

A PRELIMINARY REPORT ON

“Sentimental Analysis On Audio And Video ”

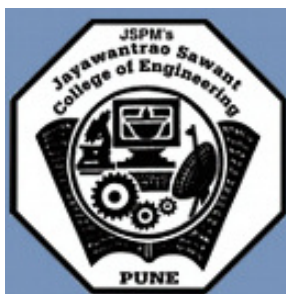
**SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY,PUNE
IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE**

OF

BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

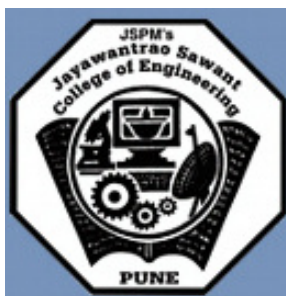
SUBMITTED BY

Ms.Shivani Raskar	Seat N0:B150404277
Mr.Vishal Waman	Seat N0:B150404303
Ms.Monali Yadav	Seat N0:B150404304



**DEPARTMENT OF COMPUTER ENGINEERING
JSCOE JAYAWANTRAO SAWANT COLLEGE OF ENGINEERING
HADAPSAR, PUNE 411028
SAVITRIBAI PHULE PUNE UNIVERSITY
[2018-19]**

CERTIFICATE



This is to certify that the Seminar Report entitles
“Sentimental Analysis On Audio And Video ”

Submitted by

Ms.Shivani Raskarr B150404277

Mr. Vishal Wamanr B150404303

Ms.Monali Yadavr B150404304

is a bonafide student of this institute and the work has been carried out by him/her under the supervision of **Dr.S.B.Chaudhari** and it is approved for the partial fulfillment of the requirement of Savitribai Phule Pune University, for the award of the degree of **Bachelor of Engineering** (Computer Engineering).

Dr.S.B.Chaudhari

Internal Guide

Dept. of Computer Engg.

Prof.N.B.Shardoor

Project Co-ordinator

Dept. of Computer Engg.

Dr.S.B.Chaudhari

H.O.D

Dept. of Computer Engg.

Dr.M.G.Jadhav

Principal

JSCOE,Pune

Prof.

External Examiner

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project report on Sentimental Analysis on Audio And video .

We would like to take this opportunity to thank our internal guide Dr.S B.Chaudhari for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful.

We are also grateful to Dr.S B.Chaudhari, Head of Computer Engineering De-partment, JSCOE for his indispensable support and suggestions.

Ms.Monali Yadav

Ms.Shivani Raskar

Mr.Vishal Waman

ABSTRACT

Sentimental Analysis is a reference to the task of Natural Language Processing that is NLP to determine that whether a text data contains objective information, subjective information and what information it expresses i.e., whether the point of view behind the text is positive, negative or neutral. Sentimental Analysis is the best way to evaluate people's opinion regarding a particular post. This paper mainly focuses on the several machine learning techniques which are used in analyzing the sentiments and in opinion mining which is present in the converted text data. Sentimental analysis with the combination of machine learning could be useful in predicting the product reviews and consumer attitude towards newly launched product. This paper takes the input as audio and video which is then converted into text and this text data is analyzed using suitable machine learning techniques. This paper presents a detailed survey of various machine learning methods and then compared with their accuracy, advantages, and limitations of each technique.

Keywords: Sentiment Analysis, Keyword spotting, Opinion, UT-Sentiment Audio Archive

Contents

ACKNOWLEDGEMENT	I
ABSTRACT	II
List of Figures	VI
1 INTRODUCTION	1
1.1 SENTIMENTAL ANALYSIS:	1
1.2 MOTIVATION:	2
1.3 PROBLEM STATEMENT:	3
1.4 PROJECT SCOPE:	3
1.5 METHODOLOGIES OF PROBLEM SOLVING:	3
2 LITERATURE SURVEY	5
3 SOFTWARE REQUIREMENT SPECIFICATION	7
3.1 ASSUMPTIONS AND DEPENDENCIES	7
3.2 FUNCTIONAL REQUIREMENTS	7
3.2.1 REGISTRATION AND LOGIN	7
3.2.2 AUDIO AND VIDEO TRACKING	8
3.3 EXTERNAL INTERFACE REQUIREMENTS	8
3.3.1 USER INTERFACES	8
3.3.2 HARDWARE INTERFACES	8

3.3.3	SOFTWARE INTERFACES	8
3.3.4	COMMUNICATION INTERFACES	9
3.4	NONFUNCTIONAL REQUIREMENTS	9
3.4.1	PERFORMANCE REQUIREMENTS	9
3.4.2	SAFETY REQUIREMENTS	9
3.4.3	SECURITY REQUIREMENTS	10
3.4.4	SOFTWARE QUALITY ATTRIBUTES	10
3.5	SYSTEM REQUIREMENTS	10
3.5.1	DATABASE REQUIREMENTS	10
3.5.2	SOFTWARE REQUIREMENTS (PLATFORM CHOICE) . . .	10
3.5.3	HARDWARE REQUIREMENTS	10
3.6	ANALYSIS MODELS	11
4	SYSTEM DESIGN	13
4.1	SYSTEM ARCHITECTURE:	13
4.2	MATHEMATICAL MODEL	15
4.3	DATA FLOW DIAGRAMS	16
4.4	UML DIAGRAMS:	17
4.4.1	USE CASE DIAGRAM:	17
4.4.2	ACTIVITY DIAGRAM:	18
4.4.3	SEQUENCE DIAGRAM:	19
4.4.4	CLASS DIAGRAM:	20

5	PROJECT PLAN	21
5.1	PROJECT ESTIMATE	21
5.1.1	RECONCILED ESTIMATES	21
5.1.2	PROJECT RESOURCES	22
5.2	RISK MANAGEMENT	22
5.2.1	RISK IDENTIFICATION	23
5.2.2	RISK ANALYSIS	23
5.2.3	OVERVIEW OF RISK MITIGATION,MONITORING,MANAGEMENT	23
5.3	PROJECT SCHEDULE	24
5.3.1	PROJECT TASK SET	24
5.3.2	TASK NETWORK	25
5.3.3	TIMELINE CHART	25
5.4	5.4 TEAM ORGANIZATION	25
5.4.1	TEAM STRUCTURE	25
5.4.2	MANAGEMENT REPORTING AND COMMUNICATION .	26
6	PROJECT IMPLEMENTATION	27
6.1	OVERVIEW OF PROJECT MODULES	27
6.2	TOOLS AND TECHNOLOGIES USED	28
6.2.1	DJANGO	28
6.3	ALGORITHM DETAILS	30
6.3.1	VADER	30

7	SOFTWARE TESTING	33
7.1	TYPE OF TESTING	33
7.1.1	UNIT TESTING	34
7.1.2	INTEGRATION TESTING	34
7.1.3	SYSTEM TESTING	35
7.1.4	VALIDATION TESTING	36
7.1.5	WHITE BOX TESTING	36
7.1.6	BLACK BOX TESTING	37
7.1.7	GUI TESTING	37
7.2	TEST CASES TEST RESULTS	38
8	OUTCOMES AND SCREENSHOTS	39
9	CONCLUSION	42
9.1	CONCLUSION	42
9.2	FUTURE WORK	42
9.3	APPLICATIONS	43
9.4	REFERENCES	44

List of Figures

3.1	SDLC	12
4.1	Proposed System Design	13
4.2	Mathematical Model	15
4.3	DFD 0	16
4.4	DFD 1	17
4.5	Use Case Diagram	17
4.6	Activity Diagram	18
4.7	Sequence Diagram	19
4.8	Class Diagram	20
6.1	Sentiment Rating	31
6.2	Sentiment Metric	32
8.1	Home Page	39
8.2	Main Page	39
8.3	SignUp Page	40
8.4	SignIn Page	40
8.5	Video Input	41

8.6	:Result	41
-----	-------------------	----

Chapter 1

INTRODUNCTION

1.1 SENTIMENTAL ANALYSIS:

With the rapid development of e-commerce websites, people can live with the web. Nowadays people are used to reviewing the comments and posts on the product which are known as opinion, emotion, feeling, attitude, thoughts or behavior of the user. Sentimental Analysis is a method for identifying the ways in which sentiment is expressed in terms of texts.

In sentiment analysis, there are three classification levels: document-level classification, sentence-level classification, and aspect-level sentiment analysis. In document-level classification, the main aim is to classify an opinion in the whole document as positive and negative. It speculates the entire document as a single unit. The aim of the sentence-level analysis is to categorize emotion expressed in the respective sentences. In sentence-level, the basic step is to recognize the sentence as subjective or objective. Suppose sentence is subjective, it will decide whether it expresses a negative or positive opinion. In the aspect-level analysis, it aims to categorize the sentiment in respect of particular entities.

Our approach towards sentiment extraction from audio and video. Here in this proposed system, we use the machine learning algorithm for analysis of the text data that is converted from audio and video. An important characteristic of our technique is the ability to identify the individual contributions of the text features towards sentiment estimation or analysis. We evaluate the proposed sentiment estimation on both publically available text databases, audios, and videos.

Here, in this proposed system we take audio and video data from different sites such as youtube, facebook, etc. In this system audio extract, the text data from that text data sentiment analysis is to be applied. In the same way, video extracts the audio and from that, it extracts the text data then after sentiment analysis is to be applied. The output is getting in terms of positive and negative percentage after applying the machine learning algorithm.

1.2 MOTIVATION:

Motivation for Sentiment Analysis is two-fold. Both consumers and producers highly value customers opinion about products and services. Thus, Sentiment Analysis has seen a considerable effort from industry as well as academia.

1.3 PROBLEM STATEMENT:

Achieve Sentimental analysis over audio and video reviews of Products on social media. Social sites like Youtube, Facebook, Twitter contain a lot of unprocessed reviews that go unchecked for sentimental analysis.

OBJECTIVE:

Select a method for Video and Audio Extraction and create an algorithm and module to merge audio and video data to create more Effective and accurate sentimental analysis.

1.4 PROJECT SCOPE:

Now a days sentimental analysis of data is require in many fields like product rating e-commerce, government polls, reality shares. Products reviews published by users are valuable as well as important for both purchasers and producers. This proposed system used in many e-commerce sites, retail price analytics,social sites,etc.

1.5 METHODOLOGIES OF PROBLEM SOLVING:

In this project we have introduced an idea about Sentimental analysis on audio and video. It consists of several distinct algorithms, which try to over-come inherent deficiencies, burdens and limitations. The overall assembly performance has proven to be

particularly satisfactory. We plan to further improve the performance of the proposed method. Improvements can be done in each step of the proposed method. For example in the previous methods sentimental analysis on separate text, audio and video would be done. But in this previous system sentimental analysis on audios and videos gives less accuracy whereas proposed system gives more accuracy on sentimental analysis in terms of audios and videos.

Chapter 2

LITERATURE SURVEY

S R. N O	PAPER NAME	FOCUS	METHO DS	ADVANTAGES	DISADVANTAGE S	FUTURE WORK
1	SENTIMENT ANALYSIS ON PRODUCT REVIEWS	The main focus of sentiment analysis on product reviews is to review different algorithm and techniques to extract feature wise summary of a product and analyze it to form an authentic review.	Naïve Bayes ⁷	The system provides more successful sentimental classification based reviews.	Only suitable for text data.	The future work is to use data from websites such as EBay and twitter to aggregate their reviews and provide a larger number of numbers of reviews because in previous work, data are extracted from Amazon and Flipkart.
2	Sentimental Analysis on Speaker specific speech data	This work presents a generalized model that takes an audio which contains a conversation between two people as input and studies the content and speakers' identity by automatically converting the audio into text and by performing speaker recognition	1.Naïve Bayes 2.Linear SVM 3.VADER	--No loss of data while performing compression and decompression. --This system was useful for audio data.	It suffers some flaws, right now the system can handle a conversation between two speakers and in the conversation only one speaker should talk at a given time, it cannot understand if two people talk simultaneously	Our future work would address the issue like only one speaker should talk at a given time, it cannot understand if two people talk simultaneously and improve the accuracy and scalability of the system.
3.	Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques	The main focus of this paper was to analyze the data based on suitable algorithm provides the results in terms of analysis.	1.Naïve Bayes 2.Linear SVM	It is suitable for audios and text and provides better results.	It gives less accuracy to adventure based data as compared to other data.	The future scope of the work is that we can explore our data to a wider genre of different products on social networking sites or e-commerce as day by day the user is moving online and they prefer buying stuff online so we can identify the accuracy rates of the products like books, games etc.

4.	Onto-based sentiment classification using Machine Learning Techniques	The main focus in on text-based emotion prediction	1.Naïve Bayes 2.SVM 3.kNN	This project is simple and suitable for text data.	Classification and clustering of semi-structured text data have some challenges.	-
5.	Audio and Text based Multimodal Sentiment Analysis using Features Extracted from Selective Regions and Deep Neural Networks	While developing multimodal sentiment analysis, instead of taking entire input and extracting several features from toolkit, we identify selective regions of an input and experiment is performed on those regions by extracting specific features.	1.Naïve Bayes 2.SVM 3. Gaussian Mixture Models (GMM) 4. Deep Neural Networks (DNN) 5. Deep Neural Network Attention Mechanism (DNNAM)	--This project is simple. -- Suitable for small database.	In this paper the selective regions concept is not applied on deep neural network classifiers because of less data. The performance of deep neural networks is depend on the amount of training data.	In this work the main focused only on two modalities text and audio. The performance can be improved by combining these two modalities with video modality.
6.	Large-scale Affective Content Analysis: Combining Media Content Features and Facial Reactions	The paper present a novel multimodal model for affective video content analysis using deep visual-sentiment descriptions and automated facial coding.	Support Vector Machine(SVM)	Suitable for both audio and videos.	It gives less accuracy.	-
7.	Comparative Study of Machine Learning Techniques in Sentimental Analysis	The paper presents a detail survey of various machine learning techniques and then compared with their accuracy, advantages and limitations of each technique.	1.Naïve Bayes 2.SVM 3.KNN	This paper includes outline of current works that done on sentimental classification and analysis.	-	A more innovative and effective techniques required to be invented which should overcome the current challenges like classification of indirect opinions, comparative sentences and sarcastic sentences.

Chapter 3

SOFTWARE REQUIREMENT SPECIFICATION

3.1 ASSUMPTIONS AND DEPENDENCIES

Assumptions

For this application we have assumed that:

1. Bit rate for audio and video should be more than 250 kbps.
2. Internet network connectivity.

Dependencies

1. Audio should be in the form of mp3 or wav format with more than 250 kbps bit rate.
2. Video should be in the form of mp4 format with more than 250kbps bit rate.
3. Network dependency.

3.2 FUNCTIONAL REQUIREMENTS

3.2.1 REGISTRATION AND LOGIN

Individual login is provided for each users .They can select the audio and video which is collected or stored in file system. The procedure of solving sentiments will not be visible to the user. The user will only get to know the status of the sentiment analysis.

3.2.2 AUDIO AND VIDEO TRACKING

The user uploads the audio and video from file system. The audio and video can also be uploaded from the storage of the device. After uploading the audio and video from device the user has to automatically get the result in the form of sentiment analysis.

3.3 EXTERNAL INTERFACE REQUIREMENTS

3.3.1 USER INTERFACES

The website will be running on any type of browser. Its user interface will provide all the features described above. The features will be seamlessly integrated in to the website in a way that the user want have to select every feature manually, they will be active all the time as a background processes. The panels provided to user login:

1. The user will be given a Posting panel for uploading audio and video files for analysis.

3.3.2 HARDWARE INTERFACES

The website will require a file system for storing data gathered by a user of various sites in the form of audio and video. Detailed system requirements are: 1. Any core processor 2. 4 GB RAM

3.3.3 SOFTWARE INTERFACES

The server will require an operating system for running all the processes operating on the data gathered from the user file system. To store the user collected data , we will be using File system. Detailed software requirements are:

1. 64-bit Windows 10
2. Django version 2.0
3. File system

3.3.4 COMMUNICATION INTERFACES

For this project we are using a server for gathering all the user data and for processing it. The user data is gathered via websites which will be running the website. A stable connection has to be established between the user website and the server. This connection is established via the Internet.

3.4 NONFUNCTIONAL REQUIREMENTS

3.4.1 PERFORMANCE REQUIREMENTS

This website should provide robust management system where user can put all the audio and video data and system should be able to classify sentiments automatically , then record all the comprehensive solutions in file system for post analysis .The application should run smoothly with minimal delay between transitions from one activity to another.

3.4.2 SAFETY REQUIREMENTS

1. There should be less breaks in connection between user application and central server.
2. Communication should be maintained at a constant rate.
3. There should be little to no data loss while transferring data from user application to central server.

3.4.3 SECURITY REQUIREMENTS

1. The administrator should know how to operate the server efficiently.
2. Certain security guidelines need to be followed to prevent theft of user data.

3.4.4 SOFTWARE QUALITY ATTRIBUTES

1. The web-application is extremely user friendly.
2. It provides ease of access to various features of the application.
3. It is accurate and reliable.
4. It operates with minimum delay between activities by efficiently utilizing resources.

3.5 SYSTEM REQUIREMENTS

3.5.1 DATABASE REQUIREMENTS

1. File System

3.5.2 SOFTWARE REQUIREMENTS (PLATFORM CHOICE)

1. Operating system: 64-bit Windows 10.
2. Coding Language : python 3.6
3. Tool Kit : Django 2.0

3.5.3 HARDWARE REQUIREMENTS

1. System : Intel(R) Core(TM) i5-6006U @ 2.00 GHz
2. Hard Disk : 1 TB
3. Ram : 4GB

3.6 ANALYSIS MODELS

The Waterfall Model: Waterfall model is also called linear sequential model or classic life cycle model. In our proposed system we need to follow a particular sequence that's why we are using waterfall model in proposed system. This waterfall model paradigm begins at system level and goes through analysis, design, coding and maintenance.

Different stages of waterfall model:

1. Communication: Project development process starts with communication between group members. We need to specify all the requirements.

2. Planning: It includes complete estimation and scheduling of project.

3. Modeling: It includes detail requirement analysis and project design (algorithm and flowchart).

4. Construction: a. Coding b. Testing

5. Deployment: It includes software delivery, support and feedback from customer.

- Advantages:

1. This model is very easy to understand and use.

2. In this approach, each phase is processed and completed at one time and thus avoids phase overlapping.

3. It is very easy to manage since all the requirements are very well understood in the beginning itself.

- Disadvantage:

1. According to this model customer must state all his requirements at beginning stage of development which is difficult for the customer.

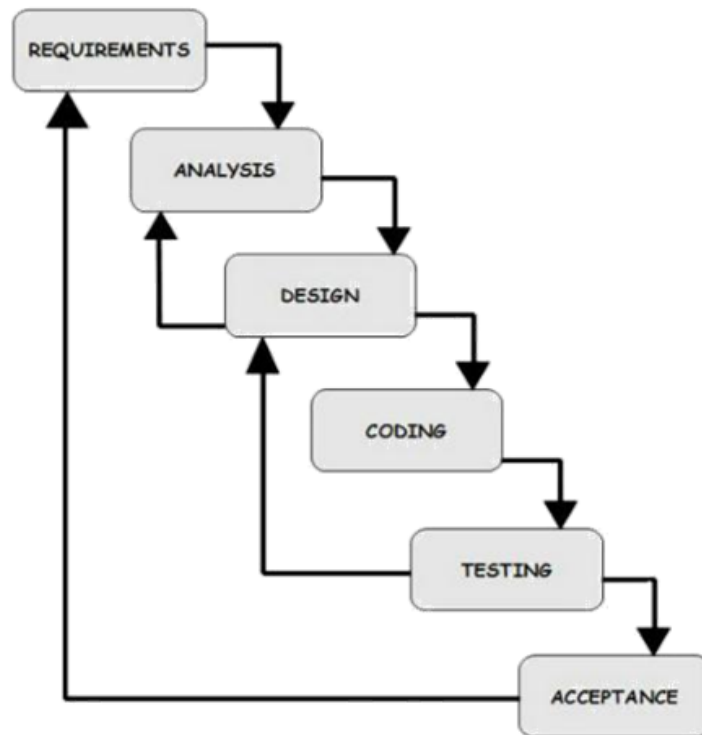


Figure 3.1: SDLC

Chapter 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE:

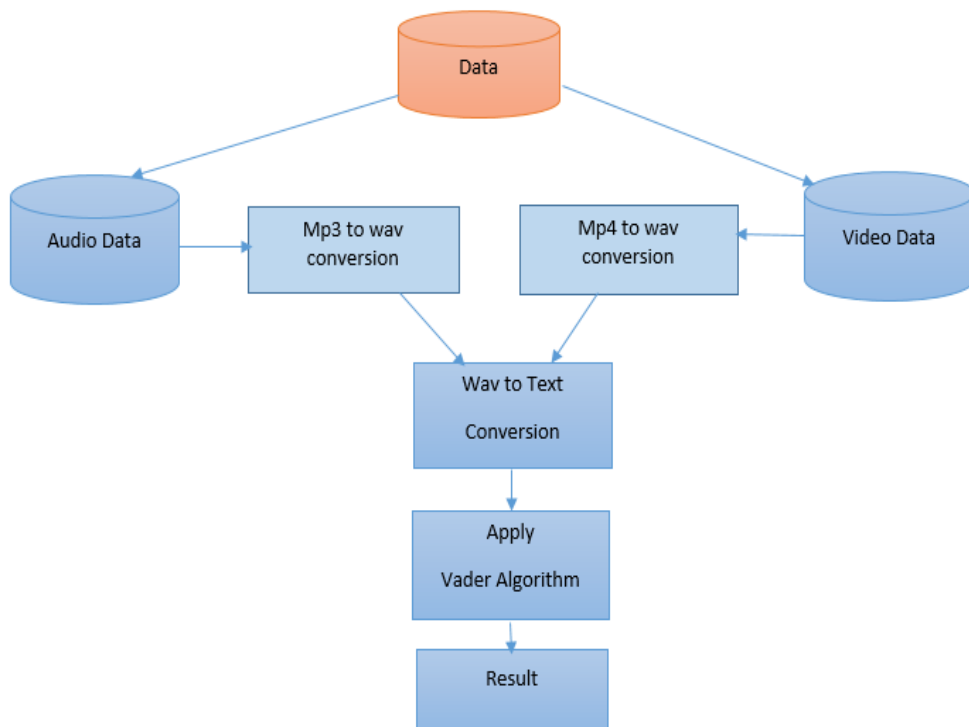


Figure 4.1: Proposed System Design

In the proposed system, we collect data from different social sites such as Facebook, Youtube to collect reviews of products. Depending upon the data such as audio and video, they are divided for further processing .system uses videos as a source for analyzing the content for sentiments. From Video, audio and will be separated to be processed. Audio and video will be converted to a .wav file and after that .wav file

can be converted into a text file. This text file is given to the VADER algorithm as input. VADER belongs to a type of sentiment analysis that is based on the lexicons of sentiment-related words. In this approach, each of the words in the lexicon is rated as to whether it is negative or positive, and neutral. From that, we found out the accuracy of the given data.

Vader working:

VADER is referred to as Valence Aware Dictionary and sEntiment Reasoner. VADER is a rule-based sentiment analysis library and lexicon. With VADER you can be up and run the performance of sentiment classification very fast even if you dont have negative and positive text samples to train a classifier or need to write a code to search for words in a sentiment lexicon. VADER is also computationally efficient when compared to other Machine Learning and Deep Learning approaches. VADER algorithm consists of an inbuilt training library that means we can not train our model separately.

VADER algorithm uses subjectivity and polarity concept. Polarity concept defines positive and negative words as well as it defines the range of polarity which is in between 0 and 1. Polarity, also known as orientation is the emotion expressed in the sentence. It can be positive, negative or neutral. Subjectivity is when text is an explanatory article which must be analyzed in context. Subjectivity is in between -1 to +1.

The results of VADER analysis are not only remarkable but also very encourag-

ing. The outcomes highlight the tremendous benefits that can be attained by the use of VADER in cases of micro-blogging sites wherein the text data is a complex mix of a variety of text.

4.2 MATHEMATICAL MODEL

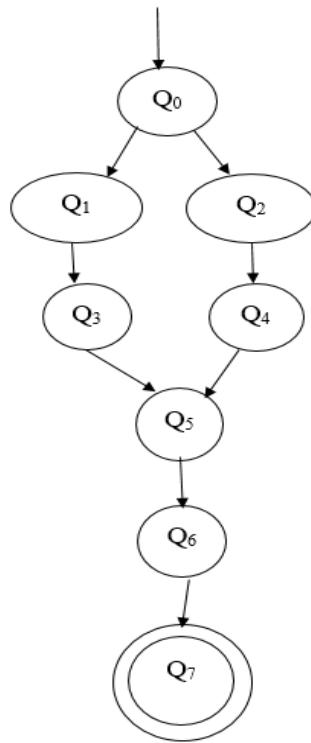


Figure 4.2: Mathematical Model

Q0 initial State

Q1 audio separation

Q2 video separation

Q3 mp3 to wav conversion

Q4 mp4 to wav conversion

Q5 wav to text conversion

Q6 Apply Algorithm

Q7 Result Display

U= Q, F,Q0, S, F,

Where,

Q= Q1, Q2,Q3, Q4, Q5, Q6, Q7

F= final State

Q0-Initial State

S: Success:

Data converted successfully

Accuracy obtained successfully

F: Failure:

If data is not converted successfully

Connection failure

4.3 DATA FLOW DIAGRAMS

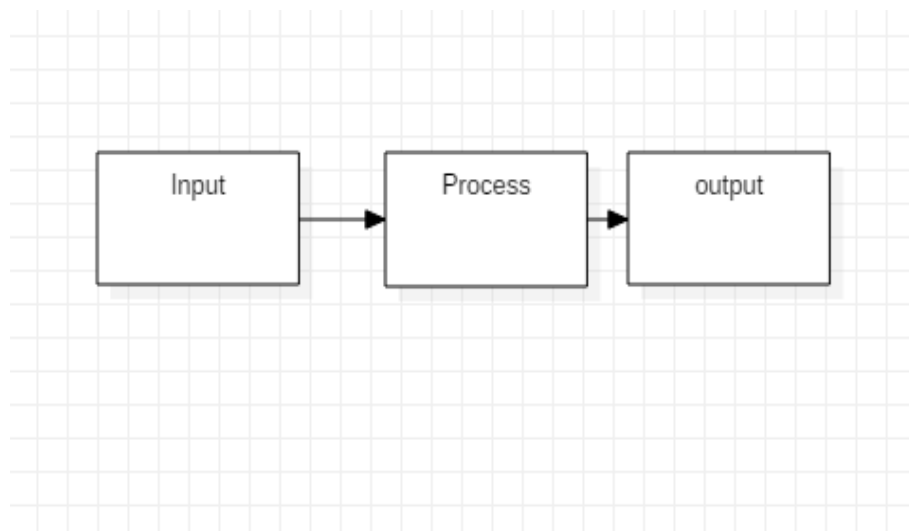


Figure 4.3: DFD 0

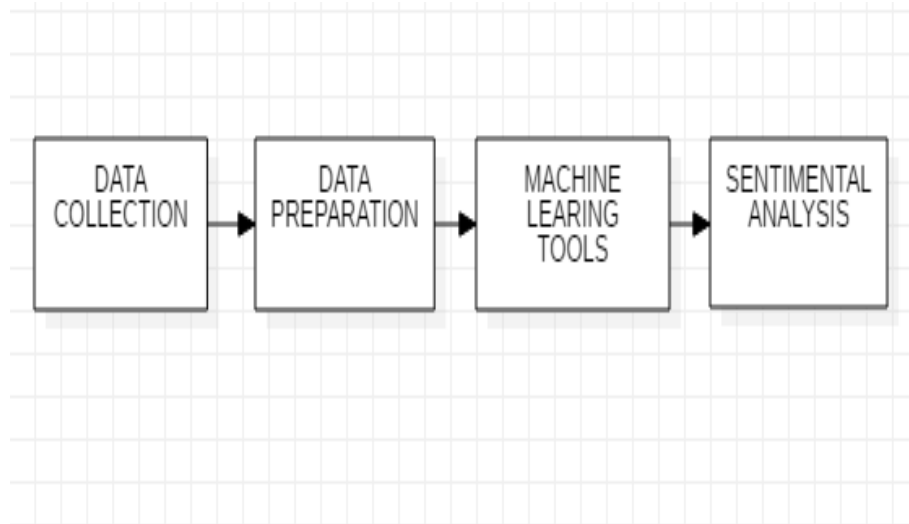


Figure 4.4: DFD 1

4.4 UML DIAGRAMS:

4.4.1 USE CASE DIAGRAM:

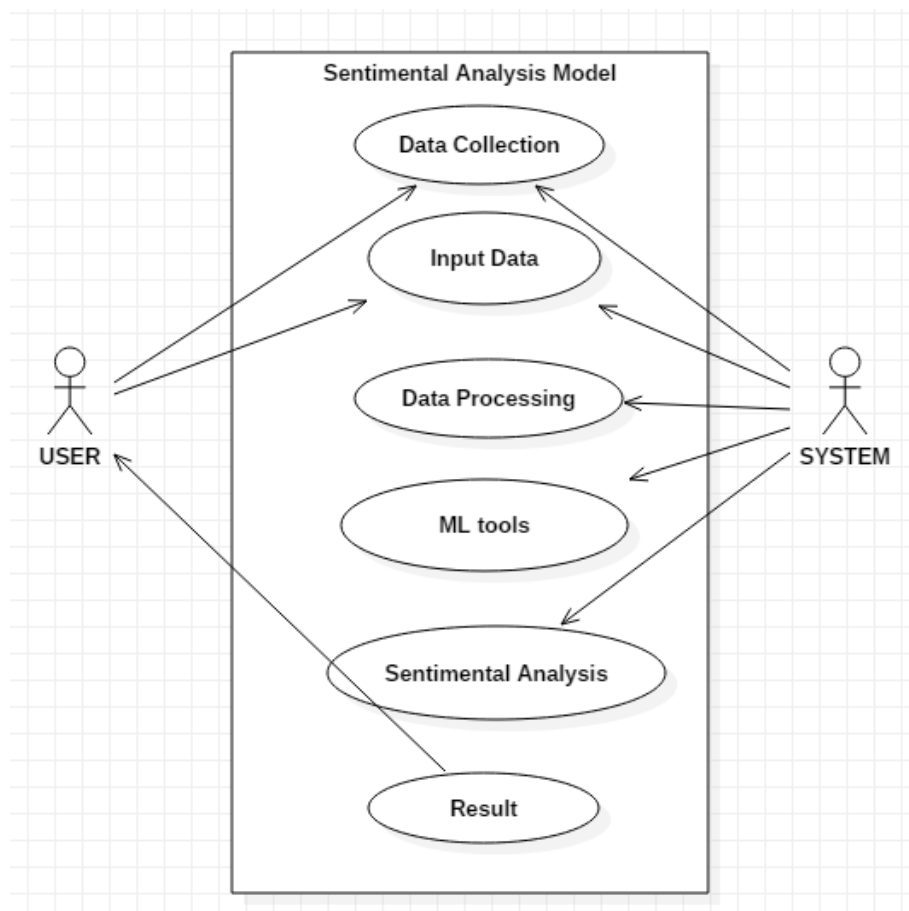


Figure 4.5: Use Case Diagram

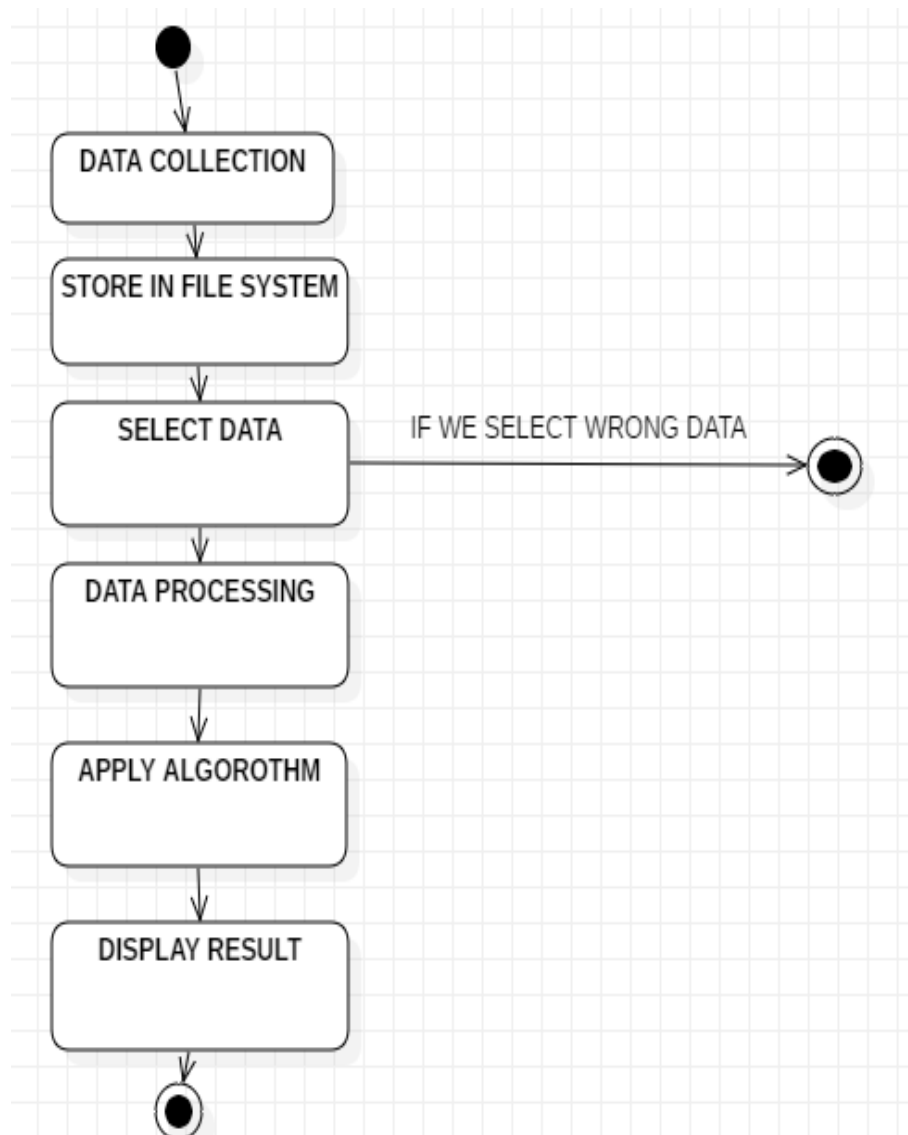
4.4.2 ACTIVITY DIAGRAM:

Figure 4.6: Activity Diagram

4.4.3 SEQUENCE DIAGRAM:

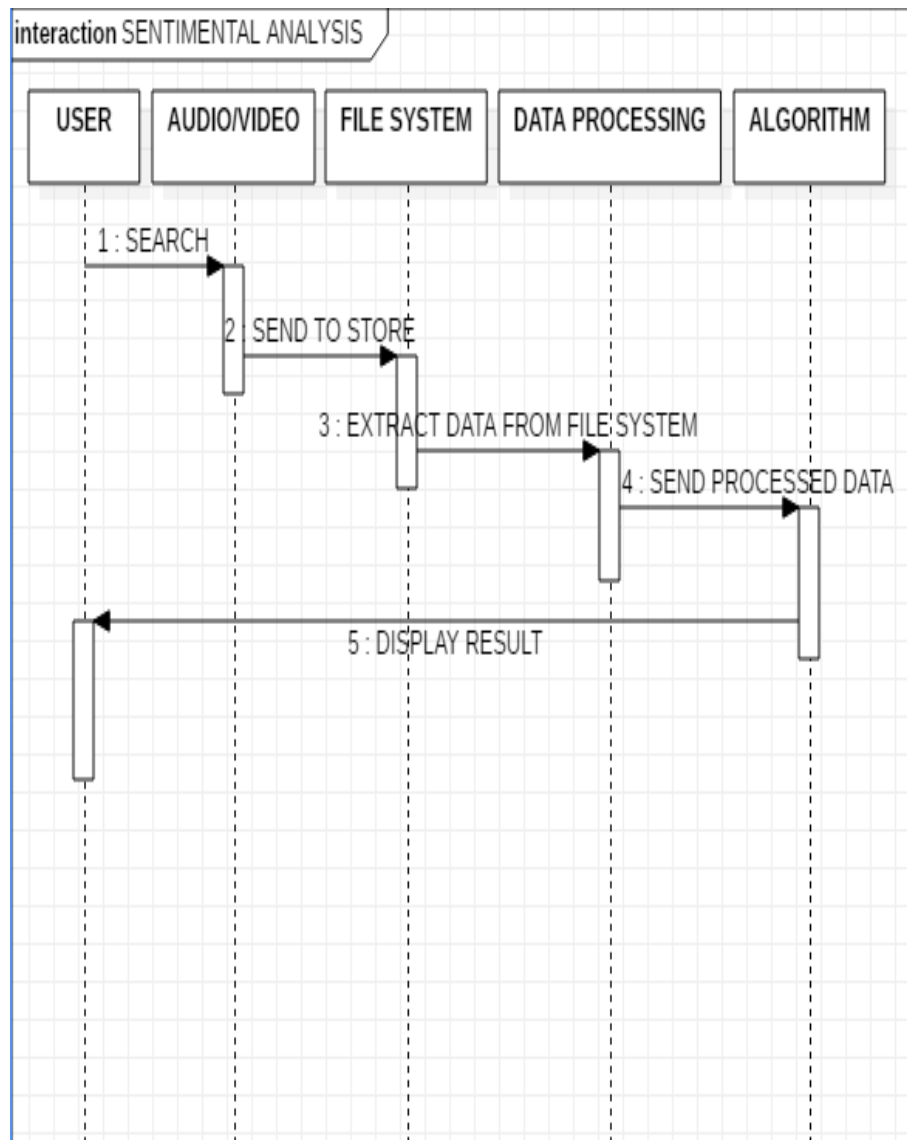


Figure 4.7: Sequence Diagram

4.4.4 CLASS DIAGRAM:

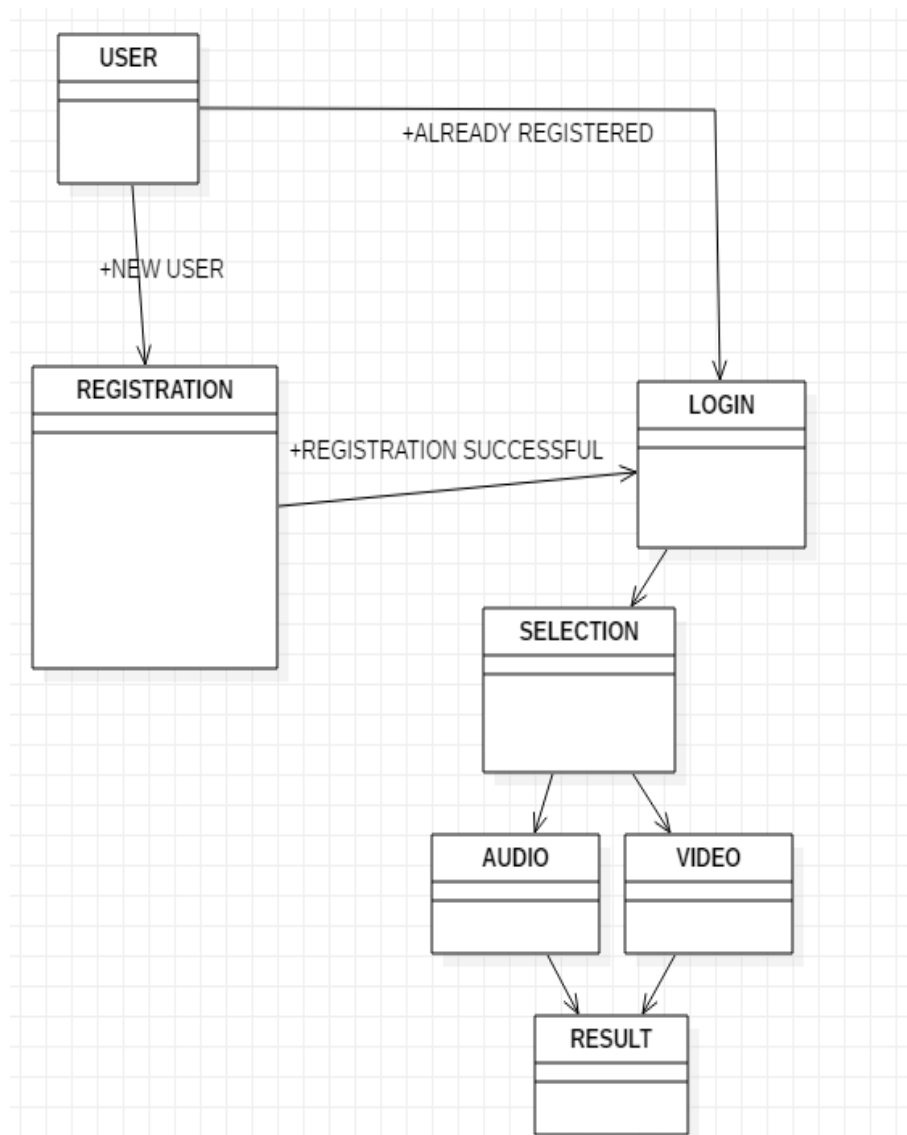


Figure 4.8: Class Diagram

Chapter 5

PROJECT PLAN

5.1 PROJECT ESTIMATE

A Project Estimate is an approximate calculation of the effort and cost it might take to complete a project. An estimate should not be a guarantee of the final cost, but provide a client with a guide of how much it might cost to do or deliver something so that they can secure budget to start the project. Budgetary estimates are made before a project is officially started.

5.1.1 RECONCILED ESTIMATES

The completed project may significantly differ from the planned tasks and projected conditions upon which the initial estimate was based. The initial estimate must be adjusted to account for these differences if a meaningful comparison between the estimate and the actual project effort is to be established. The purpose of the Reconciliation Adviser is to recalculate the estimated effort using the actual statistics and results from the completed project. The Reconciliation Adviser gathers actual project data through a question and answer process similar to that used when the system requirements were gathered for the initial estimate. The questions differ only in that the past tense is used: Did the application replace a mission critical or line of business process? instead of

Does the application replace a mission critical or line of business process?.

COST ESTIMATE

Resources required: - 1

Time require: - 6 Months.

Cost Required: - 2000

5.1.2 PROJECT RESOURCES

Type	Name	Description	Availability
Hardware	Computers	1 PCs for Server and client operation	High
Software	Django 2.0	These software for Run whole Project	High
File system	Any	To Store Data	Medium
Human	To write the code for the system	programmers	Medium
	Guide	To oversee the creation and installation of software	Low

5.2 RISK MANAGEMENT

Risk is an expectation of loss, a potential problem that may or may not occur in the future. It is generally caused due to lack of information, control or time. Loss can be anything, increase in production cost, development of poor quality software, not being able to complete the project on time. Software risk exists because the future is

uncertain and there are many known and unknown things that cannot be incorporated in the project plan. A software risk can be of two types (a) internal risks that are within the control of the project manager and (b) external risks that are beyond the control of project manager. Risk management is carried out to:

1. Identify the risk
2. Reduce the impact of risk
3. Reduce the probability or likelihood of risk
4. Risk monitoring

5.2.1 RISK IDENTIFICATION

In our project, we found out the following risks:

1. Technical risk
2. Schedule risk

5.2.2 RISK ANALYSIS

1. Technical risk:

If the data in file system is of low bit rate then it cannot be process by the application.

2. Schedule risk:

The time required to integrate the whole system will depend upon time required to construct individual components.

5.2.3 OVERVIEW OF RISK MITIGATION, MONITORING, MANAGEMENT

Following are the details for each risk.

Risk ID	1
Risk Description	Connectivity
Category	Development Environment.
Source	Software requirement Specification document
Probability	Low
Impact	High
Response	Mitigate
Strategy	Let users know in the beginning about all requirement
Risk Status	May occur sometime

Risk ID	2
Risk Description	User privacy
Category	Requirements
Source	Software Design Specification documentation review
Probability	Low
Impact	High
Response	Mitigate
Strategy	Better testing will resolve this issue.
Risk Status	Identified.

Risk ID	3
Risk Description	Data not uploaded on server
Category	Requirements
Source	This was identified during early development and testing
Probability	Low
Impact	Very High
Response	Accept
Strategy	Better testing will resolve this issue.
Risk Status	Identified

5.3 PROJECT SCHEDULE

5.3.1 PROJECT TASK SET

Major Tasks in the Project stages are:

Task 1:Requirement Analysis.

Task 2: Requirement Specification.

Task 3: System Design

Task 4: Detailed Design.

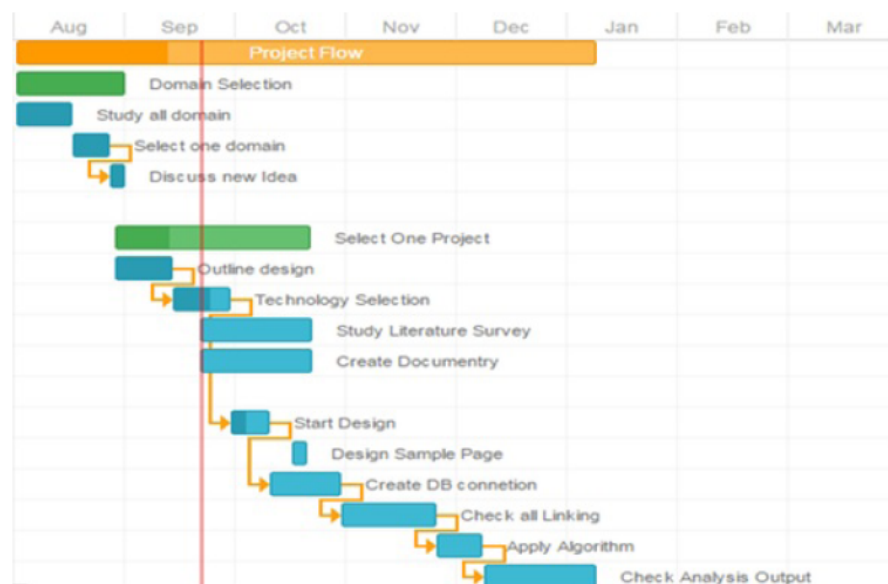
Task 5: Coding.

Task 6: Testing.

5.3.2 TASK NETWORK

The task network is a useful mechanism for depicting inter task dependencies and determining the critical path. When more people are working on a project, tasks and sub-tasks have their own inter-dependencies and also as the number of people increases, the development activities and tasks are performed in parallel. These concurrent tasks must be coordinated so that they will be complete when later tasks require their work products.

5.3.3 TIMELINE CHART



5.4 TEAM ORGANIZATION

5.4.1 TEAM STRUCTURE

Team Leader : That will divide the task.

Developr : To develop code for Execution.

Tester : To analyze error and correction its.

Analyzer : To analysis of code.

5.4.2 MANAGEMENT REPORTING AND COMMUNICATION

The team was very well organised to keep everyone in a loop. Everyone agreed on a x time to do project work. Project tasks were continued on holidays too. The overall project was divided among small tasks and distributed among all members. We communicated and shared our tasks through E-media. At the end of day everyone reported about their tasks. If any difficulty faced by any team member , we all used to gather next day and discuss the problem. Overall the team was very well coordinated.

Chapter 6

PROJECT IMPLEMENTATION

6.1 OVERVIEW OF PROJECT MODULES

1.Django:

is a Python-based free and open-source web framework, which follows the model-template-view (MTV) architectural pattern. Django's primary goal is to ease the creation of complex, database-driven websites. Database contains login details of user such as phone username, emailid and password .By using Django we can upload the audio and video file by selecting the file from the file system.

2.File System:

File System contains the audio and video data .We can store audio and video files in the file system which we collected from different sites .During the operation of upload file we choose the files from the file system. We can analyze the data which is stored in file system.

3.Browser:

Browser is just for user interface that is browser is front end side. On that browser we can performs many operations like user login, user registration, we can upload the file on that browser ,etc.Browser displays the final result.

6.2 TOOLS AND TECHNOLOGIES USED

6.2.1 DJANGO

Django is a Python-based free and open-source web framework, which follows the model-template-view (MTV) architectural pattern. Django was created in the fall of 2003, when the web programmers at the Lawrence Journal-World newspaper, Adrian Holovaty and Simon Willison, began using Python to build applications. It was released publicly under a BSD license in July 2005. The framework was named after guitarist Django Reinhardt.

In June 2008, it was announced that a newly formed Django Software Foundation (DSF) would maintain Django in the future. Django's primary goal is to ease the creation of complex, database-driven websites. The framework emphasizes reusability and "plug-gability" of components, less code, low coupling, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models. Django also provides an optional administrative create, read, update and delete interface that is generated dynamically through introspection and configured via admin models.

Despite having its own nomenclature, such as naming the callable objects generating the HTTP responses "views", the core Django framework can be seen as an MVC architecture. It consists of an object-relational mapper (ORM) that mediates between data models (defined as Python classes) and a relational database ("Model"), a system for processing HTTP requests with a web templating system ("View"), and a regular-

expression-based URL dispatcher ("Controller"). Also included in the core framework are:

- 1.A lightweight and standalone web server for development and testing
- 2.A form serialization and validation system that can translate between HTML forms and values suitable for storage in the database
- 3.A template system that utilizes the concept of inheritance borrowed from object-oriented programming
- 4.A caching framework that can use any of several cache methods
- 5.Support for middleware classes that can intervene at various stages of request processing and carry out custom functions
- 6.An internal dispatcher system that allows components of an application to communicate events to each other via pre-defined signals
- 7.An internationalization system, including translations of Django's own components into a variety of languages
- 8.A serialization system that can produce and read XML and/or JSON representations of Django model instances
- 9.A system for extending the capabilities of the template engine
- 10.An interface to Python's built-in unit test framework.

Django Extensibility:

Django's configuration system allows third party code to be plugged into a regular project, provided that it follows the reusable app conventions. More than 2500 packages are available to extend the framework's original behavior, providing solutions to issues the original tool didn't tackle: registration, search, API provision and consumption, CMS, etc.

This extensibility is, however, mitigated by internal components' dependencies. While the Django philosophy implies loose coupling, the template filters and tags assume one engine implementation, and both the auth and admin bundled applications require the use of the internal ORM. None of these filters or bundled apps are mandatory to run a Django project, but reusable apps tend to depend on them, encouraging developers to keep using the official stack in order to benefit fully from the apps ecosystem.

6.3 ALGORITHM DETAILS

6.3.1 VADER

VADER is referred to as Valence Aware Dictionary and sEntiment Reasoner. VADER is a rule-based sentiment analysis library and lexicon. With VADER you can be up and run the performance of sentiment classification very fast even if you don't have negative and positive text samples to train a classifier or need to write a code to search for words in a sentiment lexicon. VADER is also computationally efficient when compared to other Machine Learning and Deep Learning approaches. VADER algorithm consists of an inbuilt training library that means we can not train our model separately.

VADER algorithm uses subjectivity and polarity concept. Polarity concept defines positive and negative words as well as it defines the range of polarity which is in between 0 and 1. Polarity, also known as orientation is the emotion expressed in the sentence. It can be positive, negative or neutral. Subjectivity is when text is an explanatory article which must be analyzed in context. Subjectivity is in between -1 to +1.

The results of VADER analysis are not only remarkable but also very encouraging. The outcomes highlight the tremendous benefits that can be attained by the use of VADER in cases of micro-blogging sites wherein the text data is a complex mix of a variety of text. VADER belongs to a type of sentiment analysis that is based on the lexicons of sentiment-related words. In this approach, each of the words in the lexicon is rated as to whether it is negative or positive, and in many cases, how positive or negative. In the following table, you can see an excerpt from VADERs lexicon, where more positive words have higher positive ratings and more negative words have lower negative ratings.

Word	Sentiment rating
Great	3.1
Disaster	-3.1
Insane	-1.7
Tragedy	-3.4

Figure 6.1: Sentiment Rating

To work out whether these words are positive or negative the developers of these

approaches need to get a bunch of people to manually rate them, which is obviously pretty expensive and time-consuming. In addition, the lexicon needs to have good treatment of the words in your text of interest, otherwise, it won't be very accurate. On the other side, when there is a good fit between the lexicon and the text, this approach is accurate, and additionally quickly returns results even on large amounts of text. Incidentally, the developers of VADER used Amazon's Mechanical Turk to get most of their ratings, which is a very quick and cheap way to get their ratings!

When VADER analyses a chunk of text it checks to see if any of the words in the text are present in the lexicon. Consider the example, the sentence "The juice is good and the dress is nice" has two words in the lexicon (good and nice) with ratings of 1.9 and 1.8 respectively.

VADER generates four sentiment metrics from these word ratings, which you can see below. The first three, negative, neutral and positive, represent the proportion of the text that falls into those categories. As you can see, our example sentence was rated as 45

Sentiment metric	Value
Positive	0.45
Negative	0
Neutral	0.55

Figure 6.2: Sentiment Metric

Chapter 7

SOFTWARE TESTING

7.1 TYPE OF TESTING

Along with the type of testing also mention the approach to be followed for the testing, that is, Manual Testing or Automated Testing. Use Automated Testing Plan for planning automation activities in details. The different types of testing that may be carried out in the project are as follows:

- Unit Testing
- Integration Testing
- System Testing
- Validation Testing
- White Box Testing
- Black Box Testing
- GUI Testing

7.1.1 UNIT TESTING

Individual components are tested independently to ensure their quality. The focus is to uncover errors in design and implementation, including

- Data structure in component
- Program logic and program structure in a component
- Component interface
- Functions and operations of a component

7.1.2 INTEGRATION TESTING

A group of dependent components are tested together to ensure their quality of their integration unit. This approach is to do incremental integration to avoid big-bang problem. That is when the entire program is put together from all units and tested as a whole. The big bang approach usually results in chaos which incremental integration avoids. Incremental integration testing can be done in two different ways top down and bottom up. Then there is also the possibility of regression integration. The top down integration is when modules are integrated by moving downwards through the control hierarchy, beginning with the main control module. Modules subordinate to the main control module are incorporated into main structure in either depth-first or breadth-first manner. The top down integration verifies major controls or decision points early in the test process. If major control problems do exist, early recognition is essential. Bottom-up integration testing begins construction and testing with the lowest levels in the program structure.

Because modules are integrated from the bottom-up, processing required for modules subordinate to a given level is always available and the need for test stubs is eliminated.

The focus is to uncover errors in:

- Design and construction of software architecture
- Integrated functions or operations at sub-system level
- Interfaces and interaction and/or environment integration

7.1.3 SYSTEM TESTING

The system software is tested as a whole. It verifies all elements mesh properly to make sure that all system functions and performance are achieved in the target environment.

The focus areas are:

- System functions and performance
- System reliability and recoverability (recovery test)
- System behavior in the special conditions (stress and load test)
- System user operations (acceptance test/alpha test)
- Hardware and software integration collaboration
- Integration of external software and the system

7.1.4 VALIDATION TESTING

Validation can be defined in many ways, but a simple definition is that succeeds when software functions in a manner that can be reasonably expected by the customer. Software validation is achieved through a series of black-box tests that demonstrate conformity with requirements. A test plan outlines the classes of tests to be conducted and a test procedure defines specific test cases that will be used to demonstrate conformity with requirements. Both the plan and procedure are designed to ensure that all functional requirements are satisfied, all behavioral characteristics are achieved, all performance requirements are attained, documentation is correct, and human engineered and other requirements are met.

7.1.5 WHITE BOX TESTING

White-box test design allows one to peek inside the box, and it focuses specifically on using internal knowledge of the software to guide the selection of test data. Synonyms for white-box include: structural, glass-box and clear-box. White box testing is much more expensive than black box testing. It requires the source code to be produced before the tests can be planned and is much more laborious in the determination of suitable input data and the determination if the software is or is not correct. This testing is concerned only with testing the software product; it cannot guarantee that the complete specification has been implemented.

7.1.6 BLACK BOX TESTING

Black-box test design treats the system as a black-box, so it does not explicitly use knowledge of the internal structure. Black-box test design is usually described as focusing on testing functional requirements. Synonyms for black box include: behavioural, functional, opaque-box, and closed-box. Black box testing is concerned only with testing the specification ; it cannot guarantee that all parts of the implementation have been tested. Thus black box testing is testing against the specification and will discover faults of omission, indicating that part of the specification has not been fulfilled.

7.1.7 GUI TESTING

Graphical User Interface (GUIs) present interesting challenges for software engineers. Because of reusable components provided as part of GUI development environments, the creation of the user interface has become less time consuming and more precise. But, the same time, the complexity of GUIs has grown, leading to more difficulty in the design and execution of the test cases. Because many modern GUIs have the same look and same feel, a series of test cases can be derived.

7.2 TEST CASES TEST RESULTS

TEST CASE ID	TEST CASE	STEPS TO BE EXECUTED	RESULT EXPECTED
TC1	Change Path	Change the path using command prompt	Path successfully changed
TC2	Run Django	Run the Django using command	Successful Run
TC3	Browser Type	User can set browser which is available to the system	Browser will start
TC4	Register button show on home page of the website	<ul style="list-style-type: none"> Click on the Register button. Registration page should get open. Fill the Registration form and submit 	After clicking on the submit button of system should perform the respective operation.
TC5	Login button show on home page of the website	<ul style="list-style-type: none"> User fill the User name and password in text Box. User click on Login Button. 	Login Successfully
TC6	System should Verify Password and Username	User should click on submit button for verification	Fail
TC7	System should Verify all required field.	<ul style="list-style-type: none"> User fill the User name and password in text Box User email id,etc. 	Successful Registered User
TC8	select audio file	From the file system select audio file	File successfully selected
TC9	select video file	From the file system select video file	File successfully selected
T10	Analyze data	Click on upload	Display final result

Chapter 8

OUTCOMES AND SCREENSHOTS

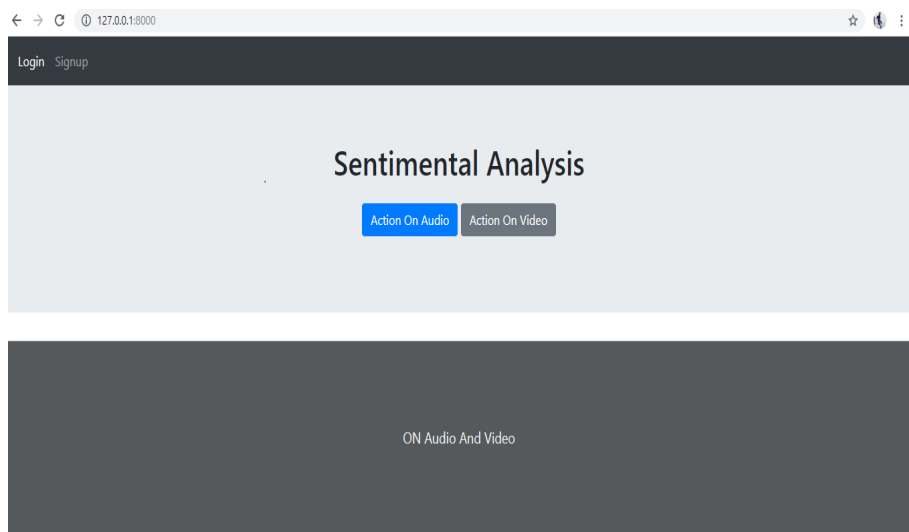


Figure 8.1: Home Page

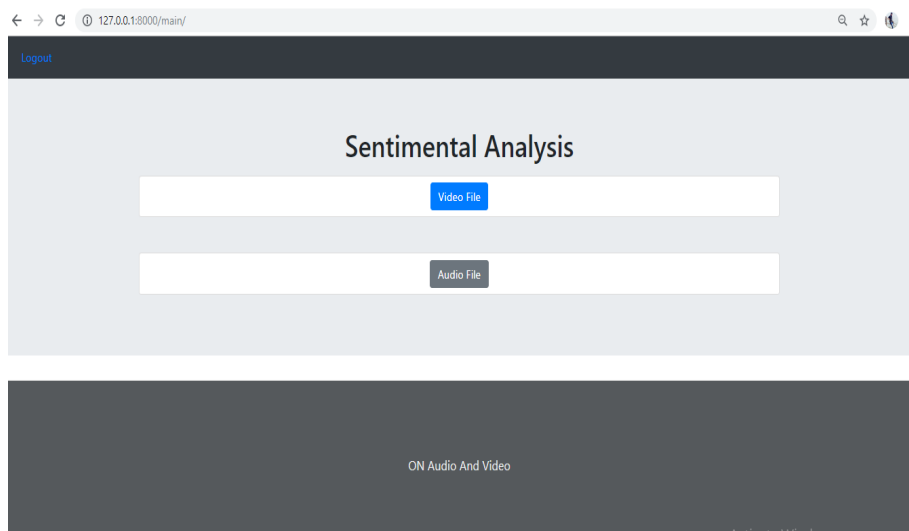


Figure 8.2: Main Page

← → 127.0.0.1:8000/signup/

Sign up [Log In](#)

Username
eg. Vishal

Email address
eg. vishalpawan123@gmail.com

Create password

Confirm Password

[Register](#)

Have an account? [Log In](#)

Activate Windows
Go to Settings to activate Windows.

Figure 8.3: SignUp Page

← → 127.0.0.1:8000/login/

Sign In

Username
monali

password

[Sign In](#)

Have an account? [Sign Up](#)

Figure 8.4: SignIn Page

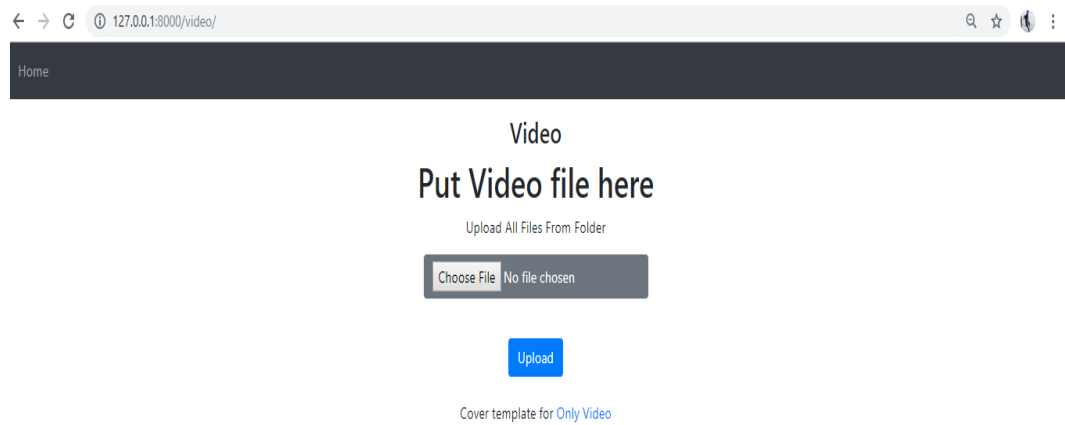


Figure 8.5: Video Input

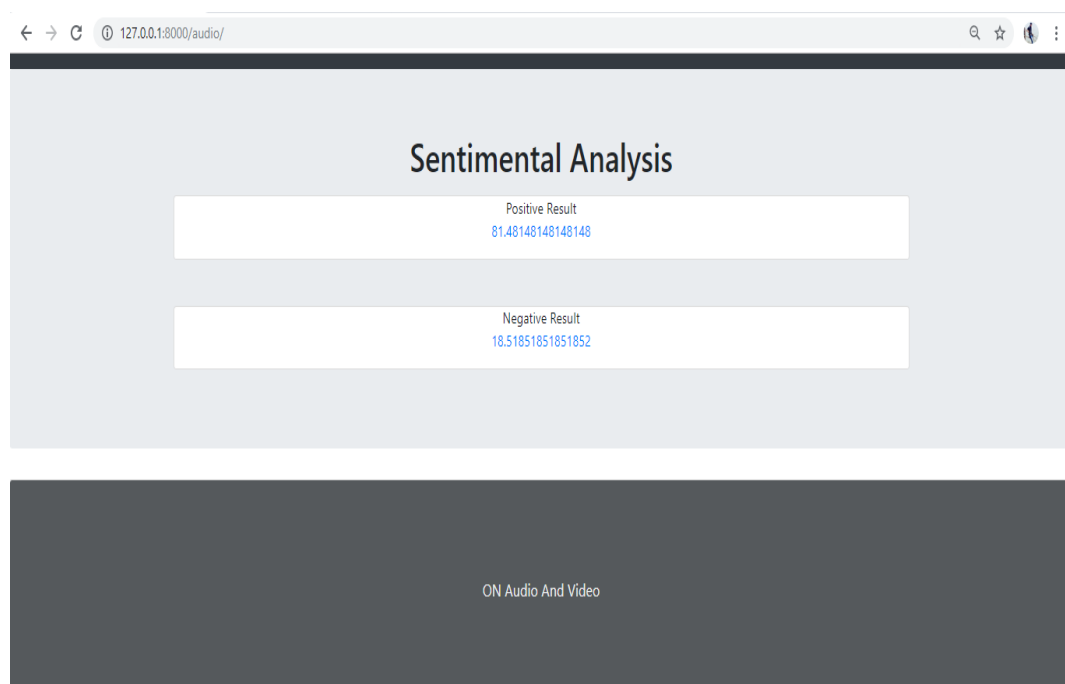


Figure 8.6: :Result

Chapter 9

CONCLUSION

9.1 CONCLUSION

Products reviews published by users are valuable for both purchasers and producers. Nowadays audio and video monitoring is vital in many social sites like facebook, tweeter, youtube, etc. Using our system we can analyze the usage and also detect the Opinion in a particular product. In this paper, we take the audio and video from different sites such as youtube, facebook, etc this audio and video files first converted into text data, then from that converted data we have applied the sentiment analysis. In this paper, we have used the VADER algorithm the functionality of the VADER algorithm is that it does not require separate data for training because the VADER algorithm already has the training data set. VADER algorithm uses the concept of polarity and subjectivity from that they find out the analysis. The proposed system gives the output in terms of positive and negative percentage.

9.2 FUTURE WORK

In this system, we apply the sentiment analysis on audio and video as audio is converted into text data and video is extracted so that we get audio data then this audio data

is converted into text data after that we apply the sentiment analysis on that text data. But the main thing is that from a video we only extract the audio data, here we have not extracted the image data. So our future work is that from a video we extract the audio data as well as image data after that we apply image processing on that image data and finally, this image data is compared with audio data so as to get a final output that is we apply sentiment analysis on that data that we get.

9.3 APPLICATIONS

- Review-related Websites
- Sub-component Technology
- Business Intelligence
- Domains
- E-commerce
- Government polls

9.4 REFERENCES

- [1] Bhavitha B K, Anisha P Rodrigues and Dr. Niranjan N Chiplunkar, Comparative Study of Machine Learning Techniques in Sentimental Analysis,IEEE2017
- [2] Harika Abburi, Audio and Text based Multimodal Sentiment Analysis using Features Extracted from Selective Regions and Deep Neural Networks,2017
- [3]Chhaya Chauhan, SmritiSehgal, SENTIMENT ANALYSIS ON PRODUCT REVIEWS,IEEE2017
- [4] Maghilnan S, Rajesh Kumar M, Sentiment Analysis on Speaker Specific Speech Data,2017(I2C2)
- [5]Daniel McDuff and Mohammad Soleymani , Large-scale Affective Content Analysis: Combining Media Content Features and Facial Reactions,2017IEEE
- [6]Shweta Rana and Archana Singh, Comparative Analysis of Sentiment Orientation Using SVM and Nave Bayes Techniques,2016IEEE
- [7]Ms.K.Saranya, Dr.S.Jayanthi, Onto-based sentiment classification using Machine Learning Techniques2017,IEEE
- [8]Marie Katsurai and Shinichi Satoh, Image Sentiment Analysis Using Latent Correlations Among Visual,Textual,And Sentiment Views,2016IEEE
- [9]Aishwarya Murarka¹, Kajal Shivarkar², Sneha³, Vani Gupta⁴,Prof.Lata SankpalSentiment Analysis of Speech,2017
- [10]Lakshmish Kaushik, Abhijeet Sangwan, John H. L. Hansen , Sentiment Extraction From Natural Audio Streams2013IEEE