A PROJECT REPORT

ON

"DETECTING FAKE REVIEW USING OPINION MINING"

Submitted to

UNIVERSITY OF MUMBAI

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER ENGINEERING

BY

ANSARI MOHD FAIZ MOHD SIDDIQUE   **14CO17**
KHAN KHALID AHMED SAJID AHMED   **16DCO54**
SHAIKH AZAMALI MOHD MAJID   **17DCO75**

UNDER THE GUIDANCE OF
Prof. MUKHTAR ANSARI

DEPARTMENT OF COMPUTER ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,
Khandagaon, New Panvel - 410206
2019-2020

AFFILIATED TO

UNIVERSITY OF MUMBAI

A PROJECT II REPORT
ON

"DETECTING FAKE REVIEW USING OPINION MINING"

Submitted to
UNIVERSITY OF MUMBAI

In Partial Fulfilment of the Requirement for the Award of

BACHELOR'S DEGREE IN
COMPUTER ENGINEERING

BY

ANSARI MOHD FAIZ MOHD SIDDIQUE   14CO17
KHAN KHALID AHMED SAJID AHMED   16DCO54
SHAIKH AZAMALI MOHD MAJID   17DCO75

UNDER THE GUIDANCE OF
Prof. MUKHTAR ANSARI

DEPARTMENT OF COMPUTER ENGINEERING
Anjuman-I-Islam's Kalsekar Technical Campus
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,

Khandagaon, New Panvel - 410206

2019-2020
AFFILIATED TO

UNIVERSITY OF MUMBAI

Anjuman-i-Islam's Kalsekar Technical Campus
Department of Computer Engineering
SCHOOL OF ENGINEERING & TECHNOLOGY
Plot No. 2 3, Sector - 16, Near Thana Naka,

Khandagaon, New Panvel - 410206

CERTIFICATE

This is certify that the project entitled

"DETECTING FAKE REVIEW USING OPINION MINING"

submitted by

ANSARI MOHD FAIZ MOHD SIDDIQUE	**14CO17**
KHAN KHALID AHMED SAJID AHMED	**16DCO54**
SHAIKH AZAMALI MOHD MAJID	**17DCO75**

is a record of bonafide work carried out by them, in the partial fulfilment of the requirement for the award of Degree of Bachelor of Engineering (Computer Engineering) at *Anjuman-I-Islam's Kalsekar Technical Campus, Navi Mumbai* under the University of MUMBAI. This work is done during year 2019-2020, under our guidance.

Date: /   /

(Prof. MUKHTAR ANSARI)   **(Prof. KALPANA BODKE)**
    Project Supervisor   **Project Coordinator**

(Prof. TABREZ KHAN)   **DR. ABDUL RAZAK HONNUTAGI**
HOD, Computer Department   **Director**

External Examiner

Acknowledgements

I would like to take the opportunity to express my sincere thanks to my guide **Prof. MUKHTAR ANSARI**, Assistant Professor, Department of Computer Engineering, AIKTC, School of Engineering, Panvel for his invaluable support and guidance throughout my project research work. Without his kind guidance & support this was not possible.

I am grateful to him/her for his timely feedback which helped me track and schedule the process effectively. His/her time, ideas and encouragement that he gave is help me to complete my project efficiently.

We would like to express deepest appreciation towards **DR. ABDUL RAZAK HONNUTAGI**, Director, AIKTC, Navi Mumbai, **Prof. TABREZ KHAN**, Head of Department of Computer Engineering and **Prof. KALPANA BODKE**, Project Coordinator whose invaluable guidance supported us in completing this project.

At last we must express our sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped me directly or indirectly during this course of work.

<div style="text-align: right;">

ANSARI MOHD FAIZ MOHD SIDDIQUE

KHAN KHALID AHMED SAJID AHMED

SHAIKH AZAMALI MOHD MAJID

</div>

Project I Approval for Bachelor of Engineering

This project entitled *ḒETECTING FAKE REVIEW USING OPINION MINING¨* by *ANSARI MOHD FAIZ MOHD SIDDIQUE, KHAN KHALID AHMED SAJID AHMED AND SHAIKIH AZAMALI MOHD MAJID* is approved for the degree of *Bachelor of Engineering in Department of Computer Engineering.*

Examiners

1. ....
2. ....

Supervisors

1. ....
2. ....

Chairman

....

Declaration

I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

ANSARI MOHD FAIZ MOD SIDDIQUE
14CO17

KHAN KHALID AHMED SAJID AHMED
16DCO54

SHAIKH AZAMALI MOHD MAJID
17DCO75

ABSTRACT

Tittle: Detecting Fake Reviews Using Opinion Mining

In today's lifeline online marketing is on it's verge and keep increasing as people buy and sell products online. Digital marketing is an interesting and trending business platform. Sellers often post fake reviews on their products or pay people to post reviews and give higher rating, most consumer usually see and select products according to that product's rating and review which can be turn into dissatisfaction of consumer as he bought that product on the basis of fake reviews. To detect such reviews various methods are being used in past works. In this paper the method is being used is Sentiment Analysis (SA). SA has become one of the most interesting topics in text analysis, due to its promising commercial benefits. SA detects fake positive and fake negative reviews based on emotions in the opinion. In this study, we used machine learning algorithm Support Vector Machine (SVM) to detect those fake negative and fake positive reviews.

Reviews can be positive or negative which helps consumers to select product. This paper aims to classify movie reviews into groups of positive or negative polarity by using machine learning algorithms. For the movies data-sets we performed some data scrapping library like Beautiful- soup and Request to scrap movies data-sets and collected data-sets from websites which pro- vides huge data-sets like dbmovies to perform our experiments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the very rapid growth area is Digital media marketing. Generally Digital media marketing provide facility for customers to write reviews related with its services. The existence of these reviews can be used as a source of information. For examples, movie makers can use it to make design decisions of their services but unfortunately, the importance of the review is misused by certain parties who tried to create fake reviews, both aimed at raising the popularity or to discredit the product. They share their thoughts on internet. Before watching anything, it is a normal human behavior to do a survey on that product. Based on reviews, customers can compare different movies and can finalize a product of their interest. These online reviews can change the opinion of a customer about the movie. If these reviews are true, then this can help the users to select proper product that satisfy their requirements. On the other hand, if the re- views are manipulated or not true then this can mislead user. This boosts us to develop a system which detect fake reviews for a product by using the text and rating property from a review. The honesty value and measure of a fake review will be measured by utilizing the data mining techniques. An algorithm could be used to track customer reviews, through mining topics and sentiment orientation from online customer reviews and will also blocked the fake reviews.

1.1 Purpose

In recent years, online reviews have been playing an important role in decision making. This is because, these reviews can provide customers with large amounts of information about the goods or services. However, to promote fictitiously or lower the quality of the products or services, spammers may forge and produce fake reviews. Due to such behavior of the spammers, customers would be mislead and make wrong decisions. We are making this work to detect fake reviews of movies on various online websites. Thus detecting fake (spam) reviews is a significant problem. Opinion spamming refers to the use of excessive and illicit methods, such as creating a large volume of fake reviews, in order to generate biased positive or negative

opinions for a target product or service with the intention of promoting or demoting it, respectively. The reviews created for this purpose are known as fake, spam or bogus reviews, and the authors responsible for composing such deceptive content are known as fake or spam reviewers.

1.2 Project Scope

Our project is to build the system that can identify and distinguish between fake reviews and genuine reviews, our system will compare the actual rating of the movie and calculated resultant rating of particular movie. Our project will be working on machine learning algorithms that will do the good work of detecting reviews and identify reviews. In today's digital world the need of real and genuine opinion is needed on any product so it will save our time and money, and our project is doing the good work at it. As in digital media marketing to degrading or upgrading the value of movie uses of fake reviews are increasing and by seeing this problem we initiated our work and come up with this idea. Our system is build on top of machine learning and python libraries using algorithms our system works very accurately. The data-set we are importing is directly from IMDB website so the data are totally user based, unlike free data-sets available on internet this data-set is completely random and user's input so the evaluation will work on real data.

1.2.1 Objectives

The identified challenges motivate to bring up a solution to all the problems stated in the above problem statement section. Following are the objectives of the proposed approach and this thesis work:

1) To detect fake positive, fake negative, real positive real negative reviews.

2) To implement algorithm to get better Spam Detection i.e.; Account used, polarity of words using text blob, negativity and positivity of content, etc.

3) Graphical representation of work.

4) To presents Opinion Mining on Spam Filtered Data.

5) To demonstrate the difference between the actual rating of the movie and the calculated resultant rating of the same.

6) To present an algorithm that does Opinion Mining with Spam Detection.

1.3   Technical details

Technologies being used for the development of this application includes,

- Apache Laravel 5.3 for web services.

- Vanilla JS, jQuery, AJAX for Client side validations .

- CSS3, Bootstrap, HTML5 for web designing.

- Flutter for mobile application development.

- MongoDB Firebase Server for databases.

- BeautifulSoup Requests libraries are used for data scrapping.

- Pandas are used for data storing in useful manner.

Chapter 2

Literature Survey

2.1 Fake Review Detection using Classification.

In today's world, where Internet has become a household convenience, online reviews have become a critical tool for businesses to control their online reputation. Reviewing has changed the face of marketing in this new era. Nowadays, most companies invest money in mining the reviews to gain insights into customer preferences as well as to gain competitive intelligence and are hiring individuals to write fake reviews. The fraudsters activities mislead potential customers and organizations reshaping their businesses and prevent opinion-mining techniques from reaching accurate conclusions. Thus, it has become essential to detect fake reviews to bring to surface the true product opinion.

2.1.1 Weaknesses

a. Only for e-commerce products eg: Amazon.

b. Architecture is complex.

c. Computer aided generation of fake reviews are not detect.

d. Naive Bayes and random Forest classification algorithm is applied which made the system slow.

2.1.2 How to Overcome

a. To detect fake reviews of IMDB movies.

b. We used SVM classification algorithm over Naive Bayes and random forest classification as this is faster comparatively.

c. System is efficient and fast.

2.2 Detecting Fake Reviews through Sentimental Analysis Using Machine Learning Techniques.

This paper presents to classify movie reviews into groups of positive or negative polarity by using machine learning algorithms. This system analysis online movie reviews using SA methods in order to detect fake reviews. SA and text classification methods are applied to a data-set of movie reviews. More specifically, compare five supervised machine learning algorithms: Naive Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK), KStar (K*) and Decision Tree (DT-J48).

2.2.1 Weaknesses

a. Used five algorithm to detect the fake predictions which makes the system slow.

b. Architecture is complex.

2.2.2 How to Overcome

a. We used only one algorithm to detect the fake predictions which makes the system work rapidly and generate output.

b. We also used the confusion matrix to count the real and fake predictions with known data.

2.3 An empirical study on detecting fake reviews using machine learning techniques.

This paper introduces a Reputation system in E-commerce (EC). The reputation systems aim at helping consumers in deciding whether to negotiate with a given party. Many factors negatively influence the sight of the customers and the vendors in terms of the reputation system. Moreover, the opinions obtained from users can be classified into positive or negative which can be used by a consumer to select a product.

2.3.1 Weaknesses

a. Only used for E-commerce sites.

b. Four different algorithms to detect whether the given opinion is genuine or not.

c. Using more algorithms to do same work require more amount of time.

2.3.2   How to Overcome

a. Using a single but more accurate algorithm to do the task.

b. Take less time to perform operation.v

c. Using SA on movies reviews to detect fake reviews and calculate precise result to give consumer satisfaction.

2.4   Technical Review

These are the following Software that we have used for our applications:-

- **Machine Learning Algorithms:-**

  Machine Learning is a system of automated data processing algorithms that help to make decision making more natural and enhance performance based on the results. The "learning" implies that the algorithm can glean new information and insights without being explicitly programmed. There are several models of machine learning:

  – Supervised ML (the outputs are labeled, and the inputs are mapped to corresponding outputs)

  – Unsupervised ML (the inputs are unlabeled, and the algorithms have to find patterns)

  – Reinforcement ML (similar to supervised ML, but in this case, instead of a labeled output, there are rewards and the algorithm's goal is to maximize rewards)

  Artificial intelligence (AI) is a part of Computer Science that focuses on systems that can solve problems and perform tasks that require human (or human-like) intelligence. What does machine learning do in this process? It learns, just like a human(like) brain, acquiring new information.

- **Web Technologies:-**

  Web technology is a term referring to multimedia packs which can be employed in conjunction to create web sites like this one and languages. Each technology will require one other such technology's use and is restricted. We can conclude that the components that make up a site all are interdependent on one another.
  The technologies we used are:-
  for the Front-End we used:-
  HTML, CSS, BOOTSRAP, LARAVEL, JAVASCRIPT
  for the Back-End we used:-
  PHP, MYSQL

- **Android Studio:-**

  Android Studio is the integrated development environment (IDE) for Google's Android operating system, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development. It is available for download on Windows, macOS and Linux based operating systems or as a subscription-based service in 2020. It is a replacement for the Eclipse Android Development Tools (E-ADT) as the primary IDE for native Android application development.

- **spyder:-**

  Spyder is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates with a number of prominent packages in the scientific Python stack, including NumPy, SciPy, Matplotlib, pandas, IPython, SymPy and Cython, as well as other open source software. It is released under the MIT license.

  Spyder uses Qt for its GUI, and is designed to use either of the PyQt or PySide Python bindings. QtPy, a thin abstraction layer developed by the Spyder project and later adopted by multiple other packages, provides the flexibility to use either backend.

2.4.1 Advantages of Technology

1. Easily identifies trends and patterns:

   Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for an e-commerce website like Amazon, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

2. No human intervention needed (automation)

   With ML, you don't need to babysit your project every step of the way. Since it means giving machines the ability to learn, it lets them make predictions and also improve the algorithms on their own. A common example of this is anti-virus softwares; they learn to filter new threats as they are recognized. ML is also good at recognizing spam.

3. Continuous Improvement

   As ML algorithms gain experience, they keep improving in accuracy and efficiency. This lets them make better decisions. Say you need to make a weather forecast model. As the amount of data you have keeps growing, your algorithms learn to make more accurate predictions faster.

4. Handling multi-dimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

5. Wide Applications

You could be an e-tailer or a healthcare provider and make ML work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

2.4.2 Reasons to use this Technology

1. Machine Learning provides best in use pre-defined algorithms to work on, these algorithms provides convenient results.

2. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Using machine learning algorithms reduces the use of man work in this project.

Chapter 3

Project Planning

3.1 Members and Capabilities

Table 3.1: Table of Capabilities

SR. No	Name of Member	Capabilities
1	Ansari Faiz	Content Management, Designer
2	Khan Khalid	Machine learning
3	Shaikh Azamali	Web development, Database Management

Work Breakdown Structure

3.2 Roles and Responsibilities

Table 3.2: Table of Responsibilities

SR. No	Name of Member	Role	Responsibilities
1	Shaikh Azamali	Team Leader	Database management, web development
2	Khan Khalid	Team Member	Machine Learning development
3	Ansari Faiz	Team Member	Documentation, UI/UX Design

3.3 Assumptions and Constraints

1. People who are using this application have no prior knowledge of given movie.

2. This application is useful for the company who gives reviews on movies and present critics to the audience.

3. user of this application may or may not have the information about the preferred movies.

4. user would like to see all the information of desired movies on this site solely no need to go to any other website.

5. this application supports all type of OS and Browser.

3.4   Project Management Approach

1. Planning of project.

2. Defining the scope of the project.

3. Estimation of time and It's management.

4. Creating Gantt Charts and properly assigning tasks to members.

5. Reporting the progress of project with the guide.

3.5   Ground Rules for the Project

1. Properly planning and gathering relevant information is very important.

2. Developing a Blueprint of the project and work accordingly.

3. All the members should report to the guide whenever required.

4. Setting up small goals every week.

5. Achieving the small goal within that span of time.

6. Keeping tracks of the progress towards project.

3.6   Project Budget

1. It is a cost-effective project.

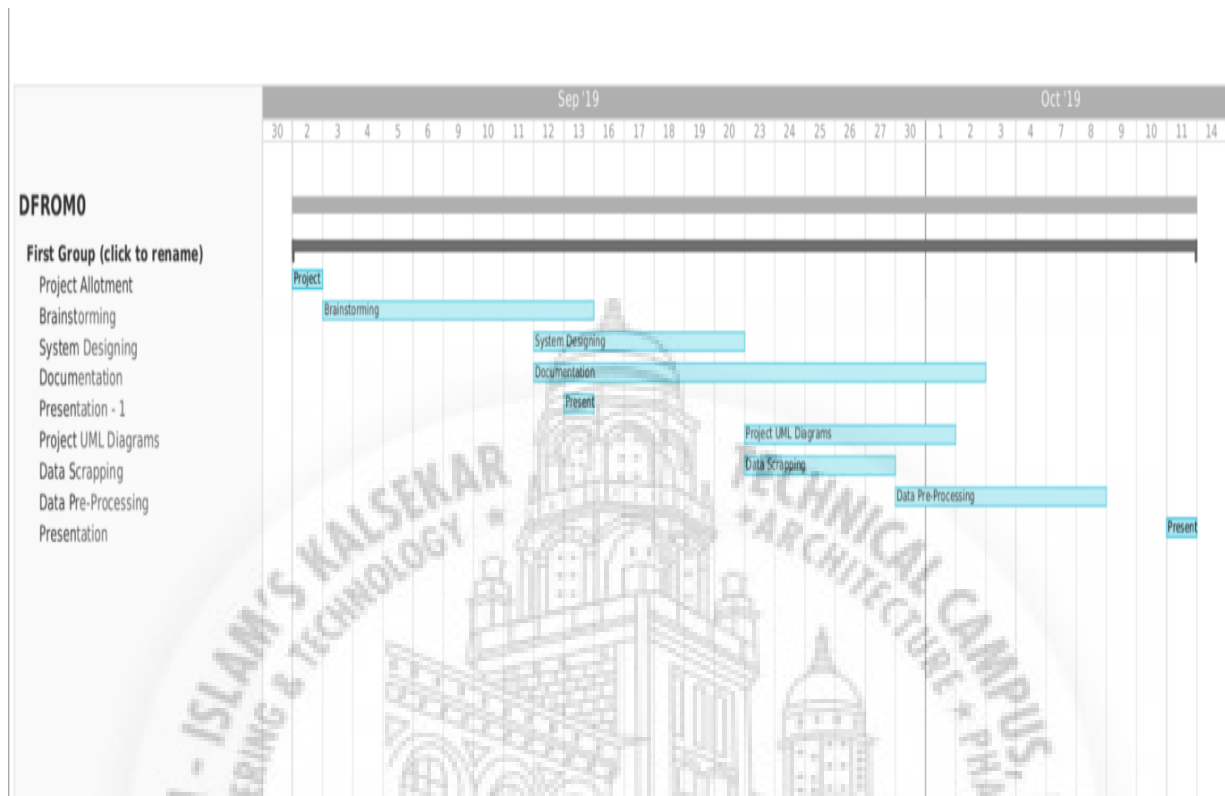2. Almost no expenses were spent during development.

3.7 Project Timeline

Figure 3.1: Project Timeline 1

Figure 3.2: Project Timeline 2

Chapter 4

Software Requirements Specification

4.1   Overall Description

4.1.1   Product Perspective

The product is an application based system implementing client-server model. This application provides service to the critics as well as users. This system is depend on IMDB for only movie's information. The main outcome of this system is to deliver unbiased information and rating of movies which could be altered by any means. Movie's information will be fetched from IMDB database so the rating and information of movies will be genuine. Machine learning algorithms will be applied on movies reviews and ratings, the result of the process will be shown to user. User will have all the details of the movies and system's calculated information.

4.1.2   Product Features

Information of movies are collected from IMDB database so user can even search/see about of any movie. Vast database provided to user with convenient user Interface so user could easily enjoy the system. Machine learning algorithms are used for accurate and precise result. Latest technology used to develop system so user would never feel any type of error or down flaw.

4.1.3   User Classes and Characteristics

This product is specifically developed for those who wants the fact checks and do not to waste their time on watching movies based on false reviews. It'll help the user to distinguish between good and bad movies.

4.1.4   Operating Environment

Software Requirement:

- Browser with HTML5 support.

Hardware Requirement:

- Internet Connectivity

- Minimum 2 GHz Processor.

- Minimum 2GB Ram.

4.1.5 Design and Implementation Constraints

System works on machine learning algorithm so it won't give 100 percent accuracy. Not giving full accuracy will be a big down flaw for certain amount of time.

4.2 System Features

Information of movies are collected from IMDB database so user can even search/see about of any movie. Vast database provided to user with convenient user Interface so user could easily enjoy the system. Machine learning algorithms are used for accurate and precise result. Latest technology used to develop system so user would never feel any type of error or down flaw.

4.2.1 System Feature

The information about any movies will be fetched from the IMDB website.

Description and Priority

This feature is the initial phase of the system. When user search for any movie information the query will passed to server and it'll in return fetch the data of desired movie from IMDB website.

Stimulus/Response Sequences

Stimulus: User will enter the name of the movie.
Response: Query then passed to the server, User will get the wait message.
Stimulus: server will search for the movie into IMDB database.
Response: Movie found, then server will fetch the data of from IMDB and processed the data to pre-processing module.

4.2.2 System Feature

Strong Machine learning algorithms provides almost accurate result.

Description and Priority

This feature is the heart of the system. Machine learning algorithm perform operation on thousands of reviews of single movie. By this feature user will get result filtered from thousands of reviews and based on the process the rating is calculated.

Stimulus/Response Sequences

Stimulus: The fetched data from IMDB will then processed to preprocessing filteration method.
Response: The filtered data without any uninformative will be the resullt.
Stimulus: Filtered data or Features will then processed to the main module which is algorithm module.
Response: The result which is calculated rating will be the outcome of the process.

Functional Requirements

REQ-1: Should there be Internet Connection.
REQ-2: Access of location and storage should be permitted.

4.3 External Interface Requirements

4.3.1 User Interfaces

This system provide a very proficient User Interface. it's lightweight web application so it will not do any harm memory. Using web technologies like laravel, php and JavaScript this system is developed. It has one front page with search bar and some logs activity like most searched movie and recent searched movie. Latter result page where result will show after search.

4.3.2 Hardware Interfaces

This application required some permission and requirement like storage and location permission to access and temporarily store data. Internet connectivity is must to connect to this product.

4.3.3 Software Interfaces

The software that needed to develop this product are spyder, which is IDE to develop programs. All the machine learning programs are made using IDE for this application. Android Studio is also used to develop the application for the same which is currently in production. Visual Studio Code is used to develop the UI of the application and Mongo DB is also used. It uses JSON like documents with schema.

4.3.4 Communications Interfaces

A welcome page is the start page with the High end User Interface for the users so users would feel the ease to use application. Latter and last page is used to show the result to the user.

4.4 Nonfunctional Requirements

4.4.1 Performance Requirements

Performance of overall system is very efficient and well optimize. From the time taken to capture and store the user query and then fetch the information of movies to calculate the final result is well organized. Processing of algorithm may take some time becuase it'll work on thousands of reviews at same time.

4.4.2 Safety Requirements

This system does not contain any critical data. Still it provide The databases that are accessed are locally executed.

4.4.3 Security Requirements

This application does not store or ask for user details so there would be no security issue. This application manage record of history for it's own to give better timing to user.

Chapter 5

System Design

5.1 System Requirements Definition

The main purpose for developing this application is to distinguish between fake and genuine reviews and based on that the rating will be generated. Application should gather all the resources for the process and to provide accurate result system should make precise decisions using machine learning algorithms, to do so one has to come up with algorithm which give more accurate result.
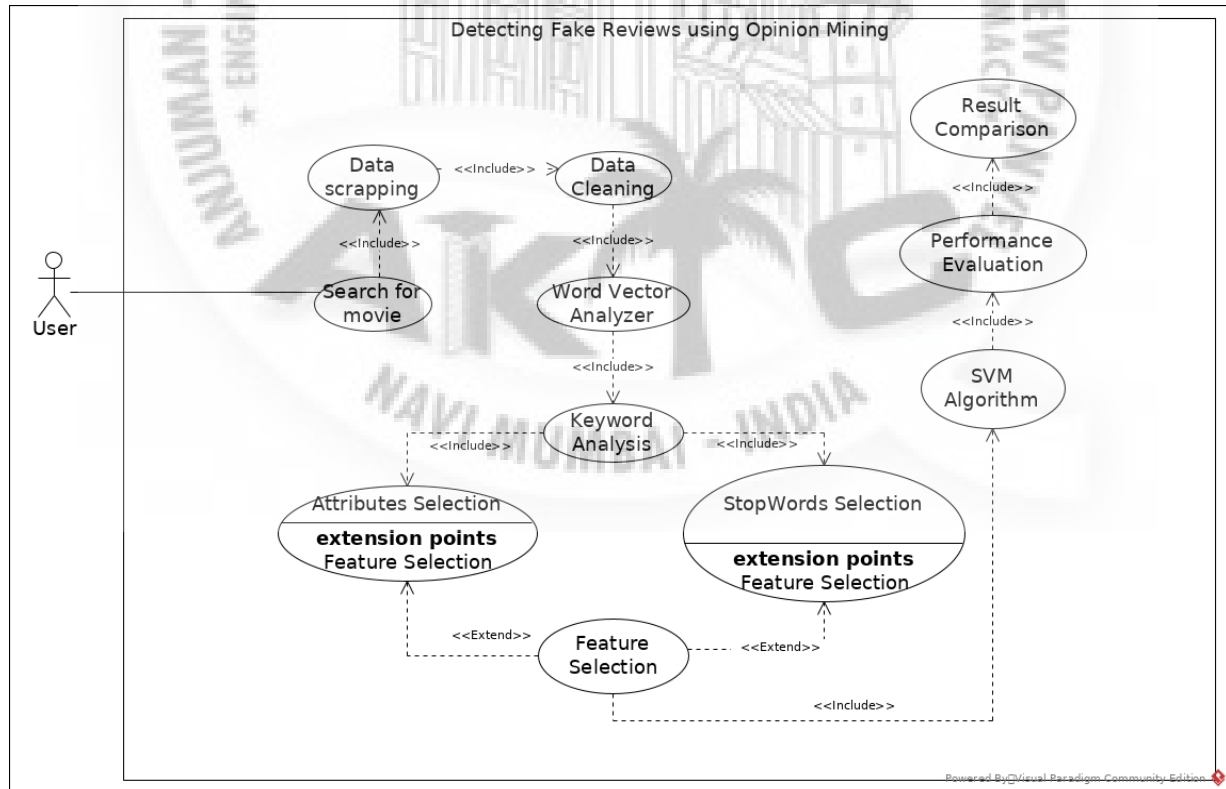
Use-case Diagram

Figure 5.1: Use Case Diagram

A use case diagram at its simplest is a representation of a user's interaction with the system that shows the relationship between the user and the different use cases in which the user is involved. In our system User will interact with use cases like Send query, show result.

Data-flow Diagram

A data-flow diagram is a way of representing a flow of a data of a process or a system. The DFD also provides information about the outputs and inputs of each entity and the process itself. Given below is Level 0 Level 1 and Level 2 DFD of system.
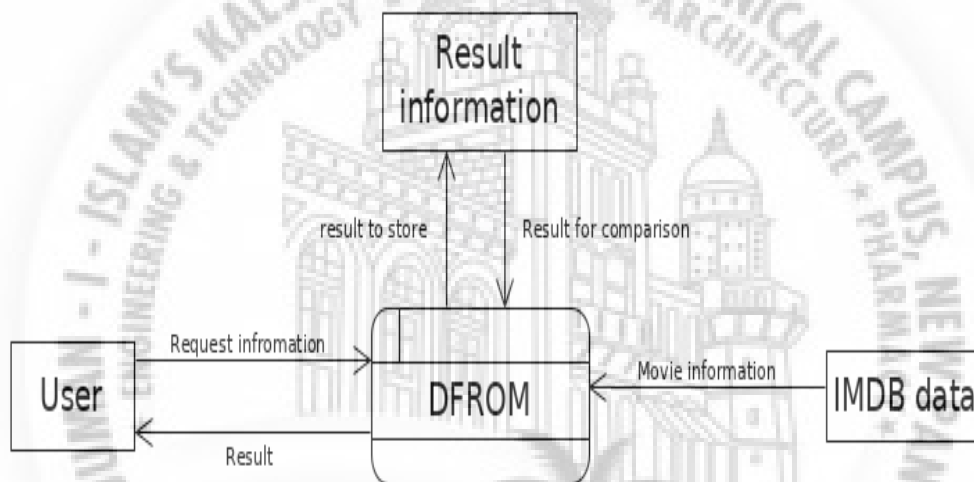
Figure 5.2: Level 0 Data Flow Diagram
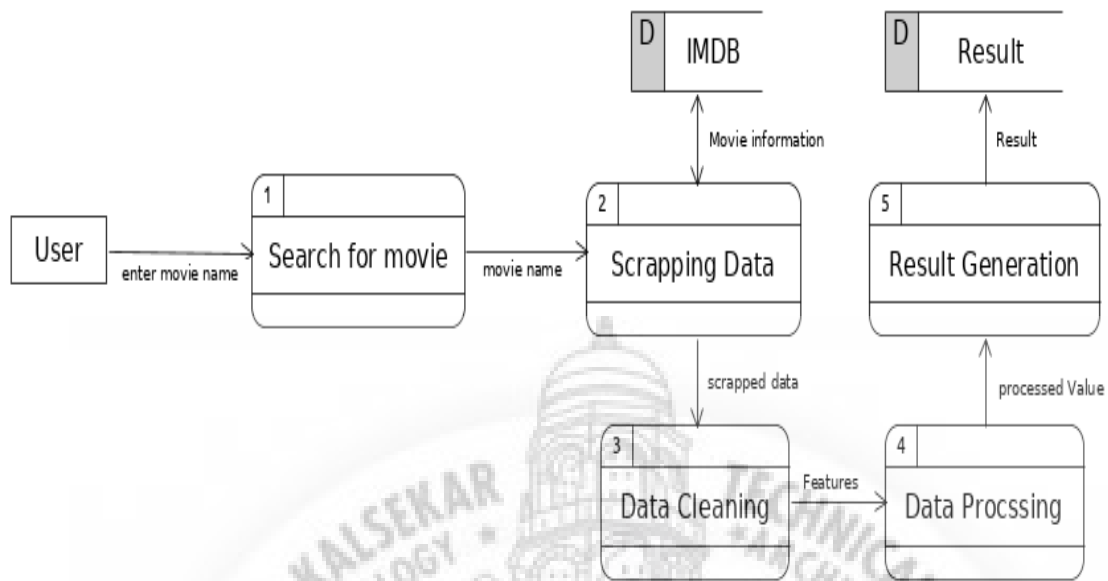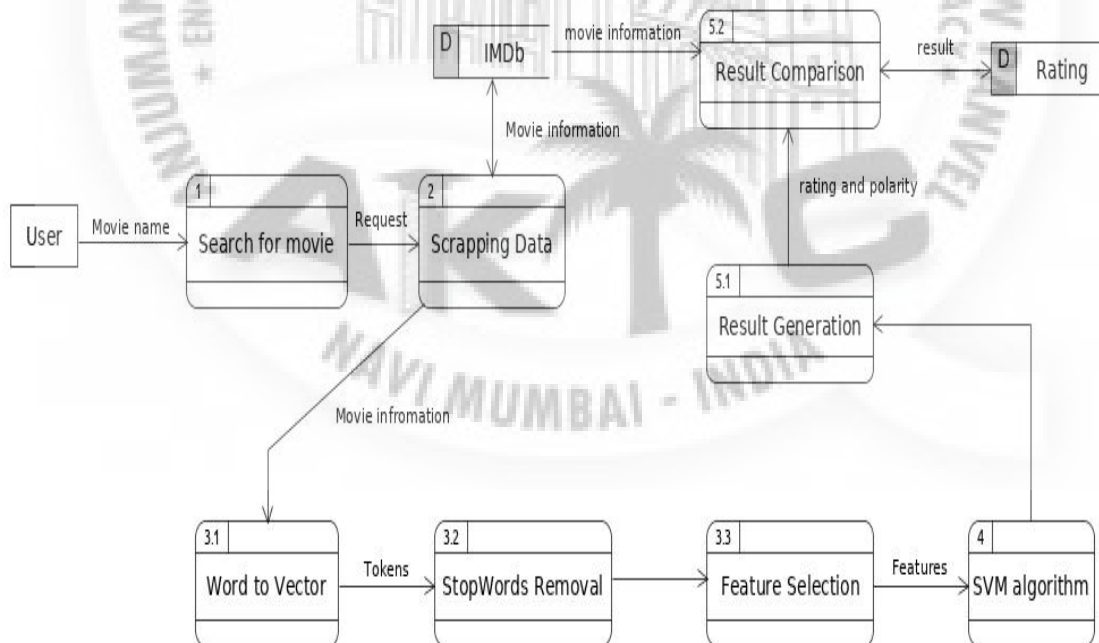
Figure 5.3: Level 1 Data Flow Diagram

Figure 5.4: Level 2 Data Flow Diagram

5.1.1   System requirements (non-functional requirements)

These are non-functional system properties such as availability, performance and safety etc. They define functions of a system, services and operational constraints in detail.

1. Time Efficiency - The execution of the system takes very less time.

2. Portability - The system needs to be work on every platform and browser and should be responsive to run on any size of device.

3. Performance - The system will work fine in every device.

4. Multi User System - This application can be used in different devices independently.

5. User Interactive - The application have a very simple User Interface.

6. Security - There are no vulnerability in its security.

5.2   System Architecture Design
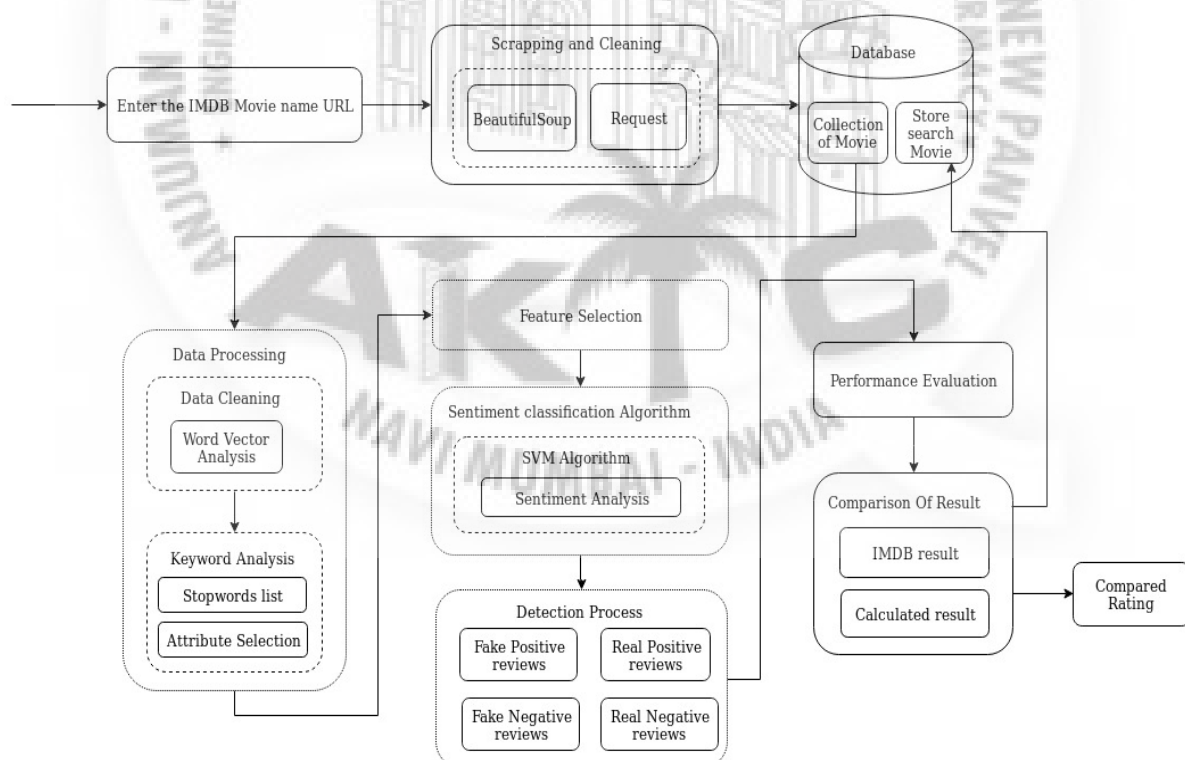
Figure 5.5: System Architecture

A system architecture is the conceptual model that defines the structure, behavior,and more views of a system. An architecture description is a formal description

and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

5.3   Sub-system Development

5.3.1   Data Scrapping

A user search for movie information by giving a movie name in search box to get the movie rating of that movie.User's requested movie name goes to the data scrapping module and then we apply our Data scrapping algorithm to fetch the user's desired movie information.Now this raw movie information will be simplified according to our need and it will proceed on our next module which is Data pre-processing module
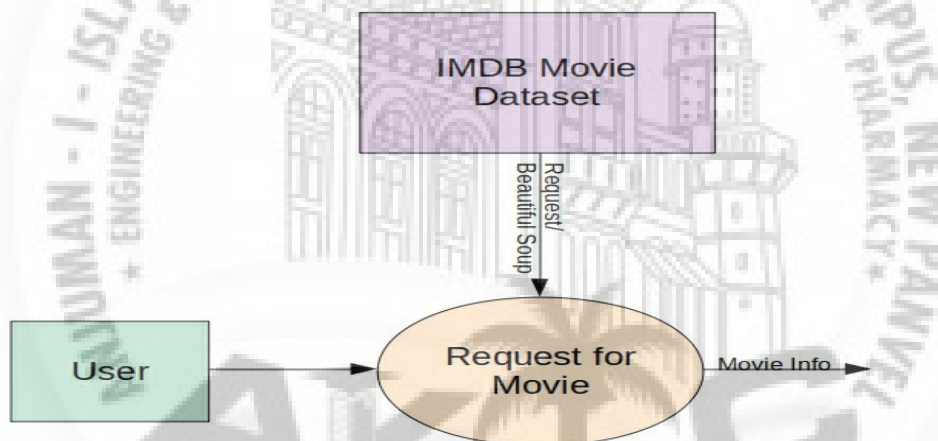

Figure 5.6: Data Scrapping Module

5.3.2   Data pre-processing

In Data pre-processing module we take the information of user search movie from the Data Scrapping module, on that information we apply algorithm for cleaning the data using various methods like word tokenizer, stop words removal and feature selection in this module.
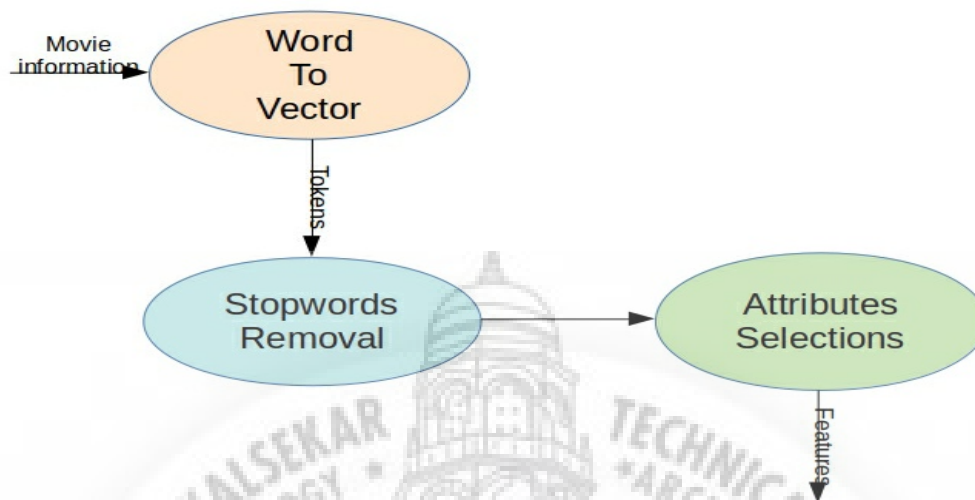
Figure 5.7: Data pre-processing Module

5.3.3   Comparison Result

After pre-processing the data we apply our algorithm to detect the fake reviews or rating in percentage and calculate the actual rating of the user search movie and show them our rating with IMDB rating.
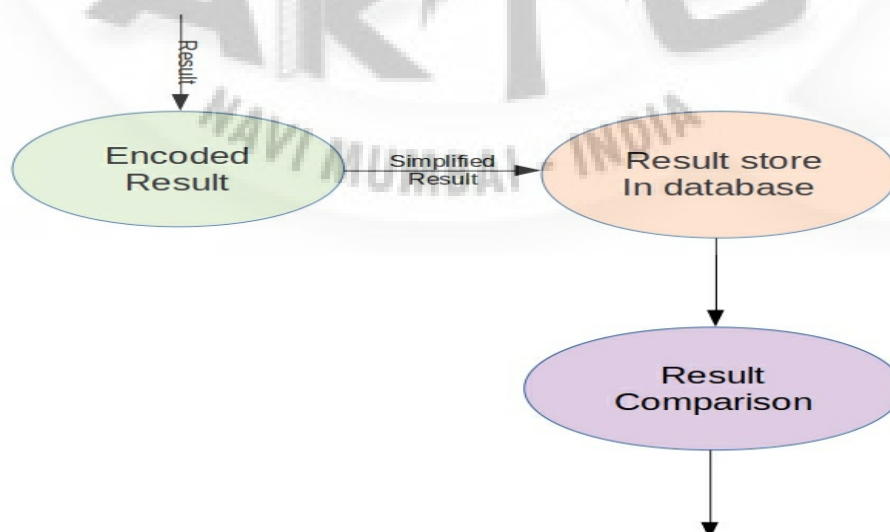

Figure 5.8: Comparison Result Module

5.4 Systems Integration

System integration (SI) is an engineering process or phase concerned with joining different subsystems or components as one large system. It ensures that each integrated subsystem functions as required. SI is also used to add value to a system through new functionalities provided by connecting functions of different systems

5.4.1 Class Diagram

In software engineering, a class diagram in the Unified Modeling Language is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations, and the relationships among objects. In our system there are classes like result optimizer, comparison, movie information, reviews, pre-processing, processing.
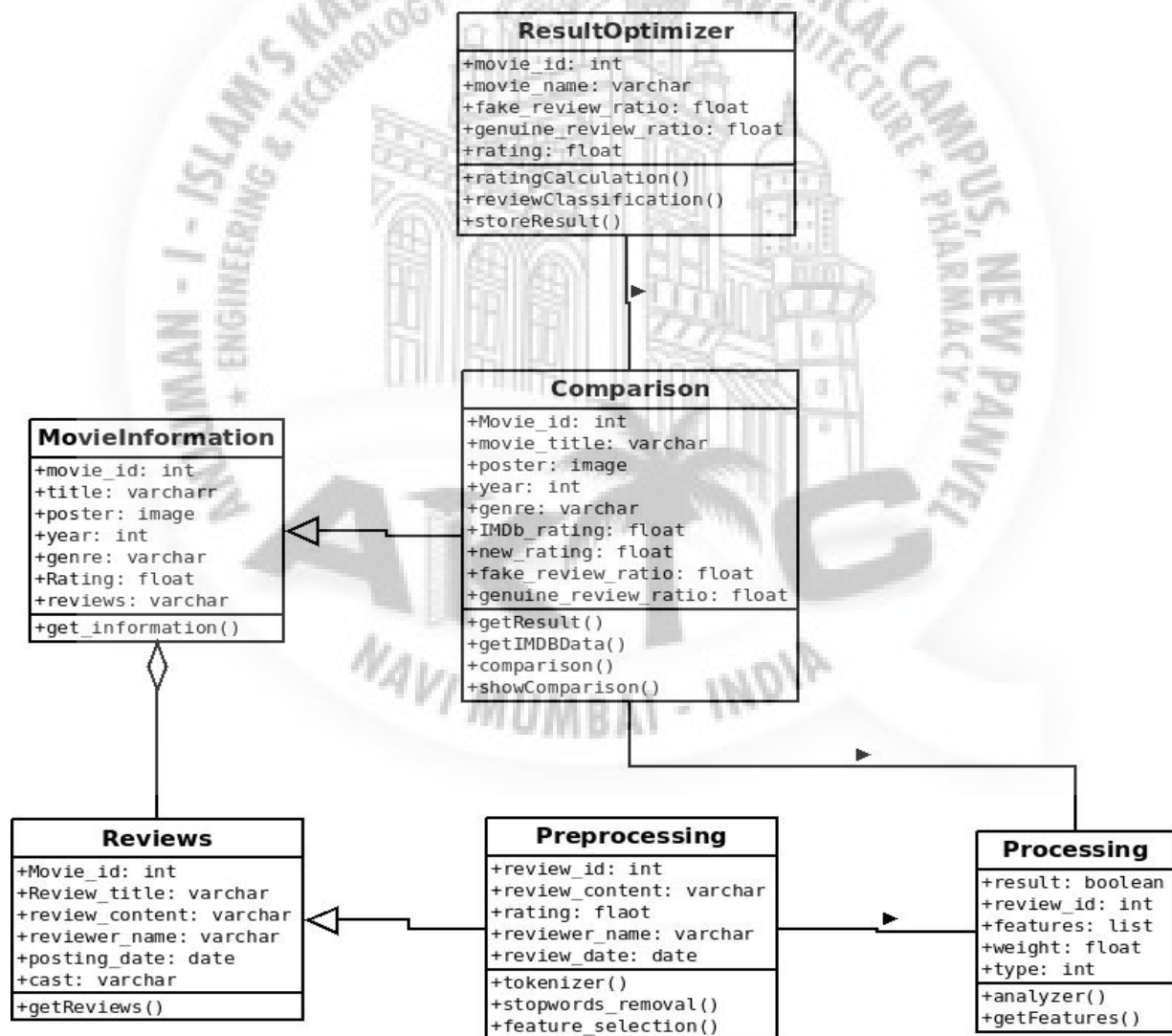

Figure 5.9: Class Diagram

5.4.2 Sequence Diagram

A sequence diagram shows object interactions arranged in time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario.
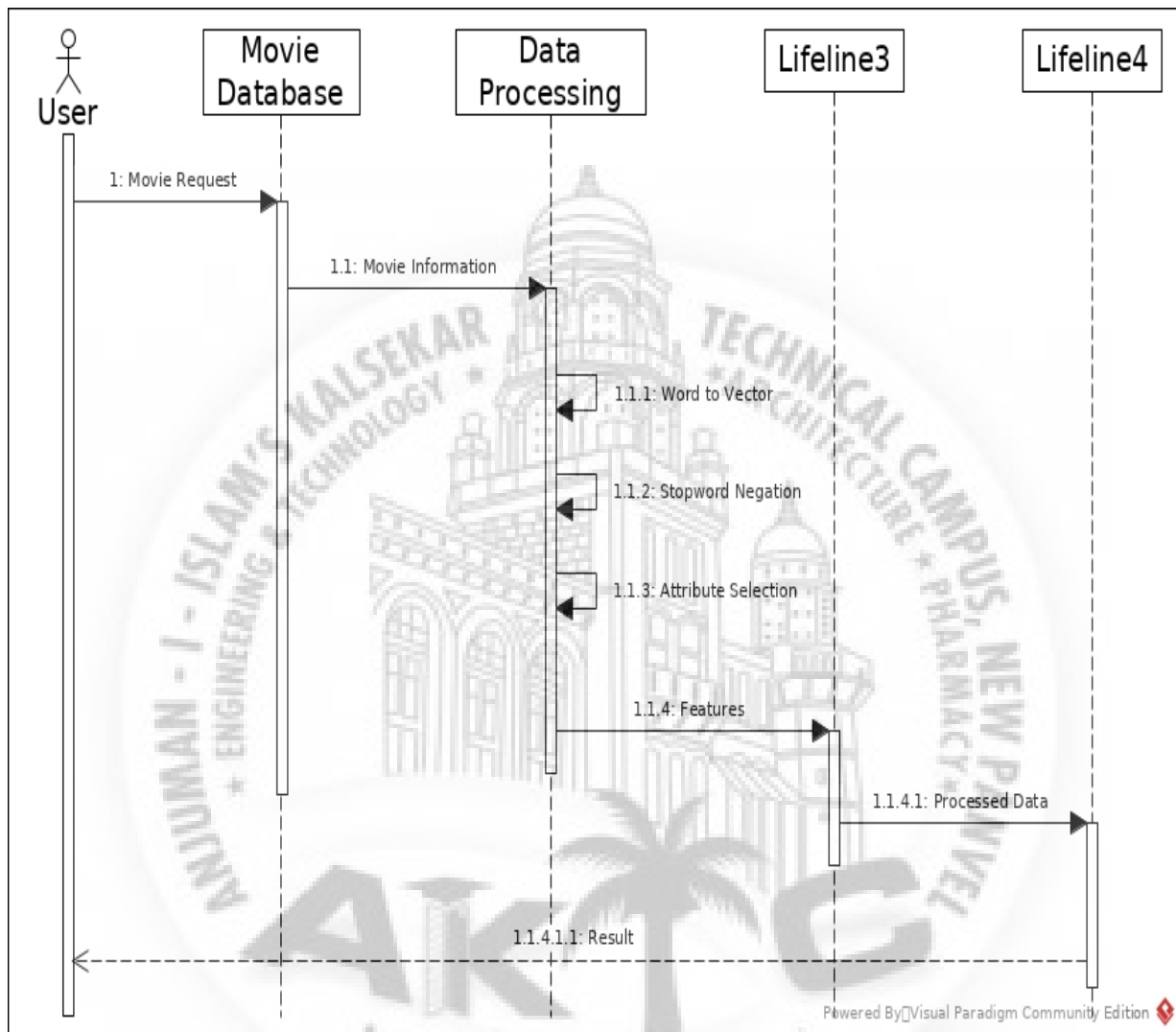

Figure 5.10: Sequence Diagram

Chapter 6

Implementation

6.1   Pre processing Module

This module is implemented using spyder IDE. In this module data filtration is being done.

```python
import nltk
import re
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer

paragraph = '''Now in no way is this show as bad as people thought it would nor
    is it as good as some of the other reviewers here think it is either.
        First of, Henry Cavill as Geralt is the best part of this show 100%. No
            doubt about that. The show is worth watching just for his portrayal.
            He really cares about the character and it shows.
        Pros:
        Henry Cavill as Geralt. Can't say this enough.
        The fight Choreography is just insane. GOT could have really used some
            of the insane fight sequences from the witcher series.
        The Really amazing moments between Jaskier(Dandelion) and Geralt. They
            did Jaskier perfectly in my opinion.
        Roach
        Cons:
        Really bad casting choices like Triss, Foltest and the Driads. Also they
            butchered an entire portion from the books which was one of my
            favorite parts from the Last Wish actually.
        Nilfgaard Armour. Like who in the right mind actually thought that it
            looked even remotely good. I wanted to skip the parts where they
            came up in the screens.
        Very unfriendly to people not familiar with the books or games. This is
            a fan-service. Not entirely a con but you get the point.
        DESTINY THIS DESTINY THAT. GOD STFU. I know destiny is a main focal
            point in the books but it was over done even there. I just got so
            tired of that word after watching for a while. All in all its worth
            the watch if you're a fan of the Witcher games/books. I can't praise
             Henry enough for his performance though. Seriously.'''

wordnet = WordNetLemmatizer()
sentences = nltk.sent_tokenize(paragraph)
corpus = []
for i in range(len(sentences)):
    review = re.sub('[^a-zA-Z]', ' ', sentences[i])
    review = review.lower()
    review = review.split()
```

```
27      review = [wordnet.lemmatize(word) for word in review if not word in set(
            stopwords.words('english'))]
28      review = ' '.join(review)
29      corpus.append(review)
30
31  cv = TfidfVectorizer()
32  X = cv.fit_transform(corpus).toarray()
33  print(X)
```

6.2 Data Scrapping Module

Movie first requested by user then the query helps in scrapping/fetching data. In this module data is being fetched/scrapped from IMDB database.

```python
import requests
import pandas as pd
from bs4 import BeautifulSoup

url = "https://www.imdb.com/search/title/?pf_rd_m=A2FGELUUNOQJNL&pf_rd_p=
    cd28805a-4e91-4f0f-b066-0db5ff4dd1a7&pf_rd_r=4FNNKR25P2Y5EQETE2G3&pf_rd_s=
    right-6&pf_rd_t=15506&pf_rd_i=boxoffice&ref_=chtbo_gnr_13&genres=horror&
    explore=title_type,genres"
print(type(url))

r = requests.get(url)
print(r)

data = pd.DataFrame(columns=['Poster', 'Title', 'Year', 'Genere', 'Rating', '
    ReviewLink'])

bs = BeautifulSoup(r.text,'html.parser')
#print(bs)
table = bs.find('div',{'class':'article'})
mdata = table.find_all('div',{'class':'lister-item mode-advanced'})
#print(mdata)

#       nextPage = 'https://www.imdb.com' + table.find('div',{'class':'desc'}).
    find('a')['href']
#       print(nextPage)

# URL for next page data

next = []
for i in range(1,1001,50):
    link = "https://www.imdb.com/search/title/?genres=horror&start="+str(i)+"&
        explore=title_type,genres"
    next.append(link)


x = 0

for url in next:

    r = requests.get(url)
    # print(r)

    bs = BeautifulSoup(r.text,'html.parser')
    #print(bs)


    table = bs.find('div',{'class':'article'})
    mdata = table.find_all('div',{'class':'lister-item mode-advanced'})


    for mlist in range(len(mdata)):
        data.at[x,'Poster'] = mdata[mlist].find('div',{'class':'lister-item-
            image float-left'}).find('img').get('src')

        # print(data.at[x,'Poster'])
```

```
50    data['Title'][x] = mdata[mlist].find('div',{'class':'lister-item-content
         '}).find('h3',{'class':'lister-item-header'}).find('a').text
51
52    data['Year'][x] = mdata[mlist].find('div',{'class':'lister-item-content'
         }).find('h3',{'class':'lister-item-header'}).find('span',{'class':'
         lister-item-year text-muted unbold'}).text.replace('(','').replace('
         )','').strip()
53
54    data['Genere'][x] = mdata[mlist].find('div',{'class':'lister-item-
         content'}).find('p',{'class':'text-muted'}).find('span',{'class':'
         genre'}).text
55    #print(genere)
56
57    try:
58        data['Rating'][x] = mdata[mlist].find('div',{'class':'lister-item-
             content'}).find('div',{'class':'ratings-bar'}).find('div',{'
             class':'inline-block ratings-imdb-rating'}).find('strong').text
59        #print(data['Rating'][mlist])
60
61    except:
62        rating = "NA"
63
64        #print('Except block')
65
66    data['ReviewLink'][x] = 'https://www.imdb.com' + mdata[mlist].find('div'
         ,{'class':'lister-item-image float-left'}).find('a')['href']+'
         reviews/'
67    print(data['ReviewLink'][x])
68    x=x+1
69
70    data.to_csv('IMDBHorrorMovies.csv')
71 print('done')
```

```
1
2 import requests
3 import pandas as pd
4 from bs4 import BeautifulSoup
5 from selenium import webdriver
6
7 url = r'https://www.imdb.com/title/tt5180504/reviews/'
8
9 wb = webdriver.Chrome("/home/azam/Downloads/chromedriver")
10 wb.get(url)
11
12 r = requests.get(url)
13
14 data = pd.DataFrame(columns=['Star','review_title','reviewr_name','review_date',
     'review_content'])
15
16 bs = BeautifulSoup(r.text,'html.parser')
17
18 main = bs.find('section',{'class':'article'})
19 rlist = main.findAll('div',{'class':'lister-item-content'})
20
21 Total_Review = main.find('div',{'class':'header'}).find('span').text
22
23
24
25 x = 0
26
```

```python
27  for rl in range(len(rlist)):
28      try:
29          data.at[x,'Star'] = rlist[rl].find('span',{'class':'rating-other-user-
                rating'}).find('span').text
30          #print(data.at[x,'Star'])
31      except:
32          star = "NA"
33
34      try:
35          data['review_title'][x] = rlist[rl].find('a',{'class':'title'}).text.
                replace('\r'," ")
36          print(data['review_title'][x])
37      except:
38          review_title = "NA"
39
40      try:
41          data['reviewr_name'][x] = rlist[rl].find('div',{'class':'display-name-
                date'}).find('span',{'class':'display-name-link'}).find('a').text
42          #print(data['reviewr_name'][x])
43      except:
44          reviewr_name = "NA"
45
46      try:
47          data['review_date'][x] = rlist[rl].find('div',{'class':'display-name-
                date'}).find('span',{'class':'review-date'}).text
48          #print(data['review_date'][x])
49      except:
50          review_date = "NA"
51
52      try:
53          data['review_content'][x] = rlist[rl].find('div',{'class':'content'}).
                find('div',{'class':'text show-more__control'}).text
54          #print(data['review_content'][x])
55
56      except:
57          review_content = "NA"
58      x=x+1
59
60  data.to_excel('IMDBHorrorMoviesReviews.xls')
61  print('done')
62
63
64
65  # In[4]:
66
67
68  import requests
69  import pandas as pd
70  from bs4 import BeautifulSoup
71  from selenium import webdriver
72  import time
73
74  url = r'https://www.imdb.com/title/tt5180504/reviews/'
75
76  wb = webdriver.Chrome("/home/azam/Downloads/chromedriver/chromedriver")
77  wb.get(url)
78
79
80  r = requests.get(url)
81
82  data = pd.DataFrame(columns=['Star','review_title','reviewr_name','review_date',
```

```
        'review_content'])
83
84  bs = BeautifulSoup(r.text,'html.parser')
85
86  main = bs.find('section',{'class':'article'})
87  rlist = main.findAll('div',{'class':'lister-item-content'})
88
89  Total_Review = main.find('div',{'class':'header'}).find('span').text
90
91  x_path_main = '//*[@id="load-more-trigger"]'
92
93  x = 0
94  rl = 0
95
96  while True:
97      #for rl in range(len(rlist)):
98      try:
99          data.at[x,'Star'] = rlist[rl].find('span',{'class':'rating-other-user-
                rating'}).find('span').text
100         #print(data.at[x,'Star'])
101     except:
102         star = "NA"
103
104     try:
105         data['review_title'][x] = rlist[rl].find('a',{'class':'title'}).text.
                replace('\r'," ")
106         #print(data['review_title'][x])
107     except:
108         review_title = "NA"
109
110     try:
111         data['reviewr_name'][x] = rlist[rl].find('div',{'class':'display-name-
                date'}).find('span',{'class':'display-name-link'}).find('a').text
112         #print(data['reviewr_name'][x])
113     except:
114         reviewr_name = "NA"
115
116     try:
117         data['review_date'][x] = rlist[rl].find('div',{'class':'display-name-
                date'}).find('span',{'class':'review-date'}).text
118         #print(data['review_date'][x])
119     except:
120         review_date = "NA"
121
122     try:
123         data['review_content'][x] = rlist[rl].find('div',{'class':'content'}).
                find('div',{'class':'text show-more__control'}).text
124         #print(data['review_content'][x])
125
126     except:
127         review_content = "NA"
128
129     x = x+1
130     rl += 1
131     if rl==25:
132         try:
133             x_path_main = '//*[@id="load-more-trigger"]'
134             xpath = x_path_main
135             wb.find_element_by_xpath(xpath).click()
136             time.sleep(4)
137             rl = 0
```

```
138
139            bs = BeautifulSoup(wb.page_source,'html.parser')
140            main = bs.find('section',{'class':'article'})
141            rlist = main.findAll('div',{'class':'lister-item-content'})
142
143            Total_Review = main.find('div',{'class':'header'}).find('span').text
144
145
146        except:
147            break
148
149
150 data.to_excel('IMDBHorrorMoviesReviews.xls')
151 print(len(data))
152 wb.close()
153 print('done')
```

6.3   Movie information fetching from database module

n This code helps in connection to database then fetch movie which is stored in our database.

```php
<?php
    require "conn.php";

    // $movie_name = $_POST["movie_name"];
    $movie_name = "The Witcher";

    if($conn){

        $movieDetails = "SELECT * FROM `movies` WHERE `title` LIKE '$movie_name'";
        $movieQuery = mysqli_query($conn, $movieDetails);

        if(mysqli_num_rows($movieQuery) > 0){

            // echo "movies tables" . "<br>";

            $reviewDetails = "SELECT * FROM `reviews` WHERE `movie_name` LIKE '$movie_name'";
            $reviewQuery = mysqli_query($conn, $reviewDetails);

            if(mysqli_num_rows($reviewQuery) > 0){

                // echo "reviews table" . "<br>";

                $resultDetails = "SELECT * FROM `result` WHERE `movie_name` LIKE '$movie_name'";
                $resultQuery = mysqli_query($conn, $resultDetails);

                if(mysqli_num_rows($resultQuery) > 0){
                    // echo "result table" . "<br>";

                    while($row = mysqli_fetch_assoc($resultQuery)){

                        $movie_name = $row['movie_name'];
                        $imdb_rating = $row['imdb_rating'];
                        $calculated_rating = $row['calculated_rating'];
                        $perc_of_neg_reviews = $row['perc_of_neg_reviews'];
                        $perc_of_pos_reviews = $row['perc_of_pos_reviews'];
                    }

                } else{
                    echo "No Result data Found!";
                }
            } else {
                echo "Oops! No Data found";
            }
        } else {
            echo "Movie not found";
        }
    } else {
        echo "Connection Error!";
    }
?>
```

Chapter 7

Screenshots of project

7.1 feature selection window

Figure 7.1: Feature Selection Window

7.2   System Front page

Figure 7.2: Search page

Figure 7.3: Search page

Figure 7.4: Search page

7.3   System Result page
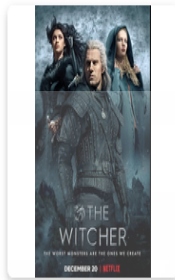
Figure 7.5: Result page

Details BETA

Most positive review ❓

Amazing Series.
Superb fantasy.
This show simply a masterpiec.
The screenplay, the characterization
is simply superb.

Most Negative review ❓

Very disappointing.
An overrated series.
overrated
Just boring

DFROM

Copyright 2020 DFROM Inc

COMPANY

DFROM.com   Careers   Press   Blog

Support

Contact   FAQ   Terms of Use   Privacy Policy

Reviews

Last Analysis   Analysis Blog

Figure 7.6: Result page

Chapter 8

Conclusion and Future Scope

8.1 Conclusion

Detecting Fake Reviews Using Opinion Mining (DFROM), is an movie review filtering system and it is cross platform digital solution to tackle the detecting fake reviews of movies.We use the IMDB data to detect the fake reviews of the movie. As per the Previous researchers no one use the Original Data to detect the fake reviews all previous researchers use the auto-mated generated data to detect the fake reviews.We are using Sentiment Analysis to determine the movie reviews whether are fake or not. In sentiment Analysis we are using SVM (Support Vector Machine) Algorithm to get the accurate result.Previous researchers are used and implemented numbers of algorithms in their paper they have concluded that SVM is most promising among them. So we are using directly SVM based on their result it will apparently save processing time of our system. Though there are several applications built similar to ours, but every application has its own pros and cons. After having a thorough reading of the other applications research papers, and based on the requirements of the customers, our applications thrives to overcome the drawbacks of the other applications.

8.2 Future Scope

- Add more variety of features for user to choose from.

- A complete fully flex mobile application for the system so user will access it from mobile.

- The users will get automatic suggestions depending on their choice of movie and the liking.

- notification and news of latest critics to user.

References

[1] *Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques*; Elshrif Elmurngi, Abdelouahed Gherbi, cole de Technologie Supérieure, June 2018

[2] *Fake Review Detection using Classification*; Neha S. Chowdhary, Anala A. Pandit, PhD, International Journal of Computer Applications, June 2018.

[3] *An empirical study on detecting fake reviews using machine learning techniques*; lshrifElmurngi, Abdelouahed Gherbi, École de Technologie Supérieure, August 2017

[4] *Feature Selection in Sentiment Analysis*; Olena Kummer, Jacques Savoy, Institut d'informatique, Université de Neuchâtel Rue Emile Argand 1, 2000 Neuchâtel, Suisse

[5] *Machine Learning Docs*; `https://ml-cheatsheet.readthedocs.io`

[6] *Scikit-learn Machine learning Docs*; `https://scikit-learn.org`

Achievements

Publications

Detecting Fake Reviews Using Opinion Mining; Azamali Shaikh, Khalid Ahmed Khan , Mohd Faiz Ansari, March 2020