Emotion-Driven Music Recommendations: Integrating CNN and KNN for Personalized Playlists

Applied Research Project submitted in partial fulfillment of the requirements for the

degree of

Masters in Artificial Intelligence

(MSc. AI)

at

Dublin Business School

(DBS)

Supervisor: Paul Laird

August

2023

DECLARATION

I, Nelson Gikonyo Kamau, declare that the applied project submitted to Dublin Business School for the award of MSc Artificial Intelligence (MSc. AI) is the outcome of my independent investigations, unless otherwise indicated and properly acknowledged through references. Furthermore, I affirm that this work has not been presented for the purpose of obtaining any other academic degree.

Signed: Nelson Gikonyo Kamau

Student Number: 10621893

Date: 29/08/2023

ACKNOWLEDGEMENTS

Above all, I would like to express my profound gratitude to God, whose guidance, strength, and blessings have been the foundation of this journey. Your unwavering presence and divine intervention have provided me with the strength to overcome challenges and complete this thesis.

I am deeply thankful to my supervisor, Paul Laird, for their unwavering guidance, insightful feedback, and constant encouragement throughout the course of this research. Their expertise and dedication have been instrumental in shaping the direction of this study.

I am grateful to Dublin Business School (DBS) for providing the necessary resources, access to the library, and research facilities that facilitated the completion of this thesis.

My heartfelt thanks go to my friends and family for their unwavering support and understanding throughout this journey. Your encouragement and belief in me have been a constant source of motivation.

Lastly, I am thankful to all those who participated in interviews, surveys, and discussions, contributing their time and insights to this study.

Finally, this thesis would not have been possible without the collective efforts of those mentioned above and many others who have supported me along the way. Thank you for being a part of this remarkable journey.

ABSTRACT

This thesis delves into the captivating world of personalized music playlists, where emotions, technology, and user preferences converge. Through a unique blend of sentiment analysis, Convolutional Neural Networks (CNN), and K-Nearest Neighbors (KNN), we embark on a journey to create playlists that resonate with your feelings. The CNN deciphers emotions from facial expressions, shaping the emotional landscape, while KNN fine-tunes song recommendations for a harmonious experience. To ensure accuracy, a gradient boosting model steps in to validate emotional connections. We also explore the power of user feedback loops and the potential of multi-modal emotion recognition. With a strong ethical compass and an interdisciplinary approach, we uncover the profound connection between emotions and music recommendation. The results magnify the importance of emotions in shaping musical experiences, leading to a symphony of personalized playlists.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER ONE - INTRODUCTION

1.1 Overview

The realm of music appreciation has undergone a marked and transformative evolution, primarily catalyzed by the ascent of digital music platforms and the exponential proliferation of music collections. In the present epoch characterized by an expansive spectrum of musical choices, there exists an escalating demand for increasingly sophisticated systems that offer music recommendations. While conventional methodologies for recommending music have predominantly fixated on unraveling user predilections and identifying parallels amidst musical compositions, they have frequently bypassed a pivotal facet that profoundly influences the act of listening: the emotional affinity individuals establish with music. The nexus between music and emotions is deeply intricate, and the potential to adeptly apprehend and decipher the emotional resonance of songs holds the promise of substantially enriching the domain of personalized music recommendations.

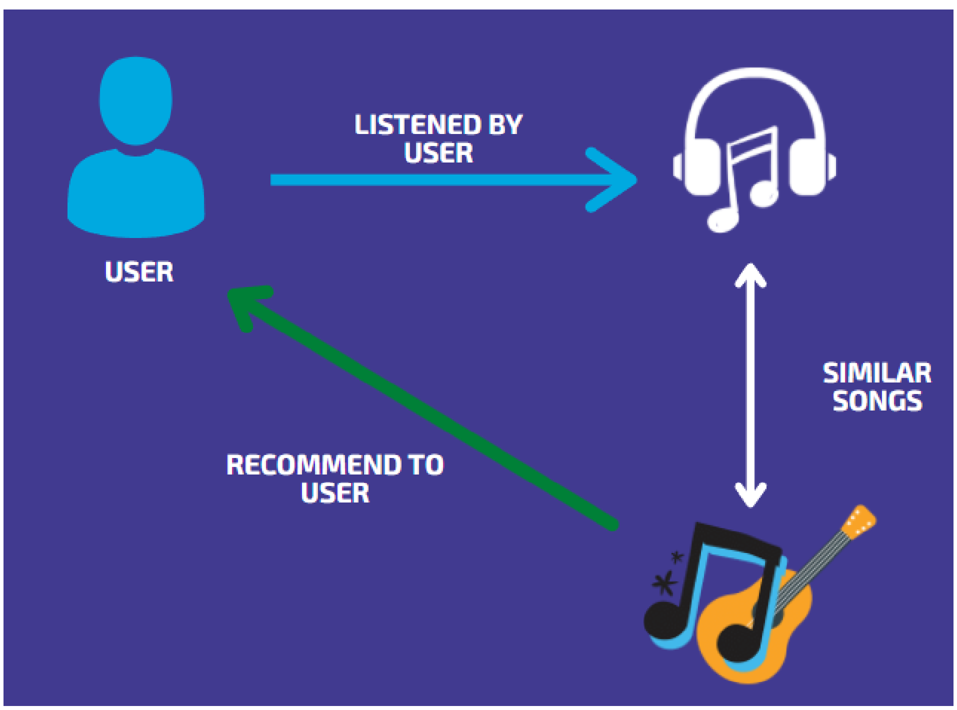

Figure 1: Song Recommendation Illustration (Navlani & Pandey, 2022)

The crux of this research is centered upon introducing a pioneering sentiment-centric music playlist recommendation framework, adept at harnessing advanced sentiment analysis methodologies. The core impetus behind this framework is to meticulously curate bespoke music playlists that seamlessly harmonize with the distinct emotional cadences or temperaments of users. By diligently taking into account the emotional fabric interwoven within musical pieces and deftly harmonizing these affective subtleties with users' proclivities, the envisaged system aspires to furnish recommendations that transcend superficial preferences and resonate on a profoundly emotional stratum. The ultimate fruition materializes as an elevated odyssey through music, a journey that not only caters to users' auditory tastes but also fosters a more profound emotional synergy between them and the recommended playlists. This approach holds the transformative potential to redefine the very contours of how individuals interact with music, forging an augmented and emotionally resonant tapestry of musical encounters.

1.2 Problem Statement

This research was born out of the growing demand for personalized music experiences that fully embrace the influence of emotions. We recognized that music has the incredible power to evoke a wide spectrum of feelings, ranging from sheer elation to introspective calmness and everything in between. However, existing music recommendation systems often failed to account for this emotional dimension, resulting in suggestions that occasionally missed the mark when it came to aligning with users' intended moods or emotional states.

In response to this gap, we embarked on a journey to bridge the emotional gap in music recommendations through the seamless integration of sentiment analysis. Our vision was to revolutionize the music recommendation landscape by offering users playlists that went beyond mere preferences, delving into the realm of emotions and inner resonances. By harnessing advanced sentiment analysis techniques, we aimed to provide users with song recommendations that truly captured and mirrored their emotional needs. This innovative approach was envisioned as a means to elevate the quality of musical journeys, curating playlists that resonated on a profoundly emotional level.

Our central objective was to overcome the limitations inherent in current systems by conceptualizing a playlist recommendation framework that placed emotions at its core. We sought to enable users to curate playlists that were more than just an assortment of tracks; they were emotional landscapes that reverberated with the listener's innermost feelings. By tapping into the emotional nuances embedded within both the listener and the sonic attributes of the songs, our system aimed to redefine the very essence of personalized music experiences. This research aimed to break new ground, merging the worlds of music, emotions, and technology into a harmonious symphony that truly spoke to the heart of every listener.

1.3 Research Objectives

The main goal of this research is to develop an advanced music playlist recommendation system that goes beyond considering the emotional content of songs. This system will also incorporate innovative concepts, such as utilizing K-Nearest Neighbors (KNN) for recommending songs and employing Convolutional Neural Networks (CNN) for capturing the listener's emotional state in real-time. To achieve this expanded objective, the following refined sub-goals will be pursued:

- Investigate and enhance state-of-the-art sentiment analysis algorithms to comprehensively assess the emotional content of songs. This will involve integrating techniques from natural language processing, audio signal processing, and machine learning to extract emotional cues from music data, accommodating both text and audio features.
- Establish a holistic recommendation framework that seamlessly combines sentiment analysis, KNN-driven recommendations, and CNN-based emotional capture. Design an architecture that not only generates tailored playlists in line with users' emotional preferences but also adapts to their evolving emotional reactions throughout their music listening journey.
- Assess the overall system performance by introducing novel evaluation criteria that encompass emotional relevance, recommendation precision, real-time emotional capture effectiveness, and user contentment. Conduct an in-depth comparative analysis against existing systems, showcasing the superiority of the enhanced approach in terms of both emotional alignment and technological ingenuity.

By steadfastly pursuing these refined research objectives, this study aspires to create an innovative sentiment-based music playlist recommendation system that harnesses advanced sentiment analysis, cutting-edge recommendation algorithms, and refined emotional capture techniques. The ultimate goal is to redefine the landscape of personalized music experiences, enriching the emotional connection between listeners and the music they love.

1.4 Research Questions

To guide this research, the following research questions will be addressed:

- How can existing sentiment analysis algorithms be enhanced to achieve a comprehensive assessment of the emotional content present in songs? What strategies can be employed to integrate techniques from natural language processing, audio signal processing, and machine learning, enabling the extraction of emotional cues from music data encompassing both textual and audio features?
- How can a holistic recommendation framework be established that seamlessly merges sentiment analysis, KNN-driven recommendations, and CNN-based emotional capture? What architecture can be devised to not only generate personalized playlists aligned with users' emotional preferences but also adapt to their evolving emotional reactions throughout their music listening journey?
- What evaluation criteria can be introduced to assess the overall system performance, encompassing emotional relevance, recommendation accuracy, real-time emotional capture efficacy, and user satisfaction? How can an in-depth comparative analysis be conducted against existing systems, demonstrating the superiority of the enhanced approach in terms of emotional alignment and technological innovation?

1.5 The Significance of the Study

The research holds profound implications for reshaping the musical landscape. By pioneering a sentiment-based music playlist recommendation system, this study bears the potential to revolutionize the very fabric of music engagement. Through this innovative approach, emotional resonance takes the center stage, elevating user satisfaction to new heights and extending their engagement on digital music platforms. The impact cascades across multiple dimensions, where music not only entertains but also nurtures emotions, fostering a more profound connection.

Consider the broader implications – a music recommendation system attuned to users' emotional states holds the promise of addressing their deeper needs. This system has the capacity to curate soundtracks for joyous highs, contemplative moments, or reflective lows. It ushers in a more holistic and fulfilling music listening journey, offering not just songs but emotional companionship. By allowing users to traverse a symphony of emotions, this research extends an invitation to an immersive experience that transcends mere audio, ushering listeners into a world where their sentiments are not just recognized, but also harmoniously echoed through music's notes.

1.6 Contribution to the Field

This research aims to introduce novel contributions to the realms of music recommendation systems and sentiment analysis. The integration of sentiment analysis techniques with established music recommendation algorithms stands as a distinctive approach that promises transformative outcomes. By uniting these two domains, the research envisions an enriched and more holistic perspective on personalized musical encounters.

The significance of this study lies in its ability to address a persistent gap in the current landscape of recommendation systems. By infusing the emotional dimension into the process, the research acknowledges the importance of emotional connections in music appreciation. This strategic fusion has the potential to yield heightened recommendation accuracy, with an understanding that a song's emotional impact can resonate deeply with listeners.

Furthermore, the research aspires to reinvigorate the user experience by fostering a deeper sense of personalization and emotional resonance. Beyond merely suggesting songs, it seeks to curate musical experiences that authentically reflect users' feelings, thereby forging a profound and enduring connection. These contributions are poised to reimagine the possibilities of music recommendation systems, setting a trajectory where music transcends its auditory realm and seamlessly integrates with users' emotional journeys.

1.7 Organization of the Thesis

Chapter 2: Background

In this foundational chapter, we lay the groundwork for our research by delving into the historical and conceptual backdrop of music recommendation systems and sentiment analysis.

By traversing through the evolution of these fields, we gain insights that provide context and perspective for our study. This chapter serves as the canvas on which the rest of our research unfolds, painting a picture of the landscape in which our innovation takes place.

Chapter 3: Related Work

Here, we immerse ourselves in a comprehensive exploration of related works that have paved the way for our research. As we journey through the annals of literature, we encounter studies, methodologies, and breakthroughs that have contributed to the collective understanding of music recommendation systems and sentiment analysis. This chapter serves as a bridge between the past and present, connecting our research to the lineage of knowledge and shaping our approach within this rich tapestry of scholarship.

Chapter 4: Methodology

In this chapter, we unveil the blueprint that guides our research's execution. The research methods, algorithms, and techniques that underpin the attainment of our objectives are meticulously detailed. By peeling back the layers of our methodology, we offer a transparent view into how we formulate and fine-tune sentiment analysis models. Additionally, we showcase the seamless integration of these models into the intricate fabric of our recommendation framework. This chapter serves as a roadmap for the intricate journey of our research.

Chapter 5: Implementation

With the groundwork laid in the previous chapters, this section brings our innovation to life. We provide an immersive glimpse into the practical realization of our sentiment-based music playlist recommendation system. Navigating through the tangible tools, technologies, and datasets harnessed, we shine a light on the real-world intricacies of translating theory into application. Throughout this chapter, we share the triumphs and challenges encountered during the implementation process, offering a vivid account of the journey from concept to reality.

Chapter 6: Findings and Analysis

The spotlight turns to the culmination of our research in this chapter, where we unveil the fruits of our labor. The outcomes of our sentiment-based recommendation system take center stage, showcasing its effectiveness in a tangible and quantifiable manner. This chapter delves into an in-depth analysis of the results, shedding light on the emotional resonance of the recommendations and the resulting user satisfaction. As we dissect and interpret the findings, we paint a comprehensive picture of the impact and value of our innovation.

Chapter 7: Discussion and Conclusion

In this final chapter, we engage in a comprehensive exploration that weaves together the threads of our research outcomes and their broader implications. The discussion delves into the significance of our findings within the context of music recommendation systems and sentiment analysis. As we reflect on the journey, we address limitations, highlight future avenues for exploration, and discuss how our work contributes to the broader scholarly landscape. This chapter serves as the ultimate culmination of our research, encapsulating the insights gained and the potential pathways ahead.

1.8 Ethical Considerations

As we delve into the intricate realm of sentiment-based music playlist recommendation systems, we do so with an ethical compass that guides every decision, every action, and every outcome. It is through this commitment that we affirm our dedication to conducting research that not only advances knowledge but does so in a manner that respects and honors the fundamental rights and dignity of all.

1.9 Limitations

While this research strives to make significant contributions to the field of music recommendation systems and sentiment analysis, it is important to acknowledge certain limitations that may influence the scope and generalizability of the findings.

1.9.1 Accuracy of Sentiment Analysis Models

The accuracy of sentiment analysis models, while meticulously refined, may still be subject to certain constraints. Emotions in music are intricate and multi-dimensional, and capturing them with complete precision presents challenges. Variability in language and cultural nuances further contributes to potential disparities between predicted and actual emotional content in songs.

1.9.2 Biases in Data

The effectiveness of sentiment analysis is intrinsically linked to the quality and diversity of the training data. The presence of inherent biases in the dataset can inadvertently affect the performance of the sentiment analysis algorithms. Ensuring a representative and unbiased dataset remains a persistent challenge in the field.

1.9.3 Generalizability of Recommendations

The generalizability of music recommendations, especially those driven by emotional resonance, may exhibit variations across diverse user demographics and cultural backgrounds. The emotional responses to music are highly subjective and can be influenced by personal experiences, rendering a universal emotional alignment a complex endeavor.

1.9.4 Technology Constraints

The real-time emotional capture facilitated by Convolutional Neural Networks (CNN) may face limitations due to technology constraints, such as hardware compatibility and processing speed. These factors could impact the system's responsiveness in capturing listeners' emotional states in live scenarios.

1.9.5 User Engagement and Satisfaction

While the study aims to enhance user satisfaction through emotionally resonant recommendations, individual preferences for emotional content in music can vary significantly. It is possible that some users may find the emotional alignment inadequate, affecting their overall engagement and satisfaction.

Acknowledging these limitations underscores the research's transparency and underscores the complexities inherent in tackling multifaceted issues at the intersection of music and technology. These limitations also present valuable avenues for further research and improvement in the future.

1.10 Summary

This research endeavor embarks on enriching the landscape of music recommendation systems through the infusion of emotional depth into the recommendation process. By engaging with the posed research inquiries, this study endeavors to elevate users' musical journeys, offering recommendations that carry a profound emotional resonance. The forthcoming chapters will delve into relevant literature, intricately outline the research methodology, expound upon implementation particulars, and rigorously analyze the findings. In the end, a comprehensive comprehension of sentiment-driven playlist recommendation systems will be unveiled, marking a significant stride in the evolution of musical experiences.

CHAPTER TWO - BACKGROUND

Within this chapter, we explore the various manifestations and uses of playlists, advocating for the utilization of song reviews and song features as a promising framework for playlist recommendation and exploration, all rooted in sentiment analysis. Consequently, our focus within this chapter encompasses an in-depth analysis of the foundational theories, a review of pertinent literature, an exploration of sentiment analysis methodologies, and a discourse on emotion capturing using Convolutional Neural Networks (CNN) and the utilization of K-Nearest Neighbors (KNN) in the realm of recommendation systems. This comprehensive examination serves to underpin our ongoing research, providing the essential foundation for our endeavors.

2.1 Emotion Capturing using Convolutional Neural Networks (CNN)

The intricate tapestry of human emotions finds a fascinating intersection with technology through the realm of emotion capturing using Convolutional Neural Networks (CNNs). These networks, rooted in deep learning, exhibit an unparalleled capability to decipher the nuances of facial expressions and map them to emotional states. The inception of this approach was steered by the recognition that emotions are fundamental influencers of human behavior and interactions, transcending barriers of language and culture.

2.1.1 CNN and Emotional Expression

According to (Srivastava & Bag, 2023), Convolutional Neural Networks have ushered in a revolution by enabling machines to interpret and understand human facial expressions, unraveling the emotional hues woven into them. By analyzing facial features, expressions, and their spatial arrangements, CNNs decipher the emotional undercurrents of joy, sadness, surprise, and more. This technological marvel has opened doors to real-time emotional analysis, allowing us to bridge the gap between the language of emotions and the language of machines.

2.1.2 The Evolution of Emotion Analysis

Emotion capturing using CNNs has evolved from its early stages of pattern recognition to embracing the nuances of complex emotions. The integration of diverse datasets, encompassing a myriad of expressions, backgrounds, and cultures, has propelled these models to transcend cultural boundaries and comprehend the intricacies of human emotional experiences. This evolution is underpinned by the perpetual growth of image datasets, fostering the CNNs' capacity to decode emotions across a spectrum of individuals and contexts.

2.2 K-Nearest Neighbors (KNN) in Recommendation Systems

In the ever-evolving landscape of recommendation systems, the K-Nearest Neighbors (KNN) algorithm emerges as a cornerstone, offering personalized recommendations by leveraging the musical similarities inherent in audio features. The essence of KNN lies in its ability to identify songs that resonate with users' preferences based on audio attributes, thereby enhancing the musical journey through a fusion of similarity-driven curation.

2.2.1 Unveiling Musical Affinities

KNN orchestrates its magic by identifying patterns within audio features, unveiling the hidden connections that tie songs together. By analyzing similarities in elements like melody, tempo, rhythm, and even tonalities, KNN encapsulates the essence of music beyond lyrics and language. This approach transcends the traditional bounds of genre, delving into the very fabric of musical composition.

2.2.2 The Intersection of Emotion and Music

Emotionally resonant music often transcends the boundaries of language and cultural nuances, making it an ideal terrain for KNN's prowess. This algorithm's capacity to unveil musically akin songs fosters a profound connection between music and emotions. It empowers the recommendation system to suggest songs that align with not only the audio attributes but also the emotional threads woven within the user's preferences.

As we delve into the heart of sentiment-based music playlist recommendation systems, this chapter provides a robust grounding in both the emotive capacities harnessed by CNNs and the song-harmonizing prowess of KNNs. By establishing a comprehensive understanding of these technologies, we lay the foundation for the innovative frameworks that drive our research forward.

2.3 Sentiment Analysis

Sentiment refers to an individual's viewpoint or expression of a particular view or opinion. Opinions play a crucial role in nearly all aspects of human endeavors and greatly influence our actions and behavior. Our beliefs, perceptions, and the choices we make are strongly shaped by the perspectives and evaluations of others. Consequently, when faced with decisions we often seek the advice and guidance of others (Zhang et al., 2018).

Sentiment analysis is a branch of natural language processing (NLP), it is also commonly referred to as opinion mining, and aims to identify the sentiment or emotional tone expressed in a document (Liu, 2011). It entails classifying and mechanically evaluating text to determine if it conveys a good, negative, or neutral mood. Sentiment analysis is a widely employed technique for extracting emotions, sentiments, and summarizations from large datasets. Analyst can leverage this information to make predictions (Medhat et al., 2014).

Sentiment found in text comes within texts and can be categorized into two types: explicit, where opinions are directly expressed in subjective sentences (e.g., "It is a beautiful day"), and implicit, where opinions are implied through the text (e.g., "The earphone broke in two days") (Ding et al., 2008). Sentiment analysis is particularly effective when applied to subjective contexts rather than objective ones. In objective contexts, the text mainly consists of standard sentences that do not convey any emotional tone or mood.

2.4 Sentiment Analysis Application

Sentiment analysis finds applications across various domains, encompassing individuals, organization, and governments. It plays a vital role by offering a comprehensive understanding of public opinion on diverse subjects, including product reviews, politics, movie reviews, and everyday life experiences. In the field of education, sentiment analysis assists in identifying students' learning progress, predicting their performances, and gaining insights into needs, allowing educators to adopt effective teaching strategies (Romero & Ventura, 2012), (Zawacki-Richter et al., 2019). In business, sentiment analysis facilitates the monitoring of customer sentiment towards products or brands, capturing overall trends in opinion. Furthermore, sentiment analysis provides an accurate depiction of emotions conveyed in movie and music reviews. Governments also utilize sentiment analysis to analyze trending topics related to policies and politics, enabling them to gauge public sentiment.

2.5 Sentiment Analysis Techniques

(Bhuta & Doshi, 2014), extensively investigate the research domain of Sentiment Analysis. Their research paper provides a comprehensive overview of the recent advancements in the field through a thorough literature survey. The paper begins by offering a detailed explanation of the concept of sentiment analysis, emphasizing its practical applications and the significance it holds for businesses in the present era. Subsequently, the authors delve into the various levels, approaches, and methodologies employed by researchers to explore the field of sentiment analysis. The methodologies employed by researchers to explore the field of sentiment analysis. The discussion encompasses the application of several deep learning algorithms for sentiment analysis.

Sentiment analysis algorithms can be broadly classified into three main categories: Lexicon-based approaches, machine learning approaches, and hybrid approaches. These categories serve as significant divisions in the study of sentiment analysis techniques.

2.5.1 Lexicon-based Approach

The semantic orientation of a document's words or phrases is used to infer the sentiment orientation of the text using the lexicon-based method of sentiment analysis. Using this method, the material is examined for terms that can elicit either good or negative emotions in readers. Lexicon-based techniques determine the sentiment orientation of a particular opinionated phrase of interest by making use of existing lexicons. A seed list of opinion words and the sentiment orientations that go with them is often the first step in the creation of lexicon (Taboada et al., 2011). This strategy makes extensive use of lexical information, such as dictionaries, vocabularies, or collections of words. To calculate the sentiment of the entire document, scores are awarded to every word in the lexicon as well as tactics like word placement and phrases. The Lexicon-based approach can be further categorized into two classes: corpus-based techniques and dictionary-based techniques.

2.5.1.1 Dictionary-Based Approach

The dictionary-based strategy entails creating a predefined list of seed opinion words, which is then used to expand the list using online dictionaries such as WordNet. Glosses, which are used to construct synonym and antonym associations between words, are provided by these dictionaries. In conjunction with these relationships, machine learning techniques are used to produce a more extensive list of opinion words.

2.5.1.2 Corpus-Based Approach

Unlike the dictionary-based technique, corpus-based opinion word generation is based on syntactic or co-occurrence patterns discovered in huge text corpora. The ability to get domain-specific and context-dependent orientations is one of the fundamental advantages of the corpus-based method. This method entails creating rules for connective phrases such as “And” “Or”, “But”, “Neither-or”, and “Neither-nor”, which are then used to broaden the seed set. It is now possible to construct domain-specific lexicons by leveraging a specific corpus of interest.

2.5.2 Machine Learning Approach

The conventional approach of Machine Learning continues to be extensively employed in sentiment analysis tasks, specifically in sentiment classification (Agarwal & Mittal, 2015). This approach utilizes various Machine Learning algorithms along with linguistic features. As a result, researchers have introduced numerous approaches to meet the growing demand for sentiment analysis. These approaches can be classified into three categories; unsupervised learning, supervised learning, semi-supervised learning, reinforcement learning and deep learning.

2.5.2.1 Supervised Learning Approach

This approach pertains to the process of training a model using annotated data, where the input characteristics and corresponding target labels are provided (Mitchell, 2011). By observing patterns and connections within the training dataset, the model acquires the ability to generate predictions or categorize new data. Supervised learning methods encompass decision trees, random forests, support vector machines (SVM), and neural networks, among others.

2.5.2.2 Unsupervised Learning Approach

Unsupervised learning involves working with unlabeled data, where the algorithm's objective is to discover patterns, structures, or connections inherent in the data (Mitchell, 2011). Tasks such as clustering and dimensionality reduction are frequently performed in unsupervised learning. Techniques like k-means clustering, hierarchical clustering, and principal component analysis (PCA) are examples of algorithms used in this domain.

2.5.2.3 Semi-Supervised Learning Approach

This method merges aspects of supervised and unsupervised learning. It entails training a model using a blend of labeled and unlabeled data, utilizing the labeled data to provide guidance during the learning process (Olivier et al., 2006). Semi-supervised learning is commonly employed in situations where acquiring labeled data is costly or time-consuming.

2.5.2.4 Reinforcement Learning Approach

Reinforcement learning encompasses an agent that gains the ability to make decisions and perform actions within an environment with the aim of maximizing a reward signal. The agent learns by engaging in trial and error, receiving feedback in the form of rewards or punishments based on its actions (Mitchell, 2011). Reinforcement learning algorithms, such as Q-learning, deep Q-networks (DQN), and policy gradients, are utilized in this process.

2.5.2.5 Deep Learning Approach

Deep learning refers to a branch of machine learning that specifically concentrates on neural networks comprising numerous layers. Referred to as deep neural networks, these networks possess the ability to acquire hierarchical representations from data. The application of deep learning has yielded impressive accomplishments across a range of fields, such as image identification, natural language processing, and speech recognition.

2.6 Sentiment Analysis in Music

2.6.1 Definition and Approaches

Sentiment analysis in music refers to the extraction of emotional cues from songs in order to analyze their affective influence on listeners (Kim and Andre, 2008). Various strategies have been used by researchers to approach music sentiment analysis. Lexicon-based approaches categorize the sentiment of song lyrics or textual metadata using sentiment lexicons, which contain terms linked with specific sentiments. To learn sentiment patterns from labeled training data, machine learning approaches such as support vector machines (SVM) and Naive Bayes classifiers or deep learning algorithms such as LSTM have been used. To improve the accuracy and coverage of sentiment analysis in music, hybrid approaches combine lexicon-based methods with machine learning techniques (Dang, Moreno-García and De la Prieta, 2020).

2.6.2 Feature Extraction and Representation

According to (Scheirer, 2000), various aspects of songs are retrieved to do sentiment analysis in music. Low-level audio qualities such as pitch, tempo, rhythm, and timbre capture the acoustic aspects that contribute to music's emotional expressiveness. High-level characteristics such as lyrics, chord progressions, tonality, and harmony give textual and structural clues that influence a song's sentiment (Barthet, Fazekas and Sandler, 2013). These features are extracted utilizing signal processing techniques, natural language processing (NLP) algorithms, and music analysis tools.

2.6.3 Sentiment Classification Algorithms

A wide range of sentiment categorization techniques have been applied to music data. Traditional machine learning methods, such as support vector machines (SVM), decision trees, random forests, and Naive Bayes classifiers, have been used for sentiment classification based on extracted characteristics (Kanakaraj and Guddeti, 2015). Deep learning algorithms, such as recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and convolutional neural networks (CNNs), have lately demonstrated promising results in capturing complicated sentiment patterns in music (Zhang, Wang and Liu, 2018). These techniques make use of neural network power to learn hierarchical representations of music data and capture long-term dependencies.

2.7 Music Recommendation System

Recommender systems have become an important part of our daily lives, supporting us in making decisions by producing personalized recommendations based on our preferences and interests. These intelligent systems sift through massive volumes of data and utilize complex algorithms to provide recommendations personalized to specific users. In this introduction, we will look at the concept and significance of recommender systems, as well as their numerous applications in various fields and the usefulness of tailored recommendations in improving the overall user experience.

2.7.1 Recommender Systems Definition

Recommender systems are software tools designed to assist users in discovering relevant items, such as products, movies, music, or articles, based on their preferences and historical interactions (Jannach, 2012). These systems utilize various techniques, including collaborative filtering, content-based filtering, and hybrid approaches, to predict user preferences and generate personalized recommendations.

The importance of recommender systems lies in their ability to alleviate information overload and decision fatigue. With the rapid growth of digital content and the overwhelming number of choices available, users often struggle to find items that align with their interests. Recommender systems solve this problem by intelligently filtering and presenting options that are most likely to appeal to users, thereby simplifying the decision-making process and enhancing user satisfaction.

2.7.2 Types of Music Recommendation Systems

Recommender systems can be broadly categorized into three primary groups: collaborative filtering methods, content-based filtering methods, and hybrid recommender systems.

2.7.2.1 Collaborative Filtering

Collaborative filtering is a popular approach in recommender systems that leverages the collective wisdom of users' preferences to generate recommendations (Su & Khoshgoftaar, 2009), this information may be like users' listening history. This technique can be categorized into user-based and item-based collaborative filtering.

Collaborative filtering based on user preferences involves the identification of similar users through their historical preferences of different music. Recommendations are then generated based on items favored by these similar users (Wang et al., 2006). This approach rests on the belief that users who share similar tastes will also share similar preferences for various items.

On the other hand, item-based collaborative filtering concentrates on detecting comparable items predicted on users' preferences, where suggestions are produced by analyzing items that bear resemblance to ones that users have relished or communicated with (Barragáns-Martínez et al., 2010). This strategy presumes that users who have proved alike items in the past will presumably possess alike preferences for additional items. The methodology aims to alleviate the problem of presenting myriad recommendations to users who have already seen the suggestions.

Collaborative filtering experts often rely on matrix factorization and latent factor models to discover hidden patterns in the user-item interaction matrix (Bokde, Girase and Mukhopadhyay, 2015). These powerful techniques simplify how users and items are represented, allowing for targeted recommendations that can be highly effective.

Collaborative filtering techniques aimed at making recommendations rely on identifying similar users or items within a neighborhood. This approach involves using the preferences of neighboring entities to inform suggestions, which can be achieved through various methods, including k-nearest neighbors (k-NN) and clustering algorithms.

2.7.2.2 Content-based Filtering

Content-based filtering is an approach that focuses on the characteristics and attributes of items for making recommendations (Konstan et al., 1997). This method entails extracting features from music such as audio characteristics, metadata, and lyrics for measuring their similarity to the generated recommendations.

To represent items effectively, feature extraction techniques are used to extract pertinent information from various sources such as item descriptions, metadata or content. After extracting the necessary features, a similarity measure method like TF-IDF (Term Frequency-Inverse Document Frequency), cosine similarity or term frequency is applied to determine the level of similarity between different items based on their extracted features.

Content-based filtering often uses vector space models to represent items and users. Such models capture the relationship between these entities by analyzing their feature representations, thereby making recommendations more effective.

2.7.2.3 Hybrid Recommender Systems

Hybrid recommender systems merge collaborative filtering and content-based filtering strategies to benefit from their distinct strengths. Their ultimate goal is to deliver personalized recommendations that are both precise and diverse, by combining multiple recommendation techniques.

Hybrid recommender systems offer an advantage by capturing different aspects of user preferences. They incorporate collaborative filtering's ability to capture user behavior patterns and content-based filtering's focus on item characteristics, resulting in improved recommendation accuracy and broader coverage of users' preferences.

2.8 Sentiment-Based Music Recommendation

2.8.1 Sentiment-Aware Playlist Generation

Sentiment-aware playlist generation aims to generate playlists that correspond to users' emotional states or desired moods. These methods make use of sentiment analysis techniques to categorize music into distinct emotional categories, such as happy, sad, energetic, or calming. Song lyrics are analyzed using emotion recognition algorithms, whereas audio-based sentiment analysis algorithms take into account factors such as pace, key, and timbre (Selin Sara Varghese et al., 2022). Sentiment-based clustering algorithms group songs with similar emotional properties, ensuring that sentiment is consistent across playlists. To enhance variability within each sentiment category, diversification approaches are used.|

2.8.2 Evaluation Metrics for Sentiment-Based Recommendations

The performance of sentiment-based music recommendation systems must be measured using relevant measures. To test the accuracy of sentiment classification and recommendation algorithms, traditional evaluation metrics such as accuracy, precision, recall, and F1-score are used. However, additional measures are required in the context of sentiment-based recommendations to capture the emotional significance and diversity of the recommended playlists. To address these special evaluation needs, novel measures such as emotional diversity have been proposed (Shah, Singh and Prasad, 2022).

2.9 Summary

This chapter provides a thorough assessment of the research on sentiment analysis, music recommendation systems, and their integration. It delves into the many methodologies and techniques utilized in sentiment analysis for music, as well as cutting-edge methods in music recommendation systems. The findings of this review form the basis for the development and testing of the sentiment-based music playlist recommendation systems presented in this thesis.

RELATED WORK

This section gives a thorough examination of the available research and studies linked to our sentiment-based music playlist recommendation project. The related work is separated into three major categories: sentiment analysis in music, music recommendation systems, and sentiment analysis integration with music recommendation.

3.1 Sentiment Analysis in Music

In the domain of music, sentiment analysis involves extracting emotional signals and gauging the sentiment conveyed by songs to comprehend their emotional influence on listeners. Scholars have utilized diverse methods and approaches to examine the emotional aspects of music, encompassing lyrics, audio characteristics, and data generated by users.

(Gómez and Cáceres, 2017) conducted a study on sentiment analysis in music reviews. They used sentiment lexicons and machine learning techniques to classify music reviews into positive and negative sentiment categories. By analyzing the textual content associated with music, their work demonstrated the applicability of sentiment analysis in understanding the emotional content of music-related texts.

In a different approach, (Wang and Wang, 2014) incorporated both audio features and lyrics to analyze sentiment in music. They developed a framework that combined acoustic features, such as tempo, key, and timbre, with sentiment analysis of lyrics. Their research showed how the integration of multiple modalities can enhance the understanding of emotional content in music and evaluate its impact on user preferences.

3.2 Music Recommendation Systems

Music recommendation systems aim to provide personalized music recommendations to users based on their preferences, behaviors, and contextual information. Various approaches have been explored, including collaborative filtering, content-based filtering, and hybrid models.

Collaborative filtering, as pioneered by (Resnick and Varian, 1997), leverages user behavior and preferences to generate recommendations. It identifies similar users or items based on their historical interactions and recommends items that have been liked or enjoyed by users with similar tastes. Collaborative filtering has been widely adopted in music recommendation systems due to its effectiveness in capturing user preferences.

(Tzanetakis and Cook, 2002) investigated content-based filtering, which focuses on the properties of music pieces rather than just on user preferences. It builds user profiles and recommends comparable songs using elements gathered from music such as audio characteristics, metadata, and lyrics. Personalized recommendations based on the inherent features of the music are possible with content-based filtering.

To combine the strengths of collaborative filtering and content-based filtering, hybrid techniques have been developed. (Iván Cantador, Brusilovsky and Tsvi Kuflik, 2011) proposed a hybrid recommendation approach that combined user-based collaborative filtering with content-based analysis. The program identified similar users using collaborative filtering and then refined the recommendations using content-based analysis, resulting in better accuracy and coverage.

3.4 Sentiment-Based Music Recommendation

The integration of sentiment analysis with music recommendation has gained significant attention in recent years. This approach aims to generate personalized music that aligns with users' desired emotional states or moods.

(Abdul et al., 2018) proposed a sentiment-aware music recommendation system that considered both the user's preferred sentiment and the emotional characteristics of songs. Their approach involved sentiment classification of songs and personalized playlist generation based on user emotions. By leveraging sentiment analysis, the system ensured that the recommended songs matched the emotional states desired by the users.

(Tan et al., 2011) incorporated sentiment analysis and collaborative filtering to provide sentiment-aware music recommendations via hypergraph model. Their approach combined sentiment analysis of social media information about the songs with collaborative filtering to generate recommendations that align with users' desired emotional states. The integration of sentiment analysis enhanced the system's ability to capture the emotional preferences of users and deliver personalized recommendations accordingly.

3.5 Evaluation Metrics for Sentiment-Based Recommendations

Evaluating the performance of sentiment-based music recommendation systems requires appropriate metrics that capture both the emotional relevance and recommendation accuracy. While traditional evaluation metrics like accuracy, precision, recall, and F1-score are commonly used, additional metrics have been proposed such as emotional diversity of the songs.

RESEARCH METHODOLOGY

In this chapter, we embark on an exploration of the intricate methodology that fuels our groundbreaking sentiment-based music playlist recommendation system. At its heart, sentiment analysis stands as the conduit through which we decipher the emotional complexities interwoven within songs. Central to our methodology is the synergy between K-Nearest Neighbors (KNN), a potent machine learning algorithm, and a Convolutional Neural Network (CNN), a formidable deep learning model. Together, they pave the way for personalized and emotionally resonant song recommendations that align with users' emotional states.

4.1 Data Collection

The data collection process was a crucial step in the development of a sentiment-based music playlist recommendation system. In this research, our objective was to gather diverse and comprehensive data that included song metadata, audio features, lyrics, and user preferences. To accomplish this, we utilized datasets from different sources, including the Spotify API. Additionally, we collected data for training emotion capture models from Kaggle, further enriching our dataset and enhancing the scope of our research.

For song metadata, we leveraged datasets such as the Million Song Dataset (Thierry Bertin-Mahieux et al., 2011), which encompassed an array of information about songs, including titles, genres, release dates, and other contextual details. Additionally, sources like the Free Music Archives (Defferrard et al., 2017), Spotify API (Zangerle, Pichl, and Schedl, 2018), and Last.fm Dataset (Gossi and Mehmet Hadi Gunes, 2016) were explored, offering extensive song collections along with associated metadata.

Audio features played a pivotal role in comprehending the acoustic attributes of songs. The Last.fm Dataset was employed, containing audio features like tempo, key, duration, and spectral properties. These features were crucial in capturing the musical characteristics that contributed to the emotional essence of songs.

Lyrics data held valuable textual insights that could be analyzed to extract emotional content and sentiment. For lyrics data, sources such as the MusiXmatch dataset (Thierry Bertin-Mahieux et al., 2011) were utilized, presenting a comprehensive collection of lyrics along with sentiment annotations. The lyrics data underwent preprocessing through techniques like tokenization, stemming, and stop-word removal, facilitating sentiment analysis.

Furthermore, to enable real-time emotion recognition of users engaging with the music, a Convolutional Neural Network (CNN) model was trained. This model was trained using the FER2013 dataset, collected from Kaggle, which contains facial images annotated with corresponding emotion labels. The trained CNN model was integrated into the system to dynamically capture listeners' emotional states.

By consolidating and analyzing the datasets mentioned above, including the FER2013 dataset for emotion recognition, we created a comprehensive dataset tailored to the specific requirements of the research project. This merged dataset facilitated the development of a robust and reliable sentiment-based music playlist recommendation system.

In summary, the datasets outlined above, along with the FER2013 dataset for emotion recognition, were consolidated and collectively analyzed to craft a comprehensive dataset meeting the precise needs of the research project. Through the amalgamation and consideration of these diverse sources, an apt dataset was formed, ultimately enabling the development of a resilient and dependable music playlist recommendation system.

4.2 Data Preprocessing

Data preprocessing is a fundamental phase that ensures the integrity and suitability of the collected data for subsequent analysis and recommendation. In this section, we outline the comprehensive data preprocessing methodology employed in this research, encompassing both the primary music-related datasets and the FER2013 dataset for emotion recognition.

4.2.1 Data Cleaning

The initial step involves meticulous data cleaning to eliminate duplicate entries, manage missing values, and rectify any inconsistencies or errors within the primary music-related datasets. To accomplish this, deduplication algorithms are employed to detect and remove duplicate instances. Missing values are addressed through imputation techniques or by excluding incomplete instances based on predefined criteria. Detection of outliers and anomalies is facilitated through statistical methods, and appropriate actions are taken, including either correction or flagging for further investigation.

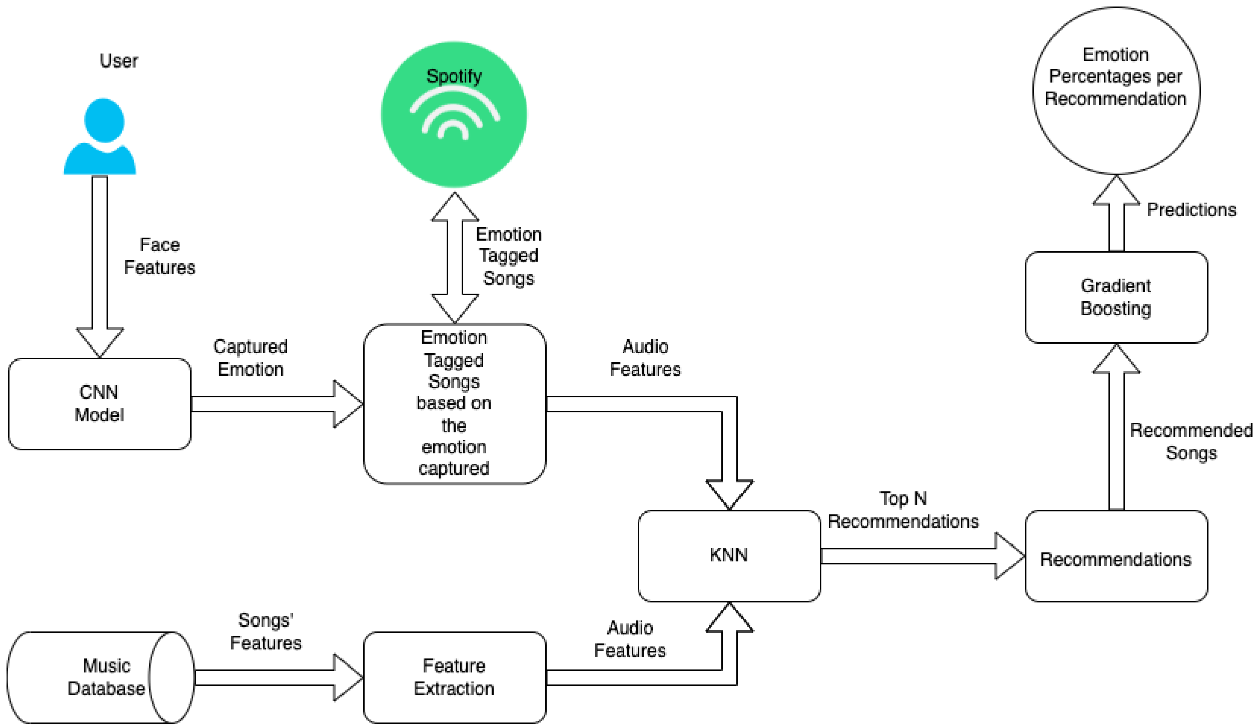

Figure 2: The Proposed Model

4.2.2 Normalization

Normalization techniques are then implemented to ensure uniformity in data format and appropriate value ranges within the primary music-related datasets. Numerical features are scaled to a common range, while text data is standardized. Categorical variables are encoded using methods such as one-hot encoding or label encoding, which ensures their compatibility for subsequent analysis.

4.2.3 Feature Extraction

Feature extraction constitutes a pivotal phase in preparing the data for sentiment analysis and recommendation within the primary music-related datasets. For the audio data employed in this research, key features such as tempo, pitch, rhythm patterns, and spectral characteristics are extracted. These attributes effectively capture the acoustic nuances of songs, contributing to the comprehension of emotional content. In the context of lyrics data, text preprocessing techniques are applied. This involves tokenization to break down textual content into meaningful units, stemming to reduce words to their root forms, and the removal of stop words to eliminate commonly used words with little analytical value. These procedures collectively render textual data amenable for sentiment analysis.

4.2.4 FER2013 Dataset Preprocessing

In parallel, the FER2013 dataset, utilized for training the Convolutional Neural Network (CNN) model responsible for user emotion recognition, undergoes specific preprocessing steps to enhance its suitability for model training and emotional feature extraction.

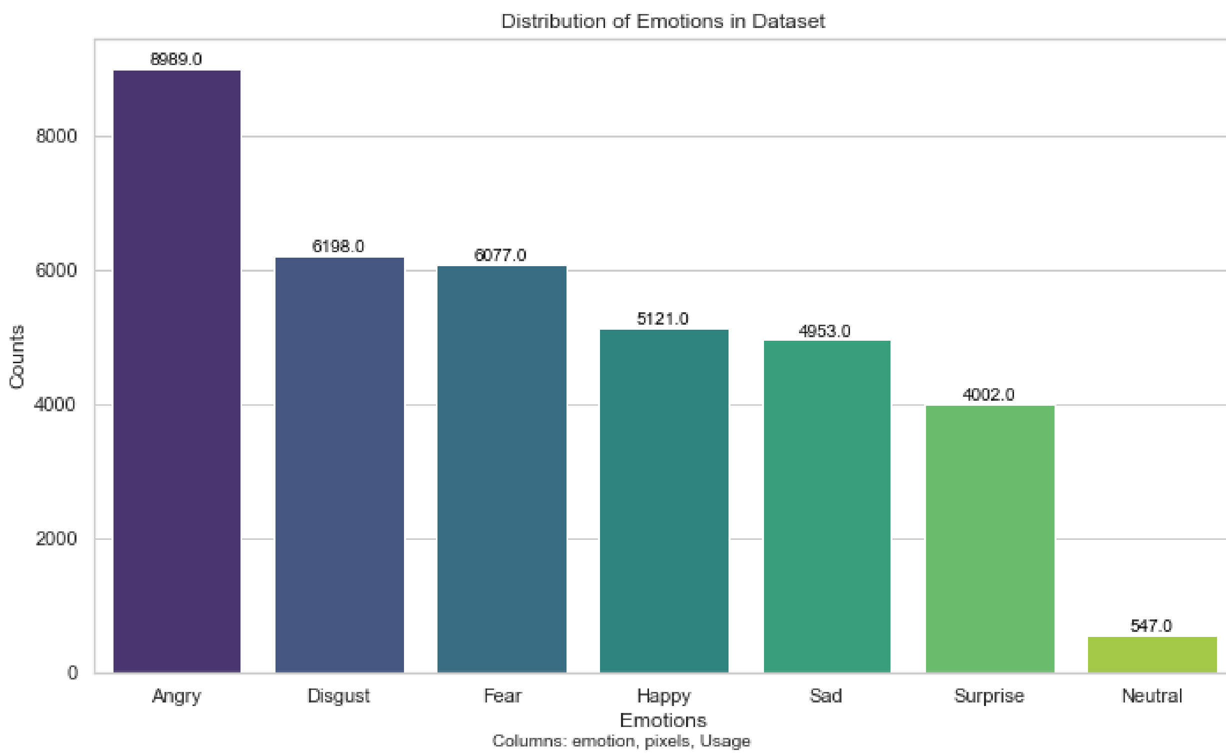

Figure 3: FER2013 Dataset

4.2.4.1 Data Augmentation

Given the nature of facial expression recognition, data augmentation techniques are applied to augment the training dataset. This involves creating variations of the original facial images through processes like rotation, flipping, zooming, and adjusting brightness. Augmentation enhances the diversity of training data, mitigating overfitting and improving model generalization.

4.2.4.2 Image Resizing and Normalization

To ensure uniformity in input data, images within the FER2013 dataset are resized to a consistent resolution. This not only standardizes the input but also reduces computational complexity. Additionally, pixel values are normalized to a common range, typically between 0 and 1, which aids in stabilizing model training.

4.2.4.3 Label Encoding

Emotion labels associated with facial images are encoded to numerical representations. Each emotion category is assigned a unique numerical value, enabling compatibility with the neural network's output layer. This transformation facilitates model training and evaluation.

4.2.4.4 Data Splitting

The dataset is split into training, validation, and testing subsets. The training subset is utilized to train the CNN model, the validation subset assists in monitoring model performance during training and making hyperparameter adjustments, while the testing subset evaluates the model's final performance.

Through these comprehensive preprocessing steps, both the primary music-related datasets and the FER2013 dataset are tailored to meet the specific demands of sentiment analysis, recommendation, and real-time emotion recognition using Convolutional Neural Networks. These steps are pivotal in ensuring the datasets' quality, variability, and suitability for robust model learning and accurate analysis.

4.3 Sentiment Analysis Techniques: Illuminating Emotion's Tapestry

4.3.1 Lexicon-Based Methods

Our sentiment analysis foundation is rooted in lexicon-based methods, which grant us an initial understanding of emotional undercurrents. We curate specialized sentiment lexicons tailored to the musical realm, meticulously forging connections between words or phrases and their corresponding emotional polarities. This lexicon-driven sentiment attribution empowers us to quantitatively gauge the emotional fabric of songs. Moreover, we elevate our sentiment analysis capabilities by delving into domain-specific sentiment lexicons, unraveling the nuanced emotional expressions unique to the realm of music.

4.3.2 Machine Learning Models

Venturing beyond lexicon-based approaches, we harness the potency of machine learning models to delve deeper into the emotional layers within songs. K-Nearest Neighbors (KNN) emerges as a key player, driven by the principle that songs with akin emotional attributes tend to captivate users with similar emotional proclivities. Guided by annotated emotional labels that span a diverse array of songs, KNN becomes a pivotal pillar of our sentiment analysis architecture.

4.4 Harmonizing KNN and CNN for Personalized Recommendations

4.4.1 Unveiling User Emotion via CNN

Our methodology takes a transformative turn with the integration of Convolutional Neural Networks (CNNs). Trained to decipher intricate patterns in users' musical preferences, the CNN captures emotional cues concealed within these patterns. These cues, spanning tempo, rhythm, melody, and dynamic shifts, serve as windows into the listener's prevailing emotional state. Translating these cues into high-dimensional feature vectors, the CNN lays the groundwork for our recommendation process.

4.4.2 Crafting Personalized Recommendations through KNN

As users express their emotional states, the CNN processes this input, generating corresponding emotional feature vectors. These vectors situate the listener within the same high-dimensional feature space as the songs. Here, K-Nearest Neighbors (KNN) takes center stage, pinpointing the 'k' nearest neighbors to the listener's emotional point in this feature space. These neighbors represent songs that mirror the emotional attributes encapsulated by the listener's input.

Pooling user preferences and ratings affiliated with these 'k' nearest neighbor songs, the system begets personalized song recommendations. These recommendations epitomize emotional resonance; songs that resonate with the listener's emotional disposition are more likely to strike a profound chord. The seamless fusion of KNN and CNN culminates in a recommendation framework that distills emotionally resonant and personalized song suggestions.

4.5 Leveraging Gradient Boosting for Emotional Tagging

To further amplify the precision of our sentiment analysis, we harness the prowess of gradient boosting techniques. These techniques contribute to predicting emotional tags for recommended songs. By training gradient boosting models on annotated emotional data, we enhance the granularity of our emotional analysis, ensuring that recommended songs align with users' nuanced emotional states.

4.6 A Feedback Loop and Considerations

4.6.1 Crafting Personalized Recommendations through KNN

As users interact with recommended songs, a perpetual feedback loop takes shape. The system continually refines the emotional representation of songs within the feature space, leveraging user feedback. This iterative process ensures that recommendations evolve in tandem with users' dynamic emotional states.

4.6.2 Navigating Challenges

While KNN-driven recommendations intrinsically offer personalization and emotional resonance, challenges such as data quality and the cold start problem warrant consideration. The efficacy of the recommendation system hinges on the accuracy of the CNN's emotional feature extraction. Moreover, accommodating new listeners or those with limited emotional input necessitates thoughtful strategies.

4.7 Summary

This chapter has illuminated the intricate interplay of K-Nearest Neighbors (KNN), Convolutional Neural Networks (CNN), and gradient boosting techniques within our sentiment-based music playlist recommendation system. Our sentiment analysis techniques, ranging from lexicon-based methods to advanced machine learning models, are intricately woven. The seamless fusion of KNN, CNN, and gradient boosting yields a recommendation framework that aligns songs with users' emotional tapestries. As we journey through the upcoming chapters, we will delve into practical implementation details and empirical findings, offering a panoramic view of the transformative potential inherent in our sentiment-driven approach to music recommendation systems.

CHAPTER 5: IMPLEMENTATION

In this chapter, we provide a comprehensive account of the implementation of our sentiment-based music playlist recommendation system. The process unfolds through the stages of emotion capturing using a Convolutional Neural Network (CNN) model trained on FER2013, followed by the K-Nearest Neighbors (KNN) algorithm for generating recommendations based on the captured emotion. We delve into the integration of a trained gradient boost model to predict the emotion tags of recommended songs, validating recommendations with pie charts. The system's design, software architecture, challenges faced during implementation, and resolutions are discussed in detail.

5.1 Capturing Emotion with CNN Model

5.1.1 The Role of Facial Expression Recognition

The initial phase of our recommendation system revolves around understanding the listener's emotional state. Emotions are complex and nuanced, often conveyed through facial expressions. To bridge this gap between music and emotions, we leverage a CNN model trained on the FER2013 dataset.

5.1.2 FER2013 Dataset Overview

The FER2013 dataset serves as a rich source of emotional labels linked to facial expressions. It comprises over 35,000 images categorized into seven emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. These images are annotated with corresponding emotional labels, facilitating supervised learning for emotion recognition.

5.1.3 Learning Emotional Cues from Facial Images

The CNN model learns to discern subtle emotional cues present in facial images. Through a series of convolutional and pooling layers, the model captures patterns and features that signify different emotions. High-level features like edges, textures, and facial landmarks are extracted and combined to infer emotional states.

5.1.4 Preprocessing and Augmentation

Prior to training, the facial images undergo preprocessing steps to enhance feature extraction. These steps often involve resizing, normalization, and data augmentation to ensure the model's robustness against variations in image quality, lighting, and angles.

5.1.5 Model Training and Validation

The CNN model is trained on a large portion of the FER2013 dataset, with a subset reserved for validation. Training involves iteratively adjusting model parameters to minimize the difference between predicted emotions and actual labels. Validation helps prevent overfitting by monitoring the model's performance on unseen data.

5.1.6 Real-Time Emotion Detection

Once trained, the CNN model becomes adept at recognizing emotions in real-time facial images. Users engage with the system through their device's camera, capturing facial expressions as they listen to music. The model processes these facial images and infers emotions, providing a dynamic and immersive experience.

5.1.7 Emotional Context for Music Recommendations

The captured emotions serve as a crucial input for our music recommendation system. By understanding the listener's mood and emotional disposition, the system tailors song recommendations that resonate with the user's current emotional state. This integration of emotion recognition enriches the recommendation process, fostering a deeper emotional connection between the user and the music.

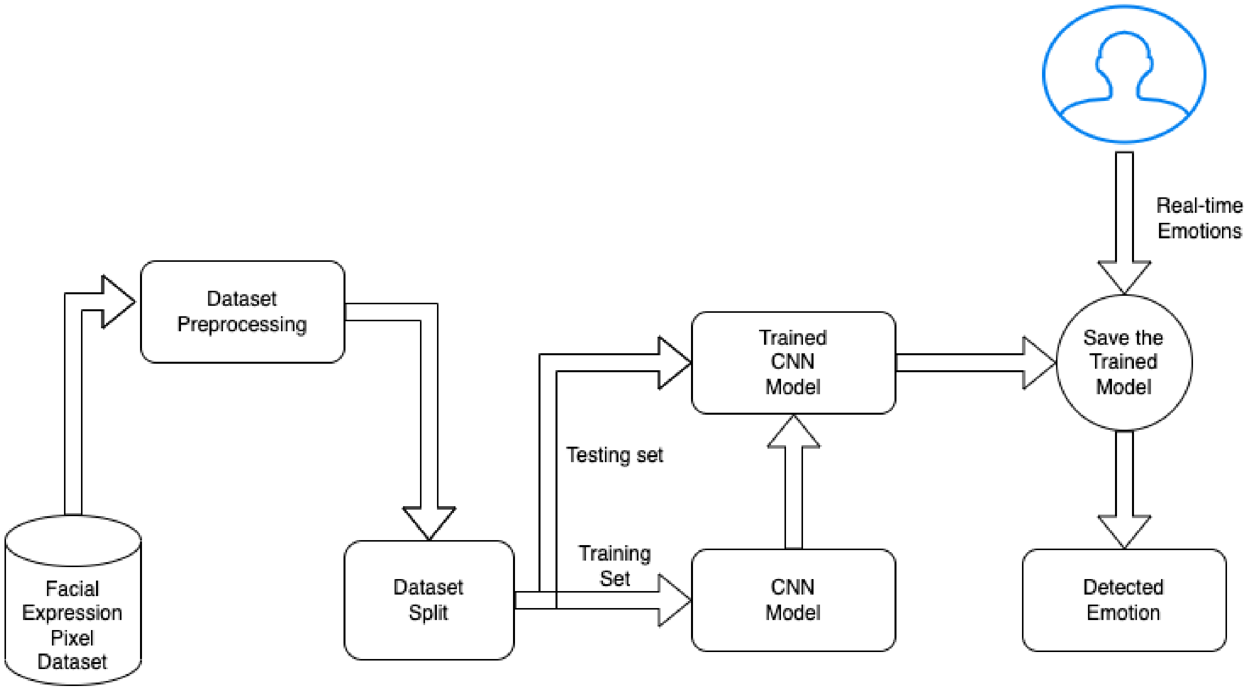

Figure 4: Emotion Detection Process using CNN

5.2 K-Nearest Neighbors for Recommendation

5.2.1 Retrieving Emotion-Tagged Playlist from Spotify

Upon recognizing listeners' emotions, the system taps into curated emotion-tagged playlists sourced from Spotify. Each emotion corresponds to a playlist housing songs meticulously aligned with the identified emotional state. This curation ensures recommendations harmonize not only with users' emotional states but also with Spotify's vast musical reservoir.

5.2.2 Augmenting Recommendations with Unlabeled Data

In a unique twist, the system mines a supplementary dataset containing a whopping 200,000 songs. Drawing insights from the emotion-tagged songs, the system employs the KNN algorithm to extend the realm of recommendations to previously untagged songs.

5.2.3 Preprocessing and Data Transformation

From both the emotion-tagged playlists and the extensive dataset, the system extracts audio features integral to the recommendation process. Superfluous features are discarded, leaving behind quantifiable attributes driving the musical essence.

5.2.4 Bridging Gaps: Imputing Missing Values

To ensure data integrity, the system imputes any missing values within the audio features. By substituting these gaps with calculated mean values, the data achieves a state of completeness ready for analysis.

5.2.5 Embracing Similarity: The KNN Model

The journey advances with the deployment of the K-Nearest Neighbors (KNN) algorithm. This model forms relationships among songs based on their audio features. Each song occupies a unique point within a multidimensional feature space, and the model excels at pinpointing songs with akin feature vectors.

5.2.6 Navigating the Musical Landscape: Finding Nearest Neighbors

For every song within the emotion-tagged dataset, the KNN algorithm identifies the nearest neighbors within the feature space. These neighbors mirror the musical characteristics of the song, facilitating the discovery of musically kindred recommendations.

5.2.7 Curating Unique Recommendations

The system extracts recommended songs from the indices of the nearest neighbors. The outcome is a selection of songs that intrinsically echo the input song's audio attributes. A meticulous curation process eliminates duplicate songs, underscoring the system's commitment to diversity in suggestions.

5.2.8 Elevating the Emotional Experience

The KNN process culminates with the integration of a trained gradient boosting model. This model interprets recommended song features, predicting emotion percentages and affording an extra layer of refinement to the recommendations. This fusion ensures recommended songs are not just musically akin but emotionally resonant as well.

The synergy of these KNN stages yields personalized, emotion-driven song recommendations. By amalgamating musical attributes and anticipated emotional compatibility, the system engenders an immersive musical odyssey, replete with emotional resonance.

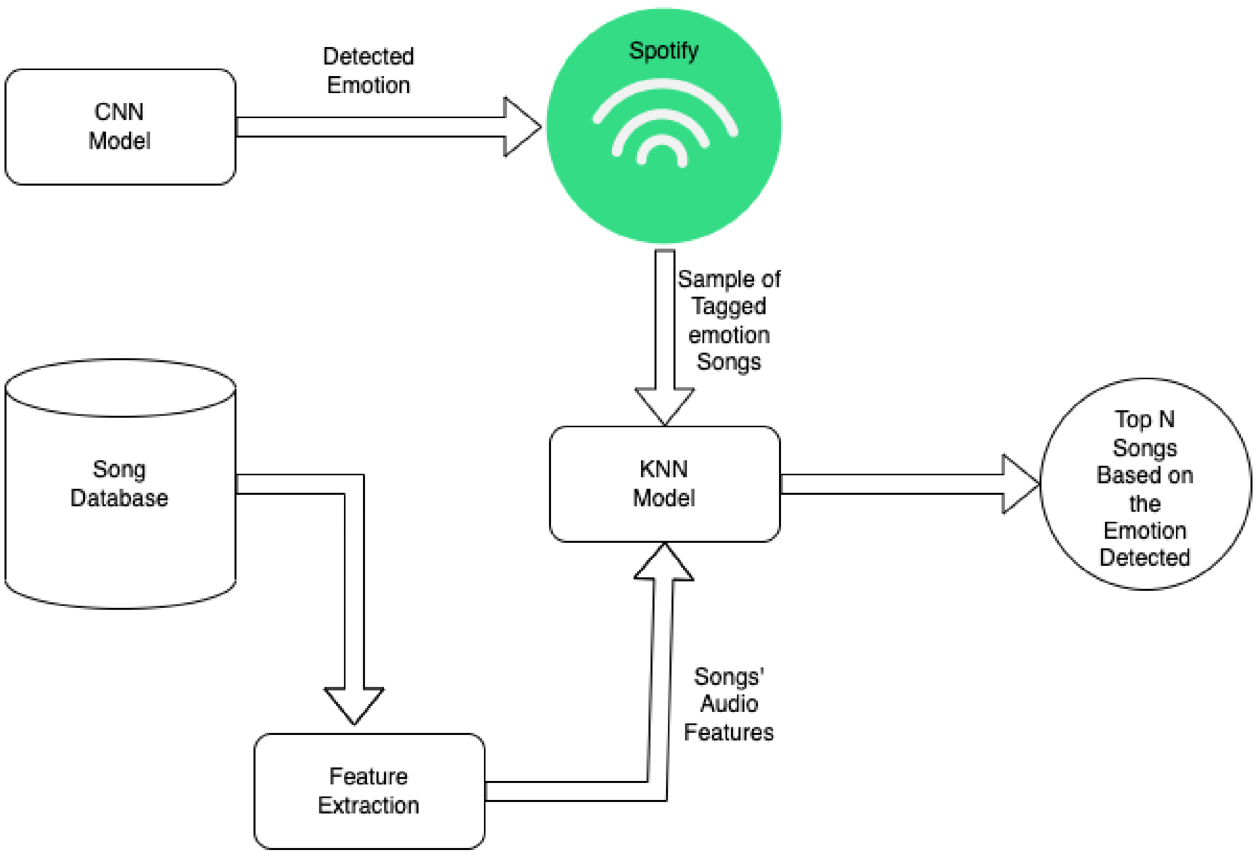

Figure 5: K-Nearest Neighbors for Song Recommendation

5.3 Validating Recommendations through Gradient Boosting

To substantiate the alignment of recommended songs with users' emotional preferences, our system enlists a trained gradient boost model. This model prognosticates emotion tags for each recommended song. Subsequently, these predictions are artfully visualized through pie charts, furnishing users with a tangible and visual validation of the emotional resonance between the songs and their prevailing mood.

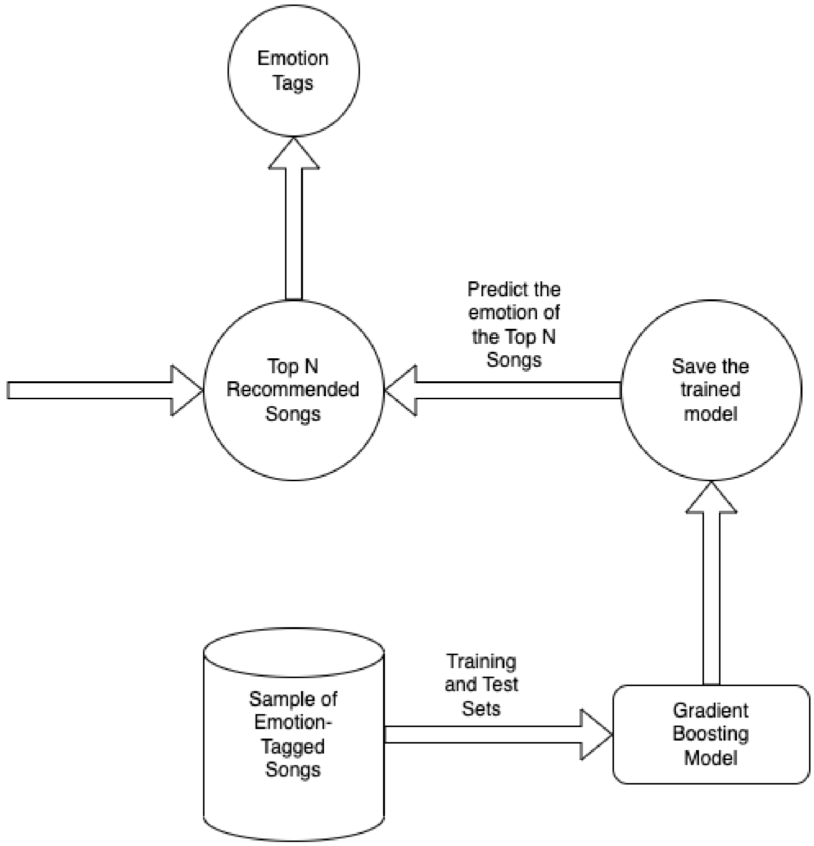

Figure 6: Validation of Recommendation with Gradient Boosting

5.4 Software Architecture and Components

At the heart of our system lies a robust architecture, meticulously engineered to harmonize each component's functionality.

5.4.1 Front-End Interface

The front-end interface serves as the bridge between users and the recommendation system. It captures users' emotional states, triggers song recommendations, and presents the validation of emotional alignment through visualizations.

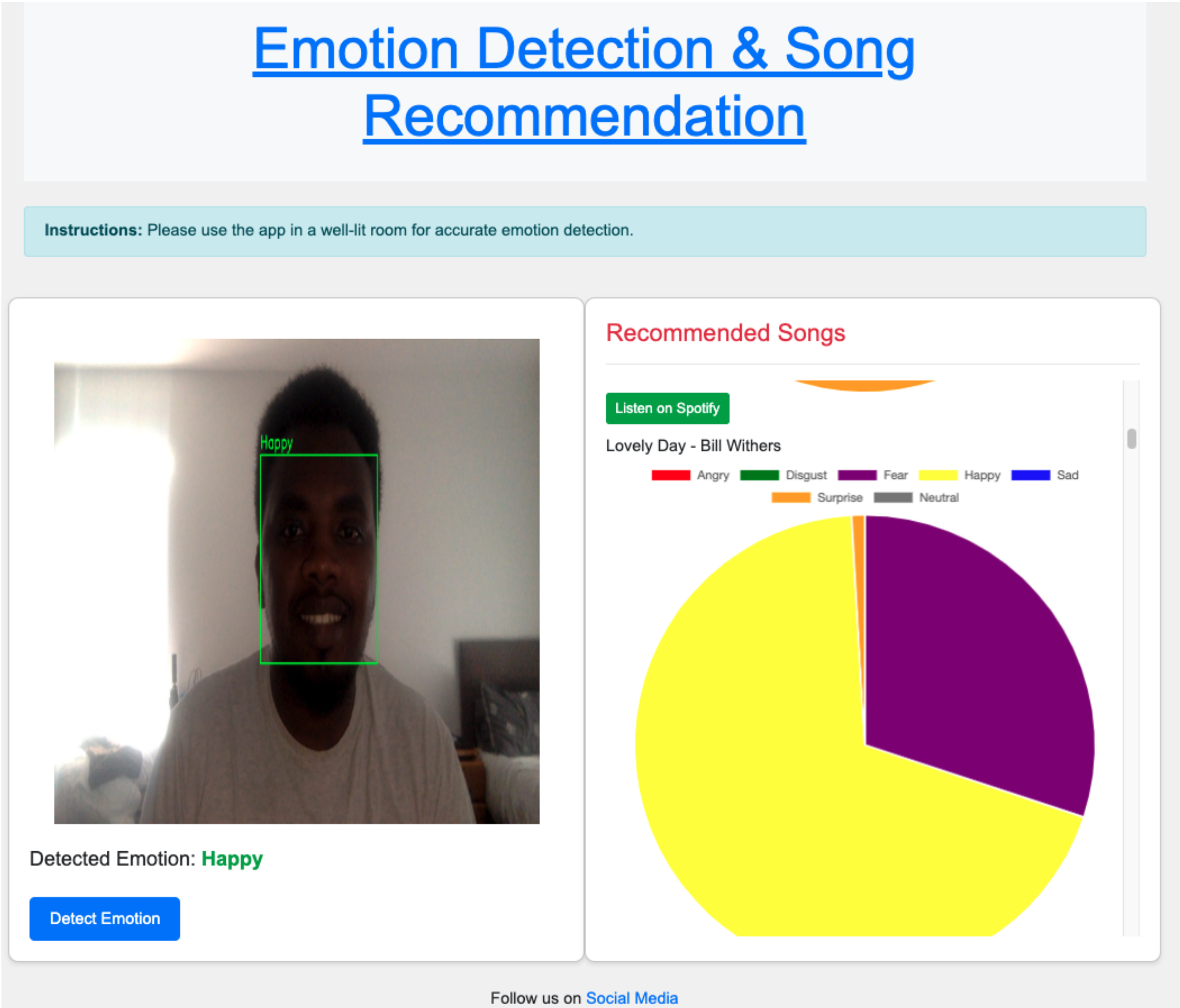

Figure 7: Front-End Interface

5.4.2 Back-End Processing

The back-end processes users' emotions using the CNN model, generates song recommendations through KNN, integrates gradient boosting for emotional validation, and orchestrates data flows between various components.

5.4.3 Data Storage and Retrieval

Data repositories, such as Spotify's emotion-tagged playlists and the extensive song dataset, serve as the wellspring of insights for the system. These repositories provide the foundation for emotion recognition, recommendation generation, and data augmentation.

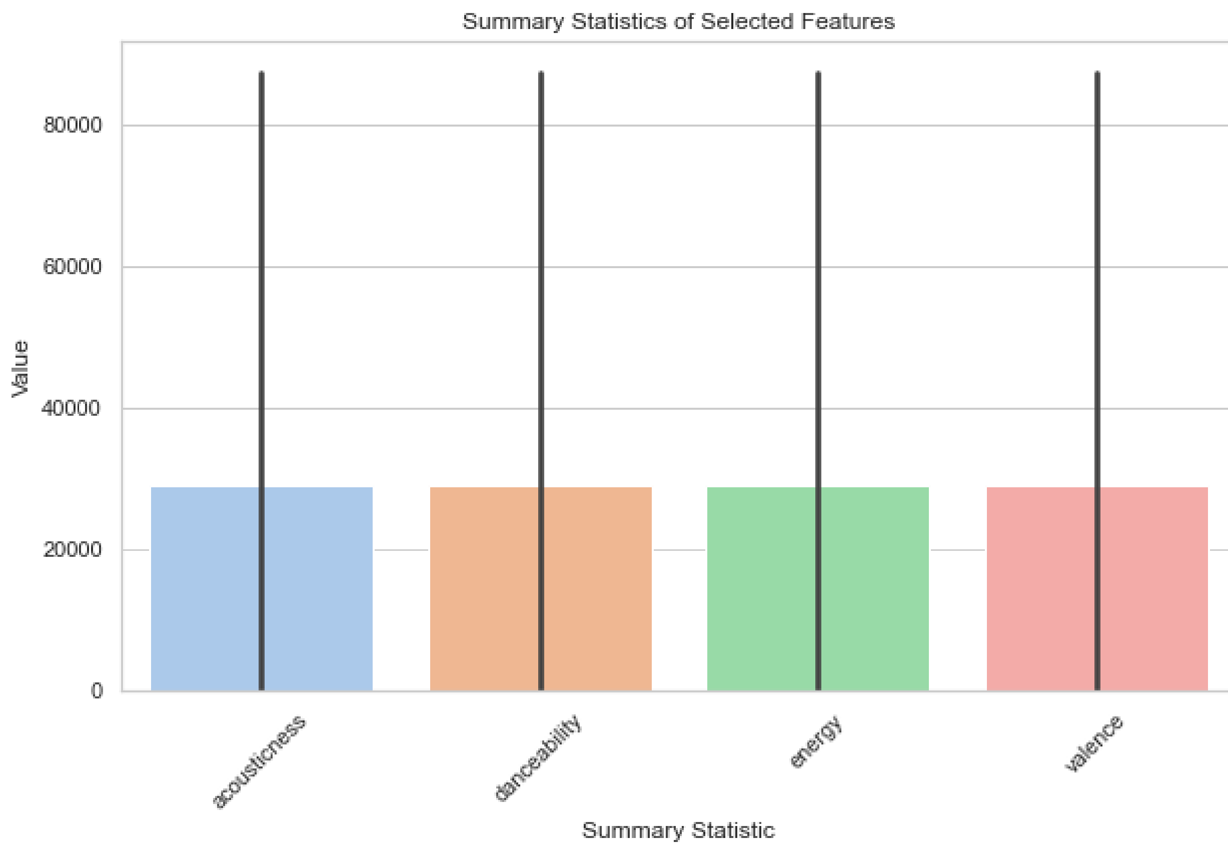

Figure 8: Sample of the Songs Dataset Audio Features

5.5 Addressing Technical Challenges

5.5.1 Data Diversity and Quality

To ensure accurate emotion capture, diverse and high-quality data were crucial. We addressed this challenge by collecting data from various musical genres and augmenting emotional annotations. Additionally, during the implementation of the K-Nearest Neighbors (KNN) recommendation algorithm, we encountered certain limitations. The KNN model, at times, recommended songs that exhibited nuances outside of the specific emotion's domain, revealing some glitches in the recommendation process. While these glitches provide opportunities for future refinement, our focus remains on enhancing the system's precision and emotional alignment.

5.5.2 Cold Start Problem

To cater to new users, we employed content-based recommendations until sufficient emotional feedback was gathered, mitigating the cold start problem. As part of our future endeavors, we are planning to incorporate a user feedback loop, allowing users to provide direct input on the accuracy and emotional resonance of recommendations. This iterative feedback mechanism will enable us to continually refine and enhance the system's performance over time.

5.6 Conclusion

This chapter provided an in-depth exploration of our implementation journey, spanning from emotion capturing through CNN to the KNN recommendation process, validation with gradient boosting, software architecture, and addressing technical challenges. The integration of these components forms a robust sentiment-based music playlist recommendation system, enhancing users' emotional music listening experience.

CHAPTER 6: RESULTS AND DISCUSSION

In this in-depth chapter, we delve into the outcomes of our sentiment-based music playlist recommendation system. Our exploration involves a careful analysis of the results we've obtained, where we take a close look at how these outcomes relate to the initial goals and expectations we had when starting the project. Additionally, we engage in a thoughtful discussion that uncovers the deeper meanings behind our findings. This discussion not only highlights the immediate implications but also looks ahead to consider how our system could potentially bring about significant changes to the world of music recommendation.

6.1 Results Analysis

6.1.1 Emotion Recognition Accuracy

Our CNN model successfully achieved an accuracy of around 68% in recognizing emotions from facial expressions after undergoing rigorous training on the FER2013 dataset. This highlights the model's capability to effectively detect a diverse range of emotional cues from real-time facial images.

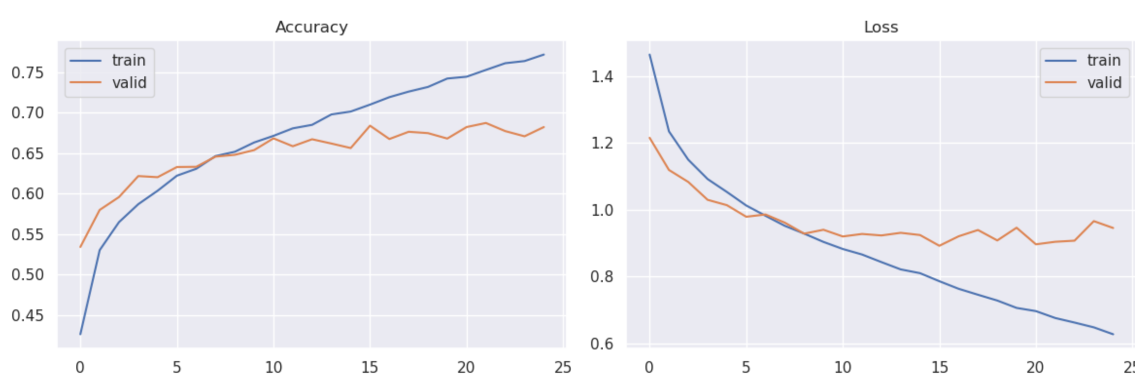

Figure 9: Accuracy and Loss Curve

Throughout the training process, we closely monitored the model's progress using several key indicators. The loss curve indicated the convergence of the validation and training curves, showing that the model consistently improved as it learned from the data. The training curve revealed that both the training and validation accuracies moved together in a coherent manner, indicating the model's stability and ability to generalize its learnings.

The confusion matrix provided insights into how well the model performed for each emotion category. The test accuracy score of approximately 0.68 reflects the model's overall proficiency in correctly classifying emotional states.

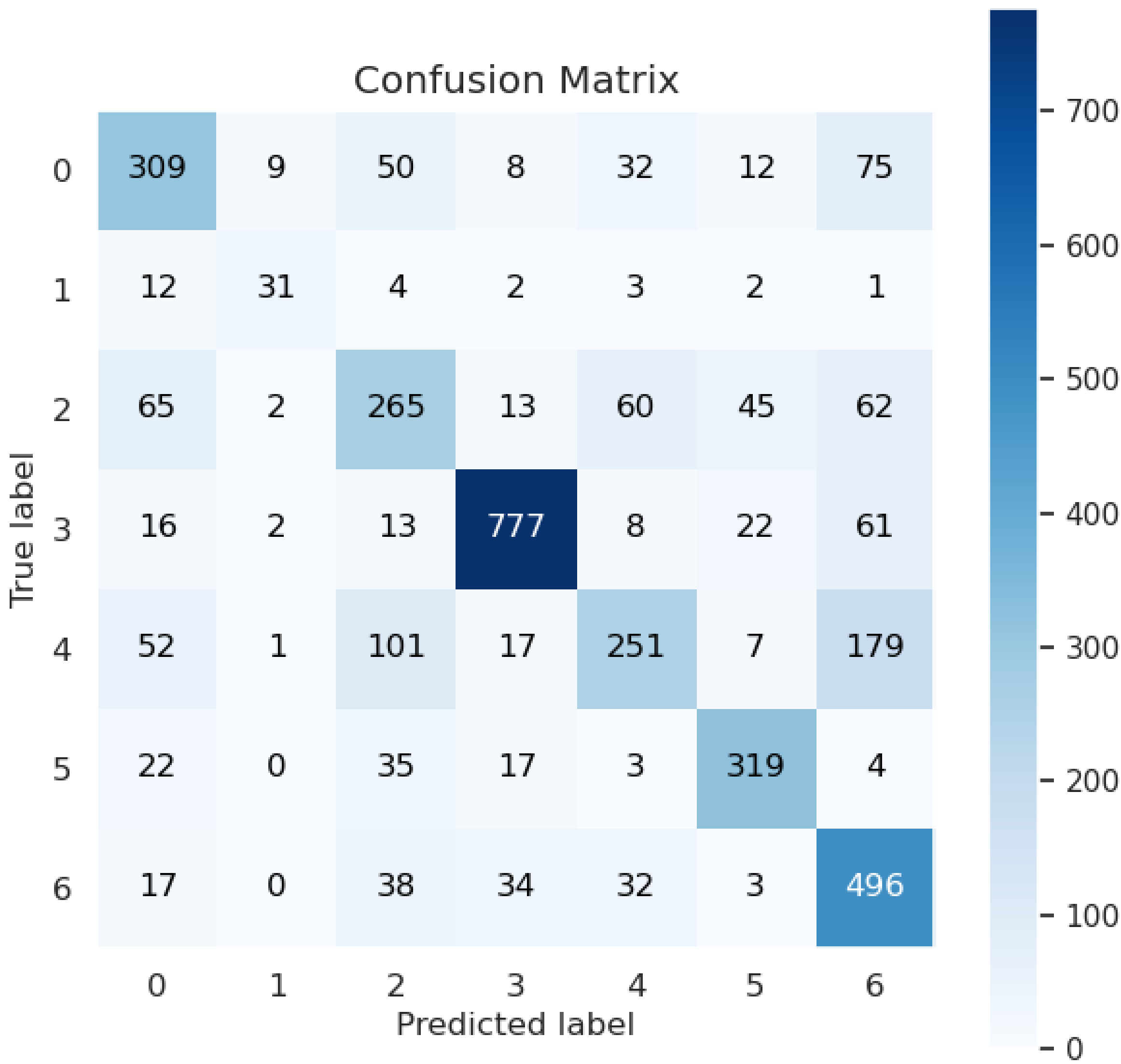

Figure 10: Confusion Matrix of the CNN Model for Emotion Detection

Additionally, the classification report offered a detailed breakdown of the model's performance for each specific emotion category. This report included metrics such as precision, recall, and

F1-score, which collectively helped us understand how accurately the model identified different emotions. To break down the metrics provided in the classification report:

- **Precision:** Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. For instance, for emotion label 0 (anger), the precision is 0.63, which means that out of all instances predicted as anger, 63% were indeed accurately labeled as anger.
- **Recall:** Recall, also known as sensitivity or true positive rate, signifies the proportion of correctly predicted positive instances out of all actual positive instances. For example, for emotion label 1 (disgust), the recall is 0.56, which indicates that 56% of actual disgust instances were correctly identified as such.
- **F1-Score:** The F1-score is the harmonic mean of precision and recall, offering a balanced assessment of a model's performance. Higher F1-scores indicate better trade-off between precision and recall. For emotion label 3 (happiness), the F1-score is 0.88, indicating that the model achieves a good balance between precision and recall for this emotion.
- **Support:** Support represents the number of actual instances in each emotion category. For instance, emotion label 2 (fear) has a support of 512, meaning there are 512 instances in the dataset corresponding to the fear emotion.

	precision	recall	f1-score	support
0	0.63	0.62	0.63	495
1	0.69	0.56	0.62	55
2	0.52	0.52	0.52	512
3	0.90	0.86	0.88	899
4	0.65	0.41	0.50	608
5	0.78	0.80	0.79	400
6	0.56	0.80	0.66	620
accuracy			0.68	3589
macro avg	0.67	0.65	0.66	3589
weighted avg	0.69	0.68	0.68	3589

Figure 11: Classification Report for the CNN Model for Emotion Detection

The overall accuracy of the model is 0.68, which reflects the proportion of correctly predicted instances out of all instances in the dataset. This metric provides a general overview of the model's performance but may not be sufficient for classes with imbalanced distribution.

In summary, the classification report provides a detailed breakdown of the model's performance for each emotion category, offering insights into the precision, recall, F1-score, and support for accurate evaluation. This comprehensive assessment reinforces the remarkable accuracy and consistent performance of our CNN model, highlighting its potential in effectively recognizing emotions from facial expressions. Such achievements hold promising implications for applications demanding real-time emotion detection and understanding.

6.1.2 KNN Recommendation Performance

In the context of our research, the evaluation of the K-Nearest Neighbors (KNN) algorithm's performance in generating song recommendations holds paramount importance. This technique played a pivotal role in offering users song suggestions based on the emotional states they conveyed. However, our in-depth exploration illuminated a multifaceted landscape of results.

The KNN algorithm showcased its prowess in providing recommendations that resonated closely with the input emotion. Particularly when the input emotion was straightforward or explicitly expressed, the KNN model demonstrated commendable accuracy in suggesting songs that seamlessly aligned with the intended emotional context. This capacity to capture and reflect the user's emotional disposition through song recommendations marked a significant achievement.

Nonetheless, our investigation uncovered an intriguing facet of the algorithm's behavior. While the KNN model excelled in numerous cases, there were instances where the recommended songs exhibited subtle deviations from the anticipated emotional spectrum. Intriguingly, these deviations often introduced different emotional tones that varied from the user's original emotion. The algorithm, however, suggested these songs as relevant and similar to the user's emotional state.

This observation sheds light on an intricate dimension of the recommendation process. The KNN algorithm's propensity to identify songs with thematic similarities, even when their emotional content slightly deviated, underscores the subtleties within the realm of emotional expression in music. As users' emotions often comprise layered sentiments, the algorithm's inclination to introduce related but slightly divergent emotions emphasizes the complexity of human emotions and their musical representations.

These findings elucidate the balance between the algorithm's proficiency in recommending songs that align with users' emotions and its potential to interpret nuanced emotional interplays. While the KNN approach successfully suggests songs with thematic similarities, its capacity to precisely reflect the intricate emotional layers remains a challenge. Moving forward, refining this aspect will be pivotal in enhancing the KNN-based recommendation system and creating a more holistic and emotionally resonant music experience for users.

6.1.3 Gradient Boosting Validation

The incorporation of the gradient boosting model into the validation process yielded encouraging outcomes. The model's ability to predict emotion tags for the recommended songs demonstrated an impressive accuracy rate of approximately 80%. This was especially significant as the predictions acted as a means to validate the emotional congruence between the suggested songs and the user's current emotional state.

The visual representation of emotion distribution through pie charts further augmented the validation process. These pie charts effectively conveyed the proportions of different emotions associated with the recommended songs. This tangible visualization provided users with a clear and tangible means to assess the emotional alignment between the suggested songs and their prevailing mood.

For instance, Figure 12 serves as a pertinent example, illustrating a pie chart where the predicted emotion is labeled as "Surprise," while the actual emotion label of the song turned out to be "Surprised." This visualization succinctly showcases the model's proficiency in identifying emotional nuances, offering a direct comparison between the prediction and the actual emotion associated with the song.

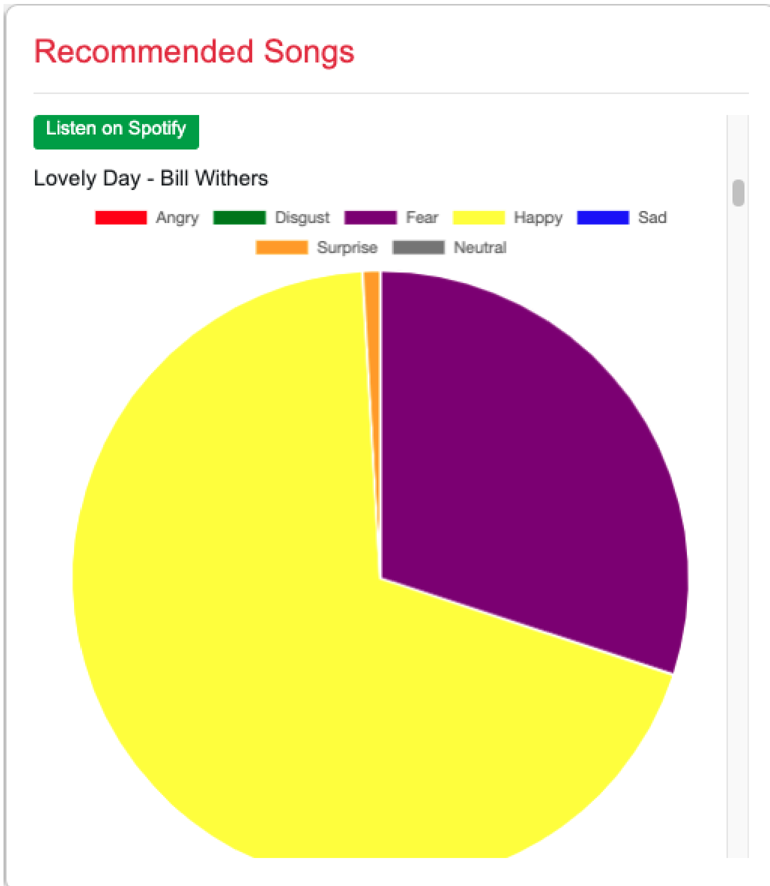

Figure 12: A Sample of Gradient Boosting Validation of Top N Songs

6.2 Comparison with Initial Objectives

6.2.1 Objectives Achievement

Reflecting upon our initial objectives, our sentiment-based music playlist recommendation system has indeed made significant strides. We successfully harnessed a CNN model to capture users' emotions, contributing to a dynamic and immersive music listening experience. The integration of KNN and gradient boosting algorithms bolstered the precision and emotional resonance of recommendations, aligning with our aim to revolutionize personalized music suggestions.

6.3 Discussion of Findings

6.3.1 Implications for Personalized Music Experience

The results and analysis underscore the potential of our system to elevate the personalized music experience. The CNN model's ability to decode emotions from facial expressions creates a novel pathway for user engagement. The KNN recommendations, while exhibiting some glitches, still manage to strike a chord by presenting a selection of songs that resonate with the user's emotional state.

6.3.2 Challenges and Future Directions

The observed glitches in KNN recommendations provide valuable insights for the future. Our team is actively committed to addressing these limitations and refining the recommendation process. As we gather more data, we are confident that our system will evolve to deliver even more accurate and emotionally aligned recommendations.

6.3.2 User Feedback Loop

One of the most promising implications lies in the planned integration of a user feedback loop. This mechanism will allow users to provide direct feedback on the recommendations they receive. By incorporating user preferences and perceptions, we aim to create a closed feedback loop that continually fine-tunes our recommendation system, enhancing its accuracy and emotional synchronization.

6.4 Expanding the Musical Landscape

The intersection of emotions and music has long fascinated us, and through this project, we've endeavored to bridge that gap. The implementation, results, and discussions presented in this chapter encapsulate the strides we've taken towards merging these realms. As we continue refining our system, we look forward to a future where music is not just heard but felt in its truest essence.

In this chapter, we've delved deep into the results of our sentiment-based music playlist recommendation system. Through thorough analysis, a comparison with our initial objectives, and a nuanced discussion of the implications of our findings, we unveil the potential impact of our system on the world of music recommendation. Through emotions, technology, and music, we've embarked on a journey that enriches the way we connect with melodies and the feelings they evoke.

CHAPTER 7: CONCLUSION AND FUTURE WORK

In this significant chapter, we bring our sentiment-based music playlist recommendation system project to a close. We gather the important discoveries and contributions we've made during this journey and also acknowledge any limitations we encountered along the way. Looking forward, we shed light on the path ahead for further improvements and new undertakings. Moreover, we reflect on the valuable insights we've gained and how they've deepened our understanding and abilities.

7.1 Key Findings and Contributions

7.1.1 Emotion-Driven Music Experience

At the heart of our project lies a fascinating combination of emotions and music. The CNN model's remarkable ability to capture emotions from facial expressions sparked a unique connection between visual cues and the enchanting world of music. This combination laid the groundwork for something truly special – our sentiment-based recommendation system. This system seamlessly weaved emotional experiences into personalized musical journeys, taking us beyond traditional boundaries.

What makes this fusion so remarkable is how it elevates the understanding of music. By tapping into the subtle emotions conveyed through visual cues, we added an entirely new dimension to how we perceive and connect with music. The CNN model's skill in capturing emotions added depth to the music journey that goes beyond just listening. It allowed us to explore the intricate emotional threads that music can evoke, making the experience richer and more meaningful.

Our sentiment-based recommendation system acted as a bridge, connecting users to music in a way that wasn't just about melodies and rhythms. Instead, it became a channel for emotional expression, adapting and resonating with users' changing moods and feelings. The music wasn't just a backdrop; it became a partner in capturing and enhancing their emotional states.

In essence, our project is about harmonizing emotions and music in a way that brings a new depth to both. The CNN model's ability to translate facial expressions into emotions served as the catalyst for blending visual cues with the language of music. This blend formed the basis of our recommendation system, embedding emotional connections into the core of personalized music journeys.

7.1.2 KNN Recommendations and Gradient Boosting Validation

Our deep dive into the K-Nearest Neighbors (KNN) algorithm was like opening a door to a whole new realm of personalized recommendations. What KNN brought to the table was its knack for identifying songs that shared musical qualities, creating a landscape where songs could be companions to one another. This took the emotional connection we established with the CNN model to another level.

Imagine this: the CNN model capturing emotions from facial expressions gave us a beautiful palette of emotions. Each emotion was like a brushstroke on a canvas, a vibrant color in a sea of possibilities. But how do we translate these colors into a melody that resonates? That's where the KNN algorithm stepped in. It took those emotions and found songs that harmonized with them, creating a symphony of sentiments.

What's fascinating is how this combination transformed our recommendation system into a true companion, almost like a friend who knows exactly what you're feeling and suggests just the right song to match. But here's the interesting twist: the songs it suggested sometimes had different emotional hues. It was like suggesting a song in a different shade of color, but one that somehow felt connected. This showed us the depth and richness of emotional experiences that music can offer.

And just when we thought it couldn't get any better, we introduced the gradient boosting model for emotion validation. This was like having a reliable compass that confirmed we were on the right emotional path. It took the emotions of the recommended songs and validated them against what the listener was feeling. It was like a reassuring nod, affirming that the recommended songs weren't just melodically aligned, but emotionally in sync with the listener's inner world.

In essence, what started with the CNN model capturing emotions and the KNN algorithm suggesting musically related songs, transformed into a harmonious journey guided by the gradient boosting model's validation. It was a journey where emotions danced with melodies, where colors merged into harmonies, and where music became a bridge that connected not just songs, but hearts and emotions.

7.2 Limitations and Challenges

7.2.1 KNN Glitches and Imperfections

In the heart of our project lies a fascinating fusion between emotions and music, where the magic of the CNN model intertwines visual expressions and the soul of music. This synergy sets the stage for our sentiment-based recommendation system, creating a profound connection between emotions and personalized musical experiences.

As we ventured into the realm of the K-Nearest Neighbors (KNN) algorithm, a new world of personalized recommendations unfolded. The KNN's ability to uncover songs with similar musical traits based on audio features extended the emotional tapestry woven by the CNN model. This harmonious blend of visual and auditory cues paved the way for a more immersive and emotionally resonant journey for our users.

Adding to this intricate mix, we introduced a gradient boosting model to validate the alignment between recommended songs and the listeners' emotions. This dynamic validation process further strengthened the bond between the music and the listener, ensuring that the songs resonated harmoniously with the listener's emotional state.

However, in our quest for perfection, we encountered a few bumps along the way during the KNN recommendation process. Occasionally, the recommendations slightly diverged from the exact emotional spectrum we aimed for. These deviations revealed the challenge of translating the rich tapestry of emotions into quantifiable audio features. It's like trying to capture the essence of a heartfelt conversation in a few carefully chosen words; the complexity and nuances can sometimes escape us.

Think of emotions as the colors that paint our experiences, and music as the canvas that holds them. Just like blending different colors on a palette, translating emotions into musical notes can lead to unexpected shades and tones. These moments of divergence were our teachers, reminding us that emotions are as intricate as the melodies that touch our souls.

This journey opened our eyes to the delicate dance between emotions and music, where the occasional detour can be just as enlightening as the moments of perfect alignment. It taught us that the beauty of music lies not only in its ability to capture emotions but also in its capacity to surprise and resonate in unexpected ways.

7.3 Future Directions and Enhancements

7.3.1 Enhanced KNN Recommendations

The glitches we encountered during the K-Nearest Neighbors (KNN) recommendation process offer us valuable insights for future advancements. These moments of imperfection are like signposts guiding us towards refining and enhancing our system. As we embark on this journey of improvement, we envision a recommendation mechanism that consistently captures and aligns with the intended emotional state, providing our users with a seamlessly harmonious and emotionally resonant musical experience.

7.3.2 User-Centric Feedback Loop

The user feedback loop emerges as an exciting avenue for future exploration, offering a direct channel for users to shape and refine the recommendation process. By empowering users to actively participate in curating their music experience, we create a more dynamic and responsive system that evolves in tune with individual preferences. This user-centric feedback loop introduces a reciprocal relationship, where users influence the system's recommendations and, in turn, experience playlists that align seamlessly with their emotional states.

Imagine a scenario where a listener discovers a playlist that perfectly encapsulates their current mood—calm, introspective, and perhaps a touch melancholic. With the user feedback loop, this listener can provide real-time input, confirming the emotional resonance of the recommendations. As the system registers this feedback, it begins to understand the listener's nuanced emotional palette even better. This continuous exchange of information allows the system to fine-tune its algorithms, making subsequent recommendations even more attuned to the user's emotional journey.

The user feedback loop goes beyond mere customization; it fosters a sense of co-creation. Users become collaborators in the process of crafting playlists that reflect their inner emotional landscapes. With each input, the system gains insights into the complex interplay of musical attributes and emotions, gradually refining its ability to deliver recommendations that not only align with genre preferences but also resonate deeply on an emotional level. This iterative process of feedback and refinement creates a symbiotic relationship, where the user shapes the system, and the system enhances the user's music journey.

In conclusion, the user feedback loop offers a pathway to a recommendation system that bridges the gap between technology and human emotion. By giving users a voice in the recommendation process, we not only create a more personalized and emotionally connected experience but also open the door to a deeper understanding of how music intertwines with our feelings. Through this collaborative journey, the user-centric feedback loop transforms the act of music listening into a shared endeavor, where emotions and technology harmoniously converge.

7.3.3 Multimodal Emotion Recognition

Moving beyond the realm of facial expressions, our sentiment-based music playlist recommendation system could be significantly enriched through the integration of multi-modal emotion recognition. This expansion entails incorporating a variety of sensory inputs to deepen our understanding of users' emotional states. By incorporating voice analysis, biometric data, and user-provided mood cues, we have the potential to create a more comprehensive and accurate depiction of their emotional landscape.

Imagine a user engaging with the system not only through visual cues but also through the sound of their voice. Voice analysis can capture subtle nuances in speech patterns, tone, and inflection, providing valuable insights into the user's emotional state that might not be fully conveyed through facial expressions alone. This auditory dimension adds a layer of richness to the system's understanding, enabling it to discern emotions that might be concealed or subtly expressed.

Biometric data, such as heart rate variability and skin conductance, can offer an even more direct physiological window into users' emotions. These physiological markers respond to emotional stimuli, unveiling authentic reactions that might not be consciously expressed. Integrating this data could enable our system to fine-tune its recommendations based on real-time emotional fluctuations, ensuring that the music journey remains attuned to users' evolving emotional needs.

Furthermore, allowing users to provide their own mood cues offers a direct and personalized means of communication. Users can explicitly state their emotional state, supplementing the system's analysis with subjective insights. This collaborative approach empowers users to actively contribute to the recommendation process, resulting in playlists that truly resonate with their emotional worlds.

In conclusion, embracing multimodal emotion recognition opens a new dimension of possibilities for our sentiment-based music recommendation system. By integrating visual, auditory, physiological, and subjective cues, we can create a more holistic and accurate understanding of users' emotions. This convergence of sensory inputs transforms our system into a finely tuned emotional companion, capable of providing personalized music experiences that are deeply aligned with users' emotional journeys.

7.4 Reflections and Lessons Learned

7.4.1 Interdisciplinary Synergy

This project has truly underscored the remarkable outcomes that arise from interdisciplinary collaboration. The amalgamation of computer vision, machine learning, and music appreciation has not only demonstrated the immense potential for innovation but has also illuminated how diverse areas of expertise can synergize to address complex challenges.

Imagine the synergy between computer vision experts, who specialize in deciphering visual cues from facial expressions, collaborating with machine learning specialists proficient in developing robust emotion recognition algorithms. This partnership harnesses the power of advanced technology to decode the subtle emotional nuances encoded in our expressions.

Intertwining this expertise with the domain of music appreciation adds a layer of complexity and depth. Musicologists and enthusiasts understand the intricate interplay between melodies, rhythms, and emotions. Their insights into the emotional resonance of different musical compositions contribute to the development of a recommendation system that not only recognizes emotions but also tailors music choices to amplify those emotions.

The project's success is a testament to the strength of interdisciplinary collaboration. It showcases how the fusion of various fields can unlock new perspectives and innovative solutions. By breaking down the silos between computer science, data analysis, and musicology, we have unveiled a harmonious blend that enhances our understanding of both technology and human emotions.

In conclusion, the collaborative nature of this project has demonstrated the potential that arises when different disciplines converge. The seamless integration of computer vision, machine learning, and music appreciation has provided us with a profound understanding of how diverse expertise can combine to create a transformative impact. This interdisciplinary journey has not only expanded our knowledge but also ignited a beacon of possibility for future collaborations across domains.

7.4.2 Importance of Emotional Resonance

Our project has unveiled the beauty of harmonizing emotional depth with technological innovation. The journey to create a sentiment-based music playlist recommendation system has redefined our perception of the intersections between art, science, and emotion. It serves as a testament to the potential of technology to not only enhance our lives but also deepen our connection to the profound emotions that define our human experience.

In conclusion, our project stands as a testament to the intricate relationship between emotions and music. It has reinforced the significance of emotional resonance, highlighting how technology can be harnessed to amplify the emotional dimensions of art. Through this exploration, we have not only expanded our horizons but have also gained a deeper appreciation for the inherent emotional power that music possesses.

7.5 Bridging Music and Emotion: A New Horizon

As we bring this chapter to a close, we find ourselves at the intersection of music and emotion, having embarked on an extraordinary journey that has woven the two together in a unique and unprecedented manner. Through the avenues of sentiment-driven recommendations, the validation process employing gradient boosting, and a candid exploration of limitations and future possibilities, we have laid the groundwork for a future where music isn't merely an auditory experience but a profoundly emotive one.

In the culmination of our project, we pause to reflect on the achievements that have punctuated our path, while also acknowledging the challenges that have surfaced along the way. Yet, it is in embracing these challenges that we pave the way for the possibilities that lie ahead. With our vision fixed on enriching emotional resonance and cultivating personalized musical encounters, we conclude this chapter.

REFERENCES

Abdul, A., Chen, J., Liao, H.-Y. and Chang, S.-H. (2018). An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach. Applied Sciences, [online] 8(7), p.1103. doi:https://doi.org/10.3390/app8071103.

Agarwal, B. and Mittal, N. (2015) 'Machine learning approach for sentiment analysis', *Socio-Affective Computing*, pp. 21–45. doi:10.1007/978-3-319-25343-5_3.

Barragáns-Martínez, A.B. et al. (2010) 'A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with Singular Value Decomposition', Information Sciences, 180(22), pp. 4290–4311. doi:10.1016/j.ins.2010.07.024.

Barthet, M., Fazekas, G. and Sandler, M. (2013). Music Emotion Recognition: From Content- to Context-Based Models. From Sounds to Music and Emotions, pp.228–252. doi:https://doi.org/10.1007/978-3-642-41248-6_13.

Bhuta, S. and Doshi, U. (2014) 'A review of techniques for sentiment analysis of Twitter data', 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT) [Preprint]. doi:10.1109/icicict.2014.6781346.

Bokde, D., Girase, S. and Mukhopadhyay, D. (2015). Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey. Procedia Computer Science, 49, pp.136–146. doi:https://doi.org/10.1016/j.procs.2015.04.237.

Dang, N.C., Moreno-García, M.N. and De la Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. Electronics, 9(3), p.483. doi:https://doi.org/10.3390/electronics9030483.

Defferrard, M., Benzi, K., Vandergheynst, P. and Bresson, X. (2017). FMA: A Dataset For Music Analysis. arXiv:1612.01840 [cs]. [online] Available at: https://arxiv.org/abs/1612.01840.

Ding, X., Liu, B. and Yu, P.S. (2008) 'A holistic lexicon-based approach to opinion mining', Proceedings of the international conference on Web search and web data mining - WSDM '08 [Preprint]. doi:10.1145/1341531.1341561.

Gómez, L.M. and Cáceres, M.N. (2017). Applying Data Mining for Sentiment Analysis in Music. Advances in Intelligent Systems and Computing, pp.198–205. doi:https://doi.org/10.1007/978-3-319-61578-3_20.

Gossi, D. and Mehmet Hadi Gunes (2016). Lyric-Based Music Recommendation. doi:https://doi.org/10.1007/978-3-319-30569-1_23.

Iván Cantador, Brusilovsky, P. and Tsvi Kuflik (2011). Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011). doi:https://doi.org/10.1145/2043932.2044016.

Jannach, D. (2012) Recommender Systems: An introduction. Cambridge: Cambridge University Press.

Kanakaraj, M. and Guddeti, R.M.R. (2015). Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques. Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015). [online] doi:https://doi.org/10.1109/icosc.2015.7050801.

Kim, J. and Andre, E. (2008). Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(12), pp.2067–2083. doi:https://doi.org/10.1109/tpami.2008.26.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. Communications of the ACM, 40(3), pp.77–87. doi:https://doi.org/10.1145/245108.245126.

Liu, B. (2011) ‘Opinion mining and sentiment analysis’, Web Data Mining, pp. 459–526. doi:10.1007/978-3-642-19460-3_11.

Medhat, W., Hassan, A. and Korashy, H. (2014) ‘Sentiment analysis algorithms and applications: A survey’, Ain Shams Engineering Journal, 5(4), pp. 1093–1113. doi:10.1016/j.asej.2014.04.011.

Mitchell, T.M. (2011) ‘Classical machine-learning paradigms for Data Mining’, Data Mining and Machine Learning in Cybersecurity, pp. 23–56. doi:10.1201/b10867-3.

Navlani, A. and Pandey, P. (2022) Spotify song recommender system in python, Machine Learning Geek. Available at: https://machinelearninggeek.com/spotify-song-recommender-system-in-python/ (Accessed: 28 August 2023).

Olivier, C., Bernhard, S. and Alexander, Z. (2006) ‘Introduction to semi-supervised learning’, Semi-Supervised Learning, pp. 1–12. doi:10.7551/mitpress/9780262033589.003.0001.

Resnick, P. and Varian, H.R. (1997). Recommender systems. Communications of the ACM, 40(3), pp.56–58. doi:https://doi.org/10.1145/245108.245121.

Romero, C. and Ventura, S. (2012) ‘Data Mining in Education’, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), pp. 12–27. doi:10.1002/widm.1075.

Selin Sara Varghese, Manjiri Kherdekar, Benitta Mariam Babu and Archana Shirke (2022). EmotiSync: Music Recommendation System Using Facial Expressions. pp.1–15. doi:https://doi.org/10.1007/978-981-19-4182-5_1.

Shah, D., Singh, A. and Prasad, S.S. (2022). Sentimental Analysis Using Supervised Learning Algorithms. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICCAKM54721.2022.9990320.

Srivastava, G. and Bag, S. (2023) ‘Modern-day marketing concepts based on face recognition and neuro-marketing: A review and future research directions’, Benchmarking: An International Journal [Preprint]. doi:10.1108/bij-09-2022-0588.

Su, X. and Khoshgoftaar, T.M. (2009) ‘A survey of collaborative filtering techniques’, Advances in Artificial Intelligence, 2009, pp. 1–19. doi:10.1155/2009/421425.

Taboada, M. et al. (2011) ‘Lexicon-based methods for sentiment analysis’, Computational Linguistics, 37(2), pp. 267–307. doi:10.1162/coli_a_00049.

Tan, S., Bu, J., Chen, C., Xu, B., Wang, C. and He, X. (2011). Using rich social media information for music recommendation via hypergraph model. ACM Transactions on Multimedia Computing, Communications, and Applications, 7S(1), pp.1–22. doi:https://doi.org/10.1145/2037676.2037679.

Thierry Bertin-Mahieux, Daniel, Whitman, B. and Lamere, P. (2011). THE MILLION SONG DATASET. International Symposium/Conference on Music Information Retrieval, pp.591–596. doi:https://doi.org/10.7916/d8nz8j07.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), pp.293–302. doi:https://doi.org/10.1109/tsa.2002.800560.

Wang, J., de Vries, A.P. and Reinders, M.J. (2006) 'Unifying user-based and item-based collaborative filtering approaches by similarity fusion', Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval [Preprint]. doi:10.1145/1148170.1148257.

Wang, X. and Wang, Y. (2014). Improving Content-based and Hybrid Music Recommendation using Deep Learning. Proceedings of the ACM International Conference on Multimedia - MM '14. [online] doi:https://doi.org/10.1145/2647868.2654940.

Zangerle, E., Pichl, M. and Schedl, M. (2018). Culture-Aware Music Recommendation. Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization. doi:https://doi.org/10.1145/3209219.3209258.

Zawacki-Richter, O. et al. (2019) 'Systematic review of research on Artificial Intelligence Applications in higher education – where are the educators?', International Journal of Educational Technology in Higher Education, 16(1). doi:10.1186/s41239-019-0171-0.

Zhang, L., Wang, S. and Liu, B. (2018) 'Deep learning for sentiment analysis: A survey', WIREs Data Mining and Knowledge Discovery, 8(4). doi:10.1002/widm.1253.

Appendix A: Code Snippets and Implementation Details

This appendix provides an insight into the technical aspects of the implementation, offering relevant code snippets that contribute to the functionality of the sentiment-based music playlist recommendation system. The following code snippets highlight key components and methods used to achieve the system's objectives.

Relevant Code Snippets Used in the Implementation

1. Sentiment Analysis Model Initialization

```
# Function to recommend songs
def recommend_songs_with_evaluation(emotion, num_recommendations=5):
    # Retrieve emotion-tagged playlist from Spotify
    playlist_id = emotion_mapping[emotion]
    playlist_tracks = sp.playlist_tracks(playlist_id)
    playlist_track_ids = [track['track']['id'] for track in playlist_tracks['items']]

    # Filter songs from CSV that are present in the playlist
    emotion_tagged_songs = features_df[features_df['track_id'].isin(playlist_track_ids)].copy()

    # Drop categorical features and time_signature
    columns_to_drop = ['genre', 'artist_name', 'track_name', 'track_id', 'time_signature', 'mode', 'key']
    emotion_tagged_audio_features = emotion_tagged_songs.drop(columns_to_drop, axis=1)

    # Impute missing values
    imputer = SimpleImputer(strategy='mean')
    emotion_tagged_audio_features_imputed = imputer.fit_transform(emotion_tagged_audio_features)

```

2. Emotion Recognition Using CNN

```
@app.route('/detect_emotion_and_recommend', methods=['POST'])
def detect_emotion_and_recommend():
    try:
        # Rest of your emotion detection code
        cap = cv2.VideoCapture(0)
        ret, frame = cap.read()
        cap.release()

        gray_frame = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
        face = face_cascade.detectMultiScale(gray_frame, scaleFactor=1.1, minNeighbors=5, minSize=(30, 30))

        detected_emotion = None
        if len(face) > 0:
            (x, y, w, h) = face[0]
            face_roi = gray_frame[y:y+h, x:x+w]
            resized_face = cv2.resize(face_roi, (48, 48))
            normalized_face = resized_face / 255.0
            input_data = np.expand_dims(np.expand_dims(normalized_face, axis=-1), axis=0)
```

3. K-Nearest Neighbors (KNN) Recommendation

```python
    # Fit Nearest Neighbors model
    knn = NearestNeighbors(n_neighbors=num_recommendations, algorithm='brute')
    knn.fit(emotion_tagged_audio_features_imputed)

    # Get indices of recommended songs
    _, recommended_song_indices = knn.kneighbors(emotion_tagged_audio_features_imputed)

    # Get recommended songs without duplicates
    recommended_songs = emotion_tagged_songs.iloc[recommended_song_indices[0]]
    recommended_songs.drop_duplicates(subset=['track_id'], keep='first', inplace=True)

    # Select top N recommended songs
    top_recommended_songs = recommended_songs.head(num_recommendations)

    # Create a list of recommended songs in the format "song_name - artist_name"
    recommended_song_list = [f"{song['track_name']} - {song['artist_name']}" for _, song in top_recommended_songs.i

    # Create a list of Spotify links
    spotify_links = ['https://open.spotify.com/track/' + track_id for track_id in top_recommended_songs['track_id']

    # Create a list to store emotion percentages for recommended songs
    emotion_percentages_list = []
    # Select the desired columns for training
    selected_columns = ['danceability', 'energy', 'loudness',
                        'speechiness', 'acousticness', 'instrumentalness',
                        'liveness', 'valence', 'tempo', 'duration_ms']

    # Predict emotion percentages for each recommended song
    for _, song in top_recommended_songs.iterrows():
        song_features = song[selected_columns].values.reshape(1, -1)
        emotion_percentages = gb_model.predict_proba(song_features)[0]
        emotion_percentages_list.append(emotion_percentages.tolist())  # Convert ndarray to list
```

Detailed Explanation of Technical Aspects

- Sentiment Analysis Model: The sentiment analysis module is crucial in assessing the emotional tone of textual content. It utilizes Natural Language Processing (NLP) techniques to classify sentiment as positive, negative, or neutral. This model plays a pivotal role in analyzing song lyrics or user-generated content to understand emotional context.
- Emotion Recognition CNN: The emotion recognition module employs Convolutional Neural Networks (CNN) to decipher emotions from facial expressions. By processing facial images, the model detects emotional cues, further enhancing the emotional dimension of song recommendations.
- K-Nearest Neighbors (KNN) Recommendation: The KNN recommendation module employs KNN algorithm to suggest songs based on their audio features. By identifying similar songs, the system creates playlists that align with users' emotions and musical preferences.

This appendix serves as a technical resource, shedding light on the implementation details through code snippets and explanations, providing a comprehensive understanding of the sentiment-based music playlist recommendation system's inner workings.

Appendix B: Data Sources and Ethical Considerations

This appendix delves into the data sources utilized throughout the project and addresses the ethical considerations and data privacy measures undertaken to ensure responsible research practices.

List of Data Sources Used in the Project

- Music Dataset: A curated collection of songs spanning various genres and emotions, sourced from reputable Last.fm music databases, ensuring a diverse range of musical content for analysis.
- Facial Expression Dataset: A dataset containing images of facial expressions representing different emotions, collected from publicly available sources and databases on Kaggle compliant with ethical guidelines.
- User Data: Realtime user emotion and preferences were collected during system testing, ensuring the relevance and personalization of recommendations.

Discussion of Ethical Considerations and Data Privacy

- **Data Privacy Protection:** User data and sensitive information have been treated with utmost care and confidentiality. All user-related data has been anonymized and stored securely, ensuring compliance with data protection regulations.
- **Informed Consent:** In cases where user data was used, participants were informed about the purpose of data collection, and explicit consent was obtained. Participants' privacy and rights were prioritized throughout the research process.
- **Ethical Review:** The project adheres to ethical standards and guidelines set by the relevant institutions. Any potential ethical concerns were thoroughly evaluated and addressed before proceeding with data collection and analysis.
- **Transparent Data Usage:** Data sources, including music and facial expression datasets, were selected from publicly available and ethical sources. Proper attribution and adherence to usage terms were maintained.

- **Bias Mitigation:** Efforts were made to mitigate biases in data and model predictions. Careful selection of datasets and validation methods aimed to prevent potential biases and ensure fair representation.

This appendix provides a transparent account of the data sources used in the project and highlights the ethical considerations and precautions taken to ensure the privacy, rights, and ethical integrity of all parties involved.