IBM SKILLSBUILD VIRTUAL INTERNSHIP PROJECT

NETWORK INTRUSION DETECTION SYSTEM

Presented By:
-Shaleen Bapna
-Techno India NJR Institute of Technology
-Computer Science & Engineering - Cybersecurity

edunet foundation

OUTLINE

- Problem Statement
- Proposed System/Solution
- Technology Used
- Algorithm & Deployment
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

Create a robust network intrusion detection system (NIDS) using machine learning. The system should be capable of analyzing network traffic data to identify and classify various types of cyber-attacks (e.g., DoS, Probe, R2L, U2R) and distinguish them from normal network activity. The goal is to build a model that can effectively secure communication networks by providing an early warning of malicious activities.

PROPOSED SOLUTION

The proposed solution is to develop a Network Intrusion Detection System (NIDS) using machine learning techniques to detect and classify malicious network traffic. The solution uses a labeled dataset containing instances of both normal and attack traffic, trains a classification model to distinguish between them, and integrates the solution into a cloud-based environment for scalable deployment. The system aims to provide early warnings of cyber-attacks like DoS, Probe, R2L, and U2R, enhancing the security of communication networks.

Key components of the solution include:

- **Data Preprocessing**: Cleaning, encoding, and normalizing network traffic data.
- **Model Training**: Using classification algorithms to learn attack patterns.
- **Evaluation**: Validating the model's performance using accuracy, F1-score, and confusion matrix.
- **Cloud Deployment**: Hosting the model on IBM Cloud Lite using watsonx.ai Runtime for prediction APIs.

SYSTEM APPROACH

The system is built on a modular approach using IBM Cloud Lite services to ensure accessibility and scalability. The key aspects include:

System Requirements:

- Python 3.x environment
- Jupyter Notebook via watsonx.ai Studio
- IBM Cloud Object Storage for dataset handling
- IBM watsonx.ai Runtime for model deployment

Cloud Services:

- IBM Cloud Object Storage
- watsonx.ai Studio
- watsonx.ai Runtime (Machine Learning)

ALGORITHM & DEPLOYMENT

Algorithm Selection:

- A **Random Forest Classifier** was selected due to its high accuracy and robustness in handling tabular classification problems with mixed feature types (numerical + categorical).

Data Input:

- Features such as protocol type, source/destination bytes, connection flags, and host behavior metrics were used.
- The output label is either 'normal' or 'anomaly'.

Training Process:

- The dataset was cleaned and label-encoded.
- 80/20 train-validation split was used.
- The model was trained using RandomForestClassifier from scikit-learn with 100 estimators.

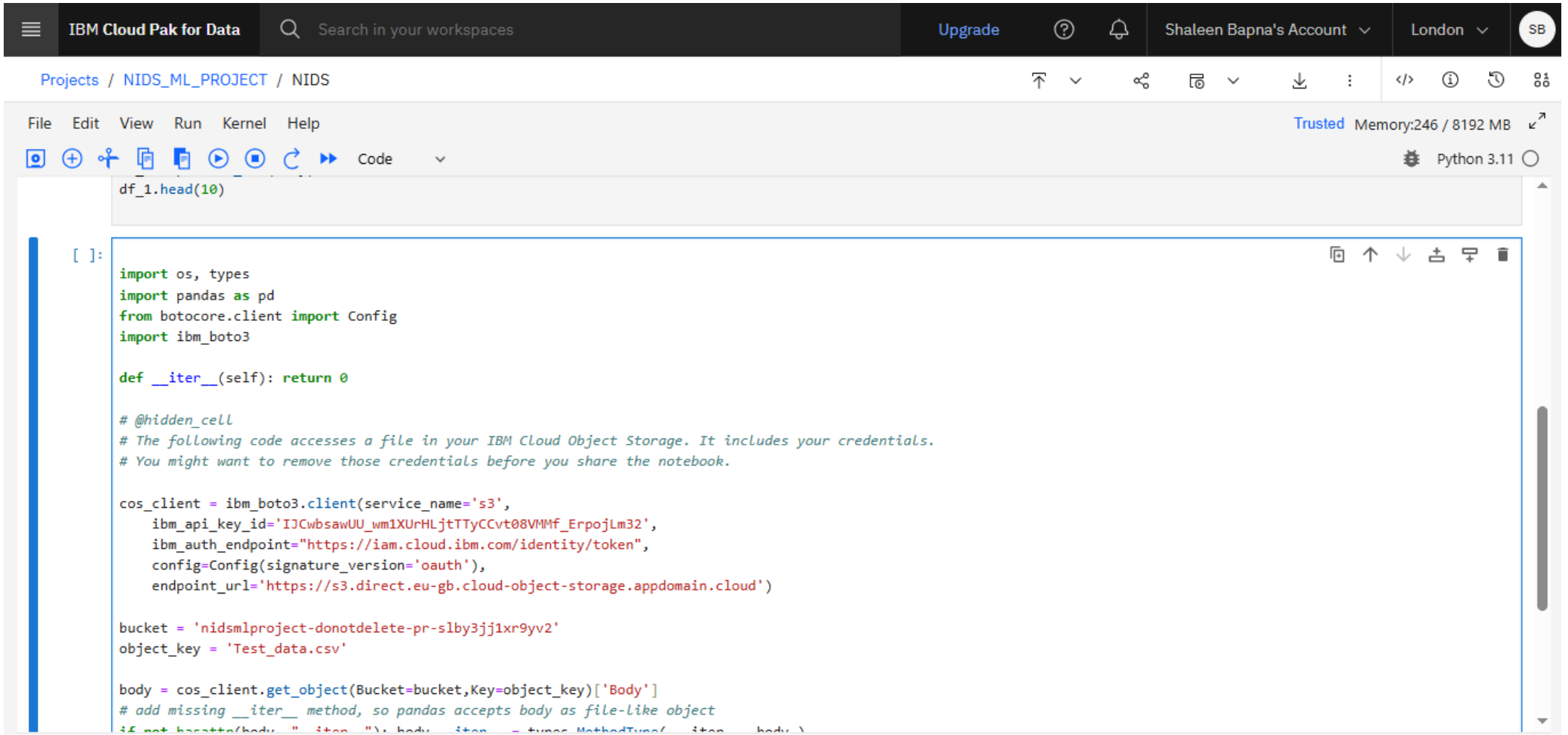
Prediction Process:

- The trained model was used to classify unseen traffic in the test dataset.
- Predictions were exported and visualized to evaluate anomaly detection.

Deployment:

- The model was saved and can be deployed via **IBM watsonx.ai Runtime** to expose an API for real-time traffic classification.
- IBM Cloud provides a seamless environment to integrate the model with external systems or dashboards.

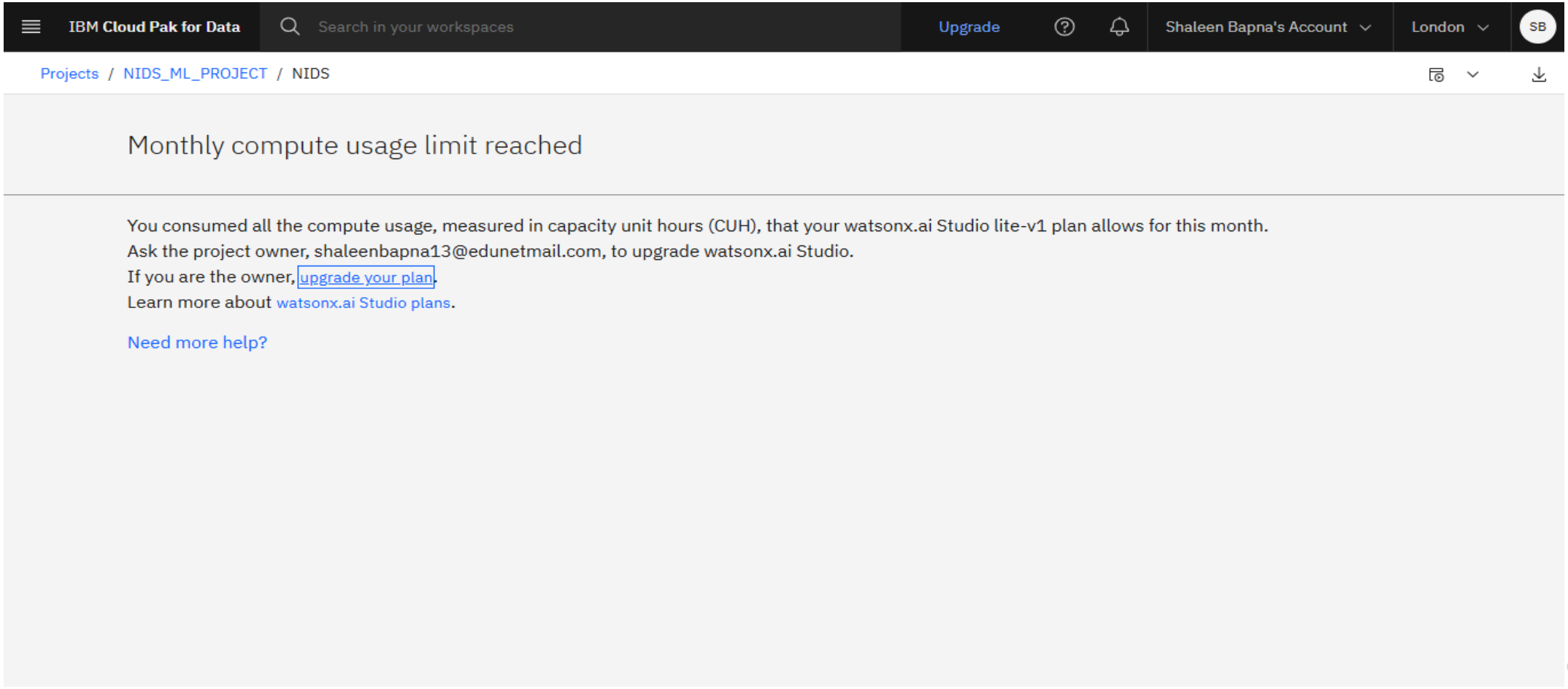
RESULT



edunet foundation

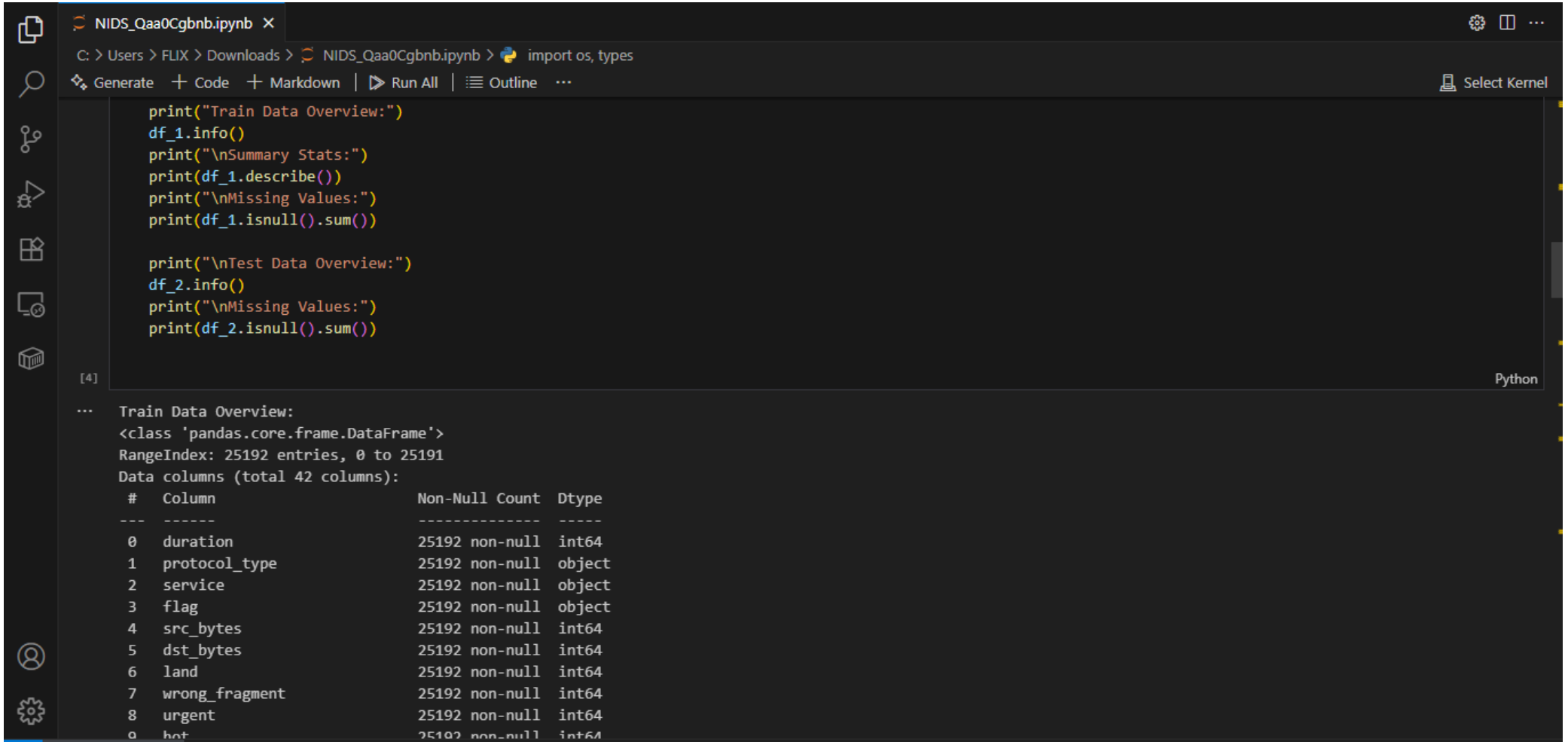
RESULT

Sir the rest of the screenshots are from vs code as I was not able to take the remaining ones from the IBM platform due to this issue, I kindly request you to consider them as I completed the same on IBM platform too.

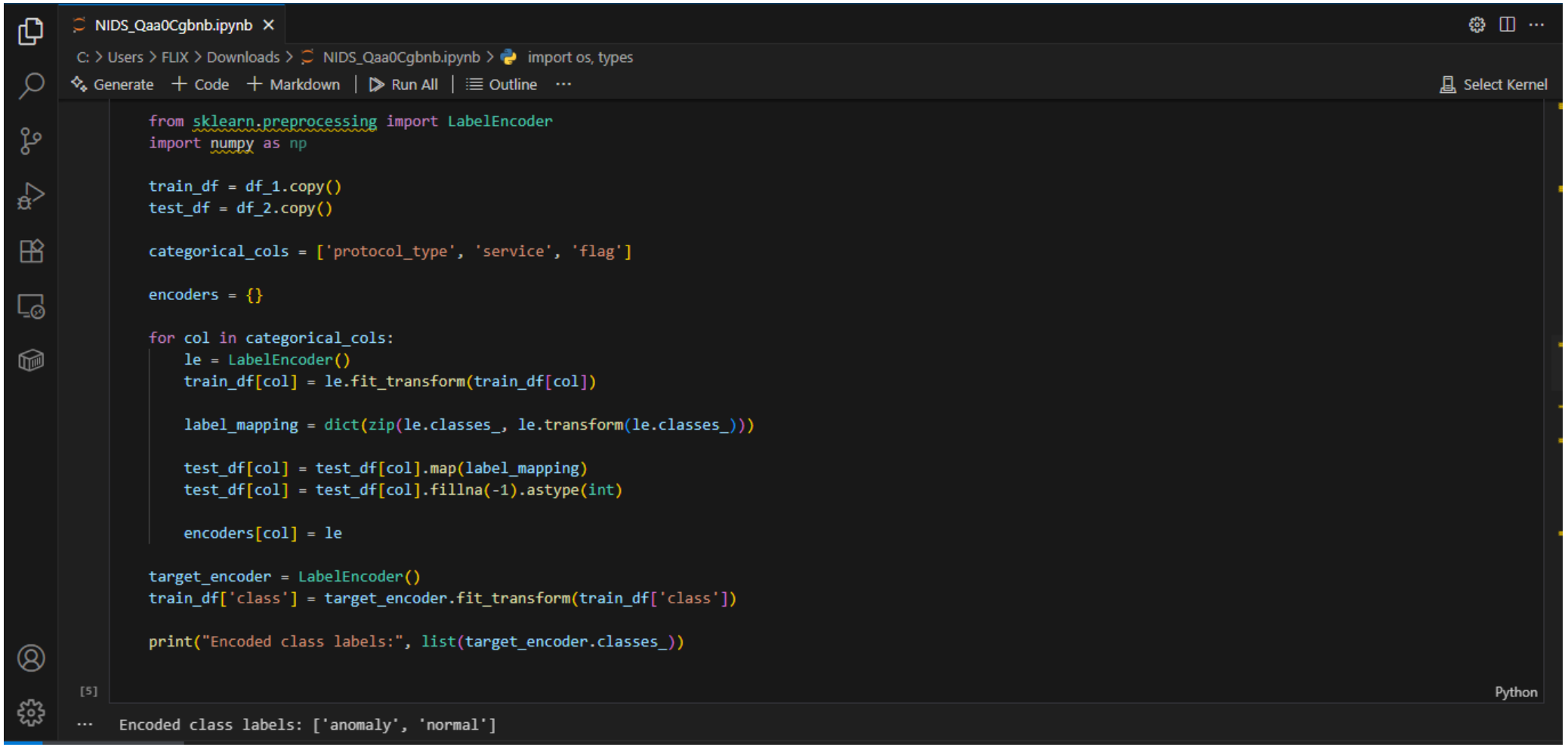


edunet foundation

RESULT

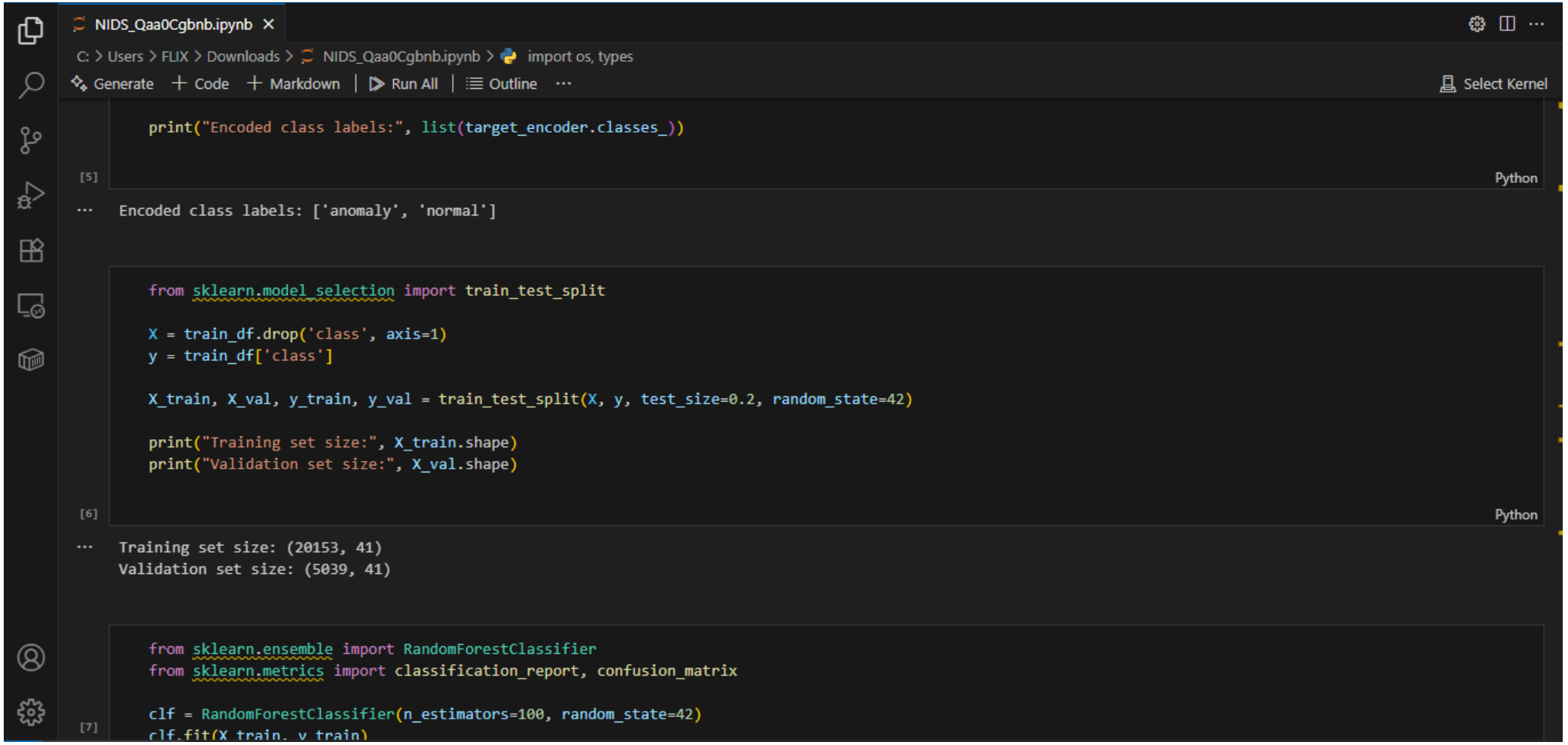

NIDS_Qaa0Cgbnb.ipynb
C: > Users > FLIX > Downloads > NIDS_Qaa0Cgbnb.ipynb > import os, types
Generate
Code
Markdown
Run All
Outline
Select Kernel
print("Train Data Overview:")
df_1.info()
print("\nSummary Stats:")
print(df_1.describe())
print("\nMissing Values:")
print(df_1.isnull().sum())
print("\nTest Data Overview:")
df_2.info()
print("\nMissing Values:")
print(df_2.isnull().sum())
[4]
Python
Train Data Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25192 entries, 0 to 25191
Data columns (total 42 columns):
Column Non-Null Count Dtype
0 duration 25192 non-null int64
1 protocol_type 25192 non-null object
2 service 25192 non-null object
3 flag 25192 non-null object
4 src_bytes 25192 non-null int64
5 dst_bytes 25192 non-null int64
6 land 25192 non-null int64
7 wrong_fragment 25192 non-null int64
8 urgent 25192 non-null int64

RESULT

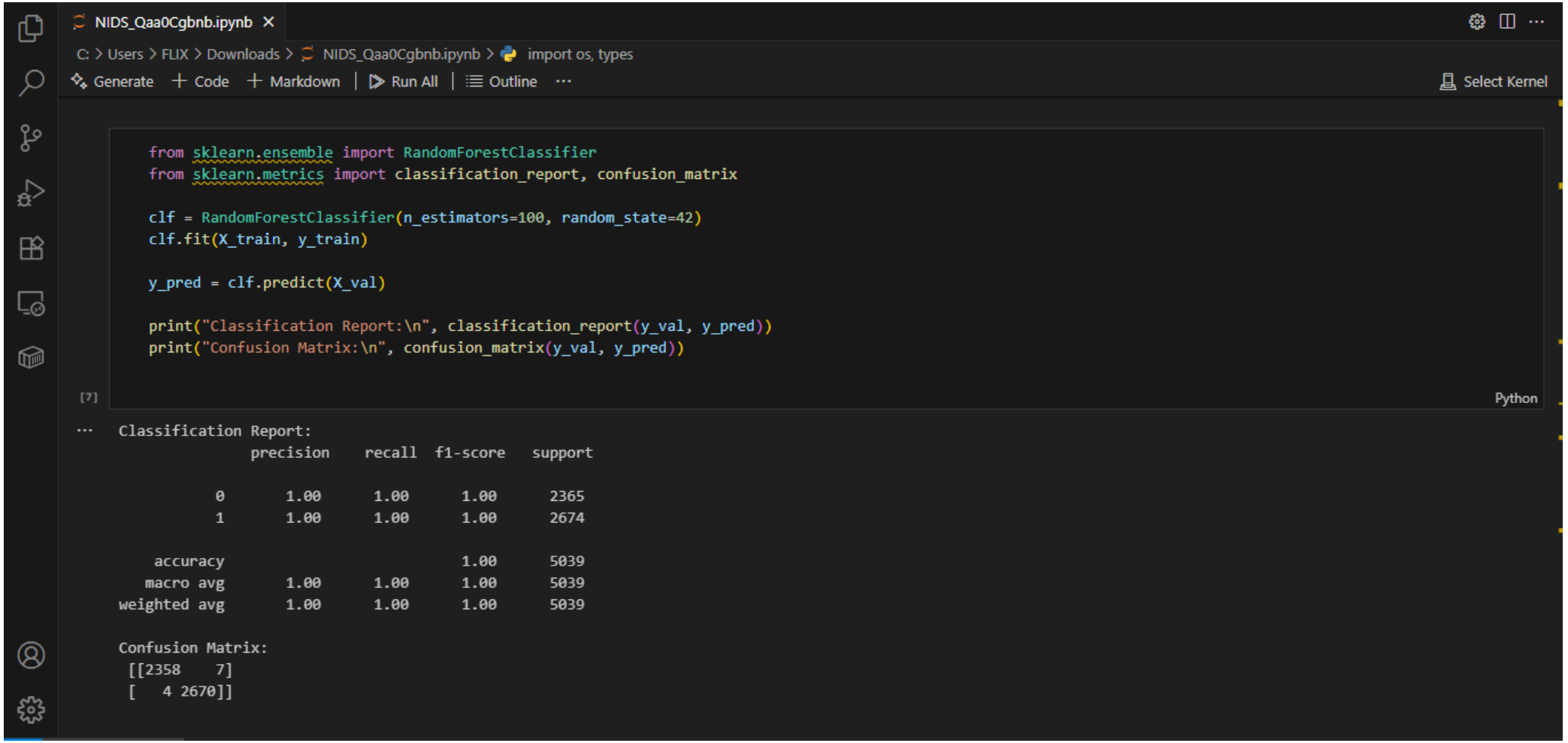


edunet foundation

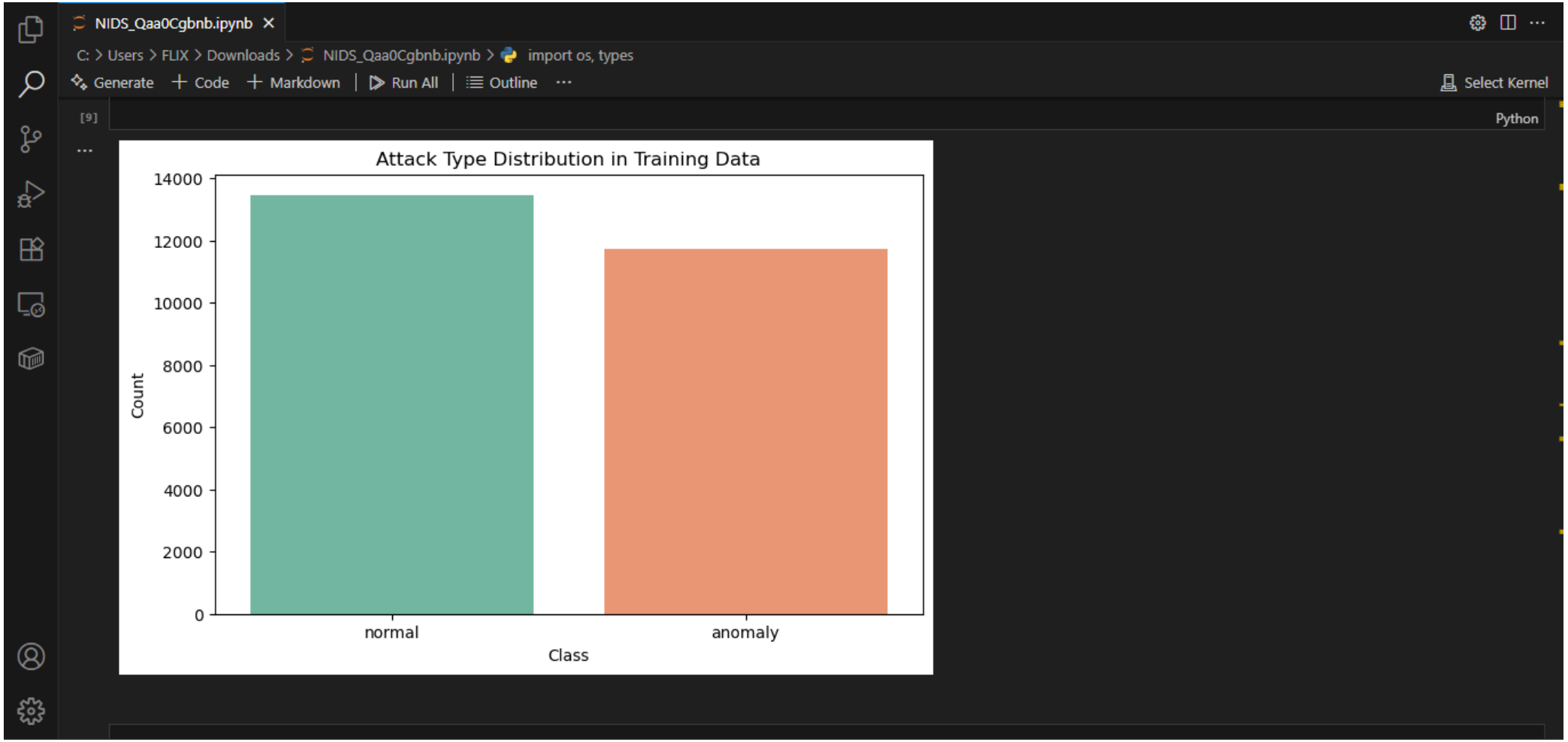
RESULT



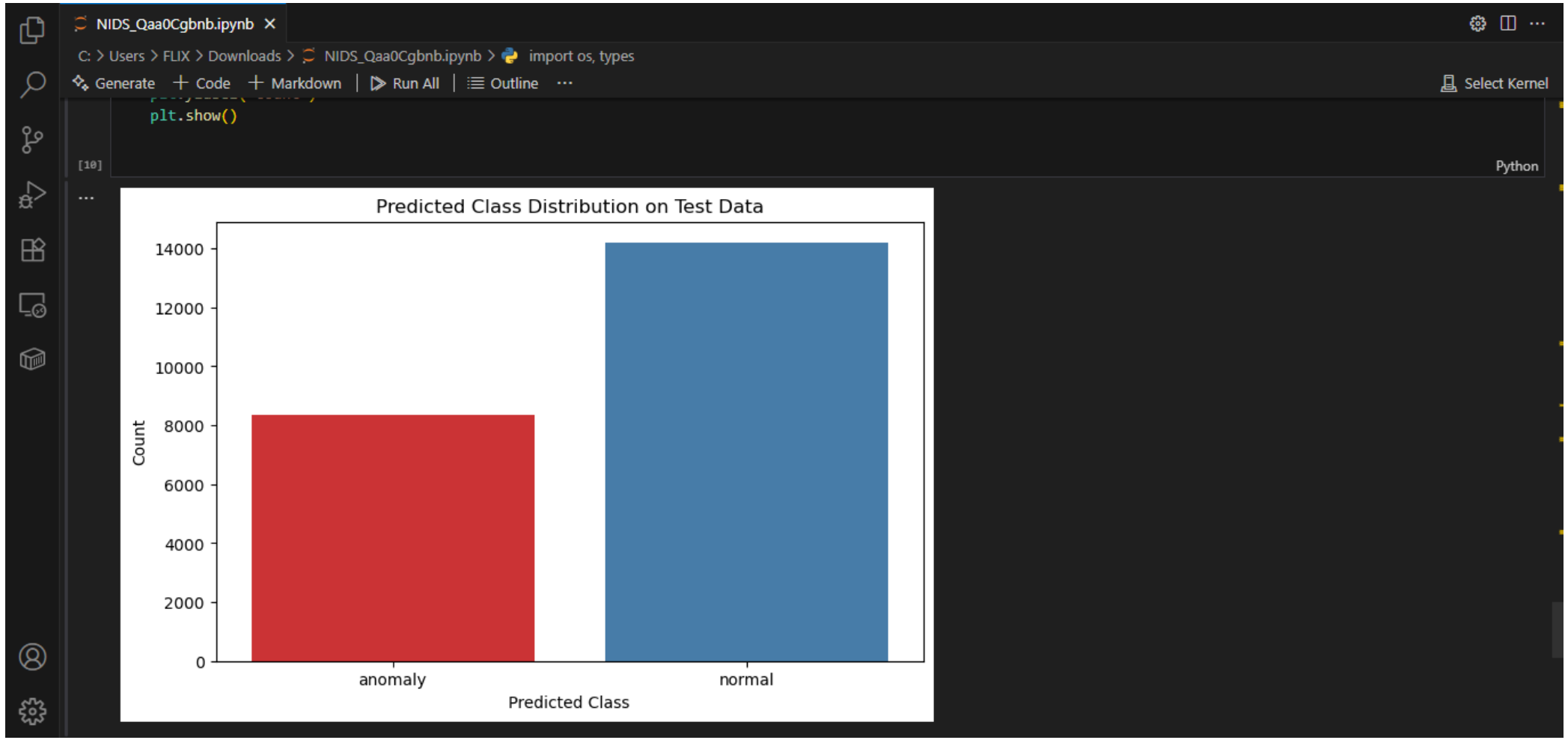
RESULT


NIDS_Qaa0Cgbnb.ipynb
C: > Users > FLIX > Downloads > NIDS_Qaa0Cgbnb.ipynb > import os, types
Generate
Code
Markdown
Run All
Outline
Select Kernel
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_val)
print("Classification Report:\n", classification_report(y_val, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_val, y_pred))
[7]
Python
Classification Report:
precision recall f1-score support
0 1.00 1.00 1.00 2365
1 1.00 1.00 1.00 2674
accuracy 1.00 5039
macro avg 1.00 1.00 1.00 5039
weighted avg 1.00 1.00 1.00 5039
Confusion Matrix:
[[2358 7]
[4 2670]]

RESULT



RESULT



CONCLUSION

The developed Network Intrusion Detection System (NIDS) using machine learning successfully demonstrates how cyber threats can be identified and mitigated in real-time through data-driven approaches. By training a Random Forest classifier on labeled network traffic data, the system achieved high accuracy in detecting and classifying intrusions such as DoS, R2L, U2R, and Probe attacks. The integration with IBM Cloud Lite services enabled scalable, cloud-based development and deployment using watsonx.ai Studio and Runtime. This solution highlights the effectiveness of machine learning in strengthening network security infrastructure and providing early alerts for potential cyber threats

FUTURE SCOPE

- Real-Time Traffic Integration
- Multi-Class Classification
- Ensemble & Deep Learning Models
- Dashboard/Visualization Tool
- Edge Computing & IoT Security
- Cross-Dataset Evaluation

REFERENCES

- GitHub Link: https://github.com/Shaleen-flix/NIDS

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

edunet foundation

IBM CERTIFICATIONS

IBM **SkillsBuild** Completion Certificate

This certificate is presented to

Shaleen Bapna

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 22 Jul 2025 (GMT) **Learning hours:** 20 mins

THANK YOU

edunet foundation