MOJTABA HEIDARYSAFA

Data Scientist

Charlottesville, Virginia

 (+1) 703-656-6964 | Email: mh4pk@virginia.edu | Github: mojtaba-Hsafa | LinkedIn: mojtaba-heidarysafa-25403a33

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, 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 scenarios.

University of Virginia   *Charlottesville, VA*
RESEARCHER(MACHINE LEARNING, DATA SCIENCE)   *January 2017 - Auguest 2021*

- **PROJECT : DATA SCIENCE JOB MARKET**
    - Collected More than 200K job descriptions in the US using cloud scrapping
    - Utilized AWS Lambda, MYSQL, and Watchcloud to process these data continuously.
    - Designed a skill extraction method to collect skills from job descriptions.
    - Deployed a web-based application using python dash framework for visual interaction on heroku at `https://dsi-usa2.herokuapp.com`
- **PROJECT : HIERARCHICAL DEEP LEARNING**
    - 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.
- **PROJECT : RANDOM MULTIMODEL DEEP LEARNING**
    - 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.
    - Contribute to writing a python package for this model (RMDL).
- **PROJECT : ANALYSIS OF RAILWAY ACCIDENTS CAUSES**
    - applied different word embeddings such as Word2Vec, Glove, and tf-idf for word representation.
    - investigate RNN, CNN, and DNN classification power using accident narratives and 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*

- **PROJECT : WOMEN IN EXTREMIST GROUPS**
    - 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
    - classified author genders by their text using deep learning

Education

University of Virginia *Charlottesville, VA*
Pʜ.D. Cᴀɴᴅɪᴅᴀᴛᴇ *Since 2016*
- **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*
Mᴀꜱᴛᴇʀ ᴏꜰ Sᴄɪᴇɴᴄᴇ ɪɴ Aᴜᴛᴏᴍᴀᴛɪᴏɴ *May 2015*

Academic Experience

Exploratory text analysis *University of Virginia*
Tᴇᴀᴄʜɪɴɢ Aꜱꜱɪꜱᴛᴀɴᴛ *Spring 2020, spring 2021*
- Supervised student Projects
- Graded weekly python coding assignments

Data mining *University of Virginia*
Tᴇᴀᴄʜɪɴɢ Aꜱꜱɪꜱᴛᴀɴᴛ *Fall 2020*
- Held TA sessions on concepts of data mining and machine learning
- Graded R markdown assignments

Capstone Project on data science *University of Virginia*
Tᴇᴀᴄʜɪɴɢ Aꜱꜱɪꜱᴛᴀɴᴛ *Spring 2016*
- Graded weekly coding assignments on data science project toolkits
- Instructed TA Sessions

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
Kᴀᴍʀᴀɴ Kᴏᴡꜱᴀʀɪ, Kɪᴀɴᴀ Jᴀꜰᴀʀɪ Mᴇɪᴍᴀɴᴅɪ, Mᴏᴊᴛᴀʙᴀ Hᴇɪᴅᴀʀʏꜱᴀꜰᴀ, Sᴀɴᴊᴀɴᴀ Mᴇɴᴅᴜ, *2019*
Lᴀᴜʀᴀ Bᴀʀɴᴇꜱ, Dᴏɴᴀʟᴅ Bʀᴏᴡɴ

Analysis of Railway Accidents' Narratives Using Deep Learning *17th IEEE ICMLA*
Mᴏᴊᴛᴀʙᴀ Hᴇɪᴅᴀʀʏꜱᴀꜰᴀ, Kᴀᴍʀᴀɴ Kᴏᴡꜱᴀʀɪ, Lᴀᴜʀᴀ Bᴀʀɴᴇꜱ, Dᴏɴᴀʟᴅ Bʀᴏᴡɴ *2018*

An improvement of data classification using random multimodel deep learning (rmdl)

Mojtaba Heidarysafa, Kamran Kowsari, Donald E Brown, Kiana Jafari Meimandi, Laura E Barnes

International Journal of Machine Learning and Computing Vol. 8 Issue 4

2018

Hdltex: Hierarchical deep learning for text classification

Kamran Kowsari, Donald E Brown, Mojtaba Heidarysafa, Kiana Jafari Meimandi, Matthew S Gerber, Laura E Barnes

16th IEEE ICMLA

2017