World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:13, No:04, 2019

Hybrid Model: An Integration of Machine Learning with Traditional Scorecards

Authors : Golnush Masghati-Amoli, Paul Chin

Abstract : Over the past recent years, with the rapid increases in data availability and computing power, Machine Learning (ML) techniques have been called on in a range of different industries for their strong predictive capability. However, the use of Machine Learning in commercial banking has been limited due to a special challenge imposed by numerous regulations that require lenders to be able to explain their analytic models, not only to regulators but often to consumers. In other words, although Machine Leaning techniques enable better prediction with a higher level of accuracy, in comparison with other industries, they are adopted less frequently in commercial banking especially for scoring purposes. This is due to the fact that Machine Learning techniques are often considered as a black box and fail to provide information on why a certain risk score is given to a customer. In order to bridge this gap between the explain-ability and performance of Machine Learning techniques, a Hybrid Model is developed at Dun and Bradstreet that is focused on blending Machine Learning algorithms with traditional approaches such as scorecards. The Hybrid Model maximizes efficiency of traditional scorecards by merging its practical benefits, such as explain-ability and the ability to input domain knowledge, with the deep insights of Machine Learning techniques which can uncover patterns scorecard approaches cannot. First, through development of Machine Learning models, engineered features and latent variables and feature interactions that demonstrate high information value in the prediction of customer risk are identified. Then, these features are employed to introduce observed non-linear relationships between the explanatory and dependent variables into traditional scorecards. Moreover, instead of directly computing the Weight of Evidence (WoE) from good and bad data points, the Hybrid Model tries to match the score distribution generated by a Machine Learning algorithm, which ends up providing an estimate of the WoE for each bin. This capability helps to build powerful scorecards with sparse cases that cannot be achieved with traditional approaches. The proposed Hybrid Model is tested on different portfolios where a significant gap is observed between the performance of traditional scorecards and Machine Learning models. The result of analysis shows that Hybrid Model can improve the performance of traditional scorecards by introducing non-linear relationships between explanatory and target variables from Machine Learning models into traditional scorecards. Also, it is observed that in some scenarios the Hybrid Model can be almost as predictive as the Machine Learning techniques while being as transparent as traditional scorecards. Therefore, it is concluded that, with the use of Hybrid Model, Machine Learning algorithms can be used in the commercial banking industry without being concerned with difficulties in explaining the models for regulatory purposes.

Keywords : machine learning algorithms, scorecard, commercial banking, consumer risk, feature engineering
Conference Title : ICCRMDA 2019 : International Conference on Credit Risk Modeling and Data Analysis
Conference Location : Paris, France
Conference Dates : April 18-19, 2019

Open Science Index, Mathematical and Computational Sciences Vol:13, No:04, 2019 publications.waset.org/abstracts/105480.pdf