

# Spam Email Detector in Azure

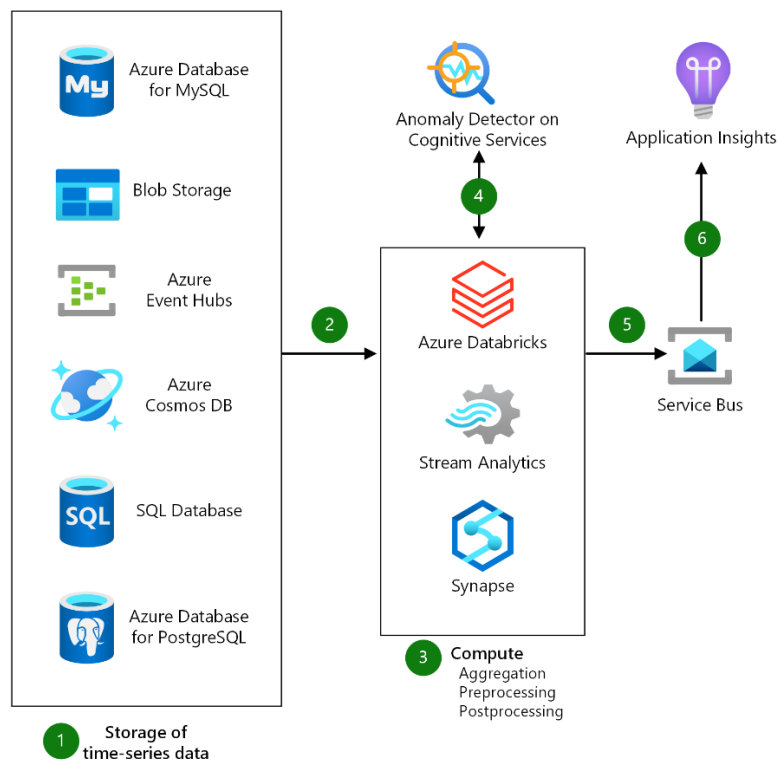
## Introduction

Spam emails can be a significant nuisance for individuals and businesses alike. However, with the advancements in machine learning and cloud computing, we can now leverage powerful tools and platforms to build robust spam detection systems. In this blog, we will explore how to build a spam detector using Azure, Microsoft's cloud computing platform.

## Problem Statement

The objective of this project is to develop a machine learning model that can accurately classify emails as spam or non-spam. By leveraging Azure's services and tools, we will create a scalable and efficient spam detector. This will enable individuals and organizations to filter out unwanted emails and enhance their productivity.

## Architecture



## Prerequisites

Azure account

## Technical Details and Implementation of Solution

### Create Resource Group

Run the following from Azure Cloud Shell

```
resourceGroupName=spam$RANDOM-rg
location=SouthCentralUS

az group create \

  --name $resourceGroupName \

  --location $location
```

### Create Azure Machine Learning Workspace

Run the following in Azure Cloud Shell to create an Azure Machine Learning Workspace

```
workspace=spam-$RANDOM

az extension add -n azure-cli-ml





az ml workspace create -w $workspace -g $resourceGroupName --sku
enterprise
```

### Data Collection

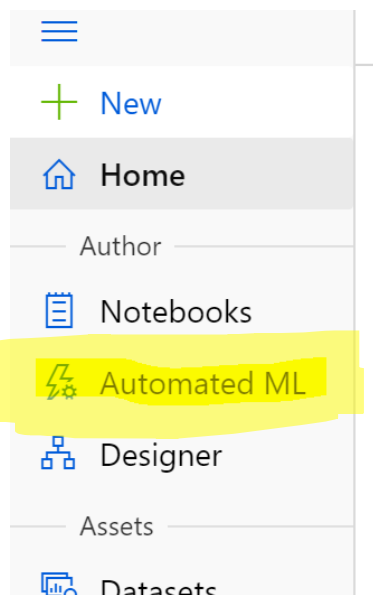
Collect a labelled dataset of spam and non-spam emails.

## Creating new Automated ML Run

Using the Azure Portal, open the Azure Resource Group. Click on your Azure Machine Learning Workspace.

<input type="checkbox"/>	Name ↑↓	Type ↑↓
<input type="checkbox"/>	 spam-17402	Machine Learning
<input type="checkbox"/>	 spam1740insights0f7e9010	Application Insights
<input type="checkbox"/>	 spam1740keyvault4b290e2c	Key vault
<input type="checkbox"/>	 spam1740storage636e41bac	Storage account

Select Automated ML from the left.



Select New Automated ML Run

Select Create dataset, then select from local files.

### Create a new Automated ML run

- Select dataset
- Configure run
- Task type and settings

#### Select dataset

Select a dataset from the list below

+ Create dataset ▾

From local files

From datastore

From web files

From Open Datasets

Select Browse and find the downloaded and unzipped spam csv file.

#### Create dataset from local files

- Basic info
- Datastore and file selection**
- Settings and preview
- Schema
- Confirm details

#### Datastore and file selection

Select a datastore \*

Previously created datastore

workspaceblobstore (Azure Blob Storage) ▾

[Refresh](#)

Create new datastore

Select files for your dataset \*

After dataset creation, these files will be uploaded to your default Blob storage and made available in your workspace. Supported file types include: binary, delimited (i.e. csv, tsv), Excel, Parquet, and plain text.

[Browse](#) 1 files selected. Total size 0.4525 MiB. 0/1 files uploaded

File name	Size (MiB)	Upload %	Status
SMSSpamCollection.csv	0.4525		

< Prev Next >

Upload path

UI Files will be uploaded to '\$(Upload path)/03-19-2020\_081054\_UTC'

Select the dataset created and then click next.

### Create a new Automated ML run

- Select dataset
- Configure run
- Task type and settings

#### Select dataset

Select a dataset from the list below, or create a new dataset. Automated ML

+ Create dataset  |  Show supported datasets only

Dataset name	Dataset type
<input checked="" type="checkbox"/> Spam Data	Tabular

Put the spam column in the target column. Give a meaningful name to the project.

Create a new compute.

### Configure run

Configure the experiment. Select from existing experiments or define a new name, select the target column and the training compute to use. [Learn more on how to configure the experiment](#)

#### Dataset

Spam Data [\(View dataset\)](#)

Experiment name \*

Target column \*

Select training cluster \*

[Create a new compute](#) [Refresh compute](#)

## New Training Cluster

ⓘ Customers should not include personal data or other sensitive information in fields marked with ⓘ because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#) ×

Compute name \* ⓘ ⓘ  
auto-ml-compute

ⓘ Machine Learning Compute is a managed training environment consisting of one or more nodes. [Learn more](#).

Region \* ⓘ  
southcentralus

Virtual Machine size \* ⓘ  
Standard\_DS12\_v2

Virtual Machine priority \* ⓘ  
 Dedicated  Low Priority

Minimum number of nodes \* ⓘ  
0

Maximum number of nodes \* ⓘ  
6

Idle seconds before scale down \* ⓘ  
120

> Advanced settings

Once the compute has been provision and added to the run configuration go to Task Type and Settings and select Classification as the task type.

 **Classification** ✔

To predict one of several categories in the target column. yes/no, blue, red, green.

---

Enable deep learning (preview) ⓘ

### Run 1 ● Starting

 Refresh  Cancel

**Details** | [Data guardrails](#) | [Models](#) | [Logs](#) | [Outputs](#)

#### Run summary

**Task type**  
Classification [View all run settings](#)

**Primary metric**  
Accuracy

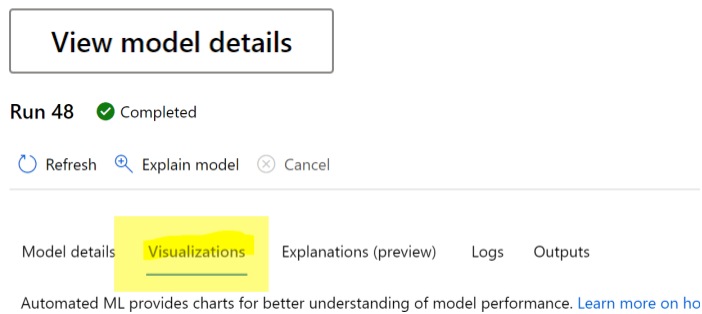
**Run status**  
Starting

**Experiment name**  
spam-experiment

**Run ID**  
AutoML\_517407...

**Input datasets**  
--

Once the run is complete, view the visualizations for the best model by selecting View model details then Visualizations



## Conclusion

In this blog, we have explored the process of building a spam detector in Azure. With this knowledge, a powerful spam detection systems can be developed to protect individuals and organizations from unwanted emails. Leveraging Azure's services, we can create scalable and efficient solutions that enhance productivity and security.

## Challenges in implementing the solution

The biggest challenge for implementing this was applying the basics of cloud computing to make a working ML Model in Azure as even after following 5+ tutorials, due to ever changing technology, many errors were encountered.

## Business Benefit

Azure ML enables businesses to automate and accelerate the machine learning model development process. It reduces time-to-market, improves accuracy, and allows organizations to leverage their data effectively. With AzureML, businesses can make data-driven decisions, optimize processes, and gain a competitive edge in their industry.

## References

<https://stochasticcoder.com/2020/03/24/spam-detection-with-azure-automated-ml/>

<https://www.codeproject.com/Articles/1229597/Detecting-Spam-with-Azure-Machine-Learning>