Microsoft Elevate

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.

### 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.

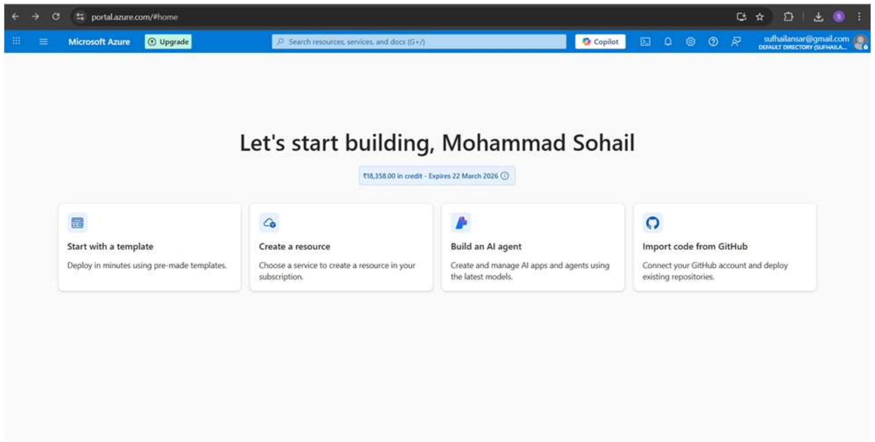
Ensured dataset is clean and ready for preprocessing.

### Two-Class Classification Algorithm:

Used supervised algorithm 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.



Let's start building, Mohammad Sohail

₹18,358.00 in credit - Expires 22 March 2026

Start with a template
Deploy in minutes using pre-made templates.

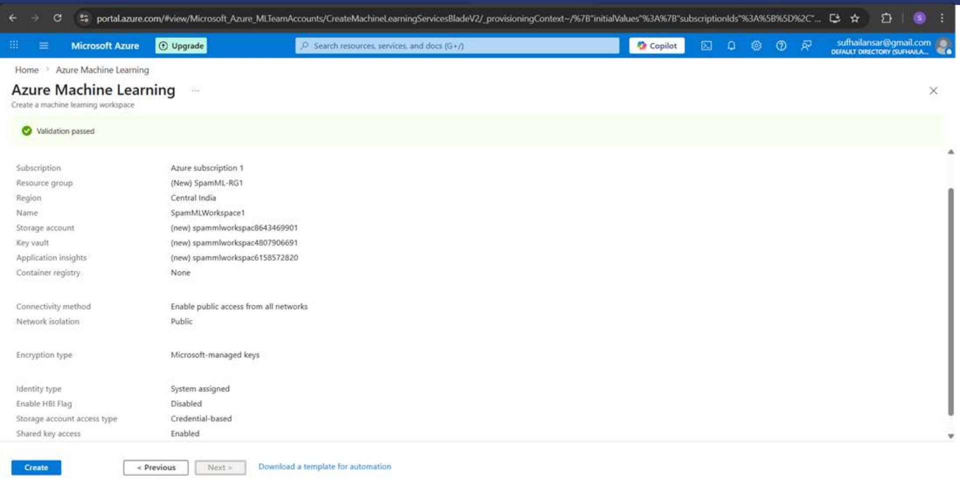
Create a resource
Choose a service to create a resource in your subscription.

Build an AI agent
Create and manage AI apps and agents using the latest models.

Import code from GitHub
Connect your GitHub account and deploy existing repositories.

# 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.


portal.azure.com/#view/Microsoft_Azure_MLTeamAccounts/CreateMachineLearningServicesBladeV2/_provisioningContext~/%7B"initialValues"%3A%7B"subscriptionIds"%3A%5B%5D%2C"...
Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
sufhailansar@gmail.com
DEFAULT DIRECTORY (SUFHAILA...
Home › Azure Machine Learning

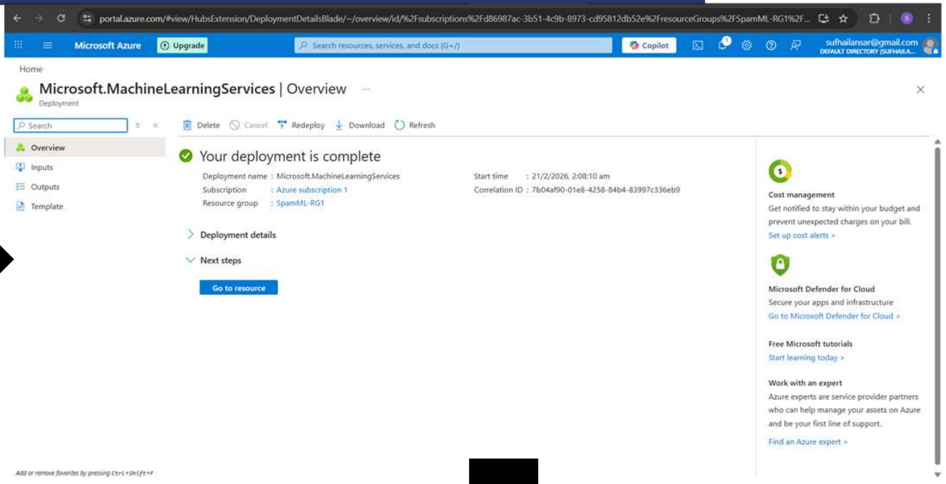
Azure Machine Learning

Create a machine learning workspace

Validation passed

| | |
|---|---|
| Subscription | Azure subscription 1 |
| Resource group | (New) SpamML-RG1 |
| Region | Central India |
| Name | SpamMLWorkspace1 |
| Storage account | (new) spammlworkspac8643469901 |
| Key vault | (new) spammlworkspac4807906691 |
| Application insights | (new) spammlworkspac6158572820 |
| Container registry | None |
| Connectivity method | Enable public access from all networks |
| Network isolation | Public |
| Encryption type | Microsoft-managed keys |
| Identity type | System assigned |
| Enable HBI Flag | Disabled |
| Storage account access type | Credential-based |
| Shared key access | Enabled |

Create
< Previous
Next >
Download a template for automation


portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2Fd86987ac-3b51-4c9b-8973-cd95812db52e%2FresourceGroups%2FSpamML-RG1%2F...
Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
sufhailansar@gmail.com
DEFAULT DIRECTORY (SUFHAILA...
Home
Search
Delete
Cancel
Redeploy
Download
Refresh
Overview
Inputs
Outputs
Template

Microsoft.MachineLearningServices | Overview

Deployment

Your deployment is complete

Deployment name : Microsoft.MachineLearningServices
Subscription : Azure subscription 1
Resource group : SpamML-RG1
Start time : 21/2/2026, 2:08:10 am
Correlation ID : 7b04af90-01e8-4258-84b4-83997c336eb9

Deployment details

Next steps

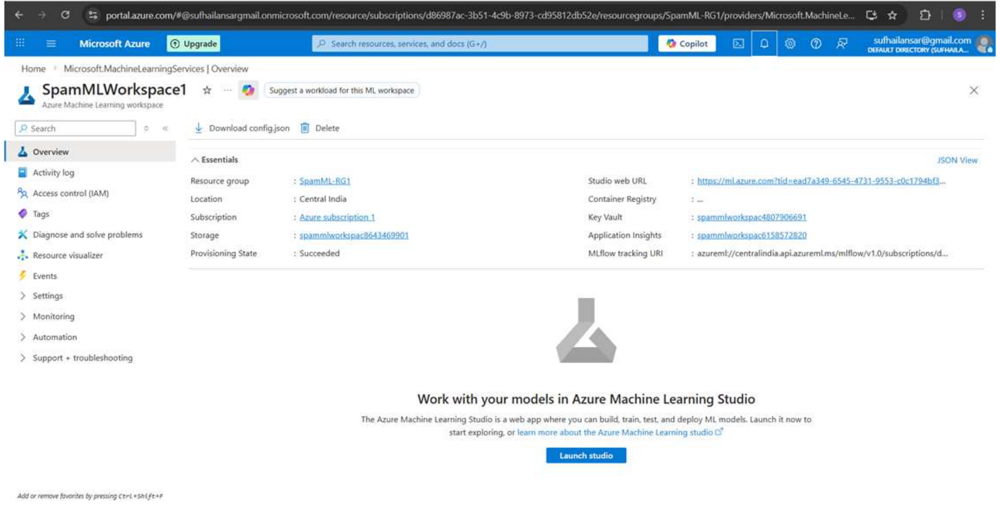
Cost management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >

Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >

Free Microsoft tutorials
Start learning today >

Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
Find an Azure expert >

*Add or remove favorites by pressing Ctrl+Shift+F*


portal.azure.com/#@sufhailansargmail.onmicrosoft.com/resource/subscriptions/d86987ac-3b51-4c9b-8973-cd95812db52e/resourcegroups/SpamML-RG1/providers/Microsoft.MachineLe...
Microsoft Azure
Upgrade
Search resources, services, and docs (G+/)
Copilot
sufhailansar@gmail.com
DEFAULT DIRECTORY (SUFHAILA...
Home › Microsoft.MachineLearningServices | Overview
Suggest a workload for this ML workspace
Search
Download config.json
Delete
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Resource visualizer
Events
Settings
Monitoring
Automation
Support + troubleshooting
JSON View

SpamMLWorkspace1

Azure Machine Learning workspace

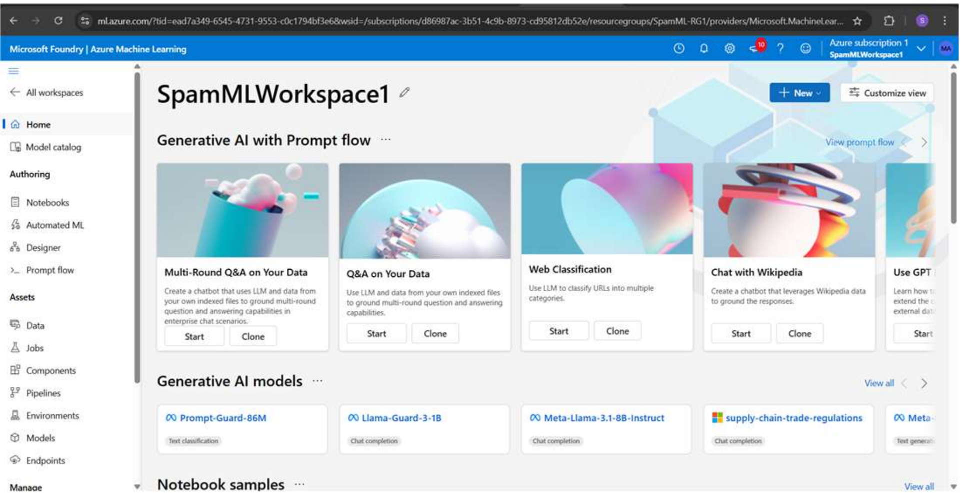
Essentials

Resource group : SpamML-RG1
Location : Central India
Subscription : Azure subscription 1
Storage : spammlworkspac8643469901
Provisioning State : Succeeded
Studio web URL : https://ml.azure.com?tid=ead7a349-6545-4731-9553-c0c1794bf3...
Container Registry : ...
Key Vault : spammlworkspac4807906691
Application Insights : spammlworkspac6158572820
MLflow tracking URI : azureml://centralindia.api.azureml.ms/mlflow/v1.0/subscriptions/d...

Work with your models in Azure Machine Learning Studio

The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning studio

*Add or remove favorites by pressing Ctrl+Shift+F*



SpamMLWorkspace1

Generative AI with Prompt flow

Multi-Round Q&A on Your Data
Create a chatbot that uses LLM and data from your own indexed files to ground multi-round question and answering capabilities in enterprise chat scenarios.

Q&A on Your Data
Use LLM and data from your own indexed files to ground multi-round question and answering capabilities.

Web Classification
Use LLM to classify URLs into multiple categories.

Chat with Wikipedia
Create a chatbot that leverages Wikipedia data to ground the responses.

Use GPT
Learn how t extend the external dat

Generative AI models

Prompt-Guard-86M
Text classification

Llama-Guard-3-1B
Chat completion

Meta-Llama-3.1-8B-Instruct
Chat completion

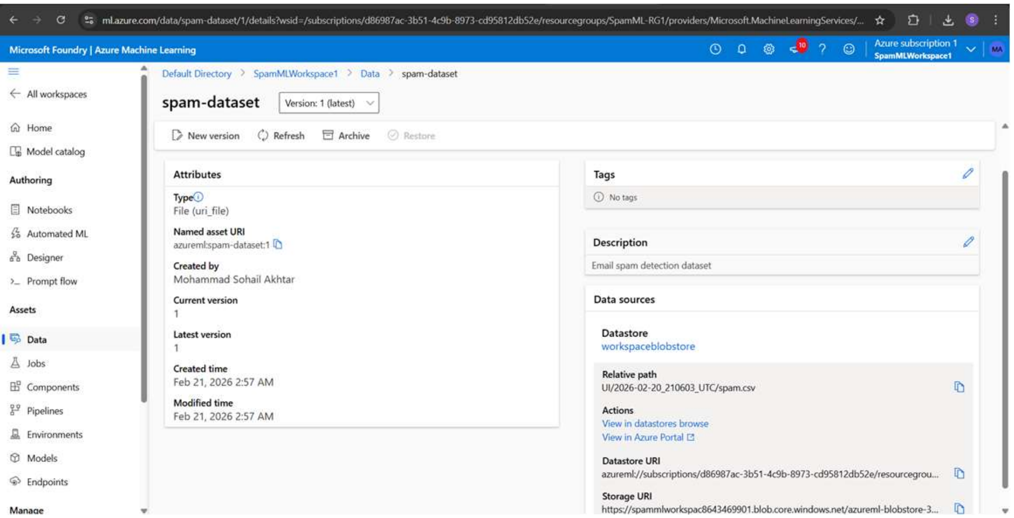
supply-chain-trade-regulations
Chat completion

Meta-
Text generati

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 model achieved high accuracy with low false positives.



Default Directory > SpamMLWorkspace1 > Data > spam-dataset

spam-dataset — Version: 1 (latest)

**Attributes**

Type
File (uri_file)

Named asset URI
azureml:spam-dataset:1

Created by
Mohammad Sohail Akhtar

Current version
1

Latest version
1

Created time
Feb 21, 2026 2:57 AM

Modified time
Feb 21, 2026 2:57 AM

**Tags**

No tags

**Description**

Email spam detection dataset

**Data sources**

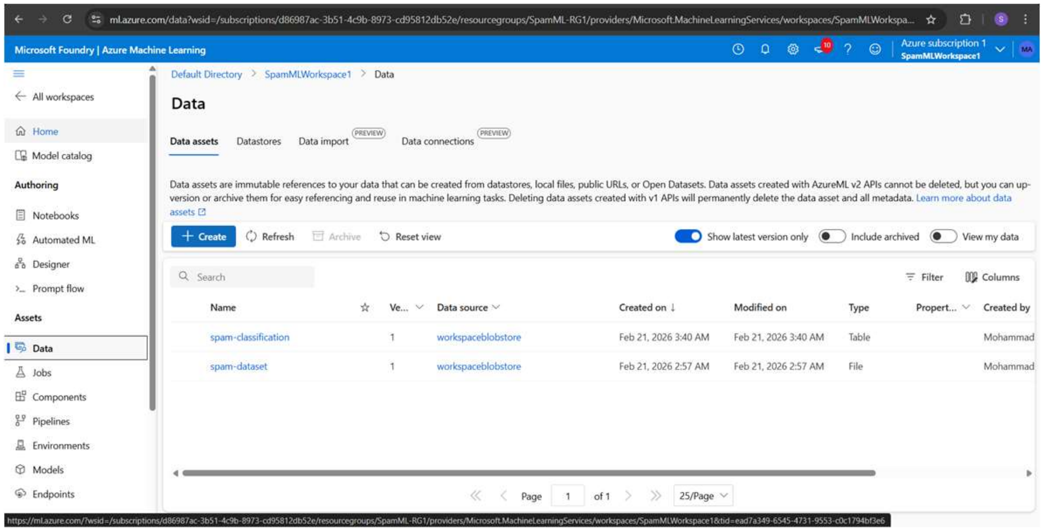
Datastore
workspaceblobstore

Relative path
UI/2026-02-20_210603_UTC/spam.csv

Actions
View in datastores browse
View in Azure Portal

Datastore URI
azureml://subscriptions/d86987ac-3b51-4c9b-8973-cd95812db52e/resourcegrou...

Storage URI
https://spammlworkspac8643469901.blob.core.windows.net/azureml-blobstore-3...



Default Directory > SpamMLWorkspace1 > Data

Data

Data assets are immutable references to your data that can be created from datastores, local files, public URLs, or Open Datasets. Data assets created with AzureML v2 APIs cannot be deleted, but you can up-version or archive them for easy referencing and reuse in machine learning tasks. Deleting data assets created with v1 APIs will permanently delete the data asset and all metadata. Learn more about data assets

● Show latest version only ○ Include archived ○ View my data

| Name | Ve... | Data source | Created on ↓ | Modified on | Type | Propert... | Created by |
|---|---|---|---|---|---|---|---|
| spam-classification | 1 | workspaceblobstore | Feb 21, 2026 3:40 AM | Feb 21, 2026 3:40 AM | Table |  | Mohammad |
| spam-dataset | 1 | workspaceblobstore | Feb 21, 2026 2:57 AM | Feb 21, 2026 2:57 AM | File |  | Mohammad |

Page 1 of 1 25/Page

https://ml.azure.com/?wsid=/subscriptions/d86987ac-3b51-4c9b-8973-cd95812db52e/resourcegroups/SpamML-RG1/providers/Microsoft.MachineLearningServices/workspaces/SpamMLWorkspace1&tid=ead7a349-6545-4731-9553-c0c1794bf3e6

# 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 Email/SMS Spam Collection Dataset [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

Microsoft Elevate

# Thank You