International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

Fake Product Identification Using Machine Learning

Deepali Godse[1], Chaitali Nigade[2], Nilofar Mulla[3], Shital Jadhav[4]
[1,2,3,4]Information Technology Department, Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, India.
Emails: *deepali.godse@bharatividyapeeth.edu[1], chaitalinigade14@gmail.com[2], nilofar.mulla2006@gmail.com3, shitaljadhav50949@gmail.com[4]*

Abstract
Recognizing counterfeit goods can be difficult in some situations. If a person does not thoroughly inspect the product's details, it becomes simpler to create and sell counterfeit goods. For less tech-savvy clients who can scan the product with the use of a smartphone application to check the authenticity of the product received, this paper offers a superior alternative employing machine learning. The detection of logos (which includes both visual and textual representations) is the main focus. The model also includes the sentiment analysis of the product's reviews. This technique is useful for predicting the validity of a product. The paper describes the Fake Product Identification Model developed using Convolution Neural Network (CNN) and Optical Character Recognition (OCR). This model determines whether a product is real or fake, and the user can make a wise decision before buying the product.

Keywords: *Convolution Neural Network; Counterfeit Goods; Optical Character Recognition; Sentiment Analysis*

1. Introduction
According to the International Anti- Counterfeiting Coalition, the global counterfeiting problem is estimated to be worth more than $1.6 trillion [1]. It is accelerating quickly, in part because there is more e-commerce. Counterfeiting is prevalent in all industries, including toothpaste, aspirin, and high-end luxury brands. Unknowingly, consumers are buying these goods, which have the potential to negatively impact their long-term health and wellbeing. The value and reputation of a company's brand are impacted by counterfeiting. Nowadays, there are numerous ways to shop, such as visiting a store or mall to purchase a specific item you require. In this type of shopping, the vendor provides you with the product's feedback, but you are unsure if it is genuine or fraudulent. Because it depends on the seller's honesty and how true to their claims they are, you must carefully inspect the merchandise as you have no other choice but to do so. If you do not pay attention when purchasing that item, it could end up being a waste for you. Today's shopping sources have been altered. You can purchase goods from several brands' internet stores. After reading the reviews and looking at the product logo, you purchase the item. As a result, you are reliant on product reviews and the logo. These reviews could be bogus or authentic. Sometimes, even if the logo is fake, the user might find it real and be tricked into buying the product. Fake product monitoring systems improve the effectiveness of the testing phase for genuine and counterfeit products. Many defect prediction models integrate well-known methodologies and algorithms, including Machine Learning and statistical methods. To determine whether models appear to be the fraudulent product, they need historical data containing inaccurate information as training data. These tools can estimate fake product modules based on training data knowledge. A recent study on defect prediction models reveals that manual code inspections can find between 35% and 60% of

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

problems, and an Artificial Intelligence (AI) based fake product monitoring system can find 70% of all faults. This paper presents a low-cost, user- friendly machine learning-based strategy that enables end users to identify and certify products without using professional equipment. This strategy uses image and language recognition to improve the detection of fake products.

2. Proposed Methodology and Algorithm

2.1. Proposed Methodology

The paper proposes an android application for logo detection that uses machine learning to identify differences between genuine and fake products based on their varied forms, text, font, and color attributes. The proposed system functions in two phases. The first stage is to identify logos using text and image recognition. The process of developing a machine learning model comes next before determining if a logo is real or fake.

2.1.1. Optical Character Recognition

A spelling detector is included as an additional feature within the scanner because a standard logo detector or scanner only collects photos. OCR (Optical Character Recognition) [2] is used for this. In a physical document, such as a scanned document or an image file, it is a software technique that recognizes (written or printed) text electronically and converts it into machine- readable text for data processing.

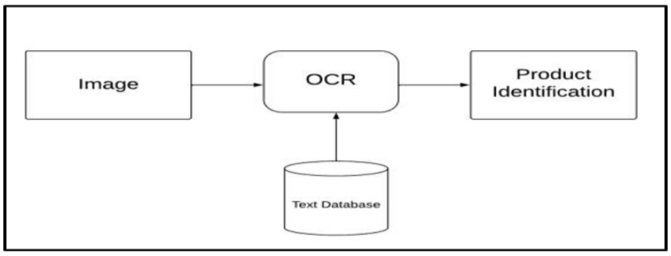

Figure 1 **Optical Character Recognition**

The proposed system takes an image as input. Fig. 1 depicts the general operation of OCR for the proposed system. To extract text from images, the Python Tesseract-OCR module, often known as Pytesseract, is used. The text is retrieved from both the template and image being tested. After that, a basic Python script is executed to determine whether the two texts are identical. If they are, the logo is considered unique; otherwise, it is considered fake. The Tesseract module applies the edge detection function to both images to determine whether they are original.

2.1.2. Sentiment Analysis

Online business is one of the business fields that is growing the fastest around the world. People buy a lot of things from online shopping sites these days. Online product sales are frequently influenced by customer reviews. As a result, spotting fake reviews is becoming increasingly important.

Sentiment analysis [3], [4] is critical in detecting false reviews. This research presents a sentiment analysis technique that can efficiently differentiate good and negative sentimental reviews. It depicts an evaluation of the sentiment distribution for false and real reviews. The flow of sentiment analysis is depicted in Fig. 2.

2.2. Algorithms

2.2.1. Convolutional Neural Network (CNN)

The Convolutional Neural Network (ConvNet/CNN) is a Deep Learning technique that takes an input image and assigns importance (learnable weights and biases) to various aspects and objects in the image, allowing them to be distinguished. [5]. A ConvNet may successfully capture the Spatial and Temporal correlations in an image by using the right filters. Because of the fewer parameters involved and the reuse of weights, the architecture better fits the image dataset. In other words, the network might be trained to better understand the image's complexity. It is used to extract the logo from the input image in the proposed system. An Accuracy score (or simply Accuracy) is a Machine Learning Classification statistic that represents the percentage of right predictions made by a model, which consists of correctly identified cases as a real product (TP), incorrectly identified cases as real product (FP), the correctly identified cases as a fake product (TN) and the incorrectly identified cases as a fake product (FN). The flow of CNN is illustrated in Fig. 3.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

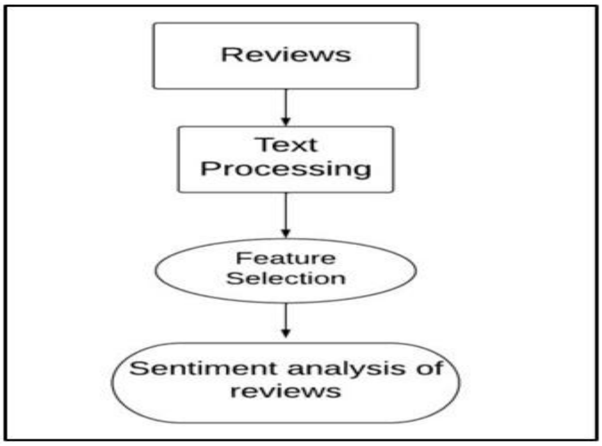

Figure 2 Sentimental Analysis

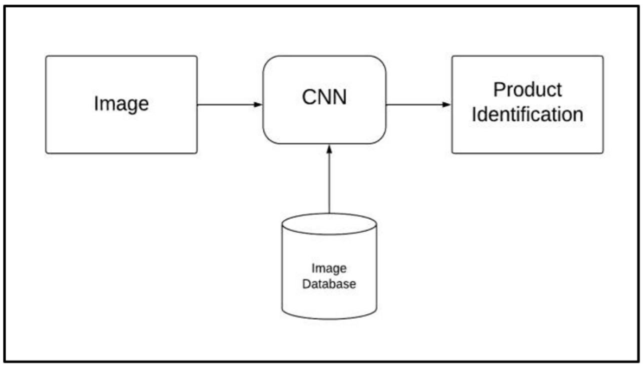

Figure 3 Convolution Neural Network

$$Accuracy = TP + \frac{TN}{FP} + FN + TN \quad (1)$$

The layer information in Table 1 is listed on the left side from first to last. The top layer is the initial layer, and the bottom layer is the last layer. Each layer's output shape is in the last column. As an example, the first Conv2D layer's output of (None, 254, 254, 16) shows the feature map's dimensions following the first convolution operation. Through the use of 16 filters, the feature map is 254 x 254 in size and has a depth of 16. The number of training examples (batch size) is indicated by the first element in the tuple, which is none. The number of parameters used in each layer is listed in the last column. Flattened and pooling layers do not have parameters.

The convolutional layer is the top layer. Filters, also referred to as "kernel," are taken by this layer. These filters expose the layer to low-level features such as edges and curves. If more convolutional layers are added, the model can better extract deep features from images and thus determine all of their characteristics. The model has added more convolutional layers over time. The filter convolutionally transforms a portion of the image. Using the filter and pixel values, the convolution operation is the multiplication and addition of individual image elements. Numerous filters from the convolutional layer have been used to derive various features. In Table 1, the output shape for each layer is shown. The max pooling layer is used after one convolution layer. This layer reduces the input's spatial dimensions (height and breadth). In a CNN, a flattened layer is located between the last pooling layer and the first dense layer. The flattened layer takes a 2D feature map and turns it into a 1D feature vector. This vector is then given to the dense layer. Fully (densely) connected layers make up CNN's final layers [8-13].

Table 1 Model Summary

Layer (type)	Pooling	Parameters	Output Shape
Convolution al	16	448	254 x 254
Max Pooling	16	0	127 x 127
Convolution al	32	4640	125 x 125
Max Pooling	32	0	62 x 62
Convolution al	16	4624	60 x 60
Max Pooling	16	0	30 x 30
Flatten	16	0	30 x 30
Dense	4	3686656	8 x 8
Dense	1	257	1 x 1

2.2.2. Sentiment Intensity Analyzer

Sentiment Intensity Analyzer [6], [7] is a tool for sentiment analysis, which is the act of evaluating a text's emotional tone or attitude. Sentiment Intensity Analyzer is a pre-trained model provided in the Natural Language Toolkit (NLTK) package that determines the sentiment of a piece of text using a

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

lexicon-based approach. The working of Sentiment Intensity Analyzer is as shown in Fig.4. A sentence is given a polarity score by the Sentiment Intensity Analyzer that ranges from -1 to +1, with -1 being the most negative, +1 being the most positive, and 0 being neutral. A graph can be plotted showing the sentiment scores for the reviews as shown in Fig. 4.

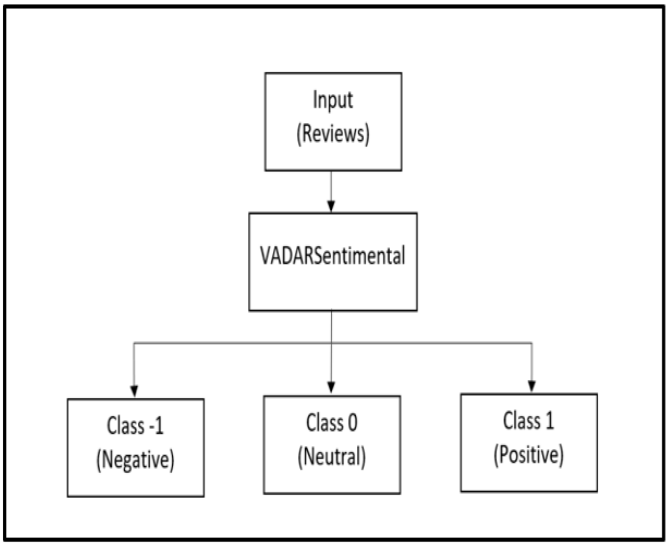

Figure 4 Sentiment Intensity Analyzer

Sentiment Intensity Analyzer provides a measure of the text's subjectivity in addition to the polarity score. This value ranges from 0 to 1, with 0 being extremely objective and 1 being highly subjective. Sentiment Intensity Analyzer uses a lexicon of phrases and their corresponding polarity ratings to analyze a piece of text and determine its sentiment. The vocabulary includes approximately 7,500 items, and the polarity scores are determined by a combination of human evaluations and machine learning techniques. Overall, Sentiment Intensity Analyzer is a valuable tool for analyzing the sentiment of a piece of text quickly and simply, but it is crucial to remember that it has some limits. It may not be accurate, for example, in texts containing sarcasm, irony, or other forms of figurative language.

3. System Architecture

3.5 billion of the 4.78 billion mobile phone users worldwide today use smartphones. Users can now get a smartphone with an internet connection and a built-in digital camera for a reasonable price. Based on this, the suggested approach will enable end users to access the product's written information, logos, and possibly certification markings or logos.

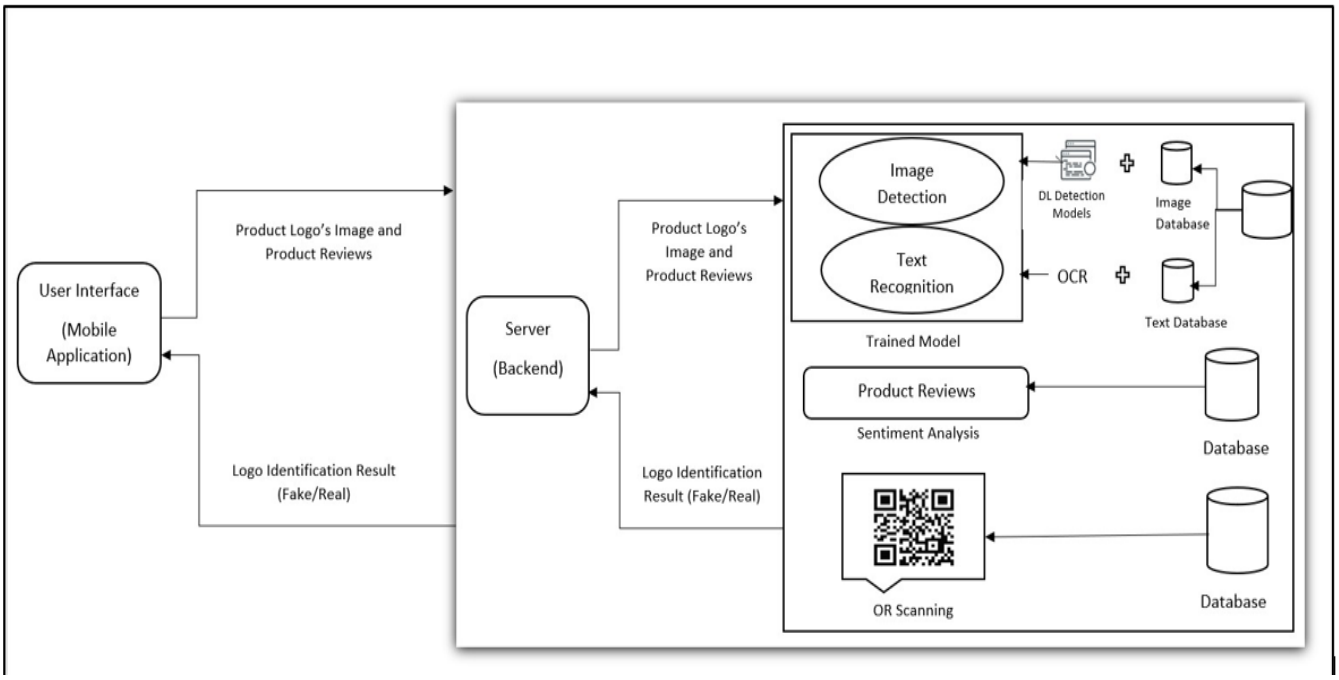

Figure 5 Proposed Architecture

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351

e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

Upon their inclusion in a request, the server will process and validate these images. The end-user will then receive the detection result, allowing them to choose their next course of action. The user will have two alternatives to determine if the product is false or real: text recognition and image recognition. The user's input will be provided to the machine learning model, which employs text detection with Optical Character Identification and image recognition with CNN to determine whether the product is legitimate or false. The user will see the logo or text identification outcome on the mobile application. The overall architecture of this solution is shown in Fig. 5. Results are shown in Fig.6 and Table 2.

4. Result

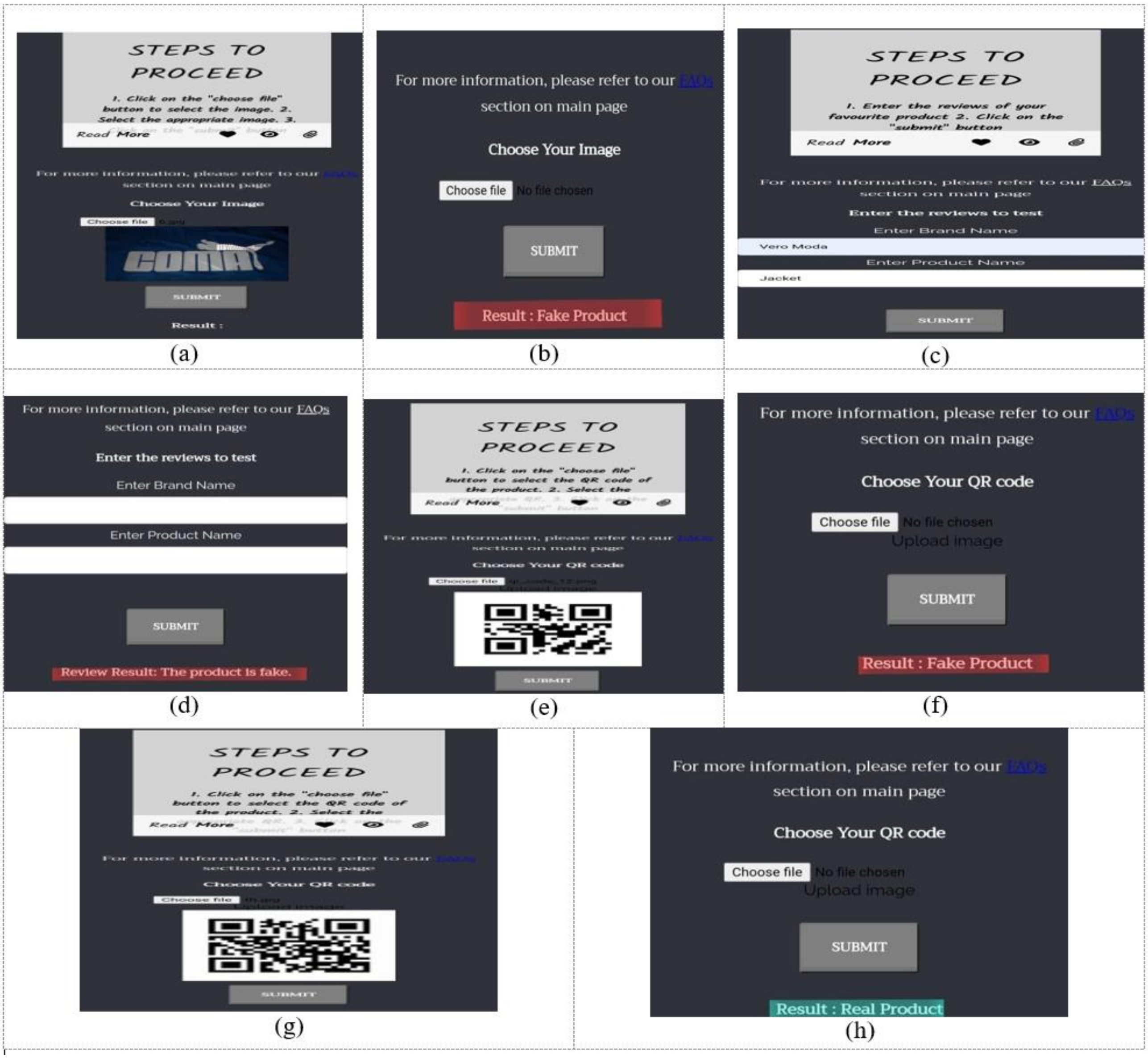

Figure 6 Result of All Modules: (A) OCR Module, (B) Result of OCR Module. (C) Sentiment Analysis Module. (D) Result of Sentiment Analysis Module. (E) QR Module Using Fake QR. (F) Result of QR Module With Fake QR. (G) QR Module Using Real QR. (H) Result of QR Module Using Real

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0351
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2437-2442

Table 2 Compiled Results of Both Models

Sr. No.	Parameter	Sentiment Intensity Analyzer (in%)	CNN (in %)
1.	Accuracy Score	91.2	95.10

Conclusion and Future Work

This paper presents a novel strategy for identifying counterfeit goods using machine learning. It is possible to draw a few implications from the new strategy, including the need for more training data to be collected prior to the system's adoption. This research aims to suggest how to build a device that can capture a product logo image and process it using artificial intelligence, along with text recognition and sentiment analysis of the product reviews, to determine whether a product is genuine or not. This application is portable and simple to use. It will be quite beneficial for those who lack technological expertise. The system focuses on only some categories of products like clothing and accessories. In future work, the system can be extended to be used for all types and categories of products.

References

[1]. Economic Times, Satyanarayan Iyer, "How you can avoid buying fake products while shopping".

[2]. What is Optical Character Recognition, 2022 [Online]. https://www.ibm.com/cloud/blog/optical-character-recognition

[3]. Syed Mohammed Anas, Santoshi Kumari, "Opinion Mining based Fake Product Review Monitoring and Removal System", 2021 6th International Conference on Inventive Computation Technologies (ICICT). IEEE, 2021.

[4]. Saumya, S., & Singh, J. P. (2018). Detection of spam reviews: a sentiment analysis approach. Csi Transactions on ICT, 6(2), 137-148.

[5]. Reema Anne Roy, Sunita R Patil, "Fake Product Monitoring System using Artificial Intelligence"

[6]. nltk.sentiment.vader module, 2023, https://www.nltk.org/api/nltk.sentiment.vader.html, Jan 2023

[7]. Sentimental Analysis Using Vader, 2020. https://towardsdatascience.com/sentimental - analysis-using-vader-a3415fef7664

[8]. Eduard Dao ud, Hung Nguyen, Martin Gaedke, Dang Vu Nguyen Hai, "Improving Fake Product Detection using AI-Based Technology", 2020.

[9]. Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter, Hamzah Al Najada, "Survey of review spam detection using machine learning techniques",Springer, 2015

[10]. Md Mahadi Hassan Sohan, Mohammad Monirujjaman Khan, Ipseeta Nanda, "Fake Product Review Detection using Machine Learning", IEEE, 2022

[11]. Fake review detection using opinion mining" by Dhairya Patel, Aishwerya Kapoor and SameetSonawane, International Research journal of Engineering and technology (IRJET), volume 5, issue 12, 2018.

[12]. Daoud, E., Vu, D., Nguyen, H., & Gaedke, M. (2020, April). Improving fake product detection using ai-based technology. In 18th International Conference e-Society (Vol. 12, pp. 1-9).

[13]. Bandakkanavar RV, Ramesh M, Geeta H (2014) A survey on detection of reviews using sentiment classification of methods. IJRITCC 2(2):310–314.