BATCH NO:18A10

A Phase-1 Project Report on

SPAM REVIEW DETECTION USING LINGUISTIC AND BEHAVIORAL METHODS

Submitted in partial fulfillment of the requirement
for the award of the degree of

Bachelor of Technology in Information Technology for AY 2021-22 By

Name of the Candidate	Registration Number
G. Chandana	18251A1210
C. Rakshitha	18251A1235
J. Sri Harshini	18251A1242
M. Deekshitha	18251A1252

Department of IT

G. Narayanamma Institute of Technology and Science

(AUTONOMOUS) (For Women)

Approved by AICTE, New Delhi & Affiliated to JNTUH, Hyderabad
Accredited by NBA, An ISO 9001:2015 Certified Institution
Shaikpet, Hyderabad – 500104, TS.

July, 2021.

CERTIFICATE

This is to certify that the Phase-1 Project Report entitled **Spam Review Detection Using Linguistic and Behavioral Methods** is a bonafide work done by

Name of the Candidate	Registration Number
G. Chandana	18251A1210
C. Rakshitha	18251A1235
J. Sri Harshini	18251A1242
M. Deekshitha	18251A1252

under the guidance of **Mrs. Ch. Sravanthi**, during **September 2021 to January 2022**, in partial fulfilment for the award of degree in B. Tech. in Information Technology, from G. Narayanamma Institute of Technology and Science.

Internal Guide
(Mrs. Ch. Sravanthi)

Head of the Department
(Dr. I. Ravi Prakash Reddy)

ACKNOWLEDGEMENT

We would like to express our sincere thanks to **Dr K. Ramesh Reddy, Principal**, GNITS, for providing the good working environment and facilities in the college. A sincere gratitude to **Dr I. Ravi Prakash Reddy**, **Professor and HOD**, Dept. of Information Technology, GNITS for encouraging us to take up the project and all the timely support during the time of mini-project. We respectfully acknowledge **Ms. Ch. Sravanthi**, Department of Information Technology for her immense support and guidance during the whole time of the phase-1 project. Finally, we would also like to thank all the faculty and staff of IT Department who helped me directly or indirectly, parents and friends for their cooperation in completing the project work.

ABSTRACT

Online reviews regarding different products or services have become the main source to determine public opinions. Consequently, manufacturers and sellers are extremely concerned with customer reviews as these have a direct impact on their businesses. Unfortunately, to gain profits or fame, spam reviews are written to promote or demote targeted products or services. This practice is known as review spamming. In recent years, the spam review detection problem has gained much attention from communities and researchers, but still there is a need to perform experiments on real-world large-scale review datasets. This can help to analyze the impact of widespread opinion spam in online reviews. In this work, two different spam review detection methods have been proposed: (1) Spam Review Detection using Behavioral Method (SRD-BM) utilizes thirteen different spammer's behavioral features to calculate the review spam score which is then used to identify spammers and spam reviews, and (2) Spam Review Detection using Linguistic Method (SRD-LM) works on the content of the reviews and utilizes transformation, feature selection and classification to identify the spam reviews. To the best of our knowledge, this is the first study of its kind which uses a large-scale review dataset to analyze different spammers' behavioral features and linguistic method utilizing different available classifiers.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

1. INTRODUCTION

1.1. General/Domain Description

Nowadays, the World Wide Web (WWW) has become the main source for individuals to express themselves. People can easily share their views about any product or service by using e-commerce sites, forums and blogs. Everybody on the web is now acknowledging the importance of these online reviews for both customers and vendors. Most people read reviews about products and services before buying them. Vendors can also design their future production or marketing strategies based on these reviews . For example, if various customers buying a specific model of a laptop, post reviews about issues related to its screen design, the manufacturer can be aware and resolve this issue to increase customer satisfaction.

Recently, the trend of spam review attacks has increased because anybody can simply write spam reviews and post them online without any constraint. Anyone can hire people to write fake reviews for their products and services, such people are called spammers. Spam reviews are usually written to gain profits or to promote a product or service. This practice is known as review spamming. The main problem with opinion sharing websites is that spammers can easily create hype about the product by writing spam reviews. These spam reviews can play a key role in increasing the value of a product or service. For example, if a customer wants to purchase a product online, he/she usually goes to the review section to know about other buyers' feedback. If the reviews are mostly positive, the user may purchase it, otherwise, he/she would not buy that specific product. This all shows that spam reviews have become the main problem in online shopping, which can cause loss to both the customers and manufacturers.

What is Machine Learning?

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.

The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers.

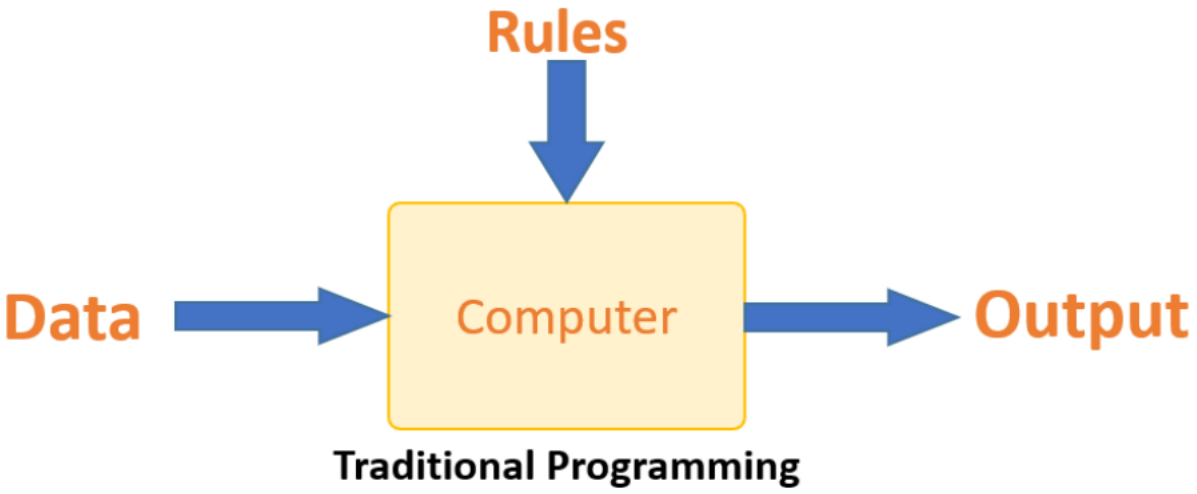
A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendations.

Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automatizing tasks and so on.

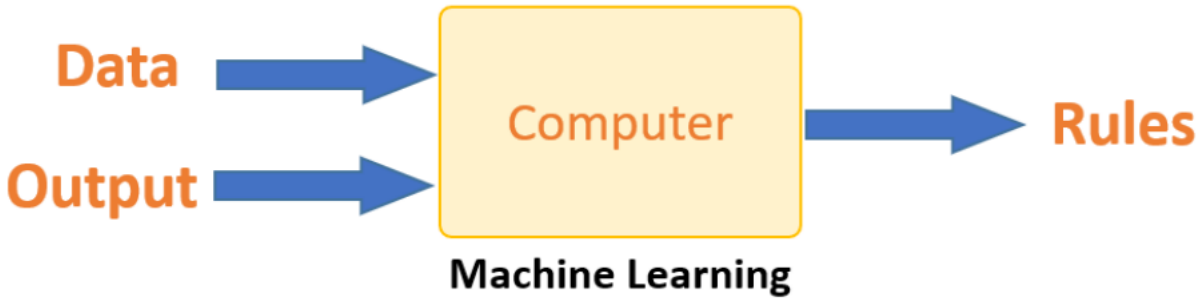
Machine Learning vs. Traditional Programming

Traditional programming differs significantly from machine learning. In traditional programming, a programmer codes all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.



Machine learning is supposed to overcome this issue. The machine learns how the input and output data are correlated and it writes a rule. The programmers do not need to write new rules each time there is new data. The algorithms adapt in response to new data and experiences to improve efficacy over time.

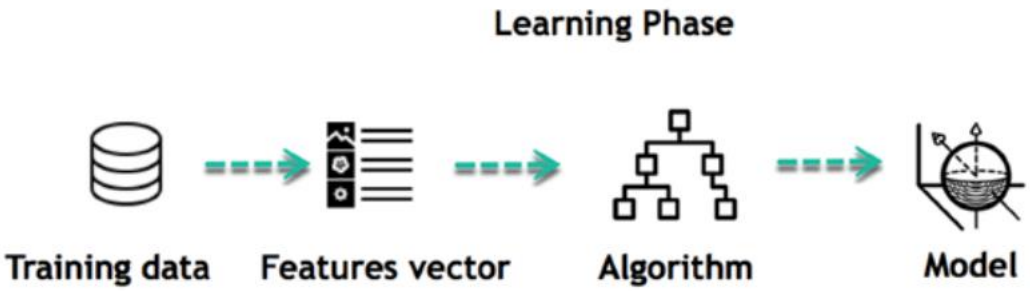


How does Machine Learning Work?

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict.

The core objective of machine learning is the **learning** and **inference**. First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the **data**. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a **feature vector.** You can think of a feature vector as a subset of data that is used to tackle a problem.

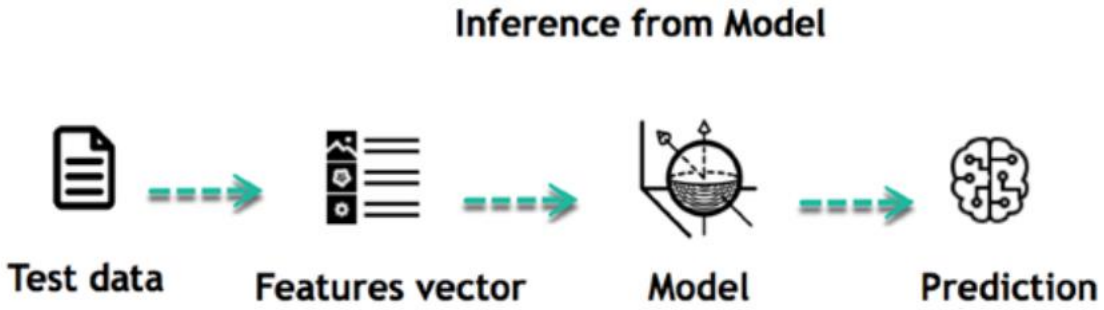
The machine uses some fancy algorithms to simplify the reality and transform this discovery into a **model**. Therefore, the learning stage is used to describe the data and summarize it into a model.



For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant: This is the model

Inferring

When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.



The life of Machine Learning programs is straightforward and can be summarized in the following points:

1. Define a question
2. Collect data
3. Visualize data
4. Train algorithm
5. Test the Algorithm

6. Collect feedback

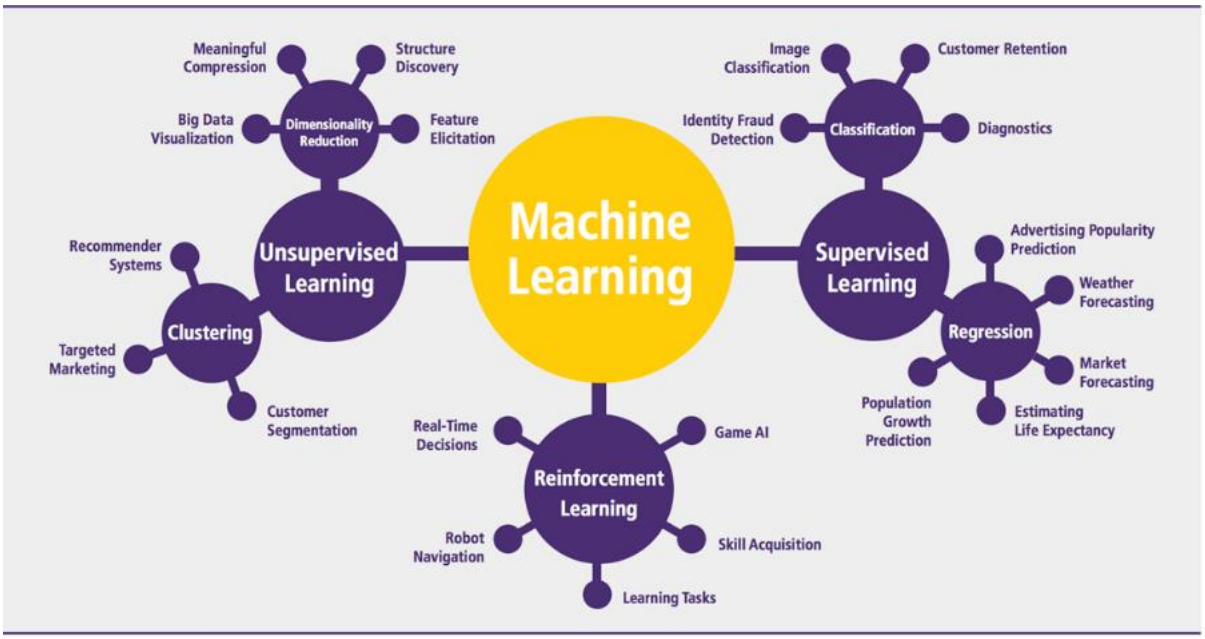
7. Refine the algorithm

8. Loop 4-7 until the results are satisfying

9. Use the model to make a prediction

Once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data.

Machine Learning Algorithms and Where they are Used?



Machine learning can be grouped into two broad learning tasks: Supervised and Unsupervised. There are many other algorithms

Supervised learning

An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output. For instance, a practitioner can use marketing expense and weather forecast as input data to predict the sales of cans.You can use supervised learning when the output data is known. The algorithm will predict new data.There are two categories of supervised learning:

- Classification task

- Regression task

Classification

Imagine you want to predict the gender of a customer for a commercial. You will start gathering data on the height, weight, job, salary, purchasing basket, etc. from your customer database. You know the gender of each of your customers, it can only be male or female. The objective of the classifier will be to assign a probability of being a male or a female (i.e., the label) based on the information (i.e., features you have collected). When the model learns how to recognize male or female, you can use new data to make a prediction. For instance, you just got new information from an unknown customer, and you want to know if it is a male or female. If the classifier predicts male = 70%, it means the algorithm is sure at 70% that this customer is a male, and 30% it is a female.

The label can be of two or more classes. The above Machine learning example has only two classes, but if a classifier needs to predict object, it has dozens of classes (e.g., glass, table, shoes, etc. each object represents a class)

Regression

When the output is a continuous value, the task is a regression. For instance, a financial analyst may need to forecast the value of a stock based on a range of feature like equity, previous stock performances, macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error.

Unsupervised learning

Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms, therefore, learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. A central application of unsupervised learning is in the field of density estimation in statistics, such as finding the probability density function. Though unsupervised learning encompasses other domains involving summarizing and explaining data features.

Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within the same cluster are similar according to one or more predesignated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some similarity metric and evaluated, for example, by internal compactness, or the similarity between members of the same cluster, and separation, the difference between clusters. Other methods are based on estimated density and graph connectivity.

Reinforcement learning

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics and genetic algorithms. In machine learning, the environment is typically represented as a Markov decision process (MDP). Many reinforcement learning algorithms use dynamic programming techniques. Reinforcement learning algorithms do not assume knowledge of an exact mathematical model of the MDP, and are used when exact models are infeasible. Reinforcement learning algorithms are used in autonomous vehicles or in learning to play a game against a human opponent.

Application of Machine Learning

Augmentation:

Machine learning, which assists humans with their day-to-day tasks, personally or commercially without having complete control of the output. Such machine learning is used in different ways such as Virtual Assistant, Data analysis, software solutions. The primary user is to reduce errors due to human bias.

Automation:

Machine learning, which works entirely autonomously in any field without the need for any human intervention. For example, robots performing the essential process steps in manufacturing plants.

Finance Industry:

Machine learning is growing in popularity in the finance industry. Banks are mainly using ML to find patterns inside the data but also to prevent fraud.

Government organization:

The government makes use of ML to manage public safety and utilities. Take the example of China with its massive face recognition. The government uses Artificial intelligence to prevent jaywalkers.

Healthcare industry:

Healthcare was one of the first industries to use machine learning with image detection.

Marketing:

Broad use of AI is done in marketing thanks to abundant access to data. Before the age of mass data, researchers developed advanced mathematical tools like Bayesian analysis to estimate the value of a customer. With the boom of data, the marketing department relies on AI to optimize the customer relationship and marketing campaign.

1.2. Objective and Scope of the Project

In this work, two different spam review detection methods have been proposed: (1) Spam Review Detection using Behavioural Method (SRD-BM) utilizes thirteen different spammer's behavioural features to calculate the review spam score which is then used to identify spammers and spam reviews, and (2) Spam Review Detection using Linguistic Method (SRD-LM) works on the content of the reviews and utilizes transformation, feature selection and classification to identify the spam reviews.

The aim of this work is to develop an SRD model adapting a vast set of behavioural and linguistic features on large-scale real-world dataset.

To develop a game that recognizes the static hand gestures which are recorded in stable lighting and simple background conditions.

1.3. Project Definition

Consequently, manufacturers and sellers are extremely concerned with customer reviews as these have a direct impact on their businesses. Unfortunately, to gain profits or fame, spam reviews are written to promote or demote targeted products or services. This practice is known as review spamming. In recent years, the spam review detection problem has gained much attention from communities and researchers, but still there is a need to perform experiments on real-world large-scale review datasets. This can help to analyze the impact of widespread opinion spam in online reviews. In this work, two different spam review detection methods have been proposed: (1) Spam Review Detection using Behavioral Method (SRD-BM) utilizes thirteen different spammer's behavioral features to calculate the review spam score which is then used to identify spammers and spam reviews, and (2) Spam Review Detection using Linguistic Method (SRD-LM) works on the content of the reviews and utilizes transformation, feature selection and classification to identify the spam reviews.

2. LITERATURE SURVEY

2.1. Existing System

Spam review detection using the spammer behavioural method finds the unusual patterns and relationships between different spammers. Only a few studies have explored spam review detection using the spammer behavioural method to date. They are:

[1]Mukherjee developed a spam review detection method using a clustering technique by modelling the spam city of the reviewer to identify spammer and non-spammer clusters.[1]

2.1.1.Heydari have proposed a model incorporating only the time series feature of the reviewer on an amazon real dataset.[2]

2.1.2 Kumar has proposed a hierarchical supervised learning method. This method analysed reviewers behavioural features and their interaction using multivariate distribution.

2.1.3 Zang recommended a supervised model based on reviewer features to identify spam reviews.

2.1.4 Li used the feature based sparse additive generative model and the svm classifier to discover the general rule for spam review detection.

[5] have proposed a hierarchical supervised-learning method. This method analysed reviewer's behavioural features and their interactions using multivariate distribution. Zhang et al.

[6] **Ahmed and Danti** used various rule-based machine learning algorithms. Moreover, the authors compared the effectiveness of the proposed method through a ten-fold cross-validation training model for sentiment classification.

2.2. Issues in Existing System

Based on the review of spammer behavioural models ,it has been observed that most of the existing studies have only utilized time series based spammer behavioural feature. It is analysed that utilizing rich set of behavioural features can help in improving the accuracy of spammer identification. Therefore, the proposed behavioural framework utilizes thirteen different spammer behavioural features to calculate spam score in spam review identification.

Based on the literature review it has been observed that most of the existing studies did not incorporate a number of important linguistic features while designing linguistic feature-based SRD models and utilized only one classifier to train their proposed models.

2.3. Proposed System

Proposed methods utilized real world large scale amazon review dataset. Proposed SRD-BM which incorporated thirteen different behavioural features to identify spammers and spam reviews. Proposed SRD-LM which utilized linguistic features and classifiers to identify spam reviews. Compared and analysed the accuracy of proposed SRD-BM and SRD-LM.

2.4 Advantages of proposed system

The linguistic feature considers review text to identify the review as spam or not spam :where as behavioural features reflect the behaviour of reviewer in terms of time stamp of the review rating user profile etc. Most of the existing systems have only utilized the uni-gram linguistic approach to classify reviews. Usually the uni-gram approach produces good results but fails in some cases.

Based on the literature review, it has been observed that most of the existing studies did not incorporate a number of important linguistic features while designing the linguistic feature based SRD models and utilized only one classifier to train their proposed models. The current study therefore extends the SRD domain to design a linguistic model utilizing several features, including stemming and N-gram techniques. These features have significantly improved the accuracy of the proposed model in spam review identification. Moreover the proposed model utilizes and compares the accuracy of four different classifiers, including Naive Bayes(NB),Logistic Regression(LR),Support Vector Machine(SVM),and Random Forest Classifier(RF) to further improve the accurate prediction of spam review.

3. REQUIREMENT SPECIFICATION

3.1. Introduction

The requirements are grouped by their stakeholders, and functional and non-functional requirements are separated.

3.2. System environment

Spam Review Detection Model environment consists of 1 active Actor and Co-operating system. The user accesses the platform using the internet. Any active user can login through the system using his saved credentials otherwise by registering himself to the system.

3.3. Functional requirements specification

3.3.1 Use case descriptions

This section outlines the brief description of the use cases for the user.

USE CASE: LOGIN/REGISTER

Use Case Name	Login/Register
Trigger	The User enters the credentials to login/register
Precondition	The User should initially have access to the website
Basic Path	User login through the website by providing his/her credentials.
Alternative Paths	The user should create an account if doesn't exists
Postcondition	The User is directed to the input grid window
Exception Paths	The User may abandon the operation at any time.
Other	None

Table 3.3.1.1 – Use case Login/Register

Diagram:

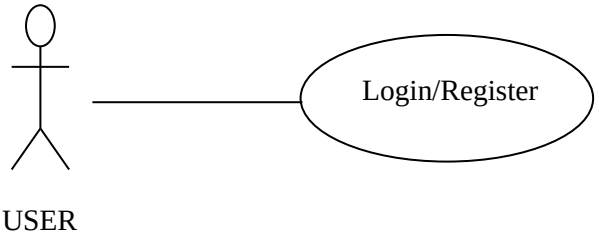


Fig. 3.3.1.1 – Login /Register

Brief Description:

The user should provide the accurate credentials to login. If the account does not exist the user should register.

Initial Step-By-Step Description:

Before, this use case is initiated, the user has already access to internet and opened the website

1. The user clicks on Login button.
2. The user enters the username in the space provided.
3. The user enters the respective password.
4. If the user forgets his/her password, they click on forgot password option.
5. The user clicks on submit button to login successfully.

USE CASE: INPUT REVIEW

Use Case Name	Input Review
Trigger	The user should provide the review and click submit button
Precondition	The user should login with correct credentials and have access to the online website
Basic Path	User enter the correct credentials and write the input review in the grid
Alternative Paths	User may logout if not
Postcondition	The system displays the review is spam or not and user can also get the spam score by clicking the spam score button.
Exception Paths	The User may abandon the operation at any time.
Other	None

Table 3.3.1.2 – Use case Input Review

Diagram:

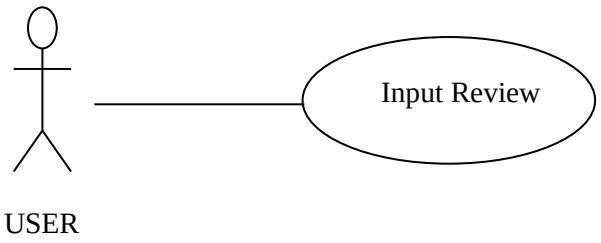


Fig. 3.3.1.2 – Use case Input Review

Brief Description:

The user has access to the website and can provide the input review to check the review is spam or not.

Initial Step-By-Step Description:

Before this use case is initiated the user is already login to the website.

1. The user sees the space provided to write the review.
2. The user enters his/her review in the input grid.
3. The user clicks on spam result button.

USE CASE: VIEW SPAM RESULT

Use Case Name	View Spam Result
Trigger	Submitting the input review by the user in the provided input grid
Precondition	The user initially should the review in the input grid
Basic Path	After submitting the system displays the spam result
Alternative Paths	The user can also get the spam score of the provided input review
Postcondition	The user can be redirected to the input grid window or can view the spam score.
Exception Paths	The attempt may be abandoned at any time.
Other	None

Table 3.3.1.3 – View Spam Result

Diagram:

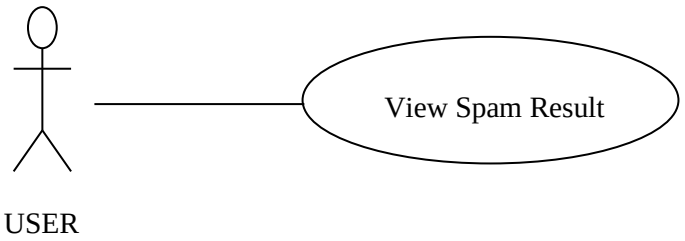


Fig. 3.3.1.3 – Use case View Spam Result

Brief Description:

The spam results can be checked by user after entering the correct input and clicking the spam result button**.**

Initial Step-By-Step Description:

Before this use case in initiated the user is already provided the review in the input grid.

1. The user can view the result of the provided review.
2. The user understands if the provided review is spam or non-spam review.

USE CASE: VIEW SPAM SCORE

Use Case Name	View Spam Score
Trigger	The user should click the spam score button
Precondition	The user should provide the input review
Basic Path	The user gets to view the pie-chart after submitting
Alternative Paths	The activity is completed ,user can go to the home page
Postcondition	User can get the spam score later redirected to the input grid page or user may logout.
Exception Paths	The attempt may be abandoned at any time.
Other	None

Table 3.3.1.4 – Use case View Spam Score

Diagram:

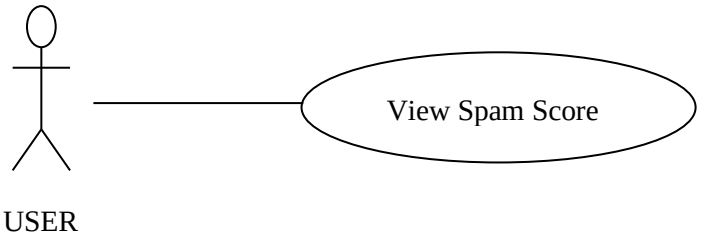


Fig. 3.3.1.4 – Use case View Spam Score

Brief Description:

The user gets the spam score in the form of pie-cart by clicking the spam score button.

Initial Step-By-Step Description:

Before this use case in initiated the user is already provided the review in the input grid.

1. The user clicks on view spam score button.
2. Pie-chart is displayed, which indicates the percentage of spam the review is.
3. User can logout of the website.

USE CASE: PROVIDE SPAM RESULT

Use Case Name	Provide Spam Result
Trigger	The admin receives the data from the user & provide the spam result
Precondition	The user initially should provide the review in the input grid
Basic Path	The admin sends the input review to the algorithm
Alternative Paths	The admin can also provide the spam score
Postcondition	The admin waits for another new input review to provide the spam result.
Exception Paths	The attempt may be abandoned at any time.
Other	None

Table 3.3.1.5 – Provide Spam Result

Diagram:

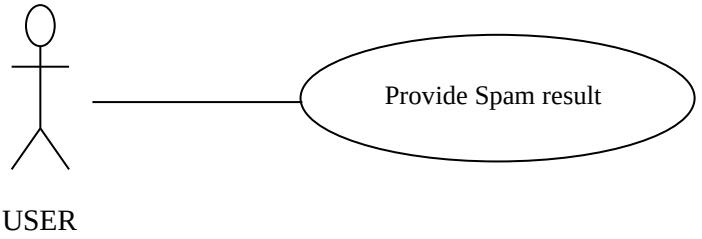


Fig. 3.3.1.5 – Use case Provide Spam Result

Brief Description:

After providing the input review by user the system provides the spam result.

Initial Step-By-Step Description:

Before this use case in initiated the admin receives the review from the input grid provided by the user

1. The admin provides the result of the provided review.

USE CASE: PROVIDE SPAM SCORE

Use Case Name	Provide Spam Score
Trigger	The admin receives the data from the user & provide the spam score
Precondition	The user initially should provide the review in the input grid
Basic Path	The admin sends the input review to the algorithm
Alternative Paths	The process can be terminated if the input is incorrect
Postcondition	The admin waits for another new input review to provide the spam score
Exception Paths	The attempt may be abandoned at any time.
Other	None

Table 3.3.1.6 – Use case Provide Spam Score

Diagram:

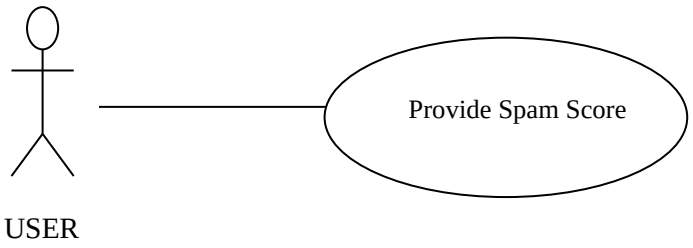


Fig. 3.3.1.6 – Use case Provide Spam Score

Brief Description:

The system provides the spam score after the user clicks the spam score button.

Initial Step-By-Step Description:

Before this use case in initiated the admin is already provided with the review from the input grid.

1. Pie-chart is displayed, which indicates the percentage of spam the review is.
2. Admin can terminate the process if the input is not provided by the user.

3.3.2 List of Functional Requirements

STAKEHOLDERS Table:3.3.2

U	User
A	Admin

Functional Requirements of User

FR1	The system must enable user to register themselves by providing their personal details.
FR2	The system must allow users to login through their username and password after registering.
FR3	The system must allow users to reset their profile.
FR4	The system must provide the spam score to the user
FR5	The system must show whether the review is spam or not

Functional Requirements of Admin

FR6	The system must allow Admin to login with valid credentials.
FR7	The System must allow Admin to access the database of the application.
FR8	The system must allow Admin to add/remove/block users profile.
FR9	The system must provide admin all control over credentials details of users.

3.4 Non-Functional Requirements for User and Admin

Table:3.4

NF1	**Usability** The system is designed with completely automated process hence there is no or less user intervention.
NF2	**Reliability** The system is more reliable because of the qualities that are inherited from the chosen platform Anaconda. The code built by using python is more reliable.
NF3	**Performance** This system is developing in the high level languages and using the advanced front-end and back-end technologies it will give response to the end user on client system with in very less time.
NF4	**Supportability** The system is designed to be the cross platform supportable.
NF5	**Security** The System shall provide security to user's data.

3.5. H/W & S/W Requirements

Software Requirements

Operating System : Windows 11

Web Framework : Flask

Frontend : Java Servlet Pages(JSP),HTML,CSS, Java Script

Database : My SQL

Hardware Requirements

Processor : Intel Core 5 (I5)

RAM : 8 GB

Monitor : 14*11 inches

Memory : 1 TB

Drive Type : Solid State Drive(SSD)

4. DESIGN SPECIFICATION

4.1 Overall Use Case Diagram

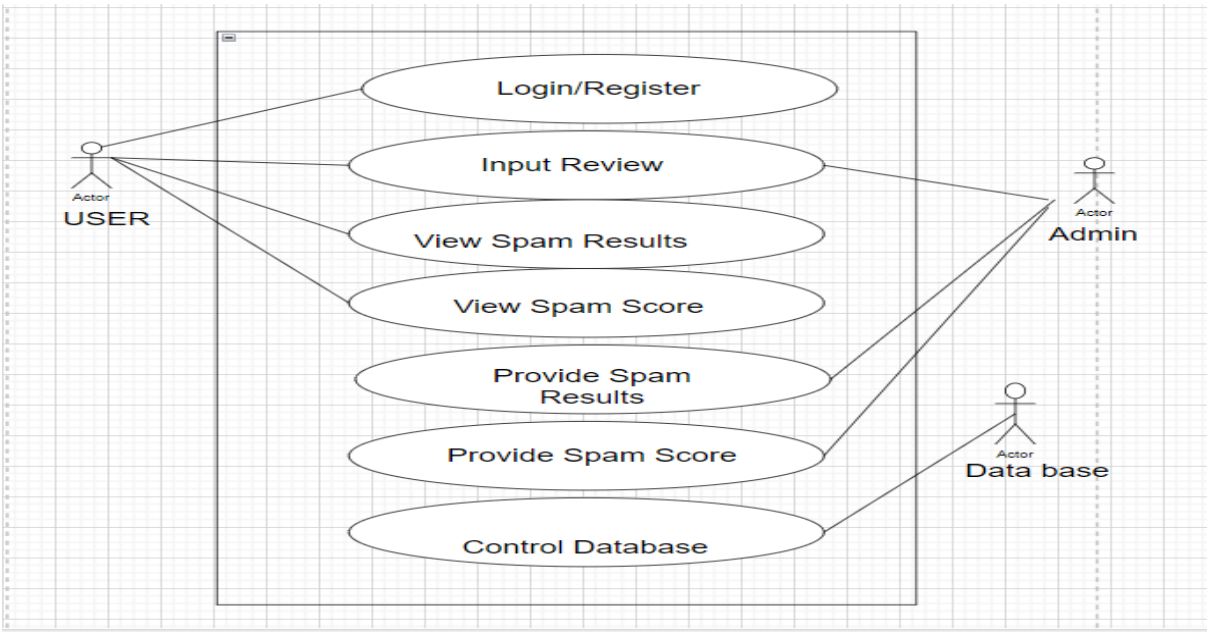


fig 4.1 – Use case Diagram

4.2 Database Description

Our dataset is of size 40,433. Data set consists of 4 attributes namely Category, Rating, Label, Review Text.

Attribute description:

1. **Category:** Various types of products are categorized into the following types .They are

- Clothing, shoes and jewels,
- Toys and games
- Kindle store
- Books
- Pet supplies
- Movies and TV
- Sports and outdoors
- Home and kitchen
- Electronics
- Tools and Home improvements

2. **Rating:** It takes a numeric value ranging between 1 to 5.

3. **Label:** Whole dataset is divided into two classes labelling CG(computer generated fake review), OR(Original Review)

4. **Text:** Review text in the dataset.

4.3 Class Diagram

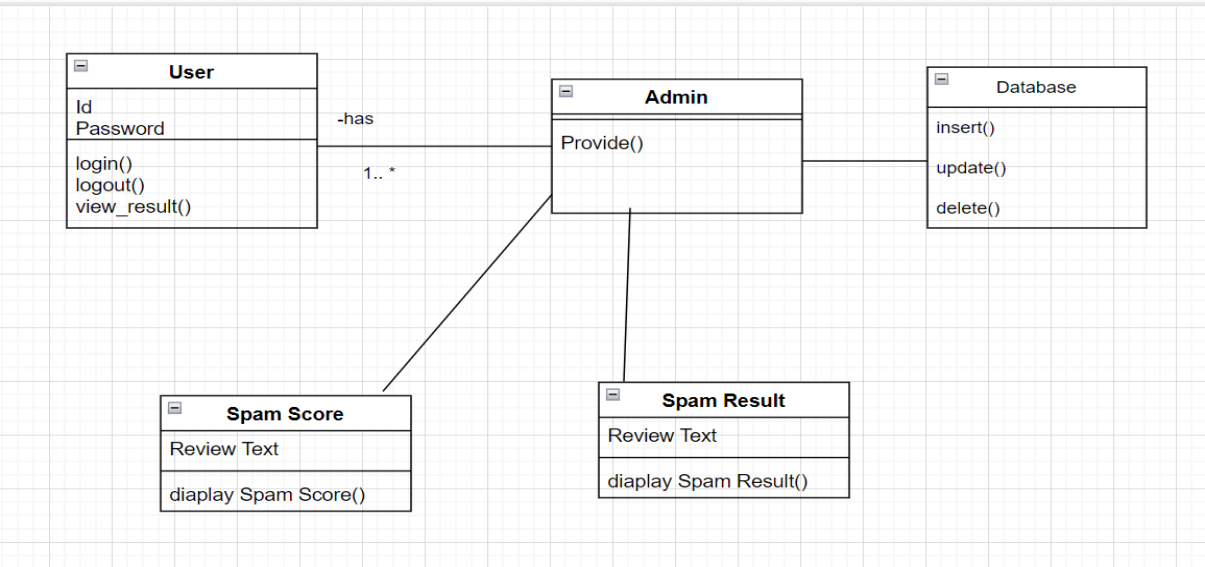


fig 4.3 - Class Diagram

4.4 Activity Diagram

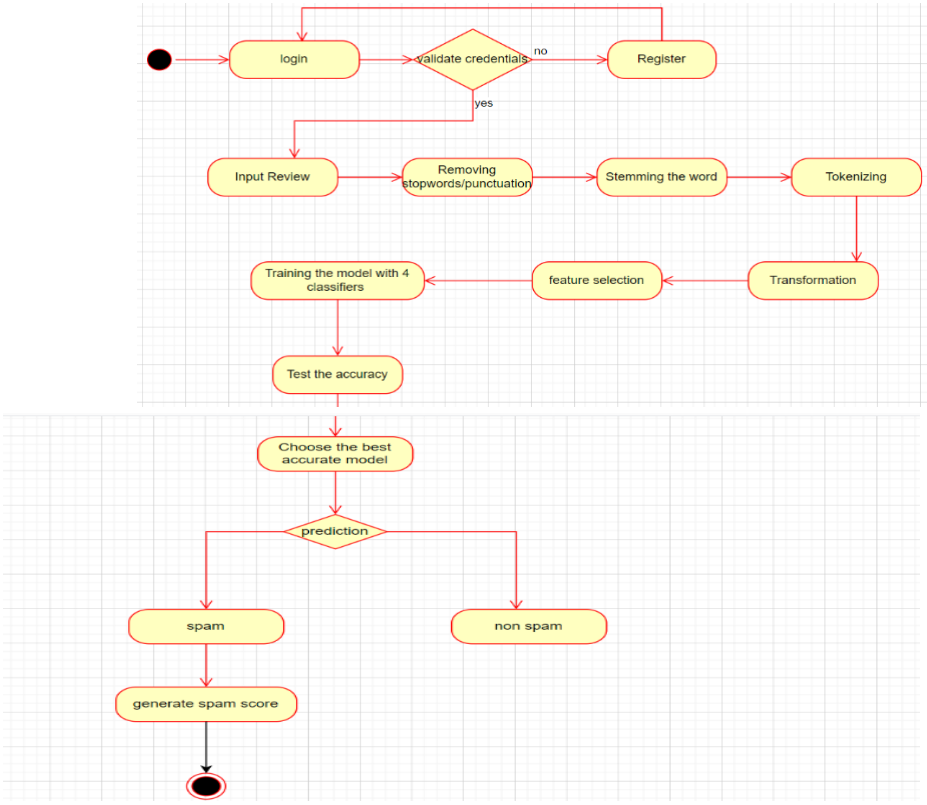


fig 4.4 - Activity Diagram

4.5 Sequence Diagram

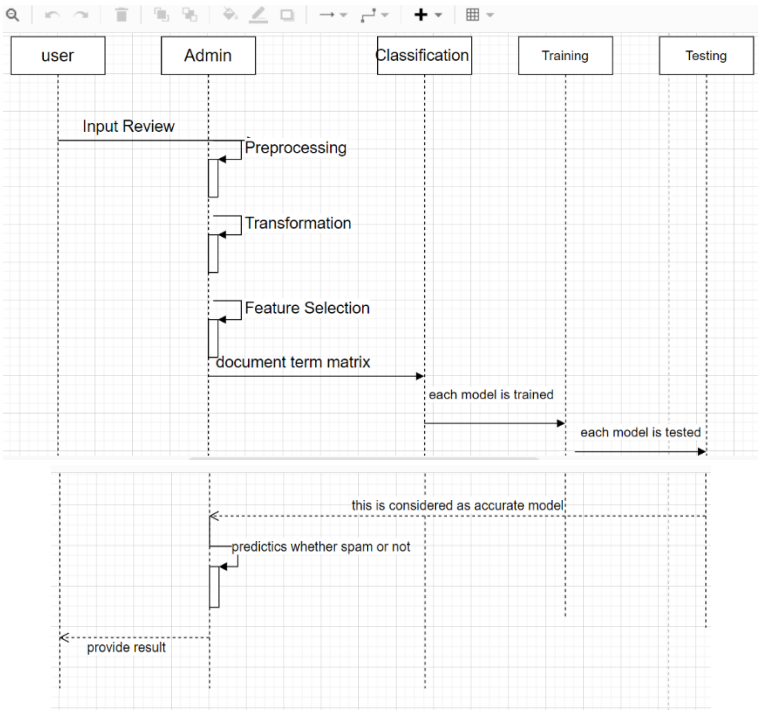


fig 4.5.1 - Sequence Diagram for Linguistic Model

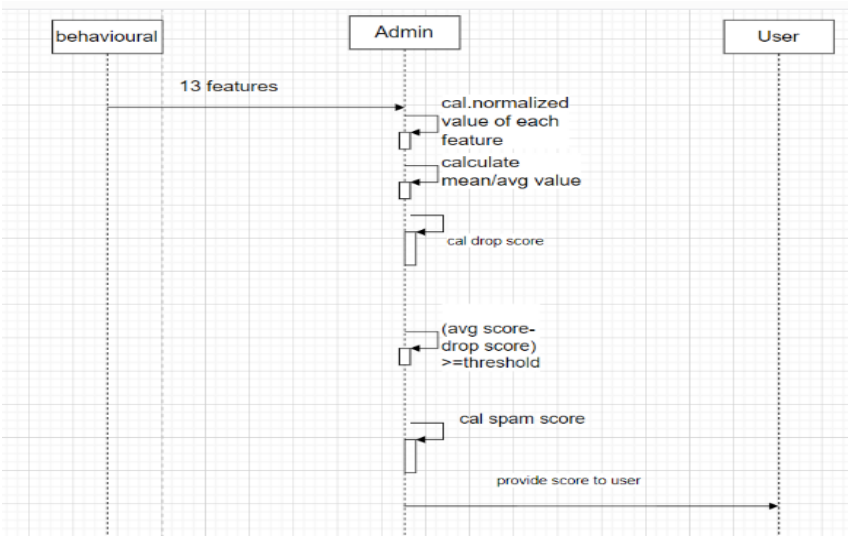


fig 4.5.2 - Sequence Diagram for Behavioral Model

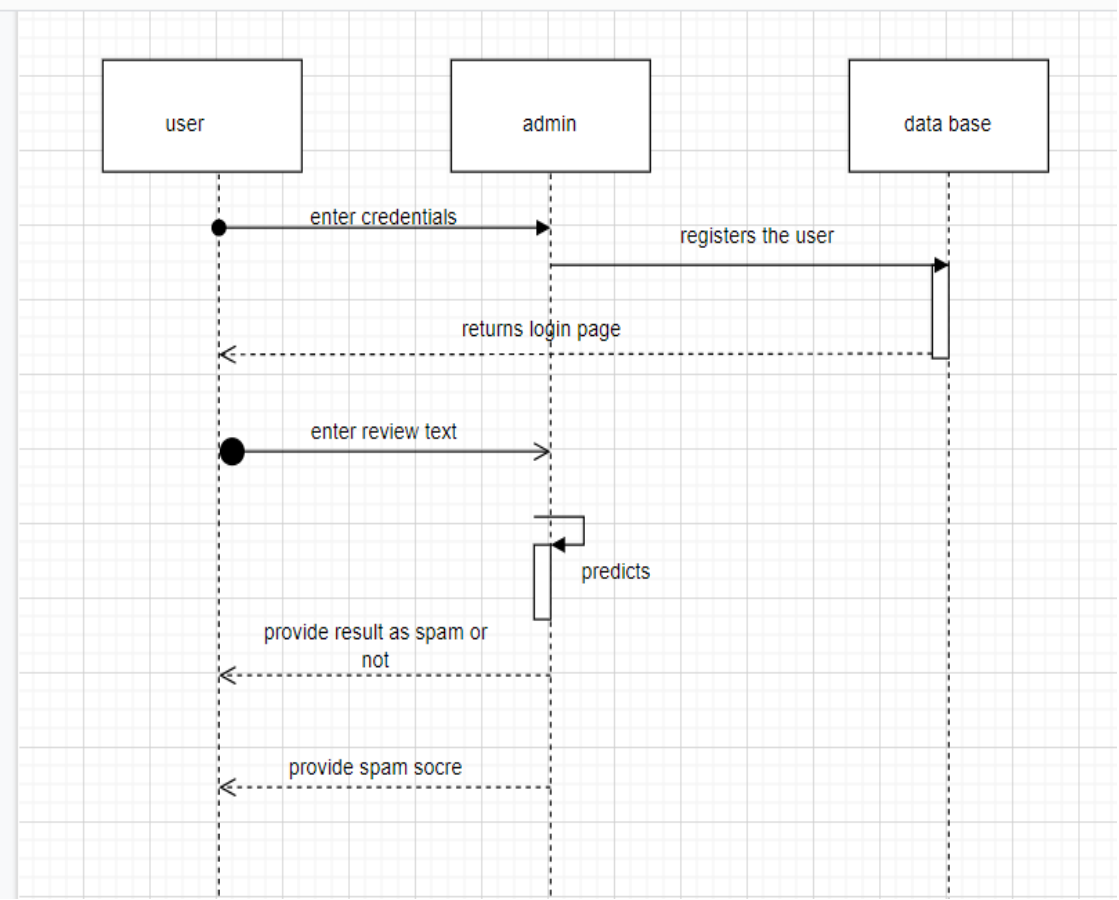


Fig 4.5.3 – sequence Diagram for website

4.6 Collaboration Diagram

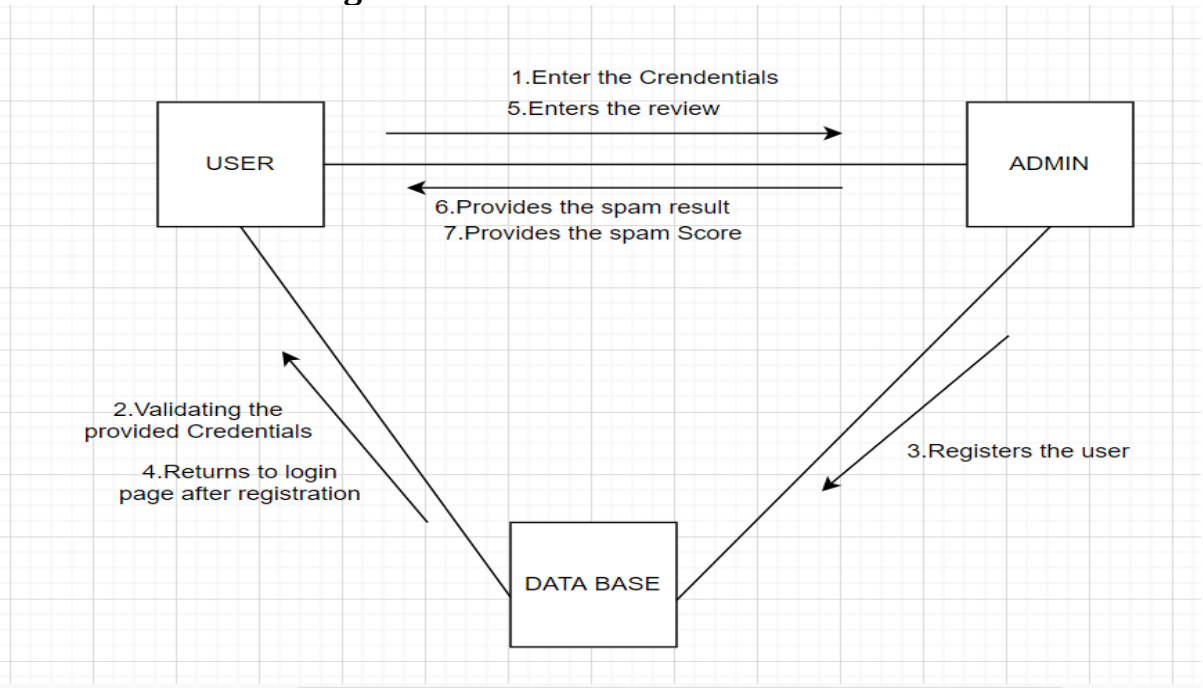


fig 4.6 - Collaboration Diagram

References:

[1].Spam Review Detection Using the Linguistic and Spammer Behavioral Methods

Author: Hamid Turab Mirza ,IEEE Paper.

[2] Y. Ren and D. Ji, ``Learning to detect deceptive opinion spam: A survey," IEEE Access, vol. 7, pp. 4293442945, 2019.

[3]] R. Barbado, O. Araque, and C. A. Iglesias, ``A framework for fake review detection in online consumer electronics retailers," Inf. Process. Manage., vol. 56, no. 4, pp. 12341244, Jul. 2019.

[4] A. C. Pandey and D. S. Rajpoot, ``Spam review detection using spiral cuckoo search clustering method," Evol. Intell., vol. 12, no. 2, pp. 147164, Jun. 2019.

[5]. N. Kumar, D. Venugopal, L. Qiu, and S. Kumar, ‘‘Detecting review manipulation on online platforms with hierarchical supervised learning,’’ J. Manage. Inf. Syst., vol. 35, no. 1, pp. 350–380, Jan. 2018.

[6]. S. Ahmed and A. Danti, ‘‘Effective sentimental analysis and opinion mining of Web reviews using rule-based classifiers,’’ in Computational Intelligence in Data Mining, vol. 1. New Delhi, India: Springer, 2016, pp. 171–179.