



Partially Supervised Learning Group Detection

¹Ajeet Singh, ²Aashish Panwar, ³Dr. Vibha Srivastava

¹Undergraduate, ²Undergraduate, ³Assistant Professor

^{1,2,3}IT Department,

^{1,2,3}SRMCEM, Lucknow, India

Abstract: With the continual evolution of E-commerce platforms, online evaluations are increasingly seen as a critical aspect in establishing and maintaining a positive reputation. Furthermore, they play an important role in the decision-making process for end users. A positive review for a specific object typically draws more customers and leads to a significant rise in sales. Deceptive or phony evaluations are now purposefully generated in order to build a virtual reputation and attract potential clients. As a result, detecting false reviews is an active and ongoing research topic. Identifying phony reviews is dependent not only on the essential elements of the reviews, but also on the reviewers' behaviour. This research provides a machine learning method for detecting false reviews. In addition to the review features extraction approach, this research employs different features engineering techniques to extract diverse reviewer behaviours. The study examines the performance of machine learning classifiers; KNN, Naive Bayes, and Logistic Regression using a genuine Yelp dataset of restaurant reviews. In terms of accuracy, the results show that Logistic Regression surpasses the other classifiers. The results demonstrate that the system is better at determining whether a review is fake or genuine.

Keywords – Machine learning, fake, reviews, Naive Bayes, KNN, Logistic Regression.

I. INTRODUCTION

Nowadays, reviews have replaced other sources of information for consumers looking to make decisions regarding services or items in the modern era. Reviews are regarded as real feedback regarding good or bad services, therefore any attempt to skew them by including false or misleading information is seen as dishonest behavior and is accompanied by the label "fake reviews." Such a situation makes us wonder what would happen if not all submitted reviews are trustworthy or sincere. What if any of these testimonials are false? As a result, it has become and continues to be a necessary and active study topic to identify false reviews.

II. OBJECTIVES

Main objectives of our project are:

- The objective of this project is to create an Effective Detection system for Fake reviews from the text in order to get rid from fake reviews while purchasing a product.
- To build an environment for the users, helping them gain more knowledge about the original reviews of the product.
- To provide users with the correct feedback so they can buy good product.
- Allows users to check the accuracy of the product review dataset using multiple algorithms.
- To provide Create, Read, Update, and Delete (CRUD) functionality to users for product review datasets.
- Allows users to predict whether the reviews are fake or original.

III. SYSTEM ANALYSIS & FEASIBILITY STUDY

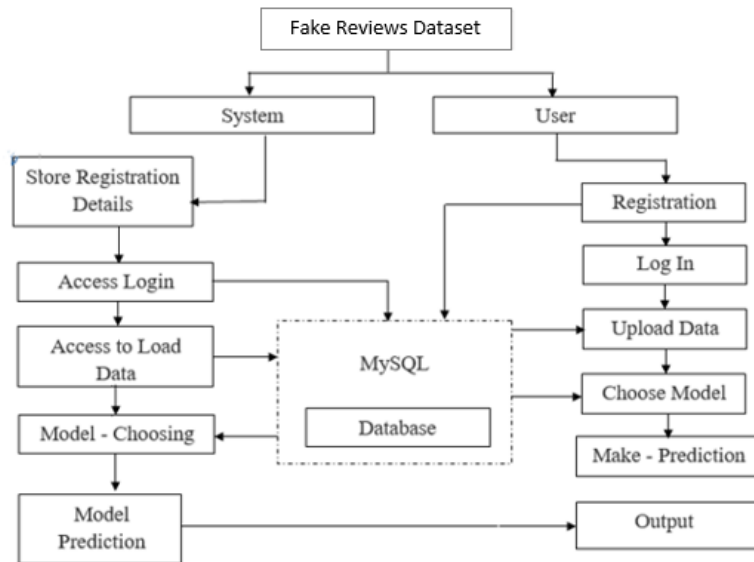
3.1 EXISTING METHODOLOGY

The Machine learning is becoming more and more prevalent, with classical and machine learning approaches used in computer science. This section discusses relevant research on detecting bogus reviews and how machine learning techniques outperform conventional ones. The project's current methodology follows a specific flow, and RF is also utilized for model creation. However, the result is inaccurate and a lot of memory is needed.

3.2 PROPOSED METHODOLOGY

We propose this application, which can be seen as a valuable system because it helps in reducing the constraints brought about by conventional and other existing ways. The goal of this project is to create an efficient, dependable approach for precisely detecting and estimating anaemia. We developed a robust method to design this system in a Python environment using the Django framework.

IV. BLOCK DIAGRAM & FEASIBILITY STUDY



V. MODULE DESCRIPTION

5.1 USER MODULE:

5.1.1 HOMEPAGE:

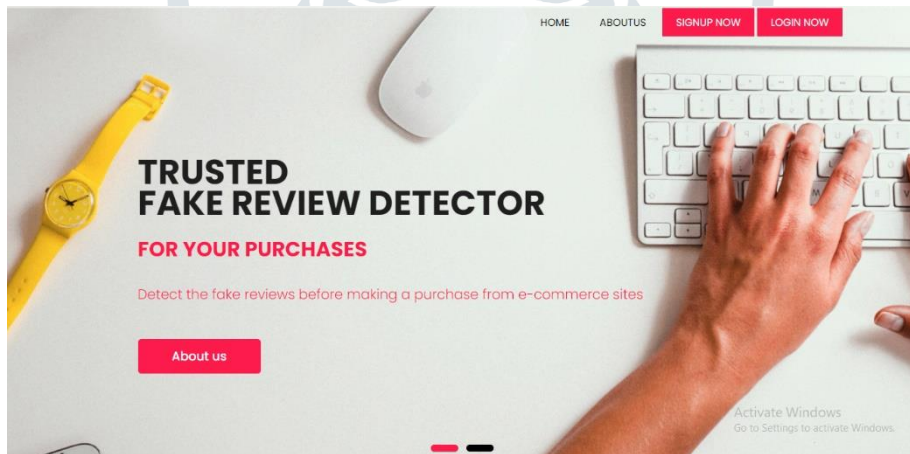


Fig 5.1.1 In our project, we are detecting the fake reviews from the review entered by the user.

5.1.2 ABOUT:



Fig 5.1.2 Here the application describes what main objective of this project is.

5.1.3 REGISTRATION:

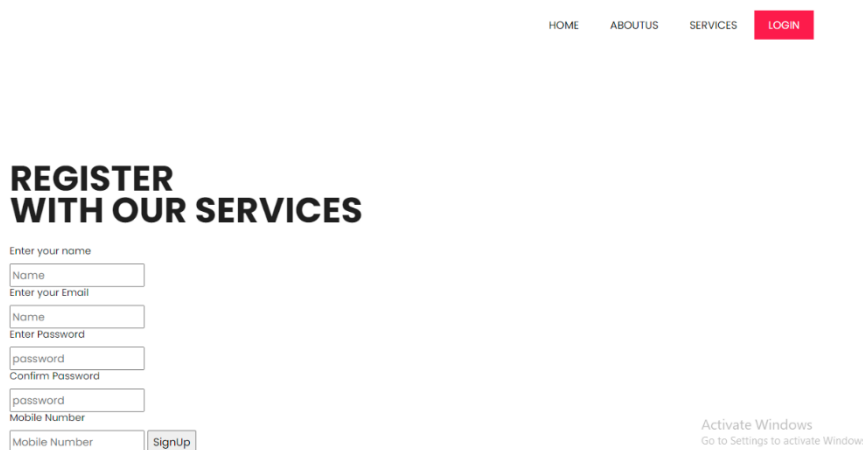


Fig 5.1.3 Registration page in which user need to register to start.

5.1.4 LOGIN:

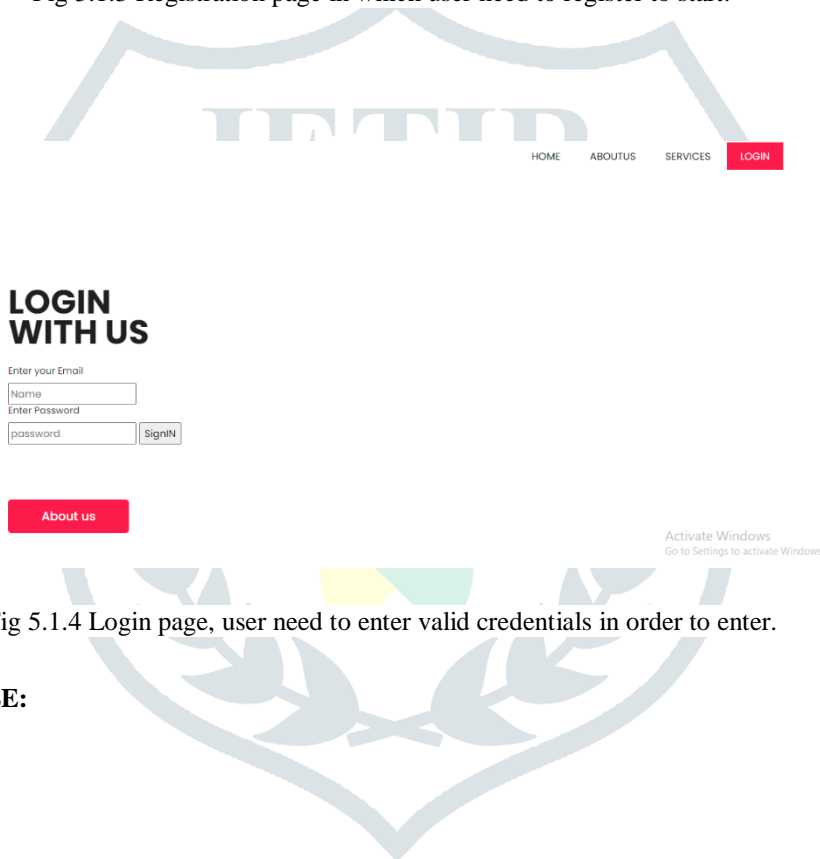


Fig 5.1.4 Login page, user need to enter valid credentials in order to enter.

5.2 SYSTEM MODULE:

5.2.1 UPLOAD:

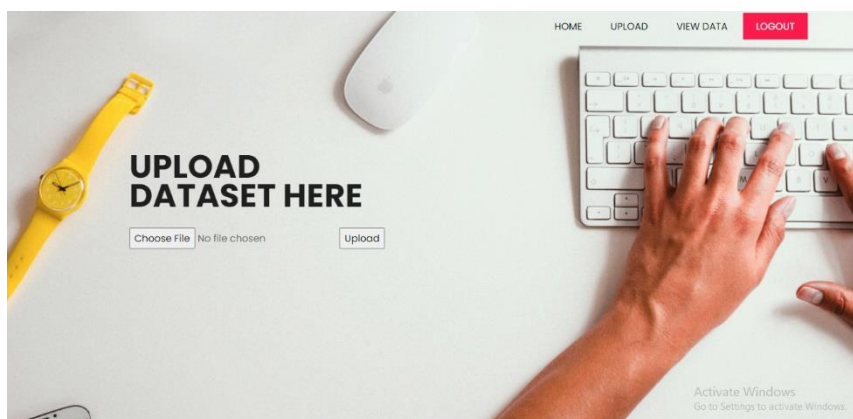


Fig 5.2.1 Upload Page in order to upload the dataset.

5.2.2 VIEW DATA:

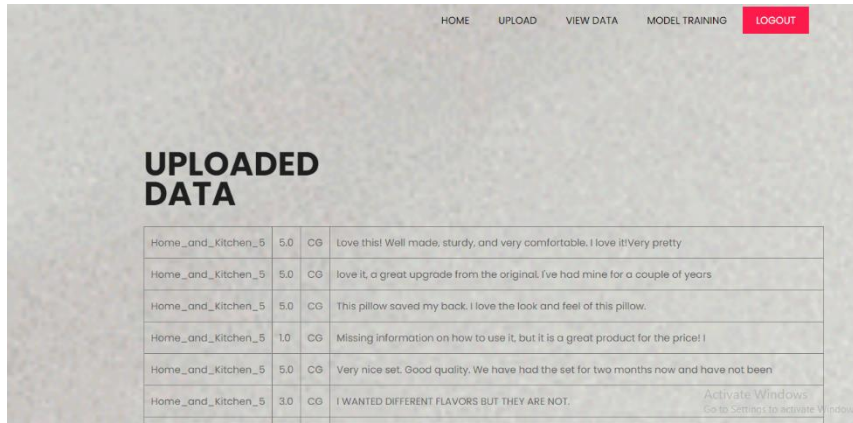


Fig 5.2.2 User views the data which he was uploaded to the system.

5.2.3 MODEL TRAINING:

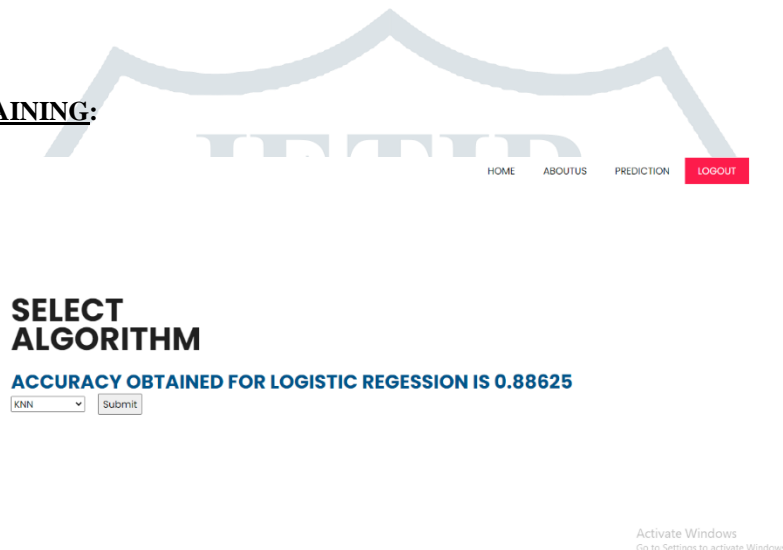


Fig 5.2.3 Here training of your model takes place and display the model's accuracy

5.2.4 PREDICTION:



Fig 5.2.4 User need to enter the required fields in order to get the response from the data whether the review is computer generated or original.

VI. METHODOLOGY AND ALGORITHMS USED

6.1 NAIVE BAYES:

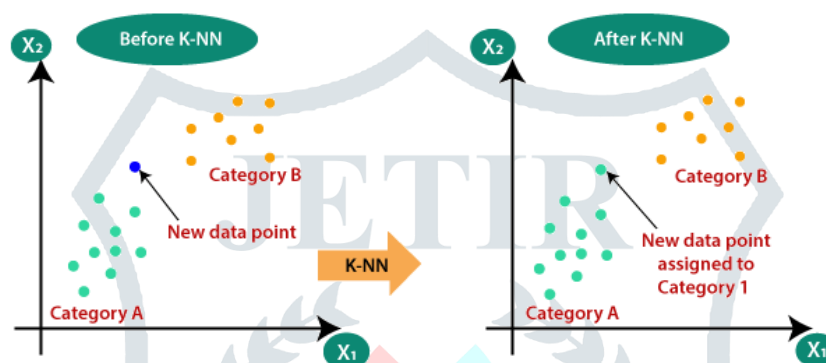
A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naive.

6.2 K-Nearest Neighbor:

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique, it assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.



K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. It is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

6.3 LOGISTIC REGRESSION:

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable (target) is categorical.

Logistic regression has become particularly popular in online advertising, enabling marketers to predict the likelihood of specific website users who will click on particular advertisements as a yes or no percentage.

VII. FUTURE WORK

This program may eventually include the capability to recognize several categories. With the updated data set, we plan to investigate the prediction approach and use the most precise and pertinent machine learning algorithms for detection.

VIII. CONCLUSION

In this application, we have successfully created a system to identify fake reviews. With Python programming and the Django framework, this is produced in a user-friendly environment. In order to assess if the review is fraudulent or not, the system is likely to collect information from the user.

REFERENCES

- [1] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," *Information Processing & Management*, vol. 56, no. 4, pp. 1234 – 1244, 2019.
- [2] Tadelis, "The economics of reputation and feedback systems in e-commerce marketplaces," *IEEE Internet Computing*, vol. 20, no. 1, pp. 12–19, 2016.
- [3] M. J. H. Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Information Retrieval*, vol. 9, no. 6, 2018.
- [4] C. C. Aggarwal, "Opinion mining and sentiment analysis," in *Machine Learning for Text*. Springer, 2018, pp. 413–434.
- [5] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in *Seventh international AAAI conference on weblogs and social media*, 2013.
- [6] N. Jindal and B. Liu, "Review spam detection," in *Proceedings of the 16th International Conference on World Wide Web*, ser. WWW '07, 2007.