Special Issue "Data Science in Insurance"

Gian Paolo Clemente [1,*] , **Francesco Della Corte** [1] , **Nino Savelli** [1] and **Diego Zappa** [2]

[1] Department of Mathematics for Economic, Financial and Actuarial Sciences, Università Cattolica del Sacro Cuore, 20123 Milano, Italy; francesco.dellacorte1@unicatt.it (F.D.C.); nino.savelli@unicatt.it (N.S.)

[2] Department of Statistical Sciences, Università Cattolica del Sacro Cuore, 20123 Milano, Italy; diego.zappa@unicatt.it

[*] Correspondence: gianpaolo.clemente@unicatt.it

Within the insurance field, the digital revolution has enabled the collection and storage of large quantities of information. This era is referred to as "big data", since the great uncertainty to be modelled is too complex for traditional data processing techniques. For insurance purposes, big data refers to unstructured and/or structured data being used to influence underwriting, rating, pricing, forms, marketing and claims handling.

The use of artificial intelligence (AI) in actuarial science is a rapidly growing area that is significantly impacting the insurance industry. This integration of technology involves both machine learning (ML) algorithms and data analysis to help actuaries develop more accurate risk assessments and predictions. With a larger possibility of accessing huge sets of data, actuaries can leverage AI to automate underwriting and claims processing, improve customer experience and provide valuable insights to insurers to perfect decision-making strategies. In this way, AI has the potential to transform traditional actuarial methodologies and provide customized solutions for individual clients.

In particular, the field of insurance has greatly benefited from the application of data science. With the use of data analytics, insurance companies can now develop predictive models to better manage risk, underwrite policies more accurately and determine valid pricing. Data science is also used to automate specific procedures such as claims processing, fraud detection, and customer retention. Insurance companies may use data science to identify and build new products that better match clients' needs and preferences. In addition, the accessibility of data can improve operational efficiency by identifying areas of improvement and revenue-generating opportunities. Overall, the use of data science is quickly becoming a crucial aspect of the insurance industry, helping insurers improve their decision making, create new products and services and enhance customer satisfaction.

This Special Issue covers methodologies and methods focused on the application of data science in the insurance and financial context. Owens et al. (2022) focus on Explainable Artificial Intelligence (XAI) (see Clinciu and Hastie 2019; Barredo Arrieta et al. 2020), which refers to the development of artificial intelligence systems that provide a clear and concise explanation of how their decisions are made. The demand for the production of more transparent models, the need fpr techniques that allow for humans to interact with them and the trustworthy inferences from such transparent models are the main justifications for the development of XAI. Starting from this consideration, the authors provide a systematic review about current applications of XAI within the insurance industry, which will contribute to the interdisciplinary understanding of applied XAI.

Sriram et al. (2023) tackle data preparation and cleaning by providing a novel application based on AI techniques and ML system architectures. In particular, the authors focus on policy listings data that pose their own unique challenges, and develop a holistic AI-based platform that standardize, improve and automate the data preparation of insurance through machine learning. Secondly, a rule-based, pairwise corporation entity resolution framework is provided that allows standardization of insured entities, enabling policy aggregations.

Neural networks have been exploited in Flaig and Junike (2022) and Jose et al. (2022). Generative adversarial networks (Goodfellow et al. 2014) are applied in Flaig and Junike (2022) to expand the scenario generation process to a complete market risk calculation for Solvency II purposes. The study shows that the proposed approach can represent a viable alternative method for market risk modelling beyond traditional economic scenario generators, which can also serve as regulatory-approved models, as they perform well in the EIOPA benchmarking study.

In Jose et al. (2022), an ensemble of statistical predictive models is developed in order to predict admission rates to hospitals (or other health facilities) related to respiratory diseases in a US population. The results indicate that the neural network-based models have better predictive performances compared to traditional GLM-type models. The proposed approach is mainly based on feed-forward neural network (FFNN) models and a combined actuarial neural network (CANN) approach (see Schelldorfer and Wuthrich 2019) under a negative binomial distributional assumption (Tzougas and Li 2021). Additionally, the approach exploits the advantages of the bias-regularised version of the negative binomial FFNN and CANN models Wüthrich (2019) and of the setup provided in Richman and Wüthrich (2023).

A support vector machine (SVM) is applied in Asimit et al. (2022). The main purpose of this paper is to reduce the effect of feature noise for binary SVM classifiers. The authors explore the internal structure of the classical SVM classifier to detect and tackle the feature noise via probabilistic arguments. In this paper, two powerful SVM-type classification algorithms are developed and discussed: the Single Perturbation SVM and Extreme Empirical Loss SVM. The authors conducted a large set of numerical experiments to test their effectiveness on synthetic and real-world datasets, both with and without noise contamination.

In Sangari et al. (2022), the authors provide a methodology which aims to correct under-reporting in cyber incidents in more than one dimension: revenue, event type and industry. The proposed approach allows us to quantify the extent of under-reporting in data sets of public cyber incidents.

In Boucher (2022b), the creation of different Bonus-Malus scales (BMS) for each type of insurance using recursive partitioning methods is proposed. Using a recursive division algorithm (see Diao and Weng 2019), several groups of the insured are created to apply a separate BMS model for each group. Recent BMS models, provided in the literature (Boucher 2022a; Boucher and Inoussa 2014; Verschuren 2021), generate the same surcharges and the same discounts for all policyholders because the transition rules within the class system do not depend on the a priori risk. In Boucher (2022b), insurers can create different BMSs to account for the differences in a priori risk.

In conclusion, this Special Issue highlights several applications of AI and ML in insurance. The AI is transforming the insurance industry by creating opportunities to improve risk management, automate key processes and gain deeper insights into customer behaviour. By leveraging data science, insurers can develop predictive models to better understand risk, automate underwriting and claims processing and create new products tailored to individual customer needs. Moreover, through the development of Explainable Artificial Intelligence, insurers can unlock the black box of AI decision making and build trust with customers and other stakeholders. The impact of AI on the insurance industry is very deep, and its full potential is only just starting to be realized.

Conflicts of Interest: The authors declare no conflict of interest.

References

Asimit, Alexandru V., Ioannis Kyriakou, Simone Santoni, Salvatore Scognamiglio, and Rui Zhu. 2022. Robust classification via support vector machines. *Risks* 10: 154. [CrossRef]

Barredo Arrieta, Alejandro, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, and et al. 2020. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion* 58: 82–115. [CrossRef]

Boucher, Jean-Philippe. 2022a. Bonus-malus scale models: Creating artificial past claims history. *Annals of Actuarial Science* 17: 1–27. [CrossRef]

Boucher, Jean-Philippe. 2022b. Multiple bonus-malus scale models for insureds of different sizes. *Risks* 10: 152. [CrossRef]

Boucher, Jean-Philippe, and Rofick Inoussa. 2014. A posteriori ratemaking with panel data. *ASTIN Bulletin* 44: 587–612. [CrossRef]

Clinciu, Miruna, and Helen Hastie. 2019. A survey of explainable ai terminology. In *Proceedings of the 1st Workshop on Interactive Natural Language Technology for Explainable Artificial Intelligence*. Stroudsburg: Association for Computational Linguistics, pp. 8–13. [CrossRef]

Diao, Liqun, and Chengguo Weng. 2019. Regression tree credibility model. *North American Actuarial Journal* 23: 1–28. [CrossRef]

Flaig, Solveig, and Gero Junike. 2022. Scenario generation for market risk models using generative neural networks. *Risks* 10: 199. [CrossRef]

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial networks. *Advances in Neural Information Processing Systems* 3: 139–44. [CrossRef]

Jose, Alex, Angus S. Macdonald, George Tzougas, and George Streftaris. 2022. A combined neural network approach for the prediction of admission rates related to respiratory diseases. *Risks* 10: 217. [CrossRef]

Owens, Emer, Barry Sheehan, Martin Mullins, Martin Cunneen, Juliane Ressel, and German Castignani. 2022. Explainable artificial intelligence (xai) in insurance. *Risks* 10: 230. [CrossRef]

Richman, Ronald, and Mario V. Wüthrich. 2023. Localglmnet: Interpretable deep learning for tabular data. *Scandinavian Actuarial Journal* 2023: 71–95. [CrossRef]

Sangari, Seema, Eric Dallal, and Michael Whitman. 2022. Modeling under-reporting in cyber incidents. *Risks* 10: 200. [CrossRef]

Schelldorfer, Jürg, and Mario Wuthrich. 2019. Nesting classical actuarial models into neural networks. *SSRN Electronic Journal*. [CrossRef]

Sriram, Varun, Zijie Fan, and Ni Liu. 2023. Eclipse: Holistic ai system for preparing insurer policy data. *Risks* 11: 4. [CrossRef]

Tzougas, George, and Ziyi Li. 2021. *Neural Network Embedding of the Mixed Poisson Regression Model for Claim Counts*. Available online: https://insurancedatascience.org/project/2021_london/ (accessed on 21 April 2023).

Verschuren, Robert. 2021. Predictive claim scores for dynamic multi-product risk classification in insurance. *Astin Bulletin* 51: 1–25. [CrossRef]

Wüthrich, Mario. 2019. Bias regularization in neural network models for general insurance pricing. *European Actuarial Journal* 10: 179–202. [CrossRef]