

Deliverable information

WP NO.	WP5
DEL. REL. NO.	D5.1
DEL. NO.	D13
TITLE	Repository of use cases and slides in big data analytics
DESCRIPTION	Repository of use cases (including paper, data and code) and slides in big data analytics shared during the Suptech and RegTech workshops.
NATURE	Website
EST. DEL. DATE	30/09/2019

Document information

DATE	25/09/2019
WRITTEN BY	Paolo Pagnottoni
APPROVED BY	Paolo Giudici

Actions and achieved results

The University of Pavia, with the support of all the project partners, has developed the repository of use cases and slides for the Big Data Analytics part. The repository includes research and teaching material, in the form of use cases, regarding big data analytics for credit risk management in peer to peer lending. The material has been developed first gathering contributions from the consortium and, then, evolving it according to the received feedbacks from all project's participants: partners, regulators, supervisors, banks, fintechs and international advisors.

The material has been prepared, shared and disseminated during:

1. The initial kick-off workshop (held in Pavia), in which European and international regulators presented priorities of work and research;
2. Suptech sessions, in which each partner discussed the developed use cases with their national Supervisors;
3. Two Regtech sessions (one in Milan and one in Frankfurt), in which FinTech companies and banks could replicate the use cases, with the provided software code;
4. a final Validation workshop (held in Winterthur), in which the project stakeholders including Regulators and the project Advisory Board exchanged feedback and views about the developed use cases.

At first, the material was updated based on internal discussion and Regulators' suggestions during the first kickoff workshop. Second, the partners from the consortium have collected feedbacks from their national Supervisors during the Suptech sessions organised in all European Union countries, and updated the material. Third, we have gathered feedbacks from participants during the Regtech sessions, mostly risk managers and IT developers from FinTechs and banks, and updated the material. Fourth, we have discussed the developed research with partners, Regulators, Advisory Board members and participants during the Validation workshop, and the material has been finalised. All intermediate outputs and the final one (attached to this report) have been made available to all the stakeholders on the project web platform.

Specifically, the final repository contains:

1. Final open source papers regarding each use case
2. Final codes related to each use case
3. Final data required for the empirical application of each use case
4. Final set of slides, gathering all use cases into one comprehensive training session.

From a technical viewpoint, the use cases selected by the partners and all the project's participants are four.. A summary description of each of them is reported below.

USE CASE 1: NETWORK BASED SCORING MODELS TO IMPROVE CREDIT RISK MANAGEMENT IN PEER TO PEER LENDING PLATFORMS

(GIUDICI P.- UNIPV, HADJI-MISHEVA B- ZHAW., SPELTA A.- UNIPV)

Financial intermediation has changed extensively over the course of the last two decades. One of the most significant change has been the emergence of FinTech. In the context of credit services, fintech peer to peer lenders have introduced many opportunities, among which improved speed, better customer experience and reduced costs. However, peer-to-peer lending platforms lead to higher risks, among which higher credit risk: not owned by the lenders, and systemic risks: due to the high interconnectedness among borrowers generated by the platform. This calls for new and more accurate credit risk models to protect consumers and preserve financial stability. In this paper we propose to enhance credit risk accuracy of peer-to-peer platforms by leveraging topological information embedded into similarity networks, derived from borrowers' financial information. Topological coefficients describing borrowers' importance and community structures are employed as additional explanatory variables, leading to an improved predictive performance of credit scoring models.

Keywords: FinTech, Peer-to-Peer Lending, Credit Scoring Models

USE CASE 2: FACTORIAL NETWORK MODELS TO IMPROVE P2P CREDIT RISK MANAGEMENT (AHELEGBEY D.F. - UNIPV, GIUDICI P.-UNIPV, HADJI-MISHEVA B.-ZHAW)

This paper investigates how to improve statistical-based credit scoring of SMEs involved in P2P lending. The methodology discussed in the paper is a factor network-based segmentation for credit score modeling. The approach first constructs a network of SMEs where links emerge from comovement of latent factors, which allows us to segment the heterogeneous population into clusters. We then build a credit score model for each cluster via lasso logistic regression. We compare our approach with the conventional logistic model by analyzing the credit score of over 15000 SMEs engaged in P2P lending services across Europe. The result reveals that credit risk modeling using our network-based segmentation achieves higher predictive performance than the conventional model.

Keywords: Credit Risk, Factor models, Fintech, Peer-to-Peer lending, Credit Scoring, Lasso, Segmentation

USE CASE 3: SPATIAL REGRESSION MODELS TO IMPROVE P2P CREDIT RISK MANAGEMENT (AGOSTO A.-UNIPV, GIUDICI P.-UNIPV, LEACH T. -UNIPV)

Calabrese et al. (2017) have shown how binary spatial regression models can be exploited to measure contagion effects in credit risk arising from bank failures. To illustrate their methodology, the authors have employed the Bank for International Settlements' data on flows between country banking systems. Here we apply a binary spatial regression model to measure contagion effects arising from corporate failures. To derive interconnectedness measures, we use the World Input-Output Trade (WIOT) statistics between economic sectors. Our application is based on a sample of 1,185 Italian companies. We provide evidence of high levels of contagion risk, which increases the individual credit risk of each company.

Keywords: credit risk, systemic risk, contagion, spatial autoregressive models, binary data

USE CASE 4: LOAN SCREENING AND DEFAULT PREDICTION WITH MACHINE LEARNING AND DEEP NEURAL NETWORKS

(TURIEL J.- UCL, ASTE T.-UCL)

Logistic Regression and Support Vector Machine algorithms, together with Linear and Deep Neural Networks, are applied to lending data. A two layer model is formed, where the first layer predicts loan rejection, while the second one further screens the loans for default risk. Logistic Regression was found to be the best performer for the first layer, with test set recall macro score of 77.4%. Deep Neural Networks, applied to the second layer alone, were the best performer, with validation set recall score of 72%, for defaults. The models were also tested on the subcategory of the dataset comprising loans taken for small businesses. Benefits of training on the whole dataset to predict the specific category were tested. The first layer of the model was found to benefit strongly from the larger training dataset, while the second layer was found to perform significantly better when trained on small business data only. This suggests a potential discrepancy between how these loans are screened and how they should be analysed in terms of default prediction.

Links

The material concerning the repository in Big Data Analytics is available on the project platform at the following links:

1. Link to the repository of Suptech material containing slides for the produced use cases:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech>
2. Link to the repository of use cases, containing related papers, replication codes and datasets:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech/SUPTECH%20BDA%20Use%20Cases>
 - 2.1. Link to the repository of use case 1, containing related paper, replication code and dataset:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech/SUPTECH%20BDA%20Use%20Cases/Use%20Case%20I>
 - 2.2. Link to the repository of use case 2, containing related paper, replication code and dataset:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech/SUPTECH%20BDA%20Use%20Cases/Use%20Case%20II>
 - 2.3. Link to the repository of use case 3, containing related paper, replication code and dataset:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech/SUPTECH%20BDA%20Use%20Cases/Use%20Case%20III>
 - 2.4. Link to the repository of use case 4, containing related paper, replication code and dataset:
<https://seafire.fintech-ho2020.eu/#org/lib/17459f7d-56d0-4c57-97a3-964c1904c106/Suptech/SUPTECH%20BDA%20Use%20Cases/Use%20Case%20IV>