



INTERNATIONAL MONETARY FUND

Caucasus, Central Asia, and Mongolia
Regional Capacity Development Center



DEVELOPMENT PARTNERS: ASIAN DEVELOPMENT BANK | CHINA | KOREA | POLAND | SWITZERLAND

MEMBER COUNTRIES: ARMENIA | AZERBAIJAN | GEORGIA | KAZAKHSTAN | THE KYRGYZ REPUBLIC | MONGOLIA | TAJIKISTAN | TURKMENISTAN | UZBEKISTAN

CCAMTAC- Regional Research/Analytical Economic Policy Seminars

“Consumer Credit Risk Analysis via Machine Learning Algorithms”

Tuesday, September 28, 2021

Introduction and Moderator:

Norbert Funke, Director, CCAMTAC

Jianping Zhou, Senior Economist, Monetary and Capital Markets Department, IMF

Presenters:

Shalkar Baikulakov, Head of Project Management Office, Payments and Fintech R&D Center
(subsidiary of the National Bank of Kazakhstan)

Zanggar Belgibayev, Senior Analyst, Monetary Policy Department, National Bank of
Kazakhstan

Discussant:

Bihong Huang, Economist, Monetary and Capital Markets Department, IMF

Serdar Guner, Financial Regulation and Supervision Advisor, CCAMTAC

In his opening remarks, Norbert Funke emphasized the importance of an accurate credit market analysis for financial stability. New techniques, such as the application of machine learning algorithms to credit risk analysis, including consumer loans, can potentially improve the assessment of the creditworthiness of individuals.

Jianping Zhou, the moderator of the seminar, pointed out that over the last decade, machine learning or Artificial intelligence (ML/AI) has been increasingly applied to finance, especially in assessing credit default risks, by banks, the central banks, rating agencies, and used by the IMF to assess financial vulnerabilities. Compared to traditional credit risk models, machine learning algorithms can handle much larger data sets, are faster in delivering results and can better capture nonlinear effects. The paper provides a useful example of how the central bank of Kazakhstan assesses the creditworthiness of individuals through machine learning algorithms, based on regulatory data provided by second-tier banks to the central bank. In turn, the assessment of the creditworthiness of borrowers can help the central bank investigate the accuracy of issued loans and predict potential systematic risks.



INTERNATIONAL MONETARY FUND

Caucasus, Central Asia, and Mongolia
Regional Capacity Development Center



DEVELOPMENT PARTNERS: ASIAN DEVELOPMENT BANK | CHINA | KOREA | POLAND | SWITZERLAND

MEMBER COUNTRIES: ARMENIA | AZERBAIJAN | GEORGIA | KAZAKHSTAN | THE KYRGYZ REPUBLIC | MONGOLIA | TAJIKISTAN | TURKMENISTAN | UZBEKISTAN

In the paper Shalkar Baikulakov and Zanggar Belgibayev developed two linear and six nonlinear classification methods (linear models – Logistic Regression, Stochastic Gradient Descent, and nonlinear - Neural Networks, kNN, Decision tree, Random Forest, XGBoost, Naïve Bayes). The algorithms were compared based on 7 classification metrics, including accuracy, precision, F1, type 2 error and others. Based on these metrics, the non-linear models illustrate more accurate predictions in comparison with the linear models. In particular, the non-linear models such as the Random Forest and kNN classifiers on oversampled data demonstrated promising outcomes. As it was explicitly stated by the presenters with regards to Kazakhstan's credit market, machine learning algorithms can help to analyze effectively large and unbalanced datasets, which include both categorical and numerical information. Overall, the results of the study showed that the machine learning models work sufficiently well on regulatory data collected by the central bank. The models based on regulatory data can provide an adequate foundation for the evaluation of credit risk of issued consumer loans by second tier banks and enable to prevent systematic risks.

In commenting on the paper, Bihong Huang, summarized the key messages and highlighted the paper's innovations. She suggested considering the possibility of providing additional descriptive statistics of input and response variables and an overview of their dynamics over time. In future extensions of this work, it may be interesting to include information on collateral (if loans are secured), income level, previous incidents of late payments, and selected macroeconomic variables. It is also important to implement the out of sample test and compare the model predictions with the real data.

Focusing on implications for supervisors and regulators, Serdar Guner emphasized the importance of assessing the efficacy of the historical data, continuous evaluation and refinement of the model, and its validation. He also reminded that from a developers' perspective it is important to be aware of privacy protection legislation when conducting research that involves sensitive information, avoid bias and be prepared to support the model development.

A discussion was followed with questions related to (i) managing the challenges of undersampled and oversampled data; (ii) possibilities for second tier banks to apply the algorithms, developed in the research, in the assessment of creditworthiness; (iii) the NPL dynamics in Kazakhstan during the pandemic; (iv) splitting the training and test data, and recommendations related to the statistical aspects of the paper.

Aliya Kistaubayeva,
Economic Analyst, CCAMTAC