Abdurrahman Munir

mannya897@gmail.com - www.linkedin.com/in/abdurrahman-munir - https://github.com/ManNy897

EDUCATION

May 2019

University of Southern California, Los Angeles, CA

- M.S. in Computer Science (NLP, Machine Learning, Database Systems, Web Technologies, Algorithms)

May 2017

University of Massachusetts, Amherst, MA,

- B.S. in Computer Science

WORK EXPERIENCE

Work Summary

- Developed scalable **distributed systems** and deployed, scaled, and managed **microservices** and **webservices**
- Strong foundation in data structures, algorithms, multi-threaded and asynchronous patterns and their applications in developing scalable systems
- Worked with project managers to evaluate product feature tradeoffs and create the system specifications for developing **cloud applications**
- Strong understanding of aspects of proper service design including operational toil, availability, and scalability
- Created resilient and reliable **cloud based microservices** using a variety of technologies including Git, AWS (Kinesis, Lambda, Dynamo, Sagemaker, S3, EMR, EC2 etc), Spark, CI/CD and more to process **massive datasets**
- Collaborated closely with cross-functional Machine Learning science and engineering teams to understand and solve customer pain points
- **Contributed to fullstack development** of front-end (Angular 2, Javascript, Typescript), middle-ware (Java, Struts 2), and back-end (Java, Scala, AWS, Spark, Springboot, Guice, Dagger, CDK)
- Drove improvement of **engineering excellence** standards through tools, documentations, presentations, discussions, and processes
- Experienced with both Scrum and Kanban Agile methodologies
- Excellent written and verbal communication skills as well as great attention to detail
- Led multiple projects, mentored junior engineers, and guided interns to receiving full time offers
- Wrote clean, readable, testable, maintainable, and extendable code following code review process

June 2019 - Current

Amazon, *Seattle, WA*

Software Development Engineer

- Key developer for design and development of the Universal Feature Catalog which was a purpose built repository that acted as the authority for vending **Machine Learning features** across different teams in Amazon
- Designed and implemented a new scalable data store snapshotting solution capable of processing **9000 TB** per day resulting in a **60% cost reduction** from the old system
- Performed trade off between distributed data processing platforms like Glue, EMR and Amazon internal platforms and different data storage options like Apache Hudi and a custom partitioning scheme to design a robust, scalable, and cost effective data store snapshotting solution
- Experience debugging Spark/Yarn jobs
- Designed and implemented a generic auto re-drive tool which was used in various services in order to automatically re-drive failures, therefore **greatly reducing operational overhead**
- Utilized ScalaPB library to serialize data into Protocol Buffers and vend to customers in de-serialized format to address customer pain-point
- Improved re-drive mechanism of Kinesis stream processor in order to successfully process records at **+15K TPS**

June - August 2019

Adobe Research, *San Jose, CA*

Video Analytics Research Intern

- Collaborated on an automatic **video summarization** microservice using a Generative Adversarial Network
- Optimized Python code written with Keras and Tensorflow to scale and run **50% faster** using GPU clusters, threading, parallelization, and code refactoring
- Deployed service on the Adobe Cloud Platform and Adobe Sensei Platform as REST API
- Configured service for deployment using Docker on both Azure andAWS

June - August 2017

Adobe, *San Jose, CA*

Machine Learning Research Intern

- Evaluated many existing matrix imputation models by observing their performance on a variety of datasets in R and writing reports comparing their performances
- Using scikit-learn, invented a categorical to numeric variable transformation method using K-means clustering, which was used in matrix imputation of high cardinality categorical variables

ACADEMIC RESEARCH

Information Extraction and Synthesis Lab, *UMass Amherst*

Undergraduate Researcher

- Performed query name expansion using Wikipedia in order to increase potential results in question answering system resulting in a 2% increase in F1 score

PUBLICATIONS

- *Standardization of Featureless Variables for Machine Learning Models Using Natural Language Processing*. [Modarresi et al.2018] Kourosh Modarresi, Abdurrahman Munir. 2018. ICCS. 2018: 234-246..
- *Generalized Variable Conversion using K-means Clustering and Web Scrapin**g***. [Modarresi et al.2018] Kourosh Modarresi, Abdurrahman Munir. 2018. ICCS. 2018: 247-258.
- *Extracting Multilingual Relations under Limited Resources*: TAC 2016 Cold-Start KB construction and Slot-Filling using Compositional Universal Schema. [Chang et al.2016] Haw-Shiuan Chang, Abdurrahman Munir, Ao Liu, Johnny Tian-Zheng, Aaron Traylor, Ajay Nagesh, Nicholas Monath, Patrick Verga, Emma Strubell, and Andrew McCallum. 2016.

PATENTS

- Categorical Data Transformation and Clustering for Machine Learning using Data Repository Systems P7569-US
- Categorical Data Transformation and Clustering for Machine Learning using Natural Language Processing P7349-US

SKILLS

- Java, Scala, Python, Spark, Apache Hudi, R, JavaScript, C++, Angular 2+, Keras (TensorFlow), Protocol Buffers
- AWS Technologies: Kinesis, Glue, EMR, Lambda, SQS, S3, CloudFormation, SageMaker, CloudWatch, EMR