

Introduction to the Machine Learning and Predictive Analytics in Accounting, Finance and Management Minitrack

Peter Sarlin
Silo.AI

RiskLab at Arcada and Hanken School of Economics
peter@siloi.ai

József Mezei
Åbo Akademi University
Avaintec Oy, Helsinki
jmezei@abo.fi

Artificial intelligence (AI) and machine learning (ML) methodologies have become increasingly used in recent years in various applications across a variety of fields. In particular, academic researchers and practitioners in the domains of accounting, finance and management now commonly use techniques from deep learning and natural language processing, among others. As a complement to traditionally used statistical techniques, AI and ML allow to address previously unsolved problems making use of structured and unstructured data sources. Company documents, data from social media or customer reviews, previously largely untapped data sources, are used nowadays to improve traditional predictive analytics models. These observations are increasingly relevant today, as the changes in digitalization, as the consequence of the COVID-19 pandemic, has strengthened the role of AI and ML at the core of business operations.

The contributions submitted to the minitrack present unique advances in the application of AI and ML, making use of state-of-the-arts developments and applying them on large datasets. The minitrack has received seven submissions, out of which three were accepted for presentation. As in the previous years, the accepted papers are representative of the various application domains (accounting, finance and marketing) addressed by the minitrack. Among the seven submissions, three specifically focused on the impact of the current pandemic and how AI and ML can help in understanding and mitigating the impact in the hospitality industry or company performance prediction, among other problems.

The first article “*Transformer-based Summarization and Sentiment Analysis of SEC 10-K Annual Reports for Company Performance Prediction*” is authored by Hsin-Ting Hsieh and Diana Hristova (HWR Berlin). In the contribution, recent advances from the domain of Natural Language Processing are applied to build models that can predict the performance of companies based on the information presented in the annual reports. By combining text summarization, topic modelling, sentiment analysis and Tree-based classification models, the authors manage to

establish a connection between the content of annual reports and the company’s future performance. The authors put special emphasis to present the results in an easily interpretable way by determining relevant company characteristics, industry-specific and risk factors.

The second article, “*Predicting the Outcome of a Debt Collection Process Using Bayesian Networks*”, is authored by Benedikt Köhler and Hansjörg Fromm (Karlsruhe Institute of Technology). As the authors illustrate, collecting outstanding debts, a task typically outsourced by most companies, is a complex process requiring the analysis of large amounts of data in order to optimize the steps of the debt collection process. The authors make use of a dataset of tens of thousands of debt claims from different industries to develop decision support systems using Bayesian networks, and achieve F1-score above 91%. As the most important outcome of the model, a recommendation is given on which of possible actions increase the probability of debt repayment.

The third article, entitled “*A Novel Deep Learning Model For Hotel Demand and Revenue Prediction amid COVID-19*”, focuses on the impact of the COVID-19 pandemic on the hospitality industry. The authors, Ashkan Farhangi, Arthur Huang, and Zhishan Guo (University of Central Florida), set out to develop interpretable forecasting models to predict hotel demand and revenue. They develop the novel DemandNet deep learning framework to make robust long-term forecasts, making use of a highly non-linear structure and a sophisticated feature selection process. The model is evaluated using hotel demand and revenue data from the US between 2013 and 2020. DemandNet is found to outperform many widely used models, such as exponential smoothing, ARIMA or MQ-transformer.