ÇANKAYA UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

CENG 407
Sentiment Analysis of the Feedback from Airplane Passengers

SDD REPORT

Merve Şirin - 201626404
İrem Şahingöz - 201611051

2020-2021

Contents

1. Introduction

1.1. Purpose

The purpose of this SDD is to describe the project that analyzes the comments made by the airport passengers on Twitter. This project aims to classify the comments made by the passengers on Twitter to evaluate the quality of the service that the American airport provides to its passengers and to conduct an emotional analysis. In this project, it is aimed to measure the satisfaction level of airport passengers based on the comments they make on Twitter to evaluate the service they receive. This project allows American airports to provide a better service to their passengers by improving the service they provide according to the results of this analysis.

1.2. Scope

This project will be very beneficial for airport companies. It will be boring for airports to read the comments made by airport passengers one by one and determine the satisfaction level of the passengers. In this project, we will use machine learning methods to classify the comments made by the passengers about the service they receive and conduct sentiment analysis. According to the comments of the passengers, we will classify their emotions as positive, negative, and neutral, and we will facilitate the work of airport operators by determining the general level of satisfaction. We need a lot of data for accurate analysis. We are considering extracting data from the Kaggle platform to satisfy our data needs. We intend to use SVM classifier algorithms, whose classifications yielded better results than the inferences we made from previous studies. We will use the SVM classifier, a controlled machine learning algorithm that can be used for classification or regression problems in our project. SVM training algorithm works by creating a model that assigns new examples to these categories by transforming it into a linear classifier when a series of training examples marked in different categories are given. SVM is an algorithm that is actively used and preferred in many machine learning projects. The biggest reason we chose it is that it has given better results in previous studies on this subject.

1.3. Glossary

Terminology	Definition
SDD	Software Design Document
SVM	Support Vector Machine
Diagram	Symbolic representation of information using visualization techniques.
Sentiment	A view or opinion that is held or expressed.

1.4. Overview of Document

The contents of the chapters that follow this document as listed below:

Chapter 2 has been prepared with diagrams and explanations to provide an overview of the work we will do, the functioning of the project and the path to be followed.

In the last chapter, in Chapter 3, the implementation of sentiment analysis process classification algorithms is mentioned.
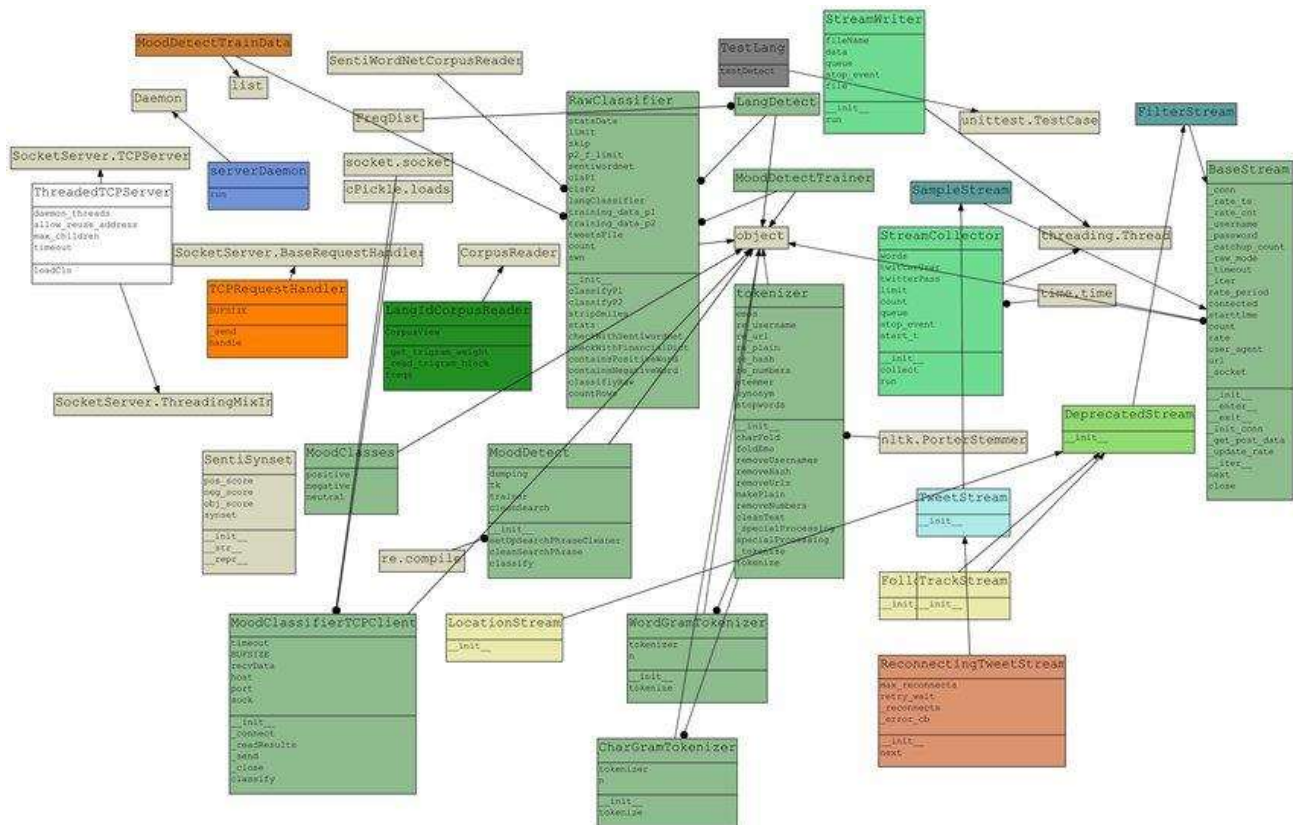
1.5. Motivation

We are senior students of the computer engineering department. We like to work on artificial intelligence and machine learning. For this reason, we chose this subject as our project topic. We are students who enjoy researching, improving ourselves, and learning new things. We wanted to improve ourselves more in the field we love by choosing a project topic that belongs to our field of interest. Our most important source of motivation is that this project will benefit airport companies by facilitating their work.
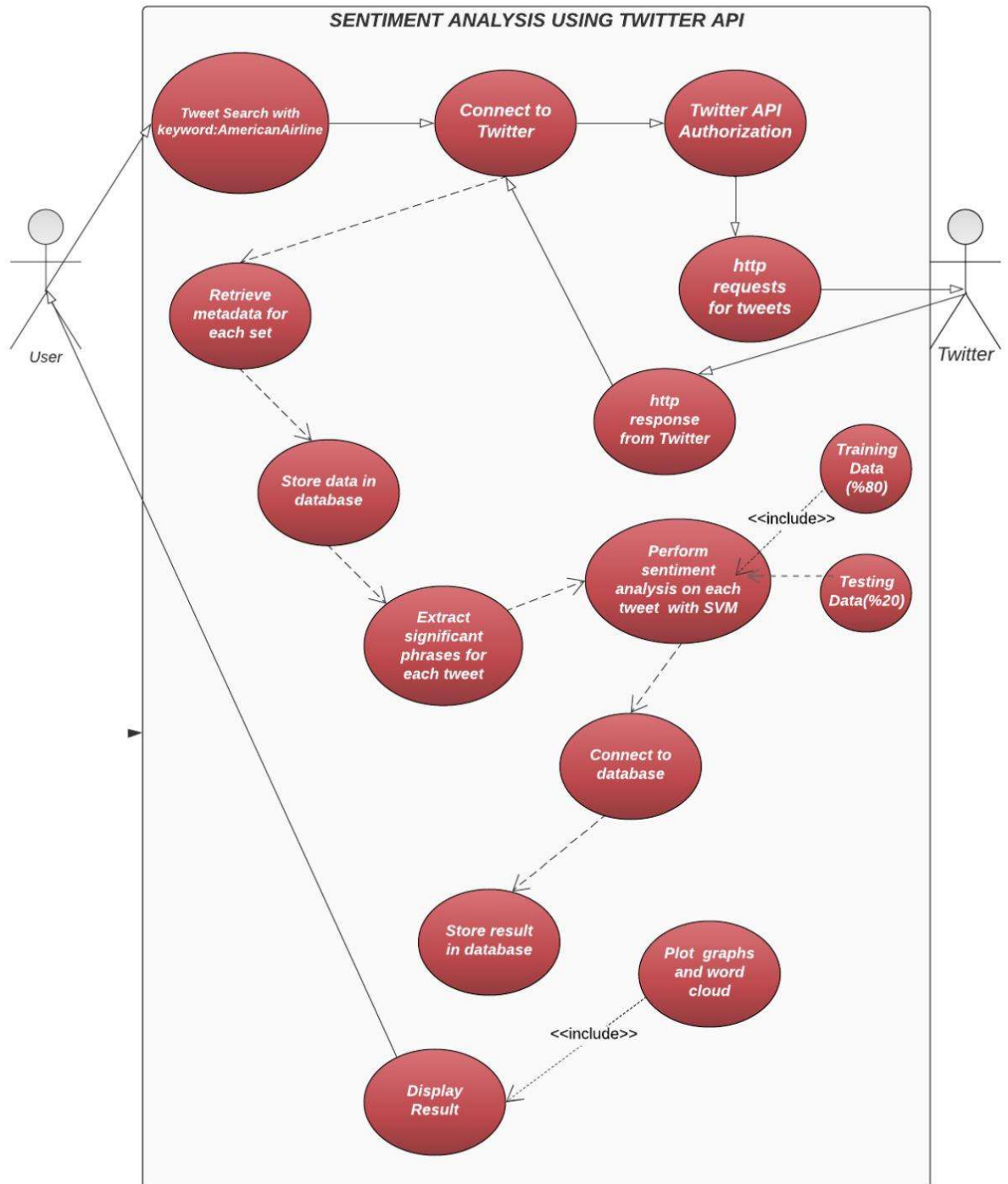
2. Design Overview

In this part, we give lots of details for our project. We explain all diagrams step by step.
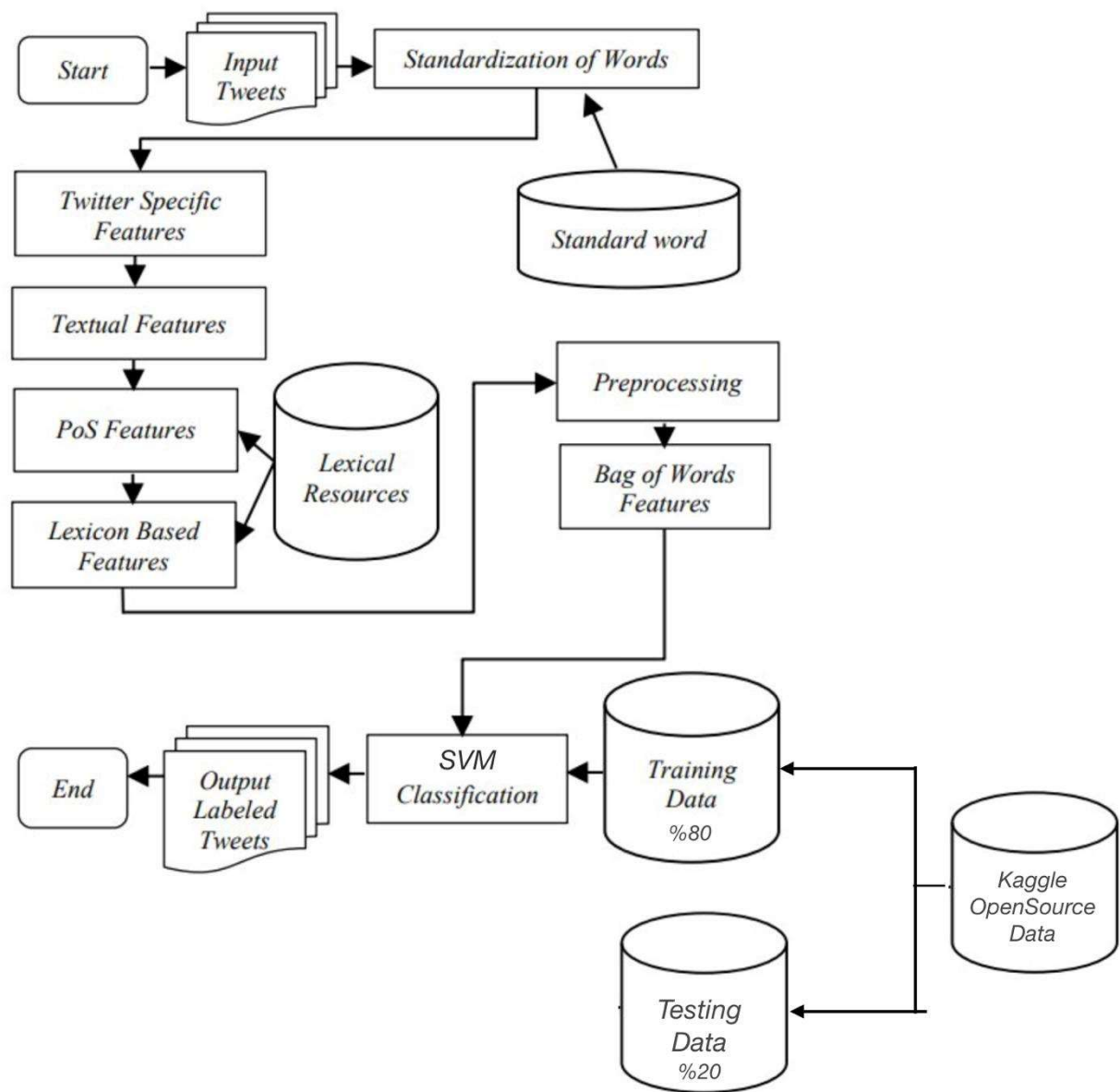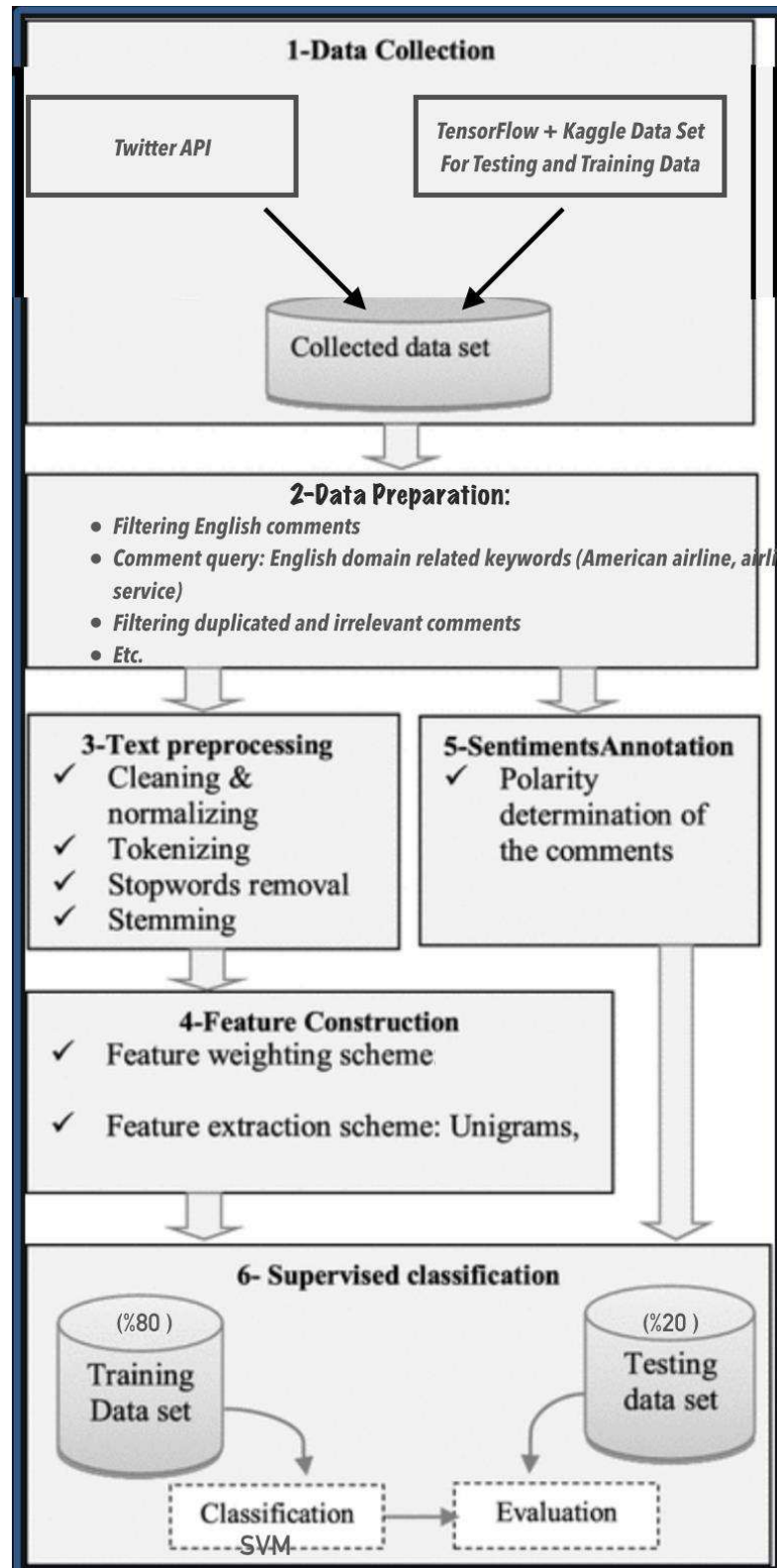
2.1. Class Diagram:

2.2. Use Case Diagram:

SENTIMENT ANALYSIS USING TWITTER API

- Tweet Search with keyword:AmericanAirline
- Connect to Twitter
- Twitter API Authorization
- http requests for tweets
- Retrieve metadata for each set
- http response from Twitter
- Training Data (%80)
- Store data in database
- <<include>>
- Perform sentiment analysis on each tweet with SVM
- Testing Data(%20)
- Extract significant phrases for each tweet
- Connect to database
- Store result in database
- Plot graphs and word cloud
- <<include>>
- Display Result

User

Twitter

2.3. System Flow Diagram:

2.4. Time Series Processing of System:

1-Data Collection

Twitter API

TensorFlow + Kaggle Data Set
For Testing and Training Data

Collected data set

2-Data Preparation:
- Filtering English comments
- Comment query: English domain related keywords (American airline, airli
  service)
- Filtering duplicated and irrelevant comments
- Etc.

3-Text preprocessing
✓ Cleaning & normalizing
✓ Tokenizing
✓ Stopwords removal
✓ Stemming

5-SentimentsAnnotation
✓ Polarity determination of the comments

4-Feature Construction
✓ Feature weighting scheme
✓ Feature extraction scheme: Unigrams,

6- Supervised classification

(%80)
Training Data set

(%20)
Testing data set

Classification SVM

Evaluation

2.5. General Natural Language Processing Perspective:

Natural Language Processing Pipeline

2.6. Text Preprocessing:
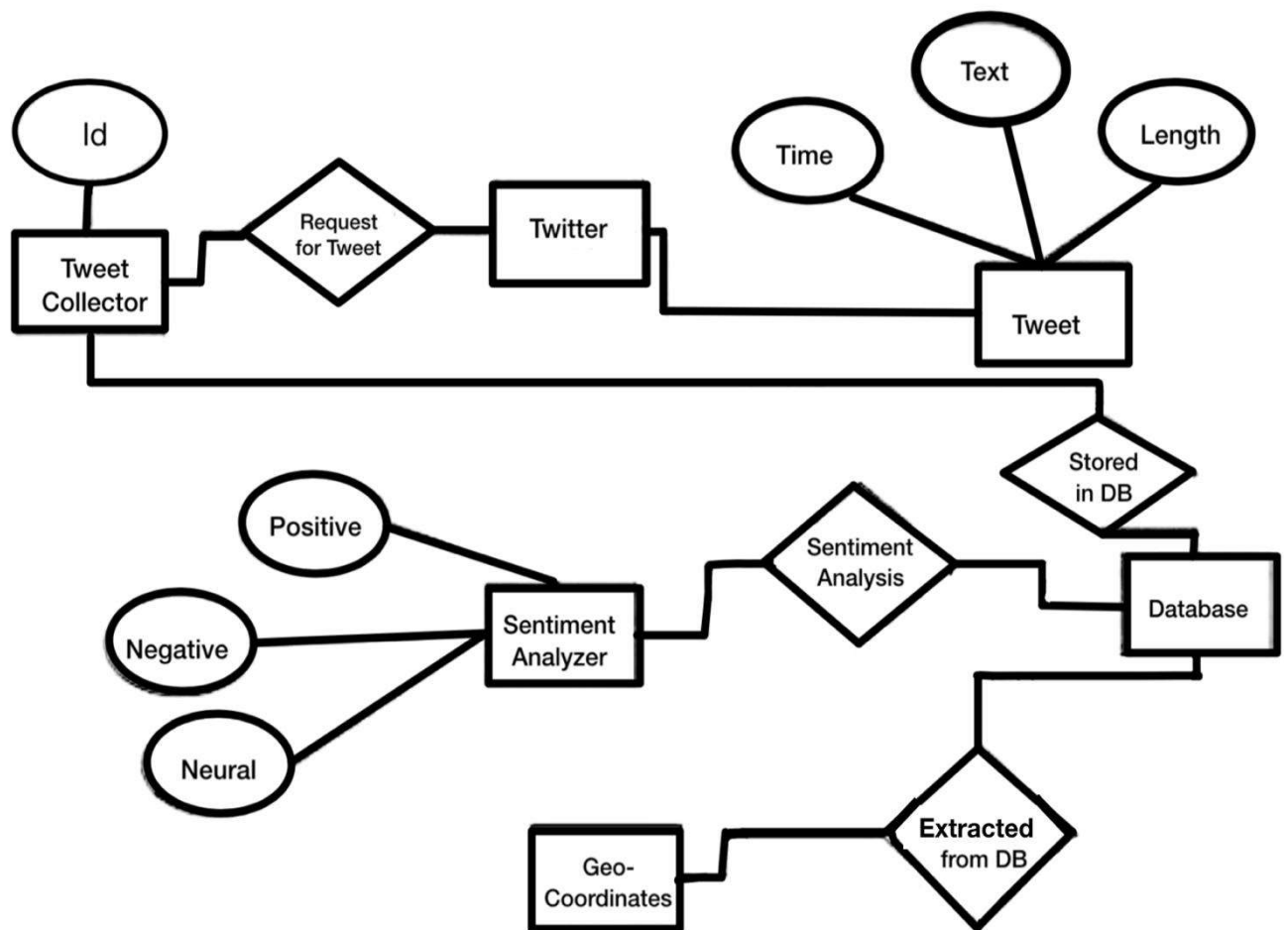

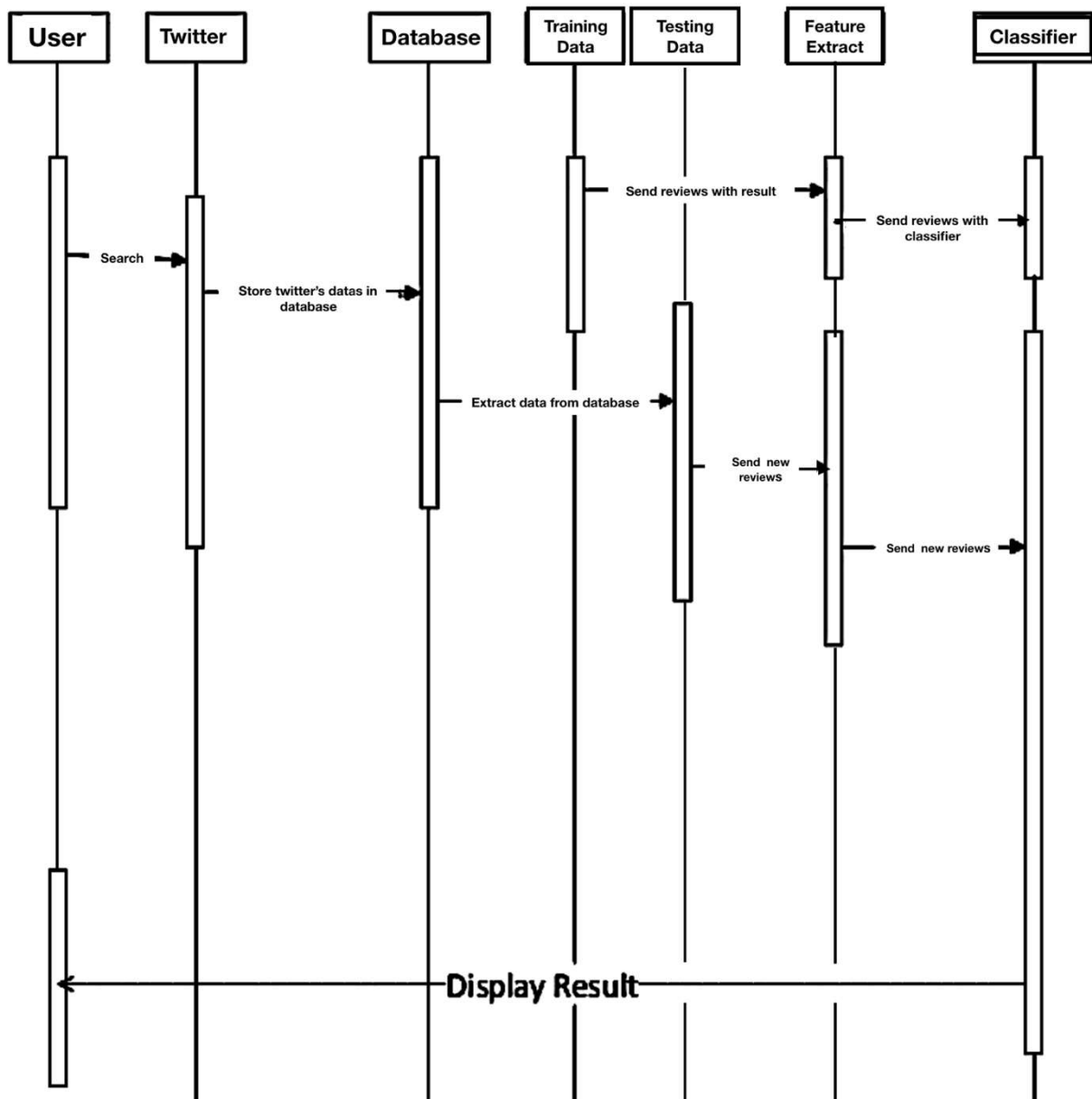Natural Language Processing Pipeline – Data Preparation
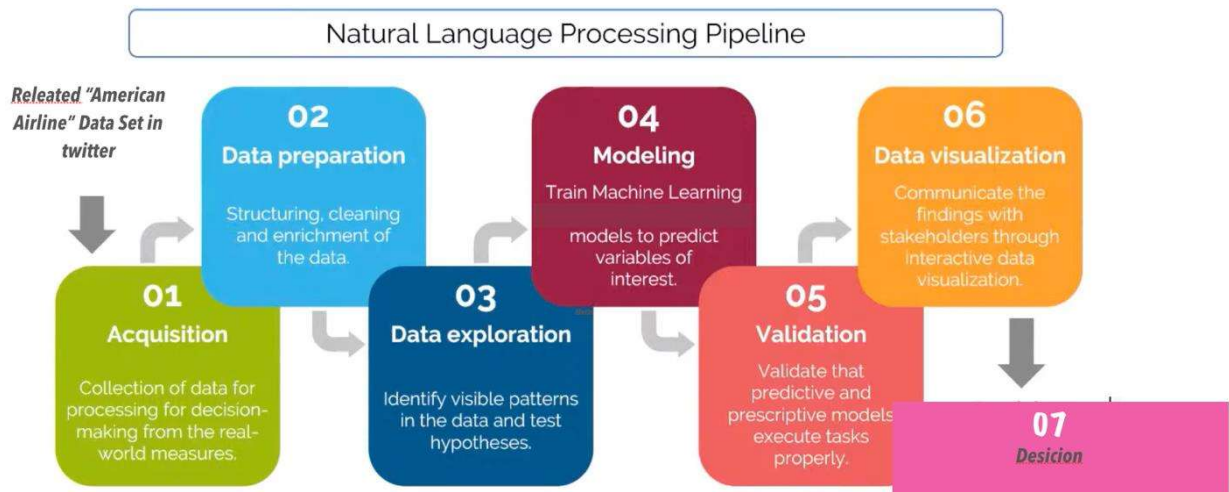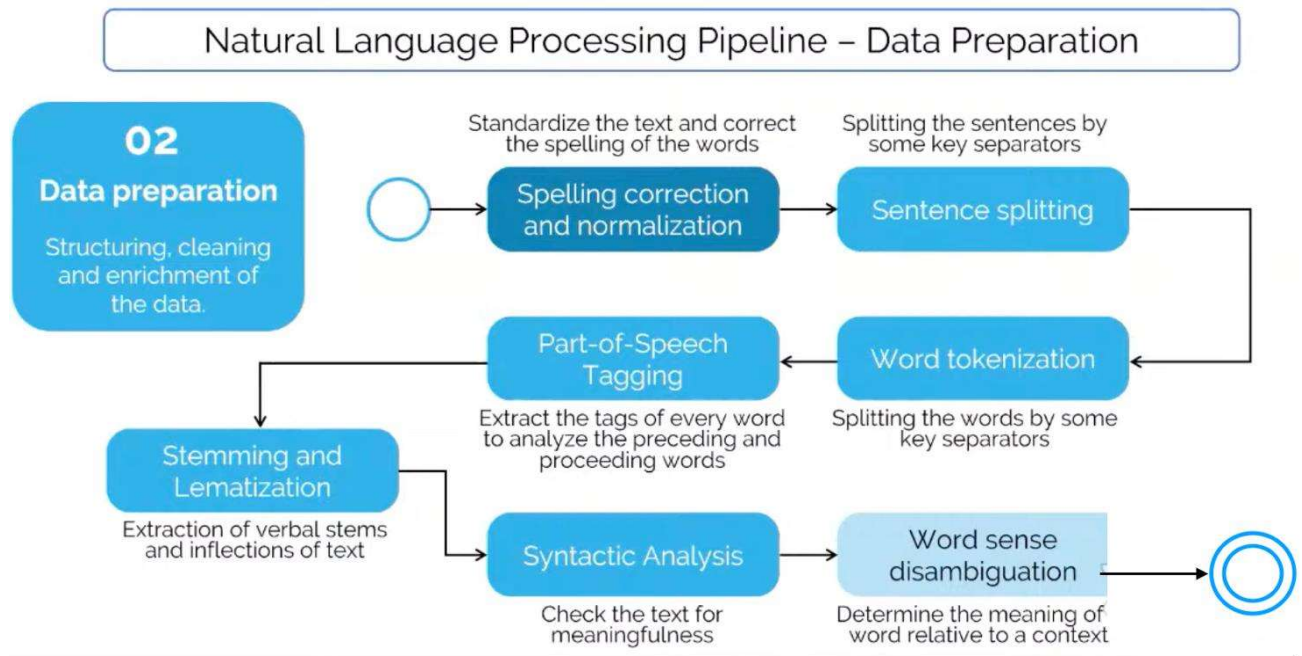
2.7. Activity Diagram:
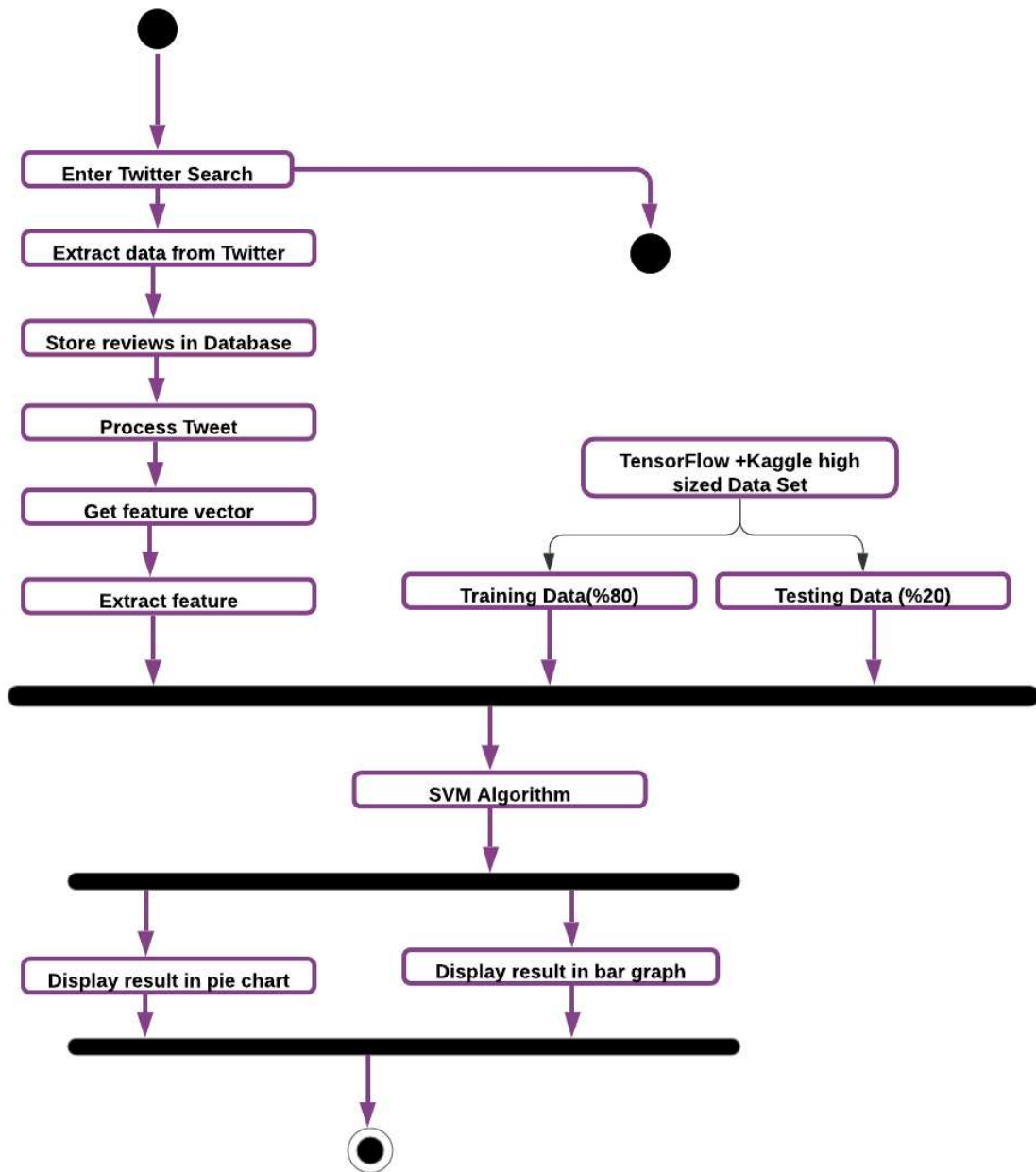
2.8. ER Diagram:

2.9. Sequence Diagram:

3. Detection

In this project, the comments made by users as a text input on Twitter are classified using machine learning methods. Twitter comments made by the airport passengers as text input are divided into smaller meaningful units by tokenization. After tokenization, words of unnecessary frequently used words, that is, stop words, should be filtered. With negation handling, negative words and their combinations with positive or negative words are recognized. Word roots are found with stemming. Then the comments are divided into classes using the classification methods. We will use the SVM classifier for classification. In this way, we will do sentiment analysis for airport passengers by following a route.

4. **References:**

[1] https://www.datacamp.com/community/tutorials/simplifying-sentiment-analysis-python

[2]   https://www.houseofbots.com/news-detail/2491-4-5-things-you-need-to-know-about-sentiment-analysis-and-classification