SAHIL WADHWA

(+1)2172086838 ◇ sahilw2@illinois.edu ◇ www.github.com/wadhwasahil ◇ www.linkedin.com/in/wadhwasahil/

EDUCATION

University of Illinois, Urbana-Champaign *Dec '20*
MS in Statistics Current GPA: 4.0/4.0

Jamia Millia Islamia, New Delhi *May '16*
Bachelor of Technology, Computer Science Overall GPA: 9.6/10

WORK EXPERIENCE

Inference Analytics Aug '19 - Dec '20
Natural Language Processing (NLP) Research Intern

- Aspect Based Sentiment Analysis: Created an end to end model to perform aspect-based sentiment analysis using pre-trained encoders and topic modeling techniques **(LDA, Rake)**
- Structured Radiology Reports: Fine-tuned pretrained embeddings of healthcare data using transfer learning to create healthcare narrative cloud to generate structure radiology reports from unstructured clinical text thus saving man hours

Seagate Technologies May '20 - Aug '20
HAMR Write Design Team, Machine Learning Engineer Intern

- Created an interactive data visualization platform that could clean the data, remove outliers if any and produce charts saving time for design owners. Wrote entire code for variability plot (also called grouped plot) in python from scratch
- Developed a web-based Machine Learning pipeline for data segmentation, wrangling and ranking. Used **AIC** and **Lasso** for feature selection and **Linear Regression**, **Gradient Boosting Methods**, **Random Forests** for feature importance. Outperformed previous JMP based models by achieving **92% R-squared** value, an increase in **10%** from previous models

School of Information Sciences UIUC Jan '20 - Present
Graduate Research and Teaching Assistant *Advisor Halil Kilicoglu*

- Implement multi-label classification models for biomedical journals on Randomised Controlled Trials (RCT) to ensure their integrity with CONSORT guidelines.
- Identifcation of cited text spans in scientific literature, using pre-trained encoders (**BERT**) in combination with different neural networks

Blackrock Feb '18 - Aug '19
Financial Modeling Group (FMG), Machine Learning Engineer

- Trained a very large scale distributed word-embedding model similar to **word2vec** for large vocabularies. Used **Ignite** as **parameter server** for asynchronous updates to model parameters across multiple nodes to avoid memory overflow.
- Developed a novel word-level **Entity Linking/Disambiguation** model by incorporating attention mechanisms using **BERT** and **bi-LSTMs**. Surpassed current state-of-the-art strong matching performance(**F1**) on **AIDA** test dataset by **2%**

Scry Analytics June '16 - Sept '17
Data Scientist

- **Named Entity Recognition**: Created an **LSTM** based model to extract information i.e entities from unstructured clinical data using **tensorflow**
- **Relation Extraction**: Developed a **Convolution Neural Network(CNN)** based deep learning model to identify relationships between entities in a text. Achieved **F1** score of **85%** surpassing previous feature-based models by **15%**
- Implemented Big Data pipelines in **Spark** for fast retrieval and processing of data residing in **HBase**. Reduced the pipeline execution time from **3 days** to **6 hours** by switching the entire pipeline from **Map-Reduce** to **Spark**

INDEPENDENT RESEARCH PROJECTS

1. **Shadow detection in images**: Implemented a U-Net based **Conditional Gated Adversarial Network** (GAN) to detect shadows in images by incorporating a sensitivity parameter to regulate the amount of shadow pixels in the predicted shadow map
2. **Event recognition in complex videos using multi-stream CNNs**: Explored fusion techniques for the spatial (static frames) and temporal (stacked optical flow) streams

RESEARCH PUBLICATIONS

1. YELM: End-to-End Contextualized Entity Linking. Asia-Pacific Chapter of the Association for Computational Linguistics AACL 2020. *link*
2. Evaluating the Readability of Force Directed Graph Layouts: A Deep Learning Approach. IEEE Computer Graphics and Applications, (CG&A). *link*

SKILLS

Languages - Python, C++, Java, R, Scala
Frameworks and Databases - Tensorflow, Spark, PyTorch, SQL, Hive, HBase, Hadoop, Docker, Kubernetes