset from the Oculus Store.

4.4.6 Dataset Problem

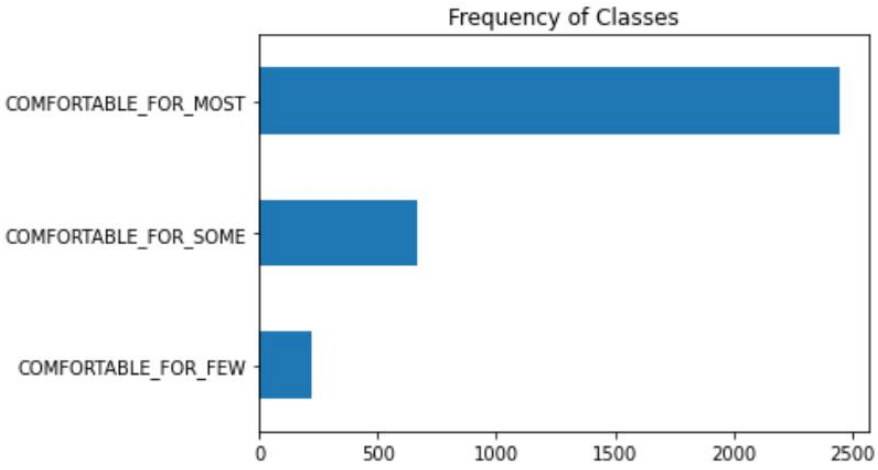

Figure 25: Oculus Store Dataset Each Class Quantity

Here, the label COMFORTABLE FOR MOST indicates comfortable, the label COMFORTABLE FOR SOME indicates moderate, and the label COMFORTABLE FOR FEW indicates intense.

According to the figure 25, The number of each category differs significantly. When encountering imbalanced data, traditional classification algorithms that focus on overall classification accuracy as the learning goal focus too much on the majority class, thus degrading the classification performance of minority class samples. The vast majority of common machine learning algorithms do not work well for unbalanced data sets.

Mitigate the Problem

Under-sampling Under-sampling the data samples of a large class reduces the number of data samples of the large class, i.e., the number of samples sampled is less than the number of samples of the class.

Over-sampling Over-sampling the data samples of a small class increases the number of data samples of the small class, i.e., the number of samples sampled is greater than the number of samples of the class.

The sampling algorithm is easy to implement and effective, but it may increase the bias of the model because the effect of enlarging or reducing some samples is equivalent to changing the distribution of the original data set. Different sampling ratios should also be taken for different classes, but generally not 1:1 because it is far from the reality [30].

Over-sampling Result For our dataset, the number of samples is not very large compared to other traditional datasets, so we use the over-sampling method. By over-sampling, we increase the number of VR applications with either moderate or intense comfort ratings label, thus making our dataset more balanced and reasonable.

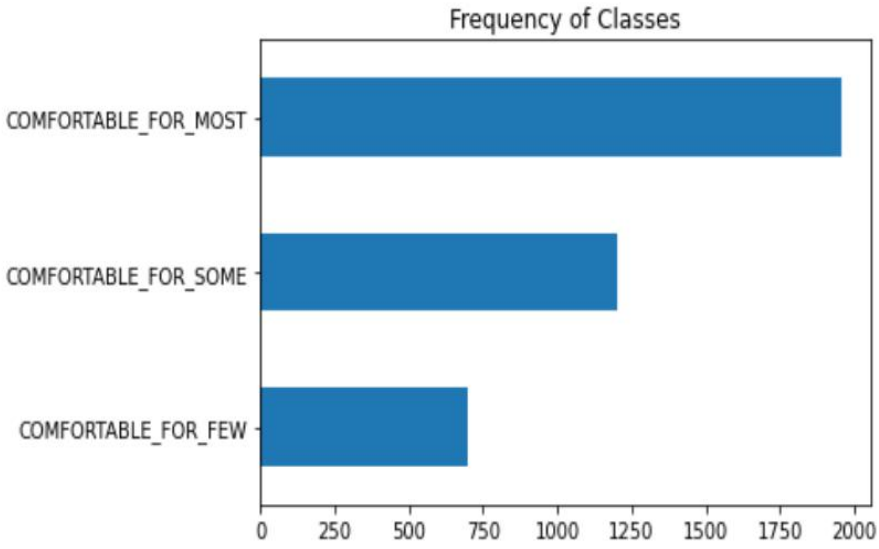

Figure 26: Oversampled Training Dataset

4.5 Text Modality for VR application

For the text modal information fed to our experiment model, we can extract it directly from our JSON dataset, and in our experiments, we have utilized the description and reviews information corresponding to the virtual reality application.

4.6 Image Modality for VR application

For the Image modal information used in the model, we target the screenshots provided by each application, and we can directly extract the links to the screenshots from our JSON dataset and download them all as part of our dataset.

4.7 Video Modality for VR application

For the Video modal information used by the model, we target the application promotional videos provided by each application, we can directly extract the links of the videos from our JSON dataset and then download them all to become part of our dataset. Inspired by UniVL [27], the video modal information can be applied in such a way that the video can be cut into many consecutive video clips and then fed into our model as input.

Video Decomposition For the task of processing video to consecutive video clips, we utlized the opencv [35] library and the YOLO [40] model. By using the opencv library, we were able to easily process the video files and decompose them into frame-by-frame images, which allowed us to better process and analyze the performance of each virtual reality application in the video. Meanwhile, we used the YOLO model to detect the informative region of vr applications in each frame and cut out the corresponding continuous video clip of each application as part of our dataset.

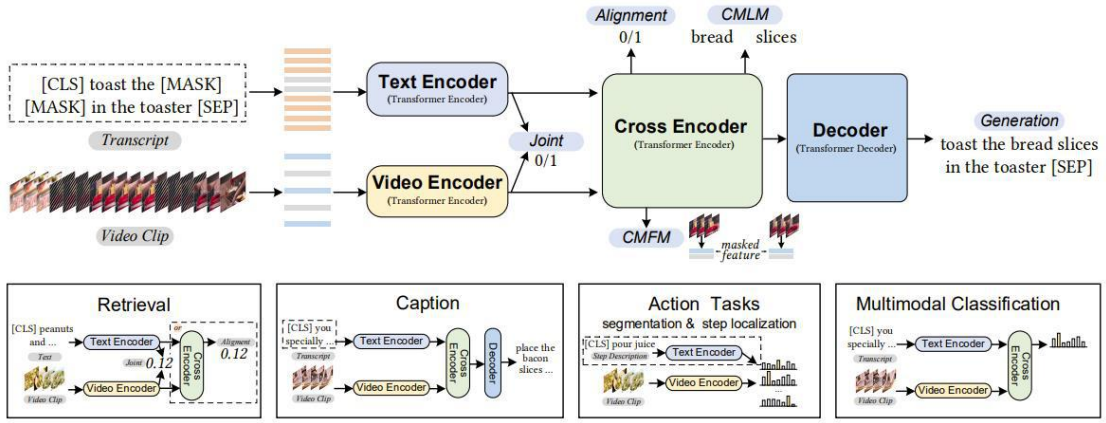

Figure 27: Main Structure of UniVL [27]

The main structure of UniVL, comprises four components, including two single-modal encoders, a cross encoder, and a decoder. The model is flexible for many text and video downstream tasks. Four possible tasks are listed as Retrieval, Caption, Action Tasks, and Multimodal Classification [27].

5 Methodology

We first try to only use the application description information to predict the VR application comfort level rating.

5.1 Feature Engineering

Feature engineering is an important part of the machine learning workflow, which is to ”translate” the raw data into a form that the model can understand. The purpose of feature engineering is to select and find a better feature for predictive models to be more accurate [63].

5.1.1 Count Vectors

Count Vector is a matrix notation of the dataset in which every row represents a document from the corpus, every column represents a term from the corpus, and every cell represents the frequency count of a particular term in a particular document.

Count Vectors is a simple and commonly used method for text feature extraction, which is often used for tasks such as text classification, sentiment analysis, and information retrieval. However, it ignores the order and grammatical structure between words and does not capture the relationship between words well, so it may have some limitations in dealing with certain natural language processing problems [65].

text	label
Interact and travel through all the solar system experiencing the amazing life of Annie Amber, from her birth until her final days, solving puzzles of increasing complexity while you unlock her memories. Use just your head to move or to solve the puzzles! From the creator of MIND: Path to Thalamus, the VR, award-winning, indie game of the year.	COMFORTABLE_FOR_SOME
Absolut unveils a new chapter in Absolut Nights. Introducing Absolut deadmau5, an interactive VR experience. Go on an unforgettable night out with deadmau5, from his studio to the club, featuring exclusive music from the artist. To learn more about Absolut deadmau5 or to get a limited edition VR headset, please visit Absolutdeadmau5.com You must be of legal drinking age to use this app. Remain seated when using the app. Stop use if you experience any discomfort.	COMFORTABLE_FOR_SOME

Figure 28: Two Description Samples

```
Vocabulary:  ['a', 'about', 'absolut', 'absolutdeadmau5', 'age', 'all', 'amazing', 'amber', 'an', 'and', 'annie', 'any', 'app', 'arti
st', 'award', 'be', 'birth', 'chapter', 'club', 'com', 'complexity', 'creator', 'days', 'deadmau5', 'discomfort', 'drinking', 'editio
n', 'exclusive', 'experience', 'experiencing', 'featuring', 'final', 'from', 'game', 'get', 'go', 'head', 'headset', 'her', 'his', 'i
f', 'in', 'increasing', 'indie', 'interact', 'interactive', 'introducing', 'just', 'learn', 'legal', 'life', 'limited', 'memories',
'mind', 'more', 'move', 'music', 'must', 'new', 'night', 'nights', 'of', 'on', 'or', 'out', 'path', 'please', 'puzzles', 'remain', 's
eated', 'solar', 'solve', 'solving', 'stop', 'studio', 'system', 'thalamus', 'the', 'this', 'through', 'to', 'travel', 'unforgettabl
e', 'unlock', 'until', 'unveils', 'use', 'using', 'visit', 'vr', 'when', 'while', 'winning', 'with', 'year', 'you', 'your']

Encoded result is:
[[0 0 0 0 0 1 1 1 0 1 1 0 0 0 1 0 1 0 0 0 1 1 1 0 0 0 0 0 0 1 0 1 2 1 0 0
  1 0 3 0 0 0 1 1 1 0 0 1 0 0 1 0 1 1 0 1 0 0 0 0 0 4 0 1 0 1 0 2 0 0 1 1
  1 0 0 1 1 6 0 1 3 1 0 1 1 0 1 0 0 1 0 1 1 0 1 1 1]
 [2 1 4 1 1 0 0 0 2 0 0 1 2 1 0 1 0 1 1 1 0 0 0 3 1 1 1 1 2 0 1 0 2 0 1 1
  0 1 0 1 1 1 0 0 0 1 1 0 1 1 0 1 0 0 1 0 1 1 1 1 1 1 1 1 1 0 1 0 1 1 0 0
  0 1 1 0 0 3 1 0 4 0 1 0 0 1 2 1 1 2 1 0 0 1 0 2 0]]
```

Figure 29: Samples after Counter Vector Encoded

5.1.2 TF-IDF Vectors

TF-IDF stands for Term Frequency — Inverse Document Frequency and is a statistic that aims to better define how important a word is for a document, while also taking into account the relation to other documents from the same corpus. This is accomplished by examining the frequency with which a term appears in a text as well as the frequency with which the same word appears in other papers within the corpus.[44].

TF-IDF is a score that is applied to every word in every document in our dataset. And for every word, the TF-IDF value increases with every appearance of the word in a document but is gradually decreased with every appearance in other documents[4].

Firstly, let's define some notations, N is the number of documents we have in our dataset, d is a given document from our dataset, D is the collection of all documents, and w is a given word in a document.

To calculate the term frequency, the formula is

$$tf(w,d) = \log(1 + f(w,d)),$$

where f(w,d) is the frequency of word w in document d. To calculate the inverse term frequency, the formula is

$$idf(w,D) = \log\left(\frac{N}{f(w,D)}\right).$$

Finally to calculate the TF-IDF score by the following formula,

$$tfidf(w,d,D) = tf(w,d) * idf(w,D).$$

```
Vocabulary:  {'interact': 44, 'and': 9, 'travel': 81, 'through': 79, 'all': 5, 'the': 77, 'solar': 70, 'system': 75, 'experiencing': 29,
'amazing': 6, 'life': 50, 'of': 61, 'annie': 10, 'amber': 7, 'from': 32, 'her': 38, 'birth': 16, 'until': 84, 'final': 31, 'days': 22,
'solving': 72, 'puzzles': 67, 'increasing': 42, 'complexity': 20, 'while': 91, 'you': 95, 'unlock': 83, 'memories': 52, 'use': 86, 'jus
t': 47, 'your': 96, 'head': 36, 'to': 80, 'move': 55, 'or': 63, 'solve': 71, 'creator': 21, 'mind': 53, 'path': 65, 'thalamus': 76, 'v
r': 89, 'award': 14, 'winning': 92, 'indie': 43, 'game': 33, 'year': 94, 'absolut': 2, 'unveils': 85, 'a': 0, 'new': 58, 'chapter': 17,
'in': 41, 'nights': 60, 'introducing': 46, 'deadmau5': 23, 'an': 8, 'interactive': 45, 'experience': 28, 'go': 35, 'on': 62, 'unforgetta
ble': 82, 'night': 59, 'out': 64, 'with': 93, 'his': 39, 'studio': 74, 'club': 18, 'featuring': 30, 'exclusive': 27, 'music': 56, 'artis
t': 13, 'learn': 48, 'more': 54, 'about': 1, 'get': 34, 'limited': 51, 'edition': 26, 'headset': 37, 'please': 66, 'visit': 88, 'absolut
deadmau5': 3, 'com': 19, 'must': 57, 'be': 15, 'legal': 49, 'drinking': 25, 'age': 4, 'this': 78, 'app': 12, 'remain': 68, 'seated': 69,
'when': 90, 'using': 87, 'stop': 73, 'if': 40, 'any': 11, 'discomfort': 24}

Encoded result is:
[[0.         0.         0.         0.         0.         0.10915389
  0.10915389 0.10915389 0.         0.10915389 0.10915389 0.
  0.         0.         0.10915389 0.         0.10915389 0.
  0.         0.         0.10915389 0.10915389 0.10915389 0.
  0.         0.         0.         0.         0.         0.10915389
  0.         0.10915389 0.15532779 0.10915389 0.         0.
  0.10915389 0.         0.32746168 0.         0.         0.
  0.10915389 0.10915389 0.10915389 0.         0.         0.10915389
  0.         0.         0.10915389 0.         0.10915389 0.10915389
  0.         0.10915389 0.         0.         0.         0.
  0.         0.31065558 0.         0.07766389 0.         0.10915389
  0.         0.21830779 0.         0.         0.10915389 0.10915389
  0.10915389 0.         0.         0.10915389 0.10915389 0.46598336
  0.         0.10915389 0.23299168 0.10915389 0.         0.10915389
  0.10915389 0.         0.07766389 0.         0.         0.07766389
  0.         0.10915389 0.10915389 0.         0.10915389 0.07766389
  0.10915389]
 [0.19265669 0.09632835 0.38531338 0.09632835 0.09632835 0.
  0.         0.         0.19265669 0.         0.         0.09632835
  0.19265669 0.09632835 0.         0.09632835 0.         0.09632835
  0.09632835 0.09632835 0.         0.         0.         0.28898504
  0.09632835 0.09632835 0.09632835 0.09632835 0.19265669 0.
  0.09632835 0.         0.13707682 0.         0.09632835 0.09632835
  0.         0.09632835 0.         0.09632835 0.09632835 0.09632835
  0.         0.         0.         0.09632835 0.09632835 0.
  0.09632835 0.09632835 0.         0.09632835 0.         0.
  0.09632835 0.         0.09632835 0.09632835 0.09632835 0.09632835
  0.09632835 0.06853841 0.09632835 0.06853841 0.09632835 0.
  0.09632835 0.         0.09632835 0.09632835 0.         0.
  0.         0.09632835 0.09632835 0.         0.         0.20561523
  0.09632835 0.         0.27415364 0.         0.09632835 0.
  0.         0.09632835 0.13707682 0.09632835 0.09632835 0.13707682
  0.09632835 0.         0.         0.09632835 0.         0.13707682
  0.        ]]
```

Figure 30: Sample Dataframe after TF-IDF Encoded

5.1.3 Word Embeddings

Word embedding is a form of representing words and documents using a dense vector representation. The position of a word within the vector space is learned from the text and is based on the words that surround the word when it is used [22] Word embeddings can be trained using the input corpus itself or can be generated using pre-trained word embeddings such as Glove[39], FastText[20], and Word2Vec[32]. Any one of them can be downloaded and used as transfer learning. Using word embedding, semantically similar words will also be more similar in vector space [7]!

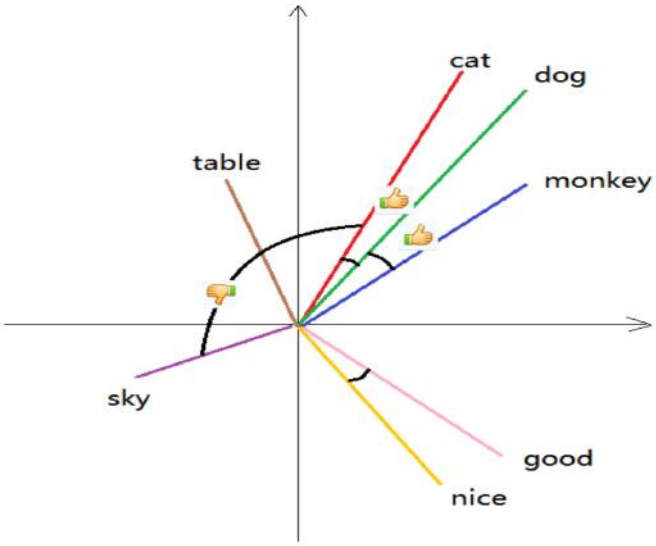

Figure 31: projection of the embedding vectors to 2D [61]

We use pre-trained word embeddings wiki-news-300d-1M.vec[33] in our experiment models, which contains 1 million word vectors trained on Wikipedia 2017, UMBC web based corpus and statmt.org news dataset (16B tokens).

5.2 Models

5.2.1 Naive Bayes Classifier

A Naive Bayes classifier is a probabilistic machine learning model that's used for the classification task. The principle of the classifier is based on the Bayes theorem.

Bayes Theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

Using Bayes theorem [21], we can find the probability of A happening given that B occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is the presence of one particular feature does not affect the other [13]. Hence it is called naive.

For one traing sample, X is given as

$$X = (x_1, x_2, x_3, \ldots., x_n)\,,$$

where x1,x2,…,xn represent the features. By substituting for X and expanding using the chain rule we get

$$P\left(y \mid x_1, \ldots, x_n\right) = \frac{P\left(x_1 \mid y\right) P\left(x_2 \mid y\right) \ldots P\left(x_n \mid y\right) P(y)}{P\left(x_1\right) P\left(x_2\right) \ldots P\left(x_n\right)}.$$

.

Now you can look at the dataset to get the values for each and then enter them into the equation. Observing the denominator does not change for any of the entries in the dataset, the denominator may be eliminated and proportionality may be added.

Finally, we can make the predicted class, using the below formula,

$$y = \text{argmax}_y \, P(y) \prod_{i=1}^{n} P\left(x_i \mid y\right).$$

.

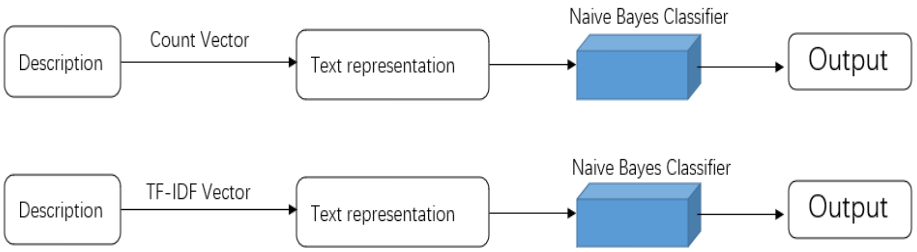

Figure 32: Counter-TFIDF Naive Bayes Classifier

5.2.2 Shallow Neural Networks

A neural network is a mathematical model that is designed to behave similarly to biological neurons and the nervous system. These models are used to recognize complex patterns and relationships that exist within labeled data. A shallow neural network [62] contains mainly three types of layers – input layers, hidden layers, and output layers.

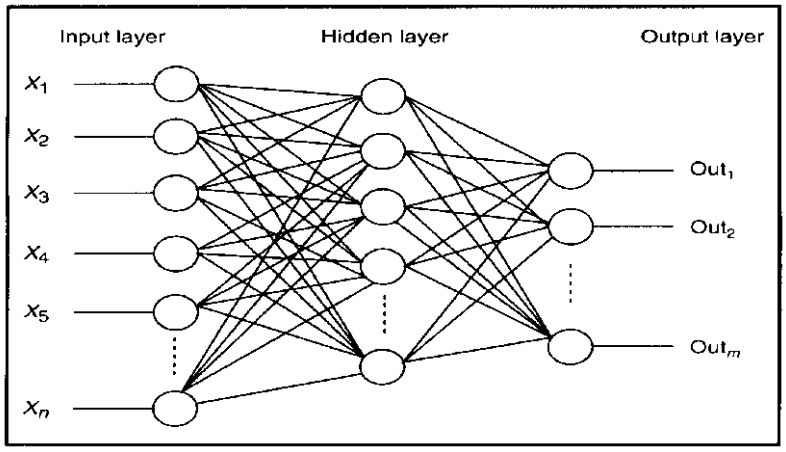

Figure 33: Shallow Neural Networks [34]

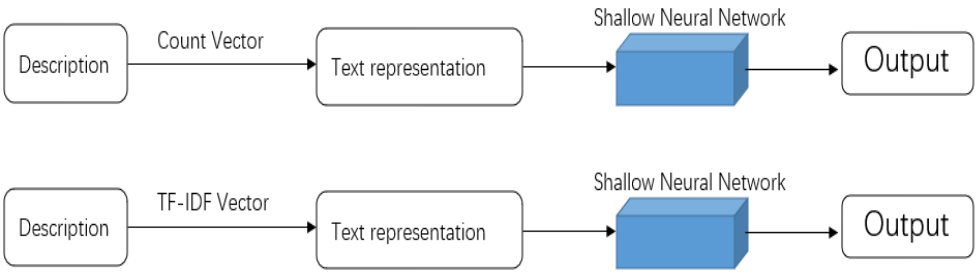

Figure 34: Counter-TFIDF Shallow Neural Network

5.2.3 Convolutional Neural Network

In convolutional neural networks, convolutions over the input layer are used to compute the output. This results in local connections, where each region of the input is connected to a neuron in the output. Each layer applies different filters and combines their results [36].

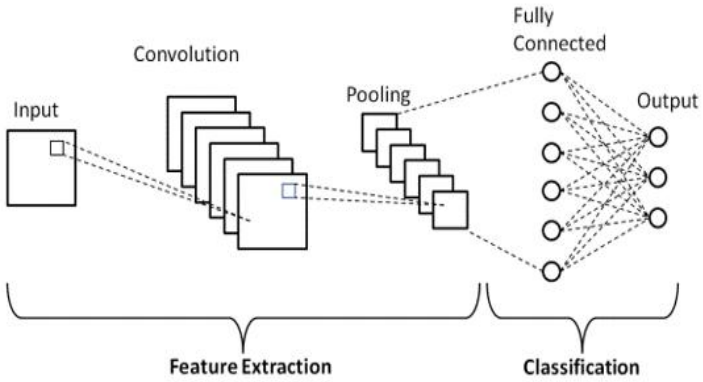

Figure 35: Convolutional Neural Network [11]

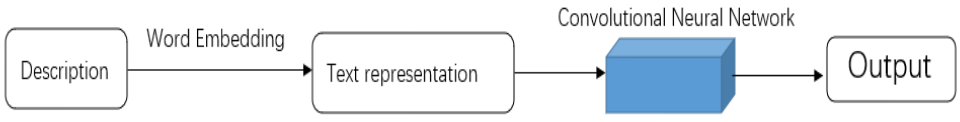

Figure 36: Word embedding Convolutional Neural Network

5.2.4 Transformer

An encoder-decoder structure is present in the transformer [52]. The encoder in the transformer transforms input discrete value sequences into an intermediary continuous value sequence. The next step is for the decoder to produce each token in the output sequence one at a time using the intermediate sequence since the previous token in the output also serves as the input for the subsequent token.

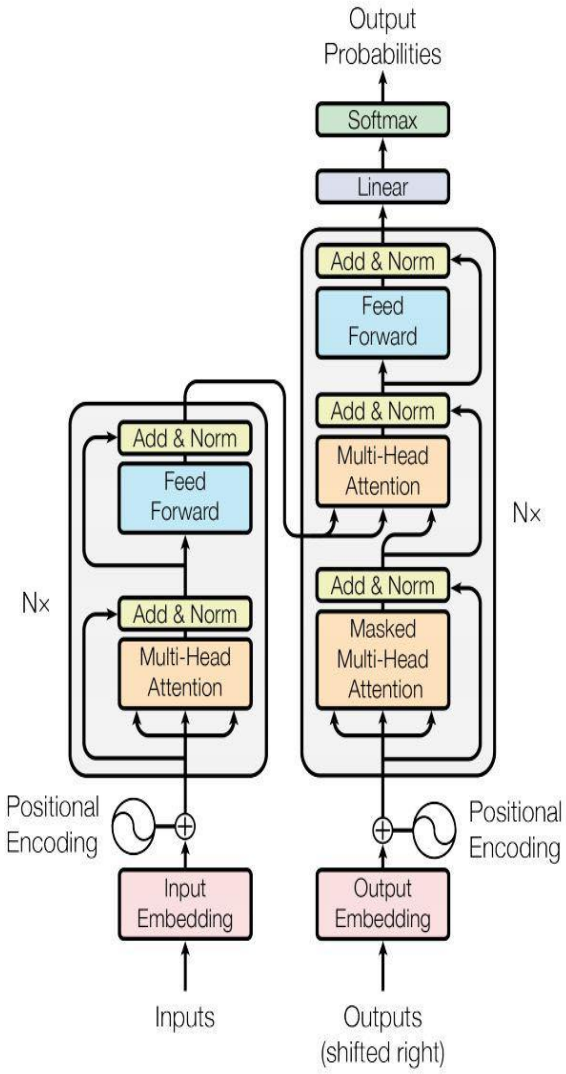

Figure 37: Transformer Model Architecture

It contains a stack of self-attention and fully connected layers in core components encoder and decoder.

Encoder The N identical layers that make up the encoder can be further divided into two sublayers for each layer in the stack. Before entering the encoder stack, the input sequence is first embedded through an embedding layer to have dimension d for each token. The multi-head attention mechanism is the first sublayer of the encoder stack that receives the input x from the layer. The original input is then combined with the output of the multi-head attention mechanism, which is fed to the normalization layer.

The normalization layer's output is then passed to a fully connected feed-forward layer, and this time, a residual connection is used so that the normalization layer follows the full feed-forward layer.

Decoder With the exception of having additional multi-head attention, the decoder is essentially identical to the encoder. The first multi-head attention in the decoder stack is introduced to a mask. The modification's goal is to ensure that output at position k can only refer to the output before position k and to stop positions from attending the unread positions.

Positional Encoding Due to the Transformer's lack of convolution and recurrence, information about the relative and absolute positions of each piece of information in the sequence is added. For this reason, positional encodings of dimension d are added to the input's embeddings before it enters the stacks in order to maintain the ordering and position information. For encoding the odd and even dimension positions, it uses two different functions.

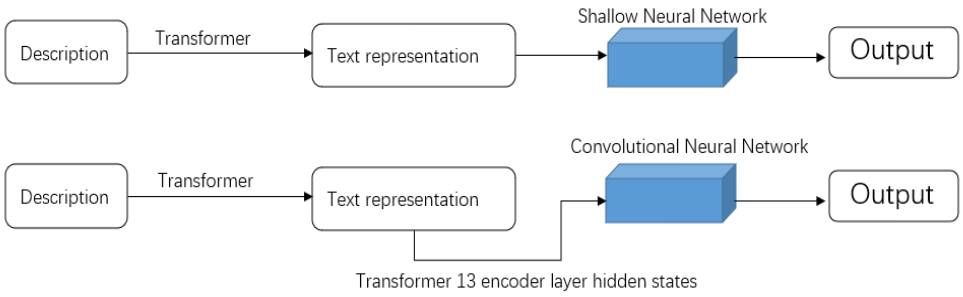

Figure 38: Transformer Shallow-Convolutional Neural Network

5.2.5 FastText

FastText is a fast text classification algorithm created by Facebook's AI Research (FAIR) lab, that has two major advantages over neural network-based classification algorithms. FastText speeds up training and testing while maintaining high accuracy, and fastText does not require a pre-trained word vector, fastText will train the word vector itself [46].

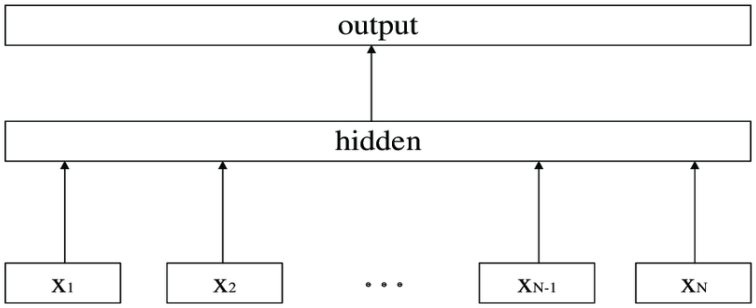

Figure 39: Model architecture of FastText [64]

5.2.6 ResNet

ResNets [14] or Residual Networks are a type of Convolutional Neural Network architecture introduced by Kaiming He. It introduces residual learning to solve the degradation problem: "Overly deep" plain nets have higher training errors.

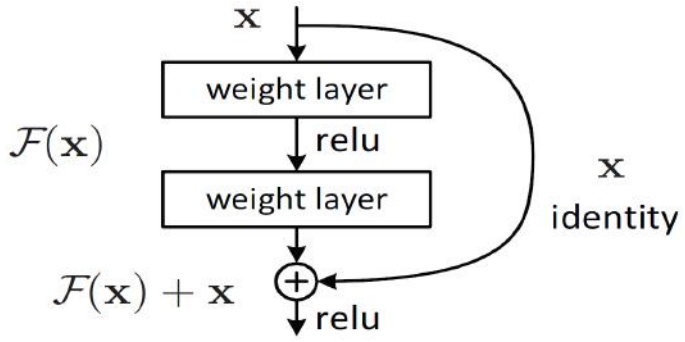

Figure 40: Residual learning: a building block [49]

5.2.7 MobileNetV2

MobileNetV2 is a convolutional neural network architecture that seeks to perform well on mobile devices. It is built on an inverted residual structure where the bottleneck layers are connected by residual connections. Lightweight depthwise convolutions are used in the intermediate expansion layer as a source of non-linearity to filter features [42].

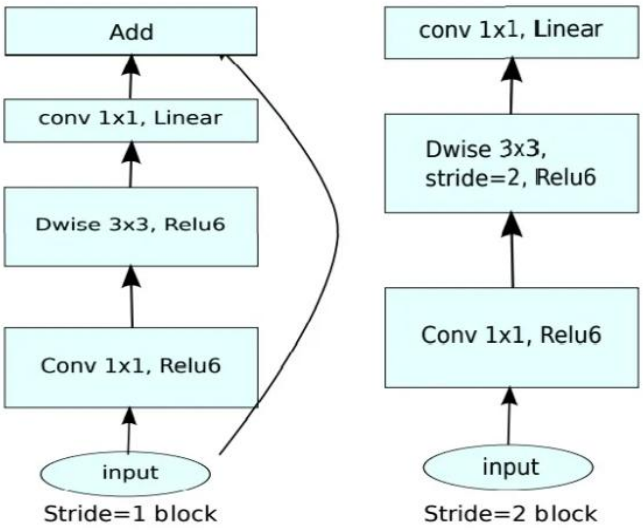

Figure 41: MobileNetV2 [48]

5.3 K-Fold Cross-Validation

One of the most often employed techniques for model evaluation is K-fold cross-validation. Although less common than the validation set technique, this can help us understand our data and model better.

The process contains a single parameter, k, that designates how many groups should be created from a given data sample. As a result, the process is frequently referred to as k-fold cross-validation. When a particular number for k is selected, it may be substituted for k in the model's reference, such as when k=10 is used to refer to a 10-fold cross-validation [37].

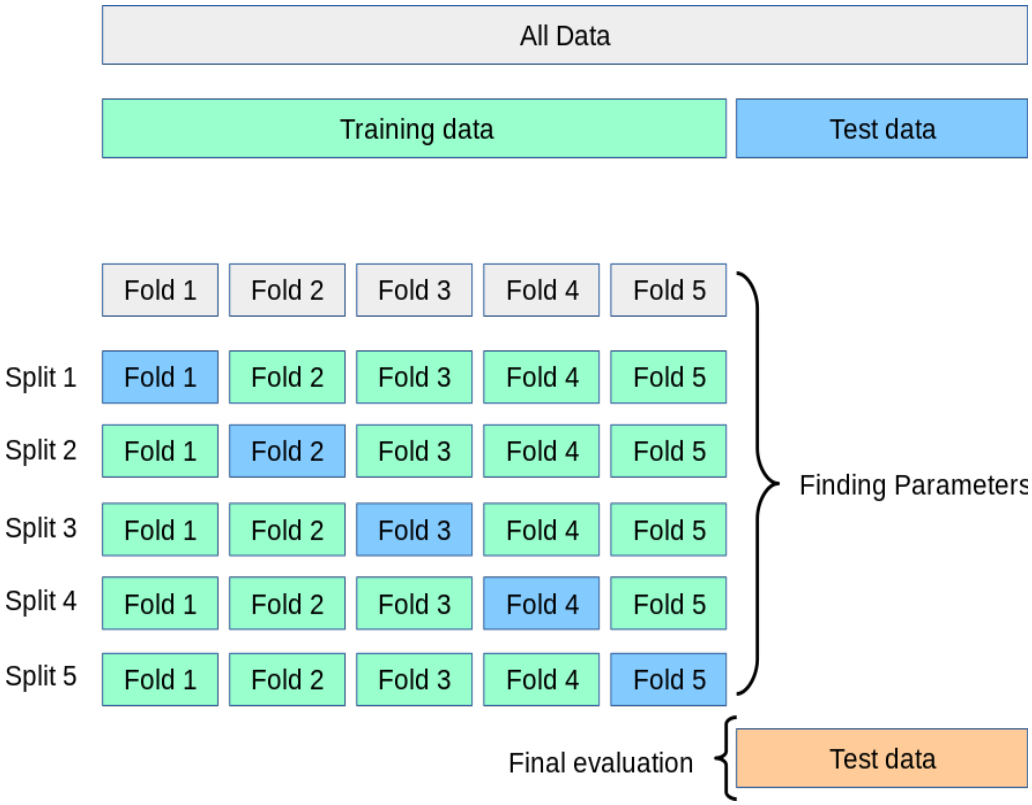

Figure 42: 5-fold cross-validation example [43]

5.3.1 Procedure

Step1: Just randomly shuffle the dataset

Step2: Split the dataset into k groups

Step3: For every distinct group:

Take this group should as a hold-out set

Take the remaining groups as training sets

Fit model to the training data, and evaluate it on the test set

Keep the evaluation result

Step4: Using evaluation results, summarize the models' performance

It's significant that every observation in the data sample is given a unique group and remains there throughout the process. This indicates that each sample has the chance to be used k-1 times to train the model and 1 time in the hold-out set [5].

5.4 Multimodal Machine Learning

Human beings come into contact with the world through multiple sensory organs, such as eyes, ears, and touch. Multimodal Machine Learning studies machine learning problems involving data of different modalities. Common modalities include: visual, text, and sound.They typically come from various sensors, and their internal structures and methods for data formation are very different. Multimodal learning involves the fusion and co-processing of different data sources, with the goal of obtaining more comprehensive information from multiple data sources to help machine learning models understand and process tasks more accurately [29].

In order to make our comfort evaluator model more accurate, we expect that if the model can use more application information such as screenshots and game operation videos instead of only texts, this should allow our model to learn more information about the application, such as teleportation approach, screen style etc., and it should make some more accurate judgments.

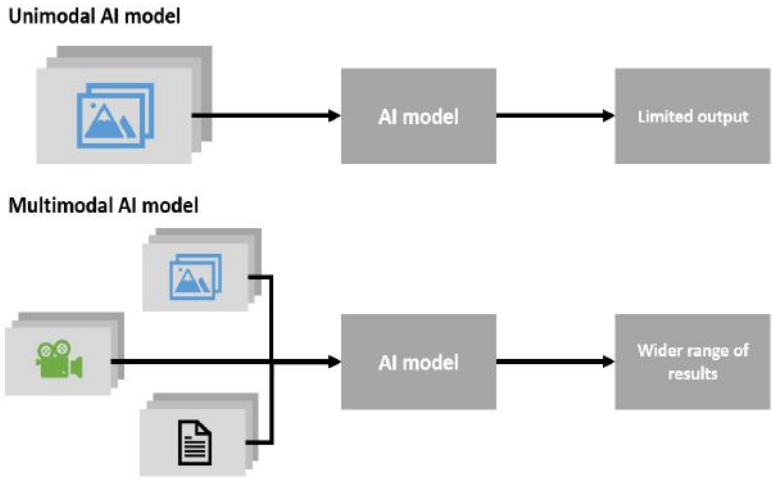

Figure 43: Unimodal AI model VS Multimodal AI model [19]

5.4.1 Fusion schemes

Multimodal fusion methods is to combine features from different modalities to form a comprehensive feature vector to improve the accuracy and reliability of classification or recognition. Among them, common multimodal fusion methods include Early-Fusion, Mid-Fusion and Late-Fusion [24].

Early Fusion Features from different modalities are fused in the input layer. For example, features of image and text information are fused as a multimodal input and then fed into a neural network for classification. Early fusion methods integrate information from different modalities directly into a complete feature vector, and are usually suitable for cases where the modality differences between data sources are small and the features are relatively simple [18].

Mid-fusion The features of different modalities are fused in an intermediate layer. For example, in a convolutional neural network (CNN), image and text information are fed into different convolutional networks, and then features of different modalities are fused in one layer and finally fed into the fully connected layer for classification. By fusing features of different modalities in an intermediate layer, the mid-term fusion method can improve the expression of features, which is suitable for cases where the modality differences of data sources are large and the features are complex [38].

Late Fusion The features of different modalities are fed into different classifiers for processing, and finally the results of different classifiers are fused. For example, image and text information are fed into different neural net-

works separately for training, and finally their outputs are weighted and averaged or stitched together to form the final classification results. By processing the features of different modalities separately and then fusing them in the last step, the late fusion method can avoid the conflict of features of different modalities, which is suitable for the cases where the modality differences of data sources are large, the features are complex, and there are strong correlations between modalities [41].

5.4.2 Text-Text Learning

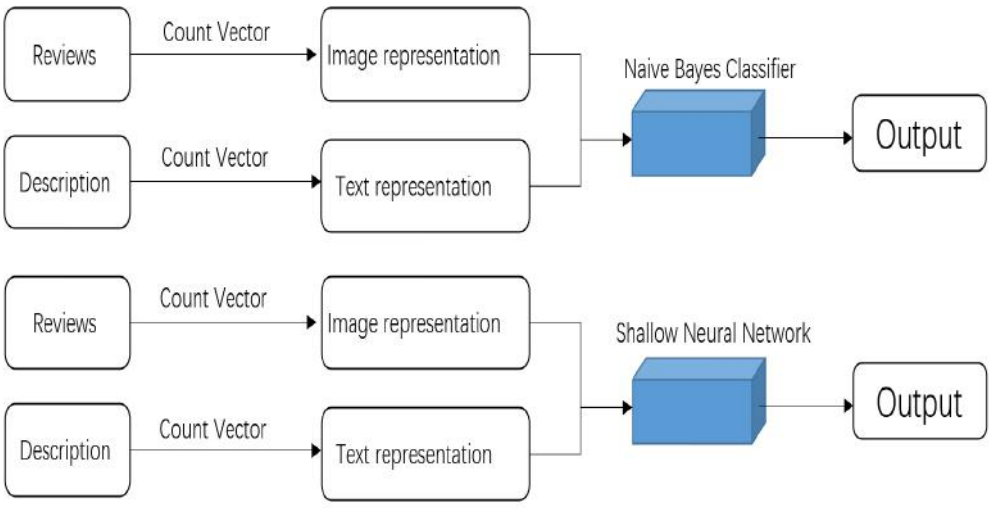

Figure 44: Review-Description Count Vector Bayes Classifier-Shallow Neural Networks

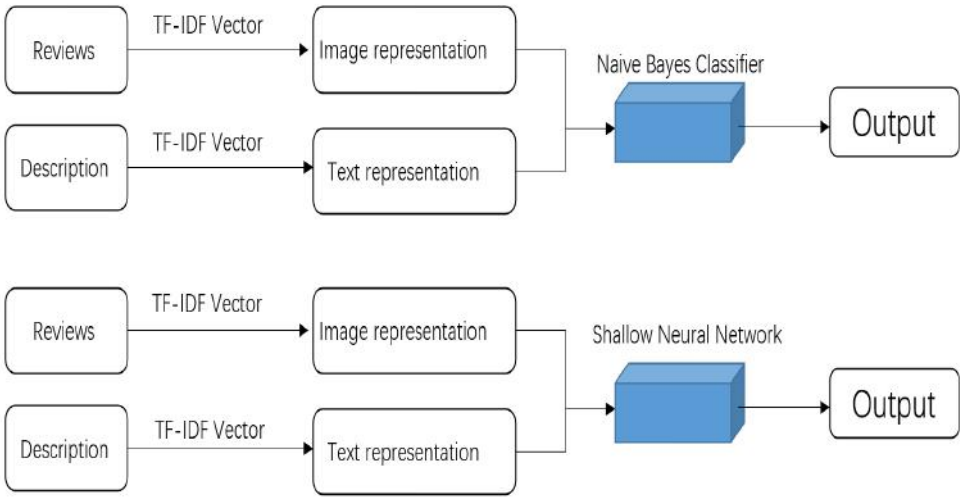

Figure 45: Review-Description TF-IDF Vector Convolutional Neural Network

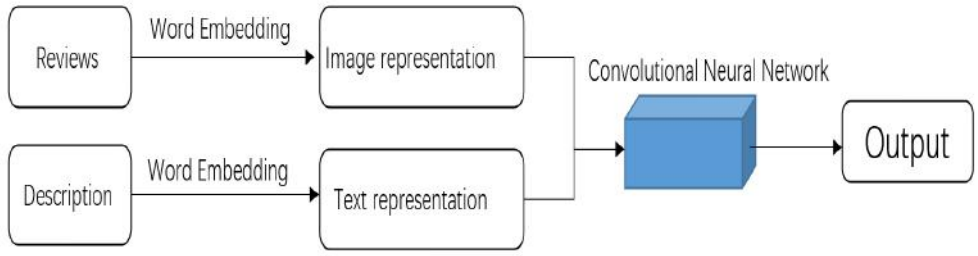

Figure 46: Review-Description Word Embedding Convolutional Neural Network

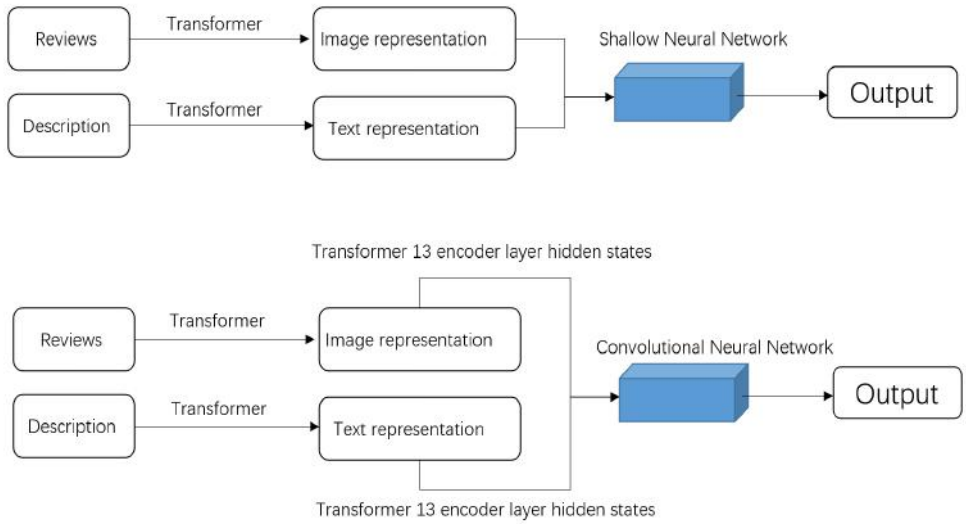

Figure 47: Review-Description Transformer Shallow-Convolutional Neural Network

5.4.3 Text-Text-Image Learning

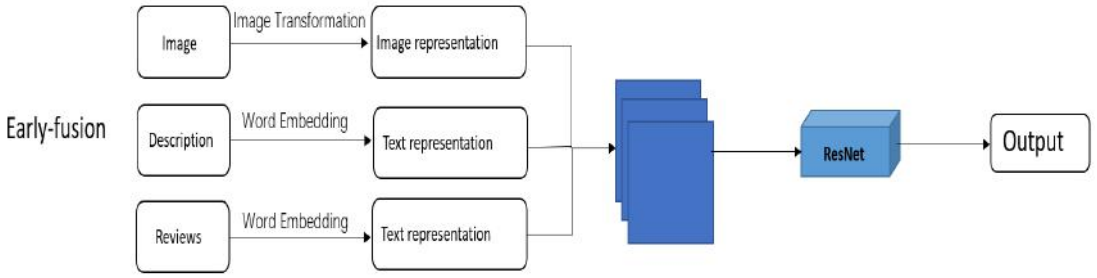

Figure 48: Text-Text-Image early-fusion approach

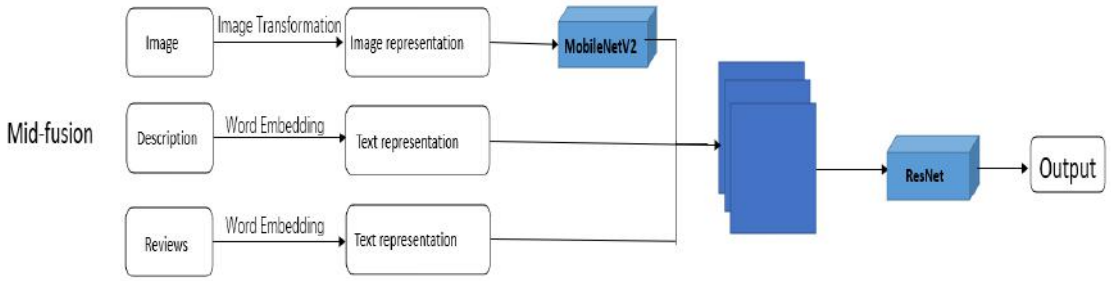

Figure 49: Text-Text-Image mid-fusion approach

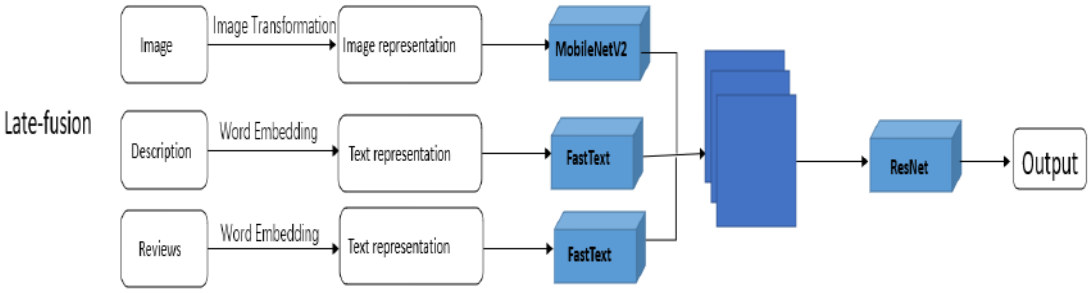

Figure 50: Text-Text-Image late-fusion approach1

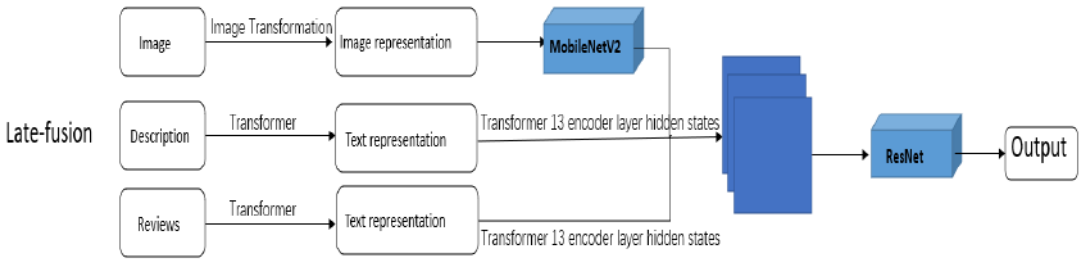

Figure 51: Text-Text-Image late-fusion approach2

5.4.4 Text-Text-Video Learning

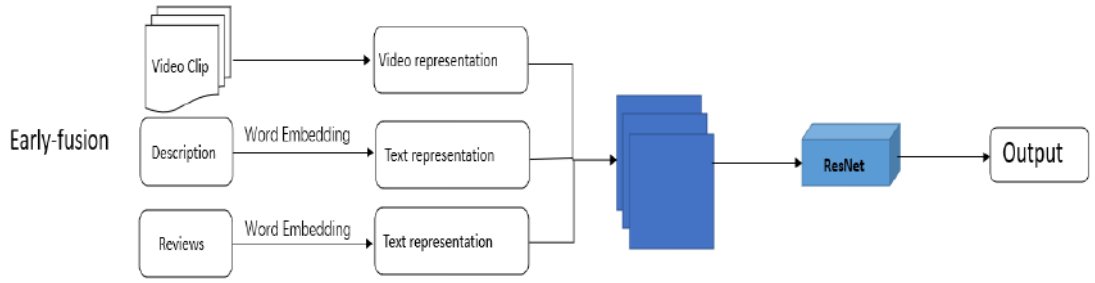

Figure 52: Text-Text-Video early-fusion

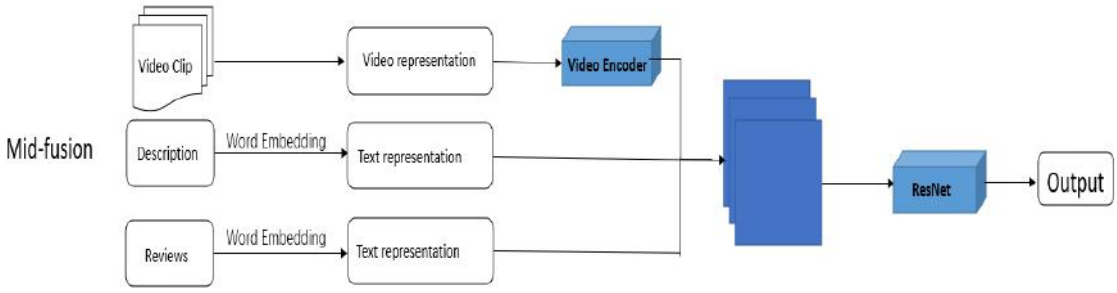

Figure 53: Text-Text-Video mid-fusion

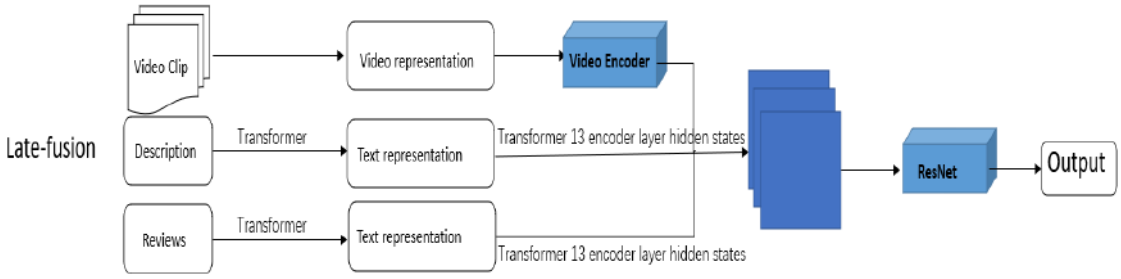

Figure 54: Text-Text-Video late-fusion

6 Experiment Result

6.1 Experiment Result

Model	Accuracy
Description Bayes Classifier (Count Vector)	0.7207
Description Bayes Classifier (Count Vector Oversampled)	0.6723
Description Bayes Classifier (TF-IDF Vector)	0.6301
Description Bayes Classifier (TF-IDF Vector Oversampled)	0.5963

Table 1: Description-Naive Bayes Classifier

Model	Accuracy
Description-Shallow Neural Network (Count Vector)	0.6483
Description-Shallow Neural Network (Count Vector Oversampled)	0.6703
Description-Shallow Neural Network (TF-IDF Vector)	0.6192
Description-Shallow Neural Network (TF-IDF Vector Oversampled)	0.6321

Table 2: Description-Shallow Neural Network

Model	Accuracy
Description-Convolutional Neural Network	0.7074
Description-Convolutional Neural Network (Oversampled)	0.7665

Table 3: Description-Convolutional Neural Network

Model	Accuracy
Description-Transformer	0.8316
Description-Transformer (Oversampled)	0.8325
Description-Transformer-CNN	0.8432
Description-Transformer-CNN (Oversampled)	0.8398

Table 4: Description-Transformers

Model	Accuracy
Description-Review Bayes Classifier (Count Vector)	0.7103
Description-Review Bayes Classifier (Count Vector Oversampled)	0.6981
Description-Review Bayes Classifier (TF-IDF Vector)	0.6502
Description-Review Bayes Classifier (TF-IDF Vector Oversampled)	0.6233

Table 5: Description-Review Bayes Classifier

Model	Accuracy
Description-Review SNN (Count Vector)	0.6501
Description-Review SNN (Count Vector Oversampled)	0.6788
Description-Review SNN (TF-IDF Vector)	0.6455
Description-Review SNN (TF-IDF Vector Oversampled)	0.6333

Table 6: Description-Review Shallow Neural Network

Model	Accuracy
Description-Review CNN (Word Embedding)	0.7221
Description-Review CNN (Word Embedding Oversampled)	0.7443

Table 7: Description-Review Convolutional Neural Network

Model	Accuracy
Description-Review Transformer	0.8401
Description-Review Transformer (Oversampled)	0.8356
Description-Review Transformer-CNN	0.8397
Description-Review Transformer-CNN (Oversampled)	0.8434

Table 8: Description-Review Transformers

Model	Accuracy
Description-Review-Image (Early-Fusion)	0.7322
Description-Review-Image MobiletNetV2 (Mid-Fusion)	0.7613
Description-Review-Image MobiletNetV2 FastText (Late-Fusion)	0.7829
Description-Review-Image MobiletNetV2 Transformer (Late-Fusion)	0.8475

Table 9: Description-Review-Image learning

Model	Accuracy
Description-Review-Video Cross Encoder (Early-Fusion)	0.8398
Description-Review-Video Cross Encoder (Mid-Fusion)	0.8492
Description-Review-Video Transformer (Late-Fusion)	**0.8514**

Table 10: Description-Review-Video learning

6.2 Analysis

In the task of predicting the comfort level for our VR application, the accuracy results of the two models, Naive Bayes Classifier and Shallow Neural Networks, show that the feature representation of the counter vector is better than that of the TF-IDF, which is more representative of some information about the comfort level.

Dealing with the data imbalance problem, for the neural network model, the training set is processed with oversampling, which can improve the accuracy of the corresponding model to some extent.

Multiple modality information fed to our model is effective, as seen by the experimental results for Text-Text, Text-Text-Image, and Text-Text-Video, the models perform better than models with description only. And the more modality information the model is fed, the more positively correlated the final accuracy of the models. The experimental results meet our initial expectation that the transformer model with review, description, and video fed has the highest accuracy rate of 85%.

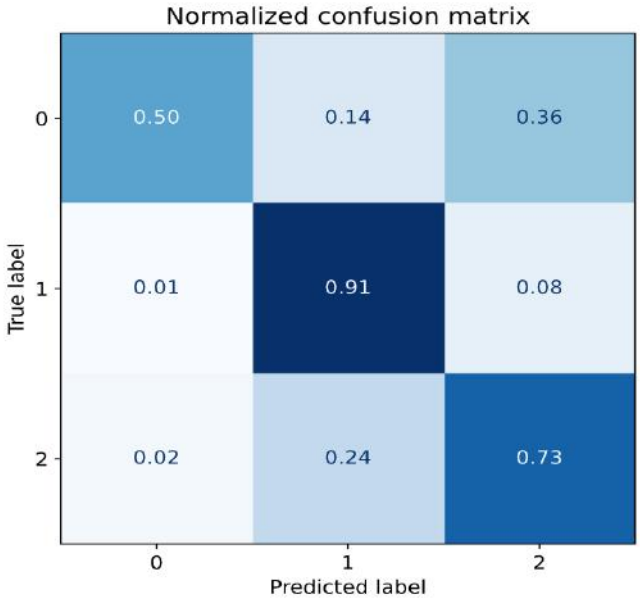

Figure 55: Description-Transformer confusion matrix

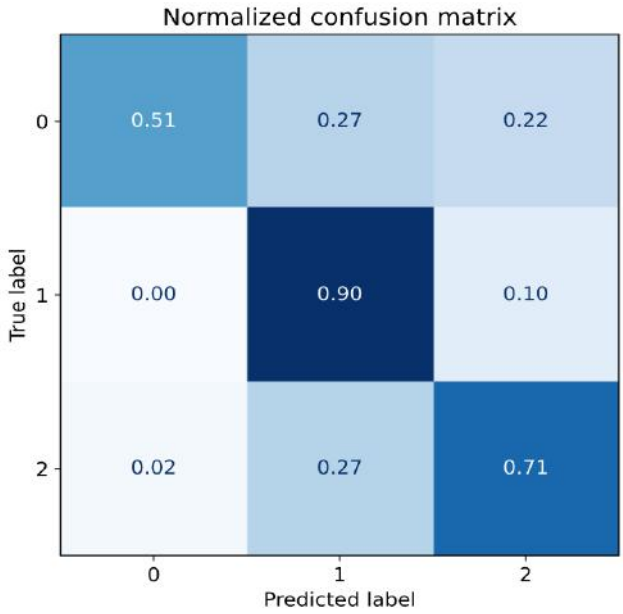

Figure 56: Description-Transformer-CNN confusion matrix

	text	label	predicted_label	loss
437	VRChat offers an endless collection of social ...	0	1	11.104321
73	Chromesthesia: seeing sounds as colors. Immers...	0	1	10.766850
707	Neverout is a unique puzzle game with innovati...	0	1	10.747291
473	Navigate city rooftops like a parkour pro. Gra...	1	0	10.579171
25	Make beats in VR - This is not a rhythm game! ...	2	1	10.390916
371	Explore a virtual anatomical heart in the 'Rea...	2	1	10.143176
195	Whirligig is an immersive media player for pla...	2	1	10.124323
135	Sharecare Windows offers an escape to immersiv...	2	1	10.124310
130	Holopoint is pure archery madness. Fight your ...	0	1	10.038361
552	War Online: Pacific is the first in a series o...	0	1	10.031110
96	You believed tower defense is only about build...	0	1	10.015985
17	Roll your marble to the goal. Try to beat the ...	2	1	9.911517
519	You are trapped by a group of mischievous monk...	2	1	9.856830
810	Explore different locations and face your fear...	1	2	9.650606
674	Digital Domain's Monkey King™ is a visually-ca...	2	1	9.506454

Figure 57: Description-Transformer Data Samples with Highest Loss (Cross Entropy)

	text	label	predicted_label	loss
42	Welcome to the world of Mini Bowling! In this ...	2	1	10.031266
244	Welcome to the world of Mini Bowling! In this ...	2	1	10.031266
217	Have you ever dreamed of what it would be to s...	2	1	10.028459
142	"All you need is 20 seconds of insane courage ...	0	1	9.992984
316	What is the site of the world's worst nuclear ...	2	1	9.945721
704	Chernobyl VR Project is a virtual experience w...	0	1	9.917863
773	Hyper Dash is a VR multiplayer team based shoo...	2	1	9.883032
203	XING: The Land Beyond is an atmospheric first-...	2	1	9.737484
183	At the edge of the future, people of all races...	1	2	9.728790
195	Inspired by classic survival horror gameplay, ...	1	2	9.675558
135	Georgian winemaking method is a part of UNESCO...	2	1	9.657598
645	Georgian winemaking method is a part of UNESCO...	2	1	9.657598
667	Gangsa Bali VR. You playing game with "Panggul...	2	1	9.608642
55	Become Orion13 – a robot mercenary who fights ...	0	1	9.596140
478	**Note: This game is a VR-only experience and ...	1	2	9.506377

Figure 58: Description-Transformer-CNN Data Samples with Highest Loss (Cross Entropy)

7 Conclusion and Discussion

We successfully trained a Transformer-based model by using multi-modal learning, combining information from three modalities: description, reviews, and video. The model has an accuracy rate of 85% and performs quite well. We can apply it to VR applications as a qualified comfort level rating evaluator to help the users, developers, and platforms.

To further improve our model, we can also consider adding information such as the category tags the VR application has, the music it uses, etc. The introduction of such information should also help enhance the classification ability of models and prediction accuracy.

References

[1] Qadir AK. Metaverse: A shared virtual environment!, 2022. [Online; accessed 22-December-2022].

[2] Thomas Alsop. Number of virtual reality (vr) and augmented reality (ar) users in the united states from 2017 to 2023, 2023. [Online; accessed 01-April-2023].

[3] ELYSE BETTERS. Htc viveport subscription: How does it work and how much is it?, 2017. [Online; accessed 22-December-2022].

[4] Marius Borcan. Tf-idf explained and python sklearn implementation, 2020. [Online; accessed 22-December-2022].

[5] Jason Brownlee. A gentle introduction to k-fold cross-validation, 2018. [Online; accessed 01-April-2023].

[6] J. Clement. Number of monthly active users on gaming platform steam worldwide from 2017 to 2021, 2022. [Online; accessed 01-April-2023].

[7] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, page 160–167, New York, NY, USA, 2008. Association for Computing Machinery.

[8] Josh Constine. The "oculus platform" marketplace for virtual reality app launches this fall, 2014. [Online; accessed 22-December-2022].

[9] International Data Corporation. Sustained prosperity, ranking first in the world - china's ar/vr market in the next five years, 2022. [Online; accessed 01-April-2023].

[10] datacamp. Making web crawlers using scrapy for python, 2019. [Online; accessed 22-December-2022].

[11] Jonathan Fregoso, Claudia I. Gonzalez, and Gabriela E. Martinez. Optimization of convolutional neural networks architectures using pso for sign language recognition. *Axioms*, 10(3), 2021.

[12] Beat Games. Beat saber, 2019. [Online; accessed 01-April-2023].

[13] Rohith Gandhi. Naive bayes classifier, 2018. [Online; accessed 22-December-2022].

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[15] James Howell. Metaverse vs. virtual reality: Key differences, 2022. [Online; accessed 22-December-2022].

[16] HTC. Viveport™, the app store for your life in virtual reality, launches today, 2016. [Online; accessed 01-April-2023].

[17] Infobae. The metaverse could contribute 2.8 percent to global gdp in its first decade, 2022. [Online; accessed 22-December-2022].

[18] Noussaiba Jaafar and Zied Lachiri. Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance. *Expert Systems with Applications*, 211:118523, 2023.

[19] Shehmir Javaid. Multimodal learning: Benefits 3 real-world examples in 2023, 2023. [Online; accessed 01-April-2023].

[20] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification, 2016.

[21] James Joyce. Bayes' Theorem. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition, 2021.

[22] Dhruvil Karani. Introduction to word embedding and word2vec, 2018. [Online; accessed 22-December-2022].

[23] Jaehyun Choi Yeongcheol Choe Mungyeong Kim, Hyun K. Park. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69:66–73, 2018.

[24] Adrienne Kline. Multimodal machine learning: Data fusion, 2022. [Online; 01-April-2023].

[25] Ben Lang. Oculus app lab lets devs distribute apps on quest while bypassing curation process, 2021. [Online; accessed 22-December-2022].

[26] David Needle Linda Tucci. Feature what is the metaverse? an explanation and in-depth guide, 2022. [Online; accessed 21-December-2022].

[27] Huaishao Luo, Lei Ji, Botian Shi, Haoyang Huang, Nan Duan, Tianrui Li, Jason Li, Taroon Bharti, and Ming Zhou. Univl: A unified video and language pre-training model for multimodal understanding and generation. *arXiv preprint arXiv:2002.06353*, 2020.

[28] Xiaoju Ma and Min Yan. Design and implementation of craweper based on scrapy. *Journal of Physics: Conference Series*, 2033:012204, 09 2021.

[29] Purvanshi Mehta. Multimodal deep learning, 2018. [Online; 01-April-2023].

[30] Giovanna Menardi and Nicola Torelli. Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28:92–122, 2012.

[31] Meta. View comfort ratings for meta quest store content, 2022. [Online; accessed 22-December-2022].

[32] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.

[33] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[34] Srija Neogi. Exploring multi-class classification with deep learning, 2020. [Online; accessed 22-December-2022].

[35] OpenCV. Opencv: Open source computer vision library, 2023. [Online; 01-April-2023].

[36] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks, 2015.

[37] Shanthababu Pandian. K-fold cross validation technique and its essentials, 2022. [Online; accessed 01-April-2023].

[38] Maciej Pawłowski, Anna Wróblewska, and Sylwia Sysko-Romańczuk. Effective techniques for multimodal data fusion: A comparative analysis. *Sensors*, 23(5), 2023.

[39] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics.

[40] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640, 2015.

[41] Gaurav Sahu and Olga Vechtomova. Dynamic fusion for multimodal data. *CoRR*, abs/1911.03821, 2019.

[42] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019.

[43] Scikit-learn. Cross-validation: evaluating estimator performance, 2019. [Online; accessed 01-April-2023].

[44] Bruno Stecanella. Understanding tf-id: A simple introduction, 2019. [Online; accessed 22-December-2022].

[45] Mike Straw. Steam sale dates for the rest of 2023 revealed, 2023. [Online; accessed 01-April-2023].

[46] Nishan Subedi. Fasttext: Under the hood, 2018. [Online; accessed 22-December-2022].

[47] VIVE Team. Introducing viveport, 2016. [Online; accessed 22-December-2022].

[48] Sik-Ho Tsang. Review: Mobilenetv2 — light weight model (image classification), 2019. [Online; accessed 01-April-2023].

[49] Muneeb ul Hassan. Resnet (34, 50, 101): Residual cnns for image classification tasks, 2019. [Online; accessed 01-April-2023].

[50] UPLOAD. What is sidequest and why is it important for quest users?, 2022. [Online; accessed 22-December-2022].

[51] Valve. Half-life: Alyx, 2020. [Online; accessed 01-April-2023].

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017.

[53] VIRTUALWARE. Virtualware partners with leading manufacturers to bring the best vr solution, 2020. [Online; accessed 22-December-2022].

[54] VIVE. Get started developing for vive, 2023. [Online; accessed 22-December-2022].

[55] Oculus VR. Introducing app lab: A new way to distribute oculus quest apps, 2021. [Online; accessed 22-December-2022].

[56] WarpFrog. Blade and sorcery, 2018. [Online; accessed 01-April-2023].

[57] Wikipedia. Metaverse, 2022. [Online; accessed 21-December-2022].

[58] Wikipeida. Oculus, 2023. [Online; accessed 01-April-2023].

[59] Wikipeida. Steam (service), 2023. [Online; accessed 01-April-2023].

[60] XpertVR. 6 best free sidequest games for oculus quest/quest 2, 2022. [Online; accessed 22-December-2022].

[61] Zack. Rasa (chatbot) mechanism/work flow simply explained!, 2018. [Online; accessed 22-December-2022].

[62] Jeremy Zhang. Build a shallow neural network, 2020. [Online; accessed 22-December-2022].

[63] Alice Zheng and Amanda Casari. *Feature engineering for machine learning: principles and techniques for data scientists*. " O'Reilly Media, Inc.", 2018.

[64] Rui Zhu, Delu Yang, and Yang Li. Learning improved semantic representations with tree-structured lstm for hashtag recommendation: An experimental study. *Information*, 10:127, 04 2019.

[65] Hafidz Zulkifli. Word vectors and lexical semantics — introduction to count-based vectors, 2019. [Online; accessed 22-December-2022].