MOJTABA HEIDARYSAFA

Data Scientist/ML Engineer

Charlottesville, Virginia

☎ (+1) 703-656-6964 | Email: mh4pk@virginia.edu | Web: heidarysafa.github.io | Github: Heidarysafa | LinkedIn: mojtaba-heidarysafa-25403a33 | **U.S. Permanent Resident**

Skills

Programming	Python, R, Java, HTML5, CSS
Data Science	Python (Pandas, scikit-learn, keras, Tensorflow, pytorch, dash), Spark, Tableau, SQL
Cloud	AWS Technologies (EC2, Cloudwatch, Lambda, databases, sagemaker)
Theory	Machine Learning, statistics, Deep Learning, Natural Language Processing, time-series, optimization
Neural Networks	Multi-layer Perceptron, CNN, RNN, LSTMs, GRU, Transformers, BERT

Work Experience

University of Virginia *Charlottesville, VA*

NLP LECTURER IN DATA SCIENCE SCHOOL *May 2021 - current*

- **Applications of natural language course**
  - Building lectures on topics such as fundamentals of natural language processing, word representation (bag of words, tf-idf, GloVe, word2Vec), text classification, sequence to sequence models for translation, new deep learning architectures for text such as transformers, Bert, GPT.
  - Developing coding assignments for applications of these methods in real-world problems.

University of Virginia *Charlottesville, VA*

RESEARCHER(MACHINE LEARNING, DATA SCIENCE) *January 2017 - Auguest 2021*

- **ANALYSIS OF DATA SCIENCE JOB MARKET**
  - Collected More than 200K job descriptions in the U.S. using cloud scrapping
  - Utilized AWS Lambda, MYSQL, and Watchcloud to process these data continuously.
  - Designed a skill extraction method and deployed the result in a web-based application using python dash framework for visual interaction at `https://dsi-usa2.herokuapp.com`
- **NEW HIERARCHICAL DEEP LEARNING ARCHITECTURE**
  - Collected 50K article abstractions with author keywords from Web of Science
  - Created a public hierarchical database with relevant labels at each level for researchers
  - Developed HDLTex model for hierarchical text classification using tf-idf, GloVe, DNN, RNN and CNN.
  - Improved classification accuracy on overage by 3-5% to 90% over multiple datasets in compare to RNN, and CNN.
- **NEW RANDOM MULTIMODEL DEEP LEARNING MODEL**
  - Designed a model by combining RNN, CNN, and DNN for classification.
  - Improved image classification error rate to 0.18 and 8.79 on MNIST and CIFAR-10 datasets.
  - Improved text classification accuracy to 90.79 and 87.91 on IMDB and 20NewsGroup datasets.
  - Contributed to writing a python package for this model (RMDL) available in PYPI.
- **ANALYSIS OF RAILWAY ACCIDENTS CAUSES**
  - Applied different word embeddings such as Word2Vec, GloVe, and tf-idf for word representation.
  - Investigated RNN, CNN, and DNN classification power using accident narratives with different word embeddings
  - Improved classification to 71% F-1 score using 10 fold cross-validation.

University of Virginia *Charlottesville, VA*

DATA SCIENCE FELLOWSHIP *August 2018 - May 2019*

- **ANALYSIS OF PROPAGANDA: EXTREMIST GROUPS APPROACH TO WOMEN**
  - Led a multi-disciplinary research resulting in two publications in 9 months.
  - Developed methods to collect text from religious and extremist groups targeting women.
  - Applied unsupervised machine learning methods such as Latent Dirichlet Allocation and Non-negative Matrix factorization to extract groups of topics
  - Analyzed emotion components of these document using Depechemood lexicon for emotions

Education _____

University of Virginia   *Charlottesville, VA*
PH.D. CANDIDATE   *August 2016-Present*
- **research area**: Application of machine learning and natural language processing
- **coursework**: Data mining, statistical learning, optimization, stochastic, time-series

Tampere University of Technology   *Tampere, Finland*
MASTER OF SCIENCE IN AUTOMATION   *May 2015*

Certificates _____

Udacity   *online*
NATURAL LANGUAGE PROCESSING NANODEGREE   *August 2021*
- Built a POS tagger using hidden markov models, and a machine translation model using deep LSTM and seq2seq with 93% accuracy.
- Built multiple speech recognition models using Deep Neural Networks.

Datacamp   *online*
PYTHON DATA SCIENTIST   *Sep 2021*
- Data wrangling with pandas, and SQL databases and deploying ML models with scikit-learn

Honors & Awards _____

2021	**Engineering Endowed Fellowship**, UVA School of Engineering	*Charlottesville, VA*
2019	**Best Survey Paper Award**, Journal of information	*Switzerland*
2018	**Presidential Fellowships in Data Science**, 30,000 $ stipend	*University of Virginia*
2018	**Best Presentation**, RMDL presentation in ICISDM	*Florida, U.S.A*

Selected Publication _____

Text classification algorithms: A survey   *Journal of information Vol. 10 Issue 4*

KAMRAN KOWSARI, KIANA JAFARI MEIMANDI, MOJTABA HEIDARYSAFA, SANJANA MENDU, LAURA BARNES, DONALD BROWN   *2019*

Analysis of Railway Accidents' Narratives Using Deep Learning   *17th IEEE ICMLA*
MOJTABA HEIDARYSAFA, KAMRAN KOWSARI, LAURA BARNES, DONALD BROWN   *2018*

An improvement of data classification using random multimodel deep learning (rmdl)   *International Journal of Machine Learning and Computing Vol. 8 Issue 4*

MOJTABA HEIDARYSAFA, KAMRAN KOWSARI, DONALD E BROWN, KIANA JAFARI MEIMANDI, LAURA E BARNES   *2018*

Hdltex: Hierarchical deep learning for text classification   *16th IEEE ICMLA*
KAMRAN KOWSARI, DONALD E BROWN, MOJTABA HEIDARYSAFA, KIANA JAFARI MEIMANDI, MATTHEW S GERBER, LAURA E BARNES   *2017*