RATE-Analytics: Next Generation Predictive Analytics for Data-Driven Banking and Insurance

Dennis Collaris
Eindhoven University of Technology, The Netherlands

Mykola Pechenizkiy
Eindhoven University of Technology, The Netherlands

Jarke J. van Wijk
Eindhoven University of Technology, The Netherlands

Abstract

We conducted the RATE-Analytics project: a unique collaboration between Rabobank, Achmea, Tilburg and Eindhoven University. We aimed to develop foundations and techniques for next generation big data analytics. The main challenge of existing approaches is the lack of reliability and trustworthiness: if experts do not trust a model or its predictions they are much less likely to use and rely on that model. Hence, we focused on solutions to bring the human-in-the-loop, enabling the diagnostics and refinement of models, and support in decision making and justification. This chapter zooms in on the part of the project focused on developing explainable and trustworthy machine learning techniques.

2012 ACM Subject Classification Computing methodologies → Philosophical/theoretical foundations of artificial intelligence; Computing methodologies → Machine learning; Human-centered computing

Keywords and phrases Visualization, Visual Analytics, Machine Learning, Interpretability, Explainability, XAI

Digital Object Identifier 10.4230/OASIcs.Commit2Data.2024.8

Acknowledgements We would like to thank numerous colleagues of the RATE project team at Tilburg University, Rabobank, Achmea, and TU/e who over the years provided endless support and facilitated collaboration. We would like to thank NWO, Rabobank and Achmea for the provided funding. This work is part of the research programme Commit2Data, specifically the RATE Analytics project with project number 628.003.001. Last, but not least we would like to thank the reviewers for providing constructive feedback.

1 Introduction

Banks and insurance companies try to make use of increasingly large volumes of customer-related data as new in-house and external data sources become available for analysis. To seize this opportunity, we conducted the RATE-Analytics project: a unique collaboration between Rabobank, Achmea, Tilburg and Eindhoven University. The overall goal of this project was to develop foundations and techniques for next generation big data analytics, facilitating the development of a wide range of applications in banking and insurance. This includes, but is not limited to: intrusion detection, transactions and insurance fraud detection, loan default prediction, prediction of extreme (insurance) claims, and data-driven product development.

In the algorithms and use cases we studied, the overarching theme is to efficiently combine (large amounts of) data and expert knowledge, where the (economic) cost of the former is rapidly decreasing and the cost of the latter increasing. We, therefore, focus on the expert in the loop, or the expert-algorithm interaction to define what a "good" algorithm is. We bring three major data-science expertise areas into this multidisciplinary project: predictive analytics, modern statistics, and visual analytics.

Commit2Data.
Editors: Boudewijn R. Haverkort, Aldert de Jongste, Pieter van Kuilenburg, and Ruben D. Vromans; Article No. 8; pp. 8:1–8:11
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

With this unique combination, we aim at breakthroughs in data-driven banking and insurance, facilitating the development of more reliable, transparent, and responsible analytics solutions and products.

In this chapter, we zoom in on the part of the project focused on developing explainable and trustworthy machine learning techniques. We focus on solutions to bring the human-in-the-loop, enabling the diagnostics and refinement of machine learning models, and support in decision making and justification.

The remainder of this chapter is structured as follows: we describe the problem addressed in Section 2. Next, in Section 3 we describe the consortium structure and way-of-working. The results of the project are highlighted in Section 4 both from a technical, and collaborative perspective. Finally, in Section 5 we conclude by discussing future prospects for the project.

2 Problem addressed

The RATE project addresses the problem of analyzing large volumes of customer-related data in the banking and insurance industries. The various sources of data available at Achmea and Rabobank contain a wealth of information about customer behavior and events. It would be advantageous (academically, for society, and for Achmea/Rabobank) to develop automated methods that take these data and analyze the possible benefits for customers of new finance and insurance products. Concretely, we focus on the problem that existing techniques to achieve this often lack transparency, explainability and trustworthiness. By involving the human-in-the-loop, we can address this and enable the diagnostics and refinement of machine learning models, and support in decision making and justification.

2.1 Challenges

The main challenge of existing data-driven approaches from the application perspective is the lack of reliability and trustworthiness: if experts do not trust a model or its particular output they are much less likely to use and rely on that model. This challenge is not unique for finance and insurance: in many other areas issues such as trustworthiness, explainability, and integration of expert knowledge play an important role. The RATE project provided, from an academic perspective, a unique opportunity to work on these challenges for real-world cases in a huge and complex domain, in close collaboration with the stakeholders.

Without a proper understanding of a machine learning model, there are several problems that can occur in the data science pipeline that may go unnoticed (summarized in Figure 1).

- **ML may be trained on unrepresentative data.** For instance, it could be based on biases, spurious correlations, and false generalizations [16]. As an example, recent work has shown that the accuracy of commercial gender classification on dark-skinned females is significantly worse than on any other group. This discrepancy was found to be largely due to unrepresentative training datasets and imbalanced test benchmarks [4].
- **ML may be using inadequate features.** There may be sufficient and representative data available, but the model could use this data in unexpected ways. Machine learning is only able to ascertain correlation amongst features, and is not able to find causal relationships. This is demonstrated in the wolf-husky problem introduced by Ribeiro et al. [23]. The authors show an example of a husky and wolf classifier, that turns out to detect the type of animal based on snow in the background of an image, instead of looking at the animal itself.

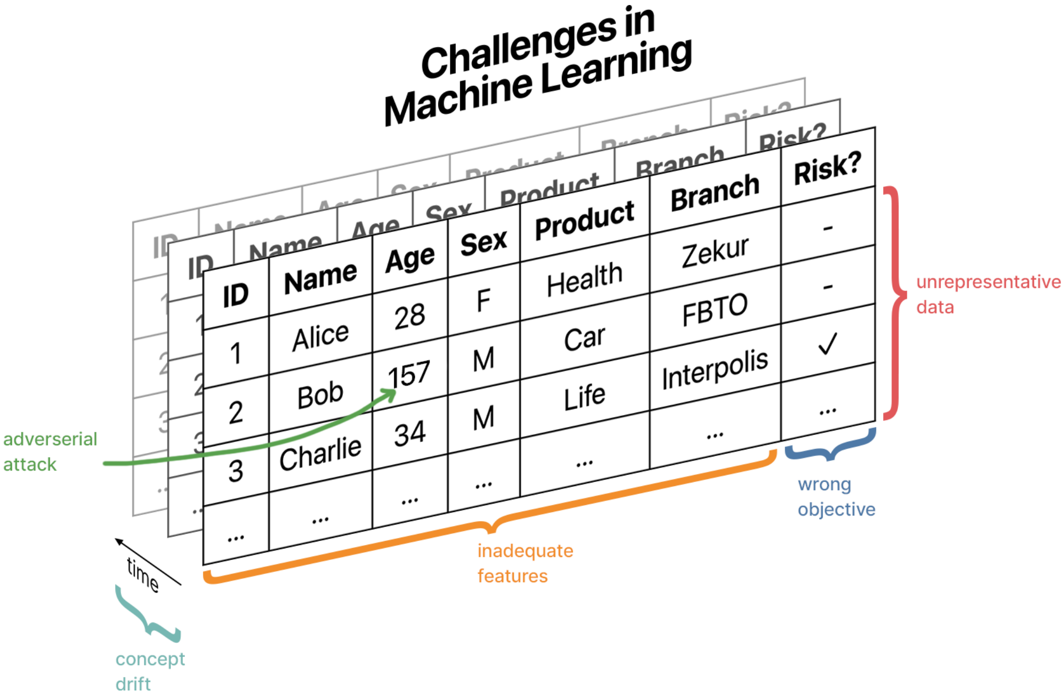

Figure 1 There are several problems that can occur in the data science pipeline that may go unnoticed without a proper understanding of the behavior of a machine learning model.

- **ML may have the wrong objective.** Even if the model scores well on a test set with any of the many performance metrics available (e.g., accuracy, precision, recall, F_1), it may still be trained to optimize the wrong objective, like trying to match doubtful labels. Consider a fraud detection model in insurance. Since no ground truth information is available about which insurance policies are fraudulent, models are typically trained with labels from human fraud experts. In such a scenario, the machine may adopt any biases the fraud experts may have, explicit or implicit. In addition, this model can only be expected to match the human performance, not exceed it.
- **ML may not be robust against concept drift.** The test set generalization during development may not match with future unseen data. This problem was the reason for the failure of the Google Flu Trends model [5], which showed promising results, but failed to predict flu trends in practice. This problem is wide-spread across real-world applications of predictive analytics [26], and is very predominant in adversarial domains (e.g., spam detection, fraud detection).
- **ML may be subject to adversarial attacks.** There is a surge of recent works showing that models may be vulnerable to adversarial attacks. For instance, authors have shown that a small perturbation in the input (e.g., a single pixel in an image or imperceptible noise) can lead to unexpected, extreme changes in the output, often leading to absurd or incorrect predictions [2, 20, 17].

2.2 Application areas

RATE-Analytics aims to address previously mentioned problems by supporting the understanding of complex machine learning models. The stakeholders can use machine learning explanations for a wide variety of applications. We identified four main categories based on discussions with data scientists at Achmea:

1. **Diagnostics** The model may not perform adequately, even though the model scores well on a test set. It could be based on biases and spurious correlations [16]. Explanations can highlight these issues such that experts can address these during model development.
2. **Refinement** Apart from identifying issues with the model, explanations can also help to improve the model. Analyzing explanations for incorrect predictions can yield insights into how to increase predictive accuracy [1, 24] or remove irrelevant features.
3. **Decision support** In high-stakes decision-making, where models make decisions that have a critical impact on real people, it is not sufficient to base decisions on the prediction score of the model alone. To qualitatively ascertain whether desiderata such as fairness, privacy, and trust are met, explanations can help internal decision-makers to verify the behavior of the model [13].
4. **Justification** Various external stakeholders may have questions about predictions by the model. For instance, customers subject to predictions may request justification, or authorities may request information to check compliance. The latter got more relevant since the recent introduction of the GDPR's "right to explanation" [14] and AI Act [15].

The results of the RATE-Anlytics project support data scientists in all these applications through providing a better understanding of machine learning models.

3 Consortium structure

The academic contribution to the project consortium consisted of internationally recognized experts in predictive analytics, visual analytics, and modern statistics who actively participate in national (NWO, STW, COMMIT) and European projects and collaborate with industry. This is a unique combination of three core data science areas in one project. Complementary expertise relevant to the project includes health economics, applied microeconomics, financial econometrics; regulation and technology; privacy, fundamental rights and Internet, and cybersecurity. All project members have, in their own field, outstanding track records and most have experience with multidisciplinary research. Besides this project, the collaboration between TU/e and TiU is facilitated by their joint data science initiative, banking and insurance being defined as a strategic area of interest. On the academic leadership side of the project:

- *Mykola Pechenizkiy* is chairing the Data and Artificial Intelligence cluster. He is leading trustworthy and responsible AI research program.
- *Bas Werker* has extensive academic leadership experience and currently works on various projects on the interplay of statistics, financial econometrics and finance and insurance. In particular, he's involved in Netspar and the committee to provide longevity predictions for the Dutch Actuarial Society.
- *Jack van Wijk* was leading the visualization group at TU/e until 2022. His group has done award-winning research on visualization. He has tight collaboration with industry through joint projects, for instance with Philips, Thales, and with spin-off companies MagnaView and SynerScope (not involved in the RATE-Analytics project). He was scientific director of the Data Science Center Eindhoven (DSC/e) from 2018 to 2020.

To represent the application domain, RATE-Analytics involved two large financial institutions in The Netherlands to provide real-world, societally relevant use cases. This ensures the research within the project generalizes and addresses a real need in industry.

- *Achmea* is a leading insurance company in The Netherlands, providing various insurance products, including life, health, and property and casualty insurance. Since the company provides a significant social value, it is under heavy regulatory scrutiny. They have a keen interest in solutions to help built trust in machine learning models.

- *Rabobank* is a Dutch bank that provides a range of financial services to individuals and businesses. It is one of the largest banks in the Netherlands. Rabobank has a global presence and operates in several countries around the world, offering a variety of banking services including loans, savings accounts, and investment products. They are specifically interested in fraud detection and predictive analytics for Know Your Customer.

In all steps along the process, Achmea and Rabobank were closely involved. The choice of specific cases and problems to address was often based on discussions with experts at these companies. For most projects, real-world data was provided to make sure our solutions generalized to their application domain, and the solutions were tested in practice at these companies. This provided actionable insights for the companies and strengthened the motivation for the academic results.

It has become evident that meaningful application of machine learning-based solutions across a wide range of considered and envisioned use cases requires certain explainability for different stakeholders, including domain experts, data scientists, and customers. We zoom in into visual analytics solutions we developed throughout the project that facilitated it and and demonstrated further promise in valorization of research results.

4 Key results

To bring the human expert in the loop and address the lack of reliability and trustworthiness of current approaches, we explored interactive visualization for machine learning. Each of our works is motivated by concrete questions from data scientists at the industrial partners, and each system has also been tested and applied in practice at projects within the company. We learned that there is no silver bullet to solve all problems. Different perspectives require different solutions, ranging from local explanation of single predictions to global explanations of the entire model. Figure 2 provides an overview of how each perspective is addressed.

4.1 Technical results

We first designed and developed ExplainExplore [8]: an interactive explanation system to explore explanations of individual predictions (i.e., local). For each explanation, it provides context by presenting similar predictions, and showing the impact of small input perturbations. We recognize many different explanations may exist that are all equally valid and useful using traditional evaluation methods. Hence, we leveraged the domain knowledge of the data scientist to determine which of these fit their preference. The local perspective is particularly useful for **decision support** and **justification** applications. In a use case with data scientists from the debtor management department at Achmea, we showed the participants could effectively use explanations to diagnose a model and find problems, identify areas where the model can be improved, and support their everyday decision-making process. To ensure these contributions can be broadly applied, we also built a software library [6] that enables interoperability with a wide range of different languages, toolkits, and enterprise software.

Next, we proposed the Contribution-Value plot [9, 10] as a new elementary building block for interpretability visualization, showing how feature contribution changes for different feature values. It provides a perspective in between local and global, as the model behavior is shown for all instances, but visualized on a per-feature basis. This perspective primarily helps to **diagnose** unexpected model behavior. In a quantitative online survey with 22 machine learning professionals and visualization experts, we show our visualization increases

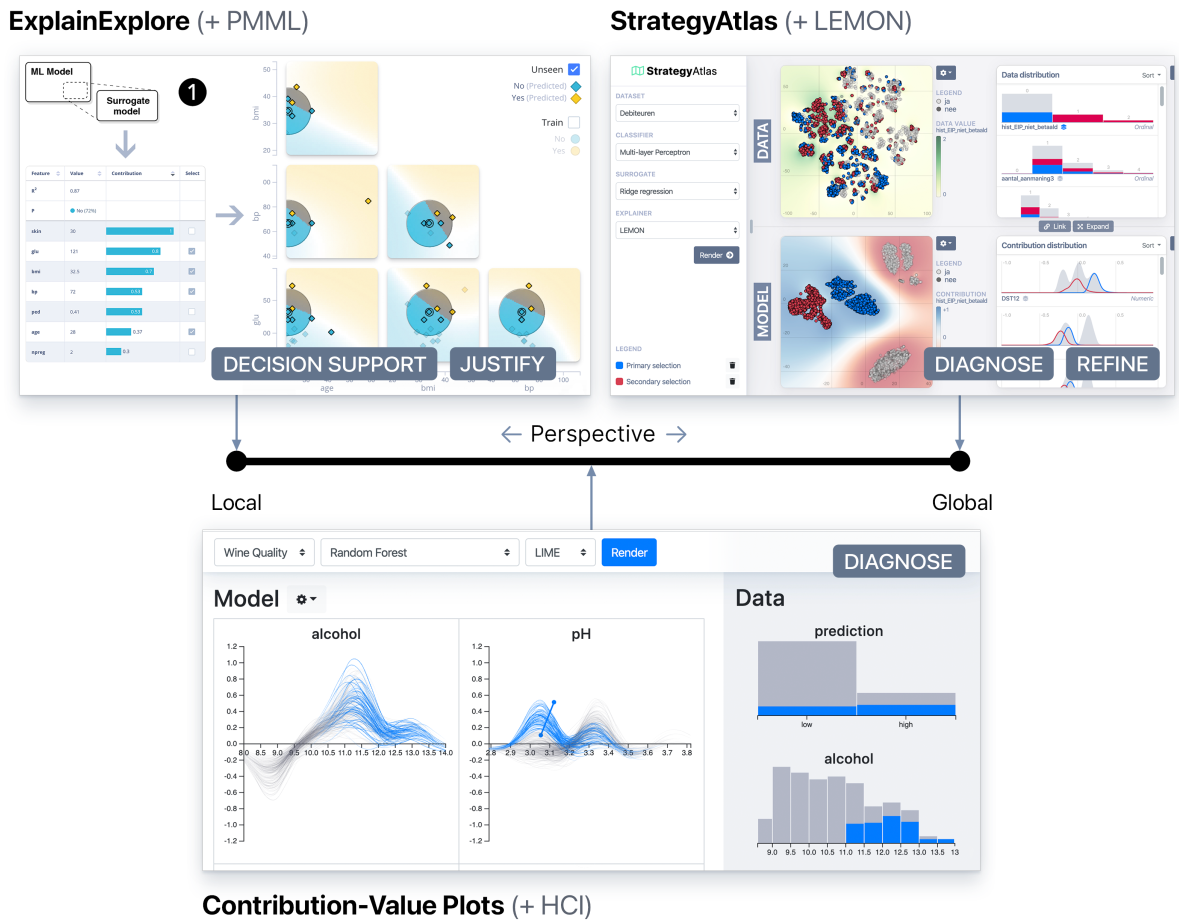

Figure 2 Overview of the different perspectives and corresponding systems developed in the RATE-Analytics project. ExplainExplore (←) provides local explanations of individual predictions with a bar chart, and provides context with class activation heatmaps; CV plots (↑) explain how feature improtance changes for different feature values; and StrategyAtlas (→) provides a global explanation of the entire model through model strategies that highlight different model behavior for subgroups. The three core VA solutions are accompanied by complementary contributions outside the visualization domain.

correctness and confidence and reduces the time needed to obtain an insight compared to previous techniques. This work highlighted that a small difference in feature importance techniques can result in a large difference in interpretation, and warranted a follow-up human-computer interaction contribution to characterize the data scientists' mental model of explanations [12], and explore the differences between existing techniques.

Finally, we designed and developed StrategyAtlas [11]: a visual analytics approach to enable a global understanding of complex machine learning models through the identification and interpretation of different model strategies. These model strategies are identified in our projection-based StrategyMap visualization. Data scientists are enabled to ascertain the validity of these strategies through analyzing feature values and contributions using heat maps, density plots, and decision tree abstractions. This global perspective is especially beneficial to **diagnose** and **refine** models. In collaboration with Achmea, we applied the system in a real-world project for automatic insurance acceptance. This showed that professional data scientists were able to understand a complex model and improve the production model based on these insights. As computing the local feature importance values for an entire dataset is computationally expensive, we complemented this work with an algorithmic contribution called LEMON [7] to improve the faithfulness of explanation results, which enables us to significantly speed up computations of StrategyMap projections.

Although the individual visualizations introduced our work are simple in nature, we have shown that the visualizations *combined* form a strong visual encoding that enables data scientists to gain an understanding of machine learning models which was not possible before. This enables data scientists to diagnose, refine models, and support decision-making and justification of predictions. Its not the individual visualizations, but the full picture - the entire interactive workflow - that yields insights.

In addition to our interactive visualization contributions, we also studied how to deal with data that is not independent and identically distributed, i.e. so-called non-IIDness. When applying predictive analytics in real-world applications, complex behavior exhibits lots of heterogeneity, including temporal dynamics and coupling relationships.

Effective anomaly detection in financial transaction networks is particularly challenging because there exist collective fraudulent behaviors at the level of subgraphs rather than individual nodes. A ring structure for money laundering and a tree structure for pyramid schemes would be common examples. In practice it is difficult to decide which features are more representative beforehand. We introduced a subgraph anomaly detection framework that allows to preserve both the local structure of subgraphs and the global structure of entire network by making use of global roles and local connections of nodes [22]. We demonstrated on both synthetic and real-world financial transaction network datasets the effectiveness of learning subgraph embeddings without requiring any prior knowledge and detecting anomalous subgraphs. In [21] we proposed a novel way of employing attention-based deep residual modeling approach that can effectively detecting anomalous nodes in attributed networks [18] by capturing the sparsity and nonlinearity, reducing the adverse effect from anomalous nodes and preventing so-called over-smoothing in representation learning. Our experiments on several real-world attributed networks demonstrated the effectiveness of detecting anomalies in these settings.

Learning from multiple time-series over an unbounded time-frame requires machine learning techniques to continuously learn from data that evolves over time, exhibits concept drifts. We explored ensembles learning frameworks that allow to capture changes in the concepts [19]. One of challenges is that supervised learning for modeling such second order dynamics is often not fully applicable since the true labels (e.g. default clients, fraudulent behavior etc) become known with considerable delay. We studied the potential of employing labelless concept drift detection, visual inspection, and explanation with use local feature importance that is used for constructing a new representation allowing effective drift detection while maintaining acceptable level of false positives [25].

All of the described work has the potential to be widely applied in various industries and domains where machine learning is used. This includes other projects at Achmea and Rabobank, but also other domains beyond the scope of RATE-Analytics. Especially in high-stakes decision-making, where models make decisions that have a critical impact on real people, it is not sufficient to base decisions on the prediction score of the model alone. To qualitatively ascertain whether desiderata such as fairness, privacy, and trust are met, our results can help to verify the behavior of models [13], through enabling understanding of complex models. To reach these other domains, we made a promotional website[1] containing simple explanations of our work, along with conference presentations, online demos and source code.

4.2 Collaboration results

The RATE-Analytics project has a rather unique positioning: aiming to advance the research area of predictive analytics, and to have an immediate impact on its acceptance in the industry. Apart from our academic contributions, the tight collaboration between universities and industry proved most fruitful.

The researchers gave several presentations at Achmea and Rabobank. This includes data science teams, senior management, innovation managers, and the Chief Information Officer (CIO) of Achmea. These presentations contributed to the ongoing internal discussion on how to use AI responsibly within Achmea and across all insurance companies in the Netherlands (through the "Verbond Van Verzekeraars").

It also enabled the industry partners to immediately incorporate and apply state-of-the-art research. The debtor management department at Achmea used ExplainExplore to analyze their model built to predict the effectiveness of a debtor contact strategy, and found that the model used several attributes in the data they did not expect were relevant for the prediction. This insight could helpt them refine their current model. Next, data scientists at Achmea used StrategyAtlas with an automatic acceptance model for car insurances. They system helped them realize their model did not detect bad actors and removed them, but rather looked at the characteristics of good customers, which was contrary to their prior beliefs.

Conversely, Achmea and Rabobank provided relevant use cases for evaluating our approaches, such as the ones mentioned above. This provided the academic partners with a strong motivation for the projects and evaluation of the published work.

[1] `https://explaining.ml`. When offline, an archived version of the website can be found at `https://web.archive.org/web/20230515141514/https://explaining.ml/`.

5 Future prospects

Since the RATE-Analytics project aimed for developing generic technology, there is a potential of utilizing the outcome of the project in other domains beyond banking and insurance. E.g. the three core research problems we formulate are relevant to the area of big data in medical diagnostics and predictive maintenance tasks in production industries.

In general, the results of this project can help companies to develop more reliable (high detection rate), more cost-effective (helping human analysts with the right tools to deal with typically large number of false alarms), and more transparent and interpretable analytics.

To explore this idea further, we applied and were granted admission to the NWO Take-off programme (phase 1). Participation in this programme and the corresponding €40.000 subsidy will help us investigate the feasibility of a startup company to help businesses understand and explain complex machine learning models. We will mainly focus on conducting interviews with potential customers and end-users, to ascertain 1) if black box models are indeed considered a problem, and by who; 2) which roles in the organisation are responsible for the buying decision (decision-making unit); 3) if our solution sufficiently addresses the customers problem (i.e., if our explanations can yield business value in practice), and 4) to find commercial commitment from an initial launching customer.

The envisioned startup helps companies that conduct data science for high-stakes decision making who want to understand machine learning models by 1) reducing the risk of (invisible) unethical decisions, and 2) reducing the risk of non-compliance with current (GDPR) and upcoming (AI Act) regulations that will restrict how companies can apply data science. This startup enables us to continue developing our solutions in industry context, make available our solutions to a broader audience, thus have a bigger impact on society, and offer Achmea and Rabobank continued support and benefit from the RATE-Analytics project results.

The RATE-Analytics project provided several benefits to these valorization efforts. The fact that the RATE-Analytics project addresses real problems in industry provides some immediate problem validation, a key first step in the development of a startup company [3]. Next, because the research contributions were all evaluated in practice at the companies, we could immediately assess the Technology Readiness Level (TRL) at grade 5. This is a helpful method for estimating the maturity of technologies, and part of the NWO Take-off application process.

To investigate the practical and commercial viability of our business model, we plan to first identify potential customers. Next, through initial meetings with these customers we plan to get early feedback on the current prototypes and business idea. While the current prototypes have been validated in the expected customer segment at Achmea, we need to make sure these prototypes are suitable for other customers too. We anticipate different customers need different levels of detail, and the prototype needs tailoring for those specific needs. Therefore, as a main work within this project, we plan to prepare a convincing and understandable demonstration of our platform to be used as the basis for the future startup.

We are looking forward to developing and expanding our ideas and technology further. The RATE-Analytics project, aimed at developing new technology in close cooperation with companies from finance, was a crucial first step. We hope that next steps will lead to much impact across a variety of application domains.

References

1 Mihael Ankerst, Martin Ester, and Hans-Peter Kriegel. Towards an effective cooperation of the user and the computer for classification. In *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 179–188. ACM, 2000. doi:10.1145/347090.347124.

2 Battista Biggio, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Srndic, Pavel Laskov, Giorgio Giacinto, and Fabio Roli. Evasion attacks against machine learning at test time. In Hendrik Blockeel, Kristian Kersting, Siegfried Nijssen, and Filip Zelezný, editors, *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III*, volume 8190 of *Lecture Notes in Computer Science*, pages 387–402. Springer, Springer, 2013. doi:10.1007/978-3-642-40994-3_25.

3 Steve Blank. *The four steps to the epiphany: successful strategies for products that win.* John Wiley & Sons, 2020.

4 Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, pages 77–91. PMLR, 2018. URL: http://proceedings.mlr.press/v81/buolamwini18a.html.

5 Declan Butler. When Google got flu wrong. *Nat.*, 494(7436):155–156, 2013. doi:10.1038/494155a.

6 Dennis Collaris. sklearn-pmml-model: Machine learning portability and interoperability using PMML, 2022. Library available at https://github.com/iamDecode/sklearn-pmml-model.

7 Dennis Collaris, Pratik Gajane, Joost Jorritsma, Jarke J van Wijk, and Mykola Pechenizkiy. LEMON: Alternative sampling for more faithful explanation through local surrogate models. In *Advances in Intelligent Data Analysis XXI: 21st International Symposium on Intelligent Data Analysis (IDA 2023)*, pages 77–90. Springer, 2023. doi:10.1007/978-3-031-30047-9_7.

8 Dennis Collaris and Jarke J van Wijk. ExplainExplore: Visual exploration of machine learning explanations. In *Proceedings of the 2020 IEEE Pacific Visualization Symposium (PacificVis)*, pages 26–35. IEEE, 2020. doi:10.1109/PacificVis48177.2020.7090.

9 Dennis Collaris and Jarke J. van Wijk. Machine learning interpretability through contribution-value plots. In Michael Burch, Michel A. Westenberg, Quang Vinh Nguyen, and Ying Zhao, editors, *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction*, pages 4:1–4:5. ACM, 2020. doi:10.1145/3430036.3430067.

10 Dennis Collaris and Jarke J. van Wijk. Comparative evaluation of contribution-value plots for machine learning understanding. *Journal of Visualization*, 25(1):47–57, 2022. doi:10.1007/s12650-021-00776-w.

11 Dennis Collaris and Jarke J. Van Wijk. StrategyAtlas: Strategy analysis for machine learning interpretability. *IEEE Transactions on Visualization and Computer Graphics*, 29(6):2996–3008, 2023. doi:10.1109/TVCG.2022.3146806.

12 Dennis Collaris, Hilde J. P. Weerts, Daphne Miedema, Jarke J. van Wijk, and Mykola Pechenizkiy. Characterizing data scientists' mental models of local feature importance. In *NordiCHI '22: Nordic Human-Computer Interaction Conference, Aarhus, Denmark, October 8 - 12, 2022*, pages 9:1–9:12. ACM, 2022. doi:10.1145/3546155.3546670.

13 Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

14 Bryce Goodman and Seth Flaxman. European union regulations on algorithmic decision-making and a "right to explanation". *AI magazine*, 38(3):50–57, 2017. doi:10.1609/aimag.v38i3.2741.

15 Philipp Hacker and Jan-Hendrik Passoth. Varieties of AI explanations under the law. from the GDPR to the aia, and beyond. In Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, and Wojciech Samek, editors, *xxAI - Beyond Explainable AI - International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria,*

Revised and Extended Papers, volume 13200 of *Lecture Notes in Computer Science*, pages 343–373, Cham, 2020. Springer. doi:10.1007/978-3-031-04083-2_17.

16 Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven Mark Drucker. Gamut: A design probe to understand how data scientists understand machine learning models. In Stephen A. Brewster, Geraldine Fitzpatrick, Anna L. Cox, and Vassilis Kostakos, editors, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, page 579. ACM, 2019. doi:10.1145/3290605.3300809.

17 Tianjin Huang, Vlado Menkovski, Yulong Pei, Yuhao Wang, and Mykola Pechenizkiy. Direction-aggregated attack for transferable adversarial examples. *ACM J. Emerg. Technol. Comput. Syst.*, 18(3):60:1–60:22, 2022. doi:10.1145/3501769.

18 Tianjin Huang, Yulong Pei, Vlado Menkovski, and Mykola Pechenizkiy. Hop-count based self-supervised anomaly detection on attributed networks. In Massih-Reza Amini, Stéphane Canu, Asja Fischer, Tias Guns, Petra Kralj Novak, and Grigorios Tsoumakas, editors, *Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases - , ECML PKDD 2022, Part I*, volume 13713 of *Lecture Notes in Computer Science*, pages 225–241. Springer, 2022. doi:10.1007/978-3-031-26387-3_14.

19 Samaneh Khoshrou and Mykola Pechenizkiy. Adaptive long-term ensemble learning from multiple high-dimensional time-series. In *Discovery Science: 22nd International Conference, DS 2019, Split, Croatia, October 28–30, 2019, Proceedings 22*, pages 511–521. Springer, 2019. doi:10.1007/978-3-030-33778-0_38.

20 Anh Mai Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 427–436. IEEE Computer Society, 2015. doi:10.1109/CVPR.2015.7298640.

21 Yulong Pei, Tianjin Huang, Werner van Ipenburg, and Mykola Pechenizkiy. ResGCN: attention-based deep residual modeling for anomaly detection on attributed networks. *Machine Learning*, 111(2):519–541, 2022. doi:10.1007/s10994-021-06044-0.

22 Yulong Pei, Fang Lyu, Werner van Ipenburg, and Mykola Pechenizkiy. Subgraph anomaly detection in financial transaction networks. In Tucker Balch, editor, *ICAIF '20: The First ACM International Conference on AI in Finance, New York, NY, USA, October 15-16, 2020*, ICAIF '20, pages 18:1–18:8, New York, NY, USA, 2020. ACM. doi:10.1145/3383455.3422548.

23 Marco Túlio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why should I trust you?" explaining the predictions of any classifier. In Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi, editors, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144. ACM, ACM, 2016. doi:10.1145/2939672.2939778.

24 Simone Stumpf, Vidya Rajaram, Lida Li, Weng-Keen Wong, Margaret Burnett, Thomas Dietterich, Erin Sullivan, and Jonathan Herlocker. Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies*, 67(8):639–662, 2009. doi:10.1016/j.ijhcs.2009.03.004.

25 Shihao Zheng, Simon P. van der Zon, Mykola Pechenizkiy, Cassio P. de Campos, Werner van Ipenburg, and Hennie de Harder. Labelless concept drift detection and explanation. In *Proceedings of NeurIPS 2019 Workshop on Robust AI in Financial Service*, 2019.

26 Indrė Žliobaitė, Mykola Pechenizkiy, and Joao Gama. An overview of concept drift applications. In *Big Data Analysis: New Algorithms for a New Society*, pages 91–114. Springer, 2016.