CURRICULUM VITAE AND PUBLICATIONS

**PART 1**

| **1a. Personal details** | | | | |
|---|---|---|---|---|
| **Full name** | *Title* Professor | *First name* Bernhard | *Second name(s)* Markus | *Family name* Pfahringer |
| **Present position** | Professor | | | |
| **Organisation/Employer** | The University of Waikato | | | |
| **Contact address** | Hillcrest Road | | | |
| | Hillcrest | | | |
| | Hamilton | | Post code | 3240 |
| **Work telephone** | 07-838 4041 | Mobile | 021-027 021 03 | |
| **Email** | bernhard@cs.waikato.ac.nz | | | |
| **Personal website (if applicable)** | http://www.cs.waikato.ac.nz/˜bernhard | | | |

## 1b. Academic qualifications

*1995, Dr.techn., Computer Science, Vienna University of Technology*
*1985, Dipl.-Ing., Computer Science, Vienna University of Technology*

## 1c. Professional positions held

*Dec 2017 - ongoing: Professor, Computer Science Department, University of Waikato*
*June 2017 - Nov 2017: Professor, Computer Science Department, University of Auckland*
*2015 - May 2017: Professor, Computer Science Department, University of Waikato*
*2007 - 2014: Associate Professor, Computer Science Department, University of Waikato*
*2000 - 2007: Senior Lecturer, Computer Science Department, University of Waikato*
*1997 - 2000: Research Fellow, Austrian Research Institute for AI*
*1996 - 1997: Postdoctoral Research Fellow, Computer Science, University of Waikato*
*1992 - 1998: Research Fellow, Austrian Research Institute for AI, Machine Learning*
*1986 - 1991: Research Assistant, Inst. for Medical Cybernetics and AI, University of Vienna*
*1985 - 1991: Research Associate, Austrian Research Institute for AI, Expert Systems*

## 1d. Present research/professional speciality

*Machine Learning, data mining.*

| **1e. Total years research experience:** | 29 years |
|---|---|

## 1f. Professional distinctions and memberships (including honours, prizes, scholarships, boards or governance roles, etc)

*2004 - ongoing: Editorial board member, Machine Learning Journal*
*2015 - ongoing: Editorial Board member, Data Mining and Knowledge Discovery Journal*

*2015 Founding member, and Co-Vice Chair: ACM SIGKDD ANZ Chapter*
*2013 - ongoing: Member of the IEEE Data Mining TC of the CI Society*
*2010 - 2015: Steering Committee Member for Asian Conference on Machine Learning*
*2010 - ongoing: Steering Committee Member for Discovery Science Conference*

*2019 invited public lecture for the Ross Ihaka Lecture series, Auckland*
*2018 invited keynote talk at the 5th Workshop on Machine Learning for Sensory Data Analytics, Wellington*

*2018 invited keynote talk at the IoT Large Scale Machine Learning from Data Streams Workshop, Dublin*
*2017 invited keynote talk at Ninth Asian Conference on Intelligent Information and Database Systems, Kanazawa*
*2017 Invited lectures on Stream Mining,Third International Winter School on Big Data, Bari*
*2016 Invited keynote talk at the Third European Network Intelligence Conference, Wroclaw*
*2015 Invited keynote talk at the MetaSel Workshop, at ECMLPKDD2015C, Porto*
*2015 Invited keynote talk at the BigMine 2015 Workshop, at KDD2015, Sydney*
*2013 invited keynote talk at the Machine Learning for Sensory Data Analysis, Dunedin*
*2012 invited keynote talk at The First International Workshop on Learning with Weak Supervision (LAWS'12), ACML, Singapore*

*2015 Program co-chair, Joint Australasian AI Conference 2015, Canberra, Australia*
*2010 Program co-chair Discovery Science 2010 (DS2010), Canberra, Australia*
*2005 Program co-chair ILP2005, Bonn, Germany*
*2016 Tutorial co-chair, Asian Conference on Machine Learning, Hamilton, NZ*
*Area Chair, ECML PKDD 2017, Skopje, Mazedonia, ECML PKDD 2016, Riva del Garda, Italy, ECML PKDD 2015, Porto, Portugal, ECML PKDD 2014, Nancy, France, ECML PKDD 2013, Prague, Czech Republic, ECML PKDD 2012, Bristol, UK, ECML/PKDD2007, Warsaw, Poland, ECML/PKDD2006, Berlin, Germany*
*2016 General Chair MLSDA Workshop, PAKDD, Auckland, NZ*
*2012 Special Session Chair, PRICAI2012, Kuching, Malaysia*
*2009 Workshop chair, PAKDD2009, Bangkok, Thailand*
*2004 Workshop chair, PRICAI2004, Auckland, New Zealand*

*2002 - ongoing: Program committee member / reviewer for multiple conferences per year, for instance KDD2018, ACML2018, ILP2018, DS2018, KDD2017, ACML2017, ILP2017, DS2017, KDD2016, ACML2016, ILP2016, DS2016, KDD2015, ACML2015, ILP2015, DS2015, KDD2014, ACML2014, ILP2014, DS2014, PRICAI2014, and many older ones.*

*Student supervision*
*PhD students: 11 (chief supervisor), 7 (co-supervisor)*
*Masters students: 8 (chief supervisor), 3 (co-supervisor)*
*Honours students: 22*

*Grants (in New Zealand, since 2000)*
*Theme Leader, MBIE grant: Precision driven healthcare initiative, 2016-2023*
*Associate Investigator, Marsden Grant: "Deep learning without the headache", 2016-2018*
*BuildIT Post-Doctoral Award for Dr.Bifet, 2010-2012*
*Objective Leader, FRST project on GCMS prediction, 2008-2012*
*Principal Investigator, Marsden Grant, 2004-2007*
*Objective Leader, FRST project on NIR prediction, 2004-2007*
*Consulting for Metrix, Orica, Crop and Food, Hill Labs and Mariner7.com*

| **1g. Total number of** *peer reviewed* **publications and patents** | Journal articles | Books, book chapters, books edited | Conference proceedings | Patents |
|---|---|---|---|---|
| | 31 | 10 | 109 | 0 |

**PART 2**

## 2a. Research publications and dissemination

Barddal J.P, Enembreck F., Gomes H.M., Bifet A., **Pfahringer B.** (**2019**) Boosting Decision Stumps for Dynamic Feature Selection on Data Streams, Inf Sys, Feb 2019.

Leathart T., Frank E., **Pfahringer B.**, Holmes G. (**2019**) Ensembles of Nested Dichotomies with Multiple Subset Evaluation, PAKDD2019.

Leathart T., Frank E., **Pfahringer B.**, Holmes G. (**2019**) On Calibration of Nested Dichotomies, PAKDD2019.

Bifet A., Gavalda R., Holmes G., **Pfahringer B.** (**2018**) Machine Learning for Data Streams with Practical Examples in MOA, MIT Press.

Gouk H., **Pfahringer B.**, Frank E., Cree M.J. (**2018**) MaxGain: Regularisation of Neural Networks by Constraining Activation Magnitudes, ECMLPKDD2018.

Peng Y., Koh Y.S., Riddle P., **Pfahringer B.** (**2018**) Using Supervised Pretraining to Improve Generalization of Neural Networks on Binary Classification Problems, ECMLPKDD2018.

van Rijn J., Holmes G., **Pfahringer B.**, Vanschoren J. (**2018**) The online performance estimation framework: heterogeneous ensemble learning for data streams. Machine Learning., 107(1):149-176.

Bravo-Marquez F., Frank E., **Pfahringer B.** (**2018**) Transferring sentiment knowledge between words and tweets. Web Intelligence 16(4): 203-220.

Yuan L., **Pfahringer B.**, Barddal J.P. (**2018**) Iterative subset selection for feature drifting data streams. SAC 2018: 510-517.

Leathart T., Frank E., Holmes G., **Pfahringer B.** (**2017**) Probability calibration trees. Proc 9th Asian Conference on Machine Learning. PMLR 77: 145-160.

Gouk H., Cree M., **Pfahringer B.** (**2017**) Learning Distance Metrics for Multi-Label Classification. ACML2016: 318-333.

Barddal J.P, Gomes H.M., Enembreck F., **Pfahringer B.** (**2017**) A survey on feature drift adaptation: Definition, benchmark, challenges and future directions. Journal of Systems and Software 127, 278-294.

Cerqueira V., Torgo L., Oliveira M., **Pfahringer B.** (**2017**) Dynamic and Heterogeneous Ensembles for Time Series Forecasting. IEEE International Conference on Data Science and Advanced Analytics (DSAA), p.242-251.

Bifet A., Zhang J., Fan W., He C., Zhang J., Qian J., Holmes G., **Pfahringer B.** (**2017**) Extremely fast decision tree mining for evolving data streams. Proceedings of the 23rd ACM SIGKDD, p.1733-1742.

Bravo-Marquez F., Frank E., **Pfahringer B.** (**2016**) From opinion lexicons to sentiment classification of tweets and vice versa: a transfer learning approach. Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI'16).

Bravo-Marquez F., Frank E., **Pfahringer B.** (**2016**) Annotate-Sample-Average (ASA): A New Distant Supervision Approach for Twitter Sentiment Analysis. ECAI 2016: 498-506.

Barddal J.P., Murilo-Gomes H., Enembreck F., **Pfahringer B.**, Bifet A. (**2016**) On Dynamic Feature Weighting for Feature Drifting Data Streams. ECML2016: 2/129-144.

Leathart T., **Pfahringer B.**, Frank E. (**2016**) Building Ensembles of Adaptive Nested Dichotomies with Random-Pair Selection. ECMLPKDD2016: 2/179-194.

Bravo-Marquez F., Frank E., **Pfahringer B.** (**2016**) Building a Twitter opinion lexicon from automatically-annotated tweets. Knowledge-Based Systems 108: 65-78.

Read J., Reutemann P., **Pfahringer B.**, Holmes G. (**2016**) MEKA: A multi-label/multi-target extension to WEKA. Journal of Machine Learning Re- search 17, 21, 1-5.

Barddal J.P., Murilo-Gomes H., Enembreck F., Pfahringer B. (**2016**) A survey on feature drift adaptation: Definition, benchmark, challenges and future directions. Journal of Systems and Software.

Torgo L., Branco P., Ribeiro R.P., **Pfahringer B.** (**2015**) Resampling strategies for regression. Expert Systems 32(3): 465-476.
Zliobaite I., Bifet A., Read J., **Pfahringer B.**, Holmes G. (**2015**) Evaluation methods and decision theory for classification of streaming data with temporal dependence. Machine Learning 98(3): 455-482.
Bifet A., de Francisci Morales G., Read J., Holmes G., **Pfahringer, B.** (**2015**) Efficient online evaluation of big data stream classifiers. 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 59-68). Sydney, Australia: ACM.
Bravo-Marquez, F. Frank, E., **Pfahringer B.** (**2015**) From unlabelled tweets to Twitter-specific opinion words. In Proc 38th International ACM SIGIR Conference on Research and Development (pp. 743-746). Santiago, Chile: ACM.
van Rijn J.N., HolmesG., **Pfahringer B.**, Vanschoren J. (**2015**) Having a Blast: Meta-Learning and Heterogeneous Ensembles for Data Streams. ICDM 2015: 1003-1008.
Bravo-Marquez F., Frank E., **Pfahringer B.** (**2015**) Positive, Negative, or Neutral: Learning an Expanded Opinion Lexicon from Emoticon-Annotated Tweets. IJCAI: 1229-1235.
Sakthithasan S., Pears R., Bifet A., **Pfahringer B.** (**2015**) Use of ensembles of Fourier spectra in capturing recurrent concepts in data streams. IJCNN 2015: 1-8.
Hapfelmeier A., **Pfahringer B.**, Kramer S. (**2014**) Pruning Incremental Linear Model Trees with Approximate Lookahead. IEEE Trans. Knowl. Data Eng. 26(8): 2072-2076
Zliobaite I., Bifet A., Read J., **Pfahringer B.**, Holmes G. (**2014**) Evaluation methods and decision theory for classification of streaming data with temporal dependence. Machine Learning, April 2014, 1–28.
Zliobaite I., Bifet A., **Pfahringer B.**, Holmes G. (**2014**) Active Learning With Drifting Streaming Data. IEEE Trans. NNLS 25(1): 27-39.
Ienco D., Bifet A., **Pfahringer B.**, Poncelet P. (**2014**) Change Detection in Categorical Evolving Data Streams, SAC2014, 792–797.
Rijn J.van, **Pfahringer B.**, Holmes G. and Vanschoren J. (**2014**) Algorithm Selection Problem in Stream Mining, Discovery Science.
Sun Q., **Pfahringer B.** (**2014**) Hierarchical Meta-Rules for Scalable Meta-Learning, PRICAI.
Sun Q., **Pfahringer B.** (**2013**) Pairwise meta-rules for better meta-learning-based algorithm ranking. Machine Learning 93(1): 141-161.
Bifet A., Read J., Zliobaite I., **Pfahringer B.**, Holmes G. (**2013**) Pitfalls in Benchmarking Data Stream Classification and How to Avoid Them. ECML/PKDD (1) 2013: 465-479
Bifet A., Frank E., Holmes G., **Pfahringer B.** (**2012**) Ensembles of Restricted Hoeffding Trees, ACM TIST, 3, 30.
Read J., Bifet A., Holmes G., **Pfahringer B.** (**2012**) Scalable and efficient multi-label classification for evolving data streams, Machine Learning Journal, 88, 243-272.
Vanschoren J., Blockeel H., **Pfahringer B.**, Holmes G. (**2012**) Experiment databases: A new way to share, organize and learn from experiments, Machine Learning Journal, 87, 127-158.
Read J., **Pfahringer B.**, Holmes G., Frank E. (**2011**) Classifier chains for multi-label classification, Machine Learning Journal, 85, 333-359.
Bifet, A., Holmes, G., Kirkby, R., **Pfahringer B.** (**2010**) MOA: Massive Online Analysis. Journal of Machine Learning Research 11, 1601-1604.
Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., **Pfahringer B.**, Reutemann, P., Witten, I. H. (**2010**) WEKA: Experiences with a Java open-source project. Journal of Machine Learning Research (JMLR) 11, 2533-2541.
Hall M., Frank E., Holmes G., **Pfahringer B.**, Peter Reutemann, Ian H. Witten (**2009**) The WEKA data mining software: an update, ACM SIGKDD Explorations, 11/10-18.

## 2b. Previous research work

**Research title:** GCMS Prediction (Objective Leader, FRST funded)
**Principal outcome:** Prediction of the concentration of various chemicals (e.g. pesticides on fruit, aromatic hydrocarbons in the soil around petrol stations, and similar) from GCMS (gas chromatography coupled with mass spectroscopy) outputs
**Principal end-user and contact:** Hill Labs, Hamilton, and BLGG, Netherlands

**Research title:** NIR Prediction (Objective Leader, FRST funded)
**Principal outcome:** Prediction of the concentration of elements or compounds in samples (e.g. total nitrogen content in a soil sample, sugar content in a kiwi fruit, and similar) from NIR (Near-Infrared) spectrograms
**Principal end-user and contact:** Hill Labs, Hamilton, and BLGG, Netherlands

**Research title:** Collective Classification (Principal Investigator, Marsden Funds)
**Principal outcome:** Methods for prediction of sets of examples at a time, that take relationships between examples into account, instead of the standard practise of predicting each example in isolation.
**Principal end-user and contact:** not applicable

**Research title:** MetaL (Objective Leader, EU funded)
**Principal outcome:** Novel meta learning methods based on a combination of statistical properties of the dataset in question, as well as the results of simple landmarking learners.
**Principal end-user and contact:** Daimler-Benz Research Centre, Ulm, Germany

## 2c. Describe the commercial, social or environmental impact of your previous research work

Both Hill Labs and BLGG use the NIR prediction system on a daily basis. For this particular task NIR prediction has replaced more than 90% of the original “wet lab” work, thus leading to much faster turn around times and lower costs. BLGG (and a new company split off from BLGG) also use the tools developed for GCMS prediction in their everyday work, again replacing a lot of ”wet lab” work. This provides for significant cost reduction, and also for some benefits to the environment, due to the reduced consumption of the various analytical chemicals needed for running a ”wet lab”.

Daimler Benz use data mining across a large number of tasks, from call centre evaluation over credit assessment all the way to engine optimisation. Meta-learning has helped them to improve their data mining outcomes and at the same time to reduce the resources needed for optimising their models.

The Marsden project has not lead to any direct commercial outcomes, and one would not expect that from a Marsden project, but in addition to research publications it is has helped improve human capital. One of our current key research programmers came to us through this project.

## 2d. Demonstration of relationships with end-users

*Consulting work (in NZ) for Metrix, Orica, Crop and Food, Hill Labs and Mariner7.com*