Project Reflection

AI Powered Threat Detection for Cloud Network

Group Number: 1

Unit: <COIT13236 or COIT20265>

Student 1: Sonam Pelden (12248722)

Project Mentor: Dr. Fariza Sabrina

Date: 02/02/2025

CQUniversity Australia

# Tasks Completed

|  |  |  |
| --- | --- | --- |
| **Task** | **Description** | **Deliverable(s)** |
| Feature Engineering | Implemented TF-IDF vectorization to convert textual data into numerical representations for model training. Stored the vectorizer for deployment (Liu et al., 2018). | TF-IDF Vectorizer model (vectorizer.pkl). |
| Label Encoding | Encoded categorical labels into numerical format for machine learning models. Stored the label encoder for later use in model prediction (Névéol et al., 2018). | Label encoder model (label\_encoder.pkl). |
| Model Building | Developed multiple machine learning models: Random Forest, SVM, Logistic Regression, and ANN. Trained and validated them using the processed dataset (Ao et al., 2019). | Trained models, including Random Forest, SVM, Logistic Regression, and ANN. |
| OTP Verification System | Implemented OTP verification for user authentication, ensuring secure access. Generated OTP, stored it temporarily, and sent it via email (Kumar et al., 2024). | Functional OTP system integrated with email notifications. |
| Admin Page Development | Created an admin dashboard to manage users, review threats, and block malicious users (Hasas et al., 2024) whereby Implemented employee management, including assigning employee IDs. | Fully functional admin panel with user and threat management. |
| Threat Detection Logging | | Integrated system to log flagged messages into the database (AdminThreat) and implemented automatic user blocking for multiple flagged submissions (Wang, Chen & Yu, 2022). | Threat logging database (AdminThreat), auto-blocking system for malicious users. |
| Data Visualization for Admin | | Implemented real-time visualization of threats using Matplotlib, generating bar charts for detected threats (Bezerra et al. 2019). | Admin visualization page displaying threat trends and statistics. |
| Employee Management | | Developed functionality for admins to add, edit, and delete employee records. Implemented ID assignment system with automatic email notifications (Hasas et al. 2024). | Employee management system integrated with admin panel. |
| AWS VPC Deployment | | Configured a secure AWS Virtual Private Cloud (VPC) to ensure network isolation and efficient cloud resource management (Miorelli et al. 2021). | AWS VPC deployed with optimized settings for secure cloud hosting. |
| Elastic IP Configuration | | Set up an AWS Elastic IP for the application to ensure persistent access, even during instance restarts | Elastic IP assigned for stable external connectivity. |

# Newly Developed Skills

|  |  |  |
| --- | --- | --- |
| **Skills** | **Related Task** | **Justification** |
| TF-IDF Vectorization | Feature Engineering | Learned how to apply TF-IDF to convert text into numerical features, which improved model accuracy by representing text data effectively (Liu et al. 2018). |
| Label Encoding | Feature Engineering | Developed expertise in encoding categorical variables into numerical form to make data compatible with machine learning models (Névéol et al. 2018). |
| Deep Learning with ANN | Model Building | Gained hands-on experience building and training an Artificial Neural Network (ANN), which helped improve classification accuracy (Ao et al. 2019). |
| OTP Authentication | OTP Verification System | Developed skills in secure user authentication by implementing OTP verification, ensuring a safer login process (Kumar et al. 2024). |
| Threat Visualization | | Data Visualization for Admin | Learned how to generate visual insights from threat detection data, using Matplotlib for real-time monitoring of security threats (Bezerra et al. 2019). |
| AWS VPC Configuration | | AWS VPC Deployment | Understood how to deploy a secure Virtual Private Cloud (VPC) for isolating and managing cloud-based applications efficiently (Miorelli et al. 2021). |
| Employee ID Management | | Employee Management | Implemented logic to generate and assign employee IDs dynamically, ensuring seamless user tracking and identification (Hasas et al. 2024). |

# Application of Existing Skills

|  |  |  |
| --- | --- | --- |
| **Skills** | **Related Task** | **Justification** |
| Flask Development | Admin Page Development | Used prior experience in Flask to build a structured and interactive admin dashboard for threat monitoring and user management (Goh et al. 2022). |
| SQLAlchemy Database Management | Threat Detection Logging | Applied knowledge of SQLAlchemy to design and implement databases for logging detected threats and managing user records efficiently (Mihaescu & Popescu, 2021). |
| Machine Learning Models | Model Building | Utilized existing expertise in machine learning to implement and fine-tune Random Forest, SVM, and Logistic Regression for better classification performance (Ao et al. 2019). |
| Email Integration | OTP Verification System | Applied previous experience in Flask-Mail to configure email notifications for OTP authentication and user communication (Hasas et al. 2024). |
| Data Preprocessing | | Feature Engineering | Used knowledge of text preprocessing techniques, such as stopword removal and lemmatization, to clean data before feeding it into models (Névéol et al. 2018). |
| AWS Elastic IP Setup | | Elastic IP Configuration | Used cloud computing experience to configure AWS Elastic IP for maintaining a stable public IP address for the application (Cloud, 2011. |
| Matplotlib for Visualization | | Data Visualization for Admin | Leveraged existing data visualization skills to create informative threat analysis charts, enhancing the usability of the admin panel (Sial et al. 2021, p. 277). |

# Contributions of Team Members

1. Sonam Pelden
2. Akshitha Ishan Ranasinghe Walgama Ranasinghe Arachchilage
3. Jay Dilipbha Yadav
4. Harshkumar Dharmendrabhai Patel

I believe I was the best contributor to the project because of my strong leadership and organizational skills, which ensured the continuous execution of our jobs. I took the initiative to design a detailed project plan, assign duties based on each team member's strengths, and regularly monitored progress to keep everyone on track. When obstacles arose, such as integrating various application components, I encouraged talks and coordinated efforts to rapidly resolve issues. By organizing team files, communicating clearly, and keeping everyone informed, I established a collaborative and cohesive workplace. I also ensured that the team remained motivated and involved throughout the project, serving as an example of dedication and commitment. My ability to identify potential roadblocks and handle them proactively was critical to the project's success, and I feel others might benefit from my approach to leadership, communication, and issue solving.

# Reflections on Project Experience

Experiencing how well models like Random Forest, SVM, Logistic Regression, and ANN worked gave me important insights into classification techniques, and I especially enjoyed feature engineering and model development, which required experimenting with various machine learning techniques, assessing their performance, and improving them (Ao et al., 2019; Ghiasi & Zendehboudi, 2020). This project was a challenging but rewarding experience that improved my technical and coordination skills.

In addition, the project included real-world Flask web application deployment on AWS, which made it more applicable and industry-relevant (Goh et al., 2022). I was able to apply TF-IDF vectorization and label encoding effectively thanks to my background in machine learning and Python, as well as my coursework in text classification and natural language processing (Liu et al., 2018; Névéol et al., 2018). I was also able to design and manage SQLAlchemy models thanks to my database management skills, and my understanding of cloud computing helped set up AWS EC2 instances, security groups, and VPCs to ensure a smooth deployment (Miorelli et al., 2021).

However, the project presented several challenges, such as optimizing the threat detection model to detect malicious messages while minimizing false positives, which I resolved by adjusting hyperparameters and enhancing text preprocessing techniques (Symeonidis et al., 2018; Liu et al., 2018). Another significant challenge was module integration because team members were working on different components, which I resolved by organizing coordination efforts, gathering files, keeping team members informed, and making sure the integration went smoothly.

The hardest part of the project was real-time threat identification and automatic user blocking while maintaining user experience and system speed. Numerous test runs were needed to execute OTP verification and email alerts securely (Hasas et al. 2024). The AWS deployment environment's machine learning backend and database need proper Flask web application communication. I resolved these issues by carefully studying logs, making incremental tweaks, and using AWS monitoring tools after patience and trial-and-error and this experience has improved my machine learning model deployment, cloud computing, database management, and full-stack web application development skills for future projects (Popescu 2021; Miorelli et al. 2021). I also enhanced my problem-solving and project management skills to tackle more complex projects.

This project highlighted the value of early logging and monitoring system implementation as well as structured problem-solving. In addition to making debugging simpler, efficient logging increased the system's overall dependability (Wang, Chen & Yu 2022). I also discovered that successful projects and smooth integration depend on regular teamwork and communication.

I would optimise model performance early if I could do this project again. To make AWS integration simpler, I would automate several deployment methods. I enjoy the project and think our approach was well-structured. One change I would make to reduce troubleshooting time is improving logging and monitoring early in deployment. This project was a terrific hands-on learning experience that improved my technical skills, problem solving abilities and teamwork.

# Portfolio Contribution

My contributions to an AI-powered Cyber Threat Detection System, where I implemented an AWS backend, carried out feature engineering, and created models for text-based malware classification and real-time threat detection, are highlighted in my public portfolio, which highlights my proficiency in machine learning, cybersecurity, and full-stack web development. I used cutting-edge machine learning techniques, such as label encoding and TF-IDF vectorization, to improve model accuracy. In web development, I created Flask-based apps, used AWS services like EC2, VPC, and Elastic IPs to establish safe, scalable cloud architecture, and integrated SQLAlchemy for database administration. These practical AI and cloud projects showcase my technological expertise, problem-solving abilities, and capacity to provide comprehensive, sector-relevant solutions.

Link for my ePortfolio: <https://portfolium.com/SonamPelden1>

**Reference**

Ao, Y, Li, H, Zhu, L, Ali, S, & Yang, Z. (2019). *The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling*. Journal of Petroleum Science and Engineering, 174, 776–789. https://doi.org/10.1016/j.petrol.2018.11.067

Bezerra, A., Silva, I., Guedes, L.A., Silva, D., Leitão, G. & Saito, K 2019, 'Extracting Value from Industrial Alarms and Events: A Data-Driven Approach Based on Exploratory Data Analysis,' *Sensors*, 19(12), p. 2772. DOI: 10.3390/s19122772.

Chavlis, S. & Poirazi, P. 2021, 'Drawing inspiration from biological dendrites to empower artificial neural networks,' *Current Opinion in Neurobiology*, 70, pp. 1–10. DOI: 10.1016/j.conb.2021.04.007.

Cloud, A.E.C, 2011. ‘Amazon web services’. *Retrieved November*, *9*(2011), p.2011. viewed 22 January 2025, <https://d1wqtxts1xzle7.cloudfront.net/27741916/running-a-website-on-amazon-ec2-libre.pdf?1390872807=&response-content-disposition=inline%3B+filename%3DAmazon_Web_Services.pdf&Expires=1738392978&Signature=FSFF2MFAfeZ3CVSL1e7cP71aOu-1QG82RYQBEl447iRGtzPJaPdeZG7VcGMw66E6gTFrdweqPOSwy5I7nClltL2mP9X1oJmaBeCUUKyuNFO5yoYPpAEMMgZz20rV4Z-TvBVsj7dYTdQttIlpqsCRM4sZ1sRnhMjxcvQMq89hjZf0kzh-rwdvREJ5VRcJwst7WI4d8lDFc0fz48s4ldT13lOmQqELjMt35d-BfrUpnLeuSfCv9fKaivj1Ebjw83z8FxHfqA~GAaccBAL6FrjgJLS1jEHZ7DjIIOvpRP2zDvOIHI1X49b0PRVa7eETAwzH0LdsixQx~OUhnM21HdD~8Q__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA>

Ghiasi, M.M. & Zendehboudi, S. 2020, 'Application of decision tree-based ensemble learning in the classification of breast cancer,' *Computers in Biology and Medicine*, 128, p. 104089. DOI: 10.1016/j.compbiomed.2020.104089.

Goh, H.-A., Ho, C.-K., & Abas, F. S. (2022). *Front-end deep learning web apps development and deployment: a review*. Applied Intelligence, *53*. https://doi.org/10.1007/s10489-022-04278-6

Hasas, A., Zarinkhail, M. S., Hakimi, M., & Quchi, M. M. (2024*). Strengthening Digital Security: Dynamic Attack Detection with LSTM, KNN, and Random Forest*. Journal of Computer Science and Technology Studies, 6(1), 49–57. https://doi.org/10.32996/jcsts.2024.6.1.6

Hicks, S.A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M.A., Halvorsen, P. & Parasa, S., 2022, 'On evaluation metrics for medical applications of artificial intelligence,' *Scientific Reports*, 12(1). DOI: 10.1038/s41598-022-09954-8.

Kibria, H.B. & Matin, A. 2022, 'The severity prediction of the binary and multi-class cardiovascular disease − A machine learning-based fusion approach,' *Computational Biology and Chemistry*, 98, p. 107672. DOI: 10.1016/j.compbiolchem.2022.107672.

Kumar, N., Kamje, R., Landge, P., Shitole, S. and Takale, B., 2024, ‘OTP-Based Authentication System’, <https://www.researchgate.net/profile/Neelam-Labhade-Kumar/publication/387079266_OTP-Based_Authentication_System/links/675face4ebc8f9797036a2e2/OTP-Based-Authentication-System.pdf>

Liu, C., Sheng, Y., Wei, Z., & Yang, Y. (2018, August 1). *Research of Text Classification Based on Improved TF-IDF Algorithm*. IEEE Xplore. https://doi.org/10.1109/IRCE.2018.8492945

Mihaescu, M.C. & Popescu, P.S. 2021, 'Review on publicly available datasets for educational data mining,' *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, 11(3). DOI: 10.1002/widm.1403.

Miorelli, R., Kulakovskyi, A., Chapuis, B., D’almeida, O. and Mesnil, O., 2021, 'Supervised learning strategy for classification and regression tasks applied to aeronautical structural health monitoring problems,' *Ultrasonics*, 113, p. 106372. DOI: 10.1016/j.ultras.2021.106372.

Névéol, A., Dalianis, H., Velupillai, S., Savova, G., & Zweigenbaum, P. (2018). *Clinical Natural Language Processing in languages other than English: opportunities and challenges*. Journal of Biomedical Semantics, 9(1). https://doi.org/10.1186/s13326-018-0179-8

Oliynyk, D., Mayer, R. & Rauber, A. 2023, 'I know what you trained last summer: a survey on stealing machine learning models and defences,' *ACM Computing Surveys*, 55(14s), pp. 1–41. DOI: 10.1145/3595292.

Perkins, J. (2010). *Python Text Processing with NLTK 2.0 Cookbook*.

Popov, A. 2023, 'Feature engineering methods,' in *Elsevier eBooks*, pp. 1–29. DOI: 10.1016/b978-0-323-85955-4.00004-1.

Rácz, A., Bajusz, D. & Héberger, K. 2021, 'Effect of dataset size and Train/Test split ratios in QSAR/QSPR Multiclass Classification,' *Molecules*, 26(4), p. 1111. DOI: 10.3390/molecules26041111.

Sial, A.H, Rashdi, S.Y.S & Khan, A.H, 2021, ‘Comparative analysis of data visualization libraries Matplotlib and Seaborn in Python’. *International Journal*, *10*(1), pp.277-281. viewed 23 January 2025, <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse391012021.pdf>

Singh, A., & Paul, S. (2020). *Hands-On Python Deep Learning for the Web Integrating neural network architectures to build smart web apps with Flask, Django, and TensorFlow*. Packt Publishing Limited.

Singla, A., & Sandarsh Chavalmane. (2023). *Automating Model Deployment: From Training to Production.* Journal of Knowledge Learning and Science Technology ISSN 2959-6386 (Online), 2(3), 340–334. <https://doi.org/10.60087/jklst.vol2.n3.p347>

Smyth, P. (2018). *Creating Web APIs with Python and Flask*. The Programming Historian, 7. https://doi.org/10.46430/phen0072

Symeonidis, S., Effrosynidis, D. & Arampatzis, A. 2018, 'A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis,' *Expert Systems With Applications*, 110, pp. 298–310. DOI: 10.1016/j.eswa.2018.06.022.

Wang, B.-X., Chen, J.-L., & Yu, C.-L. (2022). *An AI-Powered Network Threat Detection System*. IEEE Access, 10, 54029–54037. https://doi.org/10.1109/access.2022.3175886