IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Spam Review Detection using Machine Learning

Kiran Naik[1], Kajal Naik[2], Devyani More[3], Dipti Kapadi[4], Ms. Poonam Dholi[5]
Students, Department of Computer Engineering[1,2,3,4]
Professor. Department of Computer Engineering[5]
Matoshri Collage of Engineering, Nashik, India

Abstract: *Customer opinions play a vital role in buying decisions. These days most customers post their opinions about products on blogs, e-commerce sites, review sites, and social networking sites. The above information is consumed by business or corporate organizations, as they are eagerly interested in analyzing consumer views about their products, services, and support. As people buy products after reading the reviews, the kind of reviews that a product attracts are of concern to the sellers. This means that a positive review on the product would bring in sales and a negative one would reduce them. The project leverages Natural Language Processing (NLP) techniques and supervised learning algorithms to build a robust spam review detection system. The system is trained on a dataset comprising genuine and spam reviews, and it extracts various features from the textual content of reviews, such as sentiment analysis, linguistic patterns, and semantic meaning.*

Keywords: Spam Review Detection, Machine Learning, Support Vector Machines, Natural Language Processing, Text Analysis

I. INTRODUCTION

With the continuous evolution of E-commerce systems, online reviews are mainly considered as a crucial factor for building and maintaining a good reputation. Moreover, they have an effective role in the decision-making process for end users. Usually, a positive review for a target object attracts more customers and leads to a high increase in sales. Nowadays, online reviews have become one of the vital elements for customers to do online shopping. Organizations and individuals use this information to buy the right products and make business decisions. This has influenced spammers or unethical business people to create false reviews and promote their products to beat t h e competition. Sophisticated systems are developed by spammers to create bulk spam reviews on any website within hours. To tackle this problem, studies have been conducted to formulate effective ways to detect spam reviews. Various spam detection methods have been introduced in which most of them extract meaningful features from the text or use machine learning techniques. These approaches gave little importance to extracted feature type and processing rate. NetSpam defines a framework that can classify the review dataset based on spam features and maps them to a spam detection procedure that performs better than previous works in predictive accuracy. In this work, a method is proposed that can improve the processing rate by applying a distributed approach to review datasets using the MapReduce feature. Parallel programming concept using MapReduce is used for processing big data in Hadoop. The solution involves parallelizing the algorithm defined in NetSpam and it defines a spam detection procedure with better predictive accuracy and processing rate

II. LITERATURE SURVEY

This chapter discusses brief literature regarding the project. A literature survey is mainly used to identify information relevant to the project work and know the impact of it within the project area. It defines as yet how many surveys have been done w i t h knowledge of the latest technology and implementation designs

Sr. No	Title	Year and Author	Details	Advantage	Limitation
1	Machine Learning Approaches for Review Spam	2017, Wang, X., Zhang, L.,	Explores machine learning techniques for identifying spam reviews and the role of	Discusses machine learning applications Focuses on NLP and	Machine Learning, Natural Language Processing

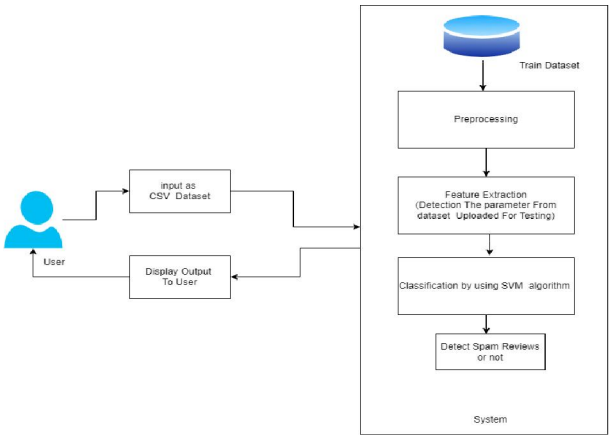
IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

	Detection		NLP and feature engineering.	feature engineering	
2	An Efficient Spam Review Detection Us-ing Active Deep Learning Classi-fiers	2021 , Mehul Bhundiya	To detect spam views we have used deep learn- ing techniques: Multi-Layer Perceptron, Convolutional Neural Network, and Long Short- Term Recur- rent Neural Network	Offers a broad range of opinion spam Exam- ines motivations of spammers	Lacks deep learning insights

III. DESIGN

This chapter introduces the architecture of the system and modules of the system. It also contains the registration, verification, authentication process, and functioning of the system.DFD and UML diagrams are explained in this chapter.

System Architecture



Data Flow Diagrams

A data flow diagram(DFD) is a graphical representation of the flow of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system without going into great detail, which can later be elaborated.

DFD 0 Level diagram

A context diagram is a top-level (also known as ”Level 0”) data flow diagram. Figure 4.2 shows a Level 0 diagram. It only contains one process node(”Process 0”) that generalizes the function of the entire system in relationship to external entities.

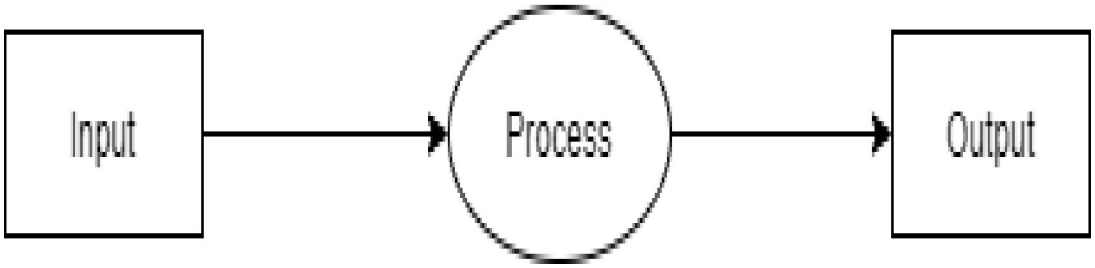


Figure: DFD Leval 0 diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

DFD 1 Level diagram

The Level 1 DFD shows how the system is divided into sub-systems (processes), each of which deals with one or more of the data flows to or from an external agent and which together provide all of the functionality of the system as a whole. Figure shows a Level 1 diagram.

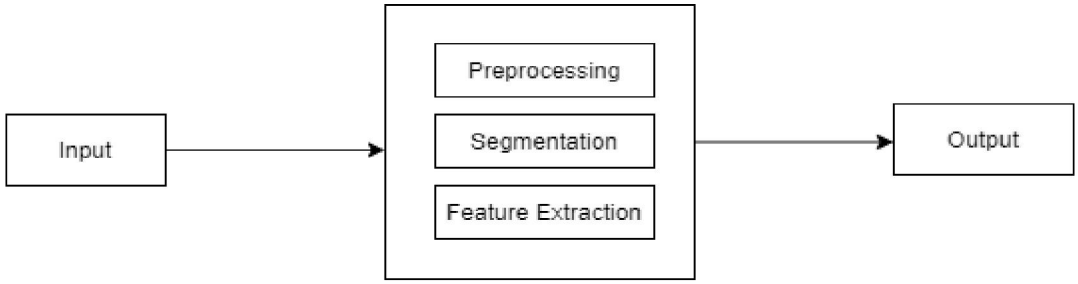


Figure: DFD Leval 1 diagram

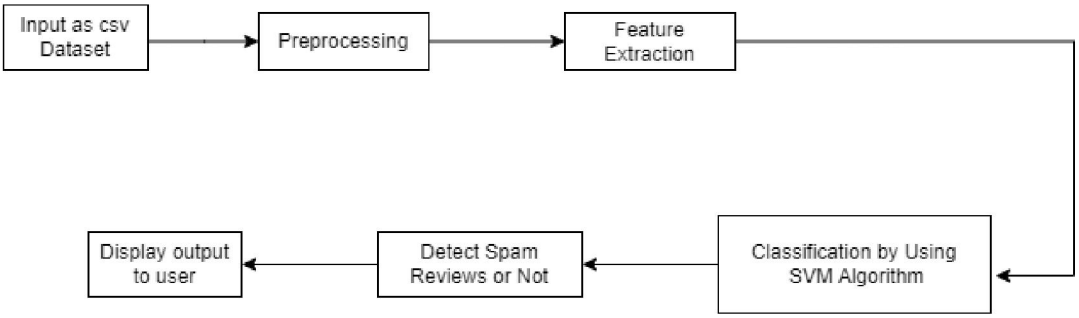


Figure : DFD Leval 1(1) diagram

Class Diagram

Class diagrams are the most common diagrams used in UML. A class diagram consists of classes, interfaces, associations, and collaboration. It represents the object-oriented view of a system that is static. Figure 4.8 shows a class diagram.

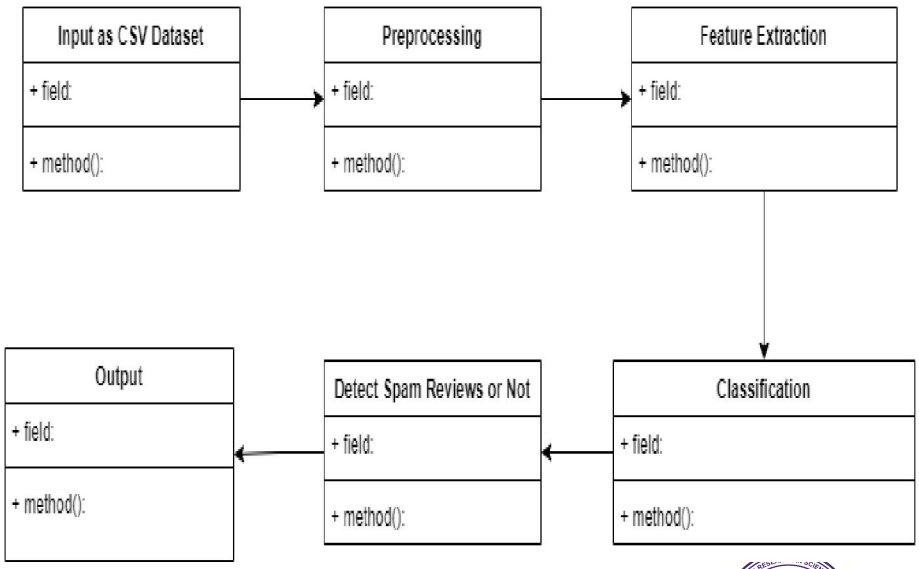


Figure: Class Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Activity Diagram

Activity diagrams are the graphical representation of workflows of step-wise activities and actions with support for choice, iteration, and concurrency. In this Unified Modeling Language, an Activity diagram can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram is constructed from a limited number of shapes, connected with arrows. The most important shape type: Arrows run from the start towards the end representing the order in which activity has been done.

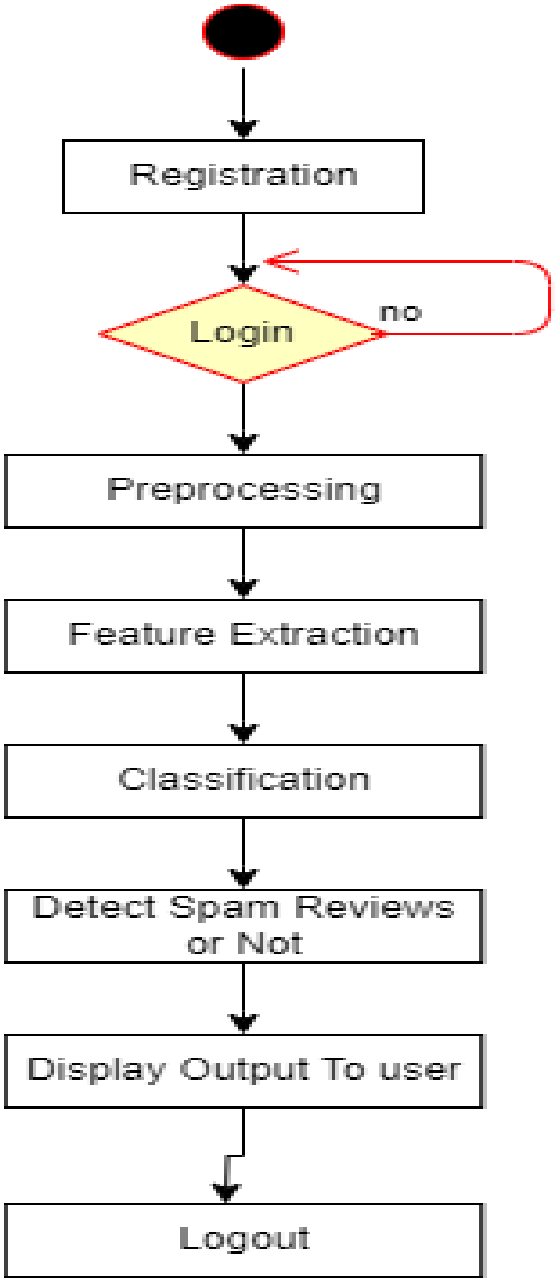
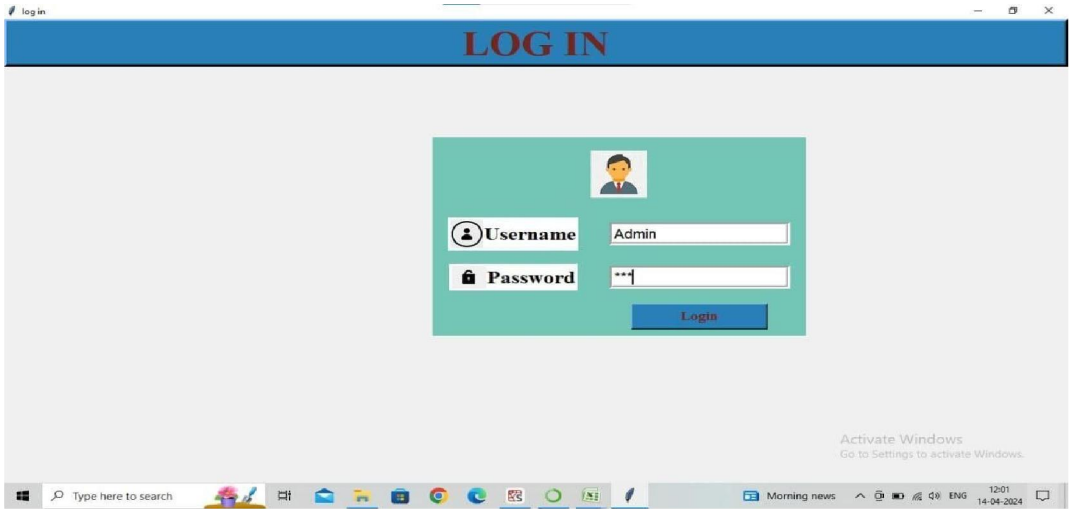


Figure 4: Activity Diagram

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

IV. OUTPUT

Log In



Registration page:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, April 2024

Spam Detection Page:

V. CONCLUSION

To detect spam reviews, supervised methods are more common than unsupervised ones. However, no previous study has addressed the problem of the Persian language. In the current study, a supervised framework is proposed for spam review detection in which different problems of the Persian language are considered. To train the supervised classifiers, a new spam review dataset is created, Spam-Per, using customers' reviews published on the digikala.com website. In the proposed framework, Na¨ıve Bayes, SVM, and decision tree classifiers are used as they have shown reasonable results for English and Arabic languages. These algorithms were trained on metadata and review-based features. Results show that the performance of the proposed system is higher when it is trained and evaluated on a balanced version of Spam Per. Also, the results suggested that SVM using metadata and combined features achieves an acceptable result when applied.

REFERENCES

[1] Lim, Ee-Peng, et al." Detecting product review spammers using rating behaviors. "Proceedings of the 19th ACM International Conference on Information and knowledge management. ACM,2010.
[2] Jindal, Nitin, and Bing Liu. "Review spam detection." Proceedings of the 16th International Conference on World Wide Web. ACM,2007.
[3] Dixit, Snehal, and A. J. Agrawal. "Survey on review spam detection." IntJ Comput Commun Technol ISSN (PRINT) 4 (2013):0975-7449.
[4] Jindal, Nitin, and Bing Liu. "Opinion spam and analysis." Proceedings of the 2008 International Conference on Web Search and Data Mining. ACM, 2008.
[5] Trivedi, Shrawan Kumar, and Shubhamoy Dey. "Effect of feature selection methods machine learning classifiers for detecting email spam."Proceedings of the 2013 Research in Adaptive and Convergent Systems. ACM