

Prof. Szymon Jaroszewicz, Ph.D. (habil.)

CURRICULUM VITAE

Year of birth 1974
E-mail s.jaroszewicz@ipipan.waw.pl

Education and Employment

2020 The title of full professor
2020 – 2023 Secretary of the Committee on Informatics of the Polish Academy of Sciences
2019 – Head of the Statistical Analysis and Modeling Group at the Institute of Computer Science, Polish Academy of Sciences
2010 – Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland, professor (profesor nadzw.)
2010 Habilitation (D.Sc.), Institute of Computer Science, Polish Academy of Sciences
2006 – 2018 National Institute of Telecommunications, Warsaw, Poland, professor (profesor nadzw.)
2004 – 2009 Szczecin University of Technology, Szczecin, Poland, assistant professor (adiunkt)
2003 Ph.D. degree, University of Massachusetts Boston. Dissertation title: “Information Theoretical and Combinatorial Methods in Data Mining”
1999 – 2003 University of Massachusetts Boston, Ph.D. Program in Computer Science.
1998 M.Sc., Eng. degree in Computer Science, Institute of Computer Science, Technical University of Szczecin, Poland
1993 – 1998 Technical University of Szczecin, Poland, Computer Science M.Sc., Eng. Program

Scholarships, Awards and Honors

2017 2007 Best paper award at the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)
2005/2006 Grant for Young Researchers from The Foundation for Polish Science
2000 The Randall G. Malbone Fellowship In Mathematics and Computer Science, Dept. of Mathematics and Computer Science, University of Massachusetts Boston
1998 Fulbright Scholarship for the academic year 1998/99 at The University of Massachusetts Boston

1997, 1996 Scholarship of The Polish Minister of Education for exceptional scientific achievements for the academic years 1996/97 and 1997/98

Grants

Leader:

2011 „Uplift Modeling for marketing and medical research” Polish National Science Centre grant no. NN516 414938
2010 „PaCal – a library for arithmetic computations with random variables” Polish National Science Centre grant no. NN516 068537
2002 Dissertation Support Grant from University of Massachusetts Boston

Participation:

2019 „An integrated system for retail price statistics” grant in the „Gospostrateg” program of the National Centre for Research and Development
2012 „Recommendations of virtual teams for complex tasks requiring open collaboration”, Polish National Science Centre grant no. 2012/05/B/ST6/03364
2010 „Future Internet Engineering”, European Regional Development Fund Grant np. POIG 01.01.02-00-045/09-00

Professional Activities

- Editorial Board Member of Data Mining and Knowledge Discovery (from 2009), Machine Learning and Fundamenta Informaticae
- Secretary of the Committee on Informatics of the Polish Academy of Sciences, member of the Presidium of the Committee (term 2020–2023)
- Member of the expert team for project assessment for the „Diamond Grant” programme at the Polish Ministry of Science and Higher Education (10.02.2012–31.12.2014)
- Member of the interdisciplinary team for project assessment for the „Iuventus Plus” Programme at the Polish Ministry of Science and Higher Education (21.11.2011–31.12.2014)
- Program committee member for major data mining conferences: KDD '14 '17 '18 '19 '20 '21, IJCAI '15 '18 '19 '20, '21 (senior PC) ICDM '08 '10 '11 '12 '13 '16, ECML/PKDD '06 '07 '08 '09 '11 '13 '14 '15 '16 '17 '18 '19 '20, ECML/PKDD Journal Track '19 '20 '21, CIKM '08 '09 '10 '17

'18 '20, SIAM DM '17 '18 '19 '20 '21, PAKDD '06 '07 '08 '09 '11 '12 '19, ISMIS '17, ACM SAC (Data Mining Track) '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19 '20 '21, KDIR '09 '10 '11 '12 '13 '14

PAKDD'06, Discovery Science'06, PKDD'06, PAKDD'07, ACM SAC'07 Data Mining Track, PKDD'07, PAKDD'08, ACM SAC'08 Data Mining Track, PKDD'08 (area chair), CIKM'08, ICDM'08, PAKDD'09, ACM SAC'09 Data Mining Track, KDIR'09, ACM CSE'09, CIKM'09, Local Patterns to Global Models Workshop'09

- Reviewer for journals: Data Mining and Knowledge Discovery, ACM Transactions on Database Systems, IEEE Transactions on Knowledge and Data Engineering, Data and Knowledge Engineering, IEEE Transactions on Information Theory, Transactions on Rough Sets, Knowledge and Information Systems, IEEE Transactions on Neural Networks and Learning Systems and others

Data Mining Related Research Cooperation

- Data-mining of continuous attributes, joint work with Bart Goethals and Toon Calders, University of Antwerp, Belgium
- Sampling algorithms in data-mining, joint work with Prof. Tobias Scheffer from Humboldt University, Berlin

Publications

- [1] B. Żogała-Siudem and S. Jaroszewicz. Fast stepwise regression based on multidimensional indexes. *Information Sciences*, 549:288–309, 2021.
- [2] R. M. Gubela, S. Lessmann, and S. Jaroszewicz. Response transformation and profit decomposition for revenue uplift modeling. *European Journal of Operational Research*, 283(2):647–661, 2020.
- [3] Krzysztof Rudaś and Szymon Jaroszewicz. Shrinkage estimators for uplift regression. In *Proc. of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'19)*, Wrzburg, Germany, Sep 2019.
- [4] Krzysztof Rudaś and Szymon Jaroszewicz. Linear regression for uplift modeling. *Data Mining and Knowledge Discovery*, 32(5):1275–1305, Sep 2018.
- [5] Oskar Jarczyk, Szymon Jaroszewicz, Adam Wierzbicki, Kamil Pawlak, and Michal Jankowski-Lorek. Surgical teams on github: Modeling performance of github project development processes. *Information and Software Technology*, 100:32–46, 2018.
- [6] M. Sołtys and S. Jaroszewicz. Boosting algorithms for uplift modeling. *CoRR*, abs/1807.07909, 2018. arXiv preprint.

- [7] L. Zaniewicz and S. Jaroszewicz. Lp-support vector machines for uplift modeling. *Knowledge and Information Systems*, 53(1):269–296, Oct 2017.
- [8] S. Jaroszewicz. Uplift modeling. In C. Sammut and G. Webb, editors, *Encyclopedia of Machine Learning and Data Mining*, pages 1304–1309. Springer US, Boston, MA, 2017.
- [9] M. Jankowski-Lorek, S. Jaroszewicz, L. Ostrowski, and A. Wierzbicki. Verifying social network models of wikipedia knowledge community. *Information Sciences*, 339:158–174, 2016.
- [10] S. Jaroszewicz and L. Zaniewicz. Székely regularization for uplift modeling. In S. Matwin and J. Mielniczuk, editors, *Challenges in Computational Statistics and Data Mining*, pages 135–154. Springer International Publishing, Cham, 2016.
- [11] M. Sołtys, S. Jaroszewicz, and P. Rzepakowski. Ensemble methods for uplift modeling. *Data Mining and Knowledge Discovery*, 29(6):1531–1559, Nov 2015.
- [12] L. Wyrwicz, S. Jaroszewicz, P. Rzepakowski, and K. Bujko. Uplift modeling in selection of patients to either radiotherapy or radiochemotherapy in resectable rectal cancer: reassessment of data from the phase 3 study. *Annals of Oncology*, 26(suppl_4):iv107, 2015. conference abstract.
- [13] S. Jaroszewicz and P. Rzepakowski. Uplift modeling with survival data. In *ACM SIGKDD Workshop on Health Informatics (HI-KDD’14)*, New York City, USA, August 2014.
- [14] M. Korzeń and Szymon Jaroszewicz. PaCAL: A python package for arithmetic computations with random variables. *Journal of Statistical Software*, 57(10), 5 2014.
- [15] O. Jarczyk, B. Gruszka, S. Jaroszewicz, L. Bukowski, and A. Wierzbicki. Github projects. Quality analysis of open-source software. In *Proc. of the 6th International Conference on Social Informatics (SocInfo’14)*, pages 80–94, Barcelona, Spain, November 2014. **Best paper nominee.**
- [16] B. Żogała-Siudem and S. Jaroszewicz. Fast stepwise regression on Linked Data. In *Proc. of the 1st Workshop on Linked Data for Knowledge Discovery (LD4KD) co-located with ECML/PKDD’14*, pages 17–26, Nancy, France, September 2014.
- [17] L. Zaniewicz and S. Jaroszewicz. Support vector machines for uplift modeling. In *The First IEEE ICDM Workshop on Causal Discovery (CD 2013)*, Dallas, Texas, December 2013.
- [18] L. Bukowski, M. Jankowski-Lorek, S. Jaroszewicz, and M. Sydow. What makes a good team of Wikipedia editors? A preliminary statistical analysis. In *Proc. of the 5th International Conference on Social Informatics (SocInfo’14)*, pages 14–28, Kyoto, Japan, November 2013.

- [19] M. Korzeń, S. Jaroszewicz, and P. Klęsk. Logistic regression with weight grouping priors. *Computational Statistics & Data Analysis*, 64:281–298, August 2013.
- [20] P. Rzepakowski and S. Jaroszewicz. Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems*, 32:303–327, August 2012.
- [21] M. Jaśkowski and S. Jaroszewicz. Uplift modeling for clinical trial data. In *ICML 2012 Workshop on Machine Learning for Clinical Data Analysis*, Edinburgh, Scotland, June 2012.
- [22] P. Rzepakowski and S. Jaroszewicz. Uplift modeling in direct marketing. *Journal of Telecommunications and Information Technology*, 2:43–50, 2012.
- [23] S. Jaroszewicz and M. Korzeń. Arithmetic operations on independent random variables: A numerical approach. *SIAM Journal on Scientific Computing*, 34:A1241–A1265, 2012.
- [24] P. Rzepakowski and S. Jaroszewicz. Decision trees for uplift modeling. In *Proc. of the 10th International Conference on Data Mining (ICDM)*, pages 441–450, Sydney, Australia, December 2010.
- [25] S. Jaroszewicz. Using interesting sequences to interactively build hidden markov models. *Data Mining and Knowledge Discovery*, 21(1):186–220, 2010.
- [26] S. Jaroszewicz. Discovering interesting patterns in numerical data with background knowledge. In Y.S. Koh and N. Rountree, editors, *Rare Association Rule Mining and Knowledge Discovery: Technologies for Infrequent and Critical Event Detection*, pages 118–130. IGI Global, 2010.
- [27] S. Jaroszewicz, T. Scheffer, and D.A. Simovici. Scalable pattern mining with bayesian networks as background knowledge. *Data Mining and Knowledge Discovery*, 18(1):56–100, 2009.
- [28] S. Jaroszewicz. Interactive HMM construction based on interesting sequences. In *Proc. of Local Patterns to Global Models (LeGo’08) Workshop at the 12th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD’08)*, pages 82–91, Antwerp, Belgium, 2008.
- [29] S. Jaroszewicz. Minimum variance associations — discovering relationships in numerical data. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 172–183, Osaka, Japan, 2008.
- [30] T. Calders and S. Jaroszewicz. Efficient auc optimization for classification. In *11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD’07)*, pages 42–53, Warsaw, Poland, 2007. **Best paper award.**

- [31] S. Jaroszewicz, L. Ivantysynova, and T. Scheffer. Schema matching on streams with accuracy guarantees. *Intelligent Data Analysis*, 12(3):253–270, 2008.
- [32] S. Jaroszewicz and M. Korzeń. Approximating representations for large numerical databases. In *7th SIAM International Conference on Data Mining (SDM'07)*, pages 521–526, Minneapolis, MN, 2007.
- [33] S. Jaroszewicz, L. Ivantysynova, and T. Scheffer. Accurate schema matching on streams. In *4th International Workshop on Knowledge Discovery from Data Streams at the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'06)*, pages 3–12, 2006.
- [34] T. Calders, B. Goethals, and S. Jaroszewicz. Mining rank-correlated sets of numerical attributes. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, 2006.
- [35] S. Jaroszewicz. Polynomial association rules with applications to logistic regression. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, 2006.
- [36] S. Jaroszewicz and M. Korzeń. Comparison of information theoretical measures for reduct finding. In *Proc. of the 8th International Conference on Artificial Intelligence and Soft Computing (ICAISC'06)*, LNAI 4029, pages 518–527, Zakopane, June 2006. Springer-Verlag.
- [37] D. A. Simovici and S. Jaroszewicz. A new metric splitting criterion for decision trees. *Parallel Algorithms Appl.*, 21(4):239–256, 2006.
- [38] D. A. Simovici and S. Jaroszewicz. Generalized conditional entropy and a metric splitting criterion for decision trees. In *10th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD'06)*, pages 35–44, Singapore, 2006.
- [39] S. Jaroszewicz and T. Scheffer. Fast discovery of unexpected patterns in data, relative to a bayesian network. In *11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2005)*, pages 118–127, Chicago, IL, August 2005.
- [40] D. Simovici and S. Jaroszewicz. A new metric splitting criterion for decision trees. *International Journal of Parallel, Emergent and Distributed Systems*, 21(4):239–256, August 2006.
- [41] S. Jaroszewicz and W. Kosiński. Machine learning for speech recognition researchers. In *Proceedings of the Speech Analysis, Synthesis and Recognition Applications of Phonetics Conference*, Kraków, Poland, September 2005. Publication on CD-ROM.

- [42] M. Korzeń and S. Jaroszewicz. Finding reducts without building the discernibility matrix. In *Proceedings of the Fifth International Conference on Intelligent Systems Design and Applications (ISDA'05)*, pages 450–455, Wrocław, Poland, 2005.
- [43] D. Simovici and S. Jaroszewicz. A metric approach to building decision trees based on Goodman-Kruskal association index. In *PAKDD 2004*, LNAI 3056, pages 181–190, Sydney, Australia, May 2004. Springer-Verlag.
- [44] S. Jaroszewicz, D. A. Simovici, and I. Rosenberg. Measures on Boolean polynomials and their applications in data mining. *Discrete Applied Mathematics*, 144(1–2):123–139, November 2004.
- [45] Szymon Jaroszewicz and Dan Simovici. Interestingness of frequent itemsets using bayesian networks as background knowledge. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, pages 178–186, Seattle, WA, August 2004.
- [46] S. Jaroszewicz, D. A. Simovici, W. P. Kuo, and L. Ohno-Machado. The Goodman-Kruskal coefficient and its applications in genetic diagnosis of cancer. *IEEE Transactions on Biomedical Engineering*, 51(7):1095–1102, July 2004.
- [47] Szymon Jaroszewicz. *Information Theoretical and Combinatorial Methods in Data Mining*. PhD thesis, University of Massachusetts Boston, December 2003.
- [48] Dan A. Simovici and Szymon Jaroszewicz. Generalized conditional entropy and decision trees. In *Journées francophones d'Extraction et de Gestion de Connaissances (EGC 2003)*, pages 369–380, Lyon, France, January 2003.
- [49] S. Jaroszewicz and D. A. Simovici. Support approximations using Bonferroni-type inequalities. In *6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2002)*, pages 212–223, Helsinki, Finland, August 2002.
- [50] S. Jaroszewicz and D. A. Simovici. Pruning redundant association rules using maximum entropy principle. In *Advances in Knowledge Discovery and Data Mining, 6th Pacific-Asia Conference, PAKDD'02*, pages 135–147, Taipei, Taiwan, May 2002.
- [51] I. Rosenberg, D. A. Simovici, and S. Jaroszewicz. On functions defined on free boolean algebras. In *IEEE International Symposium on Multiple-Valued Logic ISMVL'02*, Boston, MA, May 2002.
- [52] S. Jaroszewicz, D. A. Simovici, and I. Rosenberg. An inclusion-exclusion result for boolean polynomials and its applications in data mining. In *Workshop on Discrete Mathematics and Data Mining (DM&DM), Second SIAM International Conference on Data Mining*, Arlington, VA, April 2002.

- [53] D. A. Simovici and S. Jaroszewicz. An axiomatization of partition entropy. *IEEE Transactions on Information Theory*, 48(7):2138–2142, July 2002.
- [54] S. Jaroszewicz and D. A. Simovici. A general measure of rule interestingness. In *5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2001)*, pages 253–265, 2001.
- [55] S. Jaroszewicz and D. A. Simovici. Data mining of weak functional decompositions. In *Proc. of the 30th International Symposium on Multiple-Valued Logic (ISMVL 2000)*, pages 77–82, 2000.
- [56] D. A. Simovici and S. Jaroszewicz. On information-theoretical aspects of relational databases. In C. Calude and G. Paun, editors, *Finite versus Infinite*. Springer Verlag, London, 2000.
- [57] S. Jaroszewicz and D. A. Simovici. On axiomatization of conditional entropy of functions between finite sets. In *Proceedings of the 29th International Symposium on Multiple-Valued Logic*, pages 24–28, Freiburg, Germany, 1999.
- [58] S. Jaroszewicz. Minimization of incompletely specified multiple-valued functions in reed-muller domain. In *Proc. of the 19th International Scientific Symposium for Students and Young Research Employees*, pages 121–126, Zielona Góra, Poland, 1997. (in Polish).
- [59] S. Jaroszewicz, V. Shmerko, and S. Yanushkevich. Exact irredundant searching for a minimal reed-muller expansion for an incompletely specified mvl function. In *Proc. of the International Conference on Application of Computer Systems*, pages 65–74, Szczecin, Poland, 1996.
- [60] A. D. Zakrevskij, S. N. Yanushkevich, and S. Jaroszewicz. Minimization of reed-muller expansions for systems of incompletely specified mvl functions. In *Proc. of the International Conference on Methods and Models in Automatics and Robotics (MMAR'96)*, pages 1085–1090, Międzyzdroje, Poland, 1996.