Bekhzod Anvarov

Machine Learning Engineer
Self-supervised learning (JEPA), deep learning, rilevamento delle anomalie

Torino, Italia
virtsudo@proton.me | virtsudo.vercel.app | linkedin.com/in/virtsudo | github.com/virtsudo

PROFILO

Laurea magistrale in Ingegneria Informatica (percorso Cybersecurity) al Politecnico di Torino, con focus sul deep learning applicato ai dati di sicurezza. Nella tesi ho sviluppato in PyTorch una Joint-Embedding Predictive Architecture (JEPA) self-supervised e ho verificato se, in presenza di concept drift, la sua rappresentazione permette a un classificatore di restare più accurato che con le feature grezze. Ho inoltre addestrato modelli FFNN, RNN, GNN e BERT per compiti di sicurezza e uso Python da almeno 5 anni. Disponibile per posizioni junior e tirocini nel machine learning, in presenza o da remoto.

FORMAZIONE

Laurea Magistrale in Ingegneria Informatica (Computer Engineering), percorso Cybersecurity
Politecnico di Torino, 2026

Laurea Triennale in Ingegneria Informatica
Politecnico di Torino, 2022

ESPERIENZA DI RICERCA

JEPA Representation Learning for Intrusion Detection on Real-World Drifting Network Traffic
Tesi di laurea magistrale, Politecnico di Torino, 2026
Relatori: Prof. Guido Marchetto, Prof. Alessio Sacco, Dott. Cristiano Serra
Una rappresentazione appresa da traffico di rete non etichettato permette a un classificatore di intrusioni di restare più accurato che con le feature di flusso grezze su traffico registrato anni dopo?

- **Sviluppo** in PyTorch di un modello JEPA self-supervised con un context encoder, un target encoder EMA, un predictor e il mascheramento a gruppi delle feature correlate. Preaddestramento di un encoder per anno su flussi non etichettati di traffico reale di backbone Internet degli anni 2011, 2016 e 2021 (benchmark MAWIFlow).
- **Confronto** con le feature grezze su quattro classificatori (regressione logistica, albero decisionale, random forest, XGBoost), addestrati su ogni anno e testati su tutti gli anni (ROC-AUC, PR-AUC). Sostituzione della normalizzazione del benchmark, che faceva perdere gran parte della risoluzione delle feature.
- **Risultati.** La rappresentazione rende le classi più linearmente separabili ma non riduce il calo tra gli anni. In-domain, una regressione logistica su di essa colma gran parte del divario di ROC-AUC rispetto a XGBoost sulle feature grezze. Le etichette del 2011 hanno un significato diverso da quelle degli anni successivi. Ho inoltre implementato i regolarizzatori VICReg e SIGReg, che non hanno migliorato i classificatori.

Tecnologie: Python, PyTorch (CUDA), scikit-learn, XGBoost, Polars, NumPy, Matplotlib

PROGETTI

Laboratori di Deep Learning per la Cybersecurity
Tecnologie: Python, PyTorch, Jupyter, Hugging Face
Sviluppo di una FFNN per il rilevamento delle intrusioni di rete e di modelli FFNN, RNN e GNN per classificare malware. Fine-tuning di BERT per analizzare log SSH ed etichettarli con le tattiche MITRE ATT&CK.
GitHub: github.com/virtsudo/dl-cybersecurity-labs

COMPETENZE

Machine learning: PyTorch (CUDA), scikit-learn, XGBoost, Hugging Face, Polars, NumPy, Matplotlib, Jupyter
Programmazione: Python, C/C++, SQL (MySQL, SQLite)
Strumenti e ambienti: shell Linux, Git/GitLab, LaTeX
Aree di interesse: self-supervised learning, representation learning, rilevamento delle anomalie, concept drift

CORSI RILEVANTI

- **AI and Cybersecurity:** reti CNN, RNN e LSTM, NLP, rilevamento di anomalie, classificazione del traffico.
- **Data Science and Database Technology:** classificazione, clustering, regole associative, data warehouse, OLAP.

LINGUE

Uzbeko (madrelingua), russo (C1), inglese (B2), italiano (B1)