

Curriculum Vitae – Wouter Duivesteijn



Date of birth December 9, 1984
Place of birth Rotterdam, the Netherlands
Nationality Dutch
Email address w.duivesteijn@tue.nl
Website <http://wwwis.win.tue.nl/~wouter/>

Academic career

September 2016 Assistant Professor Data Mining at the Technische Universiteit Eindhoven (the Netherlands).
now

October 2015 Postdoctoraal Bursaal at the Universiteit Gent (Belgium), working on the FORSIED project:
August 2016 FORmalising Subjective Interestingness in Exploratory Data mining.

August 2015 See “Research Visits” section.

September 2015

August 2013 Wissenschaftlicher Mitarbeiter at the Technische Universität Dortmund (Germany),
July 2015 in Sonderforschungsbereich 876: Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung (Collaborative Research Center SFB 876: Providing Information by Resource-Constrained Data Analysis).

July 2009 PhD candidate in the Algorithms group at the Leiden Institute of Advanced Computer
July 2013 Science, Leiden University (the Netherlands), under the supervision of Joost N. Kok and Arno Knobbe (graduation date: September 17, 2013).

Education

2009-2013 PhD in Computer Science, Leiden University (graduated on September 17, 2013).
2008-2009 MSc in History and Philosophy of Science, Utrecht University (aborted when I obtained a PhD position).
2005-2008 MSc in Applied Computing Science, Utrecht University (graduated on October 15, 2008).
2005-2007 MSc in Mathematical Sciences, Utrecht University (graduated on October 30, 2007).
2002-2005 BSc in Mathematics, Utrecht University (graduated on July 7, 2005).
2002-2005 BSc in Computer Science, Utrecht University (graduated on July 7, 2005).
1996-2002 Gymnasium, C.S.G. Blaise Pascal, Spijkenisse.

Awards and funding

C.J. Kokprijs 2013: the award for the best dissertation of the year in Leiden University’s Faculty of Science. The award has been bestowed annually since 1971, and I am the first ever computer scientist to win it.

KDD 2012 student travel award: USD 1,000 plus free registration (worth USD 380). **ICDM 2011 student travel award:** CAD 500 (where CAD 1 = USD 0.96302). **ICDM 2010 student travel award:** USD 500.

Publications

Journal publications

- * L. Downar, W. Duivesteijn: Exceptionally Monotone Models — the Rank Correlation Model Class for Exceptional Model Mining. In: *Knowledge and Information Systems* 51 (2), pp. 369–394, 2017.
Impact factor: 2.004 (2016).
- * C. Pölitiz, W. Duivesteijn, K. Morik: Interpretable Domain Adaptation via Optimization over the Stiefel Manifold. In: *Machine Learning* 104 (2–3), pp. 315–336, 2016. Impact factor: 1.848.
- * W. Duivesteijn: Correction to Jin-Ting Zhang’s “Approximate and Asymptotic Distributions of Chi-Squared-Type Mixtures with Applications”. In: *Journal of the American Statistical Association* 111 (515), pp. 1370–1371, 2016. Impact factor: 2.016.
- * W. Duivesteijn, A.J. Feelders, A. Knobbe: Exceptional Model Mining — Supervised Descriptive Local Pattern Mining with Complex Target Concepts. In: *Data Mining and Knowledge Discovery* 30 (1), pp. 47–98, 2016. Impact factor: 3.160.
- * R.M. Konijn, W. Duivesteijn, M. Meeng, A. Knobbe: Cost-based Quality Measures in Subgroup Discovery. In: *Journal of Intelligent Information Systems* 45 (3), pp. 337–355, 2015. Impact factor: 1.000.
- * P. Lohuis, S. Faraj-Hakim, W. Duivesteijn, A. Knobbe, A.-J. Tasman: Benefits of a Short, Practical Questionnaire to Measure Subjective Perception of Nasal Appearance after Aesthetic Rhinoplasty. In: *Plastic and Reconstructive Surgery* 132 (6), pp. 913e–923e, 2013. Impact factor: 3.328.
- * P.J.F.M. Lohuis, S. Hakim, A. Knobbe, W. Duivesteijn, G.M. Bran: Split hump technique for reduction of the overprojected nasal dorsum – a statistical analysis on subjective body image in relation to nasal appearance and nasal patency in 97 aesthetic rhinoplasty patients. In: *Archives of Facial Plastic Surgery*, 2012, 14 (5), pp. 346–353. Impact factor: 1.463.
- * S. Hakim, A. Knobbe, W. Duivesteijn, P.J.F.M. Lohuis: Results of a screening questionnaire measuring physical perception of patients undergoing esthetic rhinoplasty: a statistical analysis. In: *Nederlands Tijdschrift voor Keel-Neus-Oorheelkunde* (Dutch Journal for Otorhinolaryngology), 2010 (2), p. 100. Impact factor: 0.

Conference publications

- * W. Duivesteijn, T. Farzami, T. Putman, E. Peer, H.J.P. Weerts, J.N. Adegeest, G. Foks, M. Pechenizkiy: Have It Both Ways — from A/B Testing to A&B Testing with Exceptional Model Mining. In: Proceedings of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2017)*, to appear.
Acceptance rate: 0.2903 (27 out of 93).
- * C. Rebelo de Sá, W. Duivesteijn, C. Soares, A. Knobbe: Exceptional Preferences Mining. In: Proceedings of the *19th International Conference on Discovery Science (DS 2016)*, pp. 3–18.
Acceptance rate: 0.5000 (30 out of 60).
- * L. Downar, W. Duivesteijn: Exceptionally Monotone Models — the Rank Correlation Model Class for Exceptional Model Mining. In: Proceedings of the *15th IEEE International Conference on Data Mining (ICDM 2015)*, pp. 111–120, 2015.
Acceptance rate: 0.0843 (68 out of 807). Including short papers: 0.1821 (147 out of 807).

Distilled from Lennart's Bachelor thesis, available at:

http://sfb876.tu-dortmund.de/PublicPublicationFiles/downar_2014a.pdf

- * W. Duivesteijn, J. Thaele: Understanding Where Your Classifier Does (Not) Work. In: Proceedings of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2015)* (III), pp. 250–253, 2015.
Acceptance rate: 0.4828 (14 out of 29).
- * W. Duivesteijn, J. Thaele: Understanding Where Your Classifier Does (Not) Work — the SCaPE Model Class for EMM. In: Proceedings of the *14th IEEE International Conference on Data Mining (ICDM 2014)*, pp. 809–814, 2014.
Acceptance rate: 0.1953 (142 out of 727).
A substantially longer version appeared as Technical Report of the SFB 876 at the TU Dortmund: http://sfb876.tu-dortmund.de/PublicPublicationFiles/duivesteijn_thaele_2014a.pdf
- * J. Witteveen, W. Duivesteijn, A. Knobbe, P. Grünwald: RealKRIMP – Finding Hyperintervals that Compress with MDL for Real-Valued Data. In: Proceedings of the *13th International Symposium on Intelligent Data Analysis (IDA 2014)*, pp. 368–379, 2014. Acceptance rate: 0.4800 (36 out of 75).
Distilled from Jouke's Bachelor thesis, available at:
<http://liacs.leidenuniv.nl/assets/Bachelorscripties/2012-13JoukeWitteveen.pdf>
- * M. Meeng, W. Duivesteijn, A. Knobbe: ROCsearch – An ROC-guided Search Strategy for Subgroup Discovery. In: Proceedings of the *2014 SIAM International Conference on Data Mining (SDM 2014)*, pp. 704–712, 2014.
Acceptance rate: 0.3084 (120 out of 389).
- * W. Duivesteijn, A. Knobbe: Exceptional Model Mining – Describing Deviations in Datasets. In: Proceedings of the *22nd Belgian-Dutch Conference on Machine Learning (BENELEARN 2013)*, p. 86, 2013.
Acceptance rate: 0.8919 (33 out of 37).
- * R.M. Konijn, W. Duivesteijn, W. Kowalczyk, A. Knobbe: Discovering Local Subgroups, with an Application to Fraud Detection. In: Proceedings of the *17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2013)*, pp. 1–12, 2013.
Acceptance rate: 0.1134 (39 out of 344). Including short presentations: 0.2849 (98 out of 344).
- * W. Duivesteijn, E. Loza Mencía, J. Fürnkranz, A. Knobbe: Multi-label LeGo – Enhancing Multi-label Classifiers with Local Patterns. In: Proceedings of the *11th International Symposium on Intelligent Data Analysis (IDA 2012)*, pp. 114–125, 2012.
Acceptance rate: 0.2250 (18 out of 80). Including poster presentations: 0.4375 (35 out of 80).
A substantially longer version appeared as Technical Report TUD-KE-2012-02 of the TU Darmstadt: <http://www.ke.tu-darmstadt.de/publications/reports/tud-ke-2012-02.pdf>
- * G. Ribeiro, W. Duivesteijn, C. Soares, A. Knobbe: Multilayer Perceptron for Label Ranking. In: Proceedings of the *22nd International Conference on Artificial Neural Networks (ICANN 2012)*, pp. 25–32, 2012.
Acceptance rate: 0.6559 (162 out of 247).
- * W. Duivesteijn, A. Feelders, A. Knobbe: Different Slopes for Different Folks – Mining for Exceptional Regression Models with Cook's Distance. In: Proceedings of the *18th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2012)*, pp. 868–876, 2012.
Acceptance rate: 0.1762 (133 out of 755).
- * W. Duivesteijn, A. Knobbe: Exploiting False Discoveries – Statistical Validation of Patterns and Quality Measures in Subgroup Discovery. In: Proceedings of the *11th IEEE International Conference on Data Mining (ICDM 2011)*, pp. 151–160, 2011.
Acceptance rate: 0.1285 (101 out of 786). Including short papers: 0.1883 (148 out of 786).

- * W. Duivesteijn, A. Knobbe, A. Feelders, M. van Leeuwen: Subgroup Discovery meets Bayesian networks – an Exceptional Model Mining approach. In: Proceedings of the 10th *IEEE International Conference on Data Mining (ICDM 2010)*, pp. 158–167, 2010.
Acceptance rate: 0.0903 (72 out of 797). Including short papers: 0.1945 (155 out of 797).
- * W. Duivesteijn, A. Feelders: Nearest Neighbour Classification with Monotonicity Constraints. In: Proceedings of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2008) (I)*, pp. 301–316, 2008.
Acceptance rate: 0.1919 (100 out of 521).

Workshop publications

- * S. van der Zon, O. Zeev Ben Mordehai, T. Vrijdag, W. van Ipenburg, J. Veldsink, W. Duivesteijn, M. Pechenizkiy: BoostEMM — Transparent Boosting using Exceptional Model Mining. In: Proceedings of the *Second Workshop on Mining Data for financial applications (MIDAS 2017)*, pp. 5–14, 2017.
- * W. Duivesteijn, M. Meeng, A. Knobbe: ROCsearch in a Wider Context – A ROC-Guided Search Strategy for Subgroup Discovery and Beyond. In: Proceedings of the *First International Workshop on Learning over Multiple Contexts (LMCE 2014)*, 2014.
- * W. Duivesteijn: A Short Survey of Exceptional Model Mining – Exploring Unusual Interactions Between Multiple Targets. In: Proceedings of the *2014 International Workshop on Multi-Target Prediction (MTP 2014)*, 2014.
- * M. Meeng, W. Duivesteijn, A. Knobbe: ROCsearch – An ROC-Guided Search Strategy for Subgroup Discovery. In: Proceedings of the *2014 Workshop on Knowledge Discovery, Data Mining and Machine Learning (KDML 2014)*, p. 180, 2014.
- * R.M. Konijn, W. Duivesteijn, M. Meeng, A. Knobbe: Cost-based Quality Measures in Subgroup Discovery. In: Proceedings of the *3rd Quality Issues, Measures of Interestingness, and Evaluation of data mining models workshop (QIMIE 2013)*, PAKDD Workshops, pp. 404–415, 2013.

Invited talk

- Oct. 2016 invited talk “Some Recent Advances in Exceptional Model Mining - Unusual Preferences and more” at the PEPS Préfute workshop, IRISA INRIA Rennes, France.

Research visits

- Aug.-Sep. 2015** For two months, visited the Department of Engineering Mathematics at the University of Bristol, UK. Collaborated with Prof. Tijn De Bie and his group on the FORSIED project: FORmalizing Subjective Interestingness in Exploratory Data mining.
- Oct. 2012** For one day, visited the Varieng Institute at the University of Helsinki, Finland. Collaborated with Prof. Terttu Nevalainen, Tanja Säily, and Mikko Hakala on subgroup discovery and data cleaning in the Parsed Corpus of Early English Correspondence.
- Feb. 2012** For two days, visited the Integrated Data Mining group at the Fraunhofer IAIS, Bonn, Germany. Exchanged ideas with Dr. Stefan Rüping, Dr. Henrik Großkreutz, and Dr. Mario Boley; collaborated with the latter and Sandy Moens on integrating coupling-from-the-past-sampling with subgroup discovery.
- Nov.-Dec. 2011** For two weeks, visited the Department of Criminology at Simon Fraser University, Burnaby, Canada. Collaborated with Dr. Richard Frank on improving a webcrawler built to detect child pornography.
- Oct.-Nov. 2010** For two weeks, visited the Knowledge Engineering group at the Technische Universität Darmstadt, Germany. Collaborated with Eneldo Loza Mencía and Prof. Johannes Fürnkranz, on integrating multi-label classification with exceptional model mining having deviating Bayesian network structure as target.

Teaching

Courses academic year 17/18:

- JM0150** Data Mining; 6 ECTS course in the Data Science and Entrepreneurship master at JADS.
- 2IID0** Web Analytics; 5 ECTS course in the Computer Science bachelors at TU/e.
- 2IAB0** Data Analytics for Engineers; 5 ECTS course in the Bachelor College, compulsory for all freshman Bachelor students TU/e-wide.

Courses academic year 16/17:

- JM0150** Data Mining; 6 ECTS course in the Data Science and Entrepreneurship master at JADS.
- PDEDMM** Data Mining Module; course in the Data Science PDEng at JADS.
- 2IID0** Web Analytics; 5 ECTS course in the Computer Science bachelors at TU/e.

Guest lecture *Exceptional Model Mining — Understanding Unusual Interactions Between Multiple Targets*; part of the master course “Advanced Data Mining”, Universiteit Antwerpen, March 8, 2016.

Supervised PhD students:

- Feb. 2017–now** Du Xin (Spatial Subgroup Discovery)
- Feb. 2017–now** Oren Zeev-Ben-Mordehai (Know Your Customer; interpretable models)
- Feb. 2017–now** Simon van der Zon (Know Your Customer; visualization & local pattern mining)

Co-supervised theses Boy Raaijmakers (TU Eindhoven, BSc, 2016/7, ongoing), Ahmet Celikkaya (TU Dortmund, MSc, 2015), Lennart Downar (TU Dortmund, BSc, 2014), and Jouke Witteveen (Universiteit Leiden, BSc, 2012).

Teaching assistant/tutor for courses:

Spring 2016	Waarschijnlijkheidsrekening en Statistiek.	Fall 2009	Concepten van Programmeertalen.
Summer 2015	Wissensentdeckung in Datenbanken.	Fall 2008	Wiskundige Technieken voor Fysici.
Winter 2014/5	Software-Praktikum.	Spring 2008	Wiskunde voor Chemici.
Summer 2014	Software-Praktikum.	Fall 2007	Wiskunde voor NW&I, Wiskunde voor Chemici.
Winter 2013/4	Software-Praktikum.	Fall 2006	Lineaire Algebra, Wiskundetoepassingen.
Fall 2012	Fundamentele Informatica voor I&E.	Fall 2005	Lineaire Algebra.
Spring 2012	Fundamentele Informatica 3.	Fall 2004	Lineaire Algebra, Wiskunde voor Informatici, Logica en Verzamelingen.
Fall 2011	Fundamentele Informatica 2.	Spring 2004	Wiskunde voor Biologen.
Spring 2011	Fundamentele Informatica 3.		
Fall 2010	Fundamentele Informatica 2.		
Spring 2010	Algoritmiek.		

Service

Conference chair of Benelearn 2017 in Eindhoven (together with Mykola Pechenizkiy).

Workshop chair of Silver 2012, collocated with ECML PKDD in Bristol, UK (together with Joaquin Vanschoren). This workshop was dedicated to learning from unexpected results, and disseminating the resulting knowledge.

Program committee member of ECML PKDD 2015–2017, ICDM 2014–2017, IDA 2014–2017, IDEA@KDD 2015–2017, IJCAI 2016, ISMIS 2017, KDD 2015–2016, KDML 2014–2015, LMCE 2015.

Reviewer for DAMI, IJBDI, INS, JMLR, KAIS, MACH, NCAA, NEUCOM, TKDD, TKDE.

Secondary reviewer for KDD 2014, ECML PKDD 2013 (journal track), PGM 2012, CIKM 2012, JMLR, ICDM 2011, MPS 2010, ECML PKDD 2010, SDM 2010.

Extracurricular activities

- 2007-2008 Chairman of the board of students' association A-Eskwadraat, for Mathematics, Physics, Computer Science, and Information Science, with more than 1700 members.
- 2006-2007 Member of the "faculteitsraad Bètawetenschappen" (science faculty council) and the "departementsraad Wiskunde" (mathematics department council) at Utrecht University.
- 2005-2006 Member of the Mathematics department's board of education at Utrecht University.
- 2004-2005 Member of the "opleidingscommissie Wiskunde" (mathematics education council) at Utrecht University.

Languages

Native in Dutch, fluent in English, good in German, can read French, may contain traces of Latin and Greek.

Computer skills

Absolutely fluent in \LaTeX . Familiar with Windows, Linux, Java, Haskell, Prolog, Matlab, R, Excel. Programmed parts of Cortana, see <http://datamining.liacs.nl/cortana.html>.