From Videos to Requirement: A Data-Driven Approach for Finding Requirements Relevant Feedback from TikTok and YouTube

by

Manish Sihag
B.Tech., IK Gujral Punjab Technical University, 2019

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

From Videos to Requirement: A Data-Driven Approach for Finding Requirements Relevant Feedback from TikTok and YouTube

by

Manish Sihag
B.Tech., IK Gujral Punjab Technical University, 2019

Supervisory Committee

Dr. Daniela Damian, Supervisor
(Department of Computer Science)

Dr. Alexandra Branzan Albu,
(Department of Electrical and Computer Engineering)

ABSTRACT

The proliferation of videos as a popular medium of engagement, influence, and communication has made them an important source of user feedback for organizations. Social media platforms, such as TikTok and YouTube, have witnessed a significant surge in user-generated video content. These platforms have become a hub for users to share their opinions, experiences, and feedback about various products and services. This wealth of user-generated video content presents a unique opportunity for organizations to utilize a vast source of feedback that can inform product development decisions. Furthermore, social media platforms offer a wide reach, with millions of users creating and sharing videos on diverse topics, industries, and products. Given the potential of social media as a source of user feedback, this study aims to explore the potential of TikTok and YouTube, two widely used social media platforms known for video content, for extracting requirements-relevant feedback from videos. The study involves analyzing audio and visual text, as well as metadata (such as descriptions and titles), from 6276 videos of 20 popular products across various industries. State-of-the-art deep learning transformer-based models are employed to classify 3097 videos containing requirements-relevant information. Relevant videos are then clustered to identify multiple themes of requirements-relevant feedback for each of the 20 products. This feedback can subsequently be refined into requirements artifacts. The findings reveal that product ratings (including features, design, and performance), bug reports, and usage tutorials are persistent themes identified from the videos. The study demonstrates that video-based social media platforms like TikTok and YouTube can provide valuable user insights, making them a powerful and novel resource for companies seeking to improve customer-centric development.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

Daniela Damian for her invaluable guidance and mentorship throughout my journey as a graduate research student.

The SEGAL Group for their support, suggestions, and consistent assistance throughout my journey as a graduate research student.

Zane Li, Nowshin Nawar Arony and Kezia Devathsan for our friendship, unwavering support, and invaluable assistance, making this research possible while creating memorable moments of joy.

My family for their support, love, and belief in me during my academic journey.

My wife for her constant support and motivation throughout my endeavors.

Set thy heart upon thy work, but never on its reward.
- Bhagavad Gita

We are kept from our goal, not by obstacles, but by a clear path to a lesser goal
- Bhagavad Gita

DEDICATION

I dedicate this thesis to my parents and my beloved wife, Jyoti Rana, whose presence has been an incredible source of strength. I am deeply grateful for their enduring love, kindness, and unwavering support

Chapter 1

Introduction

The rise of videos as a prominent form of communication and content creation cannot be understated. Videos are a very popular medium for social media and communication [64]. Social media platforms, such as TikTok and YouTube, have witnessed a significant surge in user-generated video content. With TikTok being one of the world's most popular video-based social media platforms [12,64], and YouTube reaching astronomical magnitudes in terms of user-generated content [24]. These platforms have become a hub for users to share their opinions, experiences, and feedback about various products and services. Furthermore, online videos provide an immersive experience for viewers, allowing them to visually and audibly engage with the content in a way that other forms of media may not offer. While previous research has primarily focused on analyzing the comments section of videos to gather user feedback [14, 29, 39], it is essential to recognize that videos themselves possess a wealth of valuable data. Videos combine both audio and visual components, providing a comprehensive and multifaceted source of information.

Additionally, metadata associated with videos, including descriptions, titles, and upload dates, can offer contextual insights into the content. we incorporated audio transcripts, text from captions and subtitles, and metadata into our analysis to gain an understanding of the information conveyed in videos. We converted the audio track of videos into transcripts using OpenAI's Whisper [48] which is an advanced speech recognition model, it employs a Transformer-based encoder-decoder architecture [60] to provide highly accurate transcriptions. Captions and subtitles, when available, further contribute textual information that can enhance the analysis. Furthermore, metadata provides additional context that helps interpret the content and understand factors such as the video's creation date or the creator's description. This approach

aligns with the growing recognition of the value of multimodal data analysis, which combines different types of data to uncover deeper meanings and patterns [46].

Paying attention to the direction of CrowdRE (Crowdsourcing in Requirements Engineering) research is of utmost importance for companies seeking to enhance their requirements' elicitation processes [19, 21]. CrowdRE leverages the collective intelligence of a diverse group of individuals to gather a wide range of feedback and perspectives. Organizations can utilize a vast pool of user-generated videos by harnessing the power of social media platforms such as TikTok and YouTube, thus exponentially increasing the amount of feedback available for analysis [20]. The process we propose in this study holds great potential for transforming video content into requirements-relevant feedback that can significantly impact companies' requirements and development activities.

We present a data-driven exploratory study on leveraging user-generated videos from TikTok and YouTube to identify requirements-related user feedback from 6276 videos of 20 popular products across 20 distinct products. We aim to capture a wide spectrum of user feedback and extract valuable insights that can inform the development and enhancement of these offerings by encompassing a diverse range of products. Our approach involves extracting textual data from audio and visual content from the videos and processing using natural language processing (NLP) and machine learning (ML) techniques to uncover important user feedback that may not be captured through traditional elicitation methods.

Our research adds to the existing body of knowledge by exploring the use of social media as a valuable data source for product development and analyzing user feedback. We uncover the advantages of leveraging videos as a rich data source and highlight the potential of employing advanced techniques like NLP and ML to analyze video content effectively. Through our study, we shed light on the strengths of utilizing videos and demonstrate the opportunities that arise from leveraging NLP and ML in video data analysis.

1.1 Motivation

The motivation behind this study stems from the increasing importance of online videos as a medium for engagement, communication, and content creation [64]. With the rise of social media platforms like TikTok and YouTube, videos have become a significant source of user feedback that organizations cannot afford to overlook. These

platforms attract millions of users who share their opinions, experiences, and preferences through video content. Previous research has primarily focused on analyzing the comments section of videos, neglecting the wealth of information present in the videos [14, 29, 39] and analyzing sentiment of the videos [46]. By exploring the potential of TikTok and YouTube as sources of requirements-related user feedback, this study aims to bridge the gap in understanding and leveraging the data-rich nature of videos. Furthermore, the application of natural language processing (NLP) and machine learning (ML) techniques to analyze video data opens up new opportunities for extracting relevant information and generating requirements artifacts. Leveraging these advanced techniques can enhance the efficiency and accuracy of requirement elicitation processes, enabling organizations to make informed decisions and improve their product development strategies.

1.2 Research Questions

In this study, we aimed to address two primary research questions that guided our investigation:

RQ1 How can video-based social media be used to identify requirements relevant user feedback?

RQ2 What are the main users feedback themes that we can identify?

RQ3 How do the different social media platforms and their video content affect user feedback?

These research questions formed the foundation for our investigation into the potential of video-based social media as a valuable source of requirements relevant user feedback.

1.3 Methodology

The methodology was designed to find the effective approach for identifying requirements-relevant user feedback and extracting valuable insights that can inform product development and improvement. To begin, we collected a diverse set of videos from the two popular video-based social media platforms. We selected 20 different products

across various industries to ensure a representative sample. These products included software applications, consumer electronics, automotive products, and more.

Next, we focused on the audio content of the videos. To extract the audio content, we utilized automatic speech recognition techniques. This allowed us to transcribe the spoken words within the videos and convert them into textual data for further analysis. By capturing the audio content, we could gain insights into the specific comments and opinions expressed by users regarding the products. In addition to the audio content, we employed optical character recognition (OCR) techniques to extract visual text from the videos. This involved analyzing the textual elements that appeared on-screen, such as captions, subtitles, and other visual annotations. By extracting visual text, we aimed to capture additional user feedback that may not be explicitly expressed in the audio content. To classify the extracted textual data, we utilized deep learning models GPT-2 [49], BERT [16], RoBERTa [35], XLM-RoBERTa [13], and ALBERT [32]. These models have shown effectiveness in natural language processing tasks and can accurately classify user feedback into relevant and non-relevant categories. We were able to develop a classification system that could automatically identify user feedback that is specifically related to product requirements.

Finally, we applied clustering techniques to identify common themes and patterns within the user feedback. This allowed us to group similar feedback together, providing a holistic view of the prevalent issues and concerns expressed by users across different products and industries. By identifying these themes, we could gain insights into the specific areas where products may require improvement or modification.

1.4 Research Contributions

This study provides four meaningful contributions for both researchers and practitioners. This study

- demonstrates the efficacy of leveraging videos from popular video focused social media platforms, namely TikTok and YouTube, as valuable sources of user feedback for requirements elicitation.
- contribute to the methodological domain by exploring different approaches to analyzing the textual and visual content of videos.

- presents the most effective machine learning models (GPT-2 and RoBERTa) to classify the audio and visual content into requirements relevant user feedback.
- presents a list of requirements relevant user feedback themes for software, phone, computer, and automotive industries which can be further refined into requirements.

1.5 Research Publications

Our research method and contributions culminated in the following publication:

- M. Sihag, Z. S. Li, A. Dash, N.N. Arony, K. Devathasan, N Ernst, A. Albu and D. Damian, "A Data-Driven Approach for Finding Requirements Relevant Feedback from TikTok and YouTube," in 2022 IEEE 31th International Requirements Engineering Conference (RE), 2023,

1.6 Thesis Outline

This thesis is organized as follows:

Chapter 1: Introduction chapter introduces the research topic, outlines the motivation for the study, provides an overview of the research methodology, and highlights the contributions of the research.

Chapter 2: Background and Related Work chapter explores the relevant background information and previous studies related to the research topic.

Chapter 3: Methodology chapter describes the research methodology in detail, including the data collection process, the selection of variables, the research design, and the analytical techniques used.

Chapter 4: Research Findings chapter presents the findings of the study, showcasing the results obtained from the data analysis. It highlights the key discoveries, patterns, and trends identified during the research process and provides a comprehensive overview of the research outcomes.

Chapter 5: Discussion and Implications the research findings are discussed in-depth, considering their implications and significance. It explores the implications of the research results for theory, practice, and future research directions.

Chapter 6: Threats to Validity chapter identifies and discusses the potential threats to the validity of the research findings. It examines factors that may have influenced the accuracy, reliability, or generalizability of the results and provides a critical assessment of the research's Construct, internal and external validity.

Chapter 7: Conclusion and Future Work chapter summarizes the main findings of the study, recaps the research objectives, and discusses the contributions of the research. It also offers insights into the practical implications of the findings and suggests avenues for future research to build upon the current study.

Chapter 2

Background and Related Work

In this chapter, we introduce the key concepts that provide the background for our research. We discuss CrowdRE, an approach that involves engaging a large group of users to gather feedback for product development. We also explore the limitations of traditional requirements' elicitation methods and highlight the influence of social media in user feedback.

2.1 Requirement Engineering (RE)

Requirement engineering is a critical discipline in software development that involves understanding, documenting, and managing the needs and expectations of stakeholders. According to Zave [65], it encompasses the real-world goals, functions, and constraints of software systems, along with their precise specifications and evolution over time. The process of requirement engineering comprises multiple interconnected steps, including elicitation, analysis, requirement description, system modeling, validation, and management [56] [43] [37] [44]. Each of these steps contributes to the overarching objective of capturing and defining the requirements necessary for successful software development.

During the elicitation phase, stakeholders are engaged to discover and comprehend their requirements, as well as establish the boundaries and scope of the system. To ensure a comprehensive gathering of requirements, various techniques are employed. Zhang [66] classified these techniques into four categories: conversational, observational, analytical, and synthetic approaches. Conversational techniques involve direct, face-to-face interactions with stakeholders, such as conducting interviews and organiz-

ing brainstorming sessions. Through these conversations, valuable insights are gained, allowing for a deeper understanding of stakeholder needs and perspectives. Observational techniques focus on observing and monitoring ongoing work. This can include approaches like ethnography, where analysts immerse themselves in the environment where the software will be used, observing user behavior, work practices, and interactions. By studying users in their natural settings, analysts can gain insights into their requirements and preferences. Analytical techniques rely on extracting information from existing documentation or code. This can involve analyzing previous requirements documents, system specifications, or examining the codebase of an existing system. By studying these artifacts, analysts can identify patterns, extract requirements, and leverage existing solutions or components. Synthetic techniques integrate elements from various other approaches to create a comprehensive requirement elicitation process. The synthetic method form a coherent understanding by combining conversation, observation, and analysis into a single approach [66] However, traditional requirement elicitation techniques often fall short when it comes to effectively capturing the needs and expectations of large and distributed online communities

2.2 CrowdRE in Requirements Elicitation

The rapid evolution of technology and the widespread use of social media platforms have posed challenges to traditional elicitation techniques in identifying, gathering, and formulating requirements from the large online community [19]. In response to this gap, Groen *et al.* [19] proposed CrowdRE as a semi-automated requirement engineering approach that aims to obtain and analyze user feedback from a crowd, with the goal of deriving validated user requirements. This approach leverages the collective intelligence of the crowd to address the limitations of traditional RE methods, such as the limited scope and representation of user feedback [21]. The essence of CrowdRE lies in transforming user feedback into requirements through either manual content analysis or automated approaches [18, 19, 26]. Organizations can effectively identify and prioritize user needs, leading to improved user engagement with their products by utilizing the power of the crowd. [62]. To motivate stakeholders to actively participate in the crowd, CrowdRE offers various services and tools, including gamification techniques, forums, and visuals [20].

Numerous studies have explored the potential of leveraging crowd engagement on different platforms, such as app reviews and forums, to gain valuable insights

through the analysis of user-generated conversations, comments, feedback, and suggestions [11, 17, 38, 45, 58]. Additionally, researchers have examined social media platforms as valuable sources for analyzing various aspects of requirement engineering [27,30,63]. For example, Li *et al.* [34] discovered privacy-related user feedback by studying product-related subreddits on Reddit. Kengphanphanit *et al.* [30] classified user feedback into requirements and non-requirements by scraping Twitter and Facebook, and then utilized feature extraction techniques based on factors such as polarity, subjectivity, and the number of requirement-related words. They further developed a model using these factors and the Naive Bayes method to generate requirements from user feedback.

However, despite the growing interest in leveraging user discussions on social media platforms, there is still limited research focusing on video-based social media platforms like YouTube and TikTok to identify relevant user discussions that can be refined as requirements. These platforms offer unique opportunities for capturing user feedback in the form of video content, which presents new challenges and possibilities in extracting requirements.

2.3 Social Media and User-Generated Content

Social media platforms have revolutionized the way people communicate, share information, and express their opinions. The abundance of user-generated content on these platforms has become a valuable resource for understanding user perspectives, preferences, and feedback. Extensive research has been conducted to explore the impact of social media on user decision-making processes and its potential integration into software development practices. Studies by Seyff [53], Luca [36], Cao [10], Narangajavana [42], and Meneghello [41] have highlighted the influence and impacts of social media. Additionally Guzzi *et al.* [23], Begel *et al.* [8], and Treude *et al.* [59] have put forward diverse approaches to incorporate social media into software development and integrated development environments. Bajic and Lyons [7] found that, particularly in the early stages, small companies utilize social media to gather feedback from customers. Singer *et al.* [55] investigated how developers stay updated using Twitter, while Tian *et al.* [57] analyzed a random sample of tweets containing hashtags related to development topics, they determined content categories, assessed tweet popularity, and examined which categories received the most retweets. Prasetyo *et al.* [47] developed an automatic classification system to categorize tweets according

to their relevance to software engineering.

These studies and industry practices demonstrate the increasing recognition of the potential benefits of incorporating social media into software development processes. Social media data can provide valuable insights into user needs, preferences, and trends, which can influence decision-making during the requirements engineering phase. Researchers and companies have recognized the potential of social media data for requirements engineering, with studies focusing on integrating social media into software development practices and extracting valuable insights from user-generated content.

2.4 Video Platforms

YouTube, boasting over 2.5 billion active users, and TikTok, with a user base exceeding 1 billion, have established themselves as the leading video-based social media platforms [1]. These platforms, have become significant outlets for user engagement, providing opportunities to gather user discussions that are relevant to requirements [14, 39, 61].

Madden *et al.* [39] conducted a comprehensive analysis on a large dataset of YouTube comments consisting of 66,637 comments. Through their analysis, they were able to identify 10 major categories of user discussions, shedding light on the opinions and attitudes of viewers towards the video content. This research demonstrated the potential of classifying YouTube comments as a means of understanding the sentiments and perspectives of users. Similarly, Das *et al.* delved into the analysis of comments on YouTube videos, specifically focusing on the domain of autonomous vehicles [14]. By utilizing natural language processing techniques, they were able to categorize and extract insights from the comments related to autonomous vehicles. Their findings revealed that YouTube can serve as a valuable resource for understanding consumer opinions and concerns in this specific domain.

Expanding upon the potential of YouTube comments as a source of feedback, Karras *et al.* [29] employed machine learning algorithms to analyze a substantial dataset of 4,505 comments from a YouTube video. Their objective was to classify the comments into product-relevant categories. Through their analysis, they identified various types of discussions within the relevant comments, including feature requests, problem reports, efficiency considerations, and safety concerns. This research demonstrated how YouTube comments can offer valuable insights into user feedback and

provide specific information related to product features and improvements.

In a study by Schneider *et al.* [51], the authors emphasize the effectiveness of different types of videos, such as linear videos, vision videos, and interactive videos, in providing concrete situations related to a product and engaging users to elicit feedback. Vision videos, in particular, offer a glimpse into the future by portraying a vision of a potential product or system. These videos play a crucial role in helping stakeholders better comprehend and communicate their needs [28]. As a result, researchers have delved into the use of vision videos to elicit user feedback, as users tend to actively participate in discussions and provide valuable insights in response to these videos [9, 52]. However, despite the existing literature discussing the benefits of vision videos in soliciting feedback, their full potential for CrowdRE (Crowdsourced Requirements Engineering) remains relatively unexplored, as highlighted by Karras *et al.* [29]. Their study brings attention to the fact that the extent to which videos created by content creators themselves can offer valuable insights for companies is still unclear.

Against this backdrop, our study aims to uncover the potential of video content from popular platforms like YouTube and TikTok in identifying user feedback. We focus on extracting pertinent themes from the user feedback, recognizing the significance of filtering out irrelevant data that may arise from the vastness of the crowd's input. By identifying relevant themes, companies can then employ either manual approaches, such as content analysis [18], or automated techniques, as demonstrated by Kengphanphanit *et al.* [30], to generate requirements based on these themes.

Chapter 3

Methodology

We conducted an exploratory study to investigate the feasibility of using video-based social media platforms (i.e., TikTok and YouTube) for identifying requirements relevant user feedback themes. Our approach involves extracting and analyzing textual data from the audio, visual content, and metadata of these videos. By doing so, we aim to identify key themes and patterns in the user feedback that are relevant to requirements. The Figure 3.1 summarizes the methodology workflow.

3.1 Product Selection

In order to strive for the comprehensiveness of data collection, we endeavored to cover a wide range of industries. As this was the first of its kind, our aim was to obtain an understanding of requirement relevant user feedback across various sectors. To achieve this, we conducted market research and analysis to identify the top-performing products in each industry. Our research involved examining industry reports, consumer reviews, and market trends to pinpoint the most popular and widely used products by each manufacturer of each industry. we selected four prominent industries: Software, Computers, Mobile devices, and Automotive based on the most commonly used products in North America. Within each industry, we identified the five most popular products based on market research, customer demand, and industry rankings. For example, we chose the most widely used software across different domains, including browsers such as Chrome and Firefox, tutoring applications like Duolingo, networking platforms like Discord, and productivity software like Notion. For the automotive industry, we chose the vehicles that had sold the most units in North America, while for

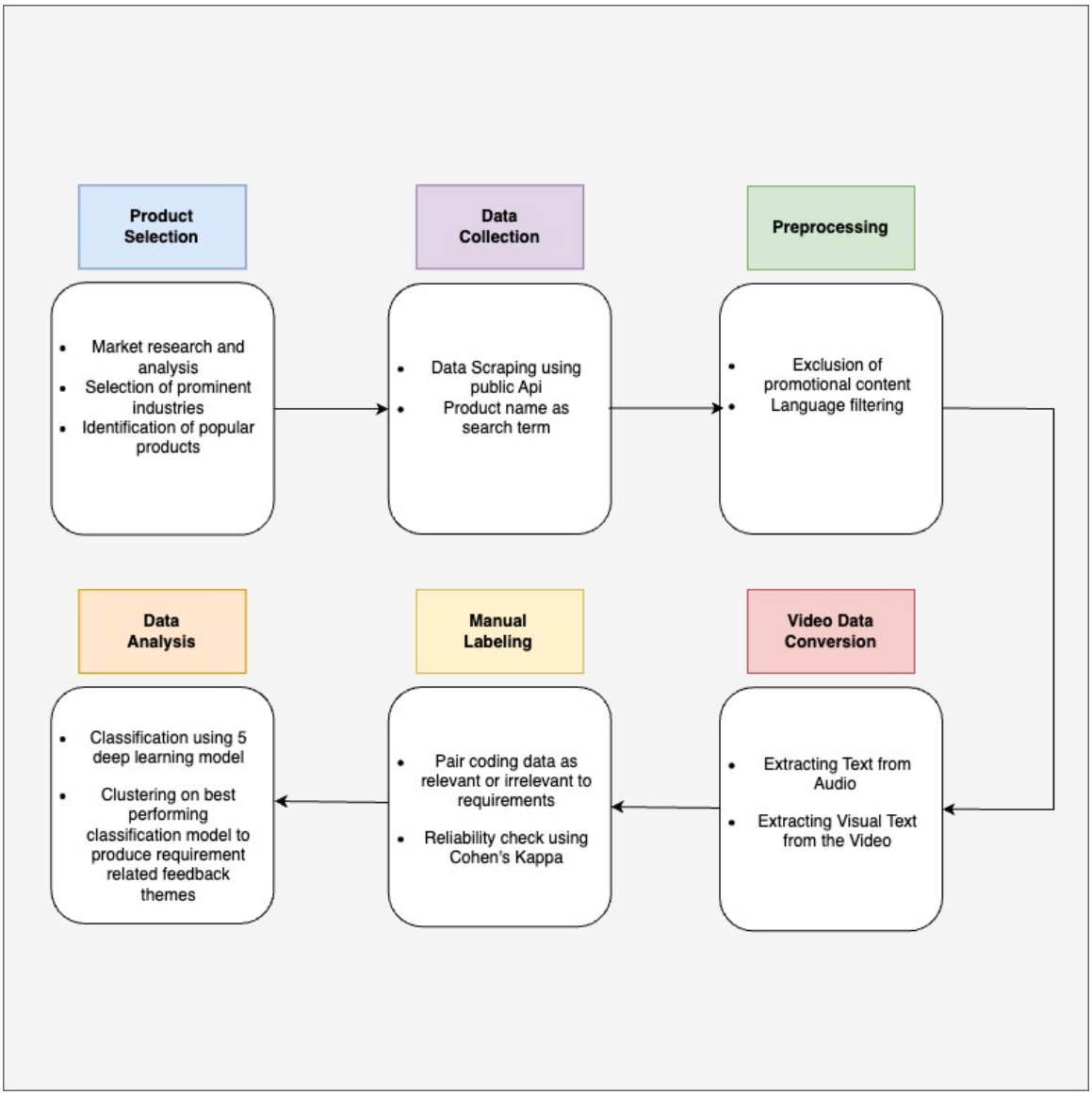

Figure 3.1: Methodology workflow

mobile phones and computers, we selected the latest releases from the most popular manufacturers. Table 3.1 shows the dataset characteristics. This approach allowed us to gather insights and identify common themes and patterns that transcend specific industries, providing a broader understanding of requirements-related user feedback. The inclusion of a diverse dataset not only enhances the validity and generalizability of our findings, but also enables us to make more informed recommendations for requirements elicitation and product development processes across different industries.

Category	Products/Search Term	TikTok Videos	YouTube Videos
Software	Notion	280	232
	Duolingo	224	217
	Discord	94	103
	Chrome	82	105
	Firefox	50	189
Phone	Google Pixel 7	223	183
	Apple Iphone 14	178	142
	Samsung Galaxy S22	162	214
	Motorolla Edge 30	76	92
	Oneplus 10	59	119
Computer	Microsoft Surface Pro 9	201	187
	Apple Macbook Air M2	193	161
	Asus Zenbook 14	119	132
	HP spectre x360 14	130	95
	Dell XPS 15	30	49
Automotive	Tesla Model 3	210	193
	BMW X5	190	197
	Ford F150	187	102
	Toyota Rav4	177	305
	Mercedes Benz GLC	154	239

Table 3.1: Products used for the analysis

3.2 Data Collection

In order to collect the necessary data for our study, we employed data scraping techniques from both TikTok and YouTube, utilizing their publicly available APIs. The data scraping process involved searching and downloading videos related to each product of interest. It typically took approximately a day to complete the scraping process for each product. To ensure accurate data collection, we utilized the exact name of each product as the search term on both platforms. By doing so, we were able to retrieve all available videos associated with the respective product. The search terms used for scraping each product are presented in Table 3.1.

For scraping YouTube, we utilized the PyTube [3] library in Python, making additional modifications to sort the search results based on views. This allowed us to prioritize the most viewed videos, ensuring that we capture the most popular and influential content related to each product. Regarding TikTok, we developed custom

Sources	Number Of Videos
TikTok	6080
YouTube	5261
Total	11341

Table 3.2: Total videos collected

code that emulated a web browser using the python library Selenium [4] and TikTok's publically available APIs, allowing us to extract data from videos and download them based on the specified search terms. This approach enabled us to access and retrieve videos from TikTok's platform in an efficient and systematic manner.

Our data collection process aimed to achieve maximum coverage by gathering all videos that were visible to an average user on both TikTok and YouTube. In total, we successfully collected 11,341 videos, with 6,080 videos obtained from TikTok and 5,261 videos from YouTube. Table 3.1 represents the number of videos downloaded for each product from both the sources. This dataset provided us with a substantial amount of user-generated video content to analyze and extract requirements-related feedback. By employing data scraping techniques from both TikTok and YouTube searching for videos associated with each product, we ensured that our dataset encompassed a wide range of videos, thereby increasing the representativeness and diversity of the collected data.

3.3 Preprocessing

To ensure the focus of our dataset on genuine user-generated content and exclude official promotional material, we implemented a robust preprocessing methodology involving a two-level data filtration process. The purpose of this preprocessing was to enhance the quality and relevance of the collected videos for our analysis.

In the first level of filtration, we excluded videos uploaded by official product handles or accounts associated with the respective products. Such videos tend to present information in a promotional manner and may not provide the authentic user feedback. We aimed to focus solely on user-generated videos and eliminate any bias introduced by promotional content and focus solely on user-generated videos. In the second level of filtration, we focused on filtering the videos to include only those in the English language. To achieve this, we employed Spacy FastLang [5], a language

detection tool, to determine the language of the video description text. Additionally, we utilized OpenAI Whisper [48], a speech recognition model, to detect the language from the audio text extracted from the videos, which will be described in much detail in Section 3.4.1. By leveraging these advanced language detection techniques, we were able to identify and retain only the videos that were predominantly in English. After the two-level data filtration process, we were left with a refined dataset consisting of 6,276 videos. These videos represented authentic user-generated content in the English language, ensuring the relevance and reliability of the collected data for our analysis.

The preprocessing helped us to enhance the quality and credibility of our dataset, ensuring that it consisted of genuine user-generated content and was focused on the English language. This approach allowed us to proceed with confidence in our subsequent analysis and requirement generation steps, knowing that the data we were working with accurately represented the user feedback we sought to extract and analyze.

3.4 Video Data Conversion

Videos are a rich source of information, encompassing audio tracks, metadata such as descriptions, and text that appears within the video itself, like captions or subtitles. In our study, we recognized the value of both the audio and visual elements of the videos, as well as the information provided by the content creators in the form of descriptions. The overview of the video data conversion processes is depicted in Figure 3.2.

To extract meaningful insights from the videos, we employed a two-fold approach. Firstly, we converted the audio content of the videos into text using advanced speech recognition technology. This allowed us to transcribe the spoken words and capture any verbal feedback expressed by the users. Secondly, we employed computer vision techniques to analyze the visual aspects of the videos. By sampling visual frames, we were able to extract any displayed text within the video, such as subtitles. This visual text provided additional context and contributed to a more comprehensive understanding of the user feedback. To augment the textual content from the audio and visual sources, we also incorporated the metadata associated with each video, including the video description and title. This information provided valuable context and further enriched our dataset for analysis. To classify the videos based on their

relevance to requirements, we employed state-of-the-art deep learning models. These models allowed us to categorize the videos into two groups: those containing user feedback that could be refined into requirements (referred to as "relevant"), and those with user feedback that was not conducive to further requirement refinement (referred to as "irrelevant") discussed in detail in Section 3.5.

In the subsequent subsections, a detailed description of the extraction techniques is provided. These methods form the foundation of our analysis and enable us to extract relevant user feedback that can inform requirement refinement and contribute to customer-centric product development.

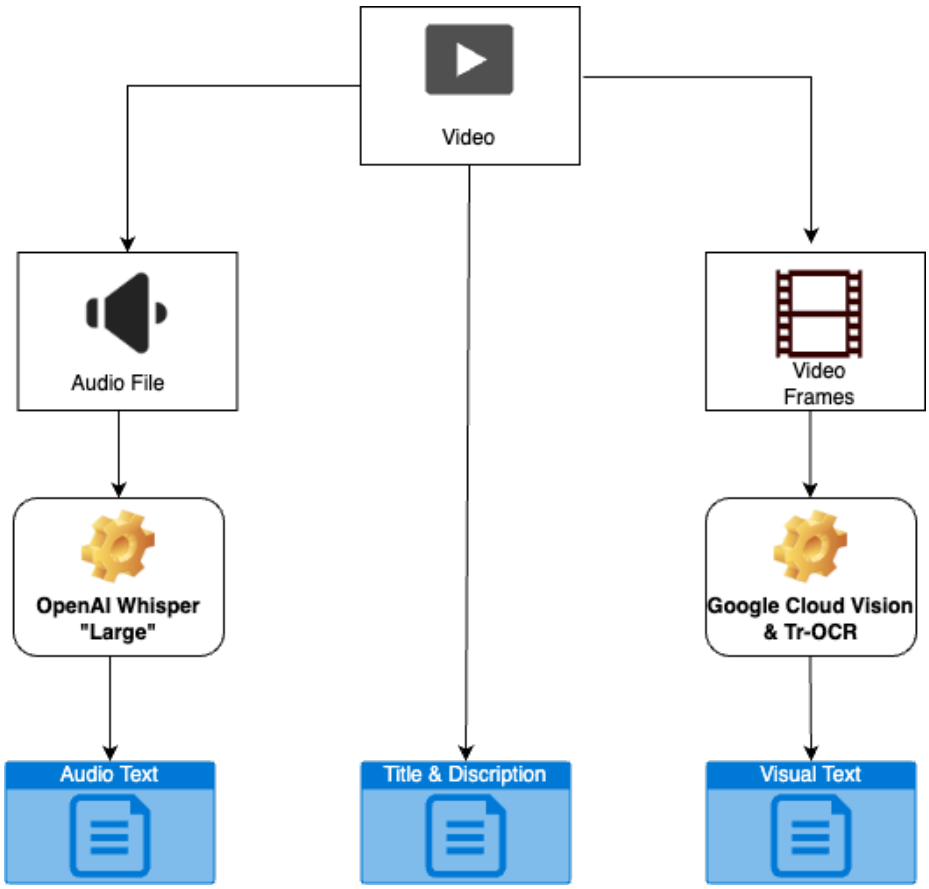

Figure 3.2: Video Data Conversion

3.4.1 Extracting Text from Audio

To extract text from the audio component of the videos, we employed OpenAI's speech recognition model called "Whisper". Whisper is a Transformer-based encoder-

decoder model that has been extensively trained on a vast corpus of web data, enabling it to provide highly accurate transcriptions in multiple languages [48]. Whisper is designed with a straightforward approach called the encoder-decoder Transformer architecture. To process the input audio, it is divided into smaller 30-second chunks. These chunks are then transformed into a log-Mel spectrogram, which captures the acoustic features of the audio. The encoder component of Whisper takes this spectrogram as input and extracts relevant information. On the other hand, the decoder is trained to predict the corresponding text captions for the input audio chunks [48]. In our study, we utilized the "Large" Whisper model, known for its exceptional transcription capabilities and suitability for high-quality transcription tasks.

We processed the audio content from the videos through the Whisper model to generate transcriptions of the spoken words. To optimize processing efficiency for longer videos, we transcribed only the first 30 minutes of audio. We made the assumption that the core topic or premise of the video would be conveyed within this timeframe. This approach helped us minimize processing time while still capturing the essential content for analysis. One of the valuable features of the Whisper model is its ability to detect the language being spoken during the transcription process. Leveraging this capability, we utilized Whisper to filter out videos that were not in the English language. This step was crucial to ensure the accuracy and relevance of the extracted text for our subsequent analysis. By focusing solely on English-language videos, we maintained consistency and effectively analyze the textual content in a language familiar to our research team. This approach enabled us to capture an understanding of requirement relevant information conveyed through the audio component.

3.4.2 Extracting Visual Text from the Video

In addition to audio, we recognized the importance of capturing visual text that may appear in the videos, such as subtitles or other relevant information displayed by content creators. To achieve this, we implemented a multistep process for extracting visual text from the videos. Since videos often contain repetitive or similar frames, we initially reduced the number of frames to be processed using motion-based video summarization techniques. This allowed us to select a representative subset of frames for further analysis. From this subset, we examined each frame to determine if it contained any visible text. To accomplish this, we utilized an Optical Character Recognition (OCR) system, which automatically recognizes and extracts text from images.

To select the candidate frames for text extraction, we modified the algorithm proposed by Dash and Albu [15]. Their approach integrates motion and saliency analysis with temporal slicing to identify frames of interest in a video. Instead of using their frame summarization technique, we utilized the saliency energy map to determine the probability divergence for the temporal slices. By calculating the Kullback-Leibler divergence $D_{KL}(.)$, of each temporal slice, $k \in \{\text{vertical}, \text{horizonal}, \text{diagonal}\}$ at time $t-1$ and t, where t is defined as the current frame. we obtained a vector $s_t \in \mathbb{R}^3$ (Eqn. 3.1) that represented the likelihood of containing text.

$$s_t^{(k)} = D_{KL}(p(k)_t || p(k)_{t-1}) \tag{3.1}$$

where $p(k)$ is the temporal slice k , normalized as a probability vector. $s_t^{(k)} \in \mathbb{R}^3$ is then thresholded by values greater than $T_h \in \mathbb{R}^3$ to select the candidate frame. For this study, $T_h = [1e-4, 1e-4, 1e-4]$.

$$\text{candidate frame} = \begin{cases} f_t & \forall k \in \{(s_t^{(k)} - s_{t-1}^{(k)}) > T_h\}, \\ \emptyset & \text{otherwise} \end{cases} \tag{3.2}$$

When the movement distribution changes significantly, a new candidate frame is selected. These candidate frames are then examined for text using CentripetalText [54]. If we don't find enough text of a suitable size, we discard the candidate frame. Next, we consider two scenarios: (1) if there is an audio track, and (2) if there is no audio track.

In situations where a video primarily relies on visual content without audio, we employed Google's commercial state-of-the-art OCR system called "Google Cloud Vision" [6] to capture all the text present in the candidate frame. However, when the video includes audio, we assume that the main content is conveyed through audio. To enhance the audio information, we utilized an open-source OCR system called "Tr-OCR" [33] by HuggingFace, specifically using the "trocr-large-printed" pretrained weights to extract larger OCR text identified by CentripetalText. It's important to note that both OCR methods are not entirely accurate, so to improve the accuracy of the extracted text, we also implemented a spelling error correction mechanism. OCR systems can sometimes generate incorrect spellings, so we applied Peter Norvig's algorithm [2] to identify and rectify common spelling mistakes in the extracted text. We chose to use multiple OCR algorithms due to budget limitations.

3.5 Manual Labelling

To ensure the accuracy and effectiveness of our classification models, we conducted a manual labeling process to create a ground truth dataset for training and evaluation. This process involved randomly selecting a subset of videos (Extracted Text) from our entire data pool and labeling them based on their relevance to our research objectives. In total, we manually labeled 1079 videos for this purpose.

The labeling process consisted of categorizing each video as either "relevant" or "irrelevant". Our criteria for determining the relevance of a video encompass various aspects that provide valuable insights to a company. These aspects include problem reports, reviews of product features, feature comparisons with competitors, and feature requests, among others. In essence, whenever a video contains content that can contribute to informed decision-making regarding positive or negative changes to a product, we classify it as "relevant." To illustrate this, consider the following example: *"To find out what the safest browser to use in 2022 is based on empirical testing techniques So we re going to go through 200 of the latest malware links... Firefox only blocked 145... Chrome not quite as good as Edge it blocked 198 links out of 200..." (Firefox).* This particular video was labeled as relevant because it provides valuable information on the performance of different browsers in terms of blocking malware.

In contrast, videos that fail to provide a substantial or meaningful description of a product, or merely touch upon the subject in a superficial manner, were categorized as irrelevant. These videos lack the necessary depth and detail to contribute to informed decision-making processes within a company. For instance, *"The new M2 MacBook Air is finally for sale. I'm not gonna buy one" (Apple MacBook Air M2).* This video exemplifies the kind of content that was classified as irrelevant. It offers minimal information about the product, consisting of a simple statement expressing the creator's personal decision not to purchase it. Such superficial commentary does not provide any valuable insights or analysis that can be utilized by a company to understand customer preferences, address potential issues, or make informed decisions regarding their products. Table 3.3 presents an example from the coded data, highlighting the disparity between requirement relevant and non-relevant content. Additionally, for further illustration, Tables A.1 and A.2 showcase additional examples.

To ensure a comprehensive and unbiased approach in our video labeling process, we implemented measures to prevent any potential bias towards specific products. This involved labeling videos from each product that we analyzed, ensuring a di-

Irrelevant	Relevant
The new M2 MacBook Air is finally for sale. I'm not gonna buy one *Audio Text*	To find out what the safest browser to use in 2022 is based on empirical testing techniques So we re going to go through 200 of the latest malware links... Firefox only blocked 145... Chrome not quite as good as Edge it blocked 198 links out of 200... *Audio Text*

Table 3.3: Example illustrating Relevant vs Irrelevant

verse representation across the dataset. To maintain consistency and reliability in the labeling process, we enlisted the expertise of two team members who possess extensive experience in requirement analysis. They collaborated through a pair coding methodology, working together to label a subset of 200 videos. The pair coding process involved jointly reviewing the videos and independently assigning them the labels of "relevant" or "irrelevant". After completing the pair coding process, we evaluated the agreement between the coders using Cohen's Kappa statistic [40], a widely accepted measure of inter-rater reliability. Cohen's Kappa assesses the level of agreement beyond what could be expected by chance alone, providing an indication of the reliability of the labeling process. To illustrate the concept, let's consider an example. Suppose two coders are tasked with labeling a set of videos as either "relevant" or "irrelevant." Each coder independently reviews the videos and assigns them labels based on their individual judgments. After labeling the videos, their responses are compared to determine the level of agreement. Cohen's Kappa takes into account both the observed agreement between the coders and the agreement expected by chance. It measures the extent to which the coders' agreement exceeds what would be expected due to random chance.

The analysis yielded an average Cohen's Kappa score of 87%, indicating a high level of agreement between the coders. This high inter-rater reliability suggests that the criteria for distinguishing between "relevant" and "irrelevant" videos were well-defined and clearly understood by both coders. Following the successful completion of the pair coding phase, one member continued with the labeling process independently, classifying the remaining 897 videos. In total, out of the 1079 videos that underwent manual labeling, 601 were categorized as "relevant" and 478 as "irrelevant." These labeled videos formed the foundation of our subsequent analyses and provided valuable

insights into the distribution of relevant and irrelevant content within the dataset.

3.6 Data Analysis of User Feedback

Data Analysis focuses on two primary objectives: identifying the best classification model for distinguishing between requirement-relevant and non-relevant videos, and uncovering feedback themes through clustering analysis within the relevant data subset.

The classification task aims to accurately classify requirement relevant and non-relevant videos. We trained and evaluated 5 classification models, to identify the most effective approach to distinguish between videos that provide valuable insights for informed decision-making and those that lack substantive content. Once we have classified the videos into relevant and non-relevant categories, the next step involves clustering analysis. Clustering allows us to group similar videos together based on shared feedback themes, providing a deeper understanding of the prevalent topics and concerns within the relevant video subset. Figure 3.3 illustrates the step-by-step process of data analysis undertaken in this study.

3.6.1 Classification

We utilized the capabilities of state-of-the-art deep learning transformer-based models to tackle the important task of video classification, distinguishing between relevant and irrelevant content. To accomplish this, we selected five cutting-edge models known for their remarkable performance in various natural language processing domains: GPT-2 (Generative Pre-trained Transformer 2) [49], BERT (Bidirectional Encoder Representations from Transformers) [16], RoBERTa (Robustly Optimized BERT Approach)) [35], XLM-RoBERTa (Cross-lingual Language Model – Robustly Optimized BERT Approach) [13], and ALBERT (A Lite BERT) [32]. Each of these models represents a significant advancement in the field, incorporating innovative techniques to effectively analyze and comprehend textual data. Given the significance of video classification in our research, it was imperative to determine which of these models would be most effective for our specific task. Therefore, we conducted an evaluation and comparative analysis to identify the model that would yield optimal performance in accurately categorizing the videos as relevant or irrelevant.

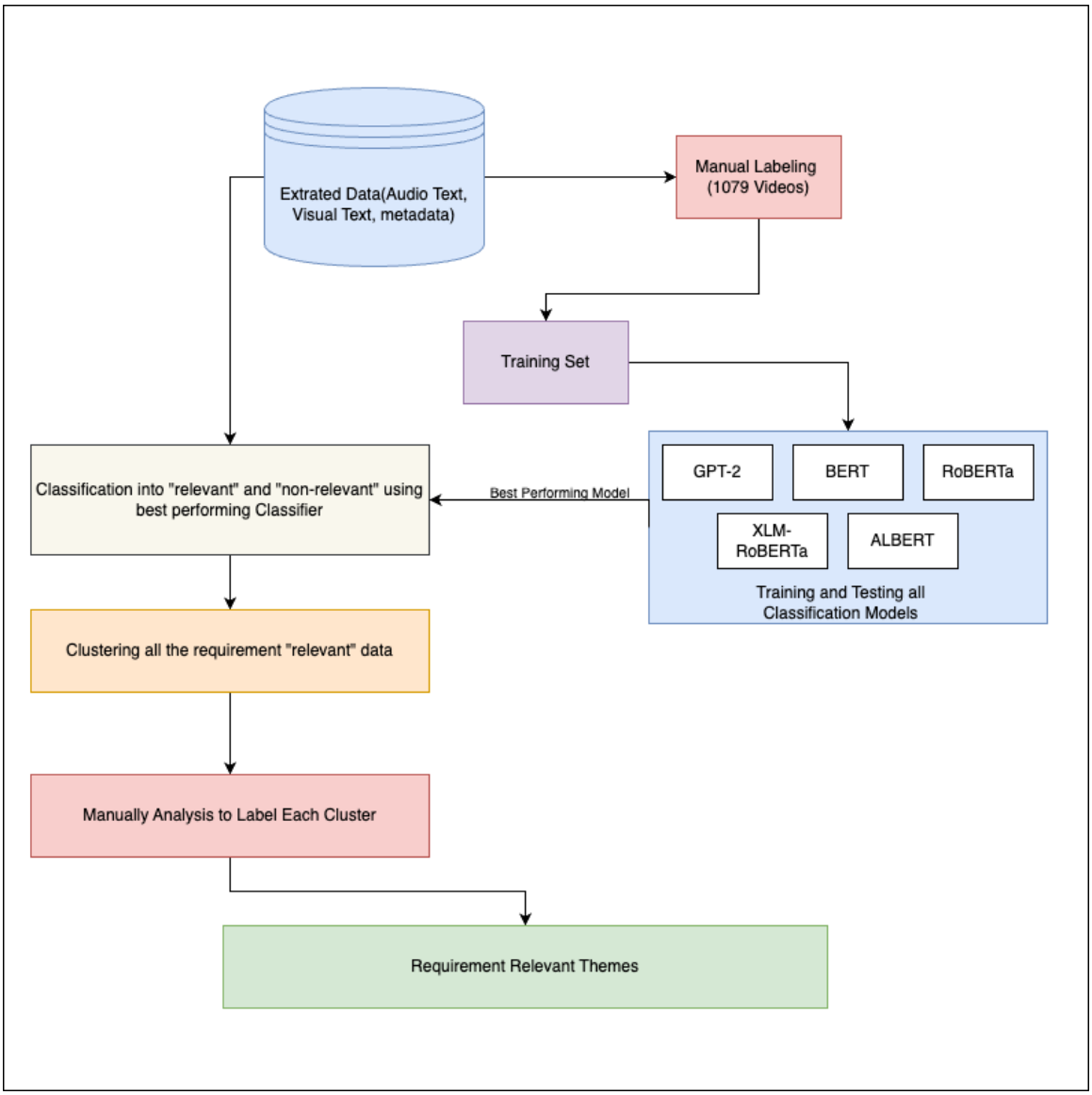

Figure 3.3: Data Analysis Process

GPT-2

GPT-2, also known as Generative Pre-trained Transformer 2 [49], is a language model developed by OpenAI. Unlike other models, GPT-2 specifically focuses on the decoder blocks of the transformer architecture. Its main function is generating sentences, treating text generation as a traditional language modeling task. The way GPT-2 works is by taking word vectors as input and estimating the probability of the next word using an autoregressive approach. It considers each token in a sentence individually, taking into account the context provided by the preceding words. This

approach allows GPT-2 to generate coherent and contextually relevant sentences. The architecture of GPT-2 primarily consists of stacked decoder blocks from the transformer framework. In a typical transformer architecture, the decoder receives a word embedding along with a context vector from the encoder. However, in GPT-2, the context vector is initialized with zeros for the first word embedding. Additionally, GPT-2 employs masked self-attention, which means the decoder can only access information from prior words in the sentence, including the current word, through obfuscation masking. GPT-2 is trained on an extensive web scrape corpus using the standard transformer training process. It operates with a batch size of 512, a predefined sentence length, and a vocabulary size of 50,000. One notable advantage of using GPT-2 for classification is its ability to capture long-range dependencies and understand the contextual relationships within the input text. By considering the entire sentence and the connections between words, GPT-2 can extract meaningful features that contribute to accurate classification. This is especially beneficial for tasks where the context and the surrounding words play a vital role in determining the class label. Despite its primary design for language generation, GPT-2 has shown remarkable effectiveness in performing classification tasks [25]. This efficacy can be attributed to its ability to learn intricate patterns and relationships within the data, thereby enabling accurate categorization of inputs into predefined classes.

BERT

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a language model that utilizes the transformer architecture [16]. Unlike GPT-2, BERT focuses on the encoder part of the transformer. This is because the transformer decoder, which considers only previous tokens, is better suited for tasks like word embedding learning. Including posterior context in determining a word can lead to target leakage, where the model unintentionally gains knowledge about the future words it is supposed to predict. Instead of predicting the next word directly, BERT learns by filling in masked-out words in sentences, which acts as a form of regularization. This prevents the model from relying solely on rote memorization and encourages it to understand the context of the sentence. Additionally, BERT also undergoes training on tasks that involve determining whether one sentence is likely to follow another. This helps improve its ability to handle two-sentence tasks and comprehend semantic relationships. By combining the transformer architecture, bidirectional training, and

masked word prediction, BERT achieves state-of-the-art performance on natural language processing tasks. Its ability to capture both prior and posterior context, coupled with effective training techniques, makes it a highly effective tool for understanding and generating language representations. For our specific task of classifying relevant and non-relevant videos, we employed BERT's pre-trained model. By fine-tuning the model on our labeled dataset, which included both relevant and non-relevant video examples, we were able to train BERT to distinguish between these two categories with high accuracy. BERT's contextual understanding of the videos' textual content enabled it to make informed predictions about their relevance.

RoBERTa

RoBERTa, short for Robustly Optimized BERT approach [35], is an advanced language model that builds upon the success of BERT. RoBERTa adopts a similar language masking strategy, where it learns to predict hidden sections of text within unannotated language examples. However, it introduces key modifications to BERT's architecture [16] and training process, resulting in improved performance on downstream tasks. Implemented in PyTorch, RoBERTa makes important adjustments to BERT's hyperparameters. Notably, it removes BERT's next-sentence pretraining objective and employs larger mini-batches and learning rates during training. These modifications enhance RoBERTa's ability to perform masked language modeling, surpassing BERT in terms of downstream task performance. One significant advantage of RoBERTa is its training on a significantly larger amount of data and for a longer duration. In addition to leveraging existing unannotated NLP datasets, RoBERTa incorporates a novel dataset called CC-News, which is extracted from public news articles. RoBERTa achieves state-of-the-art performance on various benchmark tasks such as MNLI, QNLI, RTE, STS-B, and RACE by capitalizing on this data. Notably, it demonstrates substantial improvements on the GLUE benchmark, achieving a score of 88.5 [35]. By leveraging RoBERTa's ability to understand the context and semantics of text, we trained it on our labeled dataset of videos

XLM-RoBERTa

XLM-RoBERTa [13] is a remarkable language model that combines the strengths of two influential architectures, namely Cross-lingual Language Model (XLM) and RoBERTa [35]. Designed specifically for multilingual contexts, XLM-RoBERTa en-

ables us to tackle the challenge of classifying videos that contain a mixture of languages, even though we have primarily focused on English videos. At the core of XLM-RoBERTa lies a transformer-based architecture, similar to that of RoBERTa. This architecture empowers the model to effectively capture and model complex language patterns and relationships. A key feature of XLM-RoBERTa is its utilization of masked language modeling, where it predicts intentionally hidden sections of text within unannotated language examples. By doing so, XLM-RoBERTa learns the underlying structures and representations of diverse languages, enabling it to comprehend and generate text across multiple languages [13].

To achieve its multilingual capabilities, XLM-RoBERTa employs cross-lingual training objectives and dynamically masks tokens from different languages during the training process. This approach allows the model to learn shared representations across languages and enhances its proficiency in understanding and generating text in a multitude of languages. By leveraging XLM-RoBERTa's unique capabilities, we aim to effectively classify videos that contain mixed languages, ensuring our approach is robust and inclusive of various linguistic contexts.

ALBERT

ALBERT (A Lite BERT) [32] is an innovative language model architecture specifically designed to address the challenges associated with scaling pre-trained models. It tackles the issue of model size and computational efficiency by significantly reducing the number of parameters compared to traditional BERT models [16]. This reduction in parameters is achieved through two crucial techniques, factorized embedding parameterization and cross-layer parameter sharing. Factorized embedding parameterization approach focuses on breaking down the large vocabulary embedding matrix into smaller matrices. By separating the hidden layer size from the vocabulary embedding size, ALBERT enables more straightforward scalability of the hidden size without a substantial increase in parameter size. This factorization enhances the efficiency of the model, allowing it to process and analyze language data more effectively.

Additionally, ALBERT employs the cross-layer parameter sharing technique to mitigate parameter growth with network depth. This technique involves sharing parameters across layers, which not only reduces the number of parameters but also acts as a form of regularization. By minimizing parameter redundancy, ALBERT

optimizes the model's performance while maintaining efficiency [32]. We trained the ALBERT model using a labeled dataset of video feedback, leveraging its capacity to comprehend and extract meaningful insights from the provided textual information.

3.6.2 Evaluation

To determine the most effective approach for accurately classifying textual data from popular video sharing platforms, we conducted a evaluation of five deep learning models: GPT-2, BERT, RoBERTa, XLM-RoBERTa, and ALBERT. Our evaluation encompassed different combinations of data, including visual text, audio text, and a combination of both, along with the title and description data available for all combinations. For each model, we adopted a similar training process. We utilized the pre-trained models as a starting point and fine-tuned them using our dataset, which consisted of labeled video text data. This dataset incorporated both the audio and visual text extracted from the videos, as well as the accompanying video metadata such as the title and description. Once the models were trained, we proceeded to assess their performance on a balanced test set of video text data.

To gauge the performance of the models, we employed two metrics: accuracy and area under curve (AUC). These metrics provided a evaluation of the models' effectiveness in correctly classifying the video text data. We repeated this evaluation process for each combination of data (visual text, audio text, and both audio and visual text) and for each video sharing platform (YouTube and TikTok). This allowed us to compare and analyze the performance of the models across different types of data and platforms.

3.6.3 Clustering

In order to gain insights into the user feedback themes present in our data, we employed a clustering technique. Specifically, we utilized BERTopic [22], a algorithm capable of inferring document distributions over topics and generating topic descriptions based on the BERT language model. To initiate the clustering process, we had to select a suitable cluster model. We opted for K-means as our cluster model. This model allows us to partition the data into groups based on their similarity. We then proceeded to perform the clustering process, varying the number of clusters from 2 to 6. To determine the most effective cluster configuration, we employed the Silhouette Coefficient [50], a widely-used metric for evaluating clustering quality. The Silhouette

Coefficient measures the similarity of an object to its assigned cluster compared to other clusters. By analyzing the Silhouette Coefficients for different cluster configurations, we were able to identify the configuration that yielded the highest coherence and separation between clusters. Once we obtained the clusters, we conducted a meticulous manual analysis. Our aim was to assign a relevant theme to each cluster based on the patterns and characteristics observed within the documents contained in that cluster. This manual analysis involved examining the content of the documents, identifying recurring topics, and labeling the clusters accordingly.

Chapter 4

Findings

Through rigorous data collection, preprocessing, and analysis, we delved into the realm of user-generated video feedback to gain a deeper understanding of its relevance and significance. Our objective was to extract meaningful information from the vast amount of user feedback available and decipher its implications for the domain under study. This chapter presents major findings in relation to our research questions.

RQ1 How can video-based social media be used to identify requirements relevant user feedback?

RQ2 What are the main users feedback themes that we can identify?

RQ3 How do the different social media platforms and their video content affect user feedback?

4.1 Text Extraction Approaches for Extracting User Feedback

This section will present the research findings related to the first part of our first research question (RQ1.1). It will focus on comparing classification models across various combinations of extracted data. The aim is to determine the most suitable approach by providing an answer based on the evaluation of different techniques and their effectiveness in classification tasks.

In our methodology, we employed two techniques to extract text information from the videos. Firstly, we converted the audio track of each video into text using speech

recognition model. This allowed us to transcribe the spoken words and capture any important textual content present in the audio. Additionally, we utilized computer vision algorithms to analyze the visual elements of the videos and extract any displayed text, such as captions or subtitles.

Dataset	Model	Accuracy	AUC
YouTube with only visual text	GPT-2	0.71	0.71
	BERT	0.76	0.76
	RoBERTa	0.74	0.74
	XLM-RoBERTa	0.67	0.67
	ALBERT	**0.79**	**0.79**
YouTube with only audio text	GPT-2	**0.94**	**0.94**
	BERT	0.86	0.86
	RoBERTa	0.86	0.86
	XLM-RoBERTa	0.83	0.83
	ALBERT	0.79	0.79
YouTube with both visual and audio text	GPT-2	**0.91**	**0.91**
	BERT	0.85	0.85
	RoBERTa	0.80	0.80
	XLM-RoBERTa	0.80	0.80
	ALBERT	0.79	0.79
TikTok with only visual text	GPT-2	**0.71**	**0.71**
	BERT	0.70	0.70
	RoBERTa	0.50	0.50
	XLM-RoBERTa	0.50	0.50
	ALBERT	0.70	0.70
TikTok with only audio text	GPT-2	0.92	0.92
	BERT 0.92	0.92	0.92
	RoBERTa	**0.93**	**0.93**
	XLM-RoBERTa	0.90	0.90
	ALBERT	0.90	0.90
TikTok with both visual and audio text	GPT-2	0.93	0.93
	BERT	0.95	0.95
	RoBERTa	**0.97**	**0.97**
	XLM-RoBERTa	0.90	0.90
	ALBERT	0.93	0.93

Table 4.1: Results of Deep Learning Models on Classifying between Relevant vs Irrelevant. AUC is area under curve.

In the next phase of this work, we turned our attention to the classification of the collected TikTok and YouTube videos. Our goal was to accurately categorize the user feedback contained within these videos as either “relevant” or “irrelevant” to the process of refining requirements. To achieve this, we employed five state-of-the-art deep learning models naming GPT-2, BERT, RoBERTa, XLM-RoBERTa,

and ALBERT that have demonstrated exceptional performance in various natural language processing tasks. These models were selected based on their ability to effectively capture the nuances and contextual information present in textual data. Using these advanced models, we trained them on our labeled dataset of videos, fine-tuning their parameters to optimize their performance in video classification. The videos were labeled as "relevant" if the user feedback provided valuable insights that could be further refined into requirements, and "irrelevant" if the user feedback was not useful for the requirements' refinement process. The classification results obtained from these techniques are summarized in Table 4.1. This table provides a concise overview of the performance of each model with each combination of the data in accurately classifying the videos. It showcases the effectiveness of the deep learning models in distinguishing between requirement relevant and irrelevant.

Notably, we observed a consistent trend wherein datasets incorporating audio text in paired with video metadata consistently outperformed those relying solely on visual text paired with video metadata. This finding strongly suggests that the inclusion of audio text provides valuable contextual information that enhances the classification process. When utilizing audio text paired with video metadata, we achieved impressive accuracies of 94% for YouTube videos and 93% for TikTok videos. These high accuracy levels indicate the robustness of the models in accurately identifying requirements-relevant user feedback when audio text is present. In contrast, datasets relying solely on visual text extracted from video frames proved to be less effective in distinguishing between relevant and irrelevant user feedback. Table 4.1 clearly illustrates this disparity, where the most accurate model for classifying YouTube's visual text dataset achieved an accuracy level comparable to the poorest performing model for the YouTube audio text dataset. Similarly, for TikTok videos utilizing only video text and metadata, two models exhibited notably low accuracies of 50%. In the context of a balanced dataset, this level of performance is equivalent to that of a dummy model, highlighting the inadequacy of relying solely on visual text for accurate classification.

Our findings strongly suggest that the superior performance of datasets incorporating audio text compared to those utilizing both audio and visual text can be attributed to the accuracy of audio extraction and the comprehensive nature of audio text in capturing the essence of the video's content. When a video includes a host or hosts speaking about the content, the audio text effectively encapsulates the main idea, enabling the classifiers to make accurate predictions. In contrast, relying solely

on visual text poses several challenges. The extraction of visual text relies on sampling visual frames, making assumptions that may not always hold true. 1) a video has clear subtitles that are easy to recognize, 2) a video displays visual graphics or text that pertain to the video's content. However, if a video lacks significant visual text or does not include subtitles, classifiers have limited information to base their decisions on, relying mostly on the accompanied metadata. This limitation hampers the accuracy of classifiers utilizing visual text. While audio extraction to text may face limitations such as background music masking the host's voice, the likelihood of encountering such challenges is lower compared to the no presence of visual subtitles. Extracting text from audio also has a reduced chance of encountering random audio that may confound the speech-to-text model. Consequently, datasets that rely on audio text outperformed datasets that incorporated both audio and visual text in our experiments. The only exception to this trend was observed in the "TikTok with both visual and audio text" dataset, where it achieved a 2% higher accuracy than the "TikTok with audio text" dataset. We hypothesize that the unique characteristics of TikTok videos, such as the increased use of subtitles that complement the audio content, may have contributed to the improved performance of the model in classifying videos with both visual and audio text. In the subsequent section of this chapter, we will delve deeper into the effect of video content characteristics and discuss their influence on the performance of the classifiers.

Among the various deep learning models we evaluated for identifying requirements-relevant feedback in TikTok and YouTube videos, GPT-2 and RoBERTa emerged as the most accurate classifiers. Notably, GPT-2 demonstrated exceptional performance in the YouTube dataset when utilizing audio text, surpassing the other models by a considerable margin of 8-15%. This highlights the effectiveness of GPT-2 in accurately identifying relevant feedback in YouTube videos when leveraging audio text. Furthermore, when examining the TikTok datasets, RoBERTa stood out as the top-performing classifier in two out of three scenarios. For the "TikTok with audio text" dataset, RoBERTa achieved an impressive accuracy of 93%, showcasing its ability to accurately classify TikTok videos that contain audio text relevant to requirements. In the "TikTok with both audio and visual text" dataset, RoBERTa demonstrated outstanding performance, with a remarkable accuracy of 97%. This achievement represents the highest accuracy obtained across all our tests in Table 4.1.

Once we identified the most accurate models for TikTok and YouTube datasets, our next step was to apply these models to classify the remaining unlabeled videos

Dataset	Classification Model	Manual Annotation Results
YouTube with audio text	GPT-2	98%
TikTok with both video and audio text	RoBERTa	100%

Table 4.2: Manual Annotation Results for 50 random videos

Dataset	Relevant	Irrelevant
YouTube Manual Labelling	370	167
YouTube with audio text classification via GPT-2	1691	1029
TikTok Manual Labelling	226	311
TikTok with both video and audio text classification via RoBERTa	810	1672
Total	3097	3179

Table 4.3: Result from Labelling and Classifying the Dataset

in the dataset. Additionally, we selected a random sample of 50 videos, along with their classified labels, and conducted a manual annotation process to assess the accuracy of the automated labeling. Table 4.2 illustrates the manual annotation results for random sample of 50 videos. In line with the original experiments presented in Table 4.1, the manual annotation results demonstrated consistency. For instance, the combination of "YouTube with audio text" and the GPT-2 model achieved an impressive accuracy of 98%. Similarly, the pairing of "TikTok with both video and audio text" with RoBERTa resulted in a perfect accuracy of 100%. We have summarized the distribution of relevant and irrelevant videos in Table 4.3. Notably, out of the 20 products analyzed in this work, we identified a total of 3,097 videos containing relevant information, while 3,179 videos were deemed non-relevant.

It is worth mentioning that YouTube videos exhibited a higher concentration (61%) of videos containing relevant requirements elicitation information compared to TikTok videos (34%). This discrepancy indicates that YouTube is more commonly utilized as a platform for sharing videos that contain valuable insights and feedback related to requirements. On the other hand, TikTok seems to have a lower proportion of videos containing such relevant information.

Theme	Description	Number of Products (out of 20)
Feature Ratings	Praise/criticism of product features	20
Matching Competition	Comparison with other competitor products	13
Performance Ratings	Praise/criticism of performance of the products	8
Modifications Suggestions	Suggestions for tools/upgrade	5
Bug Report	Bugs and issues of products	4
Repair & Maintenance	Videos related to fixing and preserving	3
Design Ratings	Design evaluation	3
Affordability	Cost prospects of the products	3
Usage Tutorials	Tutorials for other user to help use the product	3

Table 4.4: Requirement Relevant Themes

4.2 Thematic Analysis: Identifying User Feedback Themes

This section delves into the second part of our primary research question (RQ1.2), focusing on the identification of themes within the user feedback obtained from videos. Through the manual analysis of the clusters, we aim to uncover and categorize the prevalent themes that emerge from the feedback provided by users in the form of videos. which can play a crucial role in understanding the needs and preferences of users. By categorizing the videos based on their content, we were able to group them into distinct themes, providing a holistic understanding of the users' perspectives and experiences.

Our approach identifies 9 themes including several important themes such as **Feature Rating**, **Bug Reports**, **Usage Tutorial**, **Design Ratings** and **Performance Ratings**. These themes serve as valuable insights and can be leveraged for further requirement elicitation and refinement. These themes, which are presented in Table 4.3, lay the groundwork for uncovering specific areas of improvement and guiding the refinement of requirements for the product under evolution. This process enabled us to identify and extract the most prominent themes that emerged from the feedback. Through this thematic analysis, we gain a deeper understanding of the user's percep-

tion about the product, facilitating the development of more targeted and relevant product solutions.

4.2.1 Feature Ratings

Feature Ratings is one of the prominent themes that emerged from our analysis. It was seen in all the 20 products. This theme encompasses user feedback that provides an overall rating of the product and often highlights specific features that are perceived as strengths or weaknesses. It offers valuable insights into how users perceive and evaluate the various functionalities and capabilities of the product. For instance, in a comment about Firefox on YouTube, a user expressed their dissatisfaction by stating, *"I hate Firefox, but I'm still switching back to it. I can't stand Firefox. Every time I try to use it I just get frustrated by all of the useless features privacy invading telemetry and annoying defaults. If only there was a way to use Firefox without all the junk."* This comment not only reflects the user's negative sentiment but also highlights specific aspects of the browser that they find problematic.

Furthermore, within the broader theme of Feature Ratings, we observed a consistent pattern of user feedback related to product updates and new features. Users often offered suggestions for further improvements and voiced their opinions on the latest changes. For example, on TikTok, a user expressed their dissatisfaction with a recent update of the Duolingo language learning app, stating, *"Please boost this so that Duolingo can see this. I hate this new update so much... You can't jump between topics anymore which is so bad. Used to love how it was basically self-guided learning, this new learning plan is so restrictive and confusing. It was so simple before. Please bring back the old layout, I have Super Duolingo, and I am not happy paying for this team"*. This comment not only highlights the user's discontent with the new update, but also provides specific feedback on the changes made to the learning plan and layout. While Feature Rating was a theme found across all the products, it emerged as a particularly prominent when it came to phones and computers, suggesting that the features offered by these devices played a significant role in users' purchasing decisions.

4.2.2 Bug Report

Another significant theme that emerged from our analysis of user feedback is Bug Reports. Bug Reports encompassed videos where users highlighted issues, glitches,

or malfunctions they encountered while using a particular product. These reports shed light on the areas where a product falls short, providing valuable insights for companies to address and rectify these issues. Bug Reports were observed across software products. Users often shared their frustrations, experiences, and specific details about the bugs they encountered. These videos served as a platform for users to voice their concerns and seek resolution from both the company and other users who might have faced similar problems.

For instance, we came across a video *"Fix Discord... Discord App not launching on Windows 10... [To fix, Method 1]... Close discord in task manager and restart it. Right click on the taskbar and click on task manager. Right click on the Discord option and click on end task."* In this video, the user provided step-by-step instructions on resolving the issue of Discord not launching on Windows 10. The suggested solution involved closing Discord through the task manager and restarting it. This type of user-generated content, focused on bug reports, can be incredibly valuable for developers and software companies.

Videos that address bug reports serve as a rich source of feedback for software developers. They offer a firsthand account of the challenges and obstacles that users encounter while interacting with a product. By these videos and paying attention to the specific issues raised, developers can gain valuable insights into the areas of their software that may require improvement or troubleshooting.

4.2.3 Usage Tutorial

The Usage Tutorial theme encompasses videos where users provide guidance, tips, and demonstrations on how to effectively use a particular product. These tutorials serve as a valuable resource for both new and existing users who seek assistance in navigating the features and functionalities of a product. Video content around software products frequently showed hacks and workarounds to help users make the most of their software tools. In the digital age, where technology continues to evolve rapidly, users often encounter new software applications, gadgets, and devices that require some level of learning and understanding. Usage Tutorials play a crucial role in bridging the gap between user knowledge and product usability, empowering individuals to make the most of their products. The significance of usage tutorials lies in their ability to shed light on areas of usability that may require attention. These videos can reveal common areas of confusion or difficulty that users encounter while

utilizing a software product. By identifying these pain points, organizations can take proactive measures to address usability issues, refine user interfaces, and streamline workflows, ultimately improving the overall user experience.

As an example, there is a video titled *"How to block ads on websites in Chrome Android" (Google Chrome on YouTube)* that provides users with instructions on improving their browsing experience by eliminating ads. This type of video content holds great value for companies, as it offers insights into users' perspectives on ads and their preferences for an ad-free experience. By analyzing such videos, companies can gain a better understanding of how their users perceive ads and tailor their products accordingly to enhance consumer satisfaction. Videos like these provide practical solutions and recommendations for users to customize their browsing experience, specifically in relation to ad-blocking. By exploring these videos, companies can gather valuable information on the effectiveness and popularity of ad-blocking techniques among their user base. This insight can guide product optimization efforts and help companies align their strategies with consumer expectations. Prior research has explored the conversion of user feedback, into actionable requirements for software development [18, 26, 30]. By analyzing the content of usage tutorial videos, organizations can identify patterns, trends, and user needs that can inform the creation of more intuitive interfaces, comprehensive documentation, and targeted user training. This iterative feedback loop between users and developers ensures that software products evolve in line with user expectations and requirements.

4.2.4 Matching competition

It is one of the intriguing themes that emerged from our analysis of user feedback. In this theme, users compare and contrast different products within the same category, providing insights into how a particular product stands against its competitors. These videos offer valuable information for consumers who are in the process of making purchasing decisions, as well as for companies seeking to understand their market positioning and competitive landscape. This theme was particularly prevalent on YouTube and was seen in 13 products out of 20. This theme involves in-depth reviews of products that compare them directly to their competitors. To illustrate, *"The Pixel gets something iPhone users can only dream of and that is a 48 megapixel telephoto lens with five times optical zoom. Now on iPhone you only have a 12 megapixel telephoto that does three times optical zoom. This means so on the Pixel you can*

zoom in closer and get crisper higher resolution images than you can on the iPhone" (Google Pixel 7 on YouTube) The video explains that this difference allows the Pixel to capture closer and sharper images with higher resolution compared to the iPhone.

Such videos are extremely beneficial for both consumers and companies operating in competitive markets. Consumers can rely on these comparison videos to gain a comprehensive understanding of how different products measure up against each other. They can assess the strengths and weaknesses of each product, enabling them to make well-informed decisions based on their specific needs and preferences. From a company's perspective, these videos provide valuable feedback on their own products as well as their competitors'. Organizations can learn about the areas where their products excel or lag behind in comparison to their rivals. This information can be utilized to refine their marketing strategies, enhance product development, and effectively position themselves in the competitive landscape. In the specific example of the Pixel and iPhone comparison, it becomes evident that such videos can influence prospective customers and impact their purchase decisions. Companies, including Apple, can gather highly useful user feedback from these types of videos. This feedback serves as a valuable resource for identifying areas for improvement and staying competitive in the market.

4.2.5 Design Ratings

Design Rating was observed in videos related to three out of the twenty products. This theme revolves around creators highlighting and discussing the aesthetic appeal and visual design aspects of these devices. The Design Rating theme is particularly prevalent in the computer and mobile devices' category, where manufacturers strive to offer devices that are not only technologically advanced but also visually appealing. This theme highlights the importance of aesthetics and the impact it has on users' perceptions and preferences. Videos within this theme often showcase the physical appearance, build quality, and overall design of the computer devices. They often comment on factors such as sleekness, elegance, color options, build quality, and ergonomic considerations. The Design Rating theme reflects the significance of aesthetics in consumer decision-making. Many consumers value not only the functionality and performance of a product, but also its visual appeal and how it aligns with their personal style and preferences. One particular example that highlights the Design Rating theme, *"HP Spectre x360 now comes in a 16 inch version bringing*

big specs and a big screen... The Spectrex 360 16 comes in two color options Nightfall Black with pale brass accents and Nocturne Blue with celestial blue accents..." (HP Spectre x360 on YouTube) These videos play a significant role in catering to consumers who value not only the functional aspects but also the visual appeal of their devices. From a company's perspective, the Design Rating theme offers valuable feedback on the visual aspects of their products. By paying attention to the design elements that creators emphasize, companies can gain insights into consumer preferences and expectations regarding aesthetics. This information can guide them in product development and help refine their design strategies to cater to the target market effectively.

4.2.6 Performance Ratings

Performance Ratings often revolves around assessing the performance quality of the showcased products. This theme sheds light on users' experiences and opinions regarding the performance aspects of the devices. For example. *"There's a few things about the S22 Ultra that just drive me up the freaking war.. this is the most powerful phone Samsung has ever made except it still lags in like the most random places There's always a lag when I want to bring down the quick settings and sometimes been scrolling it just lags for seemingly no reason..." (Samsung Galaxy S22 on TikTok)* Such Performance Ratings comments are valuable for both consumers and manufacturers. For prospective buyers, these insights provide firsthand accounts of the product's performance in real-world usage scenarios, helping them make informed decisions based on their performance expectations. Users can gauge whether the device meets their performance needs and if any potential performance limitations align with their usage patterns. It is worth noting that there may be some overlap between the Design Rating and Performance Rating themes, as both aspects contribute to the overall user experience. Users often evaluate a product's performance alongside its visual appeal and design qualities. However, the Performance Ratings theme specifically delves into users' feedback on the device's functionality, responsiveness, speed, and overall performance during usage.

4.2.7 Affordability

This theme emerged in videos featuring automotive products and pertained to users' perceptions and discussions regarding the cost and value of these vehicles. Notably,

the Affordability theme was observed in three out of the twenty products we analyzed, all of which belonged to the automotive category. Users often express their opinions and concerns about the price point, value for money, financing options and overall affordability of the showcased vehicles. These discussions offer valuable insights into the perceptions of prospective buyers and shed light on the affordability aspect of the products. We identified a common theme of Affordability among expensive automotive products, including Tesla, BMW, and Mercedes. In these videos, creators are discussing ways to finance the purchase of these high-end cars, or providing cost-specific analyses to show how they may be more affordable than one might initially assume. For Example, *"'How I Bought a Tesla for 78 per Month' buying a brand new Tesla Model 3 for 78.39 per month Because I financed the cost of the car this car payment is 640 per month which works out to be 7680 per year My total cost to buy this car in the first year including monthly payments taxes license and registration comes to 12,117 However even though I ve paid that amount out of pocket I will get back 2,500 from California's EV rebate program I'll get back another 3,780 from the federal tax credit" (Tesla Model 3 on TikTok).* These types of videos are particularly useful for these companies, so they can increase their sales, focusing on buyers who may be hesitant to purchase a high-end car due to perceived high costs. The Affordability theme serves as a useful reference for both consumers and manufacturers. For potential buyers, these discussions provide real-world perspectives on the cost-effectiveness and value proposition of the showcased vehicles. Users can gauge whether the price aligns with their budget, expectations, and desired features. Additionally, insights from other users' affordability concerns can help individuals make informed decisions about the financial feasibility of their prospective purchase.

4.2.8 Modification Suggestions

The Modification Suggestions theme emerged as a prominent aspect within the user feedback videos we analyzed, particularly in the automotive category. This theme revolves around content creators and users suggesting modifications and aftermarket upgrades that can be made to the products in question, with the aim of enhancing the overall user experience. The significance of the Modification Suggestions theme lies in its potential as a valuable source of feature ideas for future product development. When users express their willingness to modify their products, it indicates that they have a clear vision of their desired features and functionalities. This feedback offers

developers an opportunity to bridge the gap between the current product offerings and the desires of their customers. By identifying the specific modifications that users propose, developers can gain insights into the areas where their products may fall short or could be further improved. For example, *"My top 20 Toyota Rav 4 aftermarket upgrades and modifications Mods...I installed [improved] driv- ing lights and I really like the driving lights. The ones that come with it are great, but of course they only work when you have it on low beam and these are of course for the fog lights" (Toyota Rav 4 on YouTube).* By closely examining these Modification Suggestions videos, developers can gain a deeper understanding of the usage context of their products. They can uncover how customers are utilizing their products and identify opportunities for customization and personalization.

4.2.9 Repair and maintenance

The Repair and Maintenance theme emerged prominently in the videos analyzed, particularly within the automotive product category. Content creators often dedicated discussions to various aspects of car repair and maintenance, sharing tips and insights on parts that commonly wear out over time and providing guidance on how to prevent and address these issues. The significance of the Repair and Maintenance theme lies in its potential as a valuable resource for automotive companies seeking to enhance the quality of their products. By analyzing the information shared in these videos, companies can gain insights into the patterns of wear and tear identified by content creators and users. This analysis can provide valuable feedback on the performance and durability of their products over time. The Repair and Maintenance theme, observed in three out of the twenty products analyzed, provides a wealth of knowledge for automotive companies. It allows them to understand the collective experiences and expertise of users and content creators in dealing with common repair and maintenance issues.

4.2.10 Evaluation of Clusters

To gauge the reliability of our approach, we adopted a manual analysis of clusters. We initiated this evaluation by randomly selecting requirement relevant classified data from four representative products from each domain, from both the TikTok and YouTube. To set the ground truth, 25 video data points were randomly selected for each product corroborated against the outcomes of the K-Means clustering. The

Product	TikTok(Accuracy)	Youtube(Accuracy)
Duolingo	0.80	0.76
Apple Iphone 14	0.84	0.72
Asus Zenbook 14	0.76	0.71
Toyota Rav4	0.84	0.76

Table 4.5: Results of Manual Analysis Comparison of Clustering Results and Ground Truth

rationale behind this manual assessment was to ascertain the degree to which our clustering aligned with the actual characteristics of the data. Through this process, we aimed to measure the precision and reliability of our clustering methodology in accurately categorizing and grouping the user-generated content. Table 4.5 showcasing the correspondence between the ground truth and our clustering results. The computed accuracy percentages is indicative of the extent to which the K-means clustering aligned with the human-assigned categories. We observed a slightly better clustering performance for TikTok videos compared to YouTube videos. This difference could be attributed to the possibility of having multiple distinct topics within a single YouTube video, making it more challenging for the clustering algorithm. The confusion matrix containing the outcomes of the manual evaluation is available in Appendix B for further reference

4.3 Impact of Social Media Platforms and Video Content on User Feedback

This section answers the second research question (RQ 2), exploring how user feedback is influenced by the distinctive characteristics and features of TikTok and YouTube platforms. It investigates the effects and implications of these platforms on user feedback dynamics.

We observed in section 4.1 how the number of videos containing relevant user feedback that can be later refined into requirements differed significantly between YouTube and TikTok. Table 4.6 provides insights into the differences in audio and visual texts found in videos from these two platforms. Notably, we discovered that YouTube videos, on average, were significantly longer than TikTok videos, with YouTube videos being at least 15 times longer. Given the substantial difference in

video duration, it is not surprising that YouTube videos have more time to cover various aspects, such as feature ratings, bug reports, and discussions on missing features in competitor products. The extended duration of YouTube videos allows content creators to delve into more detail and provide comprehensive feedback. On the other hand, TikTok videos have a much shorter average duration of approximately 33 seconds. This limited time frame presents a challenge for creators to address complex topics like feature ratings and bug reports effectively. Due to the brevity of TikTok videos, it becomes more challenging to provide in-depth analysis or discuss intricate details. The contrasting durations of videos on YouTube and TikTok highlight the differing nature of content creation and consumption on these platforms. YouTube's longer video format offers more opportunities for users to express their opinions and provide comprehensive feedback. In contrast, TikTok's shorter video format necessitates concise and visually engaging content that can be consumed quickly.

In addition, we made an interesting observation regarding the word coverage and unique words per second in TikTok and YouTube videos. TikTok videos exhibited a higher average word coverage and unique words per second compared to YouTube. Specifically, TikTok videos contained more than twice as many unique words per second on average than YouTube videos. We attribute this phenomenon to the nature of content creation on TikTok, where creators must fit a substantial amount of content within a limited time frame. To make their videos more engaging and captivating, content creators often accelerate the audio, resulting in faster speech delivery. This accelerated audio technique aims to capture the viewer's attention and ensure that the content is delivered effectively within the short duration of TikTok videos.

Interestingly, we observed a stark contrast in the amount of visual text between TikTok and YouTube videos. On average, TikTok videos contained five times more visual text per second compared to YouTube. Moreover, the number of unique visual texts per second was significantly higher on TikTok as well. This disparity can be attributed to the prevalent use of video subtitles on TikTok, where visual text often complements the audio content. The combination of visual and audio text provides a complementary factor that contributes to the accuracy of the dataset with both visual and audio text on TikTok, which achieved the highest accuracy among all TikTok datasets. In contrast, YouTube videos tend to have fewer visual subtitles, and the presence of text may actually introduce confusion for the classifying models. These differences in the usage of visual text may have influenced the performance of our classification models.

In our study, we utilized GPT-2 medium and RoBERTa base models. Notably, GPT-2 medium is a larger model with a higher number of parameters compared to RoBERTa base (GPT-2: 345M parameters vs. RoBERTa: 125M parameters). The condensed nature of TikTok videos, coupled with the abundance of visual text, results in less noisy data, requiring fewer model parameters for accurate classification. Smaller models like RoBERTa tend to be more suitable for smaller datasets, as they are less prone to overfitting due to their reduced parameter size. Our findings align with this expectation, as RoBERTa emerged as the most accurate model for two out of three TikTok datasets, while GPT-2 achieved the highest accuracy for three out of six datasets.

Platform	Product	Avg. Sec.	Views Per Video	Audio Words Per Video	Uniq. Audio Words Per Video	Audio Words Per Sec.	Uniq. Audio Words Per Sec.	Visual Words Per Video	Uniq. Visual Words Per Video	Visual Words Per Sec.	Uniq. Visual Words Per Sec.
YouTube	Software	560	0.20M	808	253	1.4	0.5	546	231	1.0	0.4
	Phone	555	1.48M	934	334	1.7	0.6	232	111	0.4	0.2
	Laptop	480	0.20M	1,356	465	2.8	1.0	319	165	0.7	0.3
	Car	450	0.22M	946	317	2.1	0.7	176	91	0.4	0.2
	Total	**509**	**0.50M**	**986**	**333**	**1.8**	**0.7**	**313**	**146**	**0.6**	**0.3**
TikTok	Software	32	1.11M	79	49	2.5	1.5	216	90	6.7	2.8
	Phone	36	1.84M	76	49	2.0	1.4	98	42	2.7	1.2
	Laptop	39	0.21M	75	48	1.9	1.2	119	46	3.0	1.2
	Car	37	1.08M	83	53	2.2	1.4	61	28	1.6	0.8
	Total	**36**	**1.07M**	**79**	**50**	**2.4**	**1.5**	**120**	**50**	**3.3**	**1.4**

Table 4.6: Video Content Statistics

Chapter 5

Discussion and Implications

The primary aim of our research was to explore the feasibility and value of leveraging video content to gain insights and enhance CrowdRE practices. Through this work, we uncovered compelling evidence that highlights the potential of video platforms in the realm of requirement elicitation and analysis. Our findings underscore the significance of video content as a valuable source of information for understanding user needs and refining requirements.

5.1 Videos: A source of Requirement Relevant User Feedback

TikTok and YouTube have become vibrant platforms where users actively participate in content creation and engage with the crowd. These platforms provide an interactive and immersive space where viewers can connect with creators through various means such as liking, commenting, sharing, and reacting to their videos. As a result, many content creators take advantage of these platforms to produce videos that include product reviews and discussions. In this work, we selected and analyzed videos on a diverse range of products from these platforms. These videos encompassed an array of product reviews, which proved to be valuable sources of user feedback. Creators shared their experiences, opinions, and insights about different products, shedding light on important aspects that influence user satisfaction. By analyzing video content available on TikTok and YouTube, we were able to uncover a wealth of information regarding user feedback. These videos offered valuable perspectives and served as a means for users to express their thoughts and opinions about various products. From

detailed reviews to lively discussions, the content creators on these platforms provided valuable insights that can inform companies and developers about the strengths, weaknesses, and overall user feedback.

An example from TikTok showcases the impact of a recent update on the user experience of Duolingo. A user expressed their frustration, stating, *"I didn't know how unmotivated I could be until this update Duolingo... The new Duolingo update is seriously messing me up I can't even get back into the lessons I was actively working on. Please revert it... Goodbye Duo it was fun. So sad that you were destroyed by an infantile update. Also note that this person has super Duolingo which means they pay for a subscription"*. This comment highlights how a seemingly innocent update can have detrimental effects on user engagement and satisfaction. The user's feedback indicates that the update introduced a series of bugs, causing inconvenience and hindering their progress within the language-learning platform. It is worth noting that the user mentions having a "super Duolingo" subscription, emphasizing their investment in the service. Despite being a paying customer, they are now considering leaving the product due to the negative impact of the update. This feedback provides a valuable opportunity for Duolingo to address the issues introduced by the update. For any organization aiming to reduce user churn, understanding and addressing bug reports and user feedback is crucial. Insights derived from these reports can inform the development of requirements that enable developers to tackle the identified issues effectively.

Furthermore, people often discuss the problems or issues they encounter while using a certain product. For instance, a user's comment about the Google Pixel 7 Pro smartphone highlights some of the issues they faced: *"Great phone, some bugs Google Pixel 7 Pro... I've just had quite a few instances where things will just randomly freeze up like apps will get stuck or I'll get stuck on just a black screen and can't get out of it. I keep swiping, trying to get out of it, and eventually, I do, but it still freezes, and things won't always work all the time, which is kind of frustrating" (TikTok).* Unlike textual feedback, videos provide a visual demonstration of the issues, allowing viewers to witness firsthand the freezing, app crashes, or other bugs mentioned by the user. Such rich details and visual evidence captured in videos can provide companies with a deeper understanding of the problems users face and the impact on their overall experience. While textual bug reports are a common form of feedback, videos on platforms like YouTube offer an additional layer of information that may not be present in other formats [31]. The combination of visual demonstration and the

user's commentary in these videos provides a comprehensive and insightful view of the encountered issues.

The videos from YouTube and TikTok offer a level of detail into problems and issues that companies can reference to understand underlying problems. Unlike traditional textual feedback, videos offer a unique advantage by providing visual and auditory context that enhances understanding. For instance, while an app review or a Reddit post may describe an issue in text or include a screenshot, a video can demonstrate how a bug occurs or its impact, allowing companies to observe and analyze the problem directly. The popularity of videos on TikTok and YouTube also adds to their significance. These videos often carry a sense of perceived honesty and objectivity, influencing both potential new users and current users. Therefore, our work emphasizes the importance of analyzing video-based content to extract requirements relevant to user feedback.

Once user feedback themes have been identified using our approach, the subsequent step for an organization is to develop requirements. This process is relatively straightforward, particularly for the product team responsible for creating user stories for the issue tracker. The user feedback themes serve as valuable guidance in determining the type of issue, such as a bug or a feature request. Moreover, the content presented in videos is typically explicit and provides clear insights into the specific issues at hand. By leveraging the user feedback themes, product personnel can easily translate them into actionable requirements. The themes act as a bridge between the captured user feedback and the subsequent steps in the requirement development process.

5.2 TikTok vs. YouTube: Impact on Requirements Elicitation

One important discussion point that arises from the research findings is the differences between TikTok and YouTube videos and their implications for requirement elicitation and analysis. The study highlights several notable distinctions, including video duration, content density, and the presence of visual text or subtitles. The implications of these differences in video duration between TikTok and YouTube highlight the need to consider the limitations and advantages of each platform when analyzing user feedback for requirement elicitation.

YouTube has more videos with relevant user feedback than TikTok, likely due to the factors of longer video and more in-depth discussions about each product. The shorter duration of TikTok videos means that users have a limited time to express their thoughts and provide feedback. This may result in more concise and condensed feedback, focusing on key issues or highlights. On the contrary, the longer duration of YouTube videos allows users to delve deeper into their experiences and provide more detailed feedback. They have more time to demonstrate specific bugs, showcase workarounds, or thoroughly explain their issues. This can provide valuable context and a more comprehensive understanding of user needs and pain points. The presence of visual text and subtitles in TikTok videos can play a crucial role in identifying relevant user feedback. Visual text complements the audio text and enhances the understanding of the video content. This suggests that the combination of audio and visual text can assist in accurately determining user feedback that can be utilized for further requirement elicitation.

When it comes to requirements elicitation, TikTok and YouTube serve different purposes. TikTok's user feedback can be particularly useful for capturing initial impressions, identifying broad areas of satisfaction or dissatisfaction, and gaining an understanding of user experiences at a high level. The concise and visually-oriented nature of TikTok videos can offer valuable insights into the perceived strengths and weaknesses of a product, as well as the overall user sentiment.

On the other hand, YouTube's user feedback lends itself well to in-depth analysis, comparison, and technical evaluation. The detailed comments, discussions, and expert opinions found on YouTube provide organizations with a wealth of information for identifying specific issues, gathering suggestions for improvements, and understanding the performance of their products in relation to competitors. By leveraging both TikTok and YouTube for requirements elicitation, organizations can gain a comprehensive understanding of user perspectives and preferences. Combining the quick, impression-based feedback from TikTok with the detailed, expert-driven feedback from YouTube allows for a more holistic approach to requirements elicitation. This multi-faceted approach ensures that organizations capture both the immediate reactions and the nuanced technical aspects of user feedback, leading to more informed decision-making in product development and refinement processes.

5.3 Implications for Practitioners

Our study's findings have significant implications for practitioners, highlighting the valuable role of user feedback from video content in the requirement generation process. By analyzing the content of videos, organizations can gain valuable insights into how users rate their products across various dimensions such as features, design, specifications, and performance. This firsthand feedback provides organizations with a direct understanding of user perspectives, enabling them to make informed decisions about product improvements.

Furthermore, the analysis of videos on platforms like YouTube can provide valuable competitor analysis. Organizations can gain insights into how their products compare to those of their competitors, identifying areas of strength and weakness. This information is crucial for companies aiming to enhance their products and gain a competitive edge in the market. For example, software products like Chrome, Firefox, Notion, and Discord often receive feature and user experience discussions in videos, which can be highly beneficial for organizations planning to introduce new features or refine existing ones.

Videos related to automotive products offer a unique opportunity for organizations to learn about user concerns regarding repair, maintenance, and efficiency. By analyzing these videos, companies can identify common issues raised by users and use this information to improve the quality and durability of their products. Moreover, the analysis of video content reveals valuable insights into user feedback on affordability, customization options, and modification suggestions. These videos provide organizations with direct access to consumer opinions and desires, enabling them to understand the specific needs and preferences of their target audience. It is worth noting that organizations have been leveraging YouTube videos for marketing and advertising purposes for some time now.

5.3.1 Implementation Cost

The implementation cost for organizations to adopt our approach is generally minimal. One key requirement is to build a web scraping pipeline that can download videos from YouTube and TikTok. Once this pipeline is set up, it can be used repeatedly for data collection. However, there may be some costs associated with extracting visual text from the videos. Third-party subscriptions for OCR extraction can be expensive, but there are alternative open-source tools available that can help mitigate

these costs. Processing a batch of 100 TikTok videos from a software organization would take approximately 15 minutes for audio text extraction and 1.5 hours for visual text extraction using hardware specifications of 2 x Intel E5-2683 v4 Broadwell @ 2.1GHz and a P100 16G RAM.

Once an organization has implemented the data analysis of user feedback, similar to our approach, they can identify the main user feedback themes. The next step would involve assigning an employee, such as a product manager or technical lead, to parse these user feedback themes into actionable requirements. Fortunately, this is a straightforward task. For instance, in the case of Duolingo's flawed new update mentioned in our sample content, a product manager can quickly identify that at the very least, the organization should roll back the changes to a previous version. Alternatively, they should focus on implementing a bug fix to allow users to access current lessons. Therefore, the actual financial cost to an organization for using our approach is limited, and the benefits of obtaining user feedback themes based on crowd insights outweigh the investment.

As users continue to actively engage in creating video content, they provide a diverse range of feedback about various products. Our study has the potential to influence industry requirements and product management practices, as practitioners can gain valuable insights into user behavior and concerns by utilizing our approach. By leveraging video content, organizations can access rich and nuanced feedback that goes beyond traditional text-based reviews and surveys. This deeper understanding of user preferences, issues, and suggestions can significantly impact product development and drive improvements aligned with user needs. Furthermore, our approach allows organizations to understand the collective intelligence of the crowd. With the increasing popularity of video content creation, the volume and diversity of user feedback continue to grow. By analyzing this vast pool of video-based feedback, organizations can gain a more comprehensive understanding of user behavior, preferences, and pain points. This, in turn, empowers them to make informed decisions, prioritize development efforts, and enhance their products to better meet customer expectations.

In conclusion, the cost of adopting our approach is minimal, primarily involving the setup of a web scraping pipeline and potentially utilizing open-source tools for text extraction. The process of deriving actionable requirements from user feedback themes is straightforward and can be undertaken by existing employees. By leveraging video content, organizations can gain valuable insights into user behavior, concerns,

and preferences. Our approach has the potential to influence industry practices, offering a fresh perspective on requirements elicitation and product management. The wealth of user feedback present in video content opens up new opportunities for organizations to better understand and serve their customers.

5.4 Implications for Researchers

5.4.1 Leveraging videos from social media platforms

Our research has significant implications for researchers and future studies, particularly in the area of leveraging videos from social media platforms as a valuable source of user feedback for identifying product requirements. The popularity and widespread use of social media, especially video-based platforms like TikTok and YouTube, offer an unprecedented opportunity to gather rich and diverse insights from users. Researchers can gain access to a wealth of user-generated content that provides valuable feedback on products by analyzing videos on these platforms. These videos allow for a more authentic and unfiltered representation of user experiences, as they capture real-time interactions, opinions, and concerns expressed by individuals from various demographics. This presents an exciting avenue for researchers to explore and extract valuable insights that can inform requirement elicitation processes. Future research should delve deeper into understanding the efficacy of leveraging videos from other emerging social media platforms. For example, researchers can delve deeper into understanding the efficacy of leveraging videos from other emerging social media platforms, such as Instagram Posts and Reels or Twitter's Vertical. Exploring alternative patterns and characteristics of these platforms can shed light on the inhibiting or enabling factors that influence the relevance and usefulness of user feedback for product development. Researchers can stay at the forefront of leveraging videos as a valuable tool for understanding perspectives by continuously adapting research methodologies to the evolving landscape of social media.

5.4.2 Advancing Methodologies for Analyzing Visual Text in Videos

Our research also highlights the need for further advancements in the methodological domain, particularly in analyzing visual text within videos. While our approach

focused on larger and more prominent visual texts, there is room for exploring alternative methods that can enhance the analysis of visual text, leading to more comprehensive insights into user feedback. Future work should explore different approaches to analyzing visual text, taking into account not only the size but also the contextual relevance and significance of the text within a video. Refining the methodologies for visual text analysis can help researchers to unlock the full potential of user feedback contained within videos. This is particularly crucial in platforms like TikTok, where visual texts are abundant and play a significant role in communicating user experiences and opinions.

One potential avenue for future research is the development of automated techniques for interpreting the visual content of videos and converting it into text. By leveraging computer vision and natural language processing algorithms, researchers can explore the possibility of extracting textual information from images and videos more accurately and efficiently. This can significantly augment the process of requirement elicitation by enabling the extraction of detailed and nuanced user feedback from the visual aspects of videos. Moreover, researchers can also explore the identification of additional information beyond textual content from the visual aspects of videos. This could involve analyzing facial expressions, gestures, or even the overall aesthetics of the video. Such analyses can provide deeper insights into user emotions, engagement levels, and preferences, which can further inform the requirement elicitation process.

5.4.3 Augmenting User Feedback Analysis through Correlation of Video Content and Accompanying Characteristics

In addition to the aforementioned implications, there is a significant opportunity for future research to explore the correlation between video content and other accompanying characteristics. By examining factors such as the number of likes, the number of comments, and the content of those comments, researchers can gain deeper insights into user feedback and refine it into actionable requirements. While previous studies have acknowledged the value of analyzing video comments [29], integrating this information with the video content itself and its associated characteristics holds the potential for even more comprehensive and insightful interpretation of user feedback. Considering the interplay between video content, engagement metrics, and user

comments, researchers can unlock additional layers of meaning and uncover valuable patterns that contribute to a more refined understanding of user requirements. One avenue for future exploration is the quantitative analysis of engagement metrics, such as the number of likes and comments in relation to specific aspects or features of the product. By identifying correlations between user feedback patterns and these engagement metrics, researchers can discern which product attributes resonate strongly with users and have a significant impact on their overall satisfaction or dissatisfaction.

Chapter 6

Conclusions and Threats To Validity

In any research study, it is crucial to assess the threats to validity to ensure the reliability and accuracy of the findings. Validity refers to the extent to which a study measures what it intends to measure and draws accurate conclusions. Threats to validity are factors or limitations that may affect the validity of the research outcomes. They can arise from various sources and can impact different aspects of the study. The following sections will examine the threats to construct validity, external validity, and internal validity. Each of these validity threats poses unique challenges and considerations that we needed to address to maintain the credibility and generalizability of their research findings.

6.1 Construct Validity

Construct validity refers to the extent to which the measurements in a study accurately capture the intended constructs or concepts. In our research, construct validity is crucial in ensuring that we measured precisely what we aimed to measure. However, there is a potential threat to construct validity concerning the manual labeling of video content as "relevant" or "irrelevant." The process of manual labeling can introduce subjective bias, potentially leading to inconsistent or inaccurate categorizations. To address this concern, we took several steps to enhance the construct validity of our study. Firstly, we established a clear definition of relevancy based on existing literature. This definition served as a guideline for the coders in determining the

appropriateness of the video content in relation to user feedback and requirements.

Furthermore, we employed two experienced coders who possessed in-depth knowledge and expertise in requirements concepts and pair coding. The inclusion of coders with domain-specific knowledge aimed to minimize subjective bias and enhance the accuracy of the labeling process. By leveraging their expertise, we aimed to ensure that the coders were capable of identifying and categorizing the video content accurately. To assess the reliability of our manual labeling process and enhance the construct validity, we employed Cohen's kappa and agreement levels as measures of inter-rater reliability. Cohen's kappa coefficient quantifies the level of agreement between coders beyond what could be attributed to chance. We obtained an objective measure of the reliability of our labeling process by calculating and evaluating the kappa coefficient. Higher agreement levels and a significant kappa coefficient would indicate a higher degree of construct validity in our study.

6.2 External Validity

External validity refers to the extent to which the findings of a study can be generalized to other settings, populations, or contexts. In our research, external validity is important for understanding whether our findings can be applied beyond the specific video platforms and software products examined in our study. One limitation to consider is that the generalizability of our findings may be limited to the specific video platforms (TikTok and YouTube) and software products we focused on. Each platform and product may have unique characteristics and user behaviors that could impact the nature of user feedback and requirements. To address this limitation, we took steps to enhance the external validity of our study. We selected TikTok and YouTube as our primary sources of video content due to their popularity and widespread use among diverse demographics. By including these two leading video-based social media platforms, we aimed to capture a broad range of user feedback and requirements. Additionally, we examined 20 leading products from four major industries, providing a diverse sample that encompasses different types of software products. This approach aimed to capture a wide range of user perspectives and experiences, increasing the likelihood of uncovering meaningful insights applicable to various software products. While our study focused specifically on TikTok and YouTube and the selected industries, we anticipate that the results and implications may be transferable to videos from other software products on these platforms. Given the

similarities in user behaviors and content creation across various software products, we expect that the patterns and themes of user feedback and requirements observed in our study may hold true for other software products on TikTok and YouTube.

6.3 Internal validity

Internal validity refers to the degree to which the conclusions drawn from a study are accurate and can be attributed to the variables being studied. In our research, there are factors that could potentially impact the internal validity of our findings, particularly regarding the visual text extraction process. One limitation that could affect the internal validity is related to the project budget constraints for optical character recognition (OCR) extraction. We utilized the "Google Vision" API, which demonstrated superior performance compared to the HuggingFace API. However, the cost associated with the "Google Vision" API, which is $1.50 per 1000 frames, posed a limitation on the extent of text extraction we could perform. This budget constraint may have influenced the completeness of the visual text extracted from the videos.

Another factor that could impact internal validity is the nature of videos as high-redundancy media. Videos often contain repetitive or redundant content, which can increase the processing time required for analysis. Despite employing algorithmic frame sampling techniques to mitigate this issue, there is still a trade-off between computation run-time and information loss. We chose parameters to sample frames at a minimum rate of 1.5 seconds per frame for TikTok and 2.5 seconds per frame for YouTube. However, this sampling rate may have resulted in missed frames that contained pertinent information relevant to user feedback and requirements. These limitations in the visual text extraction process could potentially introduce biases or gaps in the data, which may affect the internal validity of our study's conclusions. It is important to acknowledge these limitations and consider their potential impact on the accuracy and reliability of our findings.

6.4 Conclusions

Our study demonstrates the potential of video-based social media platforms, specifically TikTok and YouTube, as valuable sources for identifying user feedback and requirements relevant to various products. We employed natural language processing (NLP) and machine learning techniques to extract meaningful insights from analyzing

the audio and visual content of videos. Through our analysis of 20 different products across diverse industries, we observed that deep learning models like GPT-2 and RoBERTa proved effective in classifying video content into relevant and non-relevant user feedback. Additionally, clustering techniques allowed us to identify common user feedback themes. Popular themes that emerged from our analysis included user ratings and opinions on product features, bug reports, and concerns related to performance and efficiency. The findings of our study highlight the growing significance of videos as a medium for communication and content creation. As these platforms continue to gain popularity among users, organizations can harness the power of video-based social media to gain valuable insights into user needs and preferences for their products. Companies can refine their products and prioritize areas for improvement by understanding the discussions and opinions expressed in these videos. Furthermore, our research contributes to the field of CrowdRE by showcasing the efficacy of leveraging video content for requirement elicitation. Social media platforms provide a wealth of information, and our study demonstrates the feasibility of extracting valuable user feedback from videos to inform the product development process.

Looking ahead, there are promising avenues for future research to build upon this study. Researchers can explore alternative video platforms and analyze their suitability for requirements elicitation. Additionally, further investigation can be conducted to enhance the analysis of visual text and investigate the correlation between video content and other accompanying characteristics, such as likes, comments, and their content.

Bibliography

[1] Biggest social media platforms 2023.

[2] How to write a spelling corrector.

[3] pytube — pytube 12.1.3 documentation.

[4] Selenium.

[5] Spacy FastLang · spaCy universe.

[6] Vision AI | cloud vision API.

[7] Dejana Bajic and Kelly Lyons. Leveraging social media to gather user feedback for software development. In *Proceedings of the 2nd International Workshop on Web 2.0 for Software Engineering*, Web2SE '11, pages 1–6. Association for Computing Machinery.

[8] Andrew Begel, Robert DeLine, and Thomas Zimmermann. Social media for software engineering. In *Proceedings of the FSE/SDP workshop on Future of software engineering research*, FoSER '10, pages 33–38. Association for Computing Machinery.

[9] Melanie Busch, Jianwei Shi, Lukas Nagel, Johann Sell, and Kurt Schneider. Vision video making with novices: A research preview. In *Requirements Engineering: Foundation for Software Quality: 28th International Working Conference, REFSQ 2022, Birmingham, UK, March 21–24, 2022, Proceedings*, pages 251–258. Springer, 2022.

[10] Xiongfei Cao, Douglas R. Vogel, Xitong Guo, Hefu Liu, and Jibao Gu. Understanding the influence of social media in the workplace: An integration of media synchronicity and social capital theories. In *2012 45th Hawaii International Conference on System Sciences*, pages 3938–3947. ISSN: 1530-1605.

[11] Ning Chen, Jialiu Lin, Steven CH Hoi, Xiaokui Xiao, and Boshen Zhang. Arminer: mining informative reviews for developers from mobile app marketplace. In *Proceedings of the 36th international conference on software engineering*, pages 767–778, 2014.

[12] Salasac Clarissa, Joseph Lobo, et al. The rising popularity of tiktok during the pandemic: Utilization of the application vis-à-vis students' engagement. *American Journal of Interdisciplinary Research and Innovation*, 1(2):43–48, 2022.

[13] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. Number: arXiv:1911.02116.

[14] Subasish Das, Anandi Dutta, Tomas Lindheimer, Mohammad Jalayer, and Zachary Elgart. Youtube as a source of information in understanding autonomous vehicle consumers: natural language processing study. *Transportation research record*, 2673(8):242–253, 2019.

[15] Amanda Dash and Alexandra Branzan Albu. A domain independent approach to video summarization. In *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18*, pages 431–442. Springer, 2017.

[16] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. Number: arXiv:1810.04805.

[17] Andrea Di Sorbo, Sebastiano Panichella, Carol V Alexandru, Junji Shimagaki, Corrado A Visaggio, Gerardo Canfora, and Harald C Gall. What would users change in my app? summarizing app reviews for recommending software changes. In *Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering*, pages 499–510, 2016.

[18] Judith Gebauer, Ya Tang, and Chaiwat Baimai. User requirements of mobile technology: results from a content analysis of user reviews. *Information Systems and e-Business Management*, 6:361–384, 2008.

[19] Eduard C Groen, Joerg Doerr, and Sebastian Adam. Towards crowd-based requirements engineering a research preview. In *Requirements Engineering: Foundation for Software Quality: 21st International Working Conference, REFSQ 2015, Essen, Germany, March 23-26, 2015. Proceedings 21*, pages 247–253. Springer, 2015.

[20] Eduard C Groen and Matthias Koch. How requirements engineering can benefit from crowds. *Requirements Engineering Magazine*, 8:10, 2016.

[21] Eduard C Groen, Norbert Seyff, Raian Ali, Fabiano Dalpiaz, Joerg Doerr, Emitza Guzman, Mahmood Hosseini, Jordi Marco, Marc Oriol, Anna Perini, et al. The crowd in requirements engineering: The landscape and challenges. *IEEE software*, 34(2):44–52, 2017.

[22] Maarten Grootendorst. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. Number: arXiv:2203.05794.

[23] Anja Guzzi, Martin Pinzger, and Arie Deursen. Combining micro-blogging and IDE interactions to support developers in their quests. pages 1–5.

[24] Donna L Hoffman and Thomas P Novak. Toward a deeper understanding of social media, 2012.

[25] Athirai A. Irissappane, Hanfei Yu, Yankun Shen, Anubha Agrawal, and Gray Stanton. Leveraging GPT-2 for classifying spam reviews with limited labeled data via adversarial training.

[26] Wei Jiang, Haibin Ruan, Li Zhang, Philip Lew, and Jing Jiang. For user-driven software evolution: Requirements elicitation derived from mining online reviews. In *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part II 18*, pages 584–595. Springer, 2014.

[27] Georgi M Kanchev and Amit K Chopra. Social media through the requirements lens: A case study of google maps. In *2015 IEEE 1st International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*, pages 7–12. IEEE, 2015.

[28] Oliver Karras. *Supporting Requirements Communication for Shared Understanding by Applying Vision Videos in Requirements Engineering*. Logos Verlag Berlin GmbH, 2021.

[29] Oliver Karras, Eklekta Kristo, and Jil Klünder. The potential of using vision videos for crowdre: Video comments as a source of feedback. In *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, pages 298–305. IEEE, 2021.

[30] Natthaphon Kengphanphanit and Pornsiri Muenchaisri. Automatic requirements elicitation from social media (aresm). In *Proceedings of the 2020 International Conference on Computer Communication and Information Systems*, pages 57–62, 2020.

[31] M Laeeq Khan and Aqdas Malik. Researching youtube: Methods, tools, and analytics.

[32] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. Number: arXiv:1909.11942.

[33] Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. Trocr: Transformer-based optical character recognition with pre-trained models. *arXiv preprint arXiv:2109.10282*, 2021.

[34] Ze Shi Li, Manish Sihag, Nowshin Nawar Arony, Joao Bezerra Junior, Thanh Phan, Neil Ernst, and Daniela Damian. Narratives: the unforeseen influencer of privacy concerns. In *2022 IEEE 30th International Requirements Engineering Conference (RE)*, pages 127–139. IEEE, 2022.

[35] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. Number: arXiv:1907.11692.

[36] Michael Luca. Chapter 12 - user-generated content and social media. In Simon P. Anderson, Joel Waldfogel, and David Strömberg, editors, *Handbook of Media Economics*, volume 1 of *Handbook of Media Economics*, pages 563–592. North-Holland.

[37] Andrea Lucia and Abdallah Qusef. Requirements engineering in agile software development. 2.

[38] Walid Maalej and Hadeer Nabil. Bug report, feature request, or simply praise? on automatically classifying app reviews. In *2015 IEEE 23rd international requirements engineering conference (RE)*, pages 116–125. IEEE, 2015.

[39] Amy Madden, Ian Ruthven, and David McMenemy. A classification scheme for content analyses of youtube video comments. *Journal of documentation*, 69(5):693–714, 2013.

[40] Mary L. McHugh. Interrater reliability: the kappa statistic. 22(3):276–282.

[41] James Meneghello, Nik Thompson, Kevin Lee, Kok Wai Wong, and Bilal Abu-Salih. Unlocking social media and user generated content as a data source for knowledge management. 16(1):101–122. Publisher: IGI Global.

[42] Yeamduan Narangajavana Kaosiri, Luis José Callarisa Fiol, Miguel Ángel Moliner Tena, Rosa María Rodríguez Artola, and Javier Sánchez García. User-generated content sources in social media: A new approach to explore tourist satisfaction. 58(2):253–265. Publisher: SAGE Publications Inc.

[43] Bashar Nuseibeh and Steve Easterbrook. Requirements engineering: a roadmap. In *Proceedings of the Conference on The Future of Software Engineering*, pages 35–46. ACM.

[44] F. Paetsch, A. Eberlein, and F. Maurer. Requirements engineering and agile software development. In *WET ICE 2003. Proceedings. Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, 2003.*, pages 308–313. ISSN: 1080-1383.

[45] Dennis Pagano and Walid Maalej. User feedback in the appstore: An empirical study. In *2013 21st IEEE international requirements engineering conference (RE)*, pages 125–134. IEEE, 2013.

[46] Soujanya Poria, Erik Cambria, Amir Hussain, and Guang-Bin Huang. Towards an intelligent framework for multimodal affective data analysis. 63:104–116.

[47] Philips Kokoh Prasetyo, David Lo, Palakorn Achananuparp, Yuan Tian, and Ee-Peng Lim. Automatic classification of software related microblogs.

[48] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision.

[49] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners.

[50] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. 20:53–65.

[51] Kurt Schneider and Linda Marilena Bertolli. Video variants for crowdre: how to create linear videos, vision videos, and interactive videos. In *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pages 186–192. IEEE, 2019.

[52] Kurt Schneider, Melanie Busch, Oliver Karras, Maximilian Schrapel, and Michael Rohs. Refining vision videos. In *Requirements Engineering: Foundation for Software Quality: 25th International Working Conference, REFSQ 2019, Essen, Germany, March 18–21, 2019, Proceedings 25*, pages 135–150. Springer, 2019.

[53] Norbert Seyff, Irina Todoran, Kevin Caluser, Leif Singer, and Martin Glinz. Using popular social network sites to support requirements elicitation, prioritization and negotiation. 6(1):7.

[54] Tao Sheng, Jie Chen, and Zhouhui Lian. Centripetaltext: An efficient text instance representation for scene text detection. *Advances in Neural Information Processing Systems*, 34:335–346, 2021.

[55] Leif Singer, Fernando Figueira Filho, and Margaret-Anne Storey. Software engineering at the speed of light: how developers stay current using twitter. In *Proceedings of the 36th International Conference on Software Engineering*, ICSE 2014, pages 211–221. Association for Computing Machinery.

[56] I. Sommerville. *REQUIREMENTS ENGINEERING: A GOOD PRACTICE GUIDE*. Wiley India Pvt. Limited.

[57] Yuan TIAN, Palakorn Achananuparp, Ibrahim Nelman Lubis, David LO, and Ee Peng LIM. What does software engineering community microblog about? pages 247–250.

[58] James Tizard, Hechen Wang, Lydia Yohannes, and Kelly Blincoe. Can a conversation paint a picture? mining requirements in software forums. In *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pages 17–27. IEEE, 2019.

[59] Christoph Treude and Margaret-Anne Storey. How tagging helps bridge the gap between social and technical aspects in software development. pages 12–22.

[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[61] Peter Vistisen and Søren Bolvig Poulsen. Return of the vision video: Can corporate vision videos serve as setting for participation? 2017.

[62] Chong Wang, Maya Daneva, Marten van Sinderen, and Peng Liang. A systematic mapping study on crowdsourced requirements engineering using user feedback. *Journal of software: Evolution and Process*, 31(10):e2199, 2019.

[63] Grant Williams and Anas Mahmoud. Mining twitter feeds for software user requirements. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 1–10. IEEE, 2017.

[64] Li Xu, Xiaohui Yan, and Zhengwu Zhang. Research on the causes of the "tiktok" app becoming popular and the existing problems. *Journal of advanced management science*, 7(2), 2019.

[65] Pamela Zave. Classification of research efforts in requirements engineering. 29:315–321.

[66] Zheying Zhang. Effective requirements development–a comparison of requirements elicitation techniques.

Appendix A

Relevant Vs Non relevant Examples

Relevant	**Non-Relevant**
The new ASUS Zenbook 14 OLED...Speaking of typing the ZenBook 14 OLED has the Ergosense keyboard and touchpad installed The chiclet keys have good bounce which makes typing feel better and they are also spaced apart just right...o if you re ever stuck in a place that has no power outlet the great thing about the Asus ZenBook 14 OLED is that it has USB C easy charge support so that I can charge this up with a power bank no problem	ASUS ZenBook UM433DA 14 inch Full HD 300nits Laptop AMD Ryzen 7 16GB RAM 512GB SSD Bac ASUS ZenBook UM433DA 14 inch Full HD 300nits Laptop AMD Ryzen 7 16GB RAM 512GB SSD Backlit Keyboard Windows 10 Includes LED NumberPad Silver https day2dayshopping blogspot com search label LAPTOP bottoms
How To Add AdBlocker To Google Chrome Quick Easy...Now the first thing that you want to do is to type ad blocker into Google Then click on the first search result and you ll be redirected to the Google Chrome Web Store...	how to Design Google Chrome Logo beginners Let s Watch how to Design Google Chrome Logo beginners Design logodesign is
This Little Mistake Kills M2 MacBooks Today I will tell you when you SHOULD close your MacBook...Shutting down has several consequences for your Mac Firstly it clears the RAM of everything Open apps temporary files..You may not have known but turning off your Mac is actually doing it more harm than good And over time your Mac may load slower and have worse overall performance due to this I often go a long time without restarting my 16 inch MacBook Pro	My boss took me shopping today for a brand new macbookairm2 so grateful birthday in 3 days...
How To Turn on off Sync Across Clients on Discord App...So first thing go to discord now this is the homepage as you can see Go down and click on this option here down below And you will find many options now search for text and images And here is it now click on this option...	Hello adventurer Happy Scrolling I see well here some links that you may need on your travels Discord have you guys ever wanted to be an admin for an upcoming channel do you wanna have a drama free server...
Arabic on Duolingo An embarrassing mistake that needs to be fixed ASAP...But the word they used is "" that does not exist...	Duolingo showed me Magic Ayo Duo do a magic trick Really That was a good one

Table A.1: Examples of relevant and non-relevant videos user feedback

Relevant	**Non-Relevant**
Google Pixel 7 Pro Longterm Review...And there is no distortion with a screen edge or wasted space That's a good compromise Around back that funky camera band makes its return and honestly I don t hate it It beats having some big uneven camera bulge that throws off the weight and balance of the phone	Google Pixel 7 Pro Quick Unboxing Hazel Pixel 7 Pro Probably the Best phone of 2022 UK https amzn to 3Vn-QpDZ US https amzn to 3yx54Tsww check out...
In some countries the S23 is gonna be a HUGE upgrade So better wait a few weeks and don t buy an S22 right now...The S23 is gonna get the Snapdragon 8 Gen 2 which is first of all really powerful and can even compete with Apple s A16 bionic The special thing is that every S23 is going to get the A Gen 2...	Samsung S22 Pink Gold Hands on with the new Samsung s22 pink gold Outro Music
3 Worst Things Google Pixel 7 gcrpmedia pixel phone newtech The Pixel 7 is a great phone but here s three things that people hate about it right now First off the Pixel s 6 3 inch AMOLED display looks great but it s a bit smaller compared to the Pixel 6 Second off is a smaller battery than the Pixel 6 just by about 300 milliamps Thirdly even though it s a little nitpicky a lot of people do complain the fingerprint reader s still slow	Unboxing the BEST Color Google Pixel 7 Pro... Let s unbox the best Pixel 7 Pro color right now
How to create database templates Notion tutorial...You don t have to copy and paste or type the same thing over and over again I ll show you how in this quick Notion tutorial Go to your database and click the arrow down next to New You can edit existing templates by clicking on the three dots and select Edit or click New Template...	Notion Progress Bar without formula This new update is awesome
iPhone 14 vs iPhone 13 Review...Well almost identical because the A15 Bionic is still in the iPhone 14 The only difference is using the A15 from the iPhone 13 Pro which means you get an extra GPU core Now when it comes to day to day use of the phone you re not going to notice a difference compared to the iPhone 13...	Apple Event iPhone 14 Pro Apple Watch Ultra more What a day So much to get excited about iPhone 14 iPhone 14 Pro AirPods Pro 2nd gen Apple Watch...

Table A.2: Continued : Examples of relevant and non-relevant videos user feedback

Appendix B

Requirement Relevant themes emerged in Videos and Evaluation Matrix

Products	TikTok Themes	YouTube Themes
Discord	Feature Ratings Bug Report Usage Tutorials	Feature Ratings Bug Report Usage Tutorials
Duolingo	Feature Ratings Bug Report	Feature Ratings Usage Totorials
Chrome	Feature Ratings Bug Report Extensions Suggestions	Feature Ratings Matching Competition Extensions Suggestions
Firefox	Feature Ratings Bug Report Matching Competition	Feature Ratings Bug Report Matching Competition
Notion	Feature Ratings Usage Tutorials	Feature Ratings Usage Tutorials
Apple Iphone 14	Feature Ratings Matching Competition Older Version Comparison	Feature Ratings Matching Competition Older Version Comparison
Google Pixel 7	Feature Ratings Matching Competition	Feature Ratings Design Ratings Matching Competition
OnePlus 10	Feature Ratings Matching Competition	Feature Ratings Matching Competition
Samsung Galaxy S22	Feature Ratings Matching Competition	Feature Ratings Matching Competition

Table B.1: Major User Feedback Themes

Products	TikTok Themes	YouTube Themes
Macbook Air M2	Feature Ratings Performance Ratings Design Ratings	Feature Ratings Performance Ratings Matching Competition
Asus Zenbook 14	Feature Ratings Performance Ratings	Feature Ratings Performance Ratings Matching Competition
HP Spectre x360 14	Feature Ratings Design Ratings	Feature Ratings Matching Competition Performance Ratings
Microsoft Surface Pro 9	Feature Ratings Performance Ratings	Feature Ratings Matching Competition Performance Ratings
Dell XPS 15	Feature Ratings Performance Ratings	Feature Ratings Matching Competition Performance Ratings
Ford F150	Feature Ratings Repair & Maintenance Modifications Suggestions	Feature Ratings Repair & Maintenance Matching Competition
Tesla Model 3	Efficiency Feature Ratings Affordability	Efficiency Service Ratings Modifications Suggestions
Toyota Rav4	Feature Ratings Repair & Maintenance	Modifications Suggestions Feature Ratings
BMW X5	Feature Ratings Modifications Suggestions Affordability	Feature Ratings Matching Competition Performance Ratings
Mercedes Benz GLC	Feature Ratings Affordability	Feature Ratings Performance Ratings

Table B.2: Continued: Major User Feedback Themes

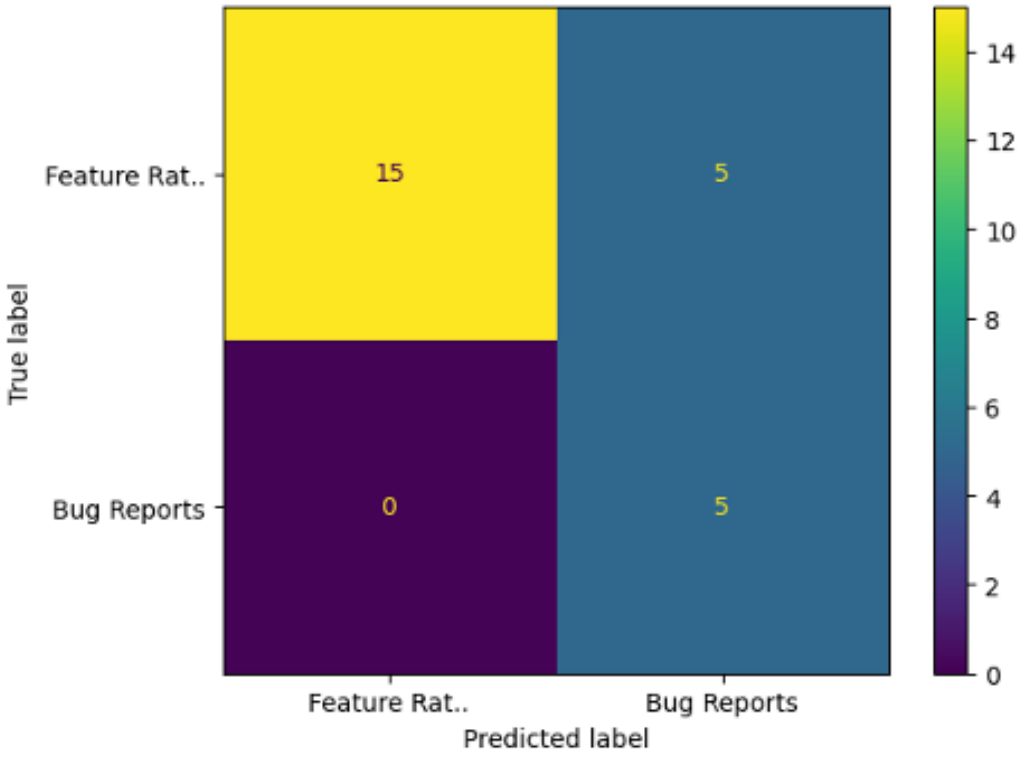

Figure B.1: Confusion Matrix by Manual Clustering Evaluation for Duolingo (TikTok) Clusters

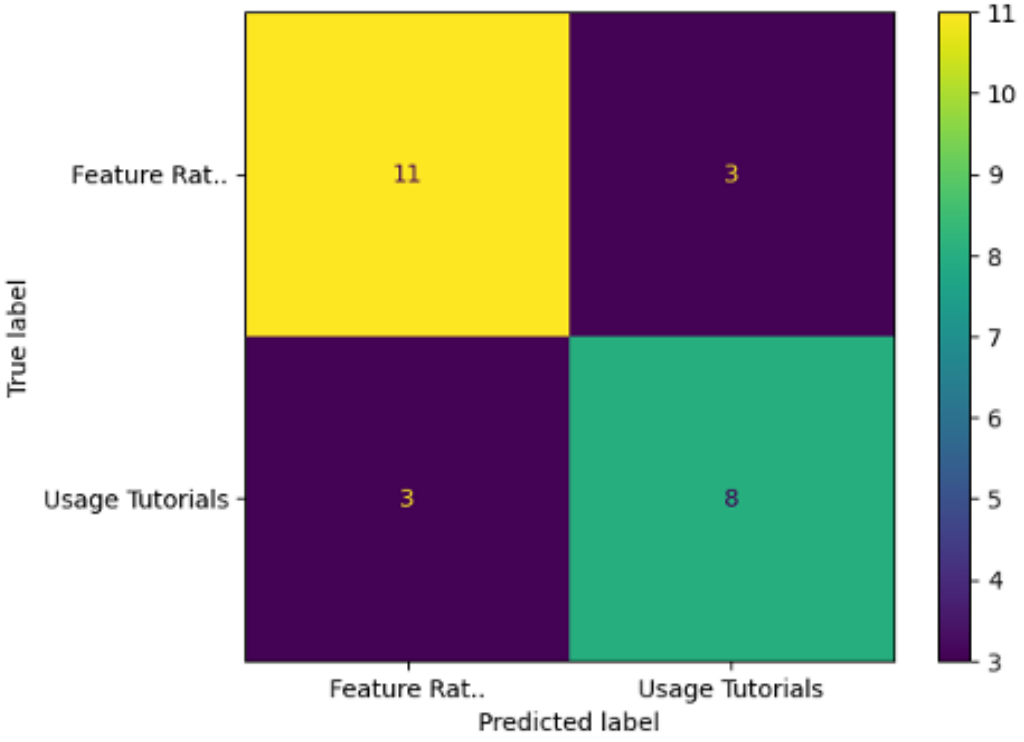

Figure B.2: Confusion Matrix by Manual Clustering Evaluation for Duolingo (YouTube) Clusters

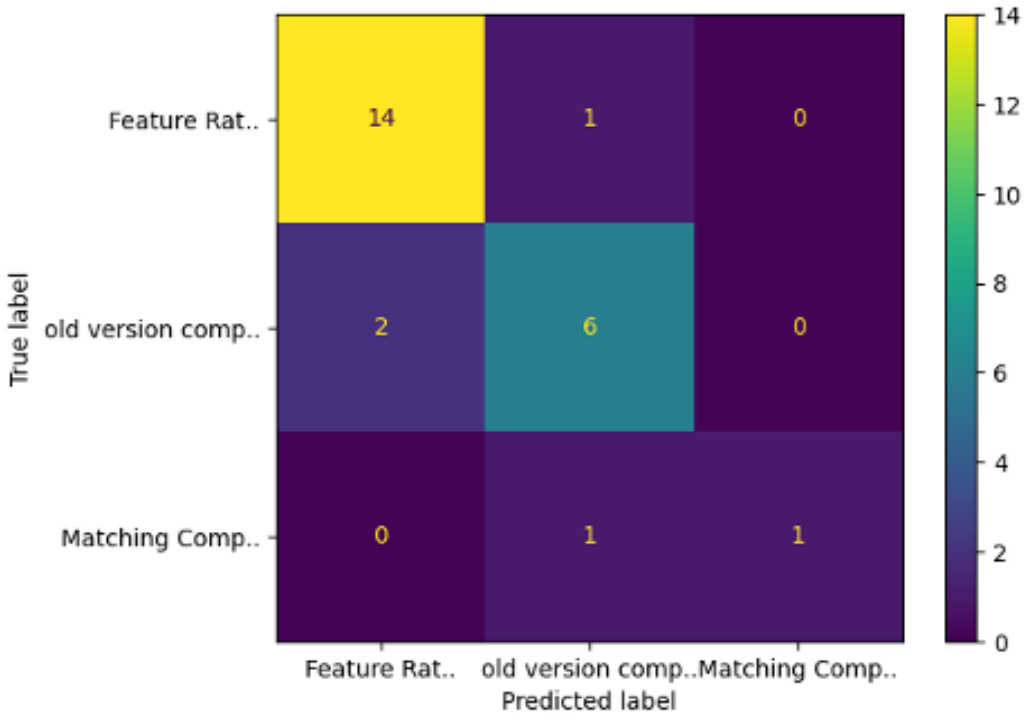

Figure B.3: Confusion Matrix by Manual Clustering Evaluation for Apple Iphone 14 (TikTok) Clusters

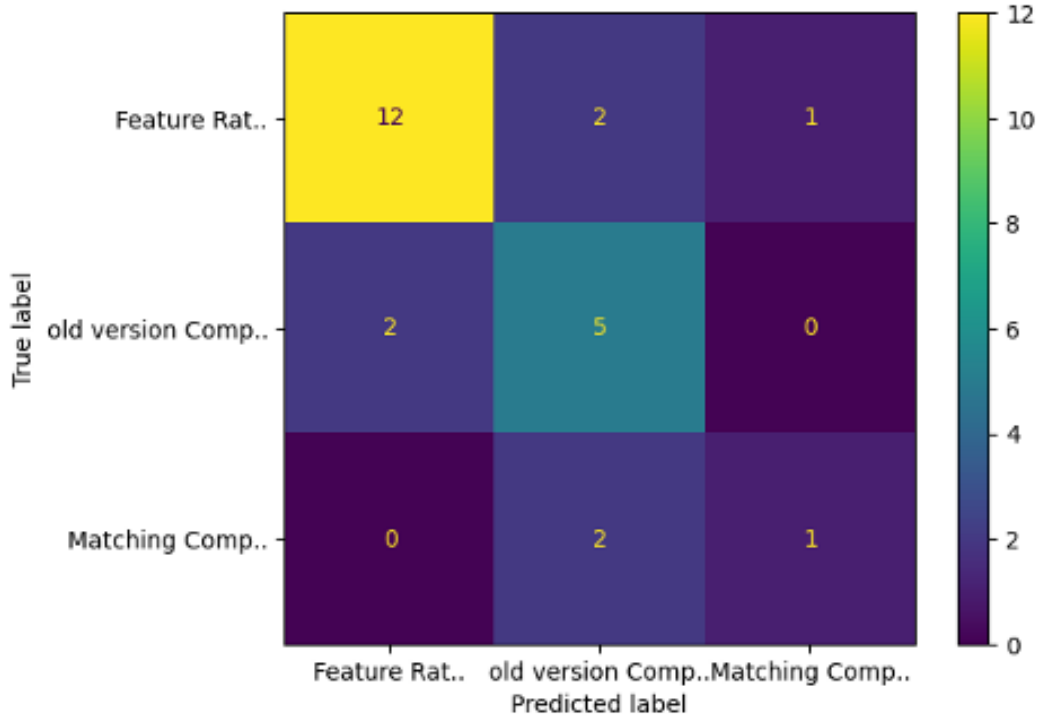

Figure B.4: Confusion Matrix by Manual Clustering Evaluation for Apple Iphone 14(YouTube) Clusters

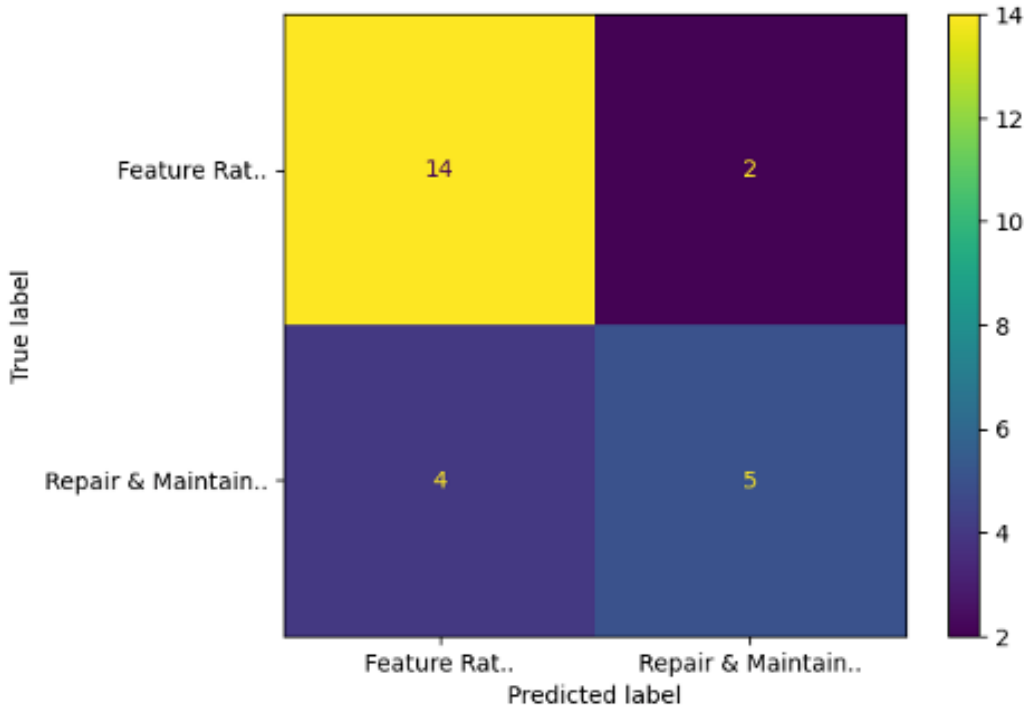

Figure B.5: Confusion Matrix by Manual Clustering Evaluation for Asus Zenbook 14 (TikTok) Clusters

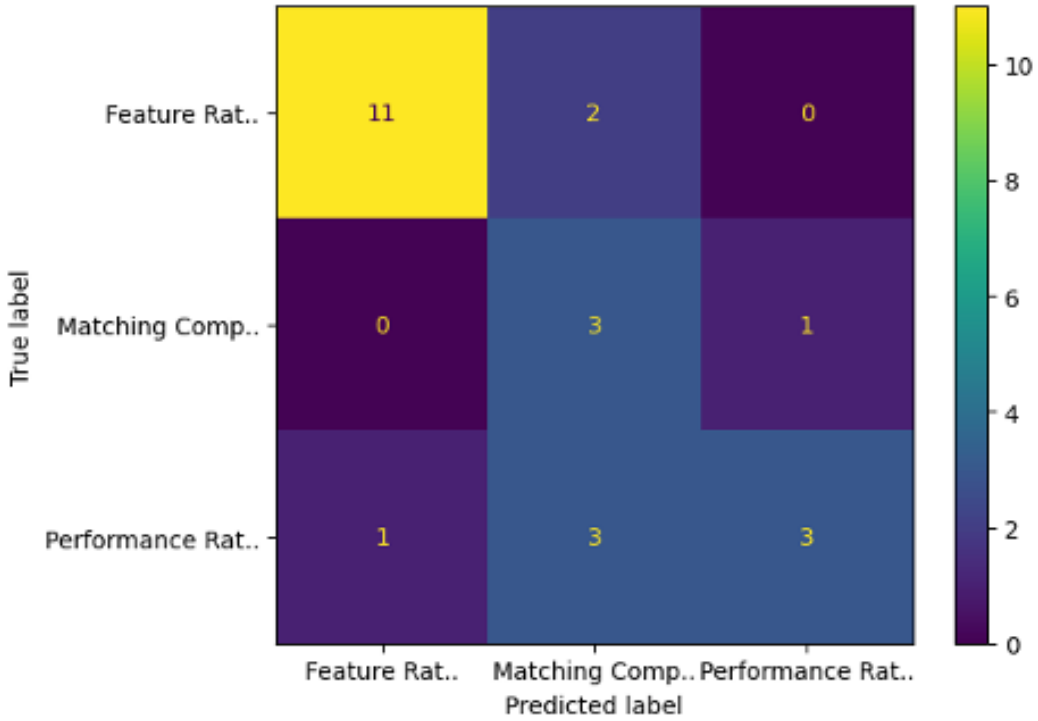

Figure B.6: Confusion Matrix by Manual Clustering Evaluation for Asus Zenbook 14 (YouTube) Clusters

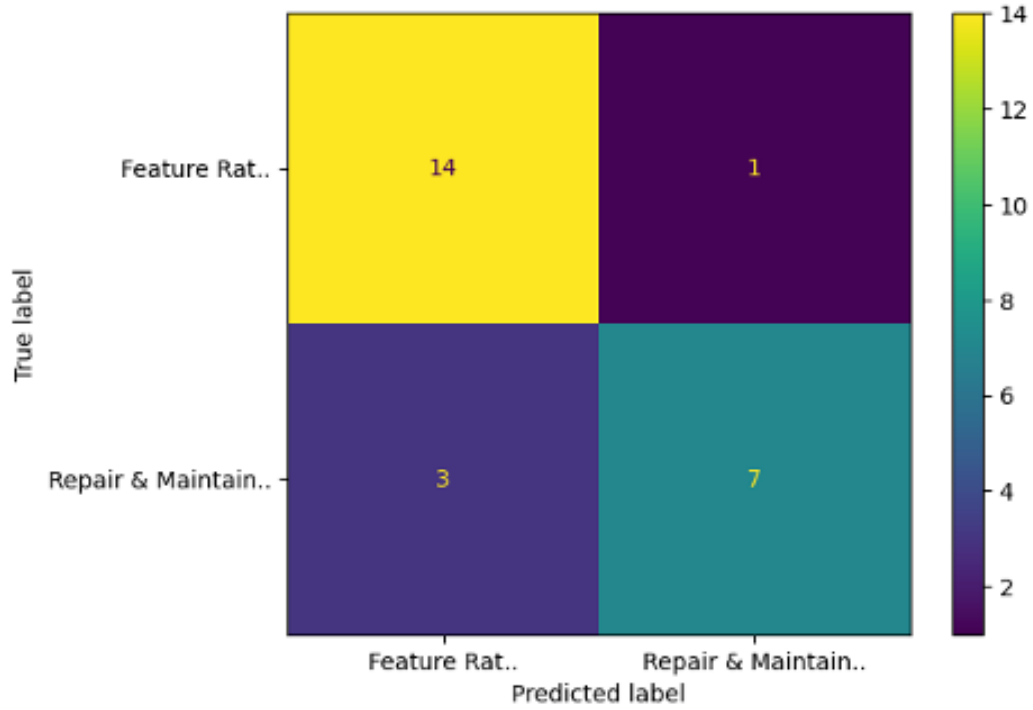

Figure B.7: Confusion Matrix by Manual Clustering Evaluation for Toyota Rav4 (TikTok) Clusters

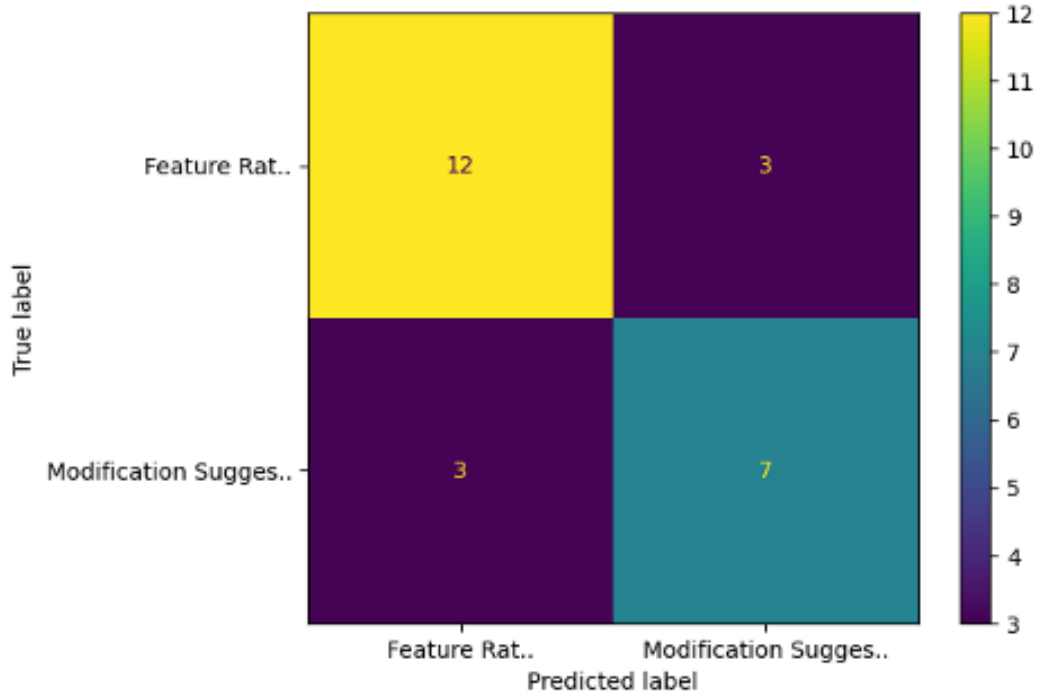

Figure B.8: Confusion Matrix by Manual Clustering Evaluation for Toyota Rav4 (YouTube) Clusters

Appendix C

Publications

A Data-Driven Approach for Finding Requirements Relevant Feedback from TikTok and YouTube

Manish Sihag, Ze Shi Li, Amanda Dash, Nowshin Nawar Arony, Kezia Devathasan, Neil Ernst,
Alexandra Branzan Albu, Daniela Damian
Department of Computer Science
University of Victoria, Victoria, Canada
{*manishsihag, lize, adash42, nowshinarony, keziadevathasan, nernst, aalbu, danielad*}*@uvic.ca*

***Abstract*—**
The increasing importance of videos as a medium for engagement, communication, and content creation makes them critical for organizations to consider for user feedback. However, sifting through vast amounts of video content on social media platforms to extract requirements-relevant feedback is challenging. This study delves into the use of TikTok and YouTube, two widely used social media platforms that focus on video content, in identifying relevant user feedback that may be further refined into requirements using subsequent requirement generation steps. We demonstrate an approach of using videos as a source of user feedback by analyzing audio and visual text, and metadata (i.e., description/title) from 6276 videos of 20 popular products across various industries. We employed state-of-the-art deep learning transformer-based models, and classified 3097 videos consisting of requirements relevant information. We then clustered relevant videos and found multiple requirements relevant feedback themes for each of the 20 products. This feedback can later be refined into requirements artifacts. We found that product ratings (feature, design, performance), bug reports, and usage tutorial are persistent themes from the videos. Video-based social media such as TikTok and YouTube can provide valuable user insights, making them a powerful and novel resource for companies to improve customer-centric development.

***Index Terms*—Requirement Elicitation, User Feedback, Video Platforms, Classification, TikTok, YouTube**

I. Introduction

Online videos are becoming more important for organizations to consider for user feedback as videos provide an immersive experience for viewers. Videos are a very popular medium for social media and communication [1]. For example, TikTok is one of the world's most popular video-based social media platforms [1], [2] and YouTube has also grown to an astronomical magnitude [3].

Previous research on requirements and videos has been limited to investigating the comments section of videos, while users engage in discussion [4]–[6]. However, videos are rich sources of data [7] with both audio and visual components, and metadata (e.g., description, title, date created). In this study we look at all three sources: the audio track of the video (converted to a transcript), any text that appears in the video, such as captions and subtitles, and the metadata.

Paying attention to the direction of CrowdRE research is critical for companies to improve requirements elicitation [8], [9]. The ability to vastly increase the amount of feedback considered [10] is extremely valuable. The process we propose converts video content to requirements-relevant feedback that can significantly impact companies' requirements and development activities.

We present a data-driven exploratory study on leveraging user-generated videos from TikTok and YouTube to identify requirements-related user feedback for 20 distinct products. This information can serve as a foundation for requirement elicitation, facilitating a more comprehensive understanding of consumer preferences and needs. We analyzed videos about products from a variety of industries, including software, consumer electronics, and automotive. Our approach involves extracting textual data from audio and visual content from the videos and processing using natural language processing (NLP) and machine learning (ML) techniques to uncover important user feedback that may not be captured through traditional elicitation methods.

Our study contributes to the growing body of research on using social media as a data source for product development and user feedback analysis. It also provides insights into the strengths of using videos as a data source and the opportunities of applying NLP and ML techniques to analyze video data. Our study was guided by three central research questions:

RQ1 How can video-based social media be used to identify requirements relevant user feedback?
RQ2 What are the main users feedback themes that we can identify?
RQ3 How do the different social media platforms and their video content affect user feedback?

From this exploration of videos contents from TikTok and YouTube, a number of findings have emerged. Our study presents the following contributions:

- An approach for identifying requirement relevant user feedback from video based social media.
- The most effective machine learning models (GPT-2 and RoBERTa) to classify the audio and visual content into requirements relevant user feedback.
- A list of requirements relevant user feedback themes for software, phone, computer, and automotive industries which can be further refined into requirements.

II. Background and Related Work

A. CrowdRE in Requirements Elicitation

With the rapid evolution of technology and social media, traditional elicitation techniques are insufficient to identify, gather and formulate requirements from the large distributed online community [8]. To address this gap, Groen *et al.* [8] proposed CrowdRE *"a semi-automated requirement engineering (RE) approach for obtaining and analyzing any kind of 'user feedback' from a 'crowd', with the goal of deriving validated user requirements."* User feedback from the 'crowd' is then transformed into requirements either through manual content analysis [11] or through automated approaches [12]. Groen *et al.* argue that CrowdRE can address the limitations of traditional RE methods, such as the limited scope and representation of user feedback [9]. By harnessing the collective intelligence of a crowd, organizations can utilize CrowdRE to identify and prioritize user needs and improve user engagement for their product [13].

The services offered by CrowdRE aim to provide motivational tools (e.g. gamification techniques, forums, visuals) that can inspire stakeholders to actively participate in a crowd [10]. A number of studies have focused on leveraging this crowd engagement on various platforms like app reviews and forums [**?**], [14]–[17], demonstrating that valuable insights can be gained by analyzing the conversations generated by users, such as their comments, feedback, and suggestions. Moreover, social media platforms have also been studied and utilized to analyze various aspects of requirement engineering [18]–[20]. Li *et al.* [21], found privacy related user feedback in a study on product related subreddits in Reddit. Kengphanphanit *et al.* [19], classified user feedback into requirements and non-requirements by scraping Twitter and Facebook, and utilized feature extraction on the user feedback based on polarity, subjective, and number of requirements word factors. Afterwards, they developed a model using the three factors and Naive Bayes method to generate requirements from the user feedback. Nevertheless, few studies have focused on exploring video based social media platforms such as YouTube and TikTok for identifying requirements relevant user discussions that have potential to be refined as requirements later on.

B. Video Platforms

YouTube with over 2.5 billion active users and TikTok with over 1 billion users, are the most popular video based social media platforms [22]. Video platforms like YouTube and TikTok have emerged as a valuable source for user engagement and obtaining requirements relevant user discussions [4], [5], [23]. In a paper by Madden *et al.* [5], the authors conducted an analysis on the content from 66,637 YouTube comments, and found 10 broad categories of user discussion, suggesting that classifying YouTube comments revealed opinions and attitudes of viewers towards the video content. In their study Das *et al.* [4] analyzed the comments generated on YouTube videos using natural language processing techniques and categorized comments on YouTube videos related to autonomous vehicles, further suggesting that YouTube can be a useful source of information for understanding consumer opinions and concerns. Karras *et al.* [6] utilized machine learning algorithms to analyze 4505 comments from a YouTube video as source of feedback and classified them into product relevant comments. The further manually analyzed the content of the relevant comments and found discussion on feature request, problem report, efficiency, and safety from the product relevant comments.

Schneider *et al.* [24], in their study describe that different types of videos (linear videos, vision videos, and interactive videos) can demonstrate concrete situations regarding a product and can be beneficial in engaging users to solicit feedback. Vision videos depict a vision of a future product or system, and they can help stakeholders to better understand and communicate their needs [25]. Hence, studies have been conducted on leveraging vision videos to elicit user feedback as users frequently engage in discussions and provide feedback on these videos [26], [27]. In another study, Karras *et al.* [6], argue that although the existing literature [26], [28] on vision videos discusses the benefit of using them in soliciting feedback, its potential for CrowdRE has not yet been fully explored. It remains unclear whether videos created by content creators themselves can provide valuable insights for companies.

Thus, in our study, we explore the potentials of video contents from YouTube and TikTok for identifying user feedback. Our work focuses on finding the pertinent themes from the user feedback. Since the information from the 'crowd' may also generate irrelevant data, extracting relevant information is a critical step before creating requirements. Once relevant themes have been identified, companies may use manual [11] or automated approaches [19] to generate requirements from the themes.

III. Methodology

We conducted an exploratory study to investigate the feasibility of using video-based social media platforms (i.e., TikTok and YouTube) for identifying requirements relevant user feedback themes. Our methodology is summarized by Figure 1.

A. Data collection

We conducted extensive market research and analysis to identify the top-performing products in each industry to build a representative dataset of twenty different products from 4 common consumer categories: software, mobile phone, computer, and automotive. Table I shows the dataset characteristics. We chose the most widely used software across different domains, including browsers such as Chrome and Firefox, tutoring applications like Duolingo, networking platforms like Discord, and productivity software like Notion. For each of the other categories, we strived to pick products that were among the best selling flagship products in North America in 2022, with on focus videos that would be in English. We chose not to select products that may otherwise sell more

units overall worldwide, such as Vivo versus Oneplus, but have a smaller English audience as they would impact the videos that we could collect. For example, for the products from the automotive industry, we chose 5 brands and their best selling model in North America in 2022. We applied the same approach to mobile phones and computers by selecting the most popular flagship product released in 2022.

To collect the data, we used the public facing APIs from TikTok and YouTube and scraped the videos by searching for each product using its name. The search term for each product is provided in the replication package [29]. We downloaded all the available videos from each search term and this process took about a day for each product. In total, we collected 11,341 videos, with 6,080 from TikTok and 5,261 from YouTube.

TABLE I: Products used for the analysis

Category	Products	TikTok Videos	YouTube Videos
Software	Notion	280	232
	Duolingo	224	217
	Discord	94	103
	Chrome	82	105
	Firefox	50	189
Phone	Google Pixel 7	223	183
	Apple Iphone 14	178	142
	Samsung Galaxy S22	162	214
	Motorola Edge 30	76	92
	Oneplus 10	59	119
Computer	Microsoft Surface Pro 9	201	187
	Apple Macbook Air M2	193	161
	Asus Zenbook 14	119	132
	HP spectre x360 14	130	95
	Dell XPS 15	30	49
Automotive	Tesla Model 3	210	193
	BMW X5	190	197
	Ford F150	187	102
	Toyota Rav4	177	305
	Mercedes Benz GLC	154	239

B. Preprocessing

The videos that we collected represented all the available videos according to our search term for each product. To ensure that our dataset is focused on user-generated content and not official promotional material, we implemented a two-level data filtration process. First, we filtered out any videos uploaded by official product accounts, as they are more likely to discuss product features in a promotional manner. The remaining videos in our dataset represented those that matched our search terms, but not from official product accounts such as Apple. Second, we filtered the videos to only include those in the English language. We used Spacy FastLang [30] to detect the language of the video description text, and OpenAI Whisper [31] to detect the language from the audio text, as described in detail in Section III-C1. We were left with 6276 videos after filtration.

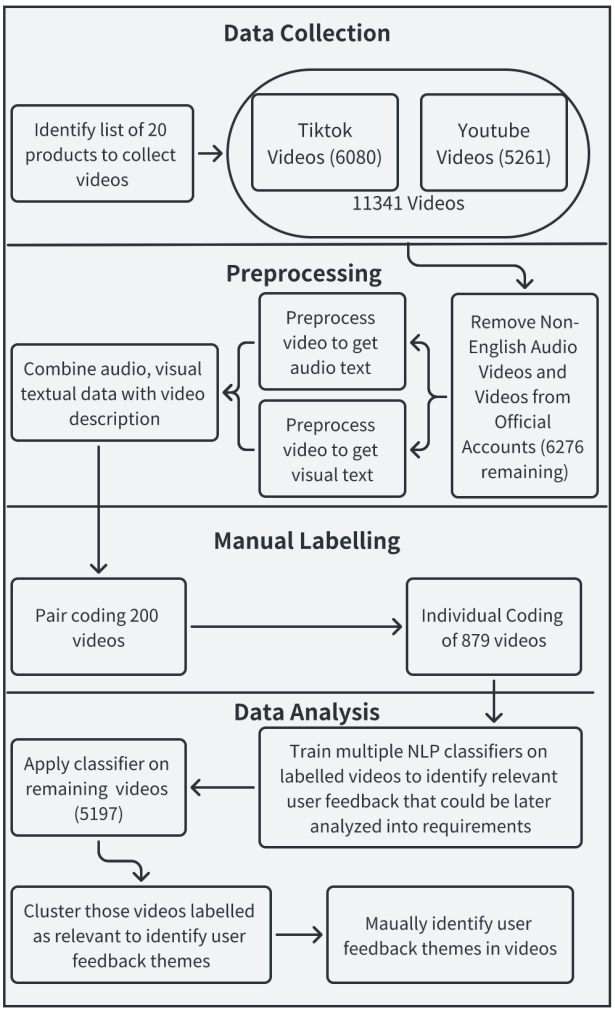

Fig. 1: Research Process

C. Analysis of Videos

Videos contain audio tracks, metadata in the form of descriptions, and finally text that appears in the video itself (such as a caption or subtitle). We used both the audio and visual text data extracted from the videos, as well as the descriptions provided by the content creators. To make sense of videos we worked with both the *visual* and *audio* elements of a video. First, we converted the audio of the video into text; secondly, we sampled visual frames and performed computer vision to collect any displayed text in a video (i.e., video subtitles). Out of the total of 6,276 videos in our dataset, 403 videos were found to have no audio content. Additionally, 101 videos did not contain any visual text. For each video we also paired this textual content with a video's metadata including video description and title where applicable. We then classified the TikTok and YouTube videos using various state-of-the-art deep learning models as user feedback that could be later refined into requirements (referred to as "relevant") or user feedback that was not useful for later refinement into requirements

("irrelevant"). We describe these techniques below.

1) Extracting Text from the Audio: OpenAI's speech recognition model "Whisper" [31] was used to extract audio text from videos. The "Large" Whisper model used in this study is one of the most accurate models and is designed for high-quality transcription tasks.

The extracted audio from videos was run on the Whisper model to generate transcriptions of the audio content. On average each TikTok videos was less a minute in duration, where as the YouTube videos were roughly eight minutes in length. Hence, some YouTube videos were a little bit longer in length. To reduce computational time, for the longer YouTube videos, we transcribe only the first thirty minutes of each audio and trim the rest. We assume that the premise of the video is conveyed in first 30 minutes, to minimize the processing time for longer videos. Whisper can detect the audio language being spoken during the transcription process and we utilized this to filter out videos that were not in English, as it is imperative to ensure the accuracy and relevance of the extracted text for our analysis.

2) Extracting Visual Text from the Video: In addition to audio, we also extracted visual text that may appear in a video as some content creators display subtitles or other important visual text. Since videos consist of many repetitive or similar frames, *motion-based video summarization* is used to select a small subset of all frames. For frames that are detected as having "text", we use an optical character recognition (OCR) system to extract the text; common spelling errors are corrected. The following section describes the video extraction pipeline.

For candidate frame selection, we use a modified version of the algorithm proposed by Dash and Albu [32]. This algorithm was chosen because it is a heuristic and not a ML-based video summarization system, therefore it is independent of the video domain. Their approach integrates motion and saliency analysis with temporal slicing to extract features and unique candidate frames from the video.

We do not use their candidate frame summarization; instead we leverage the information from the saliency energy map (instead of the slower background subtraction model) to find the probability divergence for the temporal slices. We take the Kullback-Leibler divergence, $D_{KL}(.)$, of each temporal slice, $k \in \{\text{vertical}, \text{horizonal}, \text{diagonal}\}$ at time $t-1$ and t, where t is defined as the current frame. We thus obtain a vector $s_t \in \mathbb{R}^3$ (Eqn. 1),

$$s_t^{(k)} = D_{KL}(p(k)_t || p(k)_{t-1}) \tag{1}$$

where $p(k)$ is the temporal slice k, normalized as a probability vector. The vector is then thresholded by values greater than $T_h \in \mathbb{R}^3$ to select the candidate frame (Eqn. 2); for this paper $T_h = [1e-4, 1e-4, 1e-4]$ is used.

$$\text{candidate frame} = \begin{cases} f_t & \forall k \in \{(s_t^{(k)} - s_{t-1}^{(k)}) > T_h\}, \\ \emptyset & \text{otherwise} \end{cases} \tag{2}$$

Intuitively, when the movement distribution changes significantly, a new candidate frame is selected. The candidate frames are analyzed for text using CentripetalText [33]. If no sufficient size text is detected, the candidate frame is discarded. In the next step, we consider two scenarios: (1) an audio track exists, and (2) no audio track exists.

When the primary content in a video is visual (i.e. no audio), we utilized Google's commercial state-of-the-art OCR system called "Google Cloud Vision"[1] to capture all text in the candidate frame. When audio is available, we assume the video content is primarily communicated by audio. Therefore, we supplement the audio by using HuggingFace's open-source OCR Tr-OCR [34] system with the "trocr-large-printed' pretrained weights to extract larger OCR text discovered by CentripetalText. Both OCR methods are not completely accurate, so we attempt to fix common spelling mistakes by processing the raw extract text with Peter Norvig's algorithm.[2] The choice to use multiple OCR algorithms was due to budget constraints.

D. Manual Labeling

To evaluate the accuracy and effectiveness of our classification models, we employed a manual labeling process to create a ground truth dataset for training. The dataset was randomly selected from our entire data pool, and we labeled total 1079 videos. Our labeling process consisted of labeling videos as either "relevant" or "irrelevant".

Our criteria to labelling a video as "relevant" include aspects such as problem reports, reviews of a product feature, comparison of features with competitors, feature requests, etc. In other words, any time a video included content that could be used by a company to make informative decisions regarding changes (positive or negative) to their product, we labelled it as "relevant". For example, *"To find out what the safest browser to use in 2022 is based on empirical testing techniques So we re going to go through 200 of the latest malware links... Firefox only blocked 145... Chrome not quite as good as Edge it blocked 198 links out of 200..." (Firefox)* was labeled as relevant. In contrast, a video that do not describe a product in any meaningful way or in a superficial manner (i.e., "The new M2 MacBook Air is finally for sale. I'm not gonna buy one") was labelled as irrelevant. Exemplified by these quotes, the pair coders would label a video as "relevant" if the main point of the video or substantial part of the video content (i.e., several sentences with details) discusses the product in a way that a company could take actionable steps. In the case of short videos, where the total amount of content is just a few sentences, the threshold for labeling as "relevant" could be a single sentence, but the sentence would need to provide adequate details.

To prevent bias towards any specific product, we made sure to label videos from each product that we analyzed. Two of our authors with extensive experience in requirement analysis, pair coded a set of 200 videos for the manual labeling process. The pair coding process resulted an average Cohen's Kappa

[1]https://cloud.google.com/vision

[2]https://norvig.com/spell-correct.html

score of 87%, indicating high levels of agreement between the coders. This high inter rater reliability also indicated that the separation between "relevant" and "irrelevant" was quite clear. After the successful completion of the pair coding process, one author individually labeled the remaining 897 videos. Of the 1079 videos that were manually labeled, 601 were labeled as relevant and 478 were labeled as irrelevant.

E. Data Analysis of User Feedback

1) Classification: We employed five state-of-the-art deep learning transformer-based models, namely GPT-2 (Generative Pre-trained Transformer 2) [31], BERT (Bidirectional Encoder Representations from Transformers) [35], RoBERTa (Robustly Optimized BERT Approach) [36], XLM-RoBERTa (Cross-lingual Language Model - Robustly Optimized BERT Approach) [37], and ALBERT (A Lite BERT) [38], to classify videos as either relevant or irrelevant. Fine-tuning these models allowed us to identify the most effective approach for video classification.

We evaluated the performance of these models using different combinations of data, including visual text, audio text, and both audio and visual text. Furthermore, we included title and description data for all combinations. By comparing the performance of these models, we aimed to identify the optimal approach for accurately classifying textual data from these popular video sharing platforms. For each of the five deep learning models (GPT-2, BERT, RoBERTa, XLM-RoBERTa, and ALBERT), we followed a similar training process. We used the pre-trained models and fine-tuned them on our dataset of labeled video text data, which included both the audio and visual text, as well as the video metadata (i.e., title and description). After training, we evaluated the performance of each model on a balanced test set of video text data. We measured the performance using accuracy and area under curve (AUC) metrics. We repeated this process for each combination of data (visual text, audio text, and both audio and visual text) and for each platform (YouTube and TikTok) to compare the performance of the models on each type of data and platform.

2) Clustering: We clustered the data to learn user feedback themes. We used BERTopic [39] to infer documents distribution over topics and then use BERTopic topic descriptions for clustering. BERTopic allows us to chose the cluster model. We selected K-means as our cluster model and ran the clustering process for 2 to 6 clusters. To determine the best cluster, we used the Silhouette Coefficient [40], a metric that measures how similar an object is to its own cluster compared to other clusters. After forming clusters, we conducted a manual analysis to thoroughly review the formed clusters. Based on that, we created theme names and assigned them to each cluster to represent their respective content.

TABLE II: Results of Different Deep Learning Models on Classifying between Relevant vs Irrelevant. AUC is area under curve.

Dataset	Model	Accuracy	AUC
YouTube with only visual text	GPT-2	0.71	0.71
	BERT	0.76	0.76
	RoBERTa	0.74	0.74
	XLM-RoBERTa	0.67	0.67
	ALBERT	**0.79**	**0.79**
YouTube with only audio text	GPT-2	**0.94**	**0.94**
	BERT	0.86	0.86
	RoBERTa	0.86	0.86
	XLM-RoBERTa	0.83	0.83
	ALBERT	0.79	0.79
YouTube with both visual and audio text	GPT-2	**0.91**	**0.91**
	BERT	0.85	0.85
	RoBERTa	0.80	0.80
	XLM-RoBERTa	0.80	0.80
	ALBERT	0.79	0.79
TikTok with only visual text	GPT-2	**0.71**	**0.71**
	BERT	0.70	0.70
	RoBERTa	0.50	0.50
	XLM-RoBERTa	0.50	0.50
	ALBERT	0.70	0.70
TikTok with only audio text	GPT-2	0.92	0.92
	BERT 0.92	0.92	0.92
	RoBERTa	**0.93**	**0.93**
	XLM-RoBERTa	0.90	0.90
	ALBERT	0.90	0.90
TikTok with both visual and audio text	GPT-2	0.93	0.93
	BERT	0.95	0.95
	RoBERTa	**0.97**	**0.97**
	XLM-RoBERTa	0.90	0.90
	ALBERT	0.93	0.93

IV. FINDINGS

A. RQ1: How can video-based social media be used to identify requirements relevant user feedback?

Recall that for each video we 1) converted the audio track into text and 2) performed computer vision to collect any displayed text in a video (i.e., captions or video subtitles). We then classified the TikTok and YouTube videos using various state-of-the-art deep learning models as either "relevant" or "irrelevant"; The results of the classification using these techniques are summarized in Table II. We observe that the datasets that leverage *audio* text paired with video metadata always performs extremely well. In particular, *audio* text paired with video metadata consistently performed better than *visual* text paired with video metadata. For YouTube videos and TikTok videos that utilize *audio* text paired with video metadata, accuracies of 94% and 93% were achieved.

We contrast these results with those datasets that leveraged visual text paired with video metadata. Table II shows that solely relying on text extracted from the video frames is not sufficient to identify requirements relevant user feedback. The most accurate model for classifying the dataset for YouTube's visual text was only able to achieve an accuracy equal to the worst performing model for the YouTube audio text dataset. TikTok videos using only video text and metadata is similar; 2

models had low accuracy of 50%, which for a balanced dataset means that it performs equal to a dummy model.

We surmise that the main reason for this is that audio extraction to text is quite accurate and as most videos include some host(s) speaking about the content, the audio text encapsulates the main idea of the video. In contrast, the visual text relies on sampling of visual frames to acquire the visual text, but this makes several assumptions 1) a video has clear subtitles that are easy to recognize 2) a video displays visual graphics of text that pertain to video's content. If a video's visual content had little visual text or did not include subtitles, a classifier had little to base decisions apart from accompanied metadata. While audio extraction to text suffers from potential limitations such as background music in place of a host's voice, the likelihood is lower that the existence of visual subtitles. Extracting text from audio also has less likelihood of encountering random audio that may confound the speech-to-text model.

Therefore, we found that datasets that utilized audio text performed better than datasets that utilized both audio and visual text. The only exception was "TikTok with both visual and audio text" where it actually performed 2% better than "TikTok with audio text." We believe the characteristics of TikTok videos (i.e., increased use of subtitles that complement the audio text of videos over YouTube) may be a factor for why the model could accurately classify "TikTok with both video and audio text". We expand on the effect of video content characteristics in Section IV-C.

Findings 1: Text extraction from videos using audio was highly effective for classifying videos from YouTube and TikTok. Text extraction using video on its own was not effective. However, for TikTok, the combination of text extraction using both video and audio was the most accurate option.

GPT-2 and RoBERTa models were most accurate for classifying the videos. For YouTube videos with audio text, GPT-2 significantly outperformed the other four models by 8-15%. RoBERTa was the best performing classifier for two out of the three TikTok datasets. Roberta had the highest accuracy for "TikTok with audio text" and "TikTok with both audio and visual text" with respective accuracies of 93% and 97%. The 97% RoBERTa achieved for TikTok with both video and audio text was highest accuracy we obtained in all our tests in Table II.

Findings 2: Deep learning models such as GPT-2 and RoBERTa can be utilized to perform classification of video content into relevant and non-relevant user feedback. GPT-2 was the most accurate model for classifying YouTube videos and RoBERTa was the most accurate model for classifying TikTok videos.

TABLE III: Result from Labelling and Classifying the Dataset

Dataset	Relevant	Irrelevant
YouTube Manual Labelling	370	167
YouTube with audio text classification via GPT-2	1691	1029
TikTok Manual Labelling	226	311
TikTok with both video and audio text classification via RoBERTa	810	1672
Total	3097	3179

After determining the most accurate models for TikTok and YouTube, we proceeded to classify the rest of the unlabelled dataset using these models. After classifying the rest, we also took a random sample of 50 videos with their classified labels and performed a round of manual annotation to determine the accuracy of the automated labelling.

We see consistency with the original experiments in Table II in the manual annotation. "YouTube with audio text" paired with GPT-2 achieved an accuracy of 98% and "TikTok with both video and audio text" paired with RoBERTa achieved 100%. We show in Table III the splits for *relevant* and *irrelevant* in the videos. In total, we found 3097 videos with relevant information and 3179 videos with non-relevant information for the 20 products in our study. YouTube videos (61%) had a higher concentration of requirements elicitation relevant videos compared to TikTok (34%).

Findings 3: Videos from YouTube and TikTok can be used to identify requirements relevant user feedback. Videos from YouTube (i.e., 61%) are more likely than videos from TikTok (i.e., 34%) to contain requirements relevant feedback.

B. RQ2: What are the main user feedback themes that we can identify?

We were able to group and cluster the videos based on their user feedback, allowing us to identify the most prominent themes, which are presented in Table IV. the themes were generated by reviewing the formed clusters and assigned theme names that represented each cluster accurately. These themes can serve as a foundation for further refinement and analysis for a company to derive requirement statements.

We found out that every product has videos related to the theme **Feature Ratings**. A *Feature Ratings* provides an overall rating of a product and often highlights specific features that are considered to be strengths or weaknesses. For example, *"I hate Firefox but I'm still switching back to it. I can't stand Firefox. Every time I try to use it I just get frustrated by all of the useless features privacy invading telemetry and annoying defaults. If only there was a way to use Firefox without all the junk." (Firefox on YouTube)*

Additionally, we observed a consistent theme among software product discussions regarding user feedback on updates and features, which also falls under Feature Ratings. Related to the latest updates and features, user often offered suggestions for further improvements. For instance, *"Please boost this so*

that Duolingo can see this I hate this new update so much... You can't jump between topics anymore which is so bad..." (Duolingo on TikTok).

A significant number of videos related to **Bug Reports** provided detailed explanations of the issues they encountered while using the software and may include suggestions for workarounds or solutions. To illustrate, *"Fix Discord... Discord App not launching on Windows 10... [To fix, Method 1] Close discord in task manager and restart it. Right click on the taskbar and click on task manager. Right click on the Discord option and click on end task." (Discord on YouTube)*. When it comes to software products, user-generated videos related to bug reports can be incredibly useful. These videos can provide developers with valuable insight into the issues that users are experiencing with their products.

Videos for **Usage Tutorial** help others with their user experience. Video content around software products frequently showed hacks and workarounds to help users make the most of their software tools. This highlights the importance of considering user feedback when designing and updating software products. *Usage tutorials* can be extremely useful for companies as they provide insight into how users are interacting with their products and potential areas where usability may be a concern. Previous studies have explored converting user feedback into requirements [11], [12], [19], hence an organization can similarly analyze the video content to identify common areas of confusion or difficulty in the usage of their products.

While **Feature Rating** was a theme found across all the products, it emerged as a particularly prominent theme for phones and computers. The prevalence of this theme in these product categories could be a result of features playing a crucial role in the purchase decision for these phones and computers. This theme was consistently observed across both TikTok and YouTube. However, our analysis also revealed that YouTube provides an in-depth review of a product against its competitors, which we labeled as *Matching competition*. To illustrate, *"The Pixel gets something iPhone users can only dream of and that is a 48 megapixel telephoto lens with five times optical zoom. Now on iPhone you only have a 12 megapixel telephoto that does three times optical zoom..." (Google Pixel 7 on YouTube).*

Such videos can be incredibly useful for all the related companies, as organizations can learn about the strengths and weaknesses in their product in a highly competitive market. With comparisons of similar products, a video can influence prospective customers on purchase decisions. A company can unlock highly useful user feedback regarding their own and competitor products if they collect these types of videos.

A prevalent theme among computers was the evaluation of the performance and design. **Design Ratings** revolve around creators focusing on the aesthetic appeal of these devices, while **Performance Ratings** often concern the performance quality of these products. For example, *"There's a few things about the S22 Ultra that just drive me up the freaking war.. this is the most powerful phone Samsung has ever made except it still lags in like the most random places ..." (Samsung Galaxy S22 on TikTok)* The example indicates to other users concerns over the performance of the device while using certain actions, which may be relevant user feedback for Samsung.

Automotive products also exemplified the **Feature Ratings** theme along with **Affordability**, **Performance Ratings**, and **Modification Suggestions**, which are important aspects of a product that can influence a user's decision-making. For *Modification Suggestions*, content creators often suggest modifications that other customers and users can add to their product to enhance their experience. When users and customers express their willingness to enhance their products through modifications, it indicates that they have a clear idea of their desired product features and functionalities. For example, *"My top 20 Toyota Rav 4 aftermarket upgrades and modifications Mods...I installed [improved] driving lights and I really like the driving lights. The ones that come with it are great, but of course they only work when you have it on low beam and these are of course for the fog lights." (Toyota Rav 4 on YouTube)* These videos provide opportunities for developers to better understand how customers use their products and identify areas for customization.

Furthermore, **Repair and maintenance** was also seen as a prominent theme from videos of automotive products. Video creators frequently discussed the various parts of a car that tend to wear out over time and shared tips on how to avoid and repair them. The information provided in these videos can be a valuable resource for automotive companies looking to improve the quality of their products. By analyzing the patterns of wear and tear identified by these video creators, companies can gain insights into how their products are performing over time. They can use this information to make improvements to the design and construction of their products to enhance their longevity and durability.

> **Findings 4:** User feedback themes can be generated from videos from YouTube and TikTok. These user feedback themes not only represent important aspects about products for companies to consider, but also represent relevant user feedback that companies can further refine into requirements in a subsequent step.

C. RQ3: How do the different social media platforms and their video content affect the user feedback?

We observed in RQ1 how the number of videos containing relevant user feedback that can be later refined into requirements differed significantly between YouTube and TikTok. Table V illustrates some of the differences between the audio and visual texts between YouTube and TikTok videos. In particular, we found that YouTube videos on average were at least 15 times longer than TikTok videos. It is then not surprising that higher percentage of YouTube videos would be afforded time to cover content related to feature ratings, bug reports, and discuss missing features that competitor products

TABLE IV: Requirement Relevant Themes

Theme	Description	Number of Products (out of 20)
Feature Ratings	Praise/criticism of product features	20
Matching Competition	Comparison with other competitor products	13
Performance Ratings	Praise/criticism of performance of the products	8
Modifications Suggestions	Suggestions for tools/upgrade	5
Bug Report	Bugs and issues of products	4
Repair & Maintenance	Videos related to fixing and preserving	3
Design Ratings	Design evaluation	3
Affordability	Cost prospects of the products	3
Usage Tutorials	Tutorials for other user to help use the product	2

provide. It is more difficult to cover these topics in a short span of 33 seconds, which is the average video span for TikTok.

However, we also noticed that on average TikTok video covered more words and unique words per second than YouTube. In particular, a TikTok video typically covered more than 2 times more unique words per second than YouTube. We believe this phenomenon was occurring largely because content creators have to squeeze more content in a shorter time frame. Hence, the audio is likely sped up to attract user attention.

Despite having shorter videos on TikTok, TikTok videos attract on average 2 times more views than the YouTube videos. This perhaps is not too surprising as TikTok videos are much shorter, so the theoretical videos viewed per hour watched for a typical user is likely also much higher than on YouTube. However, the higher view count is still a non-negligible characteristic that organizations should consider as higher view count may result in a bug report or feature rating becoming more influential.

TikTok videos on average had 5 times more visual texts per second than on YouTube and the number of unique visual texts per second was also significantly more. The high number of visual text on TikTok likely stems from the increased use of video subtitles on TikTok than YouTube so the visual texts often complement the audio of a video. The complementary factor between visual and audio text is potentially another reason why for TikTok the dataset with both visual and audio text had the highest accuracy among all TikTok datasets. We contrast this with YouTube videos where visual subtitles are less likely and the text that appear may actually confound the classifying models.

These differences may have contributed to the different results from our classification models. We used GPT-2 medium and RoBERTa base for our study, but GPT-2 medium is a significantly larger model with greater number of model parameters than RoBERTa base (i.e., Roberta = 125M vs GPT2=345M). The condensed nature of TikTok videos, generates less noise in the data requiring fewer model parameters. Smaller models tend to fit better to smaller data as they are less prone to overfitting, due to the fewer parameters. Our results largely reflect this expectation as RoBERTa was the most accurate model for 2 out of 3 TikTok datasets and GPT-2 was the most accurate for 3 out of 6 datasets.

Findings 5: YouTube has more videos with relevant user feedback than TikTok likely due to the factors of longer video and more in-depth discussions about each product. TikTok videos on average are more than *15 times* shorter than YouTube with much more condensed content, but has on average *double* the number of average views. Due to the frequent use of subtitles in TikTok videos, visual text can assist audio text in determining relevant user feedback that can be used later for developing requirements. These characteristics also affect the accuracy of various deep learning models that work for each platform.

V. Discussion

Our work indicates the potential to improve the practices of CrowdRE by utilizing valuable information present in video content.

A. Videos: A Source of Requirements Relevant User Feedback

TikTok and YouTube offer an interactive and immersive space for users to engage with the "crowd" through content creation. These platforms allow viewers to interact with the creator through likes, comments, shares, and reactions. In this work, we explore how videos extracted from these two social media can be used to identify requirements relevant user feedback.

For example, *"The new Duolingo update is seriously messing me up I can't even get back into the lessons I was actively working on. Please revert it... Goodbye Duo it was fun. ... Also note that this person has super Duolingo which means they pay for a subscription" (Duolingo on TikTok)* This exemplifies how a paying user is leaving the product due to a recently introduced update on Duolingo, which has resulted in a series of bugs on the platform. For Duolingo, if they considered such video, the actionable requirement would entail reverting the update or fixing the bug to not impede the user's experience. An organization that seeks to reduce user attrition may benefit from these bug report insights, and utilize them to develop requirements that developers could implement.

Furthermore, people often discuss about the problems they encounter while using a certain product, *"Great phone some bugs Google Pixel 7 Pro... I've just had quite a few instances where things will just randomly freeze up like apps will just get stuck or I'll get stuck on just a black screen and can't get out of it... things won't always work all the time which is kind of frustrating..." (Google Pixel 7 on TikTok)* Although bug report are a common issue for products in textual feedback (e.g. forums, app reviews), videos from YouTube offer extremely rich details about issues [7].

TABLE V: Video Content Statistics

Platform	Product	Avg. Sec.	Views Per Video	Audio Words Per Video	Uniq. Audio Words Per Video	Audio Words Per Sec.	Uniq. Audio Words Per Sec.	Visual Words Per Video	Uniq. Visual Words Per Video	Visual Words Per Sec.	Uniq. Visual Words Per Sec.
YouTube	Software	560	0.20M	808	253	1.4	0.5	546	231	1.0	0.4
	Phone	555	1.48M	934	334	1.7	0.6	232	111	0.4	0.2
	Laptop	480	0.20M	1,356	465	2.8	1.0	319	165	0.7	0.3
	Car	450	0.22M	946	317	2.1	0.7	176	91	0.4	0.2
	Total	**509**	**0.50M**	**986**	**333**	**1.8**	**0.7**	**313**	**146**	**0.6**	**0.3**
TikTok	Software	32	1.11M	79	49	2.5	1.5	216	90	6.7	2.8
	Phone	36	1.84M	76	49	2.0	1.4	98	42	2.7	1.2
	Laptop	39	0.21M	75	48	1.9	1.2	119	46	3.0	1.2
	Car	37	1.08M	83	53	2.2	1.4	61	28	1.6	0.8
	Total	**36**	**1.07M**	**79**	**50**	**2.4**	**1.5**	**120**	**50**	**3.3**	**1.4**

The videos from YouTube and TikTok offer a level of detail into problems and issues that companies can reference to understand underlying problems. For example, an app review may just include textual description about an issue [15], a Reddit post may include textual description along with a screenshot, but a video may include a short clip about how a bug was triggered or the outcome of the bug that the company can re-watch when they are creating actionable requirements. Since these videos are quite popular on TikTok and YouTube, they may influence potential new and current users with the perceived honest and objective opinions. Therefore, our work highlights the greater importance that organizations should place on analyzing the video based content for requirements relevant user feedback.

Example use case: We present an example scenario regarding how organizations can leverage the approach proposed in this study. Duolingo is one of the premier language learning applications in the world. If they manage their online user feedback, they could search and download for Duolingo related videos from video based social media like YouTube and TikTok. Upon converting the audio text and visual text from each video, the organization would be left with a series of videos with their corresponding audio and visual text. The organization could then run a RoBERTa model for all its TikTok video to identify the requirements relevant user feedback. Subsequently, the organization could use a topic modeling model, like BERTopic, to identify the various types of feedback. If the organization wants to identify bugs, they could expect to find bug related topics from the topic modeling. For instance, there could be multiple videos covering the issues related to the Duolingo update. If the problematic update was covered by multiple videos, the organization would find a common theme that describes the issue. The next step for the organization would be eliciting actionable requirement(s), maybe in the form of user stories, to resolve the user concern. Eliciting the actionable requirement is outside the scope of this work, but an organization can extend our approach in a further step. The work left for a product person to create a user story for their issue tracker is quite straight forward, as they could already infer the type of issue (i.e., bug, feature request, etc) and the content from a video is generally quite explicit about the specific issue.

B. Implications for Practitioners

Our findings suggest a number of implications for practitioners who can incorporate the user feedback from the video contents as part of their requirement generation process. An organization may learn about how users are rating their products in terms of features, design, specifications, and performance through the video contents itself. A company looking to improve their products in comparison to other competitors can take advantage of the competitor analysis found in these videos. The software products like Chrome, Firefox, Notion, Discord videos often contain feature and user experience related discussion, that may be beneficial for organizations to consider before rolling out a new feature. Products related to automotive companies have the potential to learn about user concerns related to repair, maintenance, and efficiency etc. Many of the videos further express the consumers feedback about affordability, customization and modification suggestions. Organizations have been analyzing YouTube videos for improving marketing and advertising purposes for a while now [41].

The cost for an organization to adopt our approach is for the most part minimal. They would need to build a web scraping pipeline to download the videos from YouTube and TikTok, but once its built, they can repeatedly use it. The main cost would likely come from extraction of visual text as OCR extraction using third-party subscriptions may be expensive, but there are alternative open-source tools that are available. Processing a software organization's sample batch of 100 TikTok videos would take approximately 15 minutes for audio text extraction and 1.5 hours for visual text extraction on 2 x Intel E5-2683 v4 Broadwell @ 2.1GHz and a P100 16G RAM. Once an organization implements data analysis of user feedback similar to our approach they can collect the main user feedback themes. The final step would involve having an employee such as product manager or technical lead to parse the user feedback themes into actionable requirements, but this should be straightforward task. For example, in the sample content regarding Duolingo's flawed new update, a product manager can quickly see that at minimum the organization should roll back changes to a previous version. Otherwise,

the organization should implement bug fix to allow users to access current lessons. Hence, the actual real dollars cost to an organization to use our approach is limited, and the organization could realize benefits of obtaining user feedback themes based on insights from the crowd.

Users are engaging in video content creation on a regular basis, providing various feedback about the products. Our study has potential to influence industry requirements and product management practices, as practitioners can gain valuable insights about user behavior and concerns.

C. Implications for Research

We believe our research has several important implications for researchers and future studies. First, we identified the efficacy of leveraging videos from two large social media platforms for identifying product requirements relevant user feedback. Social media, particularly video based on social media platforms, has exploded in popularity in recent years and their growth across different demographics provides new and important new sources for CrowdRE. Other large platforms are joining this foray, with Meta's Instagram Posts and Reels and Twitter's Vertical being the significant alternatives. Future research should explore these alternative patterns and study whether these patterns have characteristics that inhibit or enable relevant information for products.

Second, another area of future work is in the methodological domain. In our approach, we tried to focus on larger, more pronounced visual texts in videos as opposed to every possible text that may appear in a single frame, but future work should consider other approaches to analyzing the visual text. Other approaches may help to increase the usefulness of the visual text for identifying user feedback, especially on TikTok where we noticed that there were more visual texts in general. There is also the potential for researchers to explore identifying other information from the visual aspect of videos. Potential areas may involve automated interpretation of the visual content in a video and converting that into text.

While we tried to focus on user generated content, by filtering out official product account videos, it does not fully clear the dataset of potential sponsored or promotional types of videos. Future research could further separate these promotional content, perhaps through a classification filtering stage similar to the models used in our study, and explore the user generated content in more detail. Additional work from research could include automatic detection about actual feature requests or bugs as this could assist practitioners in identifying requirements from the videos.

Finally, future work can involve correlating video content with other accompanying characteristics in a video such as the number of likes, the number of comments, as well as the content of those comments. Previous work has already explored the usefulness of video comments [6], but utilizing both the video itself and its accompanying data could prove even more effective for interpreting user feedback that may be refined into requirements through subsequent steps.

VI. Threats to Validity

A. Construct Validity

Construct validity relates to whether we measured what we intended to measure. In our case, one threat relates to our manual labelling of "relevant" versus "irrelevant". There could be subjective bias introduced in this labelling, but we tried to mitigate this through a definition of this concept from literature and two coders who have in-depth experience in requirements concepts and pair coding. We used our Cohen's kappa and agreement levels as a measure for our reliability.

B. External Validity

In terms of external validity, there is the limitation that our study may not generalize to other video platforms or other software products. However, we tried to mitigate this issue by studying two leading video based social media platforms and exploring 20 leading products from 4 major industries. Therefore, we anticipate that videos from other software products on TikTok or YouTube will produce similar results.

C. Internal Validity

Our conclusions about the visual text extraction could be limited by our project budget in terms of optical character recognition (OCR) extraction. The "Google Vision" API had superior performance over the HuggingFace API, but the cost of the "Google Vision" API at $1.50 per 1000 frames was a constraining factor. Also, videos are high-redundancy media, which increases processing time, even with the algorithmic frame sampling we employed. Increasing the sampling rate will decrease the computation run-time, at the expense of the information-loss rate. We chose parameters to sample at an minimum rate of 1.5s/frame and 2.5s/frame for TikTok and YouTube, respectively. This may have caused missed frames containing pertinent information.

VII. Conclusion

Video based social media platforms generate a wide range of discussions regarding different products, and analyzing these platforms have become a popular CrowdRE practice. In this study, we explored how can we leverage videos from two of the most popular video based social media platforms: TikTok and YouTube, for identifying requirements relevant user feedback that may be refined later into requirements. We examined 20 different products across a range of industries and used NLP and machine learning techniques to analyze the audio and visual content from the two platforms. We found that deep learning models such as GPT-2 and RoBERTa are effective in classifying video content into relevant and non-relevant user feedback, and clustering techniques can be used to identify user feedback themes. Popular themes that emerged include user feedback about feature ratings, bug reports, performance and efficiency issues about the software and other consumer products. As videos continue to gain popularity as a medium of communication and content creation, organizations can benefit from leveraging this data source to gain insights into user needs for their products.

REFERENCES

[1] L. Xu, X. Yan, and Z. Zhang, "Research on the causes of the "tiktok" app becoming popular and the existing problems," *Journal of advanced management science*, vol. 7, no. 2, 2019.

[2] S. Clarissa, J. Lobo *et al.*, "The rising popularity of tiktok during the pandemic: Utilization of the application vis-à-vis students' engagement," *American Journal of Interdisciplinary Research and Innovation*, vol. 1, no. 2, pp. 43–48, 2022.

[3] D. L. Hoffman and T. P. Novak, "Toward a deeper understanding of social media," pp. 69–70, 2012.

[4] S. Das, A. Dutta, T. Lindheimer, M. Jalayer, and Z. Elgart, "Youtube as a source of information in understanding autonomous vehicle consumers: natural language processing study," *Transportation research record*, vol. 2673, no. 8, pp. 242–253, 2019.

[5] A. Madden, I. Ruthven, and D. McMenemy, "A classification scheme for content analyses of youtube video comments," *Journal of documentation*, vol. 69, no. 5, pp. 693–714, 2013.

[6] O. Karras, E. Kristo, and J. Klünder, "The potential of using vision videos for crowdre: Video comments as a source of feedback," in *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2021, pp. 298–305.

[7] M. L. Khan and A. Malik, "Researching youtube: Methods, tools, and analytics."

[8] E. C. Groen, J. Doerr, and S. Adam, "Towards crowd-based requirements engineering a research preview," in *Requirements Engineering: Foundation for Software Quality: 21st International Working Conference, REFSQ 2015, Essen, Germany, March 23-26, 2015. Proceedings 21*. Springer, 2015, pp. 247–253.

[9] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini *et al.*, "The crowd in requirements engineering: The landscape and challenges," *IEEE software*, vol. 34, no. 2, pp. 44–52, 2017.

[10] E. C. Groen and M. Koch, "How requirements engineering can benefit from crowds," *Requirements Engineering Magazine*, vol. 8, p. 10, 2016.

[11] J. Gebauer, Y. Tang, and C. Baimai, "User requirements of mobile technology: results from a content analysis of user reviews," *Information Systems and e-Business Management*, vol. 6, pp. 361–384, 2008.

[12] W. Jiang, H. Ruan, L. Zhang, P. Lew, and J. Jiang, "For user-driven software evolution: Requirements elicitation derived from mining online reviews," in *Advances in Knowledge Discovery and Data Mining: 18th Pacific-Asia Conference, PAKDD 2014, Tainan, Taiwan, May 13-16, 2014. Proceedings, Part II 18*. Springer, 2014, pp. 584–595.

[13] C. Wang, M. Daneva, M. van Sinderen, and P. Liang, "A systematic mapping study on crowdsourced requirements engineering using user feedback," *Journal of software: Evolution and Process*, vol. 31, no. 10, p. e2199, 2019.

[14] J. Tizard, H. Wang, L. Yohannes, and K. Blincoe, "Can a conversation paint a picture? mining requirements in software forums," in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 17–27.

[15] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? on automatically classifying app reviews," in *2015 IEEE 23rd international requirements engineering conference (RE)*. IEEE, 2015, pp. 116–125.

[16] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," in *2013 21st IEEE international requirements engineering conference (RE)*. IEEE, 2013, pp. 125–134.

[17] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora, and H. C. Gall, "What would users change in my app? summarizing app reviews for recommending software changes," in *Proceedings of the 2016 24th ACM SIGSOFT international symposium on foundations of software engineering*, 2016, pp. 499–510.

[18] G. M. Kanchev and A. K. Chopra, "Social media through the requirements lens: A case study of google maps," in *2015 IEEE 1st International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*. IEEE, 2015, pp. 7–12.

[19] N. Kengphanphanit and P. Muenchaisri, "Automatic requirements elicitation from social media (aresm)," in *Proceedings of the 2020 International Conference on Computer Communication and Information Systems*, 2020, pp. 57–62.

[20] G. Williams and A. Mahmoud, "Mining twitter feeds for software user requirements," in *2017 IEEE 25th International Requirements Engineering Conference (RE)*. IEEE, 2017, pp. 1–10.

[21] Z. S. Li, M. Sihag, N. N. Arony, J. B. Junior, T. Phan, N. Ernst, and D. Damian, "Narratives: the unforeseen influencer of privacy concerns," in *2022 IEEE 30th International Requirements Engineering Conference (RE)*. IEEE, 2022, pp. 127–139.

[22] "Biggest social media platforms 2023." [Online]. Available: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

[23] P. Vistisen and S. B. Poulsen, "Return of the vision video: Can corporate vision videos serve as setting for participation?" 2017.

[24] K. Schneider and L. M. Bertolli, "Video variants for crowdre: how to create linear videos, vision videos, and interactive videos," in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2019, pp. 186–192.

[25] O. Karras, *Supporting Requirements Communication for Shared Understanding by Applying Vision Videos in Requirements Engineering*. Logos Verlag Berlin GmbH, 2021.

[26] K. Schneider, M. Busch, O. Karras, M. Schrapel, and M. Rohs, "Refining vision videos," in *Requirements Engineering: Foundation for Software Quality: 25th International Working Conference, REFSQ 2019, Essen, Germany, March 18–21, 2019, Proceedings 25*. Springer, 2019, pp. 135–150.

[27] M. Busch, J. Shi, L. Nagel, J. Sell, and K. Schneider, "Vision video making with novices: A research preview," in *Requirements Engineering: Foundation for Software Quality: 28th International Working Conference, REFSQ 2022, Birmingham, UK, March 21–24, 2022, Proceedings*. Springer, 2022, pp. 251–258.

[28] M. Busch, O. Karras, K. Schneider, and M. Ahrens, "Vision meets visualization: Are animated videos an alternative?" in *Requirements Engineering: Foundation for Software Quality: 26th International Working Conference, REFSQ 2020, Pisa, Italy, March 24–27, 2020, Proceedings 26*. Springer, 2020, pp. 277–292.

[29] Anonymous, "A Data-Driven Approach for Finding Requirements Relevant Feedback from TikTok and YouTube," Jun. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.8088427

[30] "Spacy fastlang." [Online]. Available: https://spacy.io/universe/project/spacy_fastlang

[31] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision."

[32] A. Dash and A. B. Albu, "A domain independent approach to video summarization," in *Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18-21, 2017, Proceedings 18*. Springer, 2017, pp. 431–442.

[33] T. Sheng, J. Chen, and Z. Lian, "Centripetaltext: An efficient text instance representation for scene text detection," *Advances in Neural Information Processing Systems*, vol. 34, pp. 335–346, 2021.

[34] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "Trocr: Transformer-based optical character recognition with pre-trained models," *arXiv preprint arXiv:2109.10282*, 2021.

[35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," number: arXiv:1810.04805. [Online]. Available: http://arxiv.org/abs/1810.04805

[36] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," number: arXiv:1907.11692. [Online]. Available: http://arxiv.org/abs/1907.11692

[37] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," number: arXiv:1911.02116. [Online]. Available: http://arxiv.org/abs/1911.02116

[38] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," number: arXiv:1909.11942. [Online]. Available: http://arxiv.org/abs/1909.11942

[39] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," number: arXiv:2203.05794. [Online]. Available: http://arxiv.org/abs/2203.05794

[40] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," vol. 20, pp. 53–65. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0377042787901257

[41] K. Kousha, M. Thelwall, and M. Abdoli, "The role of online videos in research communication: A content analysis of y ou t ube videos cited in academic publications," *Journal of the American Society for information Science and Technology*, vol. 63, no. 9, pp. 1710–1727, 2012.