Data Science Portfolio: https://haruka-takagi-datascience.github.io/
LinkedIn: https://www.linkedin.com/in/haruka-takagi/
Email: haruka.takagi@mail.utoronto.ca
155 Yorkville Ave, Toronto, ON, M5R 1C4
Cell: +1-647-906-0192

Haruka Takagi

EDUCATION

University of Toronto St. George 09/2018 — Present
Honors Bachelor of Science Degree
Double Major in Statistics and Economics, Minor in Mathematics

- 2020 & 2021 Dean's List Scholar
- University of Toronto Excellence Award from Vice-President Research & Innovation (7500 CAD)
- Current CGPA: 3.52/4.0
- Expected Graduation: May 2023

WORK EXPERIENCE

Financial Data Analytics/Machine Learning Co-Author & Researcher 05/2020 — Present
University of Toronto Scarborough, University of Toronto

- Working on a research project with Assistant Professor Dr. Charles Martineau & Assistant Professor Dr. Pat Akey & 2 other professors as a co-author and technical lead of the team.
- The research project aims to examine group decision making dynamics in central banking by extracting agreement/dissent levels from FOMC transcript texts using natural language processing deep learning models. Currently developing domain specific dialogue act classification model to metricize dissent.
- Conducted latent dirichlet allocation topic modeling of federal meeting transcripts & Greenbook of the Federal Reserve Board of Governors.
- Executed clean text extraction from PDF FOMC transcripts, prepared texts for NLP processing (such as stemming and etc) and produced statistics & created visualizations of possible topic model outcomes using Word Clouds.
- Worked on a collaborative research project with Assistant Professor of Finance at Arizona State University Christoph M. Shiller, on a corporate finance research paper on firm reputation redemption. Akey, Pat and Lewellen, Stefan and Liskovich, Inessa and Schiller, Christoph, Hacking Corporate Reputations (January 22, 2021). Rotman School of Management Working Paper No. 3143740, Available at SSRN: https://ssrn.com/abstract=3143740 or http://dx.doi.org/10.2139/ssrn.3143740

Machine Learning & NLP Engineer 01/2022 — Present
Empath Inc, Tokyo

- NLP engineer in a software development team working on a AI equipped meeting storage software that automatically records, transcribes, analyzes, and organizes all online meetings to promote team collaboration.
- Increased accuracy of in-house automatic speech recognition model by automatic text collection for an in-domain corpus used for ASR model fine tuning.
- Implemented an unsupervised topic segmentation model for dialogue based on BERT embeddings with high accuracy results.
- Implemented dialogue topic extraction based off the previously developed topic segmentations.

Macroeconomic Data Analysis/NLP Research Assistant 11/2019 — 05/2022
Munk School of Global Affairs & Public Policy, University of Toronto

- Research Assistant for Associate Professor Dr. Mark S. Manger, applying economic news sentiment to predict central bank intervention utilizing highly supervised natural language processing models developed in R and Python.
- Created a data processing pipeline and performed extensive text preprocessing with big data, utilizing document term matrices and word embeddings.
- Developed a web scraping API to pull text data from World Bank Documents & Reports to ensemble a large dataset.
- Worked on an individual research project supervised by Professor Mark Manger on domain adapted word segmentation methods in Japanese NLP. Utilizing corpus comparison techniques for automatic term recognition with emphasis on a two tier process of mono and multi-term extraction. Corpus comparison techniques will use word embeddings and alignment methods.

HARUKA TAKAGI

WORK EXPERIENCE

Machine Learning Engineer & Data Analyst Intern 05/2020 — 08/2020
Empath Inc, Tokyo

- Team member of automatic machine learning pipeline project. Leader of feature selection machine learning automation development. Assisted & supported other team members involved in automatic feature generation & feature transformation development.
- Individually developed an automatic feature selection method for classification, tailored to csv audio data. Utilized feature methods (such as RFECV, Feature Importance, Mutual Information + Custom combinations of methods). Created auto-generated feature selection wellness reports for clientele use.
- Project leader for KDDI (company client) machine learning model development. Developed data processing methods intended for further machine learning training use. Data consisted of audio data for call center operators and customers. Main development targets was for the final dataset to reflect the dynamic speech changes & reflect the relational information across audio channels.

Analytics & Sales Intern 08/2020 — 09/2020
Bloomberg, Tokyo Office

- Completed training for Bloomberg terminal on FOREX, Equities, Commodities, Buy-side, Sell-side, fixed income, BQL, BQNT etc.
- Increased efficiency in Bloomberg sales outreach by producing a numerical metric that identifies potential Bloomberg terminal clients that may be positively receptive to BQL (Bloomberg SQL) & BQuant sales proposals by utilizing client activity data on the Bloomberg terminal.
- Delivered a Bloomberg terminal tutorial tailored to FOREX traders by utilizing usage data of the most popular terminal functions used by past FOREX traders on the platform.
- Delivered a seminar to the Analytics department on AI, ML & NLP technology to increase awareness and chances of resolution on two departmental efficiency issues. One, delivered a breakdown of the NLP ticket model and highlighted the importance of analytic department employee feedback needed to improve the model and improve efficiency in the department ticketing system. Two, increased efficiency in Bloomberg Event Driven Feed for black box applications sales by delivering a detailed list and characteristics of client questions submitted to the analytics help desk that may indicate a user's receptiveness to Bloomberg black box solutions.

Research Analyst/Undergraduate Researcher 02/2020 —05/2020
Emote AI Inc. & Ryerson University, Department of Psychology

- Former Co-investigator/undergraduate researcher of a research study under PI: Associate Professor Dr. Alexandra Fiocco from Ryerson University
- Worked on experimental design, grant application & creative research collaboration on a study investigating the psychophysiological correlates of professional E-sports performance.

SKILLS & ABILITIES

- Programming fluency in **Python**, **R**, **SQL** & **Stata**. Experience with machine learning packages in **scikit-learn**, **tensor flow**, **PyTorch**, **hugging face transformers**, **nltk** and etc.
- Education experience in traditional statistical data analysis & econometric models.
- Industry & academic research experience in data processing, data analysis & data visualization.
- Research & industry experience utilizing **compute canada HPC** cedar and graham & AWS S3.
- NLP model development experience in dialogue topic segmentation, topic extraction, dialogue act classification, and etc.
- Proficient in Microsoft Word, Excel & Powerpoint.

LANGUAGES

Fully bilingual with native proficiency in **English** and **Japanese**.

- TOEFL 117 out of 120

OTHER

Finalists in 2020 TD Data Hackathon - The Rise of Data
TD Bank, Rotman School of Management's FinHub & TDMDAL

- Employed text preprocessing, NLP methods, feature extraction to create a machine learning model that will predict the net change in a firm's stock price on the day after a quarterly earnings call.
- Our team was selected as one of the top 5 finalists in the hackathon.