

CREDIT CARD FRAUD DETECTION.

AIDI 1006 ASSIGNMENT 4

MEMBERS

AKSHAYKUMAR THAKARE

(200523970)

MANAV KUMBHANI

(200509153)

IN THIS PRESENTATION

- Selected Solution: Credit Card Fraud Detection
- Solution is based on Microsoft Azure
- Data Source: CreditCard.csv
- Programming Language: Python
- Visualization Tool: PowerPoint

OVERVIEW



DATA LOADING

Home > storageaccounttakshaynew | Containers >

mycontainernew
Container

Search resources, services, and docs (G+)

200527861@student.ge...
GEORGIAN COLLEGE (GEORGIAN...)

Home > storageaccounttakshaynew | Containers >

mycontainernew
Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot Give feedback

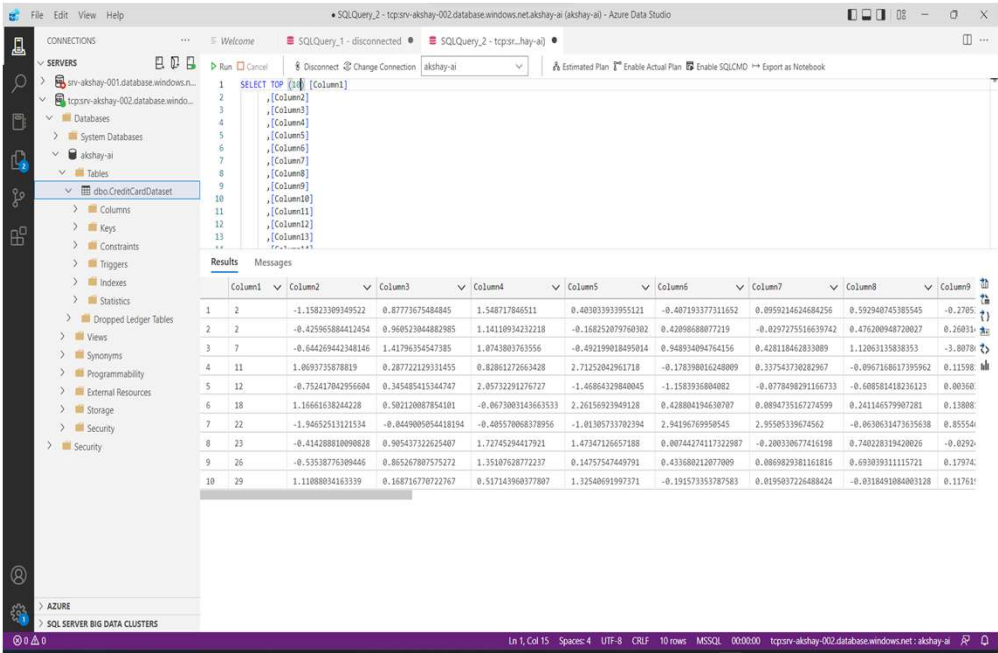
Authentication method: Access key (Switch to Azure AD User Account)
Location: mycontainernew

Search blobs by prefix (case-sensitive) Show deleted blobs

Add filter

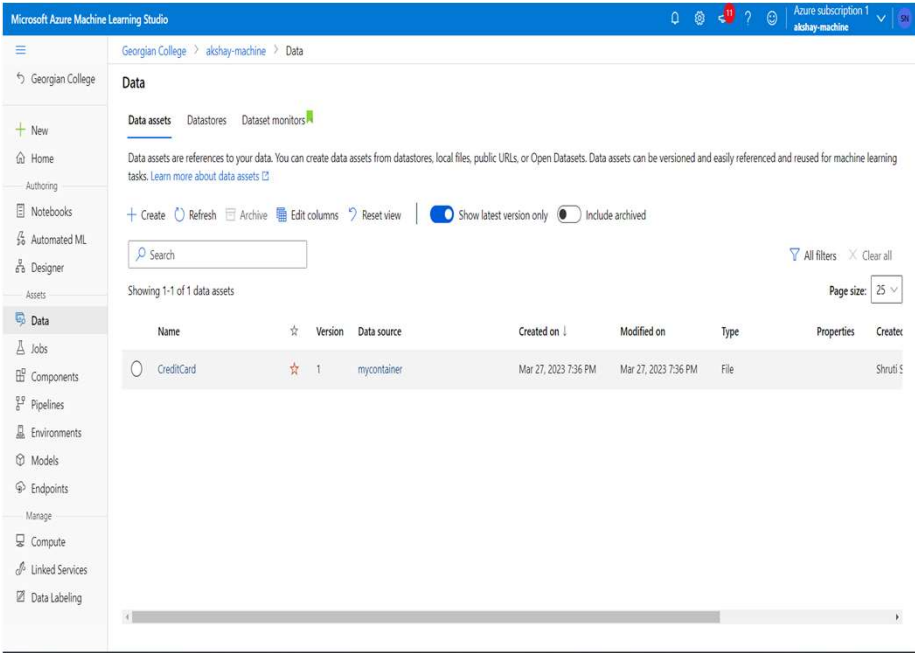
Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
<input type="checkbox"/> CreditCardDataset.csv	3/26/2023, 10:02:40 ...	Hot (Inferred)		Block blob	28.85 MiB	Available ***
<input type="checkbox"/> NOTICE.txt	3/26/2023, 9:21:17 PM	Hot (Inferred)		Block blob	500.59 KiB	Available ***

DATA LOADING



The screenshot shows the SQL Server Enterprise Manager interface. The left pane displays the server hierarchy, including 'Servers', 'Databases', and 'Tables'. The right pane shows the 'Results' tab with a table of data. The table has 10 columns and 10 rows of data. The first row of data is highlighted.

Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10
1	2	-1.15823309349522	0.87773675484845	1.548717846511	-0.483033933955121	-0.487193377311652	0.8959214624684256	0.592940745385545	-0.2785
2	2	-0.425965884412454	0.96953984882985	1.14118934232218	-0.168252097969302	-0.42898688077219	-0.8297275516636742	0.476208948720027	-0.26031
3	7	-0.644269442348146	1.41796354547385	1.8743803763556	-0.492199818495014	-0.948934894764156	0.428118462833809	1.12063135838353	-3.8078
4	11	1.0693735878819	0.287722129331455	0.82861272663428	2.712528429617718	-0.178398016248069	0.3375437730282967	-0.8967868617395962	0.11598
5	12	-0.752417842956604	0.345485415344747	2.85732291276727	-1.46864329840045	-1.1583936804082	-0.8778498291166733	-0.608581418226123	0.00360
6	18	1.16661838244228	0.582120087854101	-0.0673083143663533	2.26156923949128	0.428804194638787	0.0894735167274599	0.241146579907281	0.13808
7	22	-1.94652513121534	-0.844900854418194	-0.40557068378956	-1.81385733782394	2.94196769959545	2.95505339674562	-0.8630631473635638	0.85554
8	23	-0.414288818090828	0.905437322625407	1.72745294417921	1.47347126657188	0.00744274117322987	-0.280330677416198	0.740228319420026	-0.8292
9	26	-0.53538776309446	0.865267807575272	1.35107628772237	0.14757547449791	0.433680212077009	0.8069829381161816	0.693039311115721	0.17974
10	29	1.11088034163339	0.16876770722767	0.517143968377807	1.32540691997371	-0.191573353787583	0.8195037226488424	-0.818493884083128	0.11761



The screenshot shows the Microsoft Azure Machine Learning Studio interface. The left pane displays the navigation menu, including 'Georgian College', 'Data', 'New', 'Home', 'Authoring', 'Notebooks', 'Automated ML', 'Designer', 'Assets', 'Data', 'Jobs', 'Components', 'Pipelines', 'Environments', 'Models', 'Endpoints', 'Manage', 'Compute', 'Linked Services', and 'Data Labeling'. The right pane shows the 'Data' section with a list of data assets. The table below shows the data assets.

Name	Version	Data source	Created on	Modified on	Type	Properties	Creator
CreditCard	1	mycontainer	Mar 27, 2023 7:36 PM	Mar 27, 2023 7:36 PM	File		Shrutis

DATA LOADING

Microsoft Azure | AzDataFactorynew

Search

200527861@student.georgianc.on.ca

GEORGIAN COLLEGE

Dashboards

Runs

Pipeline runs

Trigger runs

Change Data Capture (previ...

Runtimes & sessions

Integration runtimes

Data flow debug

Notifications

Alerts & metrics

Pipeline runs

Triggered Debug Rerun Cancel options Refresh Edit columns List Gantt

Filter by run ID or name Local time : Last 24 hours Pipeline name : CopyPipeline_Test1 Status : All Runs : Latest runs Copy filters Export to CSV

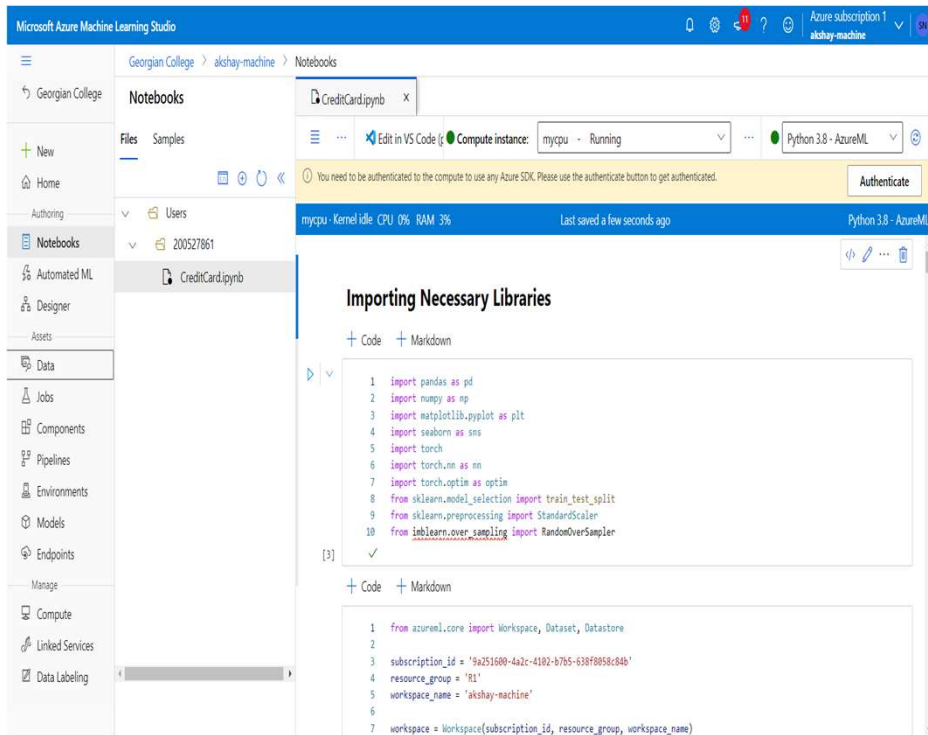
Triggered by : All Add filter

Showing 1 - 1 items Last refreshed 15 minutes ago

<input type="checkbox"/> Pipeline name ↑↓	Run start ↑↓	Run end ↑↓	Duration	Triggered by	Status ↑↓	Error	Run
<input type="checkbox"/> CopyPipeline_Test1	3/26/2023, 11:24:56 PM	3/26/2023, 11:26:20 PM	00:01:23	Manual trigger	✓ Succeeded		Original

6

MACHINE LEARNING CODE



Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Notebooks

Files Samples

Compute instance: mycpu - Running Python 3.8 - AzureML

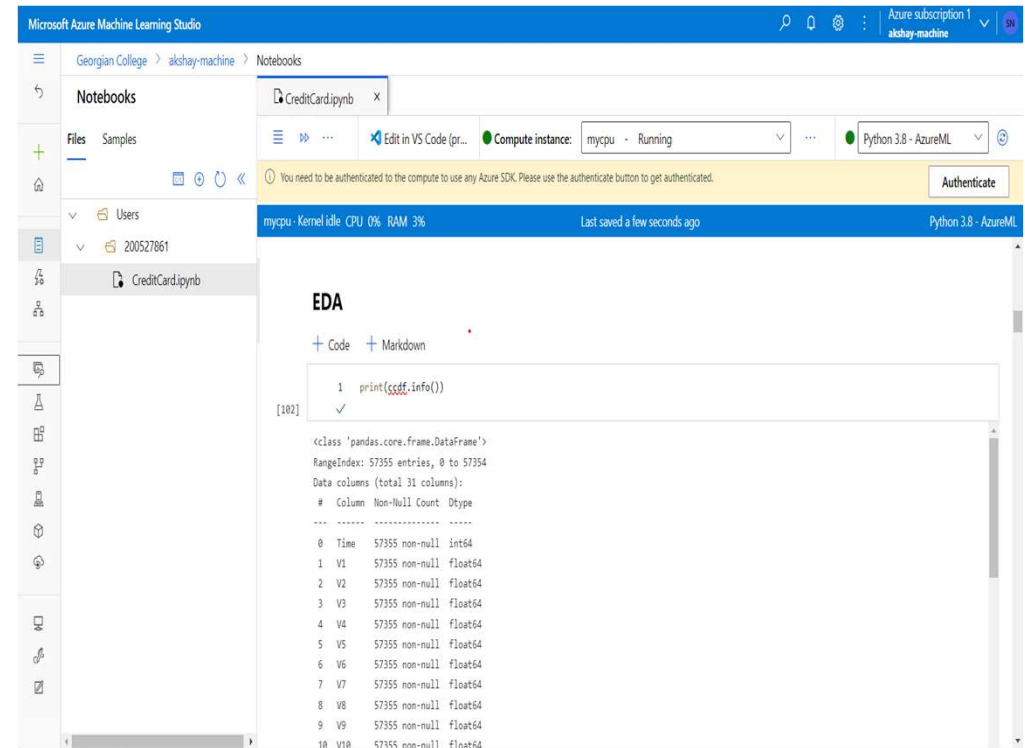
You need to be authenticated to the compute to use any Azure SDK. Please use the authenticate button to get authenticated. Authenticate

mycpu - Kernel idle CPU 0% RAM 3% Last saved a few seconds ago Python 3.8 - AzureML

Importing Necessary Libraries

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 import torch
6 import torch.nn as nn
7 import torch.optim as optim
8 from sklearn.model_selection import train_test_split
9 from sklearn.preprocessing import StandardScaler
10 from imblearn.over_sampling import RandomOverSampler
```

```
1 from azureml.core import Workspace, Dataset, Datastore
2
3 subscription_id = '9a251600-4a2c-4102-b7b5-638f8958c84b'
4 resource_group = 'RI'
5 workspace_name = 'akshay-machine'
6
7 workspace = Workspace(subscription_id, resource_group, workspace_name)
```



Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Notebooks

Files Samples

Compute instance: mycpu - Running Python 3.8 - AzureML

You need to be authenticated to the compute to use any Azure SDK. Please use the authenticate button to get authenticated. Authenticate

mycpu - Kernel idle CPU 0% RAM 3% Last saved a few seconds ago Python 3.8 - AzureML

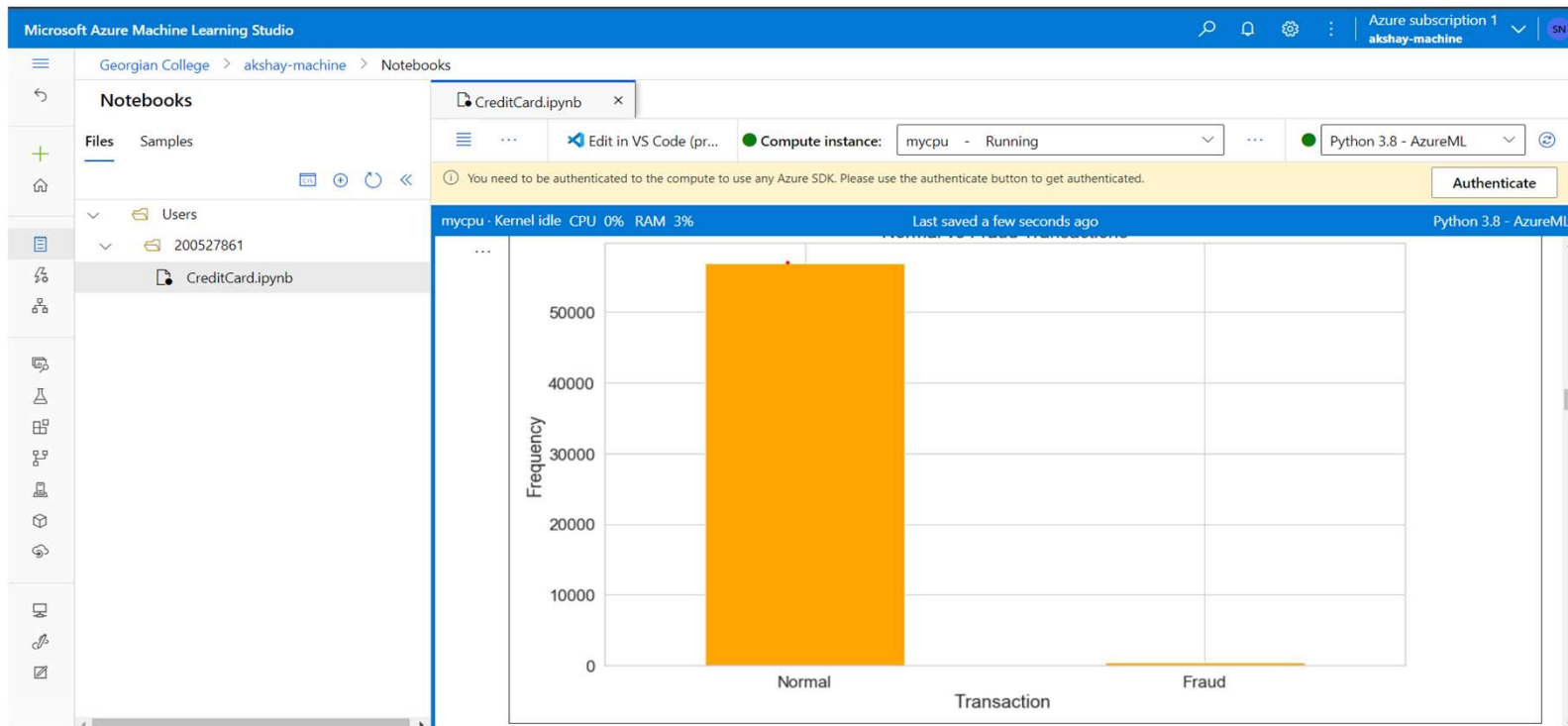
EDA

```
1 print(ccdf.info())
```

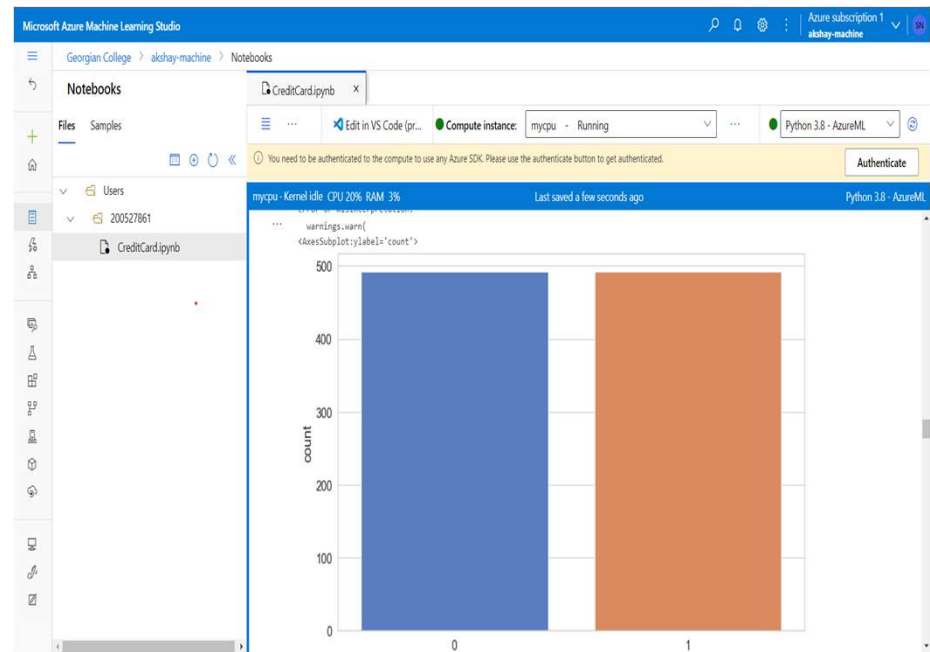
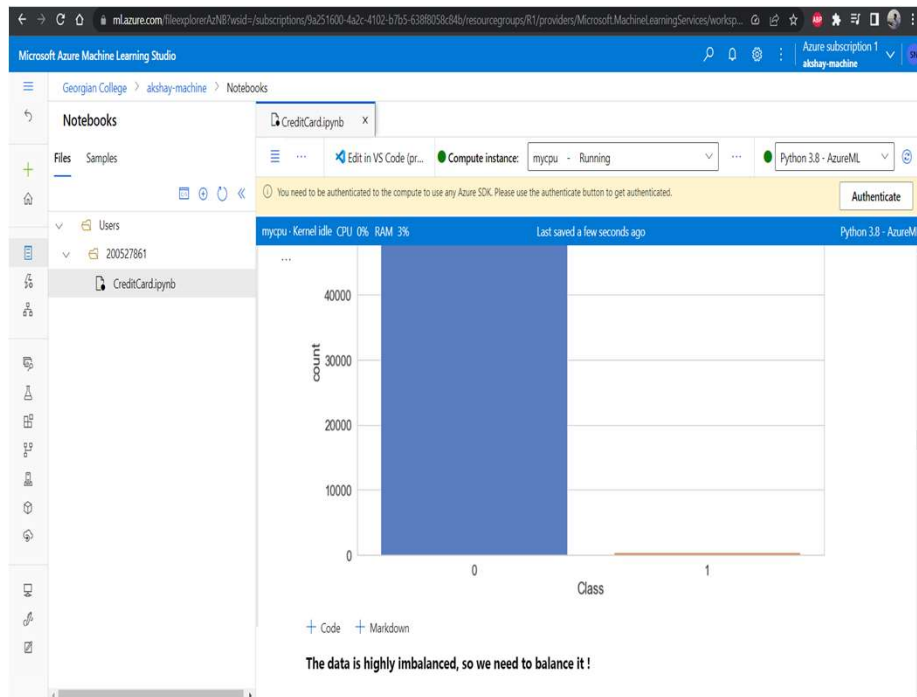
```
[182] ✓
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57355 entries, 0 to 57354
Data columns (total 31 columns):
#   Column  Non-Null Count  Dtype
---  -
0   Time    57355 non-null      int64
1   V1      57355 non-null      float64
2   V2      57355 non-null      float64
3   V3      57355 non-null      float64
4   V4      57355 non-null      float64
5   V5      57355 non-null      float64
6   V6      57355 non-null      float64
7   V7      57355 non-null      float64
8   V8      57355 non-null      float64
9   V9      57355 non-null      float64
10  V10     57355 non-null      float64
```

MACHINE LEARNING CODE



MACHINE LEARNING CODE



ML MODELLING

Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Notebooks

Notebooks

Files Samples

Users

200527861

CreditCard.ipynb

mycpu - Kernel idle CPU 0% RAM 3% Last saved a few seconds ago Python 3.8 - AzureML

You need to be authenticated to the compute to use any Azure SDK. Please use the authenticate button to get authenticated. Authenticate

```

1 # Split the dataset into features and target
2 X = ccdf.drop(['class'], axis=1)
3 y = ccdf['class']
4
5 # Split the dataset into train and test sets
6 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
7
8 # Standardize the data
9 scaler = StandardScaler()
10 X_train = scaler.fit_transform(X_train)
11 X_test = scaler.transform(X_test)
12
13 # Oversample the minority class
14 oversampler = RandomOverSampler(random_state=42)
15 X_train, y_train = oversampler.fit_resample(X_train, y_train)
16
17 # Convert the data to tensors
18 X_train = torch.Tensor(X_train)
19 y_train = torch.Tensor(y_train.to_numpy())
20 X_test = torch.Tensor(X_test)
21 y_test = torch.Tensor(y_test.to_numpy())
22

```

[119] ✓

Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Notebooks

Notebooks

Files Samples

Users

200527861

CreditCard.ipynb

mycpu - Kernel idle CPU 0% RAM 3% Last saved a few seconds ago Python 3.8 - AzureML

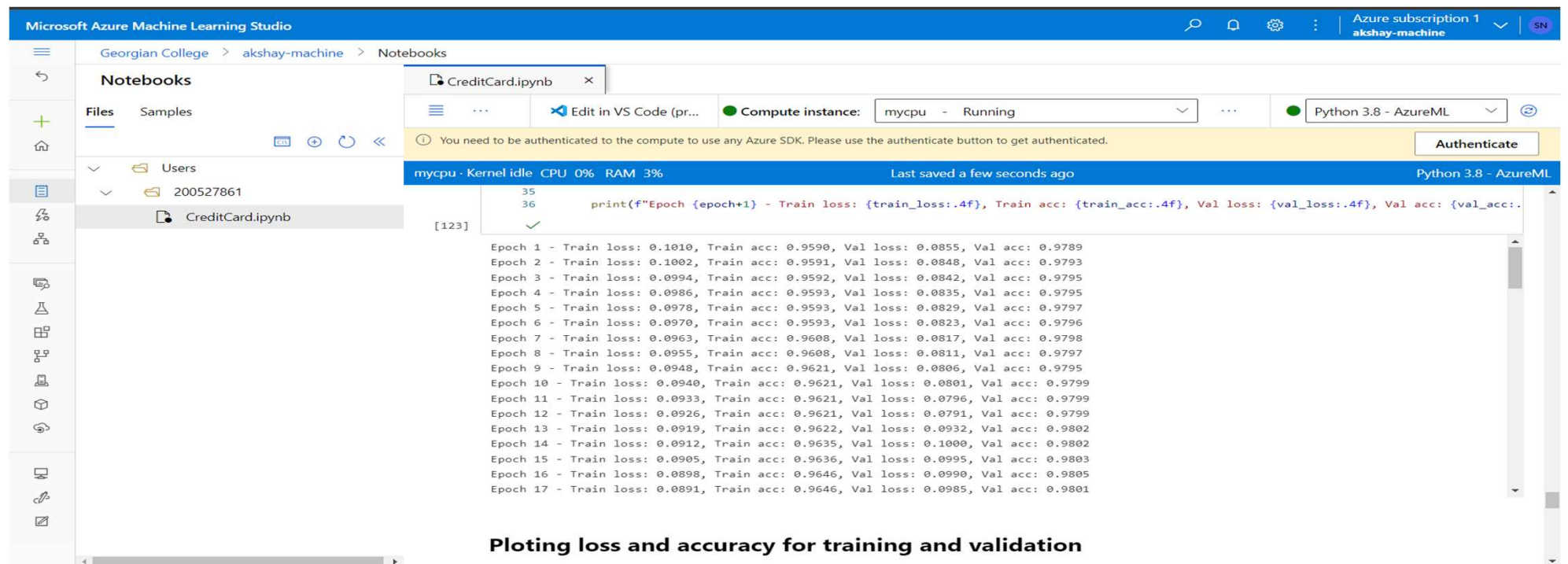
You need to be authenticated to the compute to use any Azure SDK. Please use the authenticate button to get authenticated. Authenticate

```

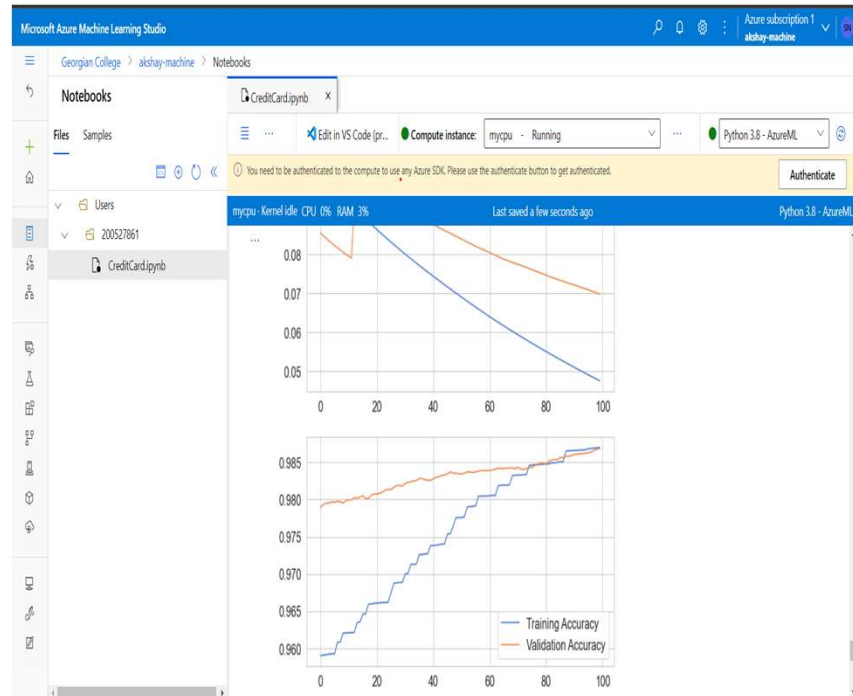
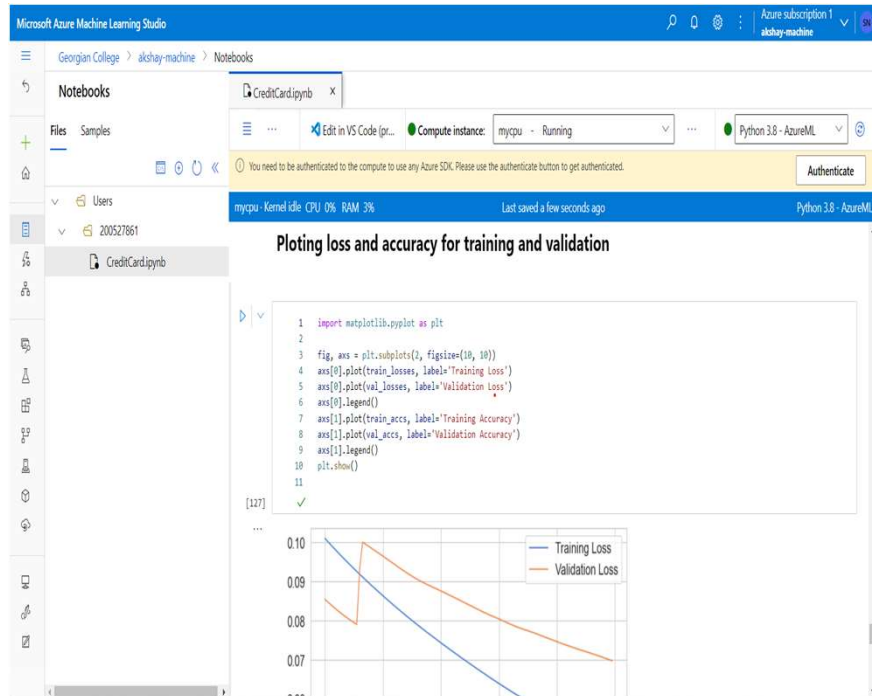
1 train_losses = []
2 train_accs = []
3 val_losses = []
4 val_accs = []
5
6 for epoch in range(100):
7     train_loss = 0.0
8     train_acc = 0.0
9     val_loss = 0.0
10    val_acc = 0.0
11    model.train()
12
13    optimizer.zero_grad() # reset the gradients
14    outputs = model(X_train)
15    loss = criterion(outputs, y_train.view(-1, 1)) # calculate the loss
16    loss.backward()
17    optimizer.step()
18
19    model.eval()
20    with torch.no_grad():
21        train_outputs = model(X_train)
22        train_loss = criterion(train_outputs, y_train.view(-1, 1)).item() # calculate the training loss
23        train_preds = (train_outputs > 0.5).float() # calculate the training predictions
24        train_acc = (train_preds == y_train.view(-1, 1)).float().mean().item() # calculate the training accuracy
25
26    val_outputs = model(X_test)
27    val_loss = criterion(val_outputs, y_test.view(-1, 1)).item()
28    val_preds = (val_outputs > 0.5).float()
29    val_acc = (val_preds == y_test.view(-1, 1)).float().mean().item() # calculate the validation accuracy

```

ML MODELLING (ACCURACY 98%)



ML MODELLING (VISUALIZING)



The background is a dark, textured surface with a network of light-colored nodes and lines. The nodes are small circles of varying sizes, and the lines are thin, connecting the nodes in a complex, web-like pattern. The overall effect is a sense of interconnectedness and complexity.

THANK YOU