**Project Proposal: Anti-Money Laundering (AML) Transaction-Level Models**

**Team Members:**

* Dmitry Pavlyuk
* Igor Rodin
* Ali Turkarslan
* Niedre Vija
* Jolanta Krastina

**1. Main Problem**

Money laundering is a significant global problem that involves concealing the origins of illegally obtained funds, typically through a complex series of financial transactions. The **layering stage** of money laundering — where illicit funds are separated from their origin by creating layers of transactions — is particularly challenging to detect due to the complexity and the number of transactions involved.

Financial institutions are under pressure to comply with strict Anti-Money Laundering (AML) regulations by monitoring financial activity and identifying suspicious behavior. However, the existing transaction monitoring systems often generate a high volume of false positives, leading to inefficiency. This project aims to improve the detection of suspicious activities during the layering stage using machine learning models at the transaction level.

**2. Initial Research on the Problem Domain**

**Relevant Literature:**

* Gupta, A., Dwivedi, D.N., Shah, J. (2023). *Artificial Intelligence Applications in Banking and Financial Services: Anti-Money Laundering and Compliance*. Springer, Singapore. This work discusses how machine learning can enhance AML processes.
* Weber, M. et al. (2018). *Scalable Graph Learning for Anti-Money Laundering*. This paper investigates graph learning methods for AML transaction monitoring.

**Existing Products:**

* **IBM AMLSim**: A multi-agent simulator used to generate synthetic transaction data and simulate money laundering behaviors, such as layering, scattering, and cycle schemes.
* **IBM Anti-Money Laundering Data**: Transaction datasets from IBM's AML solution, used for benchmarking machine learning approaches.

**3. Project Objectives**

**In-Scope:**

* To analyze and model the layering stage of money laundering using machine learning techniques.
* Develop transaction-level models capable of identifying suspicious patterns such as **scatter-gather**, **cycle**, **bipartite**, and **random** typologies.
* Use synthetic data from AMLSim for model training and testing.
* Evaluate the performance of selected models based on false positive rate, precision, and recall.

**Out-of-Scope:**

* Full implementation of AML systems beyond the layering stage.
* Detection of money laundering at the placement or integration stages.
* Analysis of non-financial transaction datasets, such as legal documents.

**4. Proposed Work Activities/Tasks**

**Work Packages:**

1. **Research & Literature Review (Weeks 1-2)**:
   * Explore existing machine learning approaches in AML.
   * Review datasets and AML simulators, such as AMLSim.
   * Deliverable: Comprehensive review document on current AML methodologies.
2. **Data Collection & Preprocessing (Weeks 3-4)**:
   * Use synthetic data from AMLSim.
   * Clean and preprocess transaction data for model training.
   * Deliverable: Clean dataset ready for modeling.
3. **Model Development (Weeks 5-8)**:
   * Implement selected machine learning models: decision trees, random forests, graph-based learning.
   * Test different transaction-level models for detecting suspicious patterns.
   * Deliverable: Trained machine learning models and their performance reports.
4. **Model Evaluation & Refinement (Weeks 9-10)**:
   * Evaluate model performance and refine based on results.
   * Conduct comparative analysis of different typologies.
   * Deliverable: Model evaluation report.
5. **Final Report & Viva Preparation (Weeks 11-12)**:
   * Prepare final project report.
   * Prepare for project viva with comprehensive results and methodology discussion.
   * Deliverable: Final project report, presentation for viva.

**5. Project Management Methodology**

We will follow an **Agile** methodology with weekly stand-up meetings to track progress and make necessary adjustments. The project will be divided into **sprints**, each focusing on one of the main work packages (e.g., data preprocessing, model development).

Tasks will be managed using a combination of GitHub Project boards for task assignment and Google Docs for collaboration. Each sprint will end with a sprint review to assess progress.

**Tools:**

* **Version Control**: GitHub repository for code and documentation.
* **Collaboration**: Google Drive for document sharing, Slack for communication.
* **Project Management**: GitHub Projects, with task boards to manage milestones and deadlines.

**6. General Implementation Plan/Timeline**

| **Task** | **Duration** | **Start Date** | **End Date** |
| --- | --- | --- | --- |
| Literature Review | 2 weeks | Week 1 | Week 2 |
| Data Collection & Preprocessing | 2 weeks | Week 3 | Week 4 |
| Model Development | 4 weeks | Week 5 | Week 8 |
| Model Evaluation | 2 weeks | Week 9 | Week 10 |
| Final Report & Viva Preparation | 2 weeks | Week 11 | Week 12 |

**Additional Material (Appendices)**

**Appendix 1: Risk Register**

| **Risk** | **Likelihood** | **Impact** | **Mitigation** |
| --- | --- | --- | --- |
| Dataset not sufficient | Medium | High | Use multiple data sources |
| Model underperforms | Medium | Medium | Iterate model improvement |
| Time constraints | High | High | Weekly sprint review & feedback |

**Appendix 2: Gantt Chart**

A detailed Gantt chart for the timeline and deliverables is available on the [GitHub Project Board](https://github.com/yourusername/aml-group-project/projects).

**Public Resources**

* GitHub Repository: <https://github.com/yourusername/aml-group-project>
* Project Management Board: [GitHub Project Board](https://github.com/yourusername/aml-group-project/projects)