Task ‘Implementing and Evaluating a Simple AI Model’ Unit 6

Introduction

Email spam is a widespread issue that results in wasted time, security risks and decreased productivity. Machine learning can help by automatically identifying and filtering spam emails using large datasets. This project uses a random forest classifier based on the Spambase dataset provided by the UCI Machine Learning Repository. The aim is to improve the effectiveness of email filtering systems.

The problem the AI model solves

This machine learning model detects email spam with the aim of predicting as accurately as possible email legitimacy as spam with (1) representing spam and (0) representing non-spam, the following approach is taken: numerical features extracted from email content. This problem is critical in real-world applications to filter out unwanted or malicious emails. A random forest classifier was used for this task, and the results and model performance metrics are produced from the Spambase dataset loaded from the UCI Machine Learning Repository URL.

The results and model performance metrics

This task was carried out using A random forest classifier is known for its robustness. and resistance to overfitting. This was done after training and testing. The model demonstrated the following strong evaluation metrics results:

- Accuracy: 0.9566
- Precision: 0.9675
- Recall: 0.9272

These values indicate that this model proves to be highly effective at identifying spam emails with minimal false positives and false negatives. The F1 score, available in the classification report "Output the results", further supports the model's balance between precision and recall.

(Oyelakin, Salau, Ogidan, Olufadi, Yusuf, and Adeinji, 2023)

Ethical considerations for the dataset and AI model use case

However, there are ethical considerations to consider. The Spambase dataset is a widely used resource for academic research, but it is from 1999 and may not reflect modern spam strategies. This could potentially introduce bias in newer contexts. Furthermore, the features are created using pre-processed email text, which may mask any linguistic or cultural biases. Any deployment of spam filters must therefore

ensure transparency and fairness, particularly when integrated into systems that affect users' communications. Over blocking legitimate content (false positives) is a risk, and this can have a negative impact on user trust and productivity. Therefore, it is vital to periodically refresh their training with up-to-date data to ensure fairness and accuracy in real-world applications.

(Hopkins, Reeber, Forman, and Suermondt, 1999)

Code for the Spambase implementation ‘Spam Classification on UCI Spambase.py’

Output the results

PS C:\Users\AmnonMalka\Documents\Code\Archive\Unit 6 - Archive> & C:/Users/AmnonMalka/AppData/Local/Programs/Python/Python313/python.exe "c:/Users/AmnonMalka/Documents/Code/Unit 6/Spam Classification on UCI Spambase.py"

Evaluation Metrics	
Accuracy	0.9566
Precision	0.9675

Recall	0.9272

Classification Report				
	precision	**recall**	**f1-score**	**support**
0	0.95	0.98	0.96	804
1	0.97	0.93	0.95	577
accuracy			0.96	1381
macro avg	0.96	0.95	0.96	1381
weighted avg	0.96	0.96	0.96	1381

References

Hopkins, M., Reeber, E., Forman, G. and Suermondt, J., (1999). *Spambase Data Set. UCI Machine Learning Repository*. Available at: https://archive.ics.uci.edu/ml/datasets/spambase [Accessed 29 August. 2025]

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, pp.2825–2830. Available at: https://scikit-learn.org/ [Accessed 29 August. 2025]

Oyelakin, Salau, Ogidan, Olufadi, Yusuf, and Adeinji, (2023) Spam email detection scheme based on random forest algorithm. LAUTECH JOURNAL OF COMPUTING AND INFORMATICS, 3(1), pp.87-97.

The pandas development team, (2023). *pandas-dev/pandas*: Pandas Zenodo (version 2.3.2) Available at: https://pandas.pydata.org/ [Accessed 29 August. 2025]

Scikit-learn, 2024. Scikit-learn: Machine learning in Python [software]. Available at: https://scikit-learn.org/ [Accessed 29 August. 2025]

Testas, A., 2023. *Distributed Machine Learning with PySpark: Migrating Effortlessly from Pandas and Scikit-Learn*. Apress. https://doi.org/10.1007/978-1-4842-9751-3 [Accessed 29 August. 2025]

This document has been written solely for educational purposes. All references, names, and trademarks mentioned here remain the property of their respective owners and are used here strictly for the educational context. Grammarly was used exclusively for proofreading and enhancing the clarity and language of the text. ChatGPT was consulted for general research. All academic writing, analysis, argumentation, and conclusions are entirely the original work of the author.