

Contact

8609909724 (Mobile)
mohammedayub44@gmail.com

www.linkedin.com/in/
mohammedayub1 (LinkedIn)

Top Skills

Python
Machine Learning
Natural Language Processing

Languages

Hindi
Urdu

Certifications

Neural Networks and Deep Learning
Intro to Statistics with R: Introduction
Introduction to R
Statistics One (Princeton University)
w/o Certificate
Sequence Models

Publications

Massachusetts' Technology, Talent
and Economic Reporting System -
MATTERS

Mohammed Ayub

Data Scientist II at NFPA
Quincy

Summary

I am a computer science engineer specializing in predictive analytics applications. Presently working in fire and life safety research to design and implement data science solutions, analytical processes and encourage data driven decisions within the organization.

Previously trained in business intelligence and data warehousing tools where I built data processing pipelines and KPI dashboards for big data and business teams in IT organizations.

Interests: Deep Learning, Natural Language Processing, Distributed Frameworks

Experience

NFPA

4 years 11 months

Data Scientist II , Data Analytics, NFPA Global

April 2019 - Present (1 year 10 months)

Quincy, Massachusetts, United States

Neural Machine Translation Tool using Fairseq, UnsupervisedMT and OpenNMT -

Co-developed with a team of data science interns a language translation service on NFPA Fire data (English-Spanish pair). Achieved avg Test BLEU score of ~73.0. Also, experimented with unsupervised techniques for multilingual models using monolingual data.

Techniques Used - Word Segmentation (Byte-Pair Encoding and SentencePiece), Embeddings (Glove, FastText), Model Tuning (Transformer), Domain Adaption, TensorFlow Serving

Created a prototype ML web application and deployed it on AWS ECS service for beta users.

Spatial Fire Analysis using Geo-coding Services on Distributed Frameworks -

Lead a team of interns to build spatial tools for fire analysis using applications like Hadoop, Spark and Hive.

Created Geo-coding service that can geo-code about 26 million address records under 8 hours. Bench-marked performance and quality against commercial (Google & Pitney Bowes) and opensource versions (Pelias).

Used data from TIGER Shapefiles and services such as Libpostal, Lucene Indexing and Spark-on-YARN.

Data Scientist, Research Data and Analytics

March 2016 - April 2019 (3 years 2 months)

Quincy, Massachusetts

Setup of Big Data platform (CDH) for ingesting various fire data sources, not limited to National Fire Incident Reporting System (NFIRS), Fire Service Inventory (FSI) and other internal CRM data sources.

Support in development of machine learning products to tackle pressing issues related Fire Community Risk and Reduction, Fire Department Similarity and Building Inspections for various fire departments across the United States.

Proactively drive new approaches and techniques for expanding the big data architecture with other Apache services (Spark, Kafka, Flume) and software efficiency for the analytics sandbox environment.

Oversee activities of data science interns, ensuring execution deadlines and alignment with business vision and initiatives.

IOMICS Corporation

Data Scientist

September 2015 - March 2016 (7 months)

Worcester, MA

- Evaluated semantic annotation tools such as IBM Watson services and Ontotext for their performance on biochemical data sets; visualized and queried results using the Neo4j graph capabilities.

- Built web application to support research chemist and biologist using Django Framework along with python's Natural Language Tool Kit (NLTK), Scikit-learn and gensim modules to analyze and classify the assay procedures stored in PubChem (database of chemical molecules and their activities)

against biological assays). Implemented techniques such LSI (Latent Semantic Indexing) and LDA (Latent Dirichlet Allocation) to generate concept modelling.

EMC

Analytics Engineering Intern

January 2015 - August 2015 (8 months)

Hopkinton, MA

As part of engineering support team for mid-range storage products, developed data processing pipelines and visualization dashboards to effectively track code development and deployment issues for all storage teams.

Used Microsoft SSIS workflows and SQL Server to pull huge amounts of data into warehouse and optimize query execution plan to reduce runtime for efficiency.

Designed Tableau dashboards for data analysis and setup alerts using scripts for automatic emails to business users.

Tools Used: Microsoft SSIS, Tableau, Alteryx, SQLServer

Dell

Technical Analysis Senior Associate

July 2012 - December 2013 (1 year 6 months)

Worked on Dell data warehouse and ETL architecture as part of Quality enablement team.

Communicate with cross functional development teams for code deployment to production environment.

Support large data movements across environments using Teradata parallel load utility.

Sourcebits Digital

Game Designer Intern

January 2012 - May 2012 (5 months)

Bangalore, India

Worked as a Game Designer on a flight simulator game using Unity3D and Blender.

Education

Worcester Polytechnic Institute

Master's Degree, Data Science · (2014 - 2015)

Nitte Meenakshi Institute of Technology

Bachelor of Engineering (B.E.), Computer Science · (2008 - 2012)