



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com



Email Spam Detection Using Machine Learning

Prof. Kamble D. R., Yewale Pradip Dashrath, Chopade Sudam Sambhaji, Gujar Dattatray Navanath,

Assistant Professor, Department of Computer Engineering, S. B. Patil Technical Campus, Indapur, India

B. E. Department of Computer Engineering, S. B. Patil College of Engineering, Indapur, India

B. E. Department of Computer Engineering, S. B. Patil College of Engineering, Indapur, India

B. E. Department of Computer Engineering, S. B. Patil College of Engineering, Indapur, India

ABSTRACT-For business purposes, email is the preferred form of formal communication. Despite there being other forms of communication, email usage keeps growing. In the modern environment, where email volume is increasing daily, automated email management is crucial. More than 55 percent of the total emails are deemed to be spam. This demonstrates how these spams suck up user resources and time while producing little helpful results. It is crucial to comprehend the various spam email classification approaches and their workings because spammers employ sophisticated and inventive methods to carry out their illicit actions via spam emails. The machine learning algorithms used for spam classification are the main topic of this paper. Additionally, this paper offers a thorough analysis and assessment of previous research on various machine learning methodologies, email properties, and techniques. Additionally, it outlines potential paths for future study as well as difficulties encountered in the field of spam classification.

KEYWORDS- email spam detection; feature selection; classification; optimization.

I. INTRODUCTION

Real-world situations often have datasets with large dimensions and superfluous, worthless data, which makes it difficult to analyse this data. Feature selection (FS) is one of the pre-processing procedures in machine learning that can eliminate pointless and unnecessary characteristics from a dataset and identify the final subset of crucial features that will improve the performance of machine learning algorithms [1–3]. In fact, FS is a crucial and frequently used technique in data mining and machine learning for the reduction of dimensions through the removal of unnecessary and redundant data from the dataset in order to achieve the optimal feature subset that increases the efficiency and effectiveness of classification algorithms. Finding the ideal feature subset however, poses a complicated optimization problem that cannot be resolved using standard techniques. In actuality, the objective of FS is to identify a subset of features from the total collection of features in order to enhance the efficiency of learning algorithms in terms of classification accuracy or learning time. Two frameworks, including the wrapper-based and filter-based techniques, have so far been suggested for effectively resolving the FS problem [6, 7]. The two essential elements of the former techniques are the search strategy and evaluation criterion [8]. The search strategy specifies the procedure for producing a solution for an ideal feature subset. A specific criterion is used to evaluate each created solution. The subset creation and evaluation procedure is continued until a stopping criterion is fulfilled in this technique, of course, in an effort to improve search methods in subsequent rounds. Contrary to the first set of approaches, the second set uses abundance, relationship, and connection between features to detect superfluous and unnecessary features. Problems can be divided into two groups: wrapper based approaches and filter based solutions.

II. EXISTING SYSTEM

Email spam has become a major problem nowadays, with rapid growth of internet users, email spam is also increasing. People are using them for illegal and unethical conducts, phishing and fraud. Sending malicious link through spam emails which can harm the system and can also seek in into the system.



III. PROPOSED SYSTEM

Due importance shall be given to experimental setups, procedures adopted, techniques developed, methodologies developed and adopted. While important derivations/formulae should normally be presented in the text of these chapters, extensive and long treatments, copious details and tedious information, detailed results in tabular and graphical forms may be presented in Appendices. Representative data in table and figures may, however, be included in appropriate chapters.

IV. ARCHITECTURE

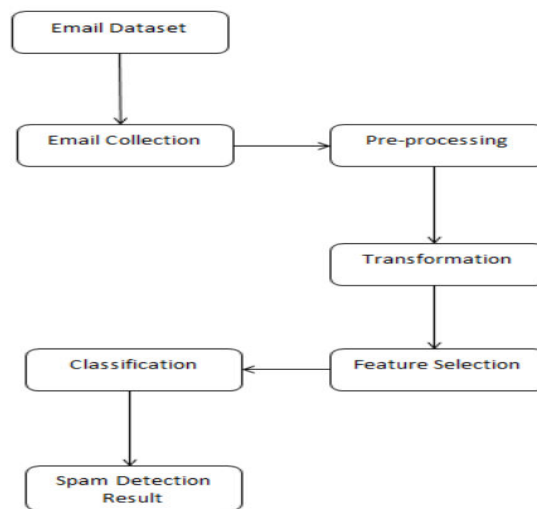


Fig 1: Architecture

V. ALGORITHM

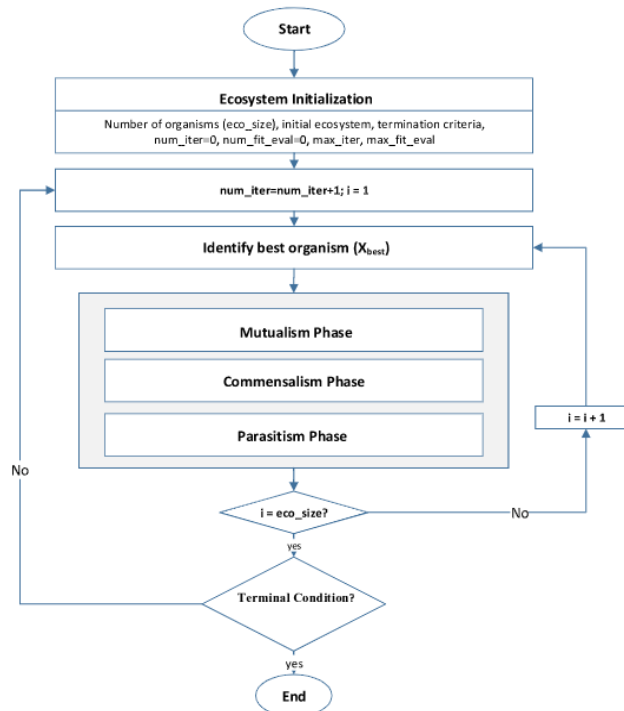


Fig 2: Algorithm

The SOS Algorithm will briefly be studied. It is meta-heuristic inspired by the opposition by the organisms presented by Cheng and Prayogo in 2014 for solving organism problem. The algorithm starts working with an initial population called the ecosystem where the group of organisms is randomly generated in the search space, and each organism is considered a solution for solving the optimization problem.

VI. DATA FLOW DIAGRAM

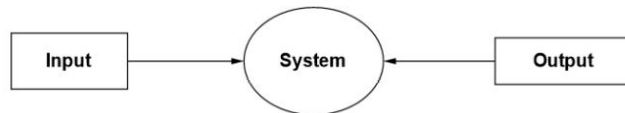


Fig 3: DFD 0

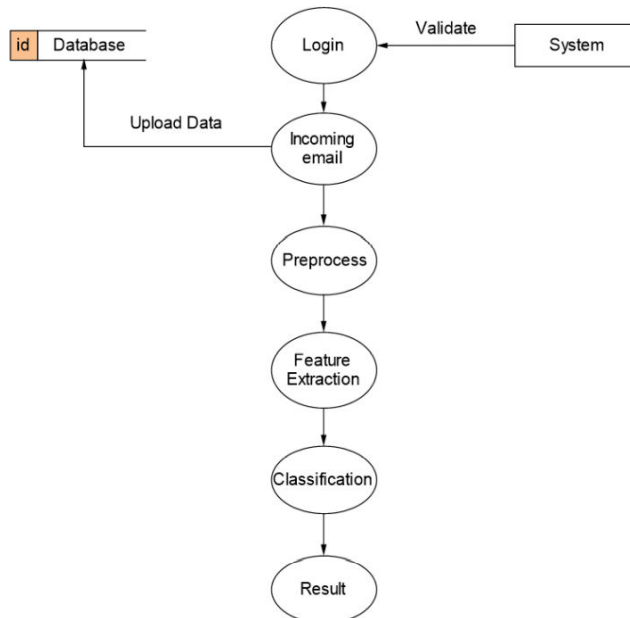


Fig 4: DFD Level 1

VII. ADVANTAGES

1. A typical office worker receives 150 emails every day, with an estimated 50 of those being spam. However, even with 60 emails a day, it is simple to miss important communications due to the volume of emails that are arriving
2. Every day, malicious emails with viruses, malware, and other threats arrive in recipients' email inboxes. The spam filters provided by your internet service provider may remove some of them with ease, but spam is becoming increasingly intelligent.

VIII. APPLICATION

1. Online Web application.
2. Spam Detection applications
3. Mail applications
4. Good Efficiency.
5. Greater Accuracy.

IX. COMPARISON

Automatic email filtering may be the most effective method of detecting spam but nowadays spammers can easily bypass all these spam filtering applications easily.



XI. RESULT

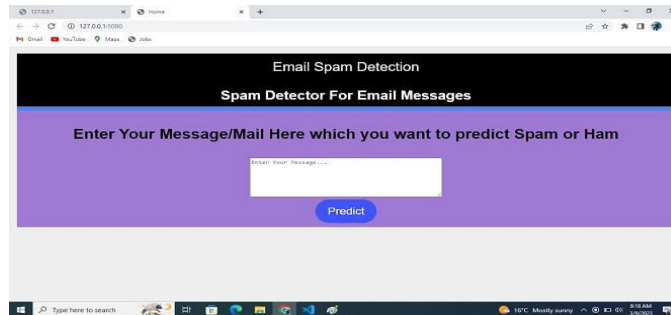


Fig 5: Home Page

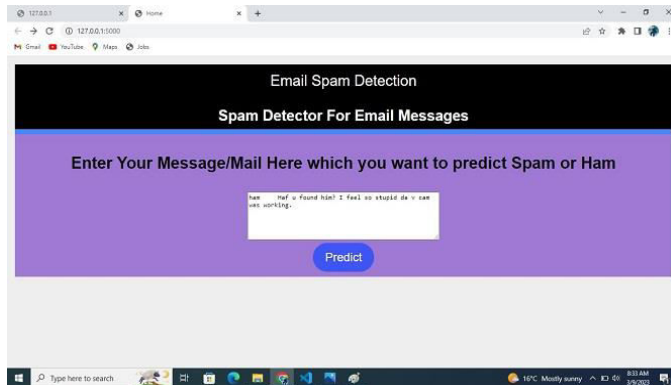


Fig 6: Message detect SPAM or HAM

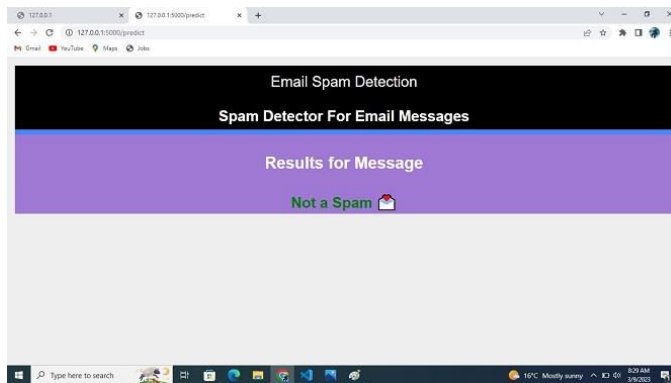


Fig 7: Result Page

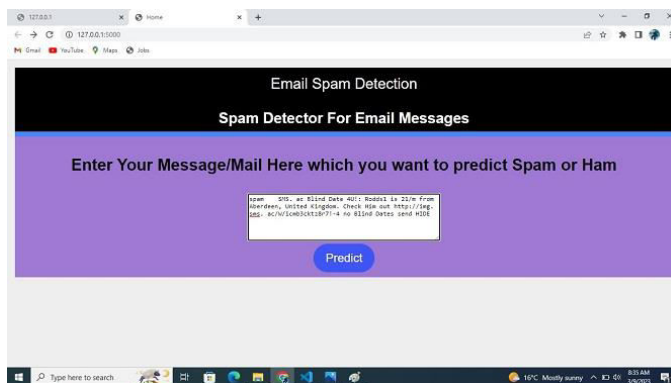


Fig 8: Message detect SPAM or HAM

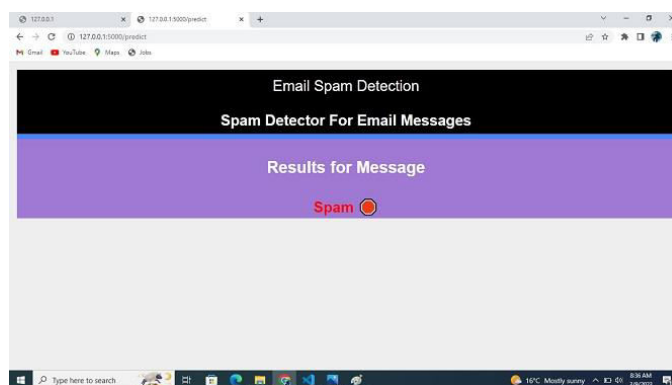


Fig 9: Result Page

XI. CONCLUSION

In order to reduce dimensions by removing redundant characteristics, FS is one of the crucial procedures in pre-processing information due to the abundance of information available. Accuracy of the classifier while choosing the crucial features and eliminating the redundant ones. As a result, FS directly affects the precision and quickness of machine learning classifiers. Recent meta-heuristic algorithms have caught the interest of numerous researchers due to their randomness and simplicity, as well as the fact that they are successful and promising FS techniques. One of the most effective meta-heuristic algorithms is called SOS, and it draws its inspiration from how organisms in nature compete with one another.

REFERENCES

1. MajdiMafarja, Ibrahim Aljarah, HossamFaris, Abdelaziz I. Hammouri, Ala' M. Al-Zoubi, SeyedaliMirjalili, Binary Grasshopper Optimisation Algorithm Approaches for Feature Selection Problems, Expert Systems With Applications (2018)
2. FarhadSoleimanian Gharehchopogh1,*, HojjatGholizadeh, A Comprehensive Survey: Whale Optimization Algorithm and its Applications
3. Maryam Samadi Bonab1 Ali Ghaffari,FarhadSoleimanianGharehchopogh,PayamAlemi,A wrapper-based feature selection for improving performance of intrusion detection systems
4. Sasan Gharehpasha1 · Mohammad Masdari1 · Ahmad Jafarian, Virtual machine placement in cloud data centers using a hybrid multi-verse optimization algorithm.
5. Mehdi Hosseinzadeh,Hawkar Kamaran Hama, · Marwan Yassin Ghafour,Mohammad Masdari, Omed Hassan Ahmed,Hemn Khezri, Service Selection Using Multi-criteria Decision Making: A Comprehensive Overview
6. SankalapAroraa, PriyankaAnand,Binary butterfly optimization approaches for feature selection,September 1, 2018
7. Md. MonirulKabir a , Md. Shahjahan b , Kazuyuki Murase, A new hybrid ant colony optimization algorithm for feature selection,2011
8. HossamFaris, Majdi M. Mafarja, Ali AsgharHeidari, Ibrahim Aljarah, Ala' M. Al-Zoubi, SeyedaliMirjalili, Hamido Fujita, An Efficient Binary Salp Swarm Algorithm with Crossover Scheme for Feature Selection Problems, Knowledge-Based Systems (2018)
9. Nailah Al-Madi,Hossam Faris,Seyedali Mirjalili.Binary multi-verse optimization algorithm for global optimization and discrete problems(2019)
10. Yongquan Zhou , Fahui Miao a , QifangLuo, Symbiotic organisms search algorithm for optimal evolutionary controller tuning of fractional fuzzy controllers



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 **6381 907 438** **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details