Universiteit Leiden
The Netherlands

Data science for tax administration

Pijnenburg, M.G.F.

Citation
Pijnenburg, M. G. F. (2020, June 24). *Data science for tax administration*. Retrieved from https://hdl.handle.net/1887/123049

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/123049

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The handle http://hdl.handle.net/1887/123049 holds various files of this Leiden University dissertation.

Author: Pijnenburg, M.G.F.
Title: Data science for tax administration
Issue Date: 2020-06-24

Acknowledgements

I am grateful for the many great moments this PhD track has brought to me. It offered an opportunity to see the richness of the field and I have been able to meet many fascinating people. In particular I would like to thank the persons below.

To start, I would like to thank Annette for encouraging me to follow the direction of my heart. Without her support, I would not have dared to start this PhD track. Also, I would like to thank Gert Willem Haasnoot from The Curious Network and Joan van der Zee, coach at the Belastingdienst. Both have encouraged me at the initial stage when the PhD track was still just an idea. With hindsight, the decision to quit my position as a CFO advisor and to start working as a data scientist and part-time PhD candidate has been a good one.

Many thanks also to the people that made possible this track, in particular Nicole Back and Cees Bredius, directors at the NTCA, who put confidence in the track and me. I also appreciate dearly the possibilities the NTCA has offered to complete this piece of research. I am also grateful for the help that was given by prof. dr. mr. Lisette van der Hel RA and prof. dr. Bert Kersten, to initialize the PhD track.

All beginnings are difficult. In the first year, Herman Hoorweg has played a pivotal role. His curiosity to know what Analytics can actually contribute to a tax administration has been the spark for the first article. After that initial paper, spotting interesting topics has been much easier.

The greatest 'thank you' goes without any doubt to my co-promotor dr. Wojtek Kowalczyk. I want to thank him for being a great master in Data Science for me. I also want to thank him for the courage to start this PhD in the first place: a track that has been surrounded by many uncertainties initially. And last but not least, for being a wonderful person. I will definitely miss our talks.

Finally, I would like to thank dr. Frank Takes for providing the latex template for this dissertation.

Curriculum Vitae English

Mark Pijnenburg was born on May 7, 1979 in Heerlen. Mark completed grammar school at the Bernardinuscollege in 1997 (summa cum laude). In 2003 he obtained his master's degree in Mathematics (cum laude) from the Radboud University.

After this, Mark worked for two years for Eindhoven University of Technology in the program 'Mathematics for Industry' of the Stan Ackermans Institute. After a final project at DSM Research, the program was completed in 2005. In 2009, the Amsterdam MBA was completed at the University of Amsterdam (cum laude).

In 2006 Mark joined the Tax Authorities in Utrecht. Various positions followed with the Tax Authorities, including 'quantitative researcher', 'business analyst' and 'advisor to the CFO'. In 2014, the position started as a Data Scientist. And later that year, the parallel PhD track started at the Leiden Institute for Advanced Computer Science (LIACS) of Leiden University, with this thesis as the final result.

Curriculum Vitae Nederlands

Mark Pijnenburg is geboren op 7 mei 1979 te Heerlen. In 1997 rondde Mark het gymnasium af bij het Bernardinuscollege (summa cum laude). In 2003 behaalde hij aan de Radboud Universiteit zijn master Wiskunde (cum laude).

Hierna werkte Mark twee jaar voor de Technische Universiteit Eindhoven in het programma 'Mathematics for Industry' van het Stan Ackermans-instituut. Na een eindopdracht bij DSM Research werd het programma in 2005 afgerond. In 2009 werd de Amsterdam MBA afgerond bij de Universiteit van Amsterdam (cum laude).

In 2006 trad Mark in dienst bij de Belastingdienst te Utrecht. Bij de Belastingdienst volgden verschillende functies, waaronder kwantitatief onderzoeker, business analist en adviseur van de Chief Financial Officer. In 2014 startte de functie als Data Scientist. En later dat jaar begon het parallelle promotietraject bij het Leiden Institute for Advanced Computer Science (LIACS) van de Universiteit Leiden, met dit proefschrift als eindresultaat.

Bibliography

[1] N. G. C. F. M. Abreu et al. *Analise do perfil do cliente Recheio e desenvolvimento de um sistema promocional*. PhD thesis, ISCTE-IUL, 2011.

[2] L. Acebedo and J. Durnall. Risk-based pricing: When does it work and when does it not? an adverse selection approach. Technical report, Experian Decision Analytics, 5 2013.

[3] O. Akbilgic, H. Bozdogan, and M. E. Balaban. A novel hybrid rbf neural networks model as a forecaster. *Statistics and Computing*, 24(3):365–375, 2014.

[4] S. Akter and S. F. Wamba. Big data analytics in e-commerce: a systematic review and agenda for future research. *Electronic Markets*, 26(2):173–194, 2016.

[5] M. G. Allingham and A. Sandmo. Income tax evasion: A theoretical analysis. *Journal of public economics*, 1(3-4):323–338, 1972.

[6] J. Alm. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International tax and public finance*, 19(1):54–77, 2012.

[7] J. Andreoni, B. Erard, and J. Feinstein. Tax compliance. *Journal of economic literature*, 36(2):818–860, 1998.

[8] M. Atzmueller. Subgroup discovery. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(1):35–49, 2015.

[9] R. Bakeman and J. M. Gottman. *Observing interaction: An introduction to sequential analysis*. Cambridge university press, 1997.

[10] D. Bassi and C. Hernandez. Credit risk scoring: results of different network structures, preprocessing and self-organised clustering. In *Decision Technologies for Financial Engineering. Proceedings of the Fourth International Conference on Neural Networks in the Capital Markets*, pages 151–61, 1997.

[11] S. Basta, F. Fassetti, M. Guarascio, G. Manco, F. Giannotti, D. Pedreschi, L. Spinsanti, G. Papi, and S. Pisani. High quality true-positive prediction for fiscal fraud detection. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on Data Mining*, pages 7–12. IEEE, 2009.

[12] S. Basu and G. B. Waymire. Recordkeeping and human evolution. *Accounting Horizons*, 20(3):201–229, 2006.

[13] Overzicht verwerkingen van persoonsgegevens door de Belastingdienst. `https://download.belastingdienst.nl/belastingdienst/docs/verw_persoonsgegevens_belastingdienst_al5303z7fd.pdf`. Accessed: 2019-07-01.

[14] Belastingdienst. Claiming refunds for vat. `https://www.belastingdienst.nl/wps/wcm/connect/bldcontenten/belastingdienst/business/vat/vat_in_the_netherlands/claiming_refund_of_vat/claiming_refund_of_vat`, 2019.

[15] V. Bentkus, G. Geuze, M. Pijnenburg, and M. van Zuijlen. Unimodality: The symmetric case. Technical report, Radboud University Nijmegen, 2006.

[16] N. C. Berkman. Value grouping for binary decision trees. Technical report, University of Massachusetts, 1995.

[17] A. Bolt, W. van der Aalst, and M. de Leoni. Finding process variants in event logs (short paper). In *Confederated International Conference On the Move to Meaningful Internet Systems, OTM 2017 held in conjunction with Conferences on CoopIS, CandTC and ODBASE 2017*. Springer Netherlands, 2017.

[18] C. Boon, F. D. Belschak, D. N. Den Hartog, and M. Pijnenburg. Perceived human resource management practices. *Journal of Personnel Psychology*, 2014.

[19] R. J. C. Bose, W. M. Van Der Aalst, I. Zliobaite, and M. Pechenizkiy. Dealing with concept drifts in process mining. *IEEE transactions on neural networks and learning systems*, 25(1):154–171, 2014.

[20] V. Braithwaite. Responsive regulation and taxation: Introduction. *Law & Policy*, 29(1):3–10, 2007.

[21] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.

[22] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000.

[23] C. Brunsdon, A. Fotheringham, and M. Charlton. An investigation of methods for visualising highly multivariate datasets. *Case Studies of Visualization in the Social Sciences*, pages 55–80, 1998.

[24] D. Burshtein, V. Della Pietra, D. Kanevsky, and A. Nadas. Minimum impurity partitions. *The Annals of Statistics*, pages 1637–1646, 1992.

[25] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[26] P. A. Chou et al. Optimal partitioning for classification and regression trees. *IEEE Transactions on pattern analysis and machine intelligence*, 13(4):340–354, 1991.

[27] R. B. Cialdini and N. J. Goldstein. Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, 55:591–621, 2004.

[28] Council of European Union. Convention for the protection of human rights and fundamental freedoms. `https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/005`, 1953.

[29] D. Cramer. *Advanced quantitative data analysis*. McGraw-Hill Education (UK), 2003.

[30] L. S. da Silva, R. N. Carvalho, and J. C. F. Souza. Predictive models on tax refund claims-essays of data mining in brazilian tax administration. In *International Conference on Electronic Government and the Information Systems Perspective (EGOVIS)*, pages 220–228. Springer, 2015.

[31] T. Davenport. *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard Business Review Press, 2014.

[32] T. Davenport and J. Harris. Competing on analytics. *Boston, MA: Harvard Business School Publishing*, 2007.

[33] T. H. Davenport, J. G. Harris, and R. Morison. *Analytics at work: Smarter decisions, better results*. Harvard Business Press, 2010.

[34] E. Dubossarsky and Y. Tyshetskiy. R package autoencoder. `https://CRAN.R-project.org/package=autoencoder`, May 2014.

[35] European Commission. Building trust in human centric artificial intelligence. `https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence`, 4 2019.

[36] W. Federer. *Statistics and Society*. Marcel Dekker Inc., 2 edition, 1991.

[37] S. Feldstein. The road to digital unfreedom: How artificial intelligence is reshaping repression. *Journal of Democracy*, 30(1):40–52, 2019.

[38] Fiscalis Risk Management Platform Group. Compliance risk management guide for tax administrations. `http://ec.europa.eu/taxation_customs/resources/documents/Ecommon/publications/info_docs/taxation/risk_managt_guide_en.pdf`, 5 2010.

[39] A. Fischer and C. Igel. An introduction to restricted boltzmann machines. In *iberoamerican congress on pattern recognition*, pages 14–36. Springer, 2012.

[40] Forum on Tax Administration. Advanced analytics for better tax administration: Putting data to work. `http://dx.doi.org/10.1787/9789264256453-en`, 2016.

[41] Forum on Tax Administration Compliance Sub-group Compliance Risk Management. Managing and improving tax compliance, guidance note. `http://www.oecd.org/tax/administration/33818656.pdf`, 2004.

[42] J. Friedman, T. Hastie, and R. Tibshirani. The elements of statistical learning: Data mining, inference, and prediction. *Springer Series in Statistics (*, 2009.

[43] M. Goldstein and S. Uchida. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, 11(4):e0152173, 2016.

[44] G. Gupta. *Introduction to data mining with case studies*. PHI Learning Pvt. Ltd., 2014.

[45] M. Gupta and V. Nagadevara. Audit selection strategy for improving tax compliance–application of data mining techniques. In *Foundations of*

Risk-Based Audits. Proceedings of the eleventh International Conference on e-Governance, Hyderabad, India, December, pages 28–30, 2007.

[46] H. Haddadi, R. Mortier, and S. Hand. Privacy analytics. *ACM SIGCOMM Computer Communication Review*, 42(2):94–98, 2012.

[47] A. Hald. Statistical theory with engineering applications. In *Statistical theory with engineering applications*. John Wiley & Sons, 1952.

[48] S. Hariri, M. C. Kind, and R. J. Brunner. Extended isolation forest. *arXiv preprint arXiv:1811.02141*, 2018.

[49] K. Hassibi et al. Detecting payment card fraud with neural networks. *World Scientific Book Chapters*, pages 141–157, 2000.

[50] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics*. Springer New York, 2 edition, 2009.

[51] G. E. Hinton. A practical guide to training restricted boltzmann machines. In *Neural networks: Tricks of the trade*, pages 599–619. Springer, 2012.

[52] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[53] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[54] K. Hsu, N. Pathak, J. Srivastava, G. Tschida, and E. Bjorklund. Data mining based tax audit selection: A case study from minnesota department of revenue. In *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2009.

[55] B. R. Jackson and V. C. Milliron. Tax compliance research: Findings, problems, and prospects. *Journal of accounting literature*, 5(1):125–165, 1986.

[56] G. K. Kanji. *100 Statistical Tests*. Sage Publications, 3 edition, 2006.

[57] G. V. Kass. An exploratory technique for investigating large quantities of categorical data. *Applied statistics*, pages 119–127, 1980.

[58] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, Aug. 2009.

[59] A. Kumar, A. Saxena, V. Suvarna, and V. Rawat. Top 10 trends in banking in 2016. `https://www.capgemini.com/wp-content/uploads/2017/07/banking_top_10_trends_2016.pdf`, 3 2016.

[60] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim. A survey of deep learning-based network anomaly detection. *Cluster Computing*, pages 1–13, 2017.

[61] D. Larose. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, 2005.

[62] K. C. Laudon, C. G. Traver, et al. *E-commerce: business, technology, society*. Pearson, 2016.

[63] J. Leskovec, A. Rajaraman, and J. D. Ullman. *Mining of massive datasets*. Cambridge university press, 2 edition, 2014.

[64] A. Liaw and M. Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002.

[65] M. Lichman. UCI machine learning repository. `http://archive.ics.uci.edu/ml`, 2013.

[66] R. Lieber. Lower your car insurance bill, at the price of some privacy. `www.nytimes.com/2014/08/16/your-money/auto-insurance/tracking-gadgets-could-lower-your-car-insurance-at-the-price-of-some-privac` `html`, 2014.

[67] G. S. Linoff and M. J. Berry. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons, 3 edition, 2011.

[68] Z. C. Lipton, J. Berkowitz, and C. Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.

[69] B. Liu. *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.

[70] F. T. Liu. R package isolationforest. `https://rdrr.io/rforge/IsolationForest/`, August 2009.

[71] F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE, 2008.

[72] A. Maaradji, M. Dumas, M. La Rosa, and A. Ostovar. Fast and accurate business process drift detection. In *International Conference on Business Process Management*, pages 406–422. Springer, 2016.

[73] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceedings*, page 89. Presses universitaires de Louvain, 2015.

[74] K. Mardia, J. Kent, and J. Bibby. Multivariate statistics, 1979.

[75] M. A. Marin, M. Hauder, and F. Matthes. Case management: an evaluation of existing approaches for knowledge-intensive processes. In *International Conference on Business Process Management*, pages 5–16. Springer, 2015.

[76] J. Meza, L. Terán, A. Piaún, and M. Tomalá. A fuzzy-based recommender system for public tax payment. In *2018 International Conference on eDemocracy & eGovernment (ICEDEG)*, pages 235–240. IEEE, 2018.

[77] D. Micci-Barreca. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *ACM SIGKDD Explorations Newsletter*, 3(1):27–32, 2001.

[78] K. P. Murphy. *Machine Learning A Probabilistic Perspective*. The MIT Press, 1 edition, 2012.

[79] Netherlands Tax and Customs Administration. `download.belastingdienst.nl/belastingdienst/docs/dutch_tax_customs_admin.pdf`, May 2019.

[80] A. Ng. Lecture notes on sparse autoencoders. `https://web.stanford.edu/class/cs294a/sparseAutoencoder-2011.pdf`, 2011.

[81] Online etymology dictionary, statistics. `https://www.etymonline.com/word/statistics`.

[82] A. Ostovar, S. J. Leemans, and M. La Rosa. Robust drift characterization from event streams of business processes. *Internal Report*, 2018.

[83] S. Pauwels and T. Calders. Detecting and explaining drifts in yearly grant applications. *arXiv preprint arXiv:1809.05650*, 2018.

[84] N. Petit. Artificial intelligence and automated law enforcement: A review paper. *Available at SSRN 3145133*, 2018.

[85] M. Pijnenburg. Simulation of rubber networks. Technical report, Stan Ackermans instituut, 2005. ISBN: 9044404784.

[86] M. Pijnenburg. Code used in experiments. `https://github.com/PijnenburgMark/anomaly_detection_benchmark`, 2019. Accessed: 2019-06-01.

[87] M. Pijnenburg, N. Kalosha, and M. C. van Zuijlen. Hoeffding-Bentkus bound in statistical auditing. Technical report, Radboud University Nijmegen, 2006.

[88] M. Pijnenburg and W. Kowalczyk. Applying analytics for improved taxpayer supervision. In *Proceedings of 16th European Conference on e-Government ECEG 2016*, pages 145–153. Academic Conferences and publishing limited, 2016.

[89] M. Pijnenburg and W. Kowalczyk. Extending logistic regression models with factorization machines. In *International Symposium on Methodologies for Intelligent Systems*, pages 323–332. Springer, 2017.

[90] M. Pijnenburg and W. Kowalczyk. Singular outliers: Finding common observations with an uncommon feature. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 492–503. Springer, 2018.

[91] M. Pijnenburg and W. Kowalczyk. Are similar cases treated similarly? a comparison between process workers. In *International Conference on Business Information Systems*, pages 1–15. Springer, 2019.

[92] M. Pijnenburg and W. Kowalczyk. Extending an anomaly detection benchmark with auto-encoders, isolation forests, and rbms. In *International Conference on Information and Software Technologies*. Springer, 2019. Best Paper Award.

[93] M. Pijnenburg, W. Kowalczyk, E. van der Hel-van Dijk, et al. A roadmap for analytics in taxpayer supervision. *Electronic Journal of e-Government*, 15:19–32, 2017.

[94] M. Pijnenburg and K. Kuijpers. Explaining risk models to the business. *Tax Tribune*, 35:57–62, 2016.

[95] M. Post. Tax data and reinforcement learning. Master's thesis, Leiden University, 2 2019. Under supervision of Mark Pijnenburg, Wojtek Kowalczyk, and Kaifeng Yang.

[96] S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. In *ACM Sigmod Record*, volume 29, pages 427–438. ACM, 2000.

[97] S. Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.

[98] S. Rendle. Factorization machines with libFM. *ACM Trans. Intell. Syst. Technol.*, 3(3):57:1–57:22, May 2012.

[99] M. Richardson and A. J. Sawyer. A taxonomy of the tax compliance literature: further findings, problems and prospects. *Austl. Tax F.*, 16:137–320, 2001.

[100] X. Rong. R package deepnet. `https://CRAN.R-project.org/package=deepnet`, March 2014.

[101] X. Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

[102] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Computer Vision and Image Understanding*, 172:88–97, 2018.

[103] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, pages 146–157. Springer, 2017.

[104] H. Shane. Insurance cheats discover social media is the real pain in the neck. `www.theguardian.com/money/2016/jul/18/insurance-cheats-social-media-whiplash-false-claimants`, 2016.

[105] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, 1 edition, 11 2017. complete draft.

[106] United nations of roma victrix (unrv), roman taxes. `https://www.unrv.com/economy/roman-taxes.php`.

[107] L. van der Maaten, E. Postma, and H. van den Jerik. Dimensionality reduction: A comparative review. Technical report, Maastricht University, 2009.

[108] F. van Dongen, B.F.; Borchert. Bpi challenge 2018. `https://doi.org/10.4121/uuid:3301445f-95e8-4ff0-98a4-901f1f204972`, 2018.

[109] S. F. Wamba and S. Akter. Big data analytics for supply chain management: A literature review and research agenda. In *Workshop on Enterprise and Organizational Modeling and Simulation*, pages 61–72. Springer, 2015.

[110] E. Wiebes. Investeringsagenda belastingdienst. `https://www.tweedekamer.nl/kamerstukken/detail?id=2015Z09033&did=2015D18368`, 2015.

[111] Wikipedia, English edition, history of statistics. `https://en.wikipedia.org/wiki/History_of_statistics#Etymology`.

[112] Wikipedia, English edition, trial of the pyx. `https://en.wikipedia.org/wiki/Trial_of_the_Pyx`.

[113] R.-S. Wu, C.-S. Ou, H.-y. Lin, S.-I. Chang, and D. C. Yen. Using data mining technique to enhance tax evasion detection performance. *Expert Systems with Applications*, 39(10):8769–8777, 2012.

[114] J. Xu, M. Saebi, B. Ribeiro, L. M. Kaplan, and N. V. Chawla. Detecting anomalies in sequential data with higher-order networks. *arXiv preprint arXiv:1712.09658*, 2017.

[115] D. Yates, D. Moore, and G. McCabe. *The Practice of Statistics*. W.H. Freeman, New York, 1 edition, 1999.

[116] M. J. Zaki. Spade: An efficient algorithm for mining frequent sequences. *Machine learning*, 42(1-2):31–60, 2001.

[117] M. Zikeba, S. K. Tomczak, and J. M. Tomczak. Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert Systems with Applications*, 2016.

Publication List Author

Below follows a chronological list of publications by the author.

1. M. Pijnenburg. Simulation of rubber networks. Technical report, Stan Ackermans instituut, 2005. ISBN: 9044404784

2. V. Bentkus, G. Geuze, M. Pijnenburg, and M. van Zuijlen. Unimodality: The symmetric case. Technical report, Radboud University Nijmegen, 2006

3. M. Pijnenburg, N. Kalosha, and M. C. van Zuijlen. Hoeffding-Bentkus bound in statistical auditing. Technical report, Radboud University Nijmegen, 2006

4. C. Boon, F. D. Belschak, D. N. Den Hartog, and M. Pijnenburg. Perceived human resource management practices. *Journal of Personnel Psychology*, 2014

5. M. Pijnenburg and W. Kowalczyk. Applying analytics for improved taxpayer supervision. In *Proceedings of 16th European Conference on e-Government ECEG 2016*, pages 145–153. Academic Conferences and publishing limited, 2016

6. M. Pijnenburg and K. Kuijpers. Explaining risk models to the business. *Tax Tribune*, 35:57–62, 2016

7. M. Pijnenburg, W. Kowalczyk, E. van der Hel-van Dijk, et al. A roadmap for analytics in taxpayer supervision. *Electronic Journal of e-Government*, 15:19–32, 2017

8. M. Pijnenburg and W. Kowalczyk. Extending logistic regression models with factorization machines. In *International Symposium on Methodologies for Intelligent Systems*, pages 323–332. Springer, 2017

9. M. Pijnenburg and W. Kowalczyk. Singular outliers: Finding common observations with an uncommon feature. In *International Conference on Information*

Processing and Management of Uncertainty in Knowledge-Based Systems, pages 492–503. Springer, 2018

10. M. Pijnenburg and W. Kowalczyk. Are similar cases treated similarly? a comparison between process workers. In *International Conference on Business Information Systems*, pages 1–15. Springer, 2019

11. M. Pijnenburg and W. Kowalczyk. Extending an anomaly detection benchmark with auto-encoders, isolation forests, and rbms. In *International Conference on Information and Software Technologies*. Springer, 2019. Best Paper Award