

Manipulation Detection of the Bank Sector

Friday 19 September 2025 14:00 (15 minutes)

The integration of Artificial Intelligence (AI) in banking has brought significant advancements in areas such as credit scoring, fraud detection, and risk management. However, the adoption of complex machine learning models has also introduced major challenges regarding transparency, regulatory compliance, and trust. This paper examines a broad set of explainability techniques applicable to banking, ranging from model-agnostic methods such as SHAP and LIME to model-specific approaches including attention mechanisms and counterfactual reasoning. We discuss their relevance, interpretability trade-offs, and integration challenges within high-stakes financial environments. Special attention is given to practical use cases and the alignment of these techniques with ethical and regulatory standards. Our analysis provides key insights into the current landscape of explainable AI and outlines future directions for trustworthy and interpretable financial systems.

Authors: BADOI, Mircea; BARCAN, Alexandra

Co-authors: TAPUS, Nicolae (Politehnica Bucharest); CIOCHIU, Marius-Daniel; ENESCU, Nicolae (University of Craiova)

Presenter: BADOI, Mircea

Session Classification: Pervasive Systems and Computing

Track Classification: Doctoral Symposium