22nd International Conference on Modelling and Applied Simulation
20th International Multidisciplinary Modeling & Simulation Multiconference

2724-0037 © 2023 The Authors.
doi: 10.46354/i3m.2023.mas.019

Innovative Machine Learning Solutions for Credit Classification of Commercial & Agricultural Vehicle Contracts

Agostino G. Bruzzone [1], Michele Arduini [1,2] *, Francesco Pandolfelli[2], Letizia Monti[2] and Federico Mini[2]

[1]DITEN Genoa University, Via Balbi 5, Genova, 16126, Italy
[2]Accenture Data + AI, Via Nino Bonnet 10, Milano, 20154, Italy

*Corresponding author. Email address: arduinimichele320@gmail.com

Abstract

This investigation explores a variety of machine learning techniques, with the goal to refine the prediction of overdue payments in installment contracts associated with agricultural and commercial vehicles. The study focuses on a leading capital goods firm as a representative example. Motivated by the need for effective liquidity management, an assortment of machine learning-based models, encompassing Logistic Regression, Random Forest, XGBoost, and LightGBM, are built and evaluated using a comprehensive dataset to anticipate payment delays. The performance of each model is meticulously examined, and a feature importance investigation is carried out to pinpoint the key determinants. In a quest to boost prediction precision, ensemble methods, especially a voting classifier and an innovative merger of a neural network and a voting classifier (NN-VC), are utilized. The investigation affirms that the amalgamation of multiple machine learning algorithms significantly augments late payment forecasting, paving the way for improved risk management and insightful financial decision-making within the firm. Through the comparison of distinct models and the unveiling of a novel ensemble technique, the research offers a unique insight and sets the stage for future applications in risk management.

Keywords: Machine Learning; Predictive Models; Late Payment Prediction; Financial Risk Management; Ensemble Methods.

1. Introduction

Timely payments are vital for the financial wellbeing of businesses such as capital goods companies, where late payments can lead to substantial difficulties. This research applies Strategic Engineering, melding Simulation, AI, and Data Analytics, to formulate machine learning algorithms capable of predicting and managing late payments.

This investigation is rooted in a capital goods company's operations and is developed in collaboration with Accenture's Applied Intelligence division. The primary objective is to construct a predictive model that can accurately determine the likelihood of delayed payments for this company. An innovative aspect of this research is the introduction of an ensemble model, which merges a Neural Network and Voting Classifier (NN-VC), to enhance prediction accuracy.

The study includes a literature review, data analysis, the creation and evaluation of various Machine Learning models, and also delves into advanced ensemble modeling techniques. The findings offer insights into the potential of Machine Learning in anticipating late payment patterns and promoting efficient financial decision-making.

2. State of the art

The groundwork for this research is a complete exploration of the existing literature relevant to late payments and the resultant implications for businesses, particularly those operating on a larger scale. This review highlights the dire necessity for accurate and timely predictive models to manage payment times and mitigate potential financial disruptions.

A significant portion of prior research and existing practices in managing payment latency heavily rely on relatively simplistic and traditional techniques. These methods, while providing a high degree of explainability, often fail to achieve optimal levels of accuracy. Often, the use of linear regression models, decision trees, and other basic machine learning models is prevalent. These models can offer valuable insights; however, they are not adept at handling complex data structures or patterns - a common characteristic of financial data.

This study delves into the exploration of machine learning techniques with an emphasis on the powerful ensemble machine learning algorithm, Gradient Boosting Machines (GBMs). A comprehensive discussion on GBMs highlights their core operational principles, potential applications, and the capability to handle large-scale data with intricate patterns.

The paper also evaluates existing research on predicting and enhancing invoice-to-cash collection through machine learning. The insights gleaned from these studies provide the necessary context and act as a guide for the development of the predictive models used in this research. This investigation underscores the potential and efficacy of machine learning methodologies in addressing late payment issues.

3. Materials and Methods

The foundational material for this study comprised two comprehensive datasets provided, categorically divided based on the age of contracts: a 'new contracts' dataset comprising contracts younger than one year, and an 'old contracts' dataset including contracts older than one year. This strategic partition was enacted to fully exploit the predictive potential of the datasets, as the 'old contracts' dataset, with its more extensive collection history, inherently carried more information than the relatively sparse 'new contracts' dataset. The 'new contracts' dataset encapsulates 10 distinct features, while the 'old contracts' dataset includes 17 features.

New Contracts	Old Contracts
District	District
LTV computed	LTV computed
House type	House type
Total amount expected	Total amount expected
Age	Age
Contract score	-
Land holding	Land holding
Family ownership of house	-
Term months (length contact)	-
Paid ratio	Paid ratio
-	Interest rate
-	Average days past due over the last four payments
-	Lag nature
-	Number of times no past due
-	Times installment has been on time
-	Variability in days past due over the last four payments
-	Number of installments past due
-	Times installment has been paid with past due between 4 and 30 dd
-	Time of last past due installment
-	Number of the installment

Table 1: Features of New and Old Contracts Dataset

For the methodology, the initial stage involved meticulous data preprocessing and cleaning, crucial to ensure the reliability and validity of the subsequent analyses. This phase was executed utilizing PySpark on the Databricks platform, an environment known for its effective handling of large datasets and complex computations.

Subsequently, a range of machine learning models, including Logistic Regression, Random Forest, XGBoost, LightGBM, and ensemble methods, were implemented on both datasets. The primary goal of employing these models was to develop a robust system capable of predicting the degree of late payment in installment contracts.

To enhance the performance of the individual models and leverage the unique strengths of each, ensemble techniques were incorporated. The utilization of these techniques was intended to augment the overall predictive accuracy of the system.

For the training, tracking and testing of the developed models, MLflow was used. This platform provided an efficient and effective means of managing the machine learning lifecycle, including experiment tracking and model deployment.

The focus of each analysis was strategically directed towards contributing to the broader research objective: developing an accurate, reliable predictive model for late payment behavior in installment contracts for agricultural and commercial vehicles, based on the distinct characteristics of 'new' and 'old' contracts. The basic findings of these analyses are further discussed and interpreted in the results section of the paper.

4. Results and Discussion

The exploration of predictive models in this research is presented in two distinct segments. The first one focuses on individual machine learning models, namely Logistic Regression, Random Forest, XGBoost, and LightGBM. This provides a comprehensive understanding of their unique attributes and potential predictive capabilities. The results garnered from each model offer a distinctive outlook on the dataset's underlying patterns, laying the groundwork for the study's subsequent phase.

The second segment builds upon these insights, venturing into the realm of ensemble methods. Here, the potential benefits of integrating multiple models are examined, aiming to mitigate individual model weaknesses and amplify their strengths. Two ensemble models are deployed: the Voting Classifier (VC) and the Neural Network Voting Classifier (NN-VC). The implementation of these ensemble models paves the way for a more nuanced understanding of the dataset and ultimately contributes to more robust and accurate predictions.

4.1. Single Machine Learning Solutions

This section evaluates the performance of four distinct machine learning models: Logistic Regression, Random Forest, XGBoost, and LightGBM. Their individual strengths and limitations in class prediction are discussed, motivating the exploration of an ensemble approach.

Logistic Regression: As a traditional linear algorithm, it performed reasonably, providing a sound baseline for more complex models. Although its prediction capabilities for Class 1 and Class 2 showed room for improvement, the performance for Class 0 was satisfactory, which is a promising start.

Random Forest: As an ensemble learning method, it displayed a moderate performance, effectively showcasing its capability to handle complex relationships between variables. The model exhibited balanced results across the classes, hinting at its potential to further enhance prediction accuracy with optimization.

XGBoost: A renowned gradient boosting model, it maintained consistent performance across all classes. While the results didn't achieve exceptional levels, the superior outcomes compared to other individual machine learning models imply that XGBoost has a better grasp of the underlying patterns within the dataset.

LightGBM: An efficient gradient boosting decision tree model, it manifested promising results, particularly with Class 0. While its precision was highest among the standalone models for Class 0, there's room for improvement when it comes to Class 1 and Class 2 predictions. Yet, it showed potential in handling multi-class classification tasks effectively.

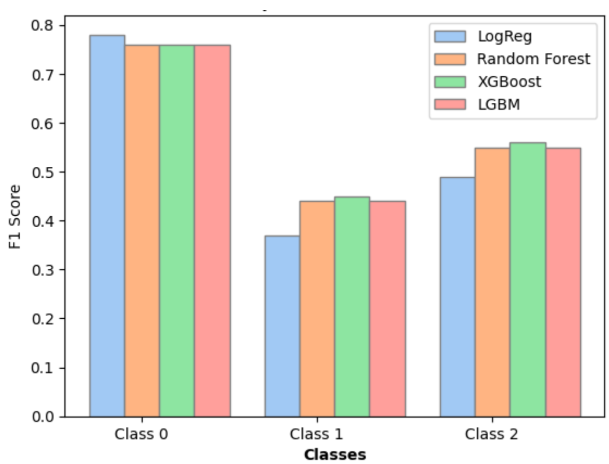

Figure 1: Comparative F1 Scores of Class Predictions by Different Machine Learning Models

In summary, the individual models demonstrated their unique strengths and offered valuable insights, even though none achieved high precision across all classes. These findings lead to an exploration of ensemble models, aiming to leverage the strengths of each model for enhanced prediction capabilities.

4.2. Ensemble Machine Learning Solutions

This section introduces two ensemble models, reflecting upon their respective architecture, design, and performance on the validation dataset. The first model, currently deployed in a production environment, amalgamates predictions from individual models, enhancing overall accuracy and reducing misclassification risk. The second model, experimental in nature, explores innovative techniques and model combinations, potentially pushing the boundaries of classification performance. These ensemble approaches are evaluated to understand their impact on the accuracy and reliability of contract classification, offering key insights for stakeholders considering the implementation of such models.

4.2.1. The Voting Classifier

The first ensemble model employs a Voting Classifier, which integrates the predictions of three individual models: Random Forest, XGBoost, and LightGBM, using a soft voting mechanism. This takes into account the confidence scores of each model, leading to more reliable predictions. Initial preprocessing steps are applied to the data for compatibility with modeling. These steps include a Column Transformer with different pipelines to handle numerical and categorical features separately, dealing with missing values and scaling the data. Random Forest uses the transformed data, while XGBoost and LightGBM can directly use the original features.

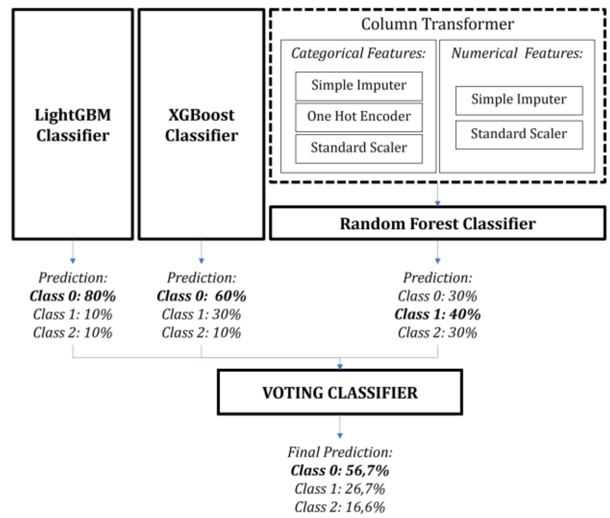

Figure 2 Structure of the Voting Classifier

This ensemble model, by harnessing diverse capabilities of the individual models, effectively recognizes patterns in the contract data, improving classification accuracy and reliability. Its performance is evaluated using various metrics such as accuracy, precision, recall, and F1-score.

The ensemble model of the Voting Classifier demonstrates promising results, particularly when assessing the F1-score across the different classes. This model brings together the predictive capabilities of three standalone machine learning models - Random Forest, XGBoost, and LightGBM. Its most notable achievement lies in its performance improvement over the best-performing single model, XGBoost, on the F1-score for classes 1 and 2.

The performance evaluation of the Voting Classifier (VC) highlights notable improvements over individual machine learning models. The overall F1-scores for the macro and weighted averages stand at 0.60 and 0.64, respectively, signifying enhanced predictive capabilities. On a more detailed analysis of class-specific F1-scores, the VC excels in predicting Class 0 with an F1-score of 0.75. Meanwhile, for Classes 1 and 2, the F1-scores are 0.47 and 0.59, respectively, signifying competitive performance levels.

4.2.2. The NN-VC approach: a Neural Network Voting Classifier

The second ensemble model, the Neural Network Voting Classifier (NN-VC), represents a novel approach in machine learning model architecture. This innovative model combines the strengths of the Voting Classifier with a feed-forward Neural Network used specifically for feature generation.

Neural Networks are known for their ability to learn complex, non-linear patterns in data. In the context of the NN-VC, the Neural Network component is not employed for direct prediction, but rather for generating a set of high-level features that capture deeper interactions within the original feature set. This additional layer of abstraction serves to enrich the feature space available to the Voting Classifier.

The enriched features, encapsulating intricate relationships within the data, are subsequently passed on to the Voting Classifier. This arrangement enables the Voting Classifier to make its predictions based on a more comprehensive understanding of the data. Thus, the NN-VC stands as a more sophisticated approach to the contract classification task, offering a compelling direction for future exploration and research.

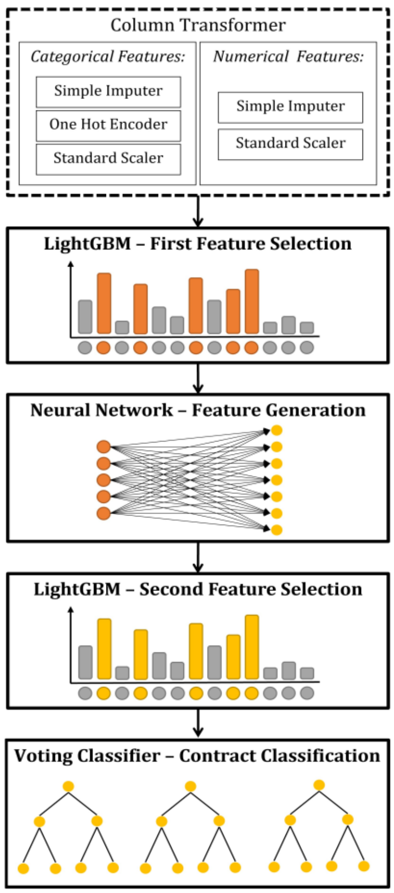

Figure 3: Structure of the NN-VC

The performance of the Neural Network Voting Classifier (NN-VC) mirrors that of the original Voting Classifier (VC). The overall F1-scores for the macro and weighted averages are 0.58 and 0.65 respectively, corresponding closely to the VC's values of 0.57 and 0.65 (This new set of values for the VC is obtained testing the model on a new wider test-set, the same on which the NN-VC has been tested) .

Looking at the class-specific F1-scores, it becomes evident that the NN-VC performs strongly for Class 0, with a score of 0.77. For Classes 1 and 2, the F1-scores are 0.46 and 0.53 respectively, which, while not exceptional, are nonetheless competitive.

When these results are compared with the Voting Classifier, the F1-score for Class 2 improves in the NN-VC model, suggesting that the incorporation of a Neural Network could potentially enhance the model's performance for this class. However, for Classes 1 and 2, the F1-scores are roughly on par between the NN-VC and the VC models, suggesting that the introduction of the Neural Network has not significantly altered the performance in these categories.

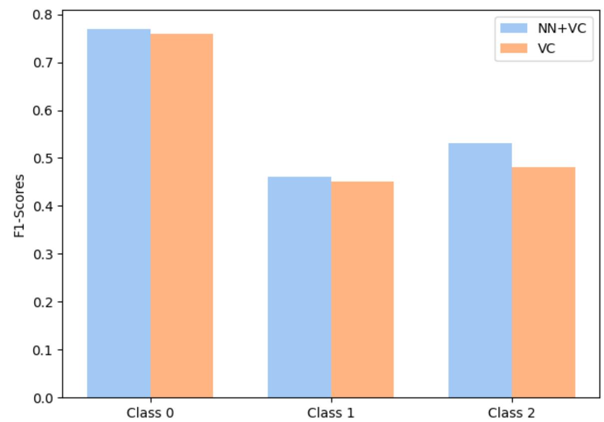

Figure 4: Comparative F1 Scores of Class Predictions by Voting Classifier (VC) and Neural Network Voting Classifier (NN-VC)

4.3. Feature Importance Analysis

The execution of this research involved the implementation of individual machine learning models, each providing an invaluable analysis of feature importance. Such analysis proved integral to discerning the variables contributing significantly to the prediction of payment lateness.

In the case of individual machine learning models, significant features were identified, such as the district (representing the geographic location of the customer), the computed Loan-to-Value (LTV), customer age, paid ratio, and the interest rate, among others. This importance was quantified using SHapley Additive exPlanation (SHAP) values, which provide a powerful approach to interpreting the output of any machine learning model. The SHAP values represent the average contribution of each feature to the prediction outcome across all possible permutations of features, thus providing a reliable estimate of feature importance. It is imperative to note that the significance of these features is not constant but varies depending on the specific dataset under consideration, owing to the unique feature sets each dataset possesses. In *Figure 5* and *6* are reported the shap values returned by LGBM for class 1 and class 2 of the old contract dataset (The varied colors for numerical features represent the feature value while categorical features are displayed).

The transition to ensemble models increases complexity in the relationship between the target variable (payment lateness) and features. Despite this complexity, it remains feasible to identify salient features within the Voting Classifier by investigating the feature importance of the individual models that form its structure. Consequently, the Voting Classifier's feature importance is understood to be a reflection of a combination of prominent features identified by its constituent models.

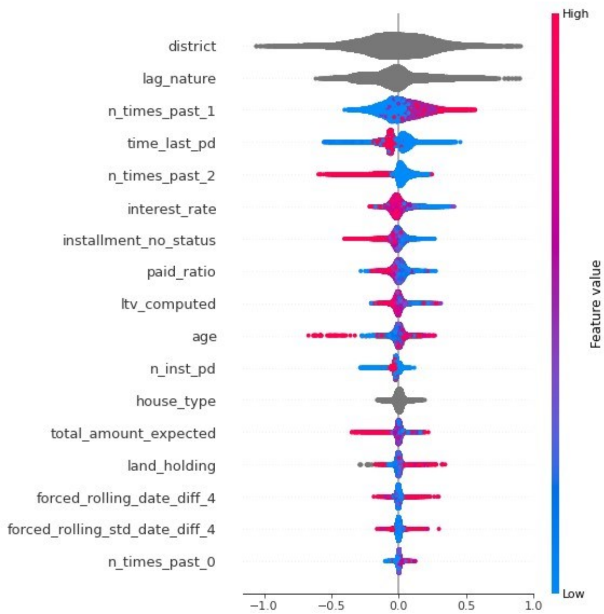

Figure 5: LGBM SHAP values showcasing feature importance for Class 1 prediction in the old contract dataset.

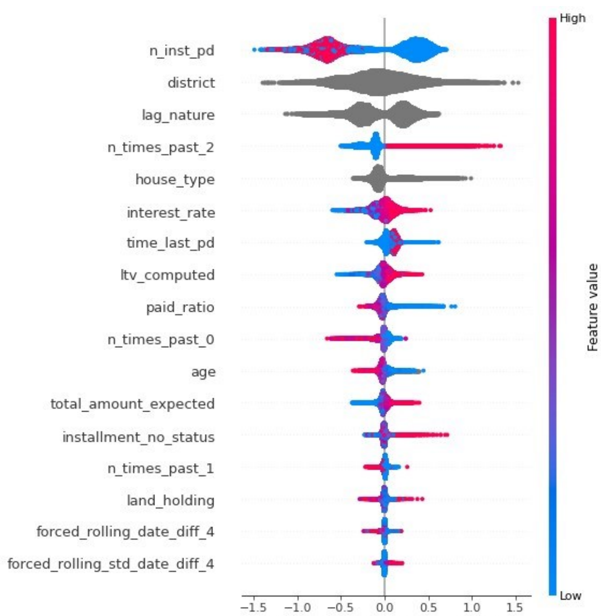

Figure 6: LGBM SHAP values showcasing feature importance for Class 2 prediction in the old contract dataset.

Interpreting the Neural Network Voting Classifier (NN-VC), however, poses a greater challenge. Neural networks are frequently characterized as 'black box' models, owing to their intricate internal computations and multiple layers of non-linear transformations. Consequently, determining the exact influence or importance of individual features on the predictions made by the NN-VC is a complex endeavour. This difficulty in discerning feature importance underscores one of the primary limitations of more intricate models, such as neural networks - their lack of transparency or 'explainability'. Despite this, the predictive accuracy of these models remains high, suggesting that a compromise between accuracy and explainability may be necessary to enhance the precision of late payment predictions.

5. Conclusions

Throughout the study, the primary focus has been on the development and comparison of machine learning models for contract classification. Single models such as Random Forest, XGBoost, and LightGBM demonstrated considerable capabilities, with XGBoost particularly standing out for its commendable performance.

The exploration of ensemble models for contract classification yielded intriguing results. Currently, the production environment of the client employs a Voting Classifier, blending predictions from three individual models: Random Forest, XGBoost, and LightGBM. This classifier has been operating reliably in the production environment for one year, delivering consistent results. It is retrained every three months to mitigate the risk of data drift and incorporate the newly collected installments. This regular retraining further strengthens the system's predictive power and accuracy.

The utilization of this approach demonstrated superior F1-scores for Classes 1 and 2 compared to the best single model, XGBoost, emphasizing the value of model diversity.

Simultaneously, an experimental Neural Network Voting Classifier (NN-VC) was also explored, using a neural network for feature generation. Despite promising potential, the NN-VC currently doesn't provide a significant performance uplift compared to the existing Voting Classifier.

The limitations of the NN-VC are primarily its experimental status and reduced explainability. In financial services, particularly, the ability to interpret and rationalize model decisions is essential, which favors the existing Voting Classifier.

In conclusion, the adoption of the Voting Classifier in the production environment represents a prudent strategy, balancing accuracy, reliability, and interpretability. Future development of the NN-VC could focus on enhancing performance and explainability, paving the way for potential deployment in contract classification tasks.

References

Hu Peiguang (2015). Predicting and Improving Invoice-to-Cash Collection Through Machine Learning. Master Thesis. Massachusetts Institute of Technology.

Paul, S. and Boden, R. (2008). The secret life of UK trade credit supply: setting a new research agenda. The British Accounting Review, 40(3):272–281.

Wilson, N. (2011). The credit management research centre. Journal of Business Credit Management, 2(2):20–23.

Carbo-Valverde, S. and Rodriguez-Fernandez, F. and Udell, G. F. (2016). Trade credit, the financial crisis, and SME access to finance. Journal of Money, Credit and Banking, 48(1):113–143.

Buzady, Z. (2012). New payment terms – an opportunity for credit managers. Journal of Business Credit Management, 3(1):24–27.

Richards, V. D and Laughlin, E. J. (1980). A cash conversion cycle approach to liquidity analysis. Financial management, 32–38. JSTOR.

Bielecki, T. R. and Rutkowski, M. (2020). Credit risk: modeling, valuation and hedging. Springer Science & Business Media.

Natekin, A. and Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in neurorobotics, 7:21. Frontiers.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232. JSTOR.

Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Ke, G. and Meng, Q. and Finley, T. and Wang, T. and Chen, W. and Ma, W. and Ye, Q. and Liu, T-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. Advances in Neural Information Processing Systems, 30.

Hosmer Jr, D. W and Lemeshow, S. and Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.

Kleinbaum, D. G and Klein, M. (2002). Logistic regression: a self-learning text. Springer Science & Business Media.

Papernot, N. and McDaniel, P. and Goodfellow, I. and Jha, S. and Celik, Z. B. and Swami, A. (2018). Stealing hyperparameters in machine learning. arXiv preprint arXiv:1802.05351.

Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32. Springer.

Chen, Z. and Zhang, Y. and Liu, J. and Hu, B. (2021). Large group activity security risk assessment and risk early warning based on random forest algorithm. Pattern Recognition Letters, 144:11–17. Elsevier.

Bruzzone, A. and Massei, M. and Sinelshnkov, K. (2020). Enabling Strategic Decisions for the Industry of Tomorrow. Procedia Manufacturing, 42:548–553. International Conference on Industry 4.0 and Smart Manufacturing (ISM 2019). doi: 10.1016/j.promfg.2020.02.028.

Bruzzone, A. G. and Di Matteo, R. and Sinelshchikov, K. (2018). Strategic Engineering & Innovative Modeling Paradigms. Workshop on Applied Modelling & Simulation, 14.

Bruzzone, A. G. and Massei, M. and Sinelshchikov, K. and Giovannetti, A. and Gadupuri, B. K. (2021). Strategic Engineering Applied to Complex Systems within Marine Environment. 2021 Annual Modeling and Simulation Conference (ANNSIM), 1–10. doi:10.23919/ANNSIM52504.2021.9552035.