

Editorial

# Special Issue on Emerging Trends and Challenges in Supervised Learning Tasks

Barbara Pes 

Dipartimento di Matematica e Informatica, Università di Cagliari, Via Ospedale 72, 09124 Cagliari, Italy; pes@unica.it

With the massive growth of data-intensive applications, the machine learning field has gained widespread popularity. It can provide solutions to a variety of real-world problems, relying on methodologies and algorithms that can automatically detect meaningful patterns in data. Supervised learning methods, in particular, have the potential of predicting the outcome of future observations, leveraging a proper set of training instances from the domain of interest.

A vast body of literature has discussed the strengths and weaknesses of the approaches devised in the context of supervised learning, as well as their applications to different fields, including disease prediction, text categorization, banking, finance, the Internet of Things, fraud detection and network intrusion detection, just to cite some examples.

As witnessed by the large amount of research in this area, the quality of supervised learning models may strongly depend on several factors, such as the number of instances available for training, their dimensionality, the number of problem classes, the level of class imbalance, and the variability of the concepts in time. Further, we often encounter datasets that are affected by data quality problems, such as incompleteness or noise. Taking such intrinsic data characteristics into account has shown to be crucial for the choice of the most appropriate learning strategy.

While new and more sophisticated learning approaches are constantly being explored, including ensemble learning and deep learning methods, many questions remain unanswered about their large-scale applicability and scalability in real-world scenarios. As well, the complexity of real-world data constantly poses new challenges for both researchers and practitioners.

This Special Issue on “*Emerging Trends and Challenges in Supervised Learning Tasks*” aims at discussing open problems and research directions in this area, especially from an application-oriented perspective. Among the ten submissions received, which went through a rigorous peer-review process, five papers have been selected for publication. A brief overview of their content is provided in what follows.

The paper “*Popularity Prediction of Instagram Posts*” [1] addresses the challenge of predicting the popularity of a future post on Instagram, by defining the problem as a classification task and by proposing an original approach based on gradient boosting and feature engineering. Such an approach exploits big data technologies for scalability and efficiency, and it is general enough to be applied to other social media as well.

The paper “*Coming to Grips with Age Prediction on Imbalanced Multimodal Community Question Answering Data*” [2] explores the use of supervised learning techniques to automatically identify age groups across community question answering (cQA) platforms. Distinct input modalities are considered, i.e., texts, profile images and metadata, discussing the effects of imbalanced class distributions, which may severely hurt performance regardless of the modality, and providing insight into how mitigating such effects.

The paper “*Individualism or Collectivism: A Reinforcement Learning Mechanism for Vaccination Decisions*” [3] investigates human voluntary vaccination through making decisions



**Citation:** Pes, B. Special Issue on Emerging Trends and Challenges in Supervised Learning Tasks. *Information* **2021**, *12*, 481. <https://doi.org/10.3390/info12110481>

Received: 17 November 2021

Accepted: 18 November 2021

Published: 19 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

based on the perspectives of individuals themselves (individualistic strategy) or individuals' local groups (collectivist strategy). Accordingly, a reinforcement learning mechanism is proposed that aims at improving the vaccine coverage level of communities. Each individual can adaptively pick one of the two strategies, individualistic or collectivist, after weighing their probabilities with a two-layer neural network whose parameters are dynamically updated with his/her more and more vaccination experience.

The paper “*Learning from High-Dimensional and Class-Imbalanced Datasets Using Random Forests*” [4] focuses on challenging classification tasks where the imbalanced distribution of the data instances is coupled with a high-dimensional feature space, which may severely degrade the generalization performance of the induced models. Leveraging the well-known Random Forest classifier, the paper investigates the effectiveness of hybrid learning strategies that involve the integration of feature selection techniques, to reduce the data dimensionality, with proper methods that cope with the adverse effects of class imbalance (in particular, both data balancing and cost-sensitive methods are considered).

The paper “*VERONICA: Visual Analytics for Identifying Feature Groups in Disease Classification*” [5] introduces a visual analytics system, called VERONICA, that utilizes the natural classification of features in electronic health records (EHRs) to identify the group of features with the strongest predictive power. VERONICA incorporates a representative set of supervised machine learning techniques—namely, classification and regression tree, C5.0, Random Forest, Support Vector Machines, and Naive Bayes to support users in developing predictive models using EHRs. The analytics results are accessible through an interactive visual interface.

**Funding:** This special issue was partially supported by Fondazione di Sardegna, within the projects “ADAM” (CUP: F74I19000900007) and “ASTRID” (CUP: F75F21001220007).

**Acknowledgments:** The guest editor would like to thank the authors for their valuable submissions, the MDPI editors for their constant support in the special issue production, and the reviewers for their precious and constructive comments.

**Conflicts of Interest:** The guest editor declares no conflict of interest.

## References

1. Carta, S.; Podda, A.S.; Recupero, D.R.; Saia, R.; Usai, G. Popularity Prediction of Instagram Posts. *Information* **2020**, *11*, 453. [[CrossRef](#)]
2. Figueroa, A.; Peralta, B.; Nicolis, O. Coming to Grips with Age Prediction on Imbalanced Multimodal Community Question Answering Data. *Information* **2021**, *12*, 48. [[CrossRef](#)]
3. Wu, C.; Qiao, T.; Qiu, H.; Shi, B.; Bao, Q. Individualism or Collectivism: A Reinforcement Learning Mechanism for Vaccination Decisions. *Information* **2021**, *12*, 66. [[CrossRef](#)]
4. Pes, B. Learning from High-Dimensional and Class-Imbalanced Datasets Using Random Forests. *Information* **2021**, *12*, 286. [[CrossRef](#)]
5. Rostamzadeh, N.; Abdullah, S.S.; Sedig, K.; Garg, A.X.; McArthur, E. VERONICA: Visual Analytics for Identifying Feature Groups in Disease Classification. *Information* **2021**, *12*, 344. [[CrossRef](#)]