



Michal Malyska


Machine Learning Team Lead,
Semantic Health

 <https://michalmalyska.com>
 malyskamichal@gmail.com

Links

 [My Github](#)
 [Academic Website](#)
 [LinkedIn](#)





Programming

 **Python** ● ● ● ● ●
 PyTorch, TensorFlow, PyMC3
 SpaCy, NLTK, Transformers
AllenNLP, Snorkel
 Pandas, Scikit-learn, LGBM
 **R** ● ● ● ● ●
 tidyverse, RSTAN, brms, INLA
 **SQL** ● ● ● ● ●





Machine Learning –

Natural Language Processing
Semi-Supervised Learning
Information Extraction
Statistical Learning
Explainable ML
Knowledge Graphs

Other

 Git, Bash
 Docker, MLFlow, WandB
 Redis, AWS EC2 / S3, Snowflake
 UMLS, SNOMED-CT

Contributions

 MedSpaCy
 AllenNLP
 SciSpaCy
 MS-BERT

Work Experience

January 2021 – Present **Machine Learning Team Lead** Semantic Health
· Responsible for R&D and delivery of company's Machine Learning offerings from idea stage to multiple successful client deployments, across various business lines, which drive a \$60M valuation.
· Manage and coach a team of 5 ML Scientists, including growth goal setting, creating and tracking OKRs, leading 1-1s, and conducting performance reviews.
· Supervised research projects with outside organizations resulting in product improvement and scientific publications

June 2019 – December 2020 **Machine Learning Scientist** Semantic Health
· Assessed, implemented, and improved deep learning algorithms from research and open source repositories to surpass current state-of-the-art performance on several clinical NLP tasks.
· Improved the data pipeline and model evaluation procedures for complicated NLP datasets in an extreme multi-label setting.

January 2018 – August 2018 **Actuarial Analyst** Deloitte
· Deployed client-site on business analytics, loss forecasting, and predictive modelling for a large insurance client.
· Created an actionable overview of internal and external data, and analytical tools for a large public sector client

Education

2019 – 2022 **Master of Science - Statistics** University of Toronto
4.0 GPA

2015 – 2019 **Honours Bachelor of Science - Statistics** University of Toronto
4.0 GPA in Statistics, 3.78 overall

Teaching and Extracurricular

2020, 2022 **Course Instructor** University of Toronto
2022 - Taught STA414 / STA2104 - Statistical Methods of Machine Learning II, a cross listed undergraduate / graduate Machine Learning course for a class of 200.
2020 - Prepared materials and taught STA220 - Practice of Statistics I, to a class of 250 students in an online setting during the pandemic

2020 **Research Visitor** St Michael's Hospital, Toronto

Publications

2022 **ICD BigBird: A Contextual Embedding Model for ICD Code Classification**
2022.bionlp-1.32
ACL BioNLP

2021 **Active learning for medical code assignment**
arXiv:2104.05741
ACM CHIL 2021 workshop track

2020 **Multiple Sclerosis Severity Classification From Clinical Text**
2020.clinicalnlp-1.2
EMNLP 2020 Clinical NLP workshop

Competitions

| | | |
|--------------|--|------------------------------|
| 2017 – 2019 | 1st. Place, Mentor | ASA Datafest |
| | Created a business case from click data for Expedia, aimed at improving the suggestion engine and customer retention, served as a mentor in the subsequent years | |
| October 2017 | 1st. Place | McKinsey Open Data Challenge |
| | Created a business case and an MVP aimed at re-routing low urgency patients to hospitals with lowest estimated combined travel and wait time | |