Insider threat detection using semi-supervised ML

Team: Noah Asing (BIOE), Diru Jia (IEOR), Wenqi Kou (IEOR), Tianxiao Gaoqu (IEOR)

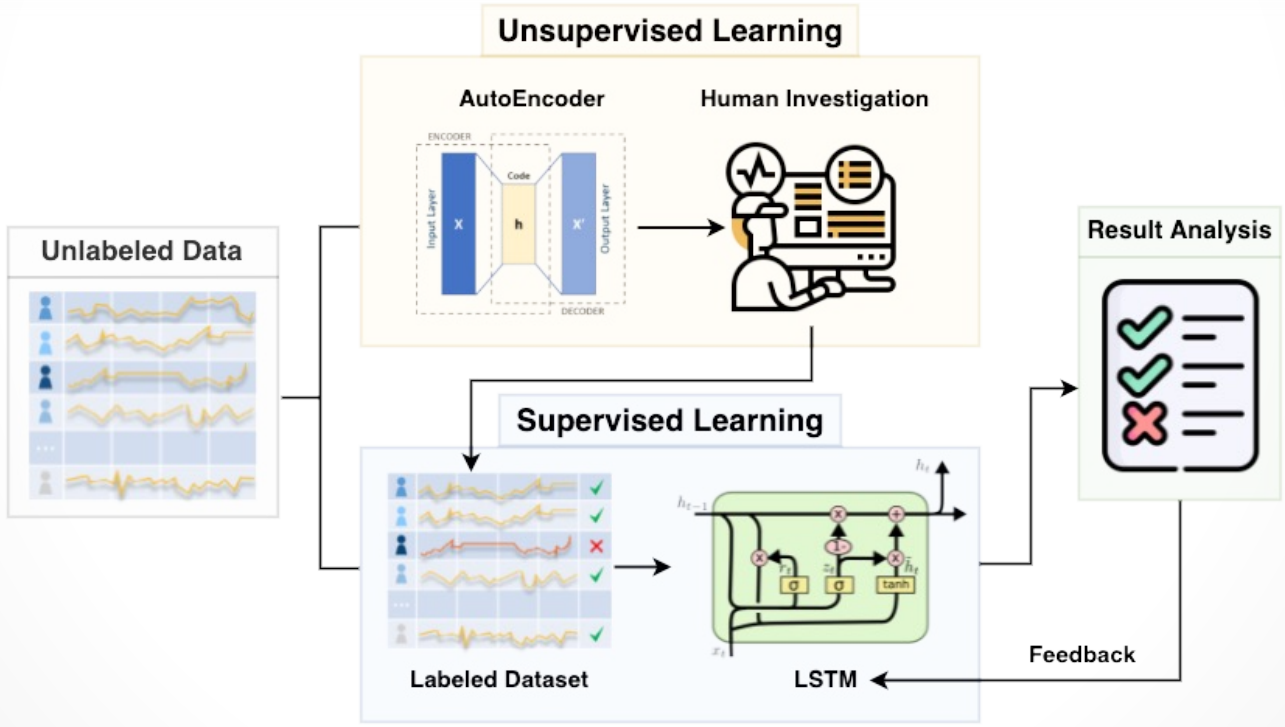
Advisors: Aditya Kapoor, Robert Molony, Patrick Crenshaw, John Stringer, Mihaela Gaman, Aisha Asanova, Dr. T.I. Zohdi

Insider threat detection is a complex problem because the human behavior is unpredictable. A 2020 Ponemon study shows following stats:

- ➔ Number of insider incidents increased by **47%** in two years
- ➔ An average insider incident costs its organization **$755,760**

Effective anomaly detection must balance between a human security team and the algorithm.

Our team presents an **Anomaly Detection System** to identify suspicious user activities, prioritize investigation resources, and incorporate human screening to avoid a bias and false positives.



Unsupervised Learning

The first part of the data is sorted by **AutoEncoder** according to degree of suspicion. Next, based on investigation budget, the upper percentile of the ranked data is manually labeled and the remaining ranked data defaults as safe behavior. Thus, this first part of the data is now labeled.

Supervised Learning

Supervised learning (SL) models are trained on the labeled data segment and applied to the remaining segment. Algorithms tested include **Random Forest**, **XGBoost** and **LSTM.** Merging predictions with labels from yield scores for the full dataset.