

Programming Assignment-2

Cloud Computing

Name – Radhika Rayala

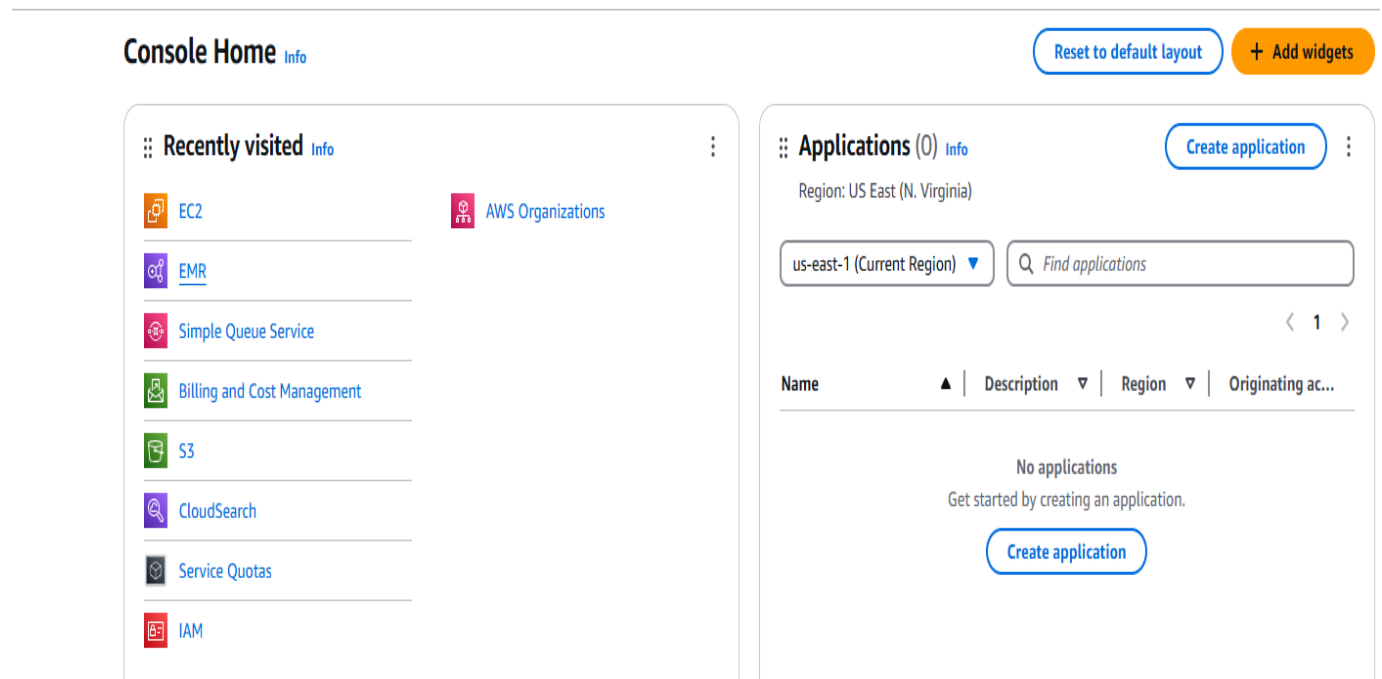
UCID – rr745@njit.edu

GitHub – https://github.com/rr745/wine_quality_predictor

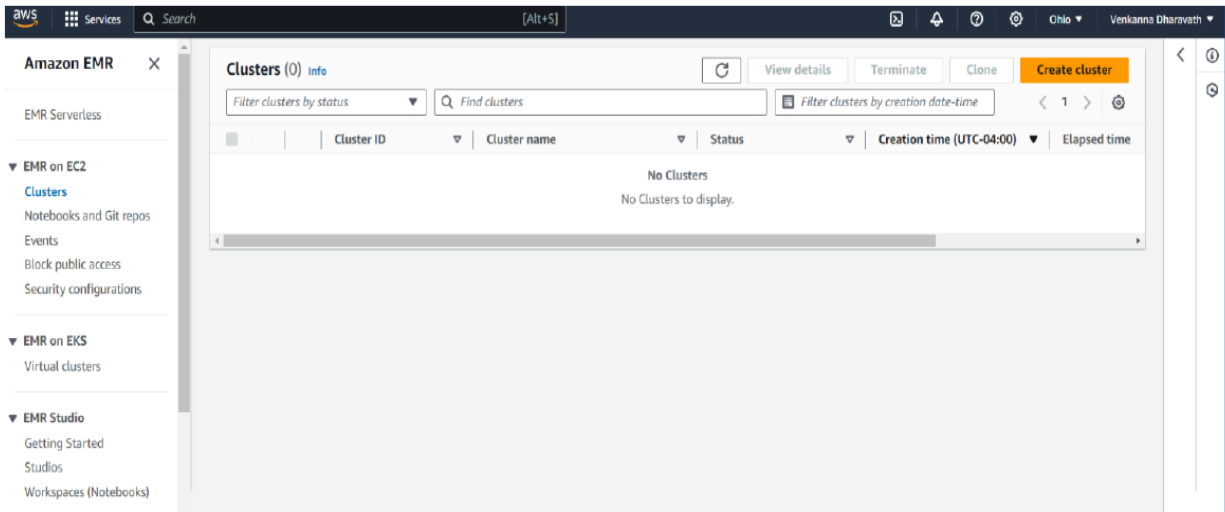
DockerHub - <https://hub.docker.com/repository/docker/rr745/cs643-programming-assignment-2/general>

Log into the AWS Management Console.

The AWS EMR service from the list of available services. Then, choose EMR on EC2 clusters.



On the Clusters page, you'll see that no clusters are currently in use. To start a new cluster, click the "Create Cluster" button.



You can name the cluster whatever you prefer.

Be sure to select **"Spark Interactive"** from the Application Bundle options and choose the latest available version of EMR.

The cluster is named wine_prediction_1112 with Amazon EMR release emr-7.5.0. **Spark Interactive** is selected as the application bundle, including Spark 3.5.2, Livy 0.8.0, and Jupyter Enterprise Gateway 2.6.0. Additional components like Hadoop and Hive are optional and can be enabled if needed. Review the configuration summary and click **Create Cluster** to finalize the setup.

Clone "My cluster" [Info](#)

▼ Name and applications - *required* [Info](#)

Name your cluster and choose the applications that you want to install on your cluster.

Name

Amazon EMR release [Info](#)

A release contains a set of applications which can be installed on your cluster.

Application bundle

Spark Interactive 	Core Hadoop 	Flink 	HBase 	Presto 	Trino 	Custom 
---	---	--	--	---	--	---

☐ AmazonCloudWatchAgent
1.300032.2

☐ HCatalog 3.1.3

☐ Hue 4.11.0

☒ Livy 0.8.0

☐ Pig 0.17.0

☐ TensorFlow 2.16.1

☐ Zeppelin 0.11.1

☐ Flink 1.19.1

☒ Hadoop 3.4.0

☒ JupyterEnterpriseGateway 2.6.0

☐ Oozie 5.2.1

☐ Presto 0.287

☐ Tez 0.10.2

☐ ZooKeeper 3.9.2

☐ HBase 2.5.10

☒ Hive 3.1.3

☐ JupyterHub 1.5.0

☐ Phoenix 5.2.0

☒ Spark 3.5.2

☐ Trino 446

AWS Glue Data Catalog settings

Use the AWS Glue Data Catalog to provide an external metastore for your application.

☐ Use for Hive table metadata

☐ Use for Spark table metadata

Operating system options [Info](#)

☒ Amazon Linux release

☐ Custom Amazon Machine Image (AMI)

Summary [Info](#)

Name and applications - *required*

Name

wine_prediction_1112

Amazon EMR release

emr-7.5.0

Application bundle

Spark Interactive (Hadoop 3.4.0, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5....)

Cluster configuration - *required*

Uniform instance groups

Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - *required*

Provisioning configuration

Core size: 1 instance

[Cancel](#)

[Clone cluster](#)

Primary

Choose EC2 instance type

m5.xlarge

4 vCore 16 GiB memory

EB5 only storage On-Demand price: -

Lowest Spot price: -

Actions

☐ Use high availability

Launch highly available, more resilient cluster with three primary nodes on On-Demand Instances. This configuration applies for the lifetime of your cluster. [Learn more](#)

► Node configuration - optional

Core

Choose EC2 instance type

m5.xlarge

4 vCore 16 GiB memory

EB5 only storage On-Demand price: -

Lowest Spot price: -

Actions

► Node configuration - optional

Task 1 of 1

Name

Task - 1

Choose EC2 instance type

m5.xlarge

4 vCore 16 GiB memory

EB5 only storage On-Demand price: -

Lowest Spot price: -

Actions

► Node configuration - optional

Remove instance group

Name and applications - required

Name

wine_prediction_1112

Amazon EMR release

emr-7.5.0

Application bundle

Spark Interactive (Hadoop 3.4.0, Hive 3.1.3, JupyterEnterpriseGateway 2.6.0, Livy 0.8.0, Spark 3.5....)

Cluster configuration - required

Uniform instance groups

Primary (m5.xlarge), Core (m5.xlarge), Task (m5.xlarge)

Cluster scaling and provisioning - required

Provisioning configuration

Core size: 1 instance

Cancel

Clone cluster

▼ Cluster scaling and provisioning - required Info

Choose how Amazon EMR should size your cluster.

Choose an option

☒ Set cluster size manually

Use this option if you know your workload patterns in advance.

☐ Use EMR-managed scaling

Monitor key workload metrics so that EMR can optimize the cluster size and resource utilization.

☐ Use custom automatic scaling

To programmatically scale core and task nodes, create custom automatic scaling policies.

Provisioning configuration

Set the size of your core and task instance groups. Amazon EMR attempts to provision this capacity when you launch your cluster.

Name	Instance type	Instance(s) size	Use Spot purchasing option
Core	m5.xlarge	1	<input type="checkbox"/>
Task - 1	m5.xlarge	2	<input type="checkbox"/>

▼ Cluster termination and node replacement [Info](#)

Choose termination settings and protect your cluster from accidental shutdown.

Termination option

- ☒ Manually terminate cluster
- ☐ Automatically terminate cluster after last step ends
- ☐ Automatically terminate cluster after idle time (Recommended)

☒ Use termination protection

Protects your cluster from accidental termination. If on, you must first turn off protection to terminate the cluster. We recommend turning on termination protection for your long running clusters.

Unhealthy node replacement - *new* | [Info](#)

- ☒ Turn on
Amazon EMR gracefully stops processes on unhealthy nodes to minimize data loss and job interruptions. It quickly replaces unhealthy nodes with new EC2 instances to keep your jobs running smoothly.
- ☐ Turn off
Amazon EMR adds unhealthy nodes to a denylist while keeping them in the cluster, allowing you continued access for troubleshooting.

▼ EC2 security groups (firewall)

[i](#) Change notice

We've updated the names of some security groups to use more inclusive language. For example, groups that included terms like "master" and "slave" now use the terms "primary" and "core" instead.

Primary node

EMR-managed security group

EMR will automatically update the selected group.

ElasticMapReduce-Primary
sg-049e6f8a21365ae26

Additional security groups - *optional*

Select up to 4 additional security groups.

Choose additional security groups

Core and task nodes

EMR-managed security group

EMR will automatically update the selected group.

ElasticMapReduce-Core
sg-0246a15d4114d42f8

Additional security groups - *optional*

Select up to 4 additional security groups.

Choose additional security groups

Create a new key pair and provide a custom name for the key pair and select 'pem' as the file format when creating the key pair.

EC2 > Key pairs > Create key pair

Create key pair [Info](#)

Key pair
A key pair, consisting of a private key and a public key, is a set of security credentials that you use to prove your identity when connecting to an instance.

Name

The name can include up to 255 ASCII characters. It can't include leading or trailing spaces.

Key pair type [Info](#)
☒ RSA ☐ ED25519

Private key file format
☒ .pem
For use with OpenSSH
☐ .ppk
For use with PuTTY

Tags - optional
No tags associated with the resource.
[Add new tag](#)
You can add up to 50 more tags.

[Cancel](#) [Create key pair](#)

The key pair named 'rr_key_pair' has been created, and the key has been downloaded and stored on the local system for connecting to the cluster using PuTTY as the SSH.

Successfully created key pair

Key pairs (3)

Info

Find Key Pair by attribute or tag

Actions

Create key pair

<input type="checkbox"/>	Name	Type	Created	Fingerprint	ID
<input type="checkbox"/>	vockey	rsa	2024/11/24 12:26 GMT-5	e1:a6:27:0f:a4:a8:97:45:ed:35:f7:ef:a4:05:...	key-05280c8e5363272a5
<input type="checkbox"/>	rr	rsa	2024/11/28 11:27 GMT-5	69:c2:06:72:d2:58:7f:89:17:6b:76:25:61:8...	key-0f97f405cd853a6c7
<input type="checkbox"/>	rr_key_pair	rsa	2024/11/24 12:23 GMT-5	26:d2:d3:1fe1:1e:40:4da7:c1:e7:de:34:06:...	key-0ed184f282d3290ce

Amazon EMR

EMR on EC2: Clusters

Create cluster

rr_key_pair

Browse

Create key pair

▼

Identity and Access Management (IAM) roles - *required*

Info

Choose or create a service role and instance profile for the EC2 instances in your cluster.

Amazon EMR service role

Info

The service role is an IAM role that Amazon EMR assumes to provision resources and perform service-level actions with other AWS services.

☒

Choose an existing service role

Select a default service role or a custom role with IAM policies attached so that your cluster can interact with other AWS services.

☐

Create a service role

Let Amazon EMR create a new service role so that you can grant and restrict access to resources in other AWS services.

Service role

EMR_DefaultRole

▼

EC2 instance profile for Amazon EMR

The instance profile assigns a role to every EC2 instance in a cluster. The instance profile must specify a role that can access the resources for your steps and bootstrap actions.

☒

Choose an existing instance profile

Select a default role or a custom instance profile with IAM policies attached so that your cluster can interact with your resources in Amazon S3.

☐

Create an instance profile

Let Amazon EMR create a new instance profile so that you can specify a custom set of resources for it to access in Amazon S3.

Instance profile

EMR_EC2_DefaultRole

▼

Custom automatic scaling role - *optional*

When a custom automatic scaling rule triggers, Amazon EMR assumes this role to add and terminate EC2 instances. [Learn more](#)

Custom automatic scaling role

Choose IAM role

Create IAM role

Click the button labeled "Create Cluster" to initiate the process of creating the cluster. Cluster has been created successfully.

wine_prediction

Updated 2 minutes ago

Terminate

Clone in AWS CLI

Clone

▼ Summary

Cluster info

Cluster ID
j-1PB7AG2E0E2MK**Cluster configuration**
Instance groups**Capacity**
1 Primary | 1 Core | 2 Task

Applications

Amazon EMR version
emr-7.5.0**Installed applications**
Hadoop 3.4.0, Hive 3.1.3,
JupyterEnterpriseGateway 2.6.0, Livy 0.8.0,
Spark 3.5.2

Cluster management

Log destination in Amazon S3
[aws-logs-763460480662-us-east-1/elasticmapreduce](#)**Persistent application UIs**[Spark History Server](#)[YARN timeline server](#)[Tez UI](#)**Primary node public DNS**[ec2-54-91-121-24.compute-1.amazonaws.com](#)[Connect to the Primary node using SSH](#)[Connect to the Primary node using SSM](#)

Status and time

Status
 Waiting**Creation time**
November 21, 2024, 12:47 (UTC-05:00)**Elapsed time**
8 minutes

Properties

Bootstrap actions

Instances (Hardware)

Steps


Applications

Configurations

Monitoring

Events

Tags (0)

Operating system [Info](#)Amazon Linux release
2023.6.20241031.0Cluster logs [Info](#)**Archive log files to Amazon S3**
Turned on**Amazon S3 location**
[s3://aws-logs-763460480662-us-east-1/elasticmapreduce/](#)**Encryption for logs**
Turned offCluster termination and node replacement [Info](#)[Edit](#)**Termination option**
Manually terminate cluster**Idle time**
- Your cluster "rr_wine_prediction" has been successfully created.

✕

rr_wine_prediction

Updated less than a minute ago

Terminate

Clone in AWS CLI

Clone

▼ Summary

Cluster info

Cluster ID
j-34IQZHM6W6VCM**Cluster configuration**
Instance groups**Capacity**
1 Primary | 1 Core | 2 Task

Applications

Amazon EMR version
emr-7.5.0**Installed applications**
Hadoop 3.4.0, Hive 3.1.3,
JupyterEnterpriseGateway 2.6.0, Livy 0.8.0,
Spark 3.5.2

Cluster management

Log destination in Amazon S3
[aws-logs-763460480662-us-east-1/elasticmapreduce](#)**Primary node public DNS**

-

[Connect to the Primary node using SSM](#)

Status and time

Status
 Starting**Creation time**
November 20, 2024, 17:06 (UTC-05:00)**Elapsed time**
26 seconds

Properties

Bootstrap actions

Instances (Hardware)

Steps

Applications

Configurations

Monitoring

Events

Tags (0)

Operating system [Info](#)Amazon Linux release
2023.6.20241031.0Cluster logs [Info](#)**Archive log files to Amazon S3**
Turned on**Amazon S3 location**
[s3://aws-logs-763460480662-us-east-1/elasticmapreduce/](#)**Encryption for logs**
Turned offCluster termination and node replacement [Info](#)[Edit](#)**Termination option**
Manually terminate cluster**Idle time**
-**Termination protection**
On**Unhealthy node replacement**
On

Go to the EC2 Instances page. As shown below, there are four EC2 instances running, one designated as the Master node and the other three as Slave nodes.

Instances (4) Info									
<div> <div> <div>Find Instance by attribute or tag (case-sensitive)</div> <div>All states</div> </div> <div> <div>Instance state = running</div> <div>Clear filters</div> </div> </div>									
<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IP
<input type="checkbox"/>		i-0123fbc385b520daa	Running	m5.xlarge	3/3 checks passed	View alarms	us-east-1d	ec2-98-81-59-58.compute-1.amazonaws.com	98.81.59.58
<input type="checkbox"/>		i-0e7a5eeffef4e6e9c	Running	m5.xlarge	3/3 checks passed	View alarms	us-east-1d	ec2-18-212-95-204.compute-1.amazonaws.com	18.212.95.204
<input type="checkbox"/>		i-0a5f8d95c7b60c181	Running	m5.xlarge	3/3 checks passed	View alarms	us-east-1d	ec2-54-91-121-24.compute-1.amazonaws.com	54.91.121.24
<input type="checkbox"/>		i-0b8b47c742ecd8d99	Running	m5.xlarge	3/3 checks passed	View alarms	us-east-1d	ec2-3-80-221-174.compute-1.amazonaws.com	3.80.221.174

To access the "ElasticMapReduce-Master" security group, go to the EC2 service and select the corresponding Security Group ID. Click on the security group to edit its inbound rules.

EC2 > Instances > i-00d4a026ed340ded5		
<div> <div>Instance summary for i-00d4a026ed340ded5 Info</div> <div> <div>Updated less than a minute ago</div> <div> <div>Connect</div> <div>Instance state</div> <div>Actions</div> </div> </div> </div>		
<div>Instance ID</div> <div>i-00d4a026ed340ded5</div>	<div>Public IPv4 address</div> <div>18.212.235.38 open address</div>	<div>Private IPv4 addresses</div> <div>172.31.30.19</div>
<div>IPv6 address</div> <div>–</div>	<div>Instance state</div> <div>Running</div>	<div>Public IPv4 DNS</div> <div>ec2-18-212-235-38.compute-1.amazonaws.com open address</div>
<div>Hostname type</div> <div>IP name: ip-172-31-30-19.ec2.internal</div>	<div>Private IP DNS name (IPv4 only)</div> <div>ip-172-31-30-19.ec2.internal</div>	
<div>Answer private resource DNS name</div> <div>–</div>	<div>Instance type</div> <div>m5.xlarge</div>	<div>Elastic IP addresses</div> <div>–</div>
<div>Auto-assigned IP address</div> <div>18.212.235.38 [Public IP]</div>	<div>VPC ID</div> <div>vpc-0ffab564311d389ad</div>	<div>AWS Compute Optimizer finding</div> <div> Opt-in to AWS Compute Optimizer for recommendations. Learn more </div>
<div>IAM Role</div> <div>EMR_EC2_DefaultRole</div>	<div>Subnet ID</div> <div>subnet-0054303fb8ecf6d46</div>	<div>Auto Scaling Group name</div> <div>–</div>
<div>IMDSv2</div> <div>Required</div>	<div>Instance ARN</div> <div>arn:aws:ec2:us-east-1:763460480662:instance/i-00d4a026ed340ded5</div>	
<div> <div>Details</div> <div>Status and alarms</div> <div>Monitoring</div> <div>Security</div> <div>Networking</div> <div>Storage</div> <div>Tags</div> </div>		
<div> <div>▼ Instance details Info</div> <div> <div>Platform</div> <div>Linux/UNIX</div> </div> <div> <div>AMI ID</div> <div>ami-03a9500db9f5e5067</div> </div> <div> <div>Monitoring</div> <div>disabled</div> </div> </div>		

Inbound rules section, select 'Edit inbound rules' and enter the port numbers 22 and 4040 with the settings indicated below, then click the add rule button and save the rules.

[EC2](#) > [Security Groups](#) > sg-049e6f8a21365ae26 - ElasticMapReduce-master

sg-049e6f8a21365ae26 - ElasticMapReduce-master

Actions ▾

Details

Security group name ElasticMapReduce-master	Security group ID sg-049e6f8a21365ae26	Description Master group for Elastic MapReduce created on 2024-11-13T17:42:04.954Z	VPC ID vpc-0ffab564311d389ad
Owner 763460480662	Inbound rules count 9 Permission entries	Outbound rules count 1 Permission entry	

Inbound rules

Outbound rules

Sharing - new

VPC associations - new

Tags

Inbound rules (9)

🔄

Manage tags

Edit inbound rules

< 1 >

⚙️

<input type="checkbox"/>	Name ▾	Security group rule... ▾	IP version ▾	Type ▾	Protocol ▾
<input type="checkbox"/>	-	sgr-00ede652c2edd6b7a	-	All TCP	TCP
<input type="checkbox"/>	-	sgr-0b95ec7cde683699e	-	All UDP	UDP
<input type="checkbox"/>	-	sgr-04147642f87fa7f38	-	All ICMP - IPv4	ICMP
<input type="checkbox"/>	-	sgr-06052890c416d5...	-	All UDP	UDP

Inbound rules (1/9)

Manage tags

Edit inbound rules

Q

Search

<

1

>

<div>▼</div>	Type	<div>▼</div>	Protocol	<div>▼</div>	Port range	<div>▼</div>	Source	<div>▼</div>
	All TCP		TCP		0 - 65535		sg-049e6f8a21365ae2...	
	All UDP		UDP		0 - 65535		sg-049e6f8a21365ae2...	
	All ICMP - IPv4		ICMP		All		sg-0246a15d4114d42...	
	All UDP		UDP		0 - 65535		sg-0246a15d4114d42...	
	SSH		TCP		22		98.110.74.115/32	
	Custom TCP		TCP		8443		pl-f8bd5e91	
	Custom TCP		TCP		4040		0.0.0.0/0	
	All ICMP - IPv4		ICMP		All		sg-049e6f8a21365ae2...	
	All TCP		TCP		0 - 65535		sg-0246a15d4114d42...	

Next step is to create a S3 bucket in AWS services to store the training & validation datasets.

Choose "Create Bucket" - Labeled the bucket as "rr-programming-assignment-2". Scroll down and click on the "Create bucket" button.

Create bucket [Info](#)

Buckets are containers for data stored in S3.

General configuration

AWS Region

US East (N. Virginia) us-east-1

Bucket type [Info](#)

☒ General purpose

Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

☐ Directory

Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name [Info](#)

rr-programming-assignment-2

Bucket name must be unique within the global namespace and follow the bucket naming rules. [See rules for bucket naming](#)

Copy settings from existing bucket - optional

Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Format: s3://bucket/prefix

Object Ownership [Info](#)

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

☒ ACLs disabled (recommended)

All objects in this bucket are owned by this account. Access to this bucket and its objects is specified using only policies.

☐ ACLs enabled

Objects in this bucket can be owned by other AWS accounts. Access to this bucket and its objects can be specified using ACLs.

Object Ownership

Bucket owner enforced

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

☒ Block all public access

Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.



Block public access to buckets and objects granted through **new** access control lists (ACLs)

S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.



Block public access to buckets and objects granted through **any** access control lists (ACLs)

S3 will ignore all ACLs that grant public access to buckets and objects.



Block public access to buckets and objects granted through **new** public bucket or access point policies

S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.



Block public and cross-account access to buckets and objects through **any** public bucket or access point policies

S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

Bucket Versioning

Versioning is a means of keeping multiple variants of an object in the same bucket. You can use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket. With versioning, you can easily recover from both unintended user actions and application failures. [Learn more](#)

Bucket Versioning

☒ Disable

☐ Enable

Tags - optional (0)

You can use bucket tags to track storage costs and organize buckets. [Learn more](#)

No tags associated with this bucket.

Tags - optional (0)

You can use bucket tags to track storage costs and organize buckets. [Learn more](#)

No tags associated with this bucket.

[Add tag](#)

Default encryption [Info](#)

Server-side encryption is automatically applied to new objects stored in this bucket.

Encryption type [Info](#)

- ☒ Server-side encryption with Amazon S3 managed keys (SSE-S3)
- ☐ Server-side encryption with AWS Key Management Service keys (SSE-KMS)
- ☐ Dual-layer server-side encryption with AWS Key Management Service keys (DSSE-KMS)

Secure your objects with two separate layers of encryption. For details on pricing, see DSSE-KMS pricing on the [Storage tab of the Amazon S3 pricing page](#).

Bucket Key

Using an S3 Bucket Key for SSE-KMS reduces encryption costs by lowering calls to AWS KMS. S3 Bucket Keys aren't supported for DSSE-KMS. [Learn more](#)

- ☐ Disable
- ☒ Enable

► Advanced settings

① After creating the bucket, you can upload files and folders to the bucket, and configure additional bucket settings.

[Cancel](#)

[Create bucket](#)

[Amazon S3](#) > Buckets

① [Help](#)

► Account snapshot - updated every 24 hours [All AWS Regions](#)

[View Storage Lens dashboard](#)

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

General purpose buckets

Directory buckets

General purpose buckets (2) [Info](#) [All AWS Regions](#)



[Copy ARN](#)

[Empty](#)

[Delete](#)

[Create bucket](#)

Buckets are containers for data stored in S3.

[Find buckets by name](#)

< 1 > [Settings](#)

	Name ▲	AWS Region ▼	IAM Access Analyzer	Creation date ▼
<input type="radio"/>	aws-logs-763460480662-us-east-1	US East (N. Virginia) us-east-1	View analyzer for us-east-1	November 14, 2024, 13:58:43 (UTC-05:00)
<input type="radio"/>	rr-programming-assignment-2	US East (N. Virginia) us-east-1	View analyzer for us-east-1	November 20, 2024, 18:07:45 (UTC-05:00)

rr-programming-assignment-2

Info

- Objects
- Properties
- Permissions
- Metrics
- Management
- Access Points

Objects (0)

Info

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 >

Name

Type

Last modified














Size

Storage class

No objects

You don't have any objects in this bucket.

Upload

Details	Status and alarms	Monitoring	Security	Networking	Storage	Tags
▼ Instance details Info						
Platform  Linux/UNIX	AMI ID  ami-03a9500db9f5e5067		Monitoring disabled			
Platform details  Linux/UNIX	AMI name  emr-7_5_0-x86_64-2023_6_20241031_0-Hadoop_Hive_Spark-2024-11-14T05-56-27.240Z		Termination protection Enabled			
Stop protection Disabled	Launch time  Wed Nov 20 2024 17:06:50 GMT-0500 (Eastern Standard Time) (38 minutes)		AMI location  amazon/emr-7_5_0-x86_64-2023_6_20241031_0-Hadoop_Hive_Spark-2024-11-14T05-56-27.240Z			
Instance auto-recovery Default	Lifecycle normal		Stop-hibernate behavior Disabled			
AMI Launch index 0	Key pair assigned at launch  rr_key_pair		State transition reason -			
Credit specification Not supported by instance type	Kernel ID -		State transition message -			
Usage operation  RunInstances	RAM disk ID -		Owner  763460480662			
Enclaves Support Disabled	Boot mode  uefi-preferred		Current instance boot mode  uefi			
Allow tags in instance metadata Disabled	Use RBN as guest OS hostname  Disabled		Answer RBN DNS hostname IPv4  Disabled			
▼ Host and placement group Info						
Host ID -	Affinity -		Placement group -			

Amazon S3 > Buckets > rr-programming-assignment-2 > Upload

Upload

info

Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDKs or Amazon S3 REST API. [Learn more](#)

Drag and drop files and folders you want to upload here, or choose **Add files** or **Add folder**.

Files and folders (0)

Remove

Add files

Add folder

All files and folders in this table will be uploaded.

Find by name

< 1 >

Name

Folder

Type

Size

No files or folders

You have not chosen any files or folders to upload.

Destination

info

Destination

s3://rr-programming-assignment-2

Destination details

Bucket settings that impact new objects stored in the specified destination.

Permissions

Grant public access and access to other AWS accounts.

Properties

Specify storage class, encryption settings, tags, and more.

Click on **"Add Files"** to select the .csv files from your local system. Then, click the **"Upload"** button to upload the datasets to the S3 bucket. Once uploaded, your S3 bucket will contain two .csv files: ValidationDataset.csv and TrainingDataset.csv.

Upload succeeded

For more information, see the **Files and folders** table.

Close

Upload: status

After you navigate away from this page, the following information is no longer available.

Summary

Destination

s3://rr-programming-assignment-2

Succeeded

2 files, 75.7 KB (100.00%)

Failed

0 files, 0 B (0%)

Files and folders

Configuration

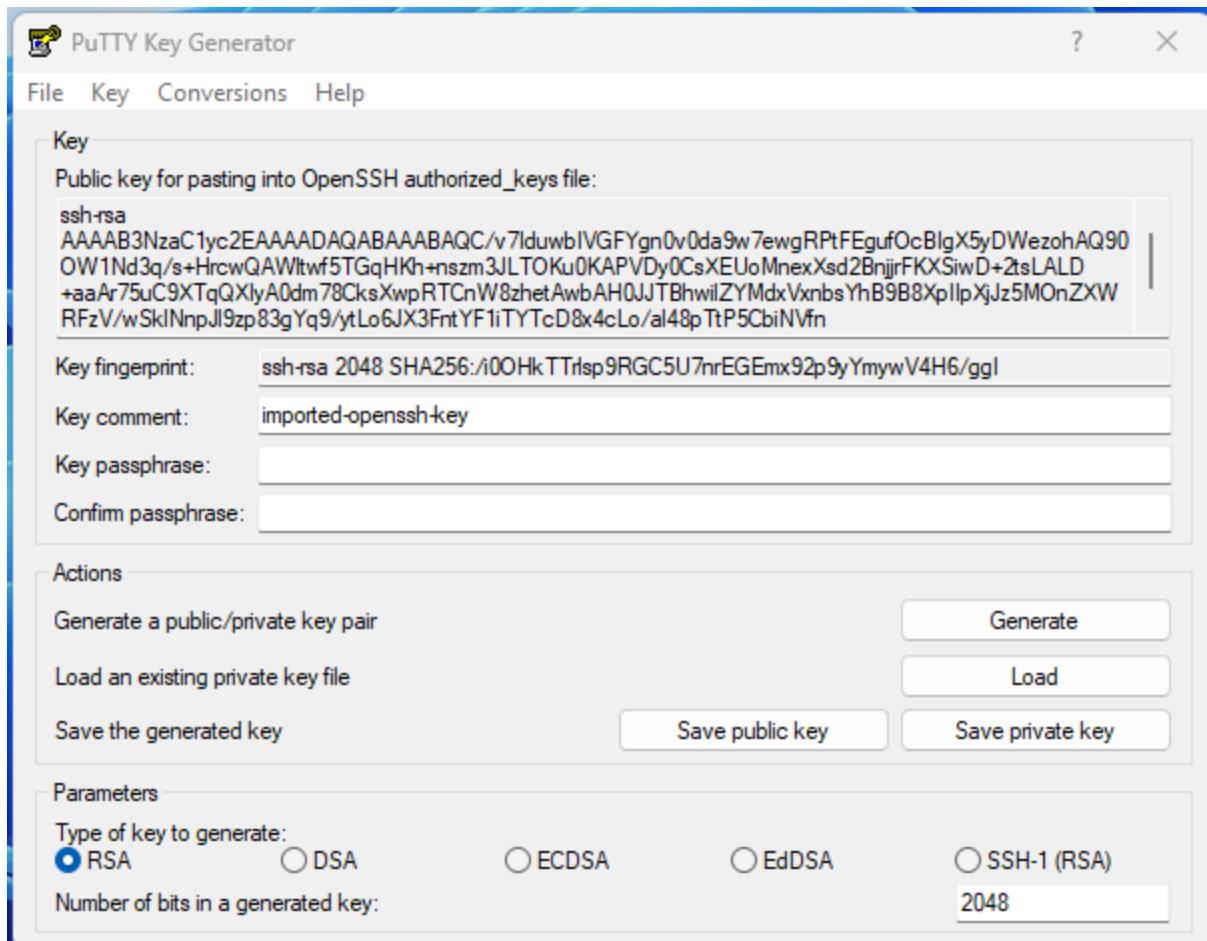
Files and folders (2 total, 75.7 KB)

Find by name

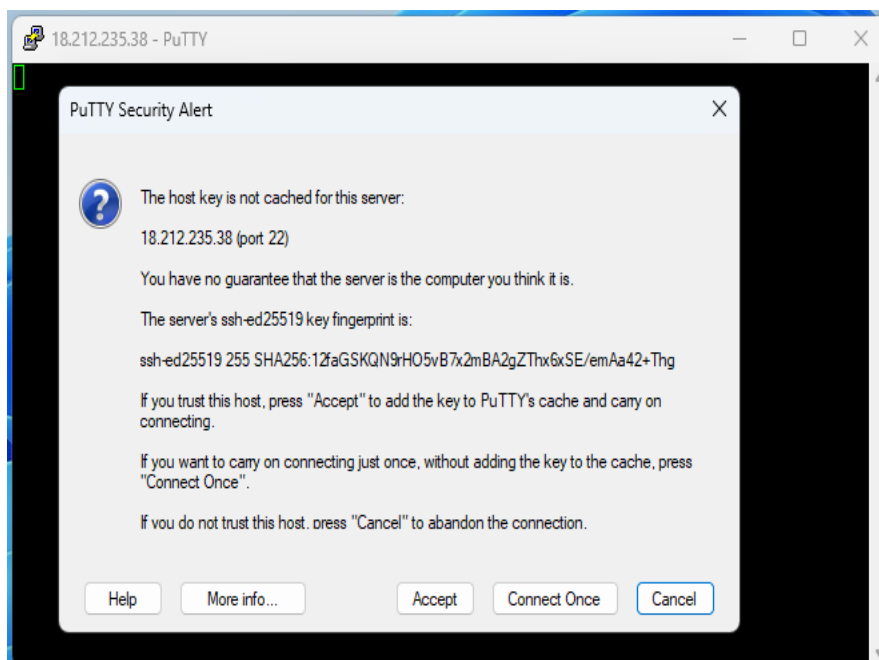
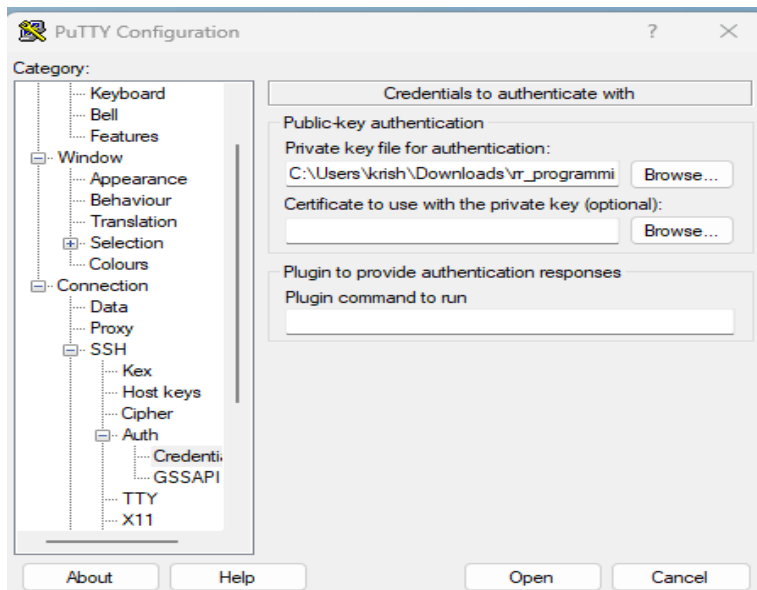
< 1 >

Name	Folder	Type	Size	Status	Error
ValidationDataset.csv	-	text/csv	8.6 KB	Succeeded	-
TrainingDataset.csv	-	text/csv	67.2 KB	Succeeded	-

Convert pem to ppk file using puttygen



Click on "SSH" under "Auth", select "Credentials", provide the path to the location of the PPK file, and then click "Open".



```

ec2-user@ip-172-31-30-19:~
login as: ec2-user
Authenticating with public key "imported-openssh-key"

A newer release of "Amazon Linux" is available.
Version 2023.6.20241111:
Run "/usr/bin/dnf check-release-update" for full release and version update info

#_
##### Amazon Linux 2023
#####\
\###|
\#/ https://aws.amazon.com/linux/amazon-linux-2023
V~' '->
~m/'

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R::::::::::::R
EE::::::::EEEEEEEE::::E M::::::::M M::::::::M R::::::::RRRRRR::::R
E:::E EEEEE M::::::::M M::::::::M RR:::R R:::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::EEEEEEEEEE M:::M M:::M M:::M R::RRRRRR::::R
E::::::::::::E M:::M M:::M M:::M R:::::::::RR
E:::EEEEEEEEEE M:::M M:::M M:::M R::RRRRRR::::R
E:::E M:::M M:::M M:::M R:::R R:::R
E:::E EEEEE M:::M MMM M:::M R:::R R:::R
EE::::::::EEEEEEEE::::E M:::M M:::M R:::R R:::R
E::::::::::::E M:::M M:::M RR:::R R:::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRR RRRRRR

[ec2-user@ip-172-31-30-19 ~]$

```

Set up the credentials for the Master node in the EC2 instance by running the following commands in the terminal:

```
# mkdir .aws
```

```
# touch .aws/credentials
```

```
# vi .aws/credentials
```

```
ec2-user@ip-172-31-30-19:~  
login as: ec2-user  
Authenticating with public key "imported-openssh-key"  
  
A newer release of "Amazon Linux" is available.  
Version 2023.6.20241111:  
Run "/usr/bin/dnf check-release-update" for full release and version update info  
  
#  
~\##### Amazon Linux 2023  
~~\#####  
~~\###|  
~~\#/ https://aws.amazon.com/linux/amazon-linux-2023  
~~V~'-'>  
~~~  
~~-.-  
~~/ /  
~~/m/'  
  
EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRRRRRRRRR  
E:::::E:::::E:::::E M:::::M M:::::M R:::::R  
EE:::::EEEEEEEEEEEE E M:::::M M:::::M R:::::R  
E:::::E EEEEE M:::::M M:::::M RR:::R R:::::R  
E:::::E M:::::M M:::::M M:::::M R:::R R:::::R  
E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR:::R  
E:::::E M:::::M M:::::M M:::::M R:::::RR  
E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR:::R  
E:::::E M:::::M M:::::M M:::::M R:::R R:::::R  
E:::::E EEEEE M:::::M MMM M:::::M R:::R R:::::R  
EE:::::EEEEEEEE:::E M:::::M M:::::M R:::R R:::::R  
E:::::E M:::::M M:::::M RR:::R R:::::R  
EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR  
  
[ec2-user@ip-172-31-30-19 ~]$ mkdir .aws  
[ec2-user@ip-172-31-30-19 ~]$ ls  
[ec2-user@ip-172-31-30-19 ~]$ ls -a  
.. .aws .bash_logout .bash_profile .bashrc .ssh  
[ec2-user@ip-172-31-30-19 ~]$ touch .aws/credentials  
[ec2-user@ip-172-31-30-19 ~]$ vi .aws/credentials
```

Copy and paste the credentials from the AWS Academy page

Cloud Access

Close

AWS CLI:
Copy and paste the following into ~/.aws/credentials

```
[default]
aws_access_key_id=ASIA3DQOKN2LHBDG7EBV
aws_secret_access_key=ZzrSxNA3ZEeDVeoMzr6yTlHIKDrXyRQyh00q2fyd
aws_session_token=IQoJb3JpZ2luX2VjEP7////////wEaCXVzLXd1c3QtMiJHMEUCIQDEUV
m6hYrDGtzB90ttj/k601ADind7IAyCguCsgYTa1wIgZV0QVzxABQI3AUQH26puzuTq+Pj6+jGCd
0Y006PQ2gqsQII1////////ARAAGgw3NjM0NjA0ODA2NjIiDMZSqb1LoP2XwwUoFCqFAqW41+
DLVuAbsB/1C1vLD1AzC+yx88Dr1VBe120dQytJ4EKbZqvtaREbptQabCqyRhAn0pv5043AWIKj/9
mo+k7Y5FosNcQLxg5Tkus3dz9wiTnoAIjGqnK41PnNETW32wRcTHWxGnk4ittBLKEm1+bdw+sqSo
BTKBjj97TNMbbF06yieQSP2MrYVnH0NZpc54aJp3ZT6s0TryzduGq0JEI/EMUzBgdxqx1bp1r0Zn
zF6qB4ZhoxFq2qGj1Px1BWiGQETg+i2przU07Eo5UmGjieUNKXhZXk8xY0V1XkHftZXWj5B1/EY
hnHPX8kwwkThTlC4QM5csn+RDLI4Y9Fe8IVR3qTCdtPm5BjqdAc1u0jEMsf45QFvkkB+h2yUism
yvE1ses5QQHvnmPeSv022/7bnFGCW9t01+68pnmNt6P0Nx5JJEq57pcjNXoqse6bok1T4xK1N1iU
i2kZBbmP8oQKH83MFue3DvQmd4dwh3vtv1VX70e91FrKHggRq1F8MajxvNJrraRGNng2lyHUL8qZ
9rHEm7AQytIoStaW10SBcwF32nSPyD0Ts=
```

Cloud Labs

Remaining session time: 00:13:32(14 minutes)
Session started at: 2024-11-20T13:52:28-0800
Session to end at: 2024-11-20T17:52:28-0800

Accumulated lab time: 3 days 13:24:00 (5124 minutes)

No running instance

SSH key

Show

Download PEM

Download PPK

AWS SSO

Download URL

Install Required Packages. Run the following commands to update the ec2 instance with necessary packages:

```
# sudo yum update
```

```
# sudo yum install git
```

```
# pip install pyspark findspark boto3 numpy pandas scikit-learn datetime
```

To clone the GitHub repository, enter the following command:

```
#git clone https://github.com/rr745/wine\_quality\_predictor
```

spark-submit \

--master yarn \

--deploy-mode cluster \

WinePredictor_Training.py

```
ec2-user@ip-172-31-30-19:~  
Total 37 MB/s | 7.1 MB 00:00  
Running transaction check  
Transaction check succeeded.  
Running transaction test  
Transaction test succeeded.  
Running transaction  
  Preparing : 1/1  
  Installing : git-core-2.40.1-1.amzn2023.0.3.x86_64 1/7  
  Installing : git-core-doc-2.40.1-1.amzn2023.0.3.noarch 2/7  
  Installing : perl-lib-0.65-477.amzn2023.0.6.x86_64 3/7  
  Installing : perl-TermReadKey-2.38-9.amzn2023.0.2.x86_64 4/7  
  Installing : perl-Error-1:0.17029-5.amzn2023.0.2.noarch 5/7  
  Installing : perl-Git-2.40.1-1.amzn2023.0.3.noarch 6/7  
  Installing : git-2.40.1-1.amzn2023.0.3.x86_64 7/7  
  Running scriptlet: git-2.40.1-1.amzn2023.0.3.x86_64 7/7  
  Verifying : git-2.40.1-1.amzn2023.0.3.x86_64 1/7  
  Verifying : git-core-2.40.1-1.amzn2023.0.3.x86_64 2/7  
  Verifying : git-core-doc-2.40.1-1.amzn2023.0.3.noarch 3/7  
  Verifying : perl-Error-1:0.17029-5.amzn2023.0.2.noarch 4/7  
  Verifying : perl-Git-2.40.1-1.amzn2023.0.3.noarch 5/7  
  Verifying : perl-TermReadKey-2.38-9.amzn2023.0.2.x86_64 6/7  
  Verifying : perl-lib-0.65-477.amzn2023.0.6.x86_64 7/7  
=====
```

WARNING:

A newer release of "Amazon Linux" is available.

Available Versions:

Version 2023.6.20241111:
Run the following command to upgrade to 2023.6.20241111:

```
dnf upgrade --releasever=2023.6.20241111
```

Release notes:
<https://docs.aws.amazon.com/linux/al2023/release-notes/relnotes-2023.6.20241111.html>

```
=====
```

Installed:

git-2.40.1-1.amzn2023.0.3.x86_64	git-core-2.40.1-1.amzn2023.0.3.x86_64
git-core-doc-2.40.1-1.amzn2023.0.3.noarch	perl-Error-1:0.17029-5.amzn2023.0.2.noarch
perl-Git-2.40.1-1.amzn2023.0.3.noarch	perl-TermReadKey-2.38-9.amzn2023.0.2.x86_64
perl-lib-0.65-477.amzn2023.0.6.x86_64	

Complete!
[ec2-user@ip-172-31-30-19 ~]\$

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
60	8.9	0.9968	3.39	0.53	9.4	6	0.077	29				
71	7.6	0.9982	3.52	0.65	9.7	5	0.082	23				
37	7.9	0.9966	3.17	0.91	9.5	5	0.106	10				
67	8.5	0.9968	3.17	0.53	9.4	5	0.084	9				
40	6.9	0.9968	3.43	0.63	9.7	6	0.085	21				
23	6.3	0.9955	3.34	0.56	9.3	5	0.08	11				
11	7.6	0.9962	3.28	0.59	9.5	5	0.106	10				
37	7.9	0.9966	3.17	0.91	9.5	5	0.106	10				
35	7.1	0.9972	3.47	0.55	9.4	5	0.08	14				
16	7.8	0.9964	3.38	0.59	9.8	6	0.082	8				
82	6.7	0.9958	3.35	0.54	10.1	5	0.105	22				
37	6.9	0.9966	3.46	0.57	10.6	6	0.083	15				
113	8.3	0.9966	3.17	0.66	9.8	5	0.073	40				
83	6.9	0.9993	3.45	0.52	9.4	6	0.103	13				
50	5.2	0.9957	3.38	0.55	9.2	5	0.086	5				
18	7.8	0.9986	3.4	0.55	9.6	6	0.086	3				
15	8.1	0.9975	3.42	0.6	10.8	6	0.066	13				
30	5.7	0.9968	3.23	0.73	9.7	7	0.172	7				
19	7.3	0.994	3.5	0.48	9.8	4	0.074	12				
87	7.3	0.9978	3.33	0.83	10.5	5						

only showing top 20 rows

```

Training DecisionTree model...
Model - DecisionTree Created
s3a://rr-programming-assignment-2/models/model_dt.model
model_dt.model
False
>>>> DecisionTree model saved
Training RandomForest model...
Model - RandomForest Created
s3a://rr-programming-assignment-2/models/model_rf.model
model_rf.model
False
Model for Random Forest algorithm

```

ec2-user@ip-172-31-24-82:~/wine_quality_predictor

```
ulfur dioxide|density|pH|sulphates|alcohol|quality|
+-----+-----+-----+-----+-----+-----+
| 8.9| 0.22| 0.48| 1.8| 0.077| 29| | |
| 60| 0.9968| 3.39| 0.53| 9.4| 6| 0.082| 23|
| 71| 0.9982| 3.52| 0.65| 9.7| 5| 0.106| 10|
| 37| 0.9966| 3.17| 0.91| 9.5| 5| 0.084| 9|
| 67| 0.9968| 3.17| 0.53| 9.4| 5| 0.085| 21|
| 40| 0.9968| 3.43| 0.63| 9.7| 6| 0.08| 11|
| 23| 0.9955| 3.34| 0.56| 9.3| 5| 0.08| 4|
| 11| 0.9962| 3.28| 0.59| 9.5| 5| 0.106| 10|
| 37| 0.9966| 3.17| 0.91| 9.5| 5| 0.08| 14|
| 35| 0.9972| 3.47| 0.55| 9.4| 5| 0.082| 8|
| 16| 0.9964| 3.38| 0.59| 9.8| 6| 0.089| 17|
| 82| 0.9958| 3.35| 0.54| 10.1| 5| 0.105| 22|
| 37| 0.9966| 3.46| 0.57| 10.6| 6| 0.083| 15|
| 113| 0.9966| 3.17| 0.66| 9.8| 5| 0.073| 40|
| 83| 0.9993| 3.45| 0.52| 9.4| 6| 0.103| 13|
| 50| 0.9957| 3.38| 0.55| 9.2| 5| 0.086| 5|
| 18| 0.9986| 3.4| 0.55| 9.6| 6| 0.086| 3|
| 15| 0.9975| 3.42| 0.6| 10.8| 6| 0.066| 13|
| 30| 0.9968| 3.23| 0.73| 9.7| 7| 0.172| 7|
| 19| 0.994| 3.5| 0.48| 9.8| 4| 0.074| 12|
| 87| 0.9978| 3.33| 0.83| 10.5| 5|
```

only showing top 20 rows

```
Training DecisionTree model...
Model - DecisionTree Created
s3a://xr-programming-assignment-2/models/model_dt.model
model_dt.model
False
>>>> DecisionTree model saved
Training RandomForest model...
Model - RandomForest Created
s3a://xr-programming-assignment-2/models/model_rf.model
model_rf.model
True
>>>> Preexisting Folder Deleted: models/model_rf.model
Model for Random Forest algorithm
```

The results will be displayed here, along with the Accuracy and F1 scores of the Machine Learning methods used.

Training DataSet Metrics

Accuracy: 0.6114151681000782

F-measure: 0.5904050519731796

fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	alcohol	label
7.4	0.7	0.0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5.0
7.8	0.88	0.0	2.6	0.098	25	67	0.9968	3.2	0.68	9.8	5.0
7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5.0
11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6.0
7.4	0.7	0.0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5.0

only showing top 5 rows

Validation Training Set Metrics

features	label	prediction
[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]	15.0	15.0
[9.8,0.68,3.2,0.9968,25.0,67.0,0.098,2.6,0.0,0.88,7.8]	15.0	15.0
[9.8,0.65,3.26,0.997,15.0,54.0,0.092,2.3,0.04,0.76,7.8]	15.0	15.0
[9.8,0.58,3.16,0.998,17.0,60.0,0.075,1.9,0.56,0.28,11.2]	16.0	15.0
[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]	15.0	15.0

only showing top 5 rows

The accuracy of the model is 0.575

F1: 0.5619407071339173

Saved models

aws

Search

[Alt+S]

N. Virginia

voclabs/user2826574=Radhika_Rayala @ 7634-6048-0662

Amazon S3

Buckets

Access Grants

Access Points

Object Lambda Access Points

Multi-Region Access Points

Batch Operations

IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

Dashboards

Storage Lens groups

AWS Organizations settings

Feature spotlight

AWS Marketplace for S3

data/

Objects

Properties

Copy S3 URI

Copy URL

Download

Open

Delete

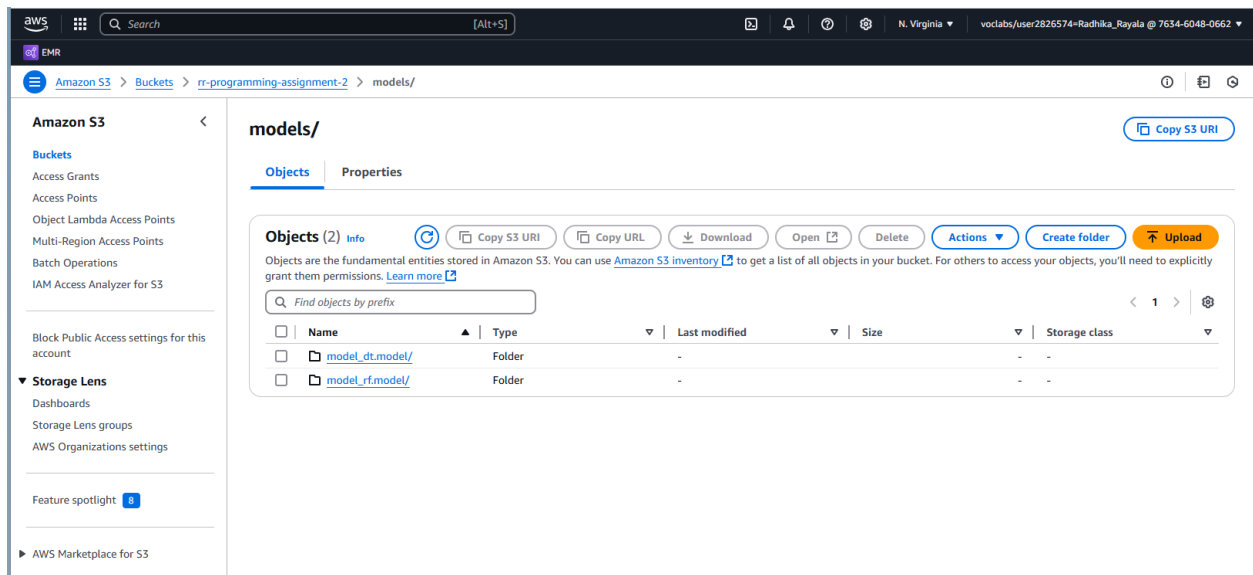
Actions

Create folder

Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
_SUCCESS	-	November 26, 2024, 21:51:17 (UTC-05:00)	0 B	Standard
part-00000-2684c659-695b-4e21-9277-4daf682cb440-c000.snappy.parquet	parquet	November 26, 2024, 21:51:16 (UTC-05:00)	19.0 KB	Standard
part-00001-2684c659-695b-4e21-9277-4daf682cb440-c000.snappy.parquet	parquet	November 26, 2024, 21:51:16 (UTC-05:00)	26.9 KB	Standard
part-00002-2684c659-695b-4e21-9277-4daf682cb440-c000.snappy.parquet	parquet	November 26, 2024, 21:51:16 (UTC-05:00)	20.7 KB	Standard
part-00003-2684c659-695b-4e21-9277-4daf682cb440-c000.snappy.parquet	parquet	November 26, 2024, 21:51:16 (UTC-05:00)	27.2 KB	Standard



DOCKER IMPLEMENTATION –

1.Update System Packages:

```
sudo yum update -y
```

2.Install Docker:

```
sudo yum install -y docker
```

3.Start Docker Service:

```
sudo service docker start
```

4.Check Docker Service Status:

```
sudo service docker status
```

```
[ec2-user@ip-172-31-24-82 wine quality predictor]$ sudo service docker status
Redirecting to /bin/systemctl status docker.service
● docker.service - Docker Application Container Engine
   Loaded: loaded (/usr/lib/systemd/system/docker.service; disabled; preset: disabled)
   Active: active (running) since Wed 2024-11-27 02:58:38 UTC; 7s ago
 TriggeredBy: ● docker.socket
   Docs: https://docs.docker.com
   Process: 26793 ExecStartPre=/bin/mkdir -p /run/docker (code=exited, status=0/SUCCESS)
   Process: 26794 ExecStartPre=/usr/libexec/docker/docker-setup-runtimes.sh (code=exited, status=0/SUCCESS)
   Main PID: 26795 (dockerd)
     Tasks: 10
    Memory: 111.2M
       CPU: 349ms
   CGroup: /system.slice/docker.service
           └─26795 /usr/bin/dockerd -H fd:// --containerd=/run/containerd/containerd.sock --default-ulimit nofile=32768:65536

Nov 27 02:58:30 ip-172-31-24-82.ec2.internal systemd[1]: Starting docker.service - Docker Application Container Engine...
Nov 27 02:58:35 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:35.911040628Z" level=info msg="Starting up"
Nov 27 02:58:36 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:36.888251597Z" level=info msg="Loading containers: start."
Nov 27 02:58:38 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:38.034338573Z" level=info msg="Loading containers: done."
Nov 27 02:58:38 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:38.108299729Z" level=info msg="Docker daemon" commit=b08a51f containerd-snapshotter=false storage-driver=overlay2 version=25.0.6
Nov 27 02:58:38 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:38.109320601Z" level=info msg="Daemon has completed initialization"
Nov 27 02:58:38 ip-172-31-24-82.ec2.internal dockerd[26795]: time="2024-11-27T02:58:38.145620170Z" level=info msg="API listen on /run/docker.sock"
Nov 27 02:58:38 ip-172-31-24-82.ec2.internal systemd[1]: Started docker.service - Docker Application Container Engine.
```

```
ec2-user@ip-172-31-89-91:~
● docker.service - Docker Application Container Engine
   Loaded: loaded (/usr/lib/systemd/system/docker.service; disabled; preset: disabled)
   Active: active (running) since Wed 2024-11-27 20:48:10 UTC; 15s ago
 TriggeredBy: ● docker.socket
   Docs: https://docs.docker.com
   Process: 24540 ExecStartPre=/bin/mkdir -p /run/docker (code=exited, status=0/SUCCESS)
   Process: 24558 ExecStartPre=/usr/libexec/docker/docker-setup-runtimes.sh (code=exited, status=0/SUCCESS)
   Main PID: 24559 (dockerd)
     Tasks: 7
    Memory: 29.6M
       CPU: 246ms
   CGroup: /system.slice/docker.service
           └─24559 /usr/bin/dockerd -H fd:// --containerd=/run/containerd/containerd.sock --default-ulimit nofile=32768:65536

Nov 27 20:48:10 ip-172-31-89-91.ec2.internal systemd[1]: Starting docker.service - Docker Application Container Engine...
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.180139959Z" level=info msg="Starting up"
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.251973003Z" level=info msg="Loading containers: start."
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.788250780Z" level=info msg="Loading containers: done."
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.815032963Z" level=info msg="Docker daemon" commit=b08a51f containerd-snapshotter=false storage-driver=overlay2 version=25.0.6
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.815113747Z" level=info msg="Daemon has completed initialization"
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal dockerd[24559]: time="2024-11-27T20:48:10.865666409Z" level=info msg="API listen on /run/docker.sock"
Nov 27 20:48:10 ip-172-31-89-91.ec2.internal systemd[1]: Started docker.service - Docker Application Container Engine.
~
~
~
~
```

Create a dockerfile and build an image with the docker build command.

`sudo docker build -t rr745/cs643-píogíamming-assignment-2 .`

To check if a Docker image was built, use the following command:

`# sudo docker image ls`

```

ec2-user@ip-172-31-89-91:~
Process: 2468 ExecStartPre=/usr/libexec/docker/docker-setup-runtimes.sh (code=exited, status=0/SUCCESS)
Main PID: 2469 (dockerd)
Tasks: 7
Memory: 109.7M
CPU: 366ms
CGroup: /system.slice/docker.service
└─2469 /usr/bin/dockerd -H fd:// --containerd=/run/containerd/containerd.sock

Nov 28 00:47:05 ip-172-31-89-91.ec2.internal systemd[1]: Starting docker.service: Docker Application Container Engine.
Nov 28 00:47:06 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:06.123456789Z" level=info msg="Starting daemon"
Nov 28 00:47:06 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:06.123456789Z" level=info msg="API listen on /var/run/docker.sock"
Nov 28 00:47:06 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:06.123456789Z" level=info msg="Listening for connections on /var/run/docker.sock"
Nov 28 00:47:07 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:07.123456789Z" level=info msg="Listening for connections on /var/run/docker.sock"
Nov 28 00:47:07 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:07.123456789Z" level=info msg="Listening for connections on /var/run/docker.sock"
Nov 28 00:47:07 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:07.123456789Z" level=info msg="Listening for connections on /var/run/docker.sock"
Nov 28 00:47:07 ip-172-31-89-91.ec2.internal dockerd[2469]: time="2024-11-28T00:47:07.123456789Z" level=info msg="Listening for connections on /var/run/docker.sock"
lines 1-23...skipping...
● docker.service - Docker Application Container Engine
   Loaded: loaded (/usr/lib/systemd/system/docker.service; disabled; preset: disabled)
   Active: active (running) since Thu 2024-11-28 00:47:07 UTC; 8s ago
 TriggeredBy: ● docker.socket
    Docs: https://docs.docker.com
 Process: 2467 ExecStartPre=/bin/mkdir -p /run/docker (code=exited, status=0/SUCCESS)
 Process: 2468 ExecStartPre=/usr/libexec/docker/docker-setup-runtimes.sh (code=exited, status=0/SUCCESS)
 Main PID: 2469 (dockerd)
Tasks: 7
Memory: 109.7M
CPU: 366ms
CGroup: /system.slice/docker.service

[ec2-user@ip-172-31-89-91 ~]$ sudo docker build -t rr745/cs643-programming-assignment-2 .
[+] Building 0.4s (8/8) FINISHED
=> [internal] load build definition from Dockerfile
=> => transferring dockerfile: 522B
=> [internal] load metadata for docker.io/library/python:3.8-slim
=> [internal] load .dockerignore
=> => transferring context: 2B
=> [1/3] FROM docker.io/library/python:3.8-slim@sha256:1d52838af602b4b5a831bebl3a0e4d073280665ea7be7f69ce2382f29c5a613f
=> [internal] load build context
=> => transferring context: 4.12kB
=> CACHED [2/3] WORKDIR /app
=> [3/3] COPY . /app
=> exporting to image
=> => exporting layers
=> => writing image sha256:15516815d2922e9578cd6cbe8a83ec2b60d48890c52a54de560da5e670c44cc0
=> => naming to docker.io/rr745/cs643-programming-assignment-2
[ec2-user@ip-172-31-89-91 ~]$ docker images
REPOSITORY          TAG                 IMAGE ID            CREATED             SIZE
rr745/cs643-programming-assignment-2  latest             15516815d292       About a minute ago  125MB
[ec2-user@ip-172-31-89-91 ~]$ docker run -it rr745/cs643-programming-assignment-2
python: can't open file 'app.py': [Errno 2] No such file or directory

```

To run the docker image, use the following command:

```
sudo docker run -it rr745/cs643-programming-assignment-2
```

Instead of using the image name, you can use the image ID: # sudo docker run -it
This results in the same return for Accuracy and F1 scores.

```

slf4j: Defaulting to no-operation (NOP) logger implementation
slf4j: See http://www.slf4j.org/codes.html#StaticLoggerBinder for further details.
root
[+] #####fixed acidity####: string (nullable = true)
[+] #####volatile acidity####: string (nullable = true)
[+] #####citric acid####: string (nullable = true)
[+] #####residual sugar####: string (nullable = true)
[+] #####chlorides####: string (nullable = true)
[+] #####free sulfur dioxide####: string (nullable = true)
[+] #####total sulfur dioxide####: string (nullable = true)
[+] #####density####: string (nullable = true)
[+] #####pH####: string (nullable = true)
[+] #####sulphates####: string (nullable = true)
[+] #####alcohol####: string (nullable = true)
[+] #####quality####: string (nullable = true)

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|#####fixed acidity####|#####volatile acidity####|#####citric acid####|#####residual sugar####|#####chlorides####|#####free sulfur dioxide####|#####total sulfur dioxide####|#####density####|#####pH####|#####sulphates####|#####alcohol####|#####quality####|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|      |      |      |      |      |      |      |      |      |      |      |      |
|3.39| 0.51| 0.53| 0.22| 0.48| 1.8| 0.077| 29| 60| 0.9968|
|      | 7.6| 0.39| 5| 0.31| 2.3| 0.082| 23| 71| 0.9982|
|3.52| 0.65| 9.7| 0.43| 0.21| 1.6| 0.106| 10| 37| 0.9966|
|      | 7.9| 0.51| 0.49| 0.11| 2.3| 0.084| 9| 67| 0.9968|
|3.17| 0.91| 9.5| 0.4| 0.14| 2.4| 0.085| 21| 40| 0.9968|
|      | 8.5| 0.51| 0.43| 0.21| 1.6| 0.106| 10| 37| 0.9966|
|3.17| 0.53| 9.4| 0.39| 0.16| 1.4| 0.08| 11| 23| 0.9955|
|      | 6.9| 0.51| 0.41| 0.24| 1.8| 0.08| 4| 11| 0.9962|
|3.43| 0.63| 9.7| 0.39| 0.16| 1.4| 0.08| 11| 23| 0.9955|
|      | 6.3| 0.56| 9.3| 0.41| 0.24| 1.8| 0.08| 4| 11| 0.9962|
|3.34| 0.56| 9.3| 0.41| 0.24| 1.8| 0.08| 4| 11| 0.9962|
|      | 7.6| 0.59| 9.5| 0.43| 0.21| 1.6| 0.106| 10| 37| 0.9966|
|3.28| 0.59| 9.5| 0.43| 0.21| 1.6| 0.106| 10| 37| 0.9966|
|      | 7.9| 0.91| 9.5| 0.71| 0| 1.9| 0.08| 14| 35| 0.9972|
|3.17| 0.91| 9.5| 0.71| 0| 1.9| 0.08| 14| 35| 0.9972|
|      | 7.1| 0.55| 9.4| 0.645| 0| 2| 0.082| 8| 16| 0.9964|
|3.47| 0.55| 9.4| 0.645| 0| 2| 0.082| 8| 16| 0.9964|
|      | 7.8| 0.59| 9.6| 0.675| 0.07| 2.4| 0.089| 17| 82| 0.9958|
|3.38| 0.59| 9.6| 0.675| 0.07| 2.4| 0.089| 17| 82| 0.9958|
|      | 6.7| 0.54| 10.11| 0.605| 0| 2.5| 0.105| 22| 37| 0.9966|
|3.35| 0.54| 10.11| 0.605| 0| 2.5| 0.105| 22| 37| 0.9966|
|      | 6.9| 0.57| 10.6| 0.655| 0.12| 2.3| 0.083| 15| 113| 0.9966|
|3.46| 0.57| 10.6| 0.655| 0.12| 2.3| 0.083| 15| 113| 0.9966|
|      | 8.3| 0.66| 9.8| 0.605| 0.12| 10.7| 0.073| 40| 83| 0.9993|
|3.17| 0.66| 9.8| 0.605| 0.12| 10.7| 0.073| 40| 83| 0.9993|
|      | 6.9| 0.52| 9.4| 0.32| 0.25| 1.8| 0.103| 13| 50| 0.9957|
|3.45| 0.52| 9.4| 0.32| 0.25| 1.8| 0.103| 13| 50| 0.9957|
|      | 5.2| 0.55| 9.2| 0.645| 0| 5.5| 0.086| 5| 18| 0.9986|
|3.38| 0.55| 9.2| 0.645| 0| 5.5| 0.086| 5| 18| 0.9986|
|      | 7.8| 0.55| 9.6| 0.6| 0.14| 2.4| 0.086| 3| 15| 0.9475|
|3.4| 0.55| 9.6| 0.6| 0.14| 2.4| 0.086| 3| 15| 0.9475|
|      | 7.8|

```

```

TestingDataSet Metrics
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|features|label|prediction|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]|15.0|15.0|
|[9.8,0.68,3.2,0.9968,25.0,67.0,0.098,2.6,0.0,0.88,7.8]|15.0|15.0|
|[9.8,0.65,3.26,0.997,15.0,54.0,0.092,2.3,0.04,0.76,7.8]|15.0|15.0|
|[9.8,0.58,3.16,0.998,17.0,60.0,0.075,1.9,0.56,0.28,11.2]|16.0|15.0|
|[9.4,0.56,3.51,0.9978,11.0,34.0,0.076,1.9,0.0,0.7,7.4]|15.0|15.0|
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 5 rows

The accuracy of the model is 0.6271186440677966
F1: 0.593151718932272

```

Run the following command to submit the produced Docker image to the DockerHub repository:

sudo docker push rr745/643-programming-assignment-2

```
[ec2-user@ip-172-31-89-91 ~]$ docker login
Authenticating with existing credentials...
WARNING! Your password will be stored unencrypted in /home/ec2-user/.docker/config.json.
Configure a credential helper to remove this warning. See
https://docs.docker.com/engine/reference/commandline/login/#credentials-store

Login Succeeded
[ec2-user@ip-172-31-89-91 ~]$ sudo docker push rr745/cs643-programming-assignment-2
Using default tag: latest
The push refers to repository [docker.io/rr745/cs643-programming-assignment-2]
b0470c1d8c15: Preparing
88f82b0a637e: Preparing
d2a2207b52a4: Preparing
5d2d143f3d7f: Preparing
c3772b569c3a: Preparing
8d853c8add5d: Waiting
...
[ec2-user@ip-172-31-89-91 ~]$
```

```
[ec2-user@ip-172-31-89-91 ~]$ docker login -u rr745
Password:
WARNING! Your password will be stored unencrypted in /home/ec2-user/.docker/config.json.
Configure a credential helper to remove this warning. See
https://docs.docker.com/engine/reference/commandline/login/#credentials-store

Login Succeeded
[ec2-user@ip-172-31-89-91 ~]$ docker push rr745/cs643-programming-assignment-2:latest
The push refers to repository [docker.io/rr745/cs643-programming-assignment-2]
fc4ddb5d5303: Pushed
88f82b0a637e: Pushed
d2a2207b52a4: Pushed
5d2d143f3d7f: Pushed
c3772b569c3a: Pushed
8d853c8add5d: Pushed
latest: digest: sha256:58e7a5a91cf558182c18635d034edd9d94fa95e2bab2ca7af63689e1652670c2 size: 1574
[ec2-user@ip-172-31-89-91 ~]$ docker pull rr745/cs643-programming-assignment-2:latest
latest: Pulling from rr745/cs643-programming-assignment-2
...
```

Download and execute the Docker image from the DockerHub repository, following the instructions provided on the website.

Git Bash:

the steps for moving code from my local system to a GitHub repository using Git Bash.

1. git init
2. git add
3. git status
4. git commit -m "Updates"
5. git remote add origin https://github.com/rr745/wine_quality_predictor.git
6. git push -u origin main