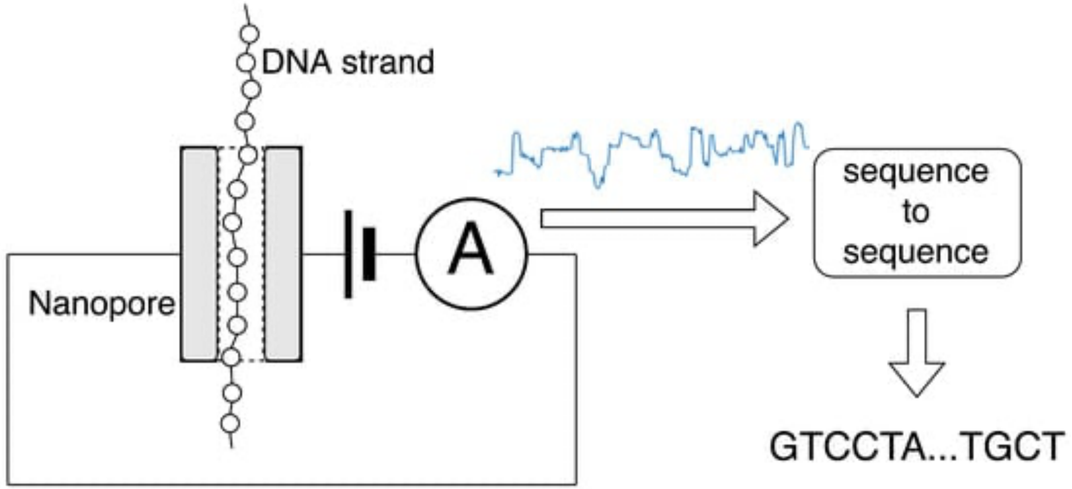
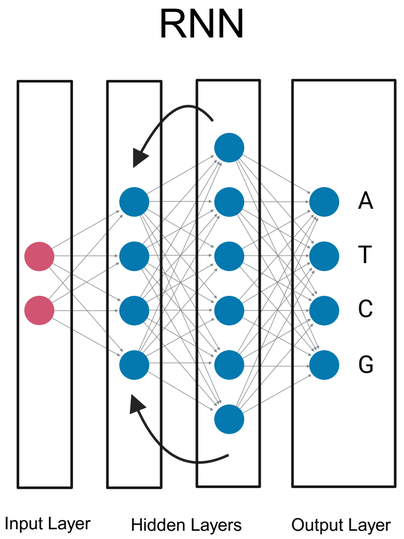
Transformer and Topological Data Analysis for Early Fraud Detection

Proposed by: Lebohang Mashatola, PhD

Source of Inspiration

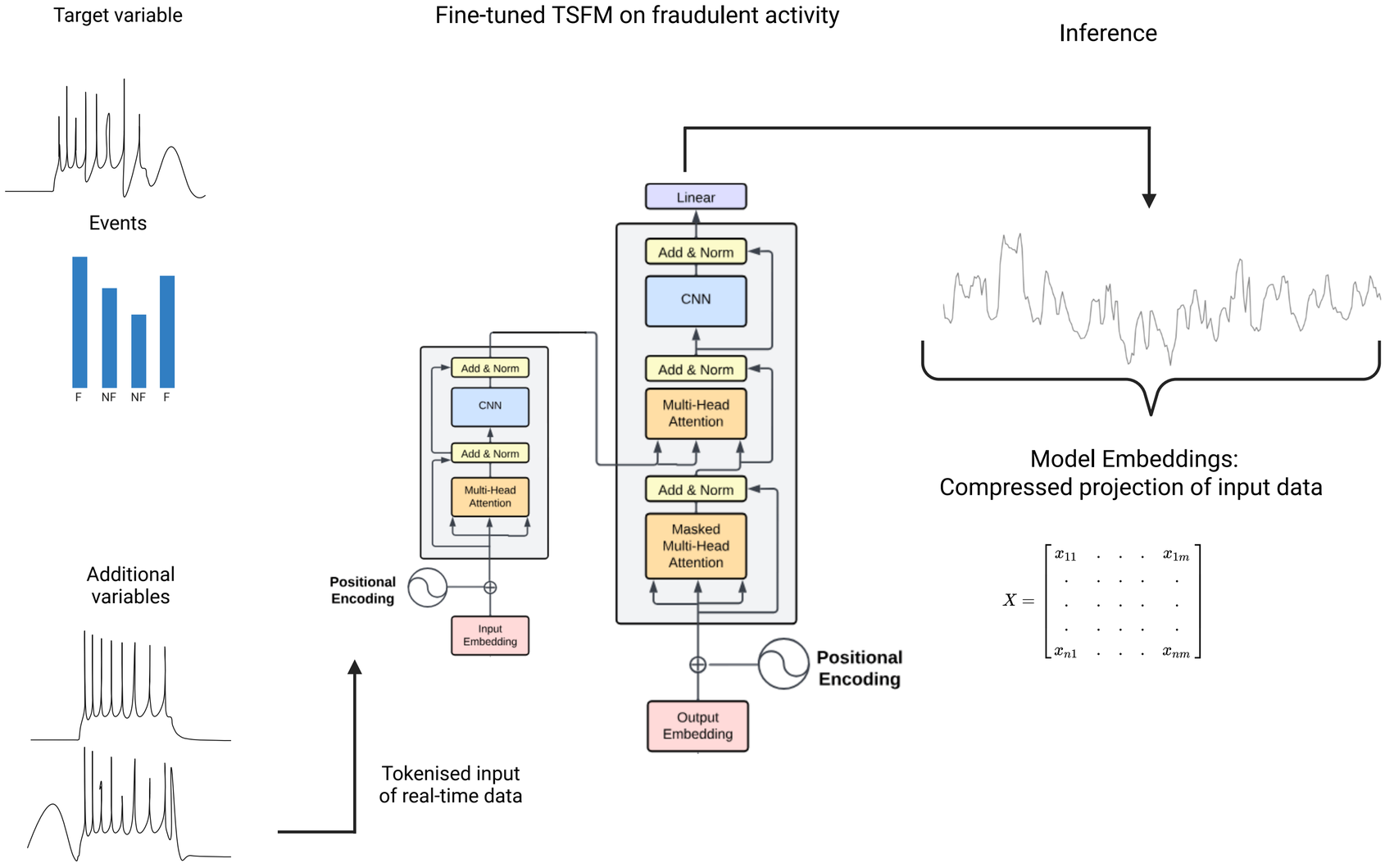
- Semiconductor DNA sequencing
- **Problem Statement:** Early-Fraud Detection
  - Mimics legitimacy
  - Small-subset of data
  - Requires high precision
- **Proposed Solution:** TDA and TSFM
  - TDA captures structure of data
  - TSFM captures long-term trends in data





Application to Banking/Finance

- Data transformation
- Inference fine-tuned TSFM on bank data
- Output model embeddings
- Use embeddings for TDA fraudulent activity detection

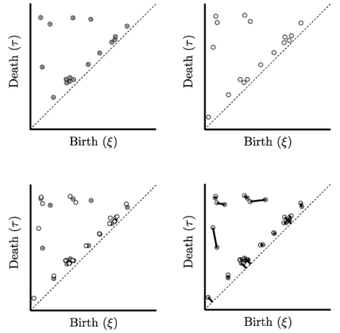


Application to Banking/Finance

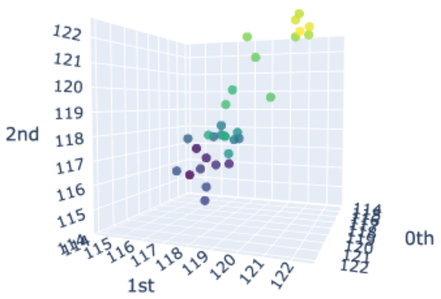
- Model embeddings transformed into Euclidean space
- Obtain topological fingerprint per time stamp
- Compute distance between topological fingerprints
- Advanced early-detection anomaly

$$X = \begin{bmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nm} \end{bmatrix}$$

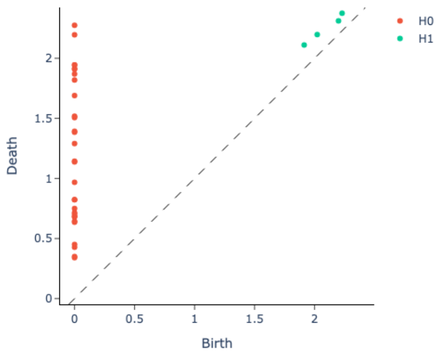
Time-series data



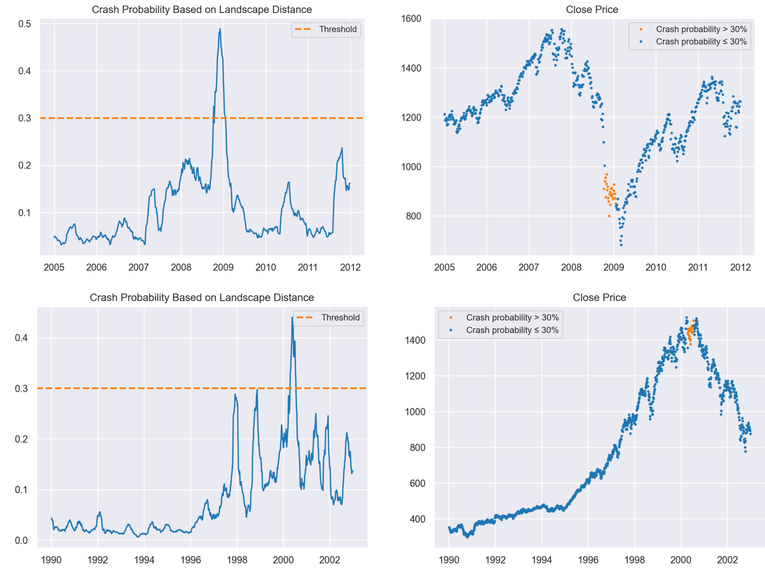
Difference between time-stamps



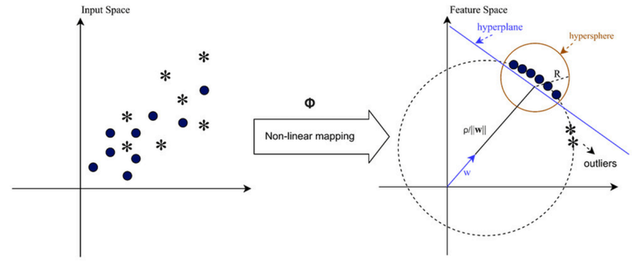
Euclidean distance measure



Persistence homology summary



Contextual anomaly detection



One-class SVM

Example adapted from: https://github.com/giotto-ai/stock-market-crashes/tree/master

Business Value

- Harness AI (long-term temporal data modelling) and TDA (high precision)
- Financial protection
- Operational and strategic insights
- Transferable use-cases within Nedbank:
    A. Cybersecurity
    B. Monitor IT systems
    C. Customer behaviour