MOVIE ANALYTIC USING CNN AND VADER

AINA SYAZZWEEN SURAYA BINTI ABDUL LATIF

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

BORANG PENGESAHAN STATUS LAPORAN

JUDUL: Movie Analytic Using CNN and Vader

SESI PENGAJIAN: 2022 / 2023

Saya: AINA SYAZZWEEN SURAYA BINTI ABDUL LATIF

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan unituk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

_____	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
_____	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan)
___✓___	TIDAK TERHAD	

Enasyazzs

_____ (TANDATANGAN PELAJAR)

Alamat tetap: NO 98, BLOK L, PERINGKAT 3, SENDAYAN, 71950, SEREMBAN, NEGERI SEMBILAN

Tarikh: 01/09/2023

_____ (TANDATANGAN PENYELIA)

PROFESSOR MADYA GS. DR. ASMALA AHMAD

Nama Penyelia

Tarikh: 13/09/2023

MOVIE ANALYTIC USING CNN AND VADER

AINA SYAZZWEEN SURAYA BINTI ABDUL LATIF

This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Artificial Intelligence) with Honours.

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

DECLARATION

I hereby declare that this project report entitled

MOVIE ANALYTIC USING CNN AND VADER

is written by me and is my own effort and that no part has been plagiarized

without citations.

STUDENT: Enasyazzs Date : 13/09/2023

(AINA SYAZZWEEN SURAYA BINTI ABDUL LATIF)

I hereby declare that I have read this project report and found

this project report is sufficient in term of the scope and quality for the award of

Bachelor of Computer Science (Artificial Intelligence) with Honours.

SUPERVISOR: _____ Date : 13/09/2023

(PROFESSOR MADYA GS. DR. ASMALA AHMAD)

DEDICATION

This dedication is dedicated to my beloved parents, who have been my unwavering source of love, support, and inspiration throughout my life. Words cannot express the depth of my gratitude for everything they have done for me. Their selfless sacrifices, tireless efforts, and unconditional love have shaped me into the person I am today. I dedicate all my accomplishments and successes to them, knowing that without their unwavering support, none of it would have been possible. They have taught me the importance of hard work, perseverance, and integrity, and I am forever grateful for the values they have instilled in me. Through all the ups and downs, they have been my pillars of strength, offering a comforting embrace and wise counsel. Their love has given me the confidence to face any obstacle and the assurance that I am never alone.

ACKNOWLEDGEMENTS

I would like to express my heartfelt gratitude to my project supervisor, Professor Madya Gs. Dr. Asmala Ahmad, for his unwavering guidance, support, and expertise throughout the entire duration of this project. His insightful feedback and constructive criticism have been invaluable in shaping the direction and quality of my work.

I am immensely thankful to my parents, Abdul Latif Bin Minhat and Norlizah Binti Anshah for their endless love, encouragement, and unwavering belief in my abilities. Their constant support and sacrifices have been the foundation of my academic journey, and I am forever grateful for their presence in my life.

I extend my appreciation to the faculty members of University Technical Malaysia Malacca for their continuous support and for providing me with an enriching learning environment. Their knowledge and dedication have played a significant role in my academic growth.

My sincere thanks go to my friends and classmates for their camaraderie, brainstorming sessions, and encouragement throughout this project. Their valuable input and discussions have contributed to the refinement of my ideas and the implementation of my project.

Lastly, I would like to express my gratitude to all the participants who willingly volunteered their time and provided valuable data for my research. Without their involvement, this project would not have been possible.

I am deeply grateful to all individuals and organizations mentioned above, as well as any others who have contributed in any way to the completion of this project. Their collective support has been invaluable, and I am truly fortunate to have had their assistance.

ABSTRACT

The Movie Analytic Using CNN And VADER system incorporates three fundamental modules which are movie recommendation, genre prediction, and review analysis. Its core objectives involve the development of a sophisticated movie analytics platform, utilizing cutting-edge techniques such as Convolutional Neural Networks (CNN), the VADER lexicon, and content-based filtering. Within the movie recommendation module, content-based filtering drives personalized movie suggestions based on user preferences and viewing history, enhancing the overall viewing experience. For genre prediction, four distinct CNN models, namely AlexNet, Keras Sequential, LeNet, and VGGNet 16, are employed, with AlexNet leading with a 91% genre prediction accuracy, offering automated genre tagging for streamlined content management. The movie review analysis module, leveraging the VADER lexicon, impressively achieves 98.67% accuracy in assessing sentiment, providing valuable insights into audience reactions and opinions. This system serves as a comprehensive toolset for movie analysis, benefiting a wide array of stakeholders including business organizations for data-driven insights, production companies for content tagging, and local residents seeking personalized movie recommendations. Ongoing refinement and adaptation to evolving user preferences and technological advancements are pivotal for continuously elevating its performance in the dynamic field of movie analytics.

ABSTRAK

Sistem Movie Analytic Menggunakan CNN Dan VADER merangkumi tiga modul utama iaitu cadangan filem, jangkaan genre, dan analisis ulasan. Objektif utamanya adalah membangunkan platform analitik filem yang canggih dengan menggunakan teknik terkini seperti Convolutional Neural Networks (CNN), kamus VADER, dan penapisan berasaskan kandungan. Dalam modul cadangan filem, penapisan berasaskan kandungan memacu cadangan filem yang dipersonalisasi berdasarkan kepada pilihan pengguna dan sejarah tontonan mereka, meningkatkan pengalaman keseluruhan. Untuk jangkaan genre, empat model CNN yang berbeza, iaitu AlexNet, Keras Sequential, LeNet, dan VGGNet 16, digunakan, dengan AlexNet memimpin dengan ketepatan jangkaan genre sebanyak 91%, menawarkan penandaan genre automatik untuk pengurusan kandungan yang lebih efisien. Modul analisis ulasan filem, dengan menggunakan kamus VADER, mencapai ketepatan sebanyak 98.67% dalam menilai sentimen, menyediakan pandangan berharga terhadap reaksi dan pendapat penonton. Sistem ini berperanan sebagai satu set alat komprehensif untuk analisis filem, memberi faedah kepada pelbagai pihak berkepentingan termasuk organisasi perniagaan untuk wawasan berasaskan data, syarikat produksi untuk penandaan kandungan, dan penduduk tempatan yang mencari cadangan filem yang dipersonalisasi. Penambahbaikan berterusan dan penyesuaian kepada pilihan pengguna yang berubah dan kemajuan teknologi adalah penting untuk terus meningkatkan prestasi sistem ini dalam bidang analitik filem yang dinamik.

TABLE OF CONTENTS

ABSTRACT ..... **V**
ABSTRAK ..... **VI**
TABLE OF CONTENTS ..... **VII**
LIST OF TABLES ..... **IX**
LIST OF FIGURES ..... **XI**
LIST OF ABBREVIATIONS ..... **XIV**
LIST OF ATTACHMENTS ..... **XV**
CHAPTER 1: INTRODUCTION **1**
1.1 INTRODUCTION 1
1.2 PROBLEM STATEMENTS 2
1.3 OBJECTIVES 2
1.4 SCOPE 3
1.5 PROJECT SIGNIFICANT 3
1.6 EXPECTED OUTPUT 3
1.7 ORGANIZATION OF REPORT 4
CHAPTER 2: LITERATURE REVEW AND PROJECT METHODOLOGY **5**
2.1 INTRODUCTION 5
2.2 FACTS AND FINDINGS 5
2.2.1 Domain 5
2.2.2 Existing System 12
2.2.3 Techniques 17
2.3 PROJECT METHODOLOGY 20
2.4 PROJECT REQUIREMENTS 22
2.4.1 Software Requirement 22
2.4.2 Hardware Requirement 23
2.5 PROJECT SCHEDULE AND MILESTONE 24
2.6 CONCLUSION 25
CHAPTER 3: ANALYSIS **26**
3.1 INTRODUCTION 26
3.2 PROBLEM ANALYSIS 26
3.3 REQUIREMENT ANALYSIS 29
3.3.1 Data Requirement 29
3.3.2 Functional Requirement 35
3.3.3 Non-Functional Requirement 36
3.3.4 Other Requirement 38
3.4 CONCLUSION 40
CHAPTER 4: DESIGN **41**
4.1 INTRODUCTION 41
4.2 HIGH-LEVEL DESIGN 41
4.2.1 System Architecture 41
4.2.2 User Interface 42

4.3 DETAILED DESIGN 49
4.3.1 Movie Recommendation Module 49
4.3.2 Movie Genre Prediction Module 50
4.3.3 Movie Review Analysis 51
4.4 CONCLUSION 52

CHAPTER 5: IMPLEMENTATION 53

5.1 INTRODUCTION 53
5.2 SOFTWARE DEVELOPMENT ENVIRONMENT SETUP 53
5.3 SOFTWARE CONFIGURATION MANAGEMENT 54
5.3.1 Version Control Procedure 54
5.4 IMPLEMENTATION STATUS 55
5.4.1 Movie Recommendation Module 55
5.4.2 Movie Genre Prediction Module 57
5.4.3 Movie Review Analysis Module 60

CHAPTER 6: TESTING 62

6.1 INTRODUCTION 62
6.2 TEST PLAN 62
6.2.1 Test Organization 62
6.2.2 Test Environment 64
6.2.3 Test Schedule 65
6.3 TEST STRATEGY 66
6.3.1 Classes of test 66
6.4 TEST IMPLEMENTATION 67
6.4.1 Test description 67
6.4.2 Test data 70
6.5 TEST RESULT AND ANALYSIS 71
6.5.1 Movie Recommendation Module 71
6.5.2 Movie Genre Prediction Module 72
6.5.3 Movie Review Analysis Module 75
6.6 CONCLUSION 79

CHAPTER 7: CONCLUSION 80

7.1 OBSERVATION ON STRENGTHS AND WEAKNESSES 80
7.2 PROPOSITION FOR IMPROVEMENT 80
7.3 PROJECT CONTRIBUTION 81
7.4 CONCLUSION 81

REFERENCES 83

APPENDIX A 85

APPENDIX B 86

LIST OF TABLES

PAGES

Table 2.1: The differences between CNN types **18-19**

Table 2.2: Software requirements to build Movie Analytic Using CNN And VADER **22**

Table 2.3: Hardware requirements to build Movie Analytic Using CNN And VADER **23**

Table 2.4: Description of activities for each phase **24-25**

Table 2.5: Gantt Chart **25**

Table 3.1: Data dictionary of TMDb Movie Dataset **30-32**

Table 3.2: Data dictionary of Movie Poster Images Dataset **33**

Table 3.3: Data dictionary of the VADER lexicon **34**

Table 3.4: Non-Functional Requirements **36-37**

Table 3.5: Other requirements **38-40**

Table 5.1: Movie Recommendation Module implementation status **56-57**

Table 5.2: Movie Genre Prediction Module implementation status **58-59**

Table 5.3: Movie Review Analysis Module implementation status **61**

Table 6.1: Test organization od Movie Analytic Using CNN And VADER **64**

Table 6.2: Test environment of Movie Analytic Using CNN And VADER **65**

Table 6.3: Test schedule of Movie Analytic Using CNN And VADER **66**

Table 6.4: Test description of Movie Recommendation Module **68-69**

Table 6.5: Test Description of Movie Genre Prediction Module **69-70**

Table 6.6: Test Description of Movie Review Analysis Module **70-71**

Table 6.7: Test data used for each module of Movie Analytic Using CNN And VADER **71**

Table 6.8: Test result of Movie Recommendation Module **72-73**

Table 6.9: Test result of Movie Genre Prediction Module **73-74**

Table 6.10: Test result of Movie Genre Prediction Module test data **74-73**

Table 6.11: Test result of Movie Review Analysis Module **76**

Table 6.12: Test result of Movie Review Analysis Module test data **77-79**

Table 7.1: Overall conclusion of the project **82**

LIST OF FIGURES

PAGES

Figure 2.1: Types of Recommendation System **6**

Figure 2.2: Content-Based Filtering Technique **7**

Figure 2.3: Collaborative Filtering Technique **7**

Figure 2.4: User-based collaborative filtering **8**

Figure 2.5: Item-based collaborative filtering **9**

Figure 2.6: Hybrid filtering technique **10**

Figure 2.7: Types of Classification **10**

Figure 2.8: Cosine similarity formula **17**

Figure 2.9: The architecture of Convolutional Neural Network **18**

Figure 2.10: Sentiment analysis using VADER **19**

Figure 2.11: Waterfall model for software development **20**

Figure 3.1: Online movie platform **26**

Figure 3.2: Online movie platform system **27**

Figure 3.3: Flowchart of movie genre prediction system **28**

Figure 3.4: Flowchart of movie review analysis system **29**

Figure 3.5: List of movie posters in Movie Poster Images Dataset **33**

Figure 3.6: Use case of Movie Analytic Using CNN And VADER **35**

Figure 4.1: System architecture of Movie Analytic Using CNN And VADER **42**

Figure 4.2: Navigation bar **42**

Figure 4.3: Input design for movie recommendation module **43**

Figure 4.4: Input design for movie genre prediction module **43**

Figure 4.5: Input design for movie review analysis **44**

Figure 4.6: Technical design for movie recommendation module **44**

Figure 4.7: Technical design for movie genre prediction module **45**

Figure 4.8: Technical design for movie review analysis **45**

Figure 4.9: Output design for movie recommendation module **46**

Figure 4.10: Error page for invalid input **46**

Figure 4.11: Output design for movie genre prediction module **47**

Figure 4.12: Output design for movie review analysis **48**

Figure 5.1: Github repository **54**

Figure 5.2: Code snippet of Movie Recommendation Module functions **56**

Figure 5.3: Code snippet of Movie Recommendation Module backend deployment **57**

Figure 5.4: Data preprocessing of Movie Poster Images Dataset **59**

Figure 5.5: The architecture of AlexNet model **59**

Figure 5.6: Code snippet of Movie Review Analysis Module backend **61**

Figure 6.1: Formula of CNN model accuracy **74**

Figure 6.2: Calculation of CNN model accuracy **74**

Figure 6.3: Formula of VADER lexicon model accuracy **79**

Figure 6.4: Calculation of VADER lexicon model accuracy **79**

LIST OF ABBREVIATIONS

FYP	-	**Final Year Project**
CNN	-	**Convolutional Neural Network**
NLP	-	**Natural Language Processing**
ML	-	**Machine Learning**
VADER	-	**Valance Aware Dictionary and Sentiment Reasoner**
ReLU	-	**Rectified Linear Unit**
API	-	**Application Programming Interface**
Sklearn		**Scikit-learn**
GUI	-	**Graphical User Interface**
TPU	-	**Tensor Processing Unit**
TMDB	-	**The Movie Database**
IMDb	-	**Internet Movie Dataset**
Pandas	-	**Python Data Analysis Library**
NumPy	-	**Numerical Python**
NLTK	-	**Natural Language Toolkit**
HTML	-	**Hypertext Markup Language**
CPU	-	**Central Processing Unit**
SSD	-	**Solid State Drive**
HDD	-	**Hard Disk Drive**
RAM	-	**Random Access Memory**
GB	-	**Gigabyte**
CSS	-	**Cascading Style Sheets**
CSV	-	**Comma Separated Environment**
VSC	-	**Visual Studio Code**
IP	-	**Internet Protocol**
TF-IDF	-	**Term Frequency-Inverse Document Frequency**
RAKE	-	**Rapid Automatic Keyword Extraction**
Hz	-	**Hertz**
SDE	-	**Software Development Environment**

LIST OF ATTACHMENTS

		PAGES
Appendix A	**Sample of Data**	**85**
Appendix B	**Sample of Source Code**	**86-93**

CHAPTER 1: INTRODUCTION

1.1 Introduction

Movies have grown in popularity and effectiveness as a source of inspiration. They provide viewers with motivational stories, positive role models, and support for personal development. As a result, movies may be an effective motivator for people to realize their maximum potential. Watching movies at home is also more convenient than going to the cinema. Viewers may view films in their own area, at their own pace, and without other people's distractions. This ease allows audiences to interact with the film and get inspired by its message. Streaming services such as Netflix, WatchTv, Rave, ViuTv, iQIYI, and Disney Plus provide a vast variety of films that cater to a wide range of interests and tastes. Viewers can select from categories such as drama, comedy, action, and documentary. This variety ensures that viewers, regardless of their personal preferences, can find inspiring and motivating films. Therefore, Movie Analytics Using CNN and VADER plays a huge role in predicting and suggesting movies that viewers are likely to enjoy.

Action, comedy, drama, romance, horror, and science fiction are some of the film genres. Each genre has its own distinct characteristics, concepts, and audience. Each genre has its own distinct characteristics, concepts, and viewers. The preference category is important in movie suggestions since it assists viewers in finding films that fit their interests and mood. Hence, a web-based movie recommendation system employs algorithms to recommend films to viewers based on characteristics such as genres, plot, and casts as well as review analysis feature.

A machine learning approach called multi-label classification is used to predict many classes or categories for a single data point. In this context of movie genre classification, this approach is used to categorize movies into different genres based on posters by extracting their visual attributes. In analyzing the quality of a film and anticipating its popularity, sentiment analysis may be a beneficial tool for both filmmakers and audiences. Positive feedback indicates that a film was well-made and of high quality while negative feedback might imply a low-quality or poorly made film.

1.2 Problem Statements

There are various elements to consider before selecting a movie or television show. The first and most crucial component is individual preferences in genre. Some people enjoy action films, while others prefer comedies, horror films, or drama. It is critical to understand viewers chosen genre in order to prevent spending time on films that viewers will not find valuable.

Another element to ponder about is the cast and crew. A strong director, writer, and cast may make a considerable difference in the overall quality of a film or television show. When deciding, ratings and reviews are also significant. Websites such as IMDB, Rotten Tomatoes, and Metacritic can give useful information and should be considered in order to determine the quality of a film or television show.

Besides that, it is critical to consider the content of the film before watching it with children, as films that are intended for adults might not be appropriate for younger viewers. One of the most difficult issues is selecting what content is acceptable for children of various ages. Despite these challenges, we need to understand movie genres in order to choose appropriate films to watch with family and friends. It might be challenging to classify films into distinct genres since they can have many different layers.

1.3 Objectives

1. To identify the requirements for developing a movie analytic.
2. To develop a movie analytics technique using Convolutional Neural Network (CNN), VADER and Content-Based Filtering.
3. To evaluate the performance of the technique using Cosine Similarity, Valence Aware Dictionary And Sentiment Reasoner (VADER) and CNN.

1.4 Scope

The web-based Movie Analytic is intended for audiences from diverse backgrounds. According to the following, there are three modules that to be developed for this system:

1. Movie genre prediction module: The system will identify and classify the genres based on movie posters using Convolutional Neural Network (CNN).
2. Movie recommendation module: The system will generate movie suggestions based on user input using a content-based filtering technique.
3. Movie review analysis module: The user may input their reviews and the system will generate the polarity of excellent movie using Valence Aware Dictionary and Sentiment Reasoner (VADER).

1.5 Project Significant

The goal of the project is to assist users in quickly and accurately identifying the films and television programs of their choosing without spending time by perusing vast amounts of movie information. Additionally, to enable users to classify movies into various genres by identifying specific elements from the movie's poster. This will enable users to decide whether they would be interested in a particular movie. The user would benefit from finding films that are appropriate for viewing with friends and family. In addition, users may see if viewers have a good, negative, or neutral opinion on a certain film before deciding whether to watch it.

1.6 Expected Output

The expected outcome of this project is a web-based movie analytic that could categorize a particular movie into different genres based on movie poster, suggest suitable movie based on user input using Content-Based Filtering technique and analyze movie reviews as positive, negative, or neutral using Valence Aware Dictionary and sentiment Reasoner (VADER).

1.7 Organization of Report

The report is structured into several chapters, each addressing different aspects of the Movie Analytic Using CNN And VADER. Chapter 1 provides an overview of the problem's background, objectives, and project scope. It includes a comprehensive literature review, project requirements, project schedule, and the methodology employed for system development. Moving on to Chapter 3, the report delves into the analysis of the problem and requirements for the project. Chapter 4 focuses on the purposeful design, system architecture, and interfaces of the individual modules within the Movie Recommender Systems. In Chapter 5, the report details the management, implementation, and integration of each module. Furthermore, Chapter 6 covers the methods employed to test the accuracy of each module and highlights the limitations and constraints of the integrated system. Lastly, Chapter 7 presents an assessment of the advantages and disadvantages of the Movie Analytic Using CNN And VADER, as well as their commercial value.

CHAPTER 2: LITERATURE REVEW AND PROJECT METHODOLOGY

2.1 Introduction

This chapter will describe the literature review and project methodology used to implement Movie Analytic Using CNN and VADER. A literature review is a critical and in-depth analysis of previous studies and publications on a certain subject. It entails locating, examining, and synthesizing pertinent academic books, journals, and other materials in order to grasp the current state of knowledge about a certain topic or research issue. Whereas methodologies are the concepts, processes, and equipment utilized to do research in each topic or discipline. It entails the methodical gathering and examination of data, as well as the use of suitable approaches and strategies to draw valuable insights and conclusions from that data. Typical components of a research methodology include a thorough explanation of the study design, sample plan, data collection methodologies, data analysis methods, and any other processes or techniques that were employed to carry out the research.

2.2 Facts and Findings

In order to offer a better understanding of the principles and methods used to implement Movie Analytic Using CNN and VADER, this part will present all the significant material that has been gathered from numerous sources, including journals, research papers, and books.

2.2.1 Domain

This section lists out all the related domains of Movie Analytic Using CNN and VADER.

2.2.1.1 Recommendation systems

A movie recommendation system is a type of information filtering system that aids users in finding goods, services, or content that they might be interested in. It is a piece of software that offers users tailored recommendations based on their previous activities, preferences, and system interactions. There are different types of recommendation systems, such as content-based filtering, collaborative filtering, and hybrid systems that combine both methods. This section will mainly introduce the fundamental principles of each type of collaborative filtering. Figure 2.1 shows the types of recommendation system.

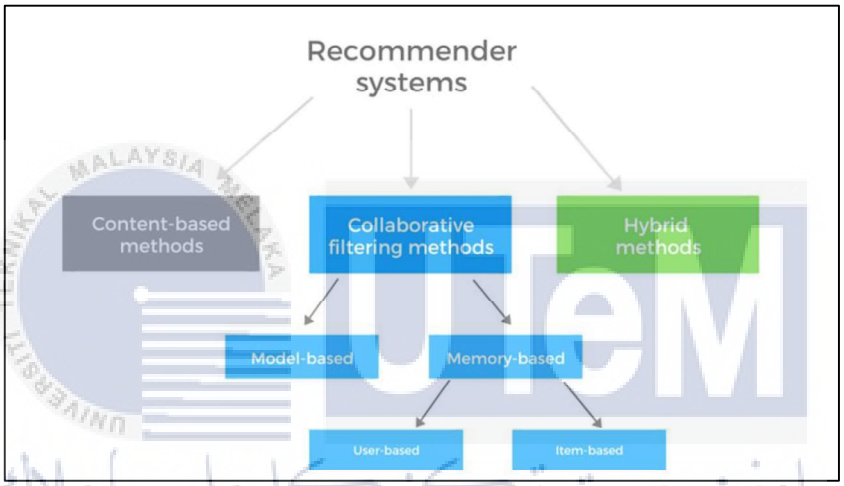

Figure 2.1: Types of Recommendation System (Quantum, 2023)

a) Content-Based Filtering

Content-based filtering is based on the properties of the recommended items. It analyses the qualities of the items and offers others that are comparable in those aspects. For example, if a viewer is viewing an action/comedy film, the content-based filtering algorithm will recommend films in comparable genres. Figure 2.2 shows content-based filtering technique.

Figure 2.2: Content-based filtering technique (Altexsoft, 2021)

b) Collaborative Filtering

Collaborative filtering is a form of recommendation system that generates recommendations for an individual user by combining input from numerous users. The algorithm recognizes trends in user behavior and preferences and utilizes them to forecast what a certain user would like. There are two forms of collaborative filtering: user-based collaborative filtering and item-based collaborative filtering. Figure 2.3 shows a collaborative filtering technique.

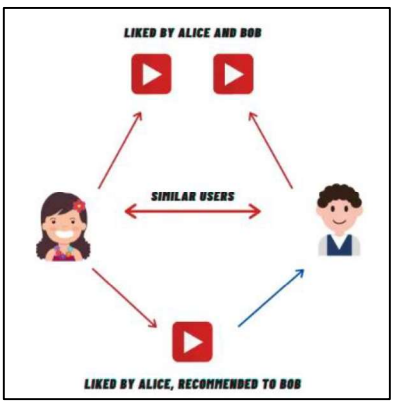

Figure 2.3: Collaborative filtering technique (Anand, 2020)