Grigorios D. Tsoumakas

Associate Professor

Department of Informatics

Aristotle University of Thessaloniki

Extensive Curriculum Vitae

Thessaloniki, April 2022

1.   Personal Information

Date of birth:	28/3/1977
Place of birth:	Thessaloniki
Marital status:	Married, with twins born 8/2013
Military obligations:	Fulfilled. Pvt Research & Informatics, 8/8/2005 - 8/8/2006
Home address:	P.O. Box 1823, Trilofos 57500, Thessaloniki
Work address:	Department of Informatics, 54124, Thessaloniki
Phone number:	+30 2310 998887 (work), +30 6977 444082 (mobile)
Email:	greg@csd.auth.gr
URL:	http://users.auth.gr/greg

2.   Education

University degrees

1994-1999	School of Informatics, Aristotle University of Thessaloniki, Greece. Honours in Informatics, Grade 8.1 out of 10, very good. Thesis: *Combining Expert and Georgraphical Information Systems for Land Evaluation.*
1999-2000	School of Informatics, University of Edinburgh, U.K. M.Sc. in Artificial Intelligence, Grade A-, excellent (distinction). Thesis: *A Knowledge Based System for Rock Fracture Classification in Deep Boreholes.*
2001-2005	PhD, School of Informatics, Aristotle University of Thessaloniki, Greece. Thesis: *Machine Learning for the Fusion of Multiple Distributed Intelligent Systems.*

Foreign languages

English: Very good reading, writing, and communicating, Cambridge Proficiency, Grade B, 1995.

German: Elementary reading, writing, and communicating, Grundstuffe, Gesamtnote Befriedigends, 1993.

Continuous education

12/2015	University of California, Berkeley, edX Verified Certificate for *The Science of Happiness*.
10/2016	ETH Zurich, edX Verified Certificate for *Agile Software Development*.

3.   Scholarships

2009-2010	IKYDA scholarship supporting my research collaboration with TU Darmstadt, Germany. IKYDA is a bilateral research promotion programme, concluded between the Greek State Scholarship Foundation (IKY) and the German Academic Exchange Service (DAAD).
2002-2004	Three scholarships for corresponding paper presentations in European Conferences from a) the European Coordinating Committee for Artificial Intelligence (ECCAI), b) the Hellenic Artificial Intelligence Society (EETN) and c) the KDNET network of excellence.
2001-2002	Two scholarships for attending corresponding schools, a) from the European Committee (project HPCF-2000-00401) for the school entitled "Intelligent Data Analysis" in Palermo, Italy and b) from Alexander S. Onassis public benefit foundation for the school entitled "The Data Avalanche: Reducing Information Overload", in Heraklion, Greece.
1999-2000	Scholarship from the Student Awards Agency for Scotland covering the tuition fees of the MSc in Artifical Intelligence.

4.   Work Experience

10/2020 – Today	Associate Professor, School of Informatics, Aristotle University of Thessaloniki.
02/2019 – Today	Academic Co-founder, Medoid AI.
06/2019 – 08/2019	Academic Visitor, IBM Researh, Zurich, Switzerland.
03/2013 – 10/2020	Assistant Professor, School of Informatics, Aristotle University of Thessaloniki.
08/2007 – 3/2013	Lecturer, Department of Informatics, Aristotle University of Thessaloniki.
2002 – 2004	Consultant for SMEs of the areas of Thessaloniki and Chalkidiki in adopting e-business practices. Programme "Go-Digital" funded by the Greek Ministry of Development.
2001 – 2003	Lecturer of seminars on C, C++, Microsoft Word/Excel in the Association of Greek Physicists, Central and Western Macedonia Branch (~ 400 hours).
2001 – 2002	Lecturer in IEK Kalamaria and IEK Thessaloniki 2, for a total of 377 hours, in the specialties Industrial Software Technician, Tele-Informatics Technician and Informatics Applications Specialist (3 semesters).
01/2000 – 09/2000	Database and web programmer (part-time). "Dynamic Blue Ltd", Edinburgh, UK.
09/1998 – 06/1999	Database and web programmer, tehnical support, network administration (part-time). Information Technology Center, Aristotle University of Thessaloniki, Greece.
03/1998 – 12/1998	Programmer of a commercial multimedia application (part-time). "Digital Arts" (in collaboration with MLS Multimedia), Thessaloniki, Greece.
10/1996 – 01/1998	Information technology manager (part-time). "STAR FM 97.1" radio station, Thessaloniki, Greece.

5.   Teaching Experience

2018-today	Teaching the postgraduate course Natural Language Processing, School of Infomatics, AUTh, Greece.
2015-2018	Teaching together with colleagues Petros Nicopolitids and Dimitris Vrakas the undergraduate course Mobile Networks & Application Development, School of Infomatics, AUTh, Greece. I'm covering the Android development part.
2009-today	Teaching the postgraduate course Advanced Machine Learning & Knowledge Discovery, School of Infomatics, AUTh, Greece.
2008-today	Teaching the undergraduate course Object Oriented Programming, School of Infomatics, AUTh, Greece. From 2008 till 2012, I was teaching this togerther with colleague Panagiotis Katsaros.
2007-2019	Teaching the undergraduate course Web Information Systems (together with colleague Athena Vakali), School of Infomatics, AUTh, Greece.
2007-today	Teaching the undergraduate courses Decision Theory and Systems (together with colleague Ioannis Vlahavas), School of Infomatics, AUTh, Greece.
2007-2015	Teaching together with colleague Dimitris Vrakas the undergraduate course Intelligent Autonomous Systems, School of Infomatics, AUTh, Greece.
2011-2013	Teaching together with colleague Athena Vakali the postgraduate course Web Information Systems, International Hellenic University, Greece.
2007-2012	Teaching the undergraduate course Multiagent Systems, School of Infomatics, AUTh, Greece.

Since the academic year 2010-2011, I have supervised 41 postgraduate theses. Since the academic year 2007-2008, I have supervised 56 graduate theses.

6. Research Projects

- *Title*: Advanced Multi-Label Learning Techniques (AMULET). *Starting date*: 12/2019. *Ending date*: 12/2022. *Budget*: € 158.270,00. *Source of funding*: Hellenic Foundation for Research and Innovation. *Role in the research team*: Principal Investigator.
- *Title*: Exploitation of Cultural Assets with computer-assisted Recognition, Labeling and meta-data Enrichment. *Starting date*: 6/2018. *Ending date*: 6/2021. *Budget*: € 133.600. *Source of funding*: Co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship, and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T1EDK-05580). *Role in the research team*: Principal Investigator.
- *Title*: A European AI On Demand Platform and Ecosystem (AI4EU). *Starting date*: 1/2019. *Ending date*: 12/2021. *Budget*: € 110.625. *Source of funding*: EU H2020. *Role in the research team*: Researcher.
- *Title*: Semantic indexing and question answering in big data. *Starting date*: 9/2018. *Ending date*: 8/2019. *Budget*: € 108.000. *Source of funding*: Atypon Systems Inc. *Role in the research team*: Researcher.
- *Title*: Semantic indexing and question answering in big data. *Starting date*: 3/2017. *Ending date*: 2/2018. *Budget*: € 120.000. *Source of funding*: Atypon Systems Inc. *Role in the research team*: Researcher.
- *Title*: Semantic indexing and question answering in big data. *Starting date*: 11/2015. *Ending date*: 10/2016. *Budget*: € 60.000. *Source of funding*: Atypon Systems Inc. *Role in the research team*: Researcher.
- *Title*: Semantic indexing and question answering in big data. *Starting date*: 1/2015. *Ending date*: 07/2015. *Budget*: € 30.000. *Source of funding*: Atypon Systems Inc. *Role in the research team*: Researcher.
- *Title*: Daily gas consumption forecasting. *Starting date*: 8/2014. *Ending date*: 12/2014. *Budget*: € 10.000. *Source of funding*: Gas Distribution Company of Thessaly. *Role in the research team*: Researcher.
- *Title*: Customer data analysis system. *Starting date*: 6/2014. *Ending date*: 12/2014. *Budget*: € 28.000. *Source of funding*: Diamantis Masoutis S.A. *Role in the research team*: Researcher.
- *Title*: Information System for productivity and customer service improvement in the Public Transportation Companies. *Starting date*: 4/2014. *Ending date*: 6/2015. *Budget*: € 191.304 (€ 66.900). *Source of funding*: Greek General Secretariat for Research and Technology. *Role in the research team*: Researcher.
- *Title*: Intelligent System in Supply Chain Monitoring and Optimization. *Starting date*: 6/2012. *Ending date*: 6/2014. *Budget*: € 338.797 (€ 27.000). *Source of funding*: Greek General Secretariat for Research and Technology. *Role in the research team*: Researcher, applying machine learning aglorithms on sales/supply data for sales prediction.
- *Title*: I-NET (Interoperability Enterprise NetWork): Consortium on Interoperability, Flexibility and Adaptability of B2B Transactions across Enterprises. *Starting date*: 1/6/2007. *Ending date*: 31/10/2008. *Budget*: € 100.750 (€ 25.340). *Source of funding*: Greek General Secretariat for Research and Technology. *Role in the research team*: Researcher, applying data mining techniques on past business transaction data with the goal of building a model that predicts the credibility of enterprises.
- *Title*: SWIM: Development of an Intelligent System for the management of Semantic Web Services. *Starting date*:1/11/2005. *Ending date*:31/10/2008. *Budget*: € 138.240 (€ 88.664). *Source of funding*: Greek General Secretariat for Research and Technology. *Role in the research team*: Researcher, developing data mining methods for automated classification of Web services according to their application domain.
- *Title*: SWEL: Planning Systems in the Semantic Web for E-Learning. *Starting date*: 1/1/2005. *Ending date*: 31/12/2007. *Budget*: € 50.000. *Source of funding*: Greek Ministry of National Education and Religious Affairs. *Role in the research team*: Researcher, responsible for the design of a text classification system that matches the textual description of an educational resource to one or more nodes of an educational terms ontology.
- *Title*: EUROCITIZEN: An interactive, agent-based, online edutainment environment to promote European intercultural education and cooperation among citizens of Europe. *Starting date*: 1/9/2002. *Ending date*: 28/2/2005. *Budget*: € 750.000 (€ 69.633). *Source of funding*: European Union under the IST program. *Role in the research team*: Developer, responsible for developing the question selection mechanism.

7.   Scientific Activities

Invited talks and lectures
- *Mining Biomedical Publications*, August 31, 2018
  2nd Workshop on Biomedical Information Management, Hamburg, Germany.
- *Machine Learning for Understanding Biomedical Publications*, October 12, 2017
  12th Conf. of the Hellenic Society for Computational Biology and Bioinformatics, Athens Greece.
- Learning from Multi-Label Data, January 12, 2017
  School of Electrical & Computer Eng., Aristotle University of Thessaloniki, Thessaloniki, Greece.
- *Exploiting Target Dependencies in Multi-Target Prediction,* September 13, 2016
  25th Belgian-Dutch Conference on Machine Learning, Kortrijk, Belgium.
- *Multi-Target Prediction and Applications in the Publishing, Energy and Retail Industries*, June 21, 2016.
  2016 South-East Forum on Data Science, Belgrade, Serbia.
- *Semantic Indexing of Biomedical Literature*, February 20, 2014
  XEROX Research Center Europe, Grenoble, France, February 20, 2014
- *Multi-Label Learning for Concept-Based Retrieval*, November 16, 2011
  Informatics and Telematics Institure, Centre for Research & Technology – Hellas, Thesaloniki, Greece.
- *Learning from Multi-Label Data*, May 10, 2011.
  Department of Informatics of the Ionian University, Kerkyra, Greece.
- *Ensemble Methods for Multi-Label Data*, September 20, 2010.
  2010 Supervised and Unsupervised Ensemble Methods Workshop of ECML PKDD, Barcelona, Spain.
- *Learning from Multi-Label Data*, August 24, 2009.
  1st Hellenic Artificial Intelligence Summer School (HAISS), Thessaloniki, Greece.

Prizes and Awards
- Marco Ramoni Best Paper Award at the 19th Internation Conference on Artificial Intelligence in Medicine (AIME 2021) for the paper "A multi-instance multi-label weakly-supervised approach for dealing with emerging MeSH descriptors" by Nikolaos Mylonas, Stamatis Karlos and Grigorios Tsoumakas.
- Winner of the Third Round of Crowdflower's AI for Everyone Challenge[1], in collaboration with Prof. A. Vakali and 4 US universities for the proposal: "Hate Speech, Cyberbullying, and Microaggressions: Creating a Taxonomic Database of Abusive Online Behavior", January 2018.
- Senior Member of ACM, December 2017.
- 2nd place among 14 teams at the Technologically Assisted Reviews (TAR) in Empirical Medicine task of the 2017 CLEF lab on eHEALTH. In collaboration with BSc student A. Anagnostou and PhD student A. Lagopoulos. Similar position in the 2018 version.
- ECML PKDD 2017 Test of Time Award for the paper "Random k-Labelsets: An Ensemble Method for Multilabel Classification" presented at ECML PKDD 2007. In collaboration with Dr. I.Vlahavas.
- Best paper award at the 2014 Brazilian Conference on Intelligent Systems (BRACIS 2014) for the paper "Label Construction for Multi-Label Feature Selection" by Newton Spolaor, Maria Carolina Monard, Grigorios Tsoumakas and Huei Lee.
- 1st place at the large-scale online biomedical semantic indexing challenge organized in 2013 by the European FP7 research project BioASQ, outweigting the state-of-the-art operational system of the National Library of Medicine and the approaches of several teams, including Mayo Clinic and the National Center for Biotechnology Information. In collaboration with Prof. Ioannis Vlahavas and US company Atypon. In 2014-2017 we achieved the 2nd place, as well as similarly good positions in other tasks (grant extraction, question answering), in collaboration with several PhD students.
- 8th place at the bird classification challenge organized by the 9th IEEE international workshop on machine learning for signal processing (MLSP 2013) and hosted by Kaggle.
- 1st place at the concept-based retrieval subtask of the photo annotation task of ImageCLEF 2011. In collaboration with MSc student Kostantinos Sechidis and PhD student Eleftherios Spyromitros-Xioufis.
- 2nd place at the music instruments recognition track of the ISMIS 2011 international contest on music information retrieval.

---

[1] https://www.prnewswire.com/news-releases/crowdflower-announces-third-wave-of-ai-for-everyone-challenge-winners-300587610.html

Technical Committee Memberships
- Data Mining and Big Data Analytics Technical Committee (DMTC), Computational Intelligence Society (CIS), IEEE, 2015 – 2016.

Action Editor
- Data Mining and Knowledge Discovery, Springer, 2019-2021.

Editorial Board Member
- Data Mining and Knowledge Discovery, Springer, Editorial Board Member, 2016-2018, 2022 - today.
- Frontiers of Computer Science, Springer, 2019-today.

Conference Organization
- ECML PKDD 2022, Research and Advanced Data Science Programm co-chair together with Asja Fischer, Tias Guns and Petra Kralj-Novak.
- Discovery Science 2020, Programm co-chair together with Annalisa Appice.
- ACML 2018 tutorials co-chair together with Liwei Wang.
- SETN 2016 workshops/tutorials chair.
- WISE 2014 challenge co-chair together with Apostolos Papadopoulos and Weining Qian.
- ECML PKDD 2012 publicity chair.
- ECML PKDD 2011 publicity co-chair together with Annalisa Appice.

Workshop Organization
- "Big Multi-Target Prediction (BigTargets)" at the ECML PKDD 2015 conference in Porto, Portugal, together with Willem Waegeman, Krzysztof Dembczynski, Tapio Pahikkala, Antti Airola, Giorgio Valentini and Massih-Reza Amini.
- "Learning from Multi-Label Data (MLD'10) at the ICML 2010 conference in Haifa, Israel, together with Min-Ling Zhang and Zhi-Hua Zhou.
- "Learning from Multi-Label Data (MLD'09)" at the ECML PKDD 2009 conference in Bled, Slovenia, together with Min-Ling Zhang and Zhi-Hua Zhou.
- "Mining Social Data (MSODA'08)" at the ECAI 2008 conference in Patras, Greece, together with Alexandros Nanopoulos.

Tutorials
- "Automatic Keyphrase Extraction from Text: A Walk-through", at ECAI 2020, Online, 29 August 2020, together with Eirini Papagiannopoulou, Ricardo Campos.
- "Advanced Topics in Ensemble Learning", at ECML PKDD 2012, Bristol, UK, 24 September 2012, together with Daniel Hernandez-Lobato, Gonzalo Martinez-Munoz, Ioannis Partalas, Alberto Suarez.
- "Mining Complex Dynamic Data", at ECML PKDD 2011, Athens, Greece, 5 September 2011, together with Myra Spiliopoulou, Irene Ntoutsi, Arthur Zimek, Hans-Peter Kriegel.
- "Learning from Multi-Label Data", at ECML PKDD 2009, Bled, Slovenia, 7 September 2009, together with Min-Ling Zhang, Zhi-Hua Zhou.

Special Issues
- "Discovery Science", Machine Learning (in press), together with Annalisa Appice.
- "Learning from Multi-Label Data", Machine Learning 88(1-2), July 2012, together with Min-Ling Zhang, Zhi-Hua Zhou.
- "ECAI'08 Workshop on Mining Social Data", Int. Journal on Data Warehousing and Mining 6(1), 2010, together with Alexandros Nanopoulos.

Selected Journal Paper Review Invitations
- *Journal of Machine Learning Research*, *Machine Learning Journal*, *Neurocomputing*, *Data Mining and Knowledge Discovery, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Selected Programme Committee Memberships

- International Joint Conference on Artificial Intelligence (IJCAI) 2021 AC, 2017-2020 SPC, 2015, 2013, 2011
- European Conference on Machine Learning & Principles and Practice of Knowldege Discovery in Databases (ECML PKDD) 2021, 2020 SPC, 2018, 2016, 2015, 2014, 2013, 2012, 2011
- European Conference on Artificial Intelligence (ECAI) 2020, 2018 SPC, 2014, 2012
- Knowledge Discovery in Databases (KDD) 2019, 2018, 2016, 2015, 2011, 2010
- IEEE International Conference on Data Mining (ICDM) 2018, 2016, 2015, 2014
- NeurIPS 2018
- Asian Conference on Machine Learning (ACML) 2018 SPC, 2017 SPC, 2014, 2012
- Association for the Advancement of Artificial Intelligence (AAAI) 2018 SPC, 2013, 2012
- SIAM International Conference on Data Mining (SDM) 2017, 2016, 2015, 2013
- International Conference on Machine Learning (ICML) 2014, 2013, 2012, 2010

Member of scientific association
- Hellenic Artificial Intelligence Society since 1999
- ACM, SIGKDD
- IEEE, Computer Society
- ACL

8. Academic Activities

Supervision of PhD Theses
1. Yannis Papanikolaou, Machine Learning Techniques for Semantic Text Indexing, 03/2014 – 9/2018.
2. Eirini Papagianopoulou, Keyphrase Extraction Techniques, 05/2014 – 01/2021.
3. Dimitris Dimitriadis, Text Mining and Natural Language Processing, 05/2015 – today.
4. Athanasios Lagopoulous, Knowledge Discovery from Scientific Publications, 03/2017 – 10/2021.
5. Bin Liu, Addressing Class Imbalance in Multi-Label Data, 07/2017 – 12/2021.
6. Anastasios Nentidis, Biomedical Knowledge Discovery, 05/2018 – today.
7. Despoina Christou, Natural Language Processing, 9/2018 – today.
8. Alexios Gidiotis, Natural Language Understanding, 05/2019 – today.
9. Nikolaos Mylonas, Machine Learning and Knowledge Discovery, 2/2020 – today.
10. Tatiana Passali, Natural Language Processing, 10/2020 – today.
11. Dimitris Papadopoulos, Machine Learning for Drug Design, 10/2020 - today

External examiner of international PhD theses
- Mattia Setzu, "Opening the Black Box: Empowering Machine Learning Models with Explanations", University of Pisa, Italy, 2022.
- Tomaž Stepišnik, "Complex Nodes in Trees for Structured Output Prediction", Jozef Stefan International Postgraduate School, Slovenia, 2021.
- Arjun Pakrashi, "Improving Multi-label Classification Using Inter-label Associations and a New Kalman Filter Based Ensemble Method", University College Dublin, Ireland, October 2019.
- Arnauld Joly, "Exploiting Random Projections and Sparsity with Random Forests and Gradient Boosting Methods", University of Liege, Belgium, March 2017.
- Jesse Read, "Scalable Multi-Label Classification", Department of Computer Science, University of Waikato, New Zealand, September 2010.
- Daniel Hernandez-Lobato, "Prediction Based on Averages over Automatically Induced Learners: Ensemble Methods and Bayesian Techniques", Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain, January 2010.

Host of PhD students
- Konstantinos Pliakos, KU Leuven, 1/2018 – 6/2018.
- Newton Spolaor, University of São Paulo at São Carlos, 4/2013 – 9/2013.
- Everton Alvares Cherman, University of São Paulo at São Carlos, 6/2012 – 10/2012.
- William Groves, University of Minnesota, 7/2011 – 8/2011.

9.   Research Interests

My research interests belong to the areas of Machine Learning and Natural Language Processing. In specific, I have worked in the past/I am still working on the following topics:

- *Learning from Multi-Target Data*. Development of efficient and effective algorithms for predictive modelling from multi-target data.
- *Ensemble Methods.* Development of algorithms for the combination of multiple predictive models with the aim of increasing predictive accuracy.
- *Natural Language Processing*. Development of algorithms for keyphrase extraction, automatic summarization, question answering and relation extraction.
- *Text Mining*. Development of algorithms for learning from textual data, including text streams and their challenges (e.g., concept drift), with emphasis on semantic indexing and hate speech detection.
- *Interpetable Machine Learning*. Development of algorithms for local explanation of black box models, such as neural networks and random forests.
- *Learning for Planning*. Development of adaptive planning systems with the aid of machine learning techniques.
- *Distributed Data Mining*. Development of algorithms that learn from geographically distributed databases. Algorithms focus on: a) discovering useful knowledge about the similarities and differences between the distributed databases, and b) construction of predictive models with high accuracy.
- *Reinforcement Learning.* Development of algorithms for efficient reinforcement learning, by utilizing transfer learning techniques.
- *Applications.* Application of machine learning algorithms for the analysis of real-world data (biology, environment, music, software engineering, recommender systems, literature), with the aim of bringing out their practical value.

10.   Publications

I have edited 2 volumes of papers and have published 2 theses and 142 papers (6 book chapters, 47 journal articles, 1 encyclopedia lemma and 88 papers in international conference/workshop proceedings). Google scholar reports more than 15.000 citations and an h-index of 45.