

# CREDIT CARD FRAUD DETECTION.

AIDI 1006 ASSIGNMENT 4

MEMBERS

AKSHAYKUMAR THAKARE

(200523970)

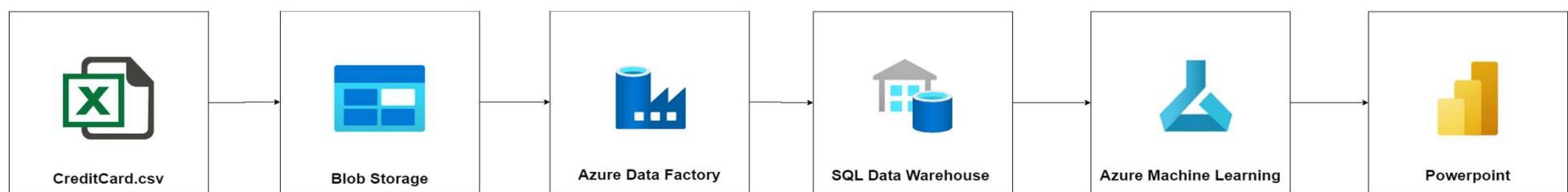
MANAV KUMBHANI

(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

The screenshot shows the Microsoft Azure Storage Container Overview page for a container named 'mycontainernew'. The left sidebar includes links for Overview, Diagnose and solve problems, Access Control (IAM), Settings (Shared access tokens, Access policy, Properties, Metadata), and a search bar. The main content area displays blob details with columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. Two blobs are listed: 'CreditCardDataset.csv' and 'NOTICE.txt'.

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

# DATA LOADING

```

File Edit View Help
CONNECTIONS
  Welcome SQLQuery_1 - disconnected SQLQuery_2 - tcpsrv-akshay-002.database.windows.net.akshay-ai (akshay-ai) - Azure Data Studio
  Run Cancel Disconnect Change Connection: akshay-ai
  Estimated Plan Enable Actual Plan Enable SQLCMD Export as Notebook
  SERVERS
    srv-akshay-001.database.windows.net
    tcpsrv-akshay-002.database.windows.net
    System Databases
    akshay-ai
      Tables
        dbo.CreditCardDataset
          Column1 Column2 Column3 Column4 Column5 Column6 Column7 Column8 Column9
        1 -1.1582309349522 -0.87773675484845 1.548717846511 0.40303933955121 -0.407193377311652 0.0959214624684256 0.592940745385545 -0.2705
        2 -0.42595884412454 0.960523044882985 1.14110934232218 -0.168252079760302 0.42896888077219 -0.029725516639742 0.476200948720027 0.2693
        3 7 -0.844269442348146 1.41796354547385 1.0743803763556 -0.492199018495914 0.948934094764156 0.428118462833089 1.12063135838353 -3.807
        4 11 1.0693735878819 0.28772212931349 0.82861272663428 2.715264294916717 -0.178390816248089 0.33754370282367 -0.09571861739596 0.121598
        5 12 -0.752417642956669 0.345485415344747 2.057321291276727 -1.46864329840495 -1.1583935880482 -0.07886498291166733 -0.608581418232123 0.083680
        6 18 1.16661638242278 0.592120087854101 -0.067308134363533 2.26156923949128 0.428864194630707 0.0894735167274599 0.24114657997281 0.133886
        7 22 -1.9465251321534 -0.849008654138194 -0.4055706883789594 -1.01389733792394 2.94196769595845 0.95505339674562 -0.063053473635639 0.85554d
        8 23 -0.414288810998826 0.905437322625407 1.7745264417921 0.4774126657108 0.80742374117322087 -0.200310677416198 0.740228319420026 -0.0292
        9 26 -0.53538776309446 0.865267807575272 1.35107628772237 0.14757547449791 0.433680212077009 0.086983931161816 0.69303931115721 0.17974
        10 29 1.11088034163339 0.168716770722767 0.517143968377887 1.32540861997371 -0.191573353787583 0.0195037226488424 -0.03184910840803128 0.11761
  
```

Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Data

**Data**

Data assets Datastores Dataset monitors

Data assets are references to your data. You can create data assets from datastores, local files, public URLs, or Open Datasets. Data assets can be versioned and easily referenced and reused for machine learning tasks. [Learn more about data assets](#)

Create Refresh Archive Edit columns Reset view Show latest version only Include archived

All filters Clear all

Showing 1-1 of 1 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		Shruti S

Jobs Components Pipelines Environments Models Endpoints Manage Compute Linked Services Data Labeling

# DATA LOADING

Microsoft Azure | AzDataFactorynew

Search

Dashboard

Runs

Pipeline runs

Trigger runs

Change Data Capture (previous)

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

Triggered by : All Add filter

Last refreshed 15 minutes ago

Pipeline name	Run start	Run end	Duration	Triggered by	Status	Error	Run
CopyPipeline_Test1	3/26/2023, 11:24:56 PM	3/26/2023, 11:26:20 PM	00:01:23	Manual trigger	Succeeded		Original

# MACHINE LEARNING CODE

Microsoft Azure Machine Learning Studio

Georgian College > akshay-machine > Notebooks

Notebooks CreditCard.ipynb

Files Samples Edit in VS Code Compute instance: mycpu - Running Python 3.8 - AzureML Authenticate

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

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

Importing Necessary Libraries

+ Code + Markdown

```
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
```

[3] ✓

+ Code + Markdown

```
1 from azureml.core import Workspace, Dataset, Datastore
2
3 subscription_id = '9a251600-4a2c-4102-b7b5-638f0058c84b'
4 resource_group = 'R1'
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

Notebooks CreditCard.ipynb

Files Samples Edit in VS Code Compute instance: mycpu - Running Python 3.8 - AzureML Authenticate

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

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

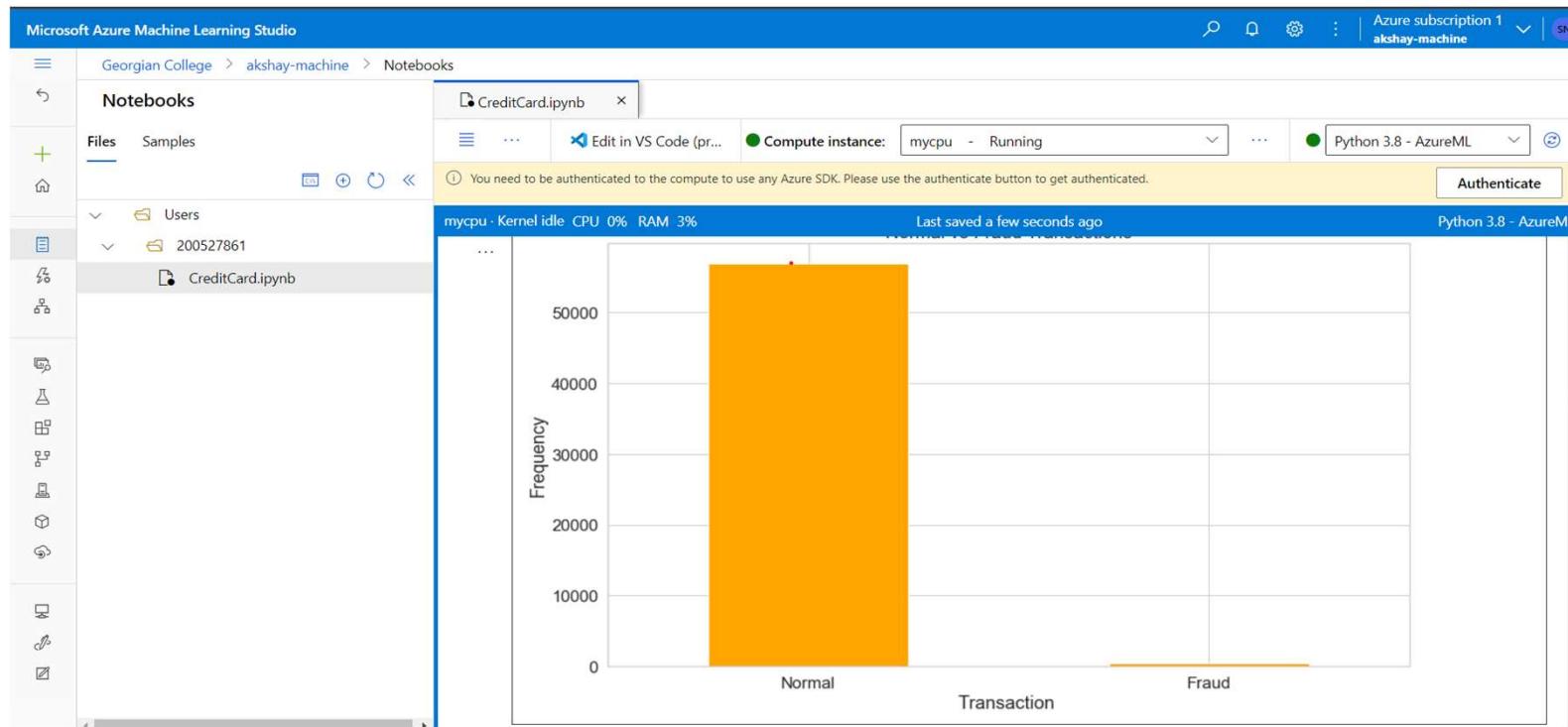
EDA

+ Code + Markdown

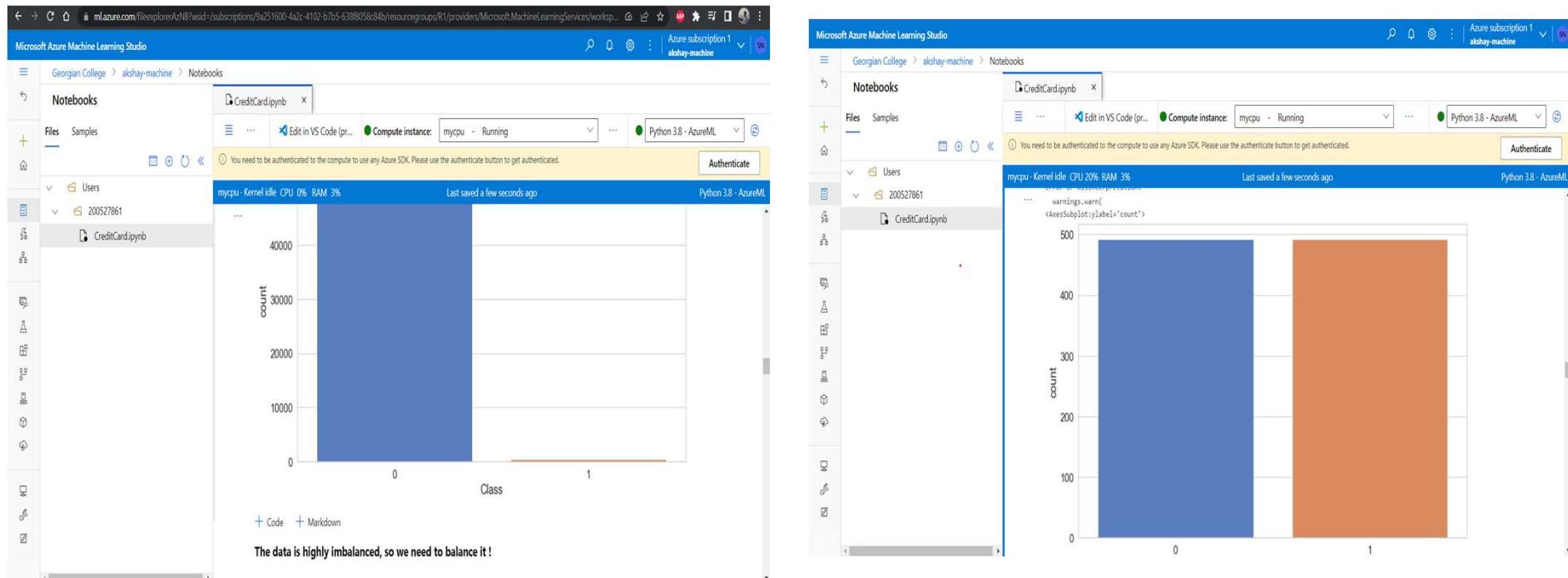
```
1 print(ccdf.info())
[182] ✓
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57355 entries, 0 to 57355
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

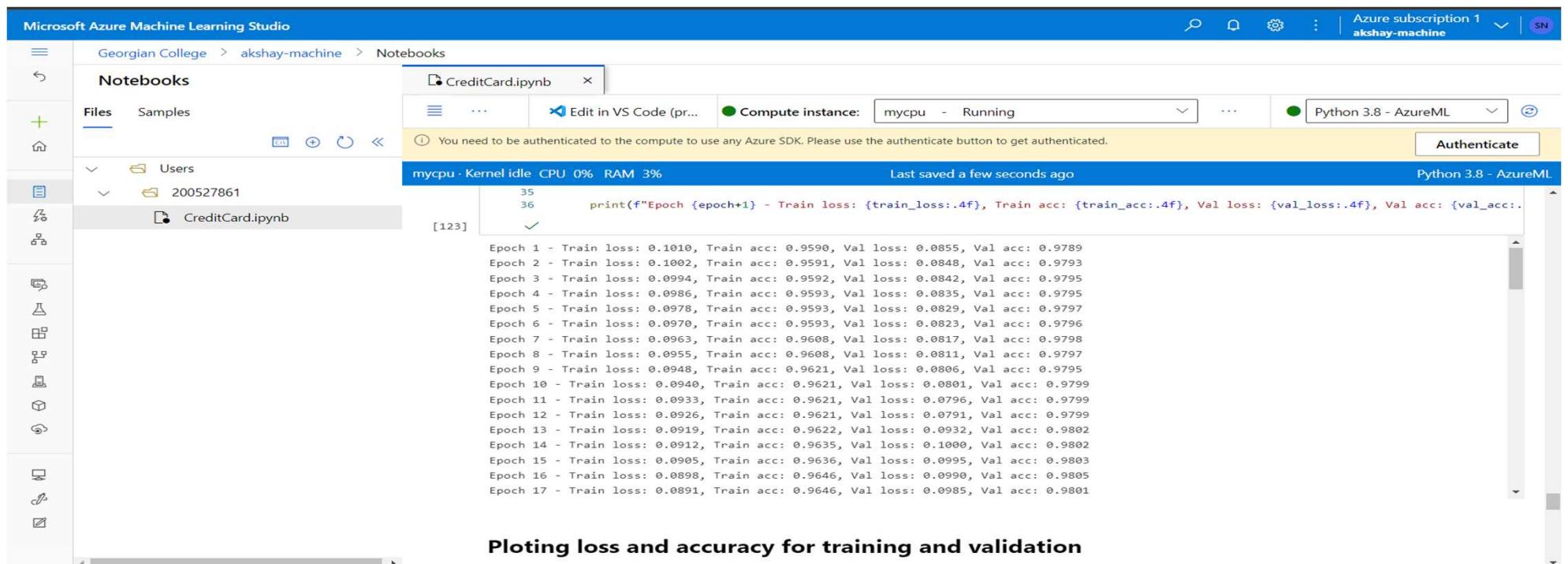
This screenshot shows the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes the URL 'ml.azure.com/ml/explorer/AzNB?wsid=/subscriptions/8a251600-4a2c-4102-b7b5-6388058c84b/resourcegroups/R1/providers/Microsoft.MachineLearningServices/worksp...', a search bar, and a user dropdown for 'Azure subscription 1 akshay-machine'. The main area is titled 'Georgian College > akshay-machine > Notebooks'. A 'Notebooks' section lists 'CreditCard.ipynb' and 'Samples'. The 'CreditCard.ipynb' notebook is currently selected, showing its code cell content. The code cell contains Python code for data preprocessing, model training, and evaluation. The status bar at the bottom indicates '[119]'.

```
1 # Split the dataset into features and target
2 X = credit.drop(['Class'], axis=1)
3 y = credit['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=0.2, 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
23
```

This screenshot shows the Microsoft Azure Machine Learning Studio interface, similar to the one above but with a different notebook content. The top navigation bar includes the URL 'ml.azure.com/ml/explorer/AzNB?wsid=/subscriptions/8a251600-4a2c-4102-b7b5-6388058c84b/resourcegroups/R1/providers/Microsoft.MachineLearningServices/worksp...', a search bar, and a user dropdown for 'Azure subscription 1 akshay-machine'. The main area is titled 'Georgian College > akshay-machine > Notebooks'. A 'Notebooks' section lists 'CreditCard.ipynb'. The 'CreditCard.ipynb' notebook is currently selected, showing its code cell content. The code cell contains Python code for training a neural network model. The status bar at the bottom indicates 'Last saved a few seconds ago'.

```
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%)



The screenshot shows the Microsoft Azure Machine Learning Studio interface. The top navigation bar includes 'Microsoft Azure Machine Learning Studio', 'Georgian College > akshay-machine > Notebooks', and 'Azure subscription 1 akshay-machine'. The main area is titled 'Notebooks' and shows a file named 'CreditCard.ipynb'. The notebook tab bar indicates it's running on 'mycpu - Running' with 'Python 3.8 - AzureML'. A message at the top of the notebook says 'You need to be authenticated to the compute to use any Azure SDK. Please use the authenticate button to get authenticated.' Below this, the notebook content shows code output for 17 epochs of training, with accuracy increasing from approximately 0.9590 to 0.9801. The bottom of the notebook has a section titled 'Plotting loss and accuracy for training and validation'.

```
Epoch 1 - Train loss: 0.1010, Train acc: 0.9590, Val loss: 0.0855, Val acc: 0.9789
Epoch 2 - Train loss: 0.1002, Train acc: 0.9591, Val loss: 0.0848, Val acc: 0.9793
Epoch 3 - Train loss: 0.0994, Train acc: 0.9592, Val loss: 0.0842, Val acc: 0.9795
Epoch 4 - Train loss: 0.0986, Train acc: 0.9593, Val loss: 0.0835, Val acc: 0.9795
Epoch 5 - Train loss: 0.0978, Train acc: 0.9593, Val loss: 0.0829, Val acc: 0.9797
Epoch 6 - Train loss: 0.0970, Train acc: 0.9593, Val loss: 0.0823, Val acc: 0.9796
Epoch 7 - Train loss: 0.0963, Train acc: 0.9608, Val loss: 0.0817, Val acc: 0.9798
Epoch 8 - Train loss: 0.0955, Train acc: 0.9608, Val loss: 0.0811, Val acc: 0.9797
Epoch 9 - Train loss: 0.0948, Train acc: 0.9621, Val loss: 0.0806, Val acc: 0.9795
Epoch 10 - Train loss: 0.0940, Train acc: 0.9621, Val loss: 0.0801, Val acc: 0.9799
Epoch 11 - Train loss: 0.0933, Train acc: 0.9621, Val loss: 0.0796, Val acc: 0.9799
Epoch 12 - Train loss: 0.0926, Train acc: 0.9621, Val loss: 0.0791, Val acc: 0.9799
Epoch 13 - Train loss: 0.0919, Train acc: 0.9622, Val loss: 0.0932, Val acc: 0.9802
Epoch 14 - Train loss: 0.0912, Train acc: 0.9635, Val loss: 0.1000, Val acc: 0.9802
Epoch 15 - Train loss: 0.0905, Train acc: 0.9636, Val loss: 0.0995, Val acc: 0.9803
Epoch 16 - Train loss: 0.0898, Train acc: 0.9646, Val loss: 0.0990, Val acc: 0.9805
Epoch 17 - Train loss: 0.0891, Train acc: 0.9646, Val loss: 0.0985, Val acc: 0.9801
```

**Plotting loss and accuracy for training and validation**

# ML MODELLING (VISUALIZING)

Microsoft Azure Machine Learning Studio

Georgian College > aishay-machine > Notebooks

CreditCard.ipynb

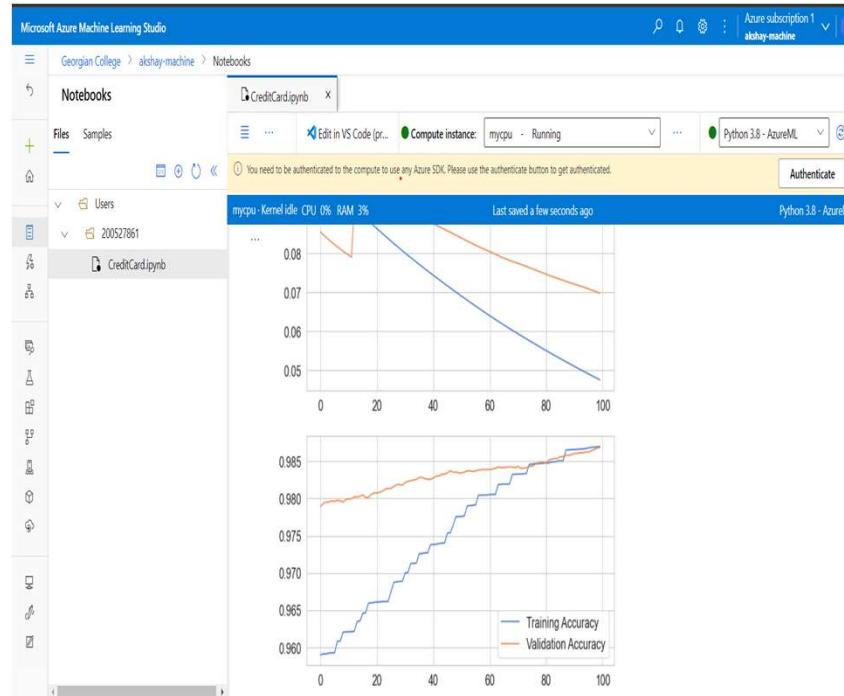
Authenticate

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

Plotting loss and accuracy for training and validation

```
1 import matplotlib.pyplot as plt
2
3 fig, axs = plt.subplots(2, figsize=(10, 10))
4 axs[0].plot(train_losses, label='Training Loss')
5 axs[0].plot(val_losses, label='Validation Loss')
6 axs[0].legend()
7 axs[1].plot(train_accs, label='Training Accuracy')
8 axs[1].plot(val_accs, label='Validation Accuracy')
9 axs[1].legend()
10 plt.show()
```

[127] ✓





THANK YOU