FAKE PRODUCT REVIEW MONITORING SYSTEM

A Project Report

submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of degree of

Bachelor of Technology

in

Computer Science and Engineering

by

KEERTHANA K(STM18CS019)

SRADHA A(STM18CS040)

ANUGRAH M(STM18CS010)

SHARUN K(STM18CS038)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ST.THOMAS COLLEGE OF ENGINEERING AND TECHNOLOGY

MATTANNUR

June 2021

DEPT. OF COMPUTER SCIENCE AND ENGINEERING ST.THOMAS COLLEGE OF ENGINEERING AND TECHNOLOGY MATTNNUR

2021 - 22

CERTIFICATE

This is to certify that the report entitled **FAKE PRODUCT REVIEW MONITORING SYSTEM** submitted by **KEERTHANA K** (STM18CS019), **SRADHA A** (STM18CS040), **ANUGRAH M** (STM18CS010) & **SHARUN K** (STM18CS038) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Computer Science and Engineering is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

MRS.DHANYAJA N
(Project Guide)
Assistant Professor
Dept.of CSE
St. Thomas College of Engineering and Technology
Mattannur

MR.RIJIN I K
(Project Coordinator)
Assistant Professor
Dept.of CSE
St. Thomas College of Engineering and Technology
Mattannur

MRS.RENJI S
Assistant Professor and Head
Dept.of CSE
St. Thomas College of Engineering and Technology
Mattannur

DECLARATION

We hereby declare that the project report **FAKE PRODUCT REVIEW MONITORING SYSTEM**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bona fide work done by us under supervision of **MRS.DHANYAJA N**

This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources.

We also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

KEERTHANA K

SRADHA A

ANUGRAH M

SHARUN K

Mattannur

15-06-2021

Acknowledgement

We take this opportunity to express my deepest sense of gratitude and sincere thanks to everyone who helped us to complete this work successfully. We are extremely thankful to our Principal **Dr. SHINU MATHEW JOHN** for giving us his consent for this project.

We express our sincere thanks to **MRS.RENJI S**, Head of Department, Department of Computer Science and Engineering for providing us with all the necessary facilities and support.

We would like to express my sincere gratitude to the Project coordinator **MR.RIJIN I K**, Asst. Professor Department of Computer Science and Engineering for the support and co-operation.

We would like to place on record my sincere gratitude to our project guide **MRS.DHANYAJA N**, Asst. Professor Department of Computer Science and Engineering for the guidance and mentorship throughout this work.

Finally I thank my family, and friends who contributed to the succesful fulfilment of this Project work.

KEERTHANA K

SRADHA A

ANUGRAH M

SHARUN K

Abstract

In the current scenario, the data on the web is growing exponentially. Social media is generating a large amount of data such as reviews, comments, and customer's opinions on a daily basis. This huge amount of user generated data is worthless unless some mining operations are applied to it. As there are a number of fake reviews so opinion mining technique should incorporate Spam detection to produce a genuine opinion.As most of the customers buy their product based on the review of the products. In such cases people go through with the rating or review of the products while observing those, people may not be able to find whether the report is real or fake. Some companies exhibit their own review for the demand of product and company rating purpose. To resolve this problem to find out fake review in the website this "Fake Product Review Monitoring" system is introduced.The proposed technique includes Ontology, Geo location and IP address tracking, Spam words Dictionary using Naive Bayes, Brand only review detection and tracking account used. We are also using Average Content Similarity, Early Time Frame and Users Spam Score for the prediction of Fake review. We plan on creating a new online shopping site to incorporate and verify our monitoring system.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

In this era of internet, everything has become very fast. People are interacting with each other across the world, using too many social networking sites, . They share their thoughts on internet. Internet provides the facility of online shopping, so related to this, on company's website or some review sites like Amazon, Yelp lots of reviews about a product are available. Before purchasing anything, it is a normal human behavior to do a survey on that product. These websites help people to check quality of the product they are going to buy. Based on reviews, customers can compare different brands and can finalize a product of their interest. These online reviews can change the opinion of a customer about the product. If these reviews are true, then this can help the users to select proper product that satisfy their requirements. On the other hand, if the reviews are manipulated or not true then this can mislead user.

Checking the reviews of a product online before purchasing is a common practice now a day. There are a number of sites which deal with these reviews. These sites provide ratings as well as comparison between different products. Some enterprises posts fake reviews to influence customers' behaviors and to increase their sale. Identification of the fake reviews is not easy for customers. In today's world of competition it is advisable for any enterprise to maintain its reputation in a market. So, it is necessary for an enterprise to identify manipulated reviews.

1.2 Problem definition

In recent years, online reviews have been playing an important role in making purchase decisions. This is because, these reviews can provide customers with large amounts of useful information about the goods or service. However, to promote factitiously or lower the quality of the products or services, spammers may forge and produce fake reviews. Due to such behavior of the spammers, customers would be mislead and make wrong decisions. Thus detecting fake (spam) reviews is a significant problem. Opinion spamming refers to the use of excessive and illicit methods, such as creating a large volume of fake reviews, in order to generate biased positive or negative opinions for a target product or service with the intention of promoting or demoting it, respectively. The reviews created for this purpose are known as fake, spam or bogus reviews, and the authors responsible for composing such deceptive content are known as fake or spam reviewers.

Chapter 2

Literature Review

2.1 Email spam detection using Naive Bayes Classifier

In this paper proposed system describing the method that is used to perform spam email classification. The first step is to select the data set file and apply the feature extraction technique for the extracted feature. For which we are using the word-count algorithm. The next step is to form the set of data that is extracted using the characteristic extraction technique. For the formation of the data we can calculate the probability of spam and not spam words in the document. The next step is to test the data with the help of Naive Bayesian Classifier for which it calculates the probability of spam and non-spam mail and make a prediction whose value is greater. If spam words are larger than words that are not spam in an email, the mail is unwanted emails. In the next step system calculating the words that are misclassified by the classifier and calculate the accuracy of the classifier and also calculate the classifier error rate by calculating the fraction of the word that is misclassified and the total number of words in the document.

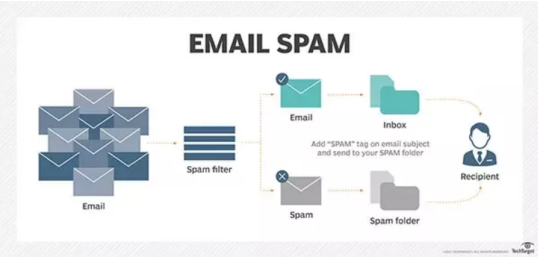

Figure 2.1: Flowchart of the Email spam detction

2.2 Spam Detection On Social Media Platforms

With the increased quality of online social platforms, spammers have come up with various techniques to lure users into accessing malicious links. This is done by generating spam on the comment section of various social media networks.In this paper, we've taken YouTube comments as the dataset and performed spam YouTube comment detection. YouTube has developed filters to combat spam. Since releasing them, they have seen a 40 percent decrease in vital spam metrics. The rules could be updated in a centralized manner by the maintainer of the spam filtering tool. One of the weak aspects of this program is that it doesn't protect a victim from new spam, i.e. it isn't an effective tool for detecting YouTube spam in real time. 90 per cent of users can visit a new spam link before the blacklist is blocked. To get past the more advanced filters that rely on word frequencies, spammers append a large quantity of "usual words" to the end of a message. We prepare our dataset by collecting 400,000 videos of videos corresponding. To get information from YouTube text, we extract certain terms that may be strong indicators for classifying YouTube into one of the classes: spam or non-spam. Feature selection is done to provide a collection of terms that can be used for classification. To do this, Bag-of Word Model's Information Gain is used. Such terms are combined here to form top-30words that we use in our feature set. After the classification process, the data is split into the training set and testing set and the best machine learning algorithm that would provide optimum accuracy in classification is selected and applied. Taking these top 30 terms into account would help us accurately identify the comments for each class. Frequent pattern mining of YouTube's text could also be the crucial factor for identifying real-time YouTube spam. Spam detection on social media platform classify the spam comment of the YouTube. Use the machine learning concept which is a subset of artificial intelligence. Four types of machine learning modules are available, namely supervised learning, semisupervised learning, unsupervised learning, and strengthening. Machine learning is the method of extraction, transforming, loading and predicting the meaningful information from huge data to extract some patterns and also transform it into understandable structure for further use. Prediction and classification and are the two kinds of data analysis techniques that are used by mine models that

identify the most relevant data classes and predict future data trends. The purpose of paper is to briefly introduce machine learning techniques as well as to outline the prediction technique.To provide a prediction so that it can perform descriptive analytics on prediction of spam comments in YouTube. Spam detection is using supervised learning. This method depends upon a huge number of labeled datasets. Propose the classification algorithm called logistic regression for classifying spam.

2.3 An ensemble machine learning approach through effective feature extraction to classify fake news

The concept of fake news has been in existence even before the emergence of Internet and other computational technologies.Dissemination of fake news and misleading information has always been used as a weapon to fulfil immoral objectives since ages.The advancement of Internet and web technologies has made it very easy for anyone to post and access anything in online platforms like blogs, comments to news articles, social media,etc.The involvement of social media replacing the traditional media has an even more catalytic effect, where both fake and authentic news are spread extremely rapid The spread of such fake news has extremely negative impact on target indviduals and also the society at large.Consequently, it also creates an impression among readers such that the general perception and responses towards authentic news also gets diluted hampering the balance of news ecosystem. the need of empirical investigation towards fake news detection, identification of source and stopping of its spreading still exists.It is very important to stop circulation of misleading falsified information.It is very important to stop circulation of misleading falsified information.Primarily can be achieved through human intervention by verifying authenticity of information using the International Fast Checking Network (ICFN) and other manual fast checking websites like Washington post, Snopes, Fast Checker,Fast Check and Truth Or Fiction These websites are quite authentic and efficient, but they have scalability issues in handling large volumes of data.To overcome this aspect, the concept of automatic fact-checking has evolved consisting of three elements,namely, identification,verification and correction.These three elements work hand in hand to 1 identify false claims, verify authenticity of the claims and delivers rectified genuine information across social media platforms. Automated fast checkers have agility in responding to information and are scalable in handling the high volume of news created across social media platforms.

The present study involves experimentation on two popular fake news datasets, ISOT and Liar datasets, using Google Colab which is an open source cloud service provided by Google Inc.The data is cleaned and feature extraction is performed resulting in selection of the most significant features contributing towards detection

of fake news. As part of the proposed algorithm, 70% of data set is used for training and remaining 30% is used to test the classification model using k-fold cross validation. The extracted features are further classified using an ensemble machine learning model comprising of Decision Tree (DT) classifier, Random Forest (RF) algorithm and Extra Tree (ET) classifier.

2.4 Fake Product Review Monitoring System

In the current scenario, the data on the web is growing exponentially. Social media is generating a large amount of data such as reviews, comments, and customer's opinions on a daily basis. This huge amount of user generated data is worthless unless some mining operations are applied to it. As there are a number of fake reviews so opinion mining technique should incorporate Spam detection to produce a genuine opinion. Nowadays, there are a number of people using social media opinions to create their call on shopping for product or service. Opinion Spam detection is an exhausting and hard problem as there are many faux or fake reviews that have been created by organizations or by the people for various purposes. They write fake reviews to mislead readers or automated detection system by promoting or demoting target products to promote them or to degrade their reputations.

A critical level at the available literature on spam detection dictates the following issues which need to be handled or taken care of

1. IP Address of the reviewer.
2. Geo location of the reviewer.
3. Account used by the reviewer to post a review.
4. The reviews that doesn't talk about the product on which it is posted.

The proposed technique includes Ontology, Geo location and IP address tracking, Spam words Dictionary using Naive Bayes, Brand only review detection and tracking account used.

Chapter 3

PROPOSED SYSTEM AND DESIGN

3.1 Proposed system

Several techniques have been introduced to detect spam. Additionally some techniques are going to be used here, to make opinion mining more accurate.
Steps of the Proposed System:-

- Spam Detection using IP address.
- Spam detection using Geo location.
- Spam Detection using Account Used.
- Spam Detection using Spam Data Dictionary using Naive Bayes.
- Spam Detection using Ontology.
- We are also using Average Content Similarity, Early Time Frame and Users Spam Score for the prediction of Fake review.
- Classify Data into Spam and Non Spam.
- Opinion Mining of Non Spam Data.

Tracking IP address of the user to detect if the reviews are from a Spammer. If multiple reviews are from the same IP address then the Reviews are considered Spam.

Using Geo Location of the user to identify whether user is same that is posting the Review.

Using Account Used to check whether the reviews are done using the same account.

Brand only Review detection:- Whether the reviews are on only Brand not the product. It's not helpful to consider only the Brand value to judge a product.

Using Spam Dictionary:- The spam words are identified in the review. If there are Spam Words then the review is a Spam. For instance, a user has posted a Review: Buying this laptop made me feel so good, it's like winning a lottery. Here Lottery is a Spam Word. So, this word will be considered a Spam, while this is not a Spam. Therefore Spam Word Dictionary will be used with Naive Bayes classifier. According to this approach, probability of 'Lottery' to be Spam is less so it will not be considered a Spam.

Using Ontology:- For instance, if the review posted on a product is not about that product but talking about something else then ontology is used to identify and classify such reviews as spam. If Class: Toshiba Context: Laptop Review: Dell is not so good. Here User is Posting Reviews about Laptop that comes under the class Toshiba. But his Review contains Dell Keyword. In order to identify this Review as Spam I am going to use Ontology.

Average Content Similarity:- Spammers, often write their reviews with same template and they prefer not to waste their time to write an original review. In result, they have similar reviews. Users have close calculated values take same values

Early Time Frame:- Spammers try to write their reviews asap, in order to keep their review in the top reviews which other users visit them sooner

Users Spam Score:- Spam score defines the history of users post, percentage of spam reviews posted by the user

Also we plan on creating a new online shopping site to incorporate and verify our monitoring system.It consist of 2 modules:-

Admin:- It consist of login, view user, view reviews, Manage Product and booking.

Login:- It consist of user name and password to login the system

View User: Admin view the username, registered email id, phone number and users product details.

View Reviews:- Admin view the users reviews and uploaded date.

Manage Product:- It consist of add product, view product, delete product and edit

product.

Booking:- It consist of booked product details and users details.

User:- It consist of Registration users registered with name, phone number, email id and create its own password.

Login:- It consist of user name and password to login the system.

View Product and Booking:- It consist of 3 steps

1) First view the product details and reviews

2) Click add to cart the move to cart page and add it's count.

3) Final booking of product

Add Reviews :- user can add reviews.

View Booking:- Booked product details.

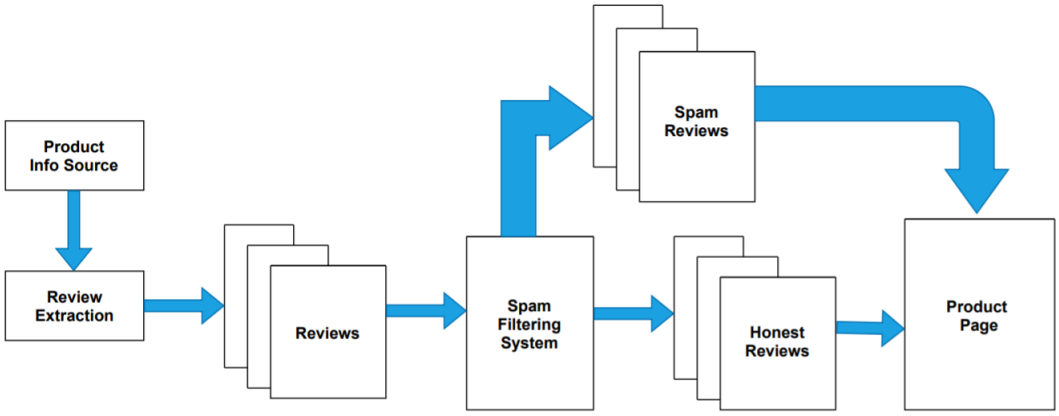

Figure 3.1: Flowchart of the proposed system

3.2 Feasibility study

Facts considered in the feasibility analysis were:-

• Operational Feasibility

• Technical Feasibility

• Economic Feasibility

• Behavioral Feasibility

The requirements of the system are specified with a set of constraints such as system objectives and the description of the out puts. Three key factors are to be considered during the feasibility study.

3.2.1 Operational Feasibility

The proposed project is operationally feasible. Because the proposed application is easily operated by the user. The proposed web site is operated by the Admin, user. The Admin, user who have basic internet knowledge, can use the application for their intended use. For operating the web content, need only the basic internet knowledge and an internet connection .The application can be can be easily operated by the Admin, user . Hence our proposed system is operationally feasible.

3.2.2 Technical Feasibility

The proposed system meets all the requirement of the Technical feasibility. Because the implementation of the project need no technical difficulties. The web part of the project can be easily implemented in a user friendly IDE that is pycharm. The development of the project does not meet any type of the technical difficult. Hence we can be say that the proposed system is technically feasible.

3.2.3 Economic Feasibility

The proposed project is economically feasible. Because once the system is put into its use in the current market the system provides economical advantage to the firm. Also the firm can afford the cost to implement the project. This system does not need any initial investments and it can improve the quality of service.

3.3 Design

3.3.1 Usecase Daigram

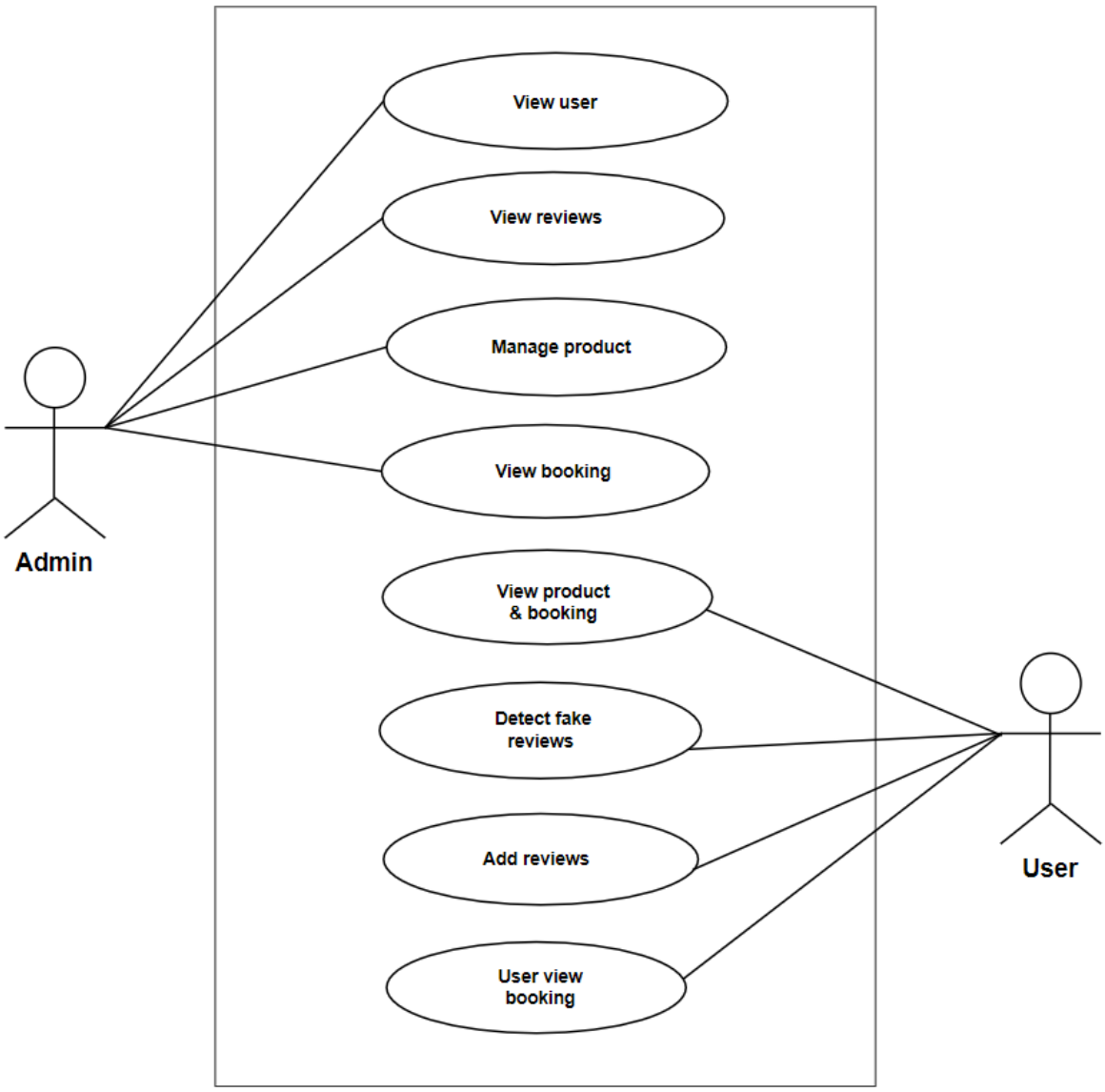

Figure 3.2: Usecase daigram

A use case diagram is the primary form of system/software requirements for a new software program underdeveloped. Use cases specify the expected behavior (what), and not the exact method of making it happen.

In this usecase we have two actors admin and user. Admin can access the features like view user, view reviews, view booking and manage proudct.

- View user:- Admin can view the user profile.
- View review:- Admin can view the user reviews.
- view booking:- Admin can check the status of product.
- Manage product:- It consist of 3 functions like add product, update product and delete product.
- add product:- admin can add product.
- delete product:- admin can delete product if it is not needed.
- update product:- Any time admin can update the product deatils.

User can access features like view product booking, Detect fake reviews, Add reviews and view booking. User access the site and register the details such as name, phone number, email id and set a strong password. After registration user can login the site using email id or phone number and password. In home page it contains view product and booking, add reviews, and view booking.

- view product and booking:- User can view all the product and it's details. If interested to buy the product, book the product and add quantity.
- Add reviews:- After buying the product user can add opinion about their product.
- view booking:- User can check the status of the booked product.
- Detect fake reviews:- Check the review is fake or not using some features like Geo location, account used, IP address, ontology, spam data dictionary using naive bayes, average content similarity, early time frame and users spam score.

3.3.2 Data Flow Daigram

The flow of data of a system or a process is represented by DFD. It also gives insight into the inputs and outputs of each entity and the process itself. DFD does not have control flow and no loops or decision rules are present.

it uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. It contains 3 levels in the data flow diagram, which are: 0-level DFD, 1-level DFD, and 2-level DFD.

0-level DFD :

DFD Level 0 is also called a Context Diagram. It's a basic overview of the whole system. It's designed to be an abstraction view, showing the system as a single process with its relationship to external entities. It represents the entire fake product system as a single bubble with input and output data indicated by incoming/outgoing arrows. It consist of 2 module admin and user. Admin module contains the fuctions like view user, view review, manage product and view booking.User module contain view product booking, add reviews and view booking.

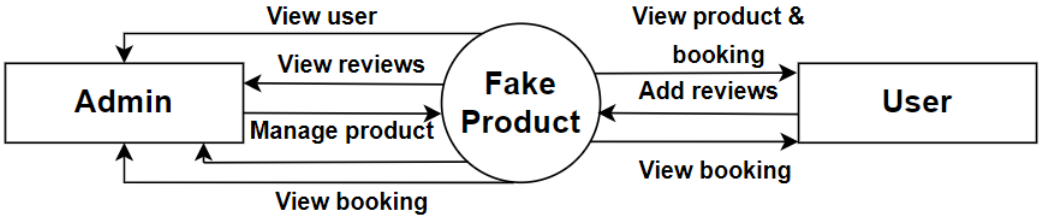

Figure 3.3: 0-level DFD

1-level DFD :

In 1-level DFD, the context diagram is decomposed into multiple processes and highlight the main functions of the system and breakdown the high-level process of 0-level DFD into subprocesses. Admin module login to the system and perform some functions like view user,view reviews,manage product and view booking. login details stored in the login table, Admin can view user details from user_reg table, view review details from user_reg and review table,manage product function handled by product table,view booking details from product, book product and master table. User register to the system, then login to the system using username and password. User can view product detais from master, product and book product table, after viewing the product user can check the reviews are fake or not. user can add reviews it will store in the reviews table , user can view booking details from master, product and book product table.

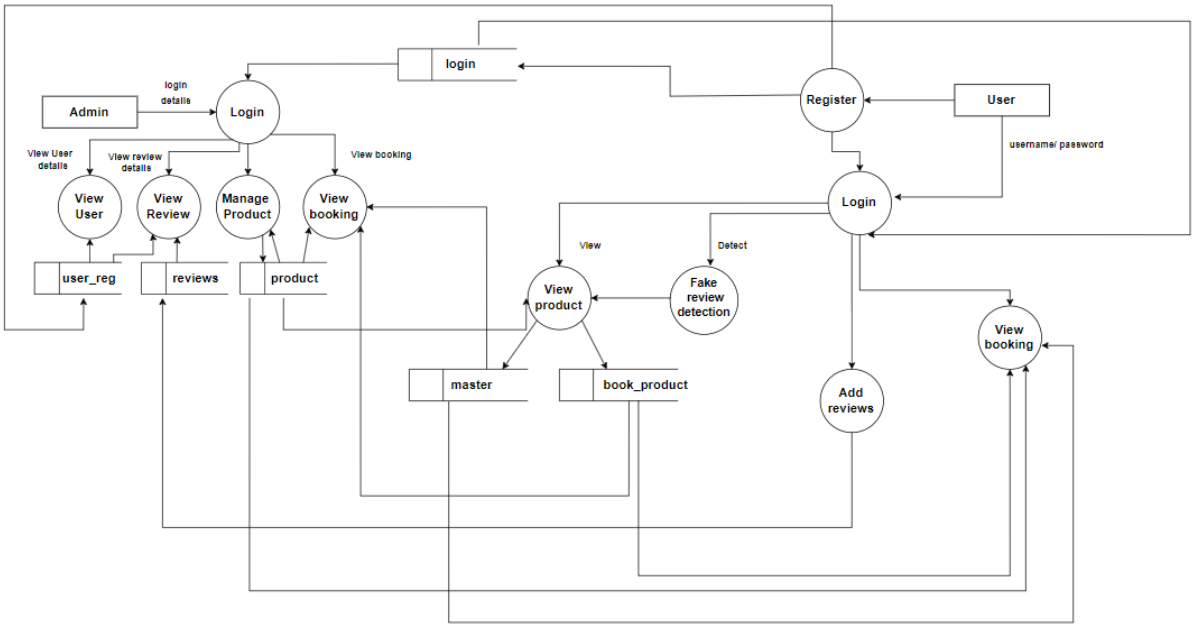

Figure 3.4: 1-level DFD

2-level DFD :

DFD Level 2 then goes one step deeper into parts of Level 1. level 2 contains admin home page and user home page .

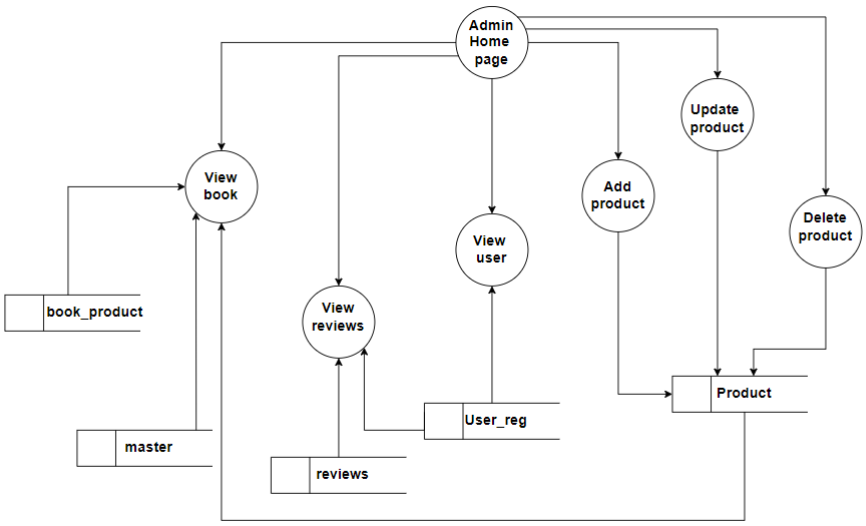

Figure 3.5: 2-level DFD(admin part)

View user details from user_reg table, view review details from user_reg and review table,view booking details from product,book product and master table. manage product consist of add product,delete product and update product, this details handled by product table.

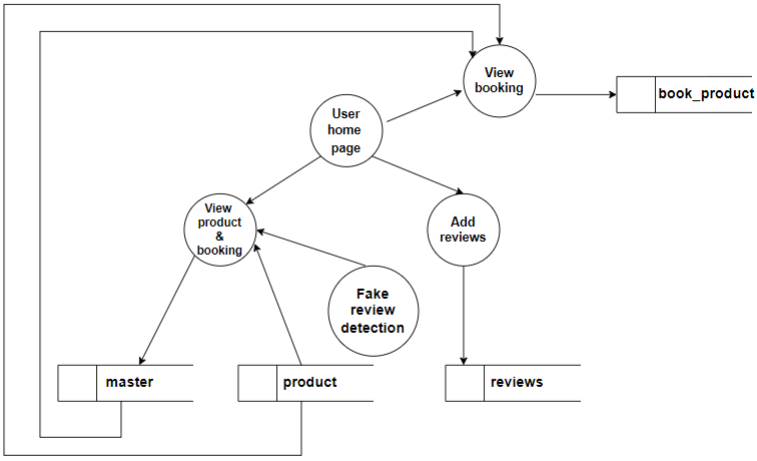

Figure 3.6: 2-level DFD(user part)

User can view product detais from master,product and book product table,after viewing the product user can check the reviews are fake or not. user can add reviews it will store in the reviews table ,user can view booking details from master, product and book product table.

- login table:- contains login id,username, password and utype.
- user_reg table:- Consist of user id , name, phone number and email id.
- reviews table:- Contain review, review id, date,user id and product id.
- master table :- master id,user id,status,date and amount.
- book product table:- product id, book_pid, count and master.
- product table:- price, pname, image, description and product id.

3.3.3 ER Diagram

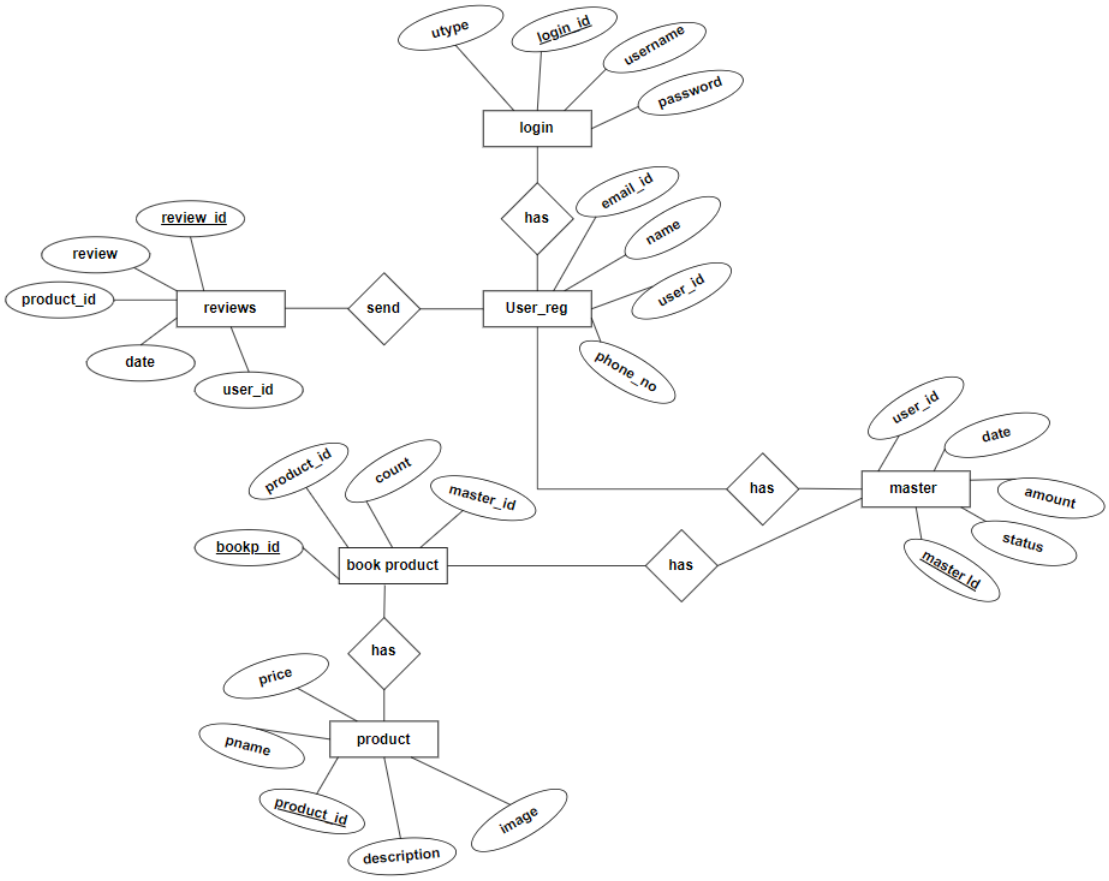

Figure 3.7: ER daigram

An Entity Relationship (ER) Diagram is a type of flowchart that illustrates how "entities" such as people, objects or concepts relate to each other within a system. ER Diagrams are most often used to design or debug relational databases in the fields of software engineering, business information systems, education and research.

ER diagram illustrates key information about Fake product review monitoring system, including entities such as login,user registration,reviews,book_product,product and master.

Entities and their Attributes are:

- login:- Attributes of login entity are username, password and login_id.login_id is primary key for login entity.
- User registration:- Attributes of user registration entity are name, email_id,

phone_number and user_id.

- Reviews:- Attributes of reviews entity are review_id, review , product_id, date and user_id. review_id is primary key for reviews entity.
- book_product:- Attributes of book_product entity are bookp_id, product_id, master_id and count .bookp_id is primary key for book_product entity.
- product:- Attributes of product entity are price, product_id, pname, image and description. product_id is primary key for product entity.
- master:- Attributes of master entity are user_id,date, amount, status and master_id.master id is primary key for master entity.

Relationship are:-

- user_registrtion has login :- only registred user can login to the system.
- user_registrtion send the reviews:- Only registred user can send reviews to the system.
- book_product work master:- master is a billing section,so book_product work under master.
- product has book_product:- Each product dependents on book_product.
- master has user_reg:- master dependents on user_reg.

3.4 Gantt Chart

A Gantt chart is a commonly used graphical depiction of a project schedule. It's a type of bar chart showing the start and finish dates of a project's elements such as resources, planning and dependencies. The project is about fake product review monitoring system,the project duration is 10 weeks. The task start and end dates, duration, and milestones appear as horizontal bars. It consist of four modules admin, user,training and detection. Admin and user modules take two weeks respectively, and remainig module take three weeks . Each task is dependent on the previous task.

Figure 3.8: Gantt chart

Chapter 4

Conclusion

The fake product review monitoring system, it will helpfull for users whether the product review is fake or not. Advantages of proposed System is

- Organizations benefits by increasing of sales of their product by genuine rating.
- User gets genuine reviews about the product.
- User can spend money on valuable products.

Also we creating a new online shopping site to incorporate and verify our monitoring system.

References

[1] Fake Product Review Monitoring System International Journal of Trend in Scientific Research and Development (IJTSRD).

[2] Opinion Mining using Ontological Spam Detection Neelam Duhan, Divya, Mamta Mitta .University of Science & Technology.

[3] Spam detection on social media platform, IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020.

[4] Alurkar A A, Ranade S B, Joshi S V, Ranade S S, Sonewar P A, Mahalle P N, and Deshpande A V, A proposed data science approach for email spam classification using machine learning techniques 2017, 2017 Inter, of Things Bus. Mod., User., and Net., 2018, 1–5.

[5] Fake Review Monitoring System NSC. Mohan RaoAsso.prof, V. Sujitha, V.RajaKumari, P. Hema, M. Tulasi, M.SivaKishore.

[6] Alurkar A A, Ranade S B, Joshi S V, Ranade S S, Sonewar P A, Mahalle P N, and Deshpande A V, A proposed data science approach for email spam classification using machine learning techniques 2017, 2017 Inter, of Things Bus. Mod., User., and Net., 2018, 1–5.

[7] Agarwal K and Tarun Kumar,Approach of Naïve Bayes and Particle Swarm Optimization 2018, 2018 Sec. Int. Conf. on Intel. Comp. and Cont. Sys., pp. 685–90.

[8] C. Buntain, J. Golbeck, Automatically identifying fake news in popular Twitter threads, in: 2017 IEEE International Conference on Smart Cloud.