

Introduction to Machine Learning and Network Analytics in Finance Minitrack

Peter Sarlin
Department of Economics
Hanken School of Economics
Risklab Finland, Arcada University of Applied Sciences
peter@risklab.fi

József Mezei
School of Business and Management
Lappeenranta University of Technology
F-Secure Corporation, Helsinki
jozsef.mezei@f-secure.com

Recent developments in both the financial domain and in machine learning techniques are continuously driving the growth of contributions utilizing computational methods in finance. This increased attention is further amplified by the availability of large datasets, composed of structured as well as unstructured data, requiring researchers and practitioners to utilize machine learning and network analysis increasingly. The contributions accepted to the minitrack represent a selection of the most widely used computational methods, including techniques from computational intelligence, network analysis, probabilistic models and classical regression-based predictions. While the main focus in all the accepted articles is on creating new prediction models, the authors put great emphasis on developing solutions that can offer insights and deep understanding to complement superior prediction performance. This is a crucial point when one considers the practical applicability of machine learning methods in the financial domain.

Four papers have been accepted to be presented in the single session of the minitrack. The articles offer a nice combination of methods and application areas dominating the field of machine learning in finance, with applications in retail banking, peer-to-peer lending, business process analysis and systemic risk analysis.

The first paper “*Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning*” is authored by József Mezei (Lappeenranta University of Technology & F-Secure Corporation), Ajay Byanjankar (Åbo Akademi University) and Markku Heikkilä (Åbo Akademi University). The paper develops a novel data transformation method to be used as preprocessing step for a supervised learning task. The article extends a previous model by incorporating fuzzy sets of higher order in the linguistic data transformation process. The proposed method shows superior classification performance when applied to a dataset from a peer-to-peer lending platform with the purpose of identifying risky clients.

The second paper “*A Partial Parameter HMM Based Clustering on Loan Repayment Data: Insights into Financial Behavior and Intent to Repay*” by Dibu Philip, Nandan Sudarsanam and Balaraman Ravindran (Indian Institute of Technology Madras) proposes novel Hidden Markov Model based clustering to understand loan repayment behaviour. Instead of purely offering a prediction on whether customers will repay their loan or not, the article focuses on understanding the various attributes that impact the repayment process. Using a real dataset from a retail bank, the authors illustrate the novel business insights that can be gained using the developed method.

The third paper “*Event entry time prediction in financial business processes using machine learning - A use case from loan applications*”, authored by Michael Frey, Andreas Emrich, Peter Fettke and Peter Loos (Saarland University), presents a framework to predict business process performance indicators. Following a design science approach, the authors develop an event entry time prediction model. A loan application process is used as a case illustration, involving performance comparison of four different regression models.

In the fourth paper, entitled “*RiskRank to predict systemic banking crises with common exposures*”, Peter Sarlin (Hanken School of Economics & Risklab at Arcada), Paolo Giudici (University of Pavia), Alessandro Spelta (University of Pavia) and Kaj-Mikael Björk (Arcada University of Applied Sciences) focus on the interconnectedness of financial markets using systemic risk modelling. The authors study the utility of the network dimension of common exposures (funding composition and portfolio overlap) utilizing the RiskRank measure. The numerical analysis shows that funding and portfolio composition overlap are significant channels of contagion and should be accounted for when measuring systemic risk.