Cairo University
Faculty of Computers and Artificial intelligence
Department of Computer Sciences

SENTIMENT ANALYSIS ON CITIZENS REVIEWS

Supervised by

Dr. Hanaa Bayoumi

TA. Mohamed Atta

Implemented by

20190041	Ahmed Tarek Fawzy Ibrahem
20190562	Mo'men Hatem Mohamed Ali
20190019	Ahmed Badr Shaaban
20190632	Youssef Khaled Abd-ElShafi
20190050	Ahmed Essam-Eldin Abdelfattah

Graduation Project

Academic Year 2022-2023

Midyear Short Documentation

Project Short Documentation

Table of Contents

Abstract 3
Background 4
Main technique and application: 4
Problem definition 5
Related Work 5
– Amazon Product Review 5
– Rotten Tomatoes Movie Reviews 5
– Sentiment Analysis on News Topics during COVID-19 5
– Analyze a Company's Reputation 6
Project Specifications 6
Machine Learning Architecture 6
Project Cycle 6
Data Collection 6
Data Preprocessing and Feature Extraction 7
System Architecture 7
Stackholders 8
Functional requirement 8
Non-Functional requirement 8
Use Case Diagram 9
Class Diagram 10
Sequence Diagram 10
Work Plan & Gantt Chart 12

Abstract

The problem which always existed is that there are a lot of government institutions with a certain number of managers who try to manage and lead this institution, but because there are a lot of aspects which need to be observed, they always need to evaluate their work to put them in the right track.

Citizens usually come with many complaints about some services that are not working satisfactorily for everyone. With so many complaints, it becomes difficult for officials to study all these complaints and find out the reason for the failure or inability to provide the service as expected.

This service allows citizens to view all reviews related to a specific sector / service, as well as view previous analyses based on complaints of others, to enable him to know if he is the only one who sees that there is a problem like him. for example, in the ministry of health services, based on previous reviews on the system {60% negative , 20% positive , 20% neutral }.

Here comes our application which takes feedback from citizens in different governments, departments, sections and analyzes them, does some probabilities to them then sends them to the managers to observe if they are doing well or if something is wrong in a certain department. In addition to this, when the manager resolves a problem he can post a blog to all the citizens that the problem was taken care of which assures the citizens that their feedback is taken care of.

The problem with those in charge of certain sectors and monitoring user reviews is that it's hard for them to look at all the reviews. Therefore, the system arranges the complaints from the most important to the least important, to make it easier for officials to read the complaints, find out the reason for the bad service, and work on developing or resolving it. Example When a manager opens reviews for his department, he sees all the reviews but in an ordered manner as prioritizing negative then neutral then positive.

Implementing this application needs tools like:

Backend (using Django framework).

Frontend (using Angular)

NLP (natural language processing).

Background

Previously in government departments, if someone wanted to file a complaint somewhere or express his opinion about the service that was provided to him, he would write an opinion poll if the government department cared about the opinions of citizens about this service.

Responsible managers rarely follow up with numerous complaints and surveys to track customer satisfaction. They may choose a random number of reviews to consider instead of reading all of them to represent average user satisfaction.

Any user or consumer would like to know people's opinions and their previous experience with this service, but it is not easy until collecting information from several people to know the situation. This service provides all statistics and analyzes based on the opinions of previous users of the service for both the user (to see if it is more appropriate for him) and the manager (to improve the service).

Main technique and application:

in Machine learning algorithms, we use:

- Long short-term memory: using recurrent neural networks that learn long term dependencies, has feedback connection, capable of processing long length sequences of data. We expect to get more accurate predictions and a better understanding of what choices to make. Used in handwriting recognition, speech recognition and machine translation.

- Transformer model: neural network that learns context and therefore meaning by tracing relationships in sequential data such as the words in this sentence. Transformers eliminate the need to train neural networks with large, labeled data sets that were costly and time consuming by mathematically finding patterns between elements.

- Convolution neural network: unlike LSTM, CNN is useful for short text data and may not be suitable for long reviews. They perform well on classification tasks, such as sentiment analysis, spam detection, or topic classification.

Problem definition

Nowadays, there are a lot of departments (of different fields in life like e.g.: education, health, …) in the whole country which includes tons of institutions among cities but only a few managers who can control and monitor them, so there is a great need to get feedback from citizens who are using those services consistently. The population of citizens is increasing tremendously so it is almost impossible for the managers to read all citizens' feedback for all institutions in every department. Even so, the managers would not be able to tell the citizens about any updates in a certain institution in a certain department easily, it would cost them a lot to reach the citizens concerning an issue they raised earlier. Feedback also needs analysis from the managers to be able to figure out the source of the problem so if they solve it will prevent further problems in the future.

Related Work

– Amazon Product Review

Amazon is one of the biggest e-commerce stores. it uses sentiment analysis to evaluate the products which it shows using Product Review from the different customers.

It Take the reviews and Use TFIDF (Term Frequency Inverse Document Frequency) to convert Review into integer and then use classification models like SVM to know if the Review is positive or negative.

– Rotten Tomatoes Movie Reviews

Its platform has review of every TV series , show or drama , it benefits from critics and Movie fans reviews to know if this Movie or series will succeed or not by using sentiment analysis Take the review and preprocessing it using NLTK library and tokenize it after make preprocessing This tokenize will enter the multi class model and classify it negative , somewhat negative, neutral, fairly positive , and positive.

– Sentiment Analysis on News Topics during COVID-19

Covid-19 is a global epidemic that has spread and caused a lot of problems. This global epidemic affects not only people’s physical health but because social distance during this period causing problems and threats to our emotional stability.

Therefore, they took advantage of social media such as Facebook, Twitter, and LinkedIn to learn about people's feelings and reactions to this global epidemic and use it in sentiment analysis.

– Analyze a Company's Reputation

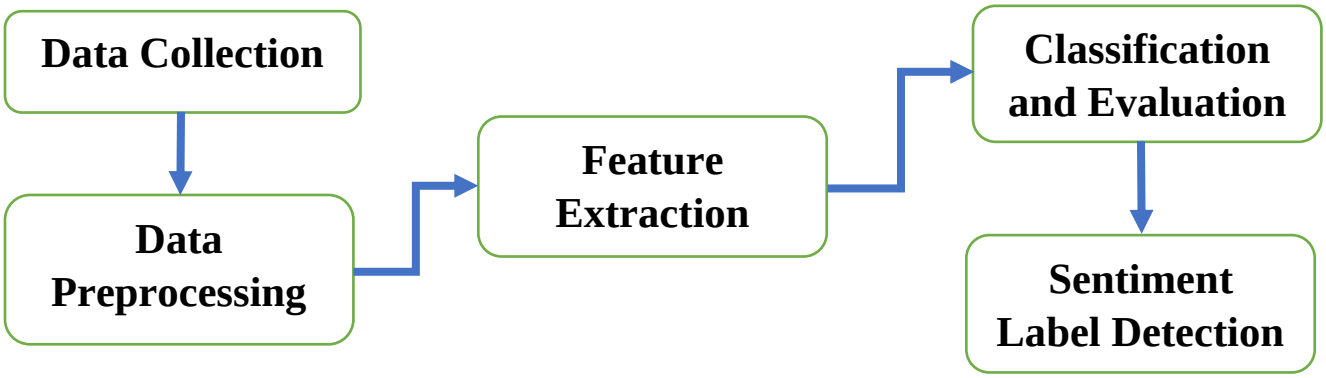
Many of people wanted to know specific company has good Reputation or not and they consume more time to know the feedback of this company in more than social media platform.

So, they use intermediate sentiment analysis and collect data from different platforms like Facebook, Twitter, and LinkedIn to know the opinion of people about any company and its product.

Project Specifications

Machine Learning Architecture

Project Cycle



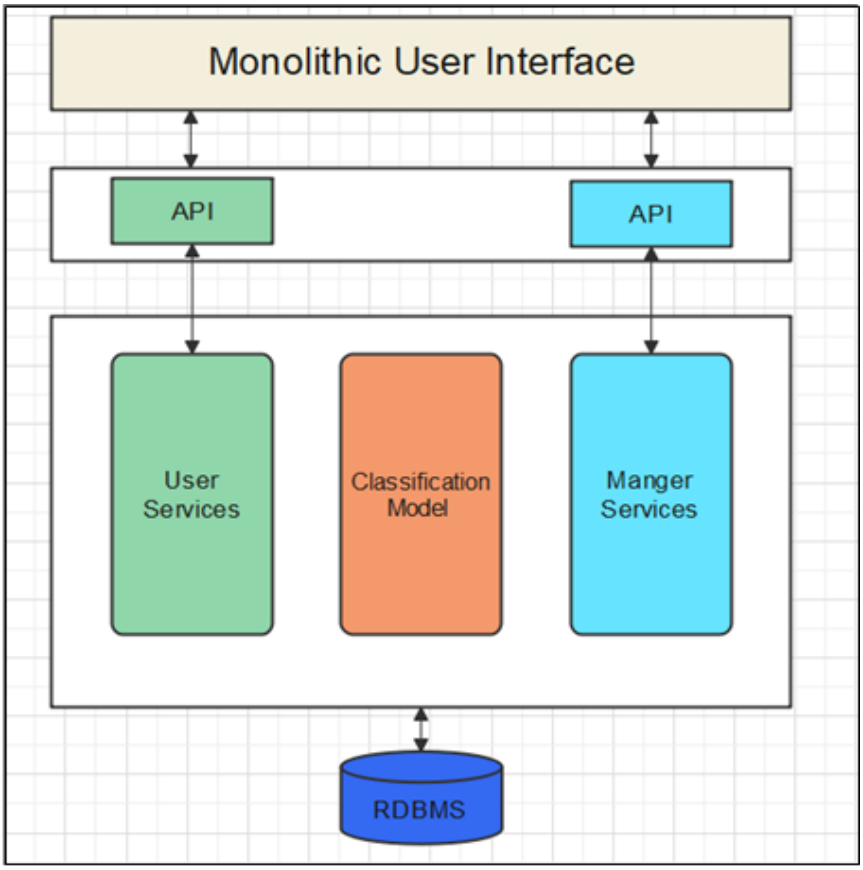
Data Collection

Dataset Name	Size	Language
BRAD: Books Reviews in Arabic	500K	MSA, Dialects
HARD: Hotel Reviews in Arabic	370K	MSA, Dialects
LABR: A Large-Scale Arabic Book Reviews	63K	MSA, Dialects
ASTD: Arabic social sentiment analysis	10K	MSA, Dialects

Data Preprocessing and Feature Extraction

Data-set	Preprocessing	Feature Extraction	Training and classification
Subset of all previous data sets	Text Tokenization	Word Embeddings (Word to Vector)	LSTM = 73%
	Stop Words Removal		
	Punctuation Removal		
	Digits and English letters Removal		
	Emoji Removal / Replacement		
	Tashkeel Removal(Normalization)		
	Word Stemming		

System Architecture



Stackholders

- Developers : they are responsible for developing, testing, handling errors and maintaining the software.
- Citizen : they will write their feedback and get some stars.
- Managers : they will view the analysis result and can write blogs.
- Admin: the administrator of a system that can remove/add managers and access the data in the system.

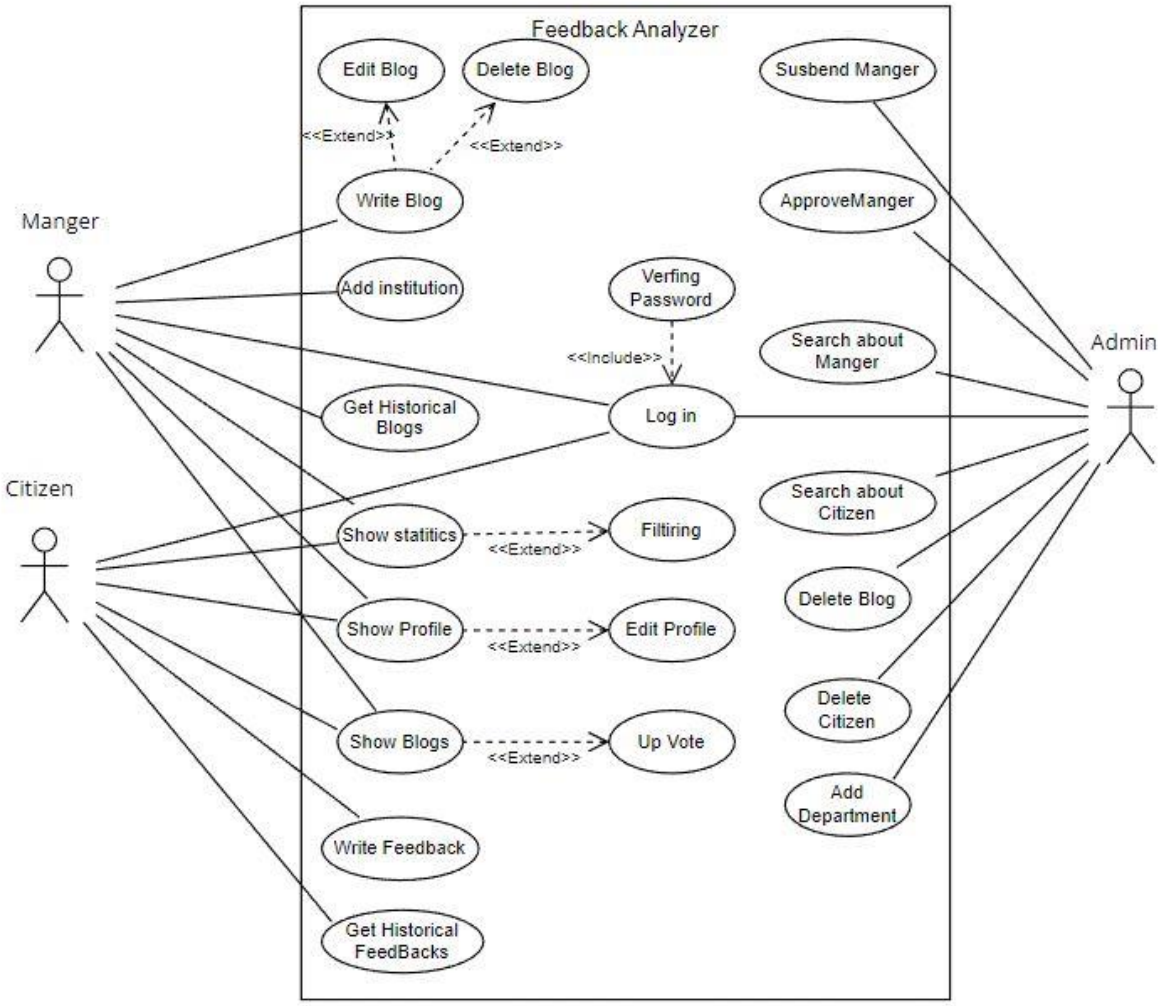
Functional requirement

- Manger
  - The manager can add and view blogs to describe his action about some Feedback.
  - The manager can create a new institution for the department.
  - The manager can show feedback about specific institutions.
  - The manager can show the statistics of department/City.
- Citizen
  - The citizen can show all blogs and Filter the blogs.
  - The citizen can give feedback about an institution in a specific department in a certain city.
  - The citizen can view his level.
  - The citizen can show the statistics of department/City.

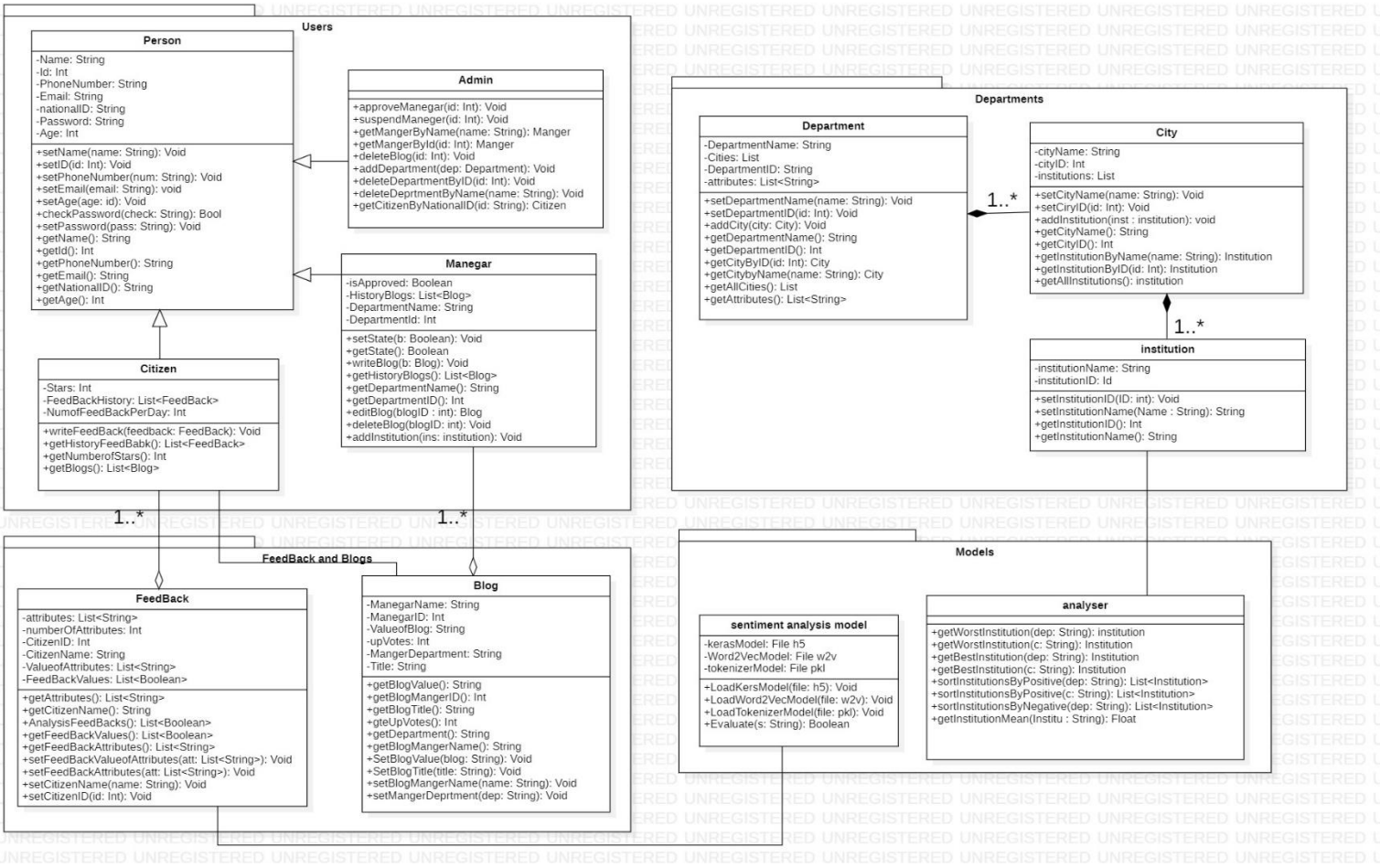
Non-Functional requirement

- Performance and scalability
  - The system is fast at processing and analyzing feedback.
  - The system will allow manger can create a new institution for the department.
- Usability
  - The system will be easy to use and the interface will be familiar to all kinds of users.
- Availability
  - The system should available 24/7 for all Egyptians.
- Safety
  - The feedback is only accessed by the manager.
- Cultural and Political
  - Customer personal information is protected in compliance with the Data Protection by the government.

Use Case Diagram

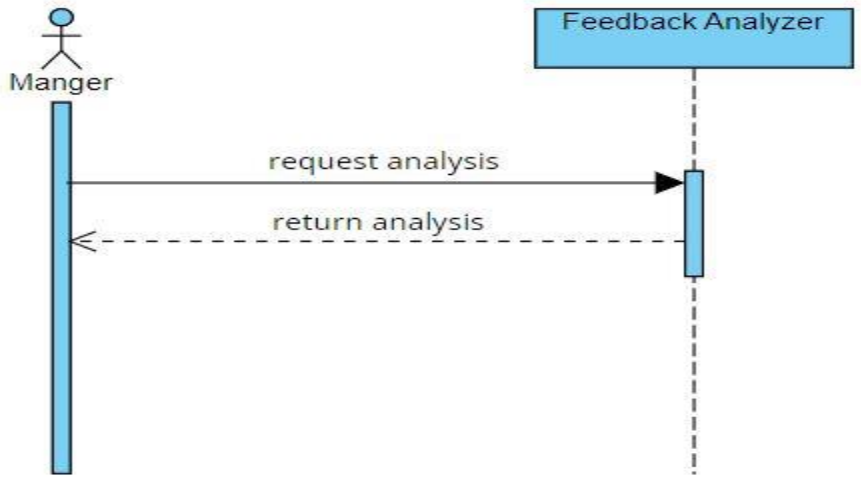


Class Diagram

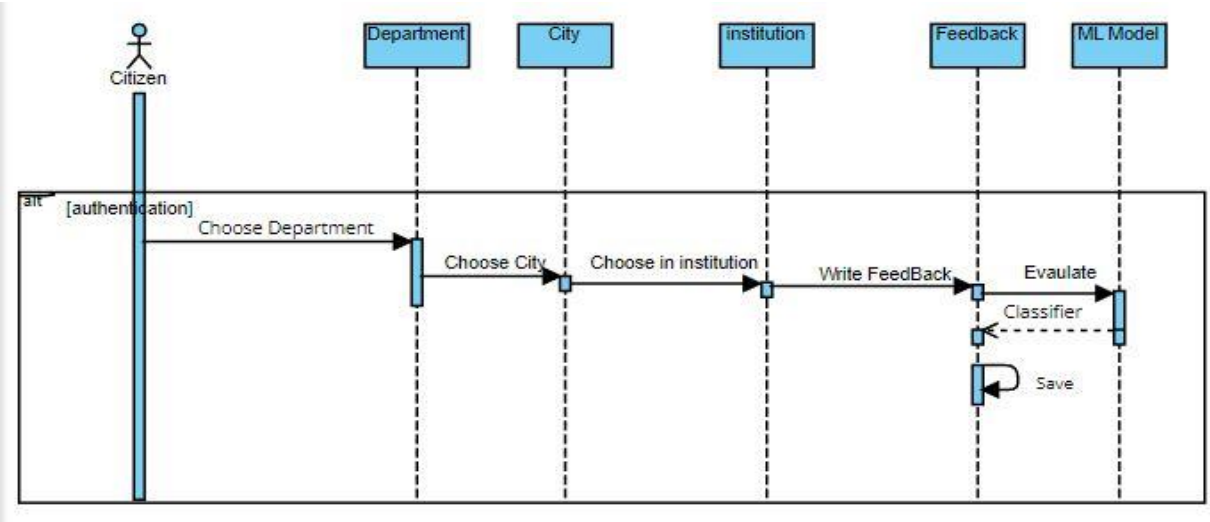


Sequence Diagram

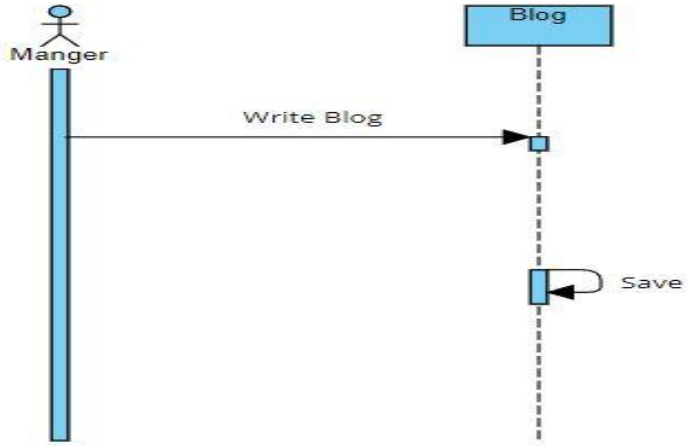
1- The manager looks for analysis based on previous reviews statistics.



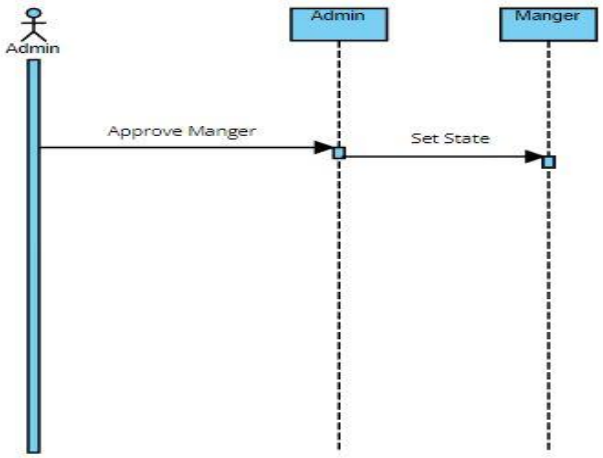
2- Citizens write Feedback after choosing a certain department.



3- Manager posts a blog about his section.



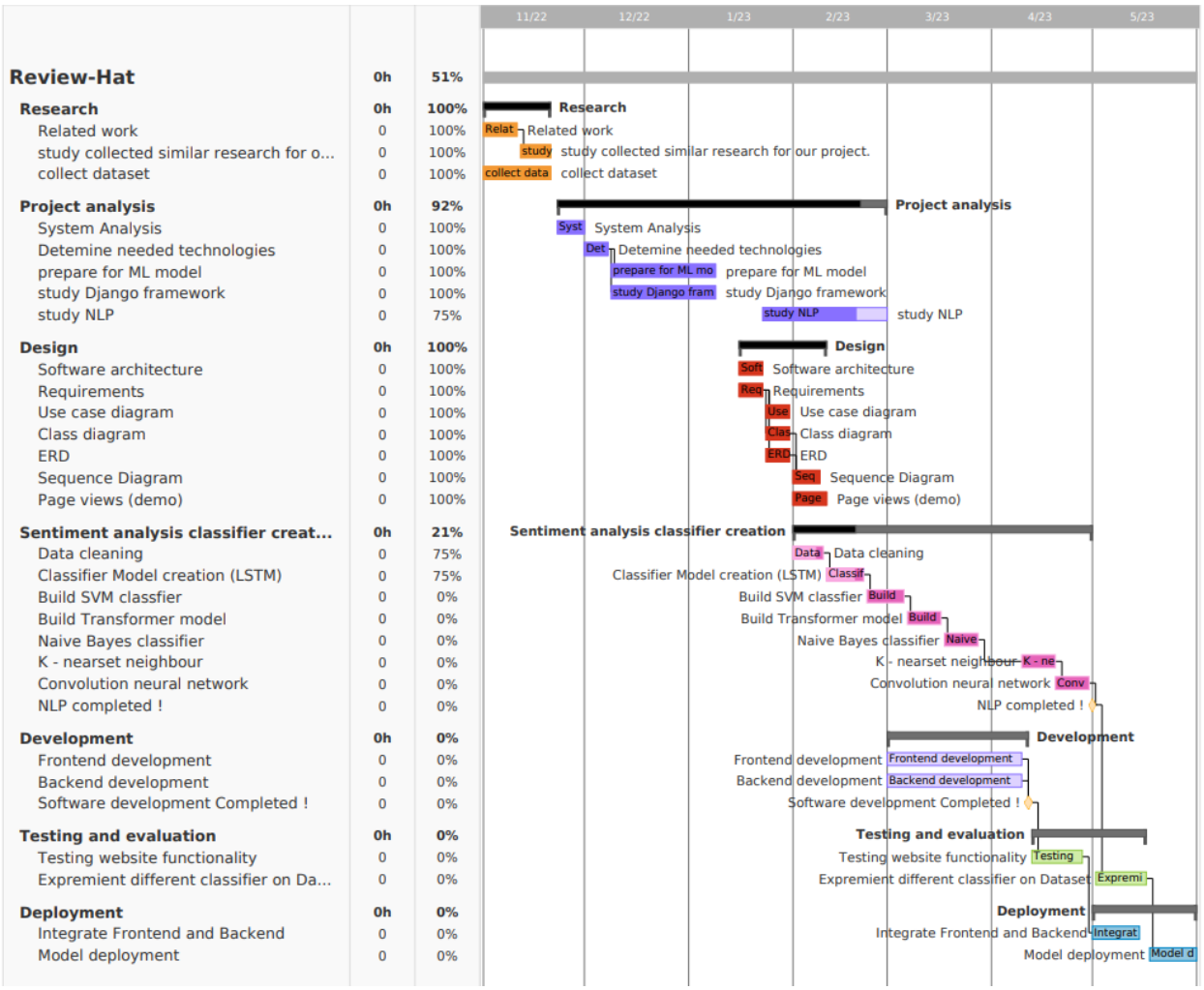
4- Admins approve manager.



Work Plan & Gantt Chart

Task	Task Title	Description	Task status
1-	Searching for related work	Search on the internet for some research like our project.	Completed
		Studying and understanding our collected similar research to our project.	
2-	Collecting Datasets	Searching for different useful datasets to help us in our project.	Completed
3-	Project Analysis	Making System Analysis.	Completed
4-	Study the needed Technologies	Studying Machine Learning and Deep Learning and Natural language processing (NLP).	Completed
5-	Study the needed Technologies	Studying Django and Mango DB and Transformers.	Working on
6-	Experiment different models on Dataset	Try different classifier on the dataset and compare results.	Working on
7-	Page views	Implement the frontend of different web pages.	Completed
8-	Mid-year Document	Writing the Mid-year Document to describe our project.	Completed

9-	Project Design	Making design for the project.	Completed
10-	Build our ML Do the	Start writing our own code to build ours.	Working on
11-	requirements for the final discussion	Write the final documentation and recap the requirements.	Planned
12-	Build Backend	Build authentication, Database, and restrictions.	Planned
13-	Integration	Integrate the frontend with backend and ML model.	Planned
14-	Testing	Evaluating and verifying that a software application does what it is supposed to do	Planned


11/22
12/22
1/23
2/23
3/23
4/23
5/23
Review-Hat 0h 51%
Research 0h 100%
Related work 0 100%
study collected similar research for o... 0 100%
collect dataset 0 100%
Project analysis 0h 92%
System Analysis 0 100%
Detemine needed technologies 0 100%
prepare for ML model 0 100%
study Django framework 0 100%
study NLP 0 75%
Design 0h 100%
Software architecture 0 100%
Requirements 0 100%
Use case diagram 0 100%
Class diagram 0 100%
ERD 0 100%
Sequence Diagram 0 100%
Page views (demo) 0 100%
Sentiment analysis classifier creat... 0h 21%
Data cleaning 0 75%
Classifier Model creation (LSTM) 0 75%
Build SVM classfier 0 0%
Build Transformer model 0 0%
Naive Bayes classifier 0 0%
K - nearset neighbour 0 0%
Convolution neural network 0 0%
NLP completed ! 0 0%
Development 0h 0%
Frontend development 0 0%
Backend development 0 0%
Software development Completed ! 0 0%
Testing and evaluation 0h 0%
Testing website functionality 0 0%
Expremient different classifier on Da... 0 0%
Deployment 0h 0%
Integrate Frontend and Backend 0 0%
Model deployment 0 0%
Research
Relat
Related work
study
study collected similar research for our project.
collect data
collect dataset
Project analysis
Syst
System Analysis
Det
Detemine needed technologies
prepare for ML mo
prepare for ML model
study Django fram
study Django framework
study NLP
study NLP
Design
Soft
Software architecture
Req
Requirements
Use
Use case diagram
Clas
Class diagram
ERD
ERD
Seq
Sequence Diagram
Page
Page views (demo)
Sentiment analysis classifier creation
Data
Data cleaning
Classifier Model creation (LSTM)
Classif
Build SVM classfier
Build
Build Transformer model
Build
Naive Bayes classifier
Naive
K - nearset neighbour
K - ne
Convolution neural network
Conv
NLP completed !
Development
Frontend development
Frontend development
Backend development
Backend development
Software development Completed !
Testing and evaluation
Testing website functionality
Testing
Expremient different classifier on Dataset
Expremi
Deployment
Integrate Frontend and Backend
Integrat
Model deployment
Model d