National College of Ireland

# Understanding the influence of behaviour monitoring tools towards mitigating insider threats in hybrid working environments through behavioural analytics

MSc Research Project
MSCCYB1JAN25B_O

Sravan Kumar Murimuri
Student ID: 23358556

School of Computing
National College of Ireland

Supervisor: Jawad Salahuddin

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

**Student Name:** Sravan Kumar Murimuri

**Student ID:** 23358556

**Programme:** MSCCYB1JAN25B_O **Year:** 2025 - 2026

**Module:** MSc (Research) Practicum Part 2

**Lecturer:** Jawad Salahuddin

**Submission Due Date:** 28/01/2026

**Project Title:** Understanding the influence of behaviour monitoring tools towards mitigating insider threats in hybrid working environments through behavioural analytics

**Word Count:** 591 **Page Count:** 2

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: M. Sravan Kumar

Date: 28/01/2026

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple | □ |

| | |
|---|---|
| copies) | |
| Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| Office Use Only | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Understanding the influence of behaviour monitoring tools towards mitigating insider threats in hybrid working environments through behavioural analytics

**Student Name: Sravan Kumar Murimuri**
**Student ID: 23358556**
**MSCCYB1JAN25B_O**

# 1 System Overview

This configuration guide contains detailed instructions on deploying a machine learning-oriented insider threat detection system that is intended to work with a hybrid workplace. The system integrates behavioural analytics with multifaceted ML-based models, such as Random Forest, Stacking Classifier, and XGBoost with SHAP, to respond and combat insider threats based on the network flow analysis and user behaviour profiling. SHAP is important for interpreting ML models, providing local as well as global, theoretically aligned insights into feature importance (Bobadilla et al., 2024). This remedy solves fundamental security threats of a distributed workforce that works outside the office.

# 2 Technical Requirements

## 2.1 Software Dependencies

- Python version 3.8 or more with scientific code libraries such as NumPy, Pandas and Scikit-learn.
- Machine learning: XGBoost, Imbalanced-learn to use SMOTE, and SHAP.
- Analytics dashboard Data visualisers: Matplotlib and Seaborn.
- Monitoring Network: Network vistas sensitive to the flow of data.

# 3 Data Collection Configuration

Turn on the network flow logs to record user transactions between hybrid working environments. Data attributes that are essential are packet-level measurements, type of protocol, duration of flow, bytes transferred in forward and back directions, number of packets, and inter-arrival time. The system will need 86 different attributes, one of them being flow identifiers, source/destination, and timestamp data, as well as derived behavioural metrics that the system uses to effectively detect (Kaggle, 2023). To improve this, install a continuous monitoring system where the data is pre-processed with automated data pre-process routines to accept the null values, remove duplicates and flag statistical outliers by a detected threshold in order to maintain integrity and quality of the data.

# 4 Model Configuration

Implement three complementary machine learning systems in parallel mode that is used to detect the threats. Train a Random Forest classifier that has ensemble decision trees that are

optimised to find behavioural patterns (Barreñada et al., 2024). Use the Stacking Classifier architecture that involves the use of Logistic Regression, Support Vector Classifier and Decision Tree (Xuan et al., 2025). The baseline category is XGBoost, which involves the use of gradient boosting and regularisation parameters (Khan, Chaudhari and Chandra, 2023). In the case of categorical features (such as Flow ID, Source IP, and Timestamp variables), consider label encoding. Filter using correlation to eliminate highly correlated features that are stronger than 0.95. Do stratified data splitting and have an 80-20 split of train-test. Use the SMOTE synthetic oversampling method to balance the classes of the benign and malicious activities, obtain an equal sample in the training datasets, and overcome the problem of class imbalance. SHAP (SHapley Additive exPlanations) interpreter of XGBoost helps to create summary plots where features contribute towards their values, which seem to be Flow Bytes/s (mean SHAP 4.0), Fwd IAT Total (mean SHAP 2.6). Assess the use of waterfall plots to explicitly predict the threats to be used in operational decision-making.

# 5 Operational Guidelines

Implement ongoing monitoring systems in a hybrid work setup through automated anomaly detection systems. XGBoost is more operational with 99.20% classification accuracy and 100% recall rates, which cannot miss a single instance of malicious activity. Have a regular model retraining, using new behavioural trends. Install multi-level escalation procedures of identified threats that come with automated reporting to the security operational teams. If false positive rates are observed, continuously monitor them, and dynamically vary classification thresholds to trade-offs between detection sensitivity to suit overall system performance to organisational security needs.

# 6 References


Barreñada, L., Dhiman, P., Timmerman, D., Anne-Laure Boulesteix and Calster, B.V. (2024). Understanding overfitting in random forest for probability estimation: a visualization and simulation study. *Diagnostic and Prognostic Research*, 8(1). doi:https://doi.org/10.1186/s41512-024-00177-1.

Bobadilla, A.V.P., Schmitt, V., Maier, C.S., Mensing, S. and Stodtmann, S. (2024). Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. Clinical and Translational Science, [online] 17(11). doi:https://doi.org/10.1111/cts.70056.

Kaggle (2023). *Insider Threat*. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/nuhamintamirat/insider-threat.

Khan, A.A., Chaudhari, O. and Chandra, R. (2023). A Review of Ensemble Learning and Data Augmentation Models for Class Imbalanced problems: Combination, Implementation and Evaluation. *Expert Systems with Applications*, [online] 244(122778), p.122778. doi:https://doi.org/10.1016/j.eswa.2023.122778.

Xuan, F., Su, W., Chen, Z., Huang, X., Zhai, W., Li, X., Zeng, Y., Li, Z., Li, J. and Huang, J. (2025). Performance of stacking machine learning and volume model for improving corn above ground biomass prediction. *Plant Phenomics*, [online] 7(3), p.100068. doi:https://doi.org/10.1016/j.plaphe.2025.100068.