iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: https://doi.org/10.22214/ijraset.2022.41687

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

Research Paper on Fake Online Reviews Detection using Semi-supervised and Supervised learning

Ajanta Chettri[1], Amal George[2], Dr. A. Rengarajan[3], Feon Jaison [4]
[1, 2]*School of CS and IT, Jain University, Bangalore*
[3, 4]*Professor, School of CS and IT, Jain University, Bangalore*

Abstract: ***Today's business and commerce are heavily influenced by online reviews. Most online product purchase decisions are based on customer reviews. As a result, opportunistic individuals or groups seek to shake product reviews in their favor. Fake online reviews have a significant impact on the efficiency of online consumers, merchants and e-commerce markets. Despite academic efforts to study fake reviews, there remains a need for research that can systematically analyze and summarize their causes and consequences. This task provides a semi-supervised and supervised text mining model for detecting fake web reviews and comparing their effectiveness to hotel review datasets.***
Keywords: ***Semi-Supervised. Supervised, Detection, Fake Review, Marketing***

I. INTRODUCTION

When buying or ordering something online, first check the ratings and reviews from the people who tried or bought the product. If your rating is high and your review meets or exceeds your expectations, don't hesitate to choose Buy Now. However, the problem that arises here is that the customer does not know the true identity of the other person. It is almost impossible to determine if a published review is from a trusted party. It could be a paid review to promote your product, or a review posted by a loved one to promote something. This increases the likelihood of false assumptions about product quality and unwanted advertising. Consumers learn more about their products from seller descriptions and user reviews published on their e-commerce platforms. Reviews allow criminals to influence the sale of goods. These false reviews certainly influence consumer purchasing decisions. Studies show that consumers are very sensitive to both positive and negative feedback on their products. Review information can influence consumers' purchasing decisions, transforming product sales and generating significant economic benefits. It encourages fraudulent reviews to some extent. Due to the large amount of product evaluation data, manual screening is costly. Therefore, detecting fake reviews has become a valuable study. It takes millions of online users to investigate how these reviews are created and complete a solution to recognize and classify them as fake reviews and true reviews. Helps you choose accurate and high quality products.

As the Internet economy develops, new forms of online fraud are constantly emerging: False verification methods and formats are becoming more diverse. Research to detect fake reviews will also be improved and deepened. Most of the true assessments include nouns, adjectives, prepositions, determiners, and parallel conjunctions to describe clear and sensory properties. Verbs, adverbs, and pronouns are all used in some false reviews. Fake reviews can be identified based on the distribution of part of speech in the review text. The majority of fake reviewers participate in the creation of fake reviews by accepting the work of brushing. They try to give positive or negative ratings to items they are unfamiliar with, and they must have certain characteristics because they have never used the item. Reviewer behavioral characteristics such as B. Reviewer activity window, maximum number of reviews per day, total number of reviews, percentage of positive and deviant reviews of users, based on the metadata information above. It can be derived after a thorough examination.

II. EXISTING METHODOLOGIES

Creating fake reviews and ratings to support products on your site to improve reputation and sales without genuine feedback can be unfair and misleading. This is a common practice today and the need for fake review detectors is increasing. The content-based approach focuses on the content of the review. This is the text of the review, or what is stated in it. By assessing the linguistic aspect of the review, Heydari sought to detect fake reviews. Ott etc. to classify data in three ways. Genre identification, psycholinguistic deception detection, and text classification are three strategies. Reviewers are the focus of behavioral trait-based research that incorporates the traits of the reviewer. Limetal. Addressed specific issues with users who are the source of review spammer detection or spam reviews from var.

©IJRASET: All Rights are Reserved

People who intentionally post fake reviews behave very differently than the average user. The following misleading evaluations and evaluation behaviors were noticed by them.Identifying misleading internet reviews is often considered a classification issue. Supervised text classification algorithms are the main solution. When training in Large dataset of labeled instances from both classes, misleading opinions (positive) Instances) and true opinions (negative instances), these strategies are resilient (negative) example).

According to researchers' observations and experimental results, existing systems use naive Bayes classifiers to classify spam and non-spam, which is very inaccurate and may not provide accurate results to users. The semi-supervised classification approach has also been used by several researchers.

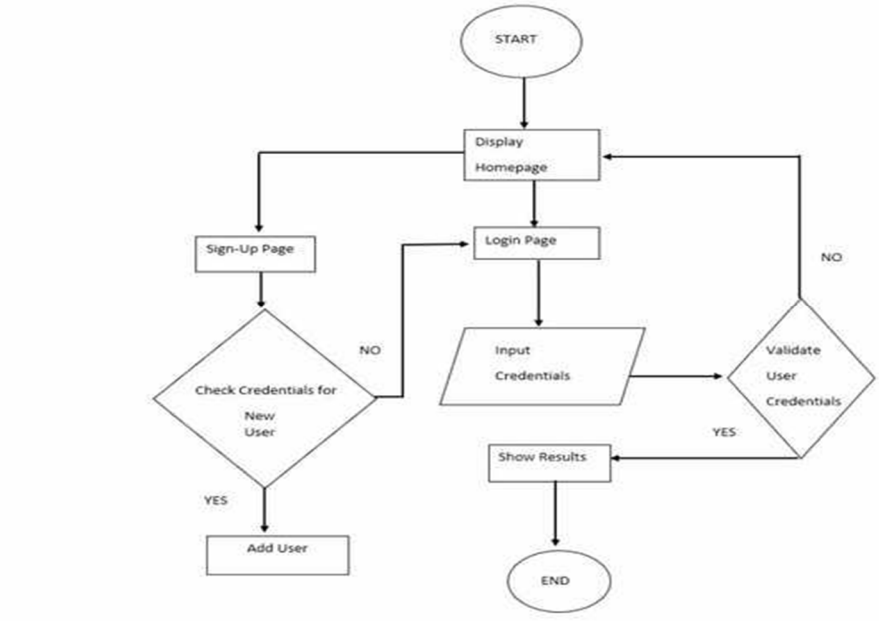
Logistic regression analysis is an existing method of identifying scores using parameters. Determines the importance of the element. Naive Bayes algorithm uses conditional Probability of function and non-aligned assumptions. There is also the use of decisions A tree that selects the use of new instances depending on the attribute value. Fake review authors are always looking for more efficient ways to create large-scale fake reviews with minimal human intervention. To defend against these approaches, researchers need to experiment with computer-generated fake reviews and create classifiers based on synthetic output to mimic the behavior of real fake review attackers.

III. PROPOSED SYSTEM

Each rating reflects the continued use of the word or its impact on the product. Therefore, all validation of the proposed system goes through the tokenization process first. Given the Data for immediate response, ratings will be validated and subject to review Because of the need and method of cooking words. Abolished after deletion Words form potential feature words. In most cases users tend to use Immediately created words that do not reflect the original aspects of the product. Or Expose them to the use of words with similar meanings, copy, write under coercion, or do one Favor to loved ones. This will create and enable an extensible version of the possible word functionality. Identify the proposed system by recognizing the text and comparing it to the proposed system dictionary. All possible feature words are collated against this dictionary to see if they are displayed There, its frequency is counted and added to the corresponding column of the feature vector. For a numerical map of words. The length of the view is measured and added to the feature Vector in addition to count frequency. Finally, the emotion score is from the dataset Included in the feature vector. In the feature vector, we assigned zero to negative emotions Value and good atmosphere are of positive value.

A. How System Works

System starts by displaying the home page that leads to the login or signup page. If the user is an existing user, he can directly login by using the login credentials or else user has to create one account. After signing in, user can use the review on hand or the review ID to check whether it’s fake or not and the result is displayed in bold.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

IV. SYSTEM DESIGN

A. *Input Design*

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system.

The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple.

The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

1) What data should be given as input?
2) How the data should be arranged or coded?
3) The dialog to guide the operating personnel in providing input.
4) Methods for preparing input validations and steps to follow when error occur.

B. *Objectives*

1) Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.
2) It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3) When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

C. *Output Design*

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs.

In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1) Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2) Select methods for presenting information.
3) Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

a) Convey information about past activities, current status or projections of the
b) Future.
c) Signal important events, opportunities, problems, or warnings.
d) Trigger an action.
e) Confirm an action.

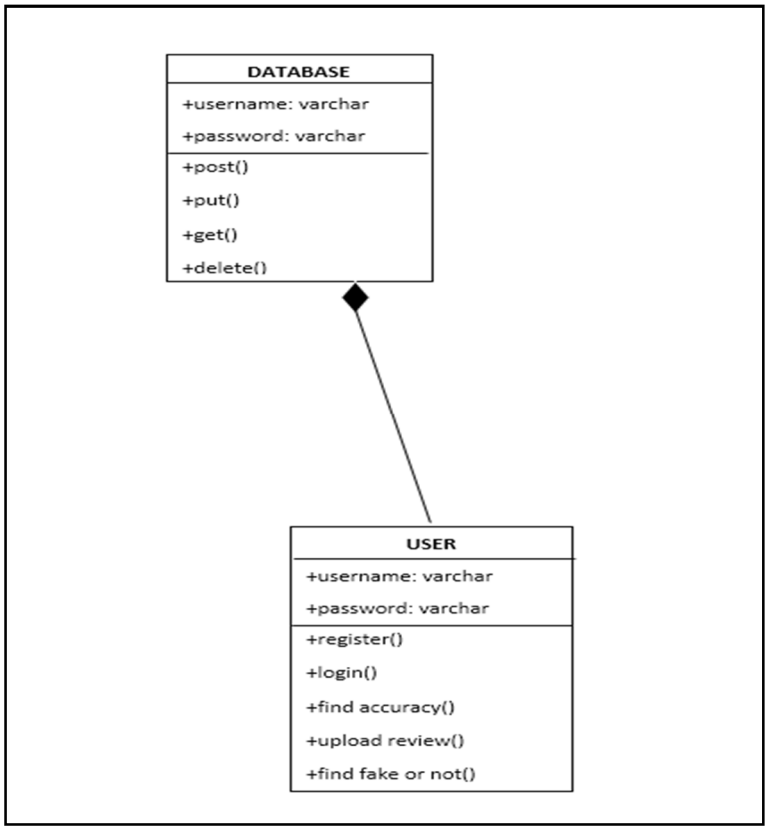
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

V. UML DIAGRAMS

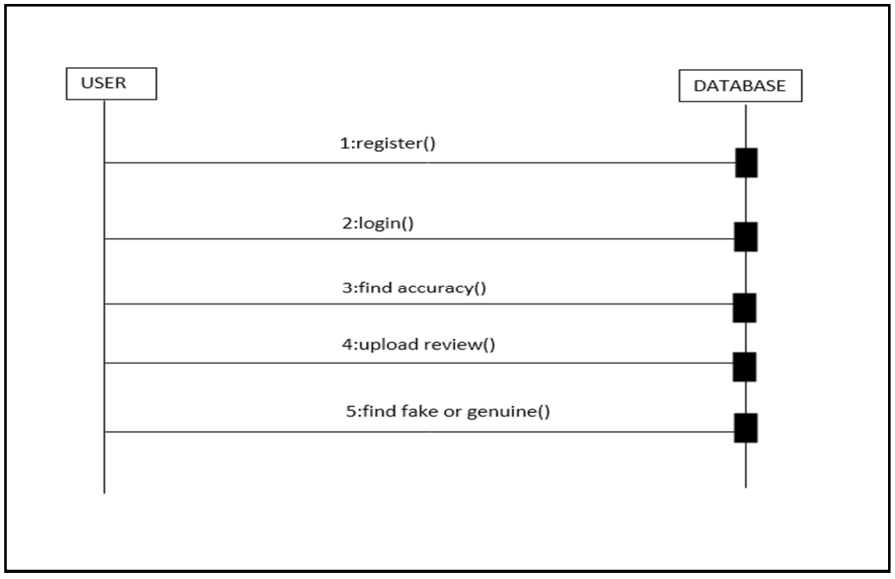
A. Class Diagram

It is the main building block of object-oriented modelling. It is used for general conceptual modelling of the structure of the application, and for detailed modeling, translating the models into programming code.



B. Sequence Diagram

It is a type of interaction diagram because it describes how and what in order. It is used to understand requirements for a new system or to document an existing process.

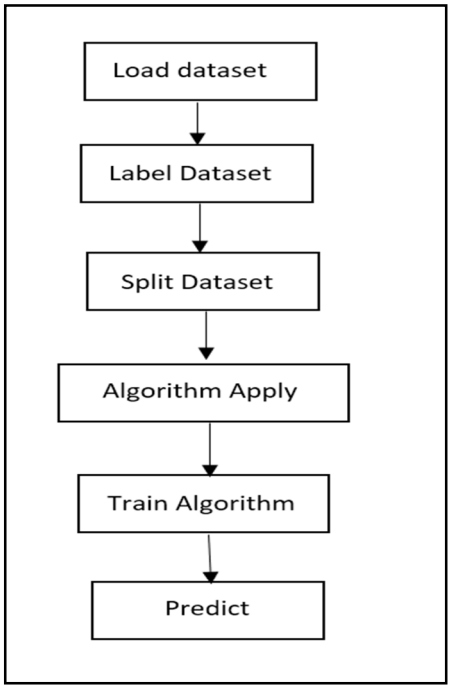


©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IV Apr 2022- Available at www.ijraset.com

C. System Architecture

It is the conceptual model that defines the structure, behavior and more views of a system. An architectural description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system



VI. RESULT AND SCREENSHOTS

Detection of online fake reviews helps to detect and filter fake reviews in online websites and e-commerce platform. This project has been implemented based on Python programming language on DJANGO framework with SQL database in the backend.

This system filters all authentic reviews with any fake reviews in any online platform such as websites and e-commerce sites. Any user can register with the system and make use of the reviews and review id to detect if it is an authentic review or not. This system will generate the results based on the calculation from the machine learning algorithm used with it.



Fig 1:Home Page

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

This is the first page of the web application where the user can navigate and go the login page and the sign-up page.

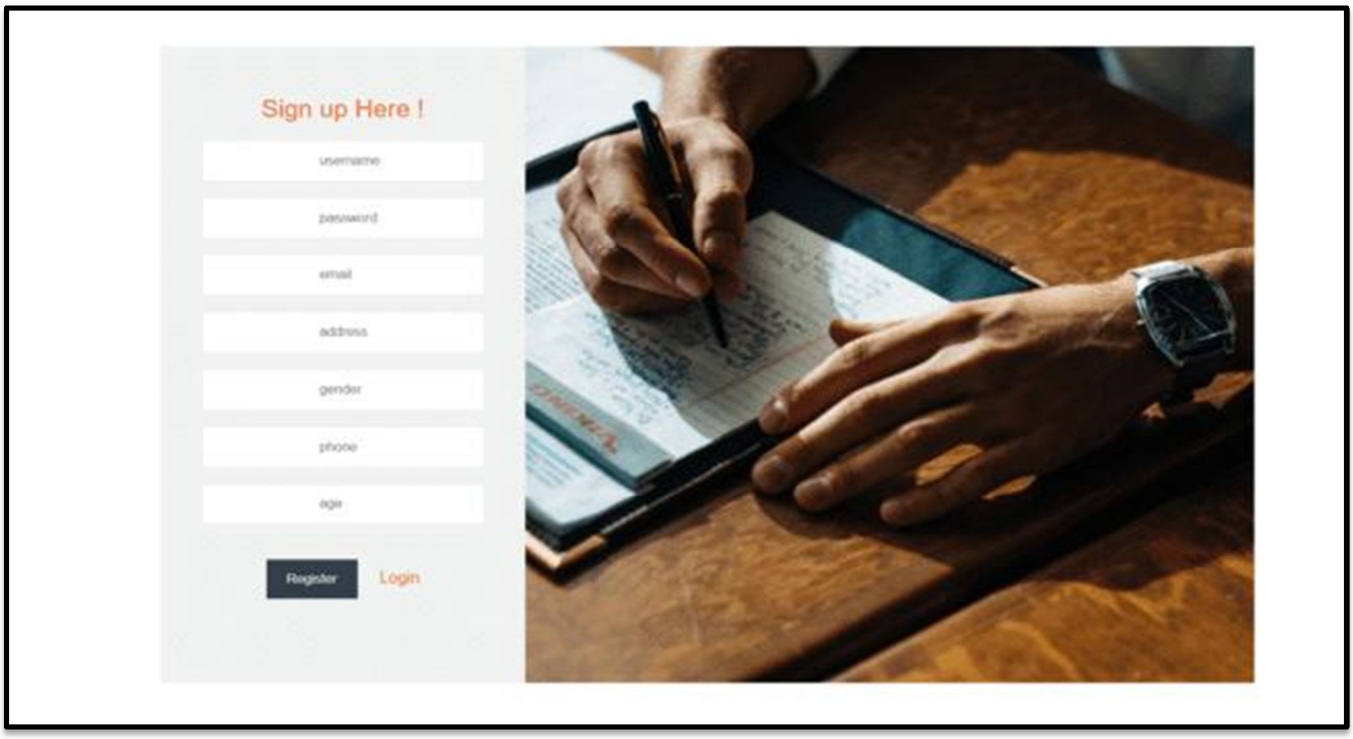


Fig 2:Sign-up Page

This is the Sign-Up page where any user can sign up with the web application to find the authenticity of a review.

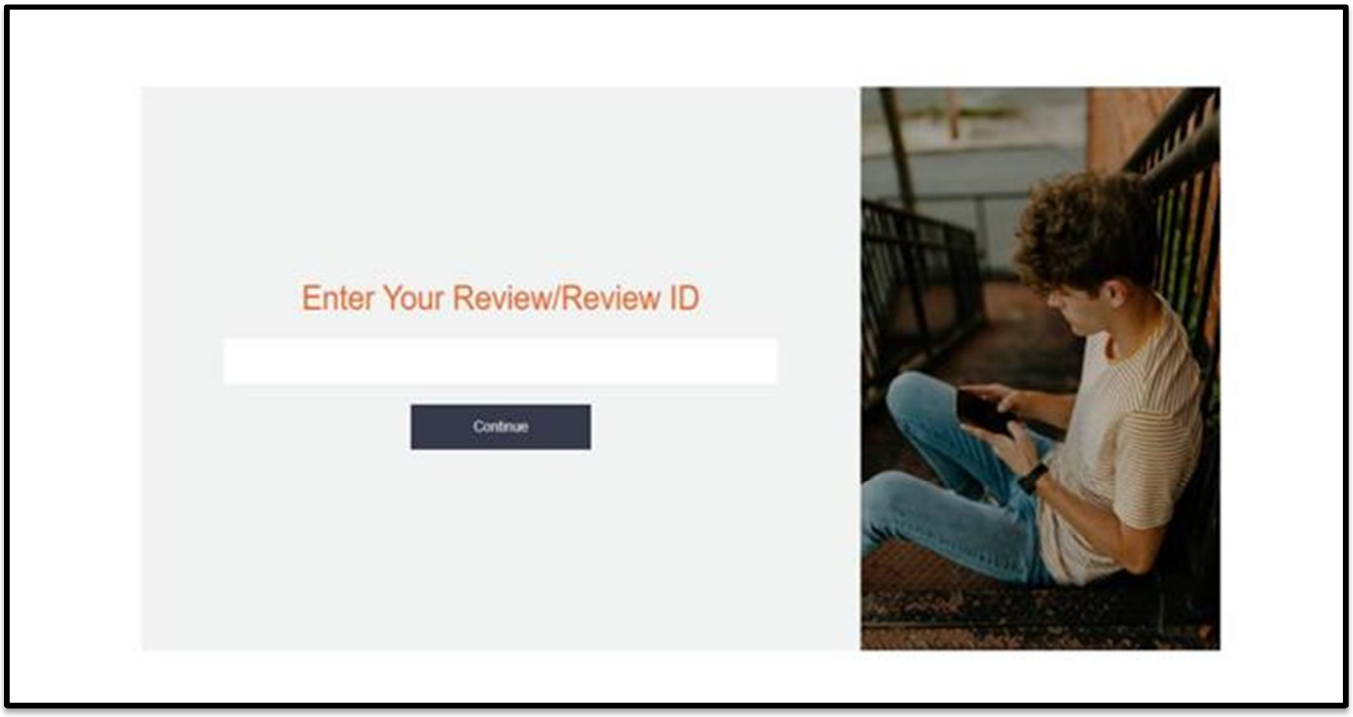


Fig 3:Review Page

After successful login, a user is directed to this page. A user can enter a review or a review id to check if it is a fake review or not.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

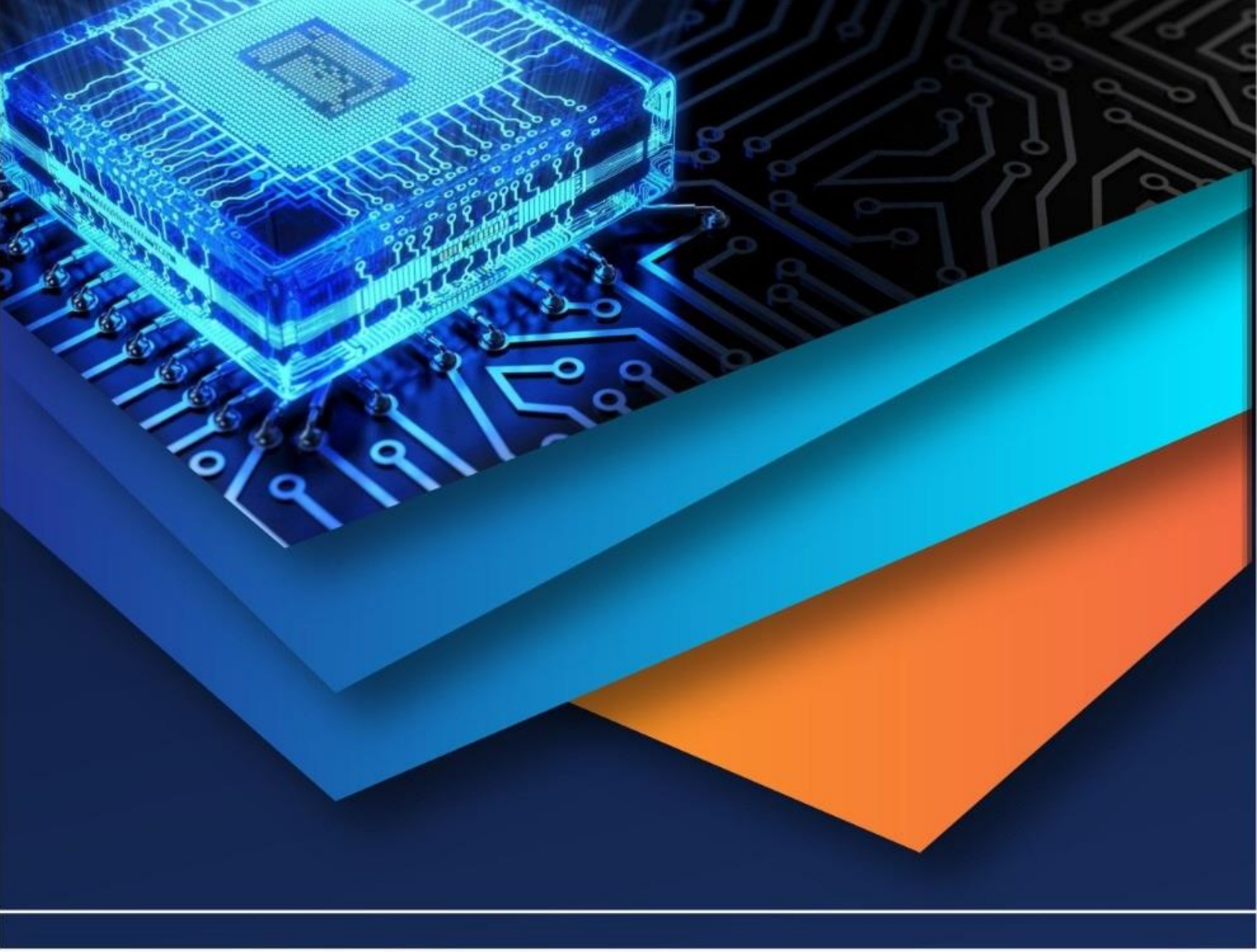
VII. CONCLUSION

In this study, we demonstrated different semi-supervised and supervised text mining approaches for detecting false internet reviews. To generate a advanced characteristic set, we merged capabilities from one of a kind studies studies. We additionally applied a one of a kind classifier that turned into not used withinside the preceding study. As a result, we have been capable of enhance the accuracy of Jiten et al in advance semi-supervised approaches. We additionally determined that the maximum correct classifier is the supervised Naive Bayes classifier. This guarantees that our dataset is efficiently labelled, as we recognize that semi-supervised fashions carry out properly while honest labelling is unavailable. We centred completely on consumer opinions in our investigation. In the future, consumer movements and texts can be included to create a greater correct category algorithm. To make the dataset greater exact, superior training methods for tokenization would possibly be applied. A large statistics set may be used to evaluate the achievement of the proposed methodology. In future, prediction with accurate decimal display system and extended library of dataset will be introduced.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, “Exploiting Product Related Review Features for Fake Review Detection,” Mathematical Problems in Engineering, 2016.
[2] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ”Detection of review spam: a survey”, Expert Systems with Applications, vol. 42, no. 7, pp. 3634–3642, 2015.
[3] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, “Finding deceptive opinion spam by any stretch of the imagination,” in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), vol. 1, pp. 309–319, Association for Computational Linguistics, Portland, Ore, USA, June 2011.
[4] J. W. Pennebaker, M. E. Francis, and R. J. Booth, ”Linguistic Inquiry and Word Count: Liwc,” vol. 71, 2001.
[5] S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, Vol. 2, 2012.
[6] J. Li, M. Ott, C. Cardie, and E. Hovy, “Towards a general rule for identifying deceptive opinion spam,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.
[7] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detecting product review spammers using rating behaviors,” in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.
[8] J. K. Rout, A. Dalmia, and K.-K. R. Choo, “Revisiting semi-supervised learning for online deceptive review detection,” IEEE Access, Vol. 5, pp. 1319–1327, 2017
[9] Beutel A, Murray K, Faloutsos C, Smola AJ (2014) CoBaFi - Collaborative Bayesian filtering. In: Proceedings of 23rd international conference on world wide web, pp 97–108
[10] Cao Q, Sirivianos M, Yang X, Pregueiro T (2012) Aiding the detection of fake accounts in large scale social online services. In: Proceedings of 9th USENIX symposium on networked systems design and implementation, pp 197–210
[11] Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Al Najada H (2015) Survey of review spam detection using machine learning techniques. J Big Data 2(1):23
[12] Harris CG (2012) Detecting deceptive opinion spam using human computation. In: Proceedings of workshops at the 26th AAAI conference on artificial intelligence, vol WS-12-08, pp 87–93
[13] Aghakhani H, Machiry A, Nilizadeh S, Kruegel C, Vigna G (2018) Detecting deceptive reviews using generative adversarial networks.
[14] Badresiya A, Vohra S, Teraiya J (2014) Performance analysis of supervised techniques for review spam detection.
[15] Banerjee S, Chua A, Kim, J (2015) Using supervised learning to classify authentic and fake online reviews.
[16] Bhattarai A, Dasgupta D (2012) A self-supervised approach to comment spam detection based on content analysis.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)