

[Review Form 2](#)

Book Name:	Mathematics and Computer Science: Contemporary Developments
Manuscript Number:	Ms_BPR_3819
Title of the Manuscript:	Logistic regression and its implementation for email spam filtering
Type of the Article	Book chapter

PART 1: Review Comments

Compulsory REVISION comments	Reviewer's comment	Author's Feedback <i>(Please correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)</i>
Please write a few sentences regarding the importance of this manuscript for the scientific community. Why do you like (or dislike) this manuscript? A minimum of 3-4 sentences may be required for this part.	This manuscript is valuable for the scientific community as it states the practical application of logistic regression in email spam filtering, emphasizing the critical role of data cleaning and token frequency analysis in improving model performance. By using the UCI dataset and evaluating the model with the ROC curve, it provides a reproducible framework for researchers, though it could benefit from comparative insights with other algorithms to broaden its applicability.	
Is the title of the article suitable? (If not please suggest an alternative title)	The title "Logistic Regression and its Implementation for Email Spam Filtering" is on the overview, suitable as it clearly indicates the primary focus of the article: the application of logistic regression in filtering email spam. However, it could be improved to make it more specific and engaging by reflecting the unique aspects of the study, such as data preprocessing, token frequency distribution, and performance evaluation. The current title lacks emphasis on the experiment's context, methodology, or findings, which could make it less compelling for potential readers. Suggested Alternative Titles: "Enhancing Email Spam Detection: Logistic Regression and the Role of Data Cleaning" This title emphasizes both the use of logistic regression and the importance of data preprocessing, which is a key focus of the study.	
Is the abstract of the article comprehensive? Do you suggest the addition (or deletion) of some points in this section? Please write your suggestions here.	The provided abstract gives a broad overview of the topic, focusing on the experiment with logistic regression for spam filtering, the use of the UCI dataset, and the importance of data cleaning. While it provides some context and mentions key aspects such as token frequency distribution and ROC curve evaluation, it lacks sufficient depth and specificity to fully reflect the scope and objectives implied by the title, "Logistic Regression and its Implementation for Email Spam Filtering." The abstract does not outline the implementation process of logistic regression, nor does it elaborate on the practical or theoretical insights gained from the study. The omission of details regarding model performance, comparative analysis with other techniques, and specific challenges addressed by the implementation leaves the abstract incomplete. Another deficiency lies in the lack of a clear structure. The abstract does not highlight why logistic regression was chosen over other algorithms or how its implementation contributes to advancing email spam filtering techniques. There is also no mention of metrics used beyond the ROC curve, such as accuracy, precision, or recall, which are critical for evaluating a spam filter's performance. Additionally, the abstract does not clearly convey how the findings could be applied in real-world scenarios or their implications for improving spam detection systems. Overall, it lacks a comprehensive summary of the study's objectives, methods, results, and significance, which are essential for providing readers with a clear and compelling overview of the research.	
Are subsections and structure of the manuscript appropriate?	The subsequent subsections and structure of the manuscript is good but could be restructured to conform to the known standard of Introduction Literature review, Methods Results and findings	

[Review Form 2](#)

	Recommendations and future directions conclusion	
Please write a few sentences regarding the scientific correctness of this manuscript. Why do you think that this manuscript is scientifically robust and technically sound? A minimum of 3-4 sentences may be required for this part.	The manuscript is scientifically robust as it leverages a well-established method, logistic regression, for a critical application—email spam filtering. It provides a systematic approach by incorporating detailed steps for tokenization, data preprocessing, model training, and evaluation using established datasets like UCI and Enron, ensuring methodological transparency. The use of confusion matrices, ROC curve, and related metrics for model performance validation reflects a sound statistical framework that aligns with best practices in machine learning applications. However, there are areas for improvement to enhance scientific rigor. While the manuscript discusses the importance of token features and evaluates model performance, it lacks comparative analysis with other classifiers like Naive Bayes or Support Vector Machines, which are commonly used for spam filtering. Additionally, the static nature of the logistic regression model and its inability to dynamically adapt to evolving spam patterns could be highlighted more critically, and potential improvements, such as updating features dynamically, could be explored	
Are the references sufficient and recent? If you have suggestions of additional references, please mention them in the review form. :	The references included in the manuscript are relevant but could benefit from updates to include more recent studies in spam filtering and advancements in machine learning techniques. While logistic regression is a foundational model, recent work in deep learning, ensemble methods, and hybrid models for email filtering could provide a broader context for comparison and potential improvement. Additionally, the manuscript might include references discussing modern datasets beyond UCI and Enron to highlight the applicability of the approach to contemporary and diverse datasets.	
<u>Minor REVISION</u> comments	Yes	
Is the language/English quality of the article suitable for scholarly communications?		
<u>Optional/General</u> comments		

PART 2:

	<u>Reviewer's comment</u>	<u>Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)</u>
<u>Are there ethical issues in this manuscript?</u>	<u>(If yes, Kindly please write down the ethical issues here in details)</u>	

Reviewer Details:

Name:	Joseph Chukwunweike
Department, University & Country	United Kingdom