

CAPSTONE PROJECT

# **PROJECT TITLE**

## AI-Based Spam Email Detection System using Microsoft Azure Machine Learning

**PRESENTED BY**

**STUDENT NAME: MOHAMMAD SOHAIL  
AKHTAR**

**COLLEGE NAME: PATLIPUTRA  
UNIVERSITY**

**DEPARTMENT: BSC ZOOLOGY**

**EMAIL ID:  
SUFHAILANSAR@GMAIL.COM**



# OUTLINE:

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach**
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

# PROBLEM STATEMENT:

Email users receive many unwanted spam emails daily. These emails can contain malicious links, fraud messages, and promotional content. Manually identifying spam emails is inefficient and time-consuming.

# PROPOSED SOLUTION:

- The proposed system uses Artificial Intelligence and Machine Learning to automatically classify emails as Spam or Not Spam. The system is trained using labeled email data and deployed on Microsoft Azure Machine Learning for real-time prediction.:
  - Data Collection:
    - Collected a dataset of emails labeled as Spam or Not Spam.
    - Ensured diversity by including emails from different sources and categories.
  - Data Preprocessing:
    - Cleaned and preprocessed email text to remove noise, symbols, and irrelevant content.
    - Converted text into a suitable format for machine learning (feature extraction).
  - Machine Learning Algorithm:
    - Trained a supervised learning algorithm (e.g., Logistic Regression) for binary classification.
    - The model learns patterns in email text to accurately predict Spam or Not Spam.
  - Deployment:
    - Deployed the trained model as a real-time endpoint on Microsoft Azure Machine Learning.
    - Users can send email text to the endpoint and get instant predictions.
  - Evaluation:
    - Evaluated model performance using metrics like accuracy and confusion matrix.
    - Fine-tuned the model for better prediction and reduced false positives.
    - Result: AI model accurately classifies emails as Spam or Not Spam.

# SYSTEM APPROACH:

## SYSTEM DEVELOPMENT APPROACH (TECHNOLOGY USED)

### Microsoft Azure Portal:

Used Azure portal to manage and monitor machine learning resources  
Enabled cloud-based storage, computation, and model deployment.

Enabled cloud-based storage, computation, and model deployment.

### Azure Machine Learning Studio :

Designed and trained the ML model using the drag-and-drop interface.

Facilitated model experimentation, versioning, and testing easily.

### Dataset (CSV File):

Uploaded labeled email dataset (Spam/Not Spam) for model training.

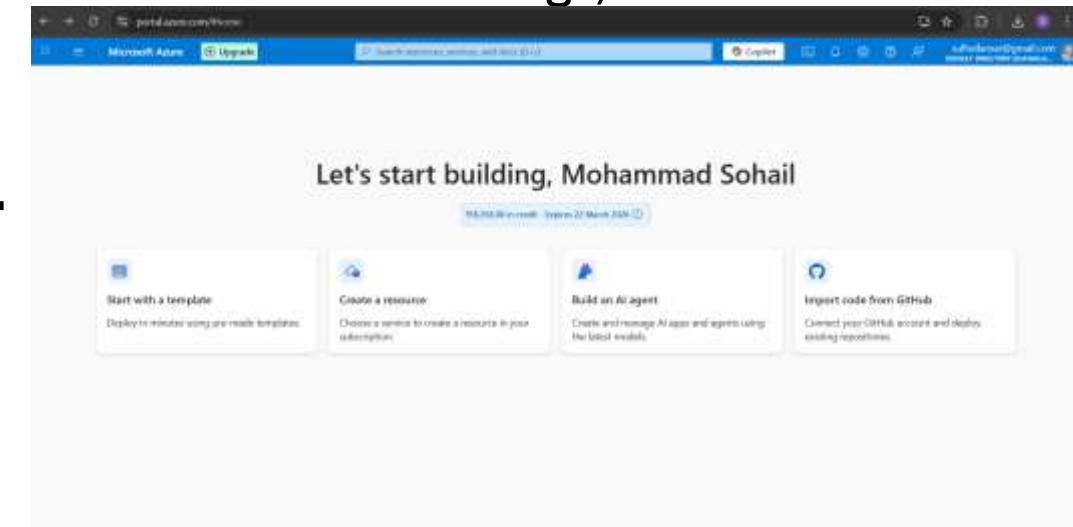
Ensured dataset is clean and ready for preprocessing.

### Two-Class Classification Algorithm:

Used supervised algorithm (e.g., Logistic Regression) for binary classification.

Model predicts email as Spam or Not Spam based on text patterns.

**Cloud Deployment:** Deployed trained model as a real-time endpoint in Azure.



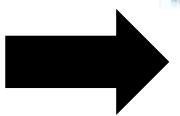
# ALGORITHM & DEPLOYMENT:

- The project uses a Supervised Machine Learning algorithm such as Logistic Regression for binary classification. The model analyzes text patterns and predicts whether an email is Spam or Not Spam:
- **Algorithm Selection:**
  - Selected a supervised learning algorithm (Logistic Regression) for binary classification. Chosen due to its efficiency, simplicity, and good performance in text classification tasks.
- **Data Input:**
  - Input data consists of email text in CSV format with Spam/Not Spam labels. Text data is converted into numerical features using text vectorization techniques.
- **Training Process:**
  - The dataset is split into training and testing sets. The algorithm learns patterns from labeled data to classify emails accurately.
- **Prediction Process:**
  - New email text is sent to the trained model endpoint. The model analyzes patterns and predicts whether the email is Spam or Not Spam.

The screenshot shows the Azure Machine Learning studio interface. At the top, there's a navigation bar with links like 'Home', 'Compute', 'Datasets', 'Experiments', 'Notebooks', 'ML Flow', 'Metrics', 'Logs', and 'Metrics'. Below the navigation bar, the main title is 'Azure Machine Learning' with a subtitle 'Create a machine learning workspace'. A green status bar at the top says 'Validation passed'. The main content area contains several configuration sections with dropdown menus and input fields:

- Subscription:** Azure subscription 1
- Resource group:** Default (westus2, RG)
- Region:** Central India
- Name:** Sample-Workshop1
- Storage account:** (read) spawneverlast9431409701
- Key vault:** (read) spawneverlast407396691
- Application insights:** (read) spawneverlast218517638
- Container registry:** None
- Connectivity journal:** Enable public access from all networks
- Network isolation:** Public
- Encryption type:** Microsoft-managed keys
- Identity type:** System assigned
- Enable HDFS:** Disabled
- Storage account access type:** Confidential-based
- Shared key storage:** Unlisted

At the bottom, there are 'Create' and 'Next >' buttons, along with a link to 'Download a template for automation'.



The screenshot shows the Microsoft Azure Machine Learning Services Overview page. At the top, there's a navigation bar with links like Home, Microsoft Azure, Upgrade, and Help & support. Below the navigation is a search bar and a toolbar with icons for Delete, Create, Refresh, Download, and Copy. The main content area has a title "Your deployment is complete" with a green checkmark icon. It displays deployment details: Deployment name - Microsoft.MachineLearningServices, Subscription - Azure subscription 1, Resource group - SpaceMLO-RST. To the right, there are sections for Cost management, Microsoft Defender for Cloud, and Work with an expert. A sidebar on the left lists Overview, Inputs, Outputs, and Templates.



Microsoft Foundry | Azure Machine Learning

# SpamMLWorkspace1

Generative AI with Prompt flow

Multi-Round Q&A on Your Data

Q&A on Your Data

Web Classification

Chat with Wikipedia

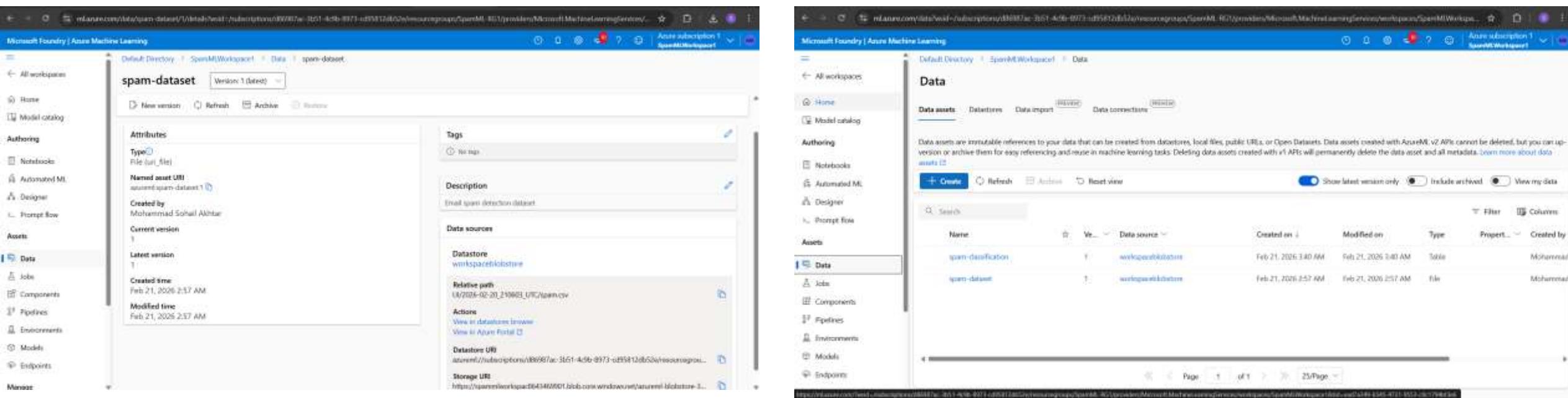
Generative AI models

Notebook samples



# RESULT:

The AI model successfully classifies emails as Spam or Not Spam with good accuracy. The system reduces manual effort and improves email security.



The image shows two side-by-side screenshots of the Microsoft Foundry Azure Machine Learning interface. Both screenshots are titled "Microsoft Foundry | Azure Machine Learning" and show the "Data" workspace.

**Left Screenshot (spam-dataset):**

- Attributes:**
  - Type: File (uri\_file)
  - Named asset URI: [spam-dataset](#)
  - Created by: Mohammad Sohal Akbar
- Assets:**
  - Current version: 1
  - Latest version: 1
  - Created time: Feb 21, 2026 2:57 AM
  - Modified time: Feb 21, 2026 2:57 AM

**Right Screenshot (spam-classification):**

- Data assets:**

Name	Type	Created on	Modified on	Type	Properties	Created by
spam-classification	File	Feb 21, 2026 3:40 AM	Feb 21, 2026 3:40 AM	Table		Mohammad
spam-dataset	File	Feb 21, 2026 3:57 AM	Feb 21, 2026 3:57 AM	file		Mohammad

# CONCLUSION:

This project demonstrates how Artificial Intelligence and Machine Learning can be effectively used to detect spam emails using Microsoft Azure Machine Learning. By training the model on labeled email data, the system is able to accurately classify emails as Spam or Not Spam. The cloud-based deployment ensures real-time prediction and scalability. Overall, the solution reduces manual effort, enhances email security, and showcases the practical implementation of ML in solving real-world problems.

# FUTURE SCOPE:

In the future, this system can be integrated into real-time email platforms such as Gmail and Outlook to provide automatic spam filtering for users. The model can be further improved by using advanced techniques like Natural Language Processing (NLP) and Deep Learning algorithms to increase accuracy and reduce false predictions. Additionally, the system can be trained with larger and more diverse datasets to make it more robust, scalable, and efficient for large-scale applications.

# REFERENCES:

This project was developed using resources and documentation from Microsoft Azure Machine Learning and Microsoft Learn for understanding cloud-based model training and deployment. The spam email dataset was obtained from Kaggle, and machine learning concepts such as Logistic Regression were studied using official Scikit-learn documentation.

GitHub Link: <https://github.com/sufhailansar-bit/spam-email-classification-azure>

# Thank You