Tae Yano

Seattle, WA,

tyyano at gmail.com
tyyano.me
linkedin.com/in/tae-yano-48005162/

SKILL SUMMARY

- **Programming Languages:** C#, C and Python (5+ years); C++ (1+ years); experience in Perl, Matlab, R, HTML, awk and various shells; Experienced with development on UNIX/Linux and Windows
- **ML and NLP:** Fluent in core NLP techniques such as Language Modeling, Topic Modeling, Word Embedding (e.g., Word2Vec, GloVe) and statistical techniques therein (e.g., MLE, MAP, MCMC, Bayesian Networks); fluent both in generative (e.g., Naive Bayes, HMM, LDA) and discriminative (e.g., Logistic Regression, Decision Trees, CRF) approaches; solid understanding of DNN theory and practice; Completed 5-course DNN specialization with Corsera
- **Databases:** Scope and Cosmos DB (4+ years); experience with PL/SQL and PostgreSQL
- **Tools:** Vim, Pandas, NumPy, NLTK, Scikit-Learn, Keras, TensorFlow, Flask, Git, AWS, Visual Studio

EXPERIENCE

Insight Data Science Seattle, WA
Data Science Fellow *2018 - Present*
- Built *Talking Point Sampler*, a tweet aggregator which enables users to interface with political issue ads from across the spectrum; deployed an interactive web app using Flask and Bootstrap on AWS
- Developed a political ad language model, a Kneser-Ney N-gram model, using 66K+ Facebook ads

Microsoft Bing Bellevue, WA
Software Development Engineer *2013 - 2018*
- 5+ years experience in ML+NLP product development for Bing's web text mining team; including data collection, algorithm design, EDA, training, experimentation, feature engineering, and deployment to Terabyte-scale data pipelines
- Trained various information extraction models for local business attributes (e.g, address) using 100+GB of web contents
- Delivered novel iterative structure inference schemes for top impression chain Businesses
- Developed news recommendation features using sentiment analysis, NER, and other NLP techniques

Microsoft Research Redmond, WA
Research Intern *2012*
- Contributed to the development of a novel algorithm to identify and extract entities in documents

Carnegie Mellon University, School of Computer Science Pittsburgh, PA
Graduate Research Assistant *2007 - 2013*
- Statistical Text Analysis, Latent Variable Models, Information Extraction, Probabilistic Topic Models, Text Forecasting
- Developed novel graphical model for text generation in political blogs, trained with 600K+ user comments and posts
- Developed novel text forecasting models for Congressional votes, trained using text from 50K+ congressional bills
- 20+ publications and reports; 10+ conference talks; released 3 public data sets

Columbia University New York, NY
Master of Science Teaching Assistant Fellow *2005 - 2007*
- Courses Taught: Undergraduate Programming, Data Structures, Programing Languages & Compilers

Ricoh Corporation West Caldwell, NJ
System Engineer *2001 - 2005*
- Developed embedded applications and APIs in C/C++ for UNIX based document management devices

EDUCATION

Carnegie Mellon University, School of Computer Science Pittsburgh, PA
Ph.D. in Language and Information Technologies *2007 – 2013*

Columbia University, School of Engineering and Applied Science New York, NY
M.S. in Computer Science *2005 – 2007*

Hunter College New York, NY
B.A. in Computer Science with honors *1998 – 2001*

PATENTS

Identifying salient items in documents: U.S. Patent 9251473 B2 (patents.google.com/patent/US9251473)