Umang Mehta

Data Scientist & **AI Research Engineer** with **2.5 years** of active research & development experience in **Machine Learning (ML)**, **Neural Networks**, **Natural Language Processing (NLP)**, **Social Media Mining**, **Conversational AI**, **Information Retrieval (IR)**, **Knowledge Graph**, **Predictive Analytics**, **Big Data Engineering, Python & Scala** and **5 years** of professional experience as **Full-Stack Software Engineer** & **Team Mentor** with expertise in **Database, Java** & **Web Development**.

WORK EXPERIENCE

Capital One, Chicago, IL — *Data Engineer (Contractor)*

JANUARY 2020 - PRESENT

- Developed Data Processing Pipeline using Spark-Scala to process Credit Card Requests joining data from multiple sources & APIs and deliver the final data to Embossing Vendor
- This cloud-based system replaced the legacy mainframe system improving efficiency by 70%
- Built Data Processing Framework for Analytics using Spark & Snowflake for Anti-Money Laundering & Fraud Detection.

Hello Nesh Inc. (Nesh), Houston, TX — *Data Scientist - NLP & AI*

JUNE 2019 - OCTOBER 2019

- Developed Conversational AI agent in Python for Oil & Gas domain
- Built Knowledge Extraction pipeline with Semantic Extraction & Topic Modeling and representation into Knowledge Graph built using Dgraph & GraphQL.
- Extracted knowledge from 100k+public documents into 1M+ entities & topics.
- Trained & deployed text classifiers with ELMo, BERT, LSTM, CNN using TensorFlow, Keras on AWS Sagemaker
- These modules replaced existing disjoint processes to save more than 90% of manual effort.
- Built PoC of Oil Well Failure Diagnosis & Prediction using EM, MLE & MAP

Kelley School of Business at Indiana University, Bloomington, IN — *Graduate Research Assistant under Prof. Matthew Josefy*

JUNE 2018 - MAY 2019

- Researched & implemented NLP methods to extract relevant information from 50k+ SEC Filings using Stanford CoreNLP & Spacy in Python
- Developed models using ML & NLP to analyze the business model and board leadership structure of companies with Text Classifiers using NLTK & SciKit-Learn with 90% accuracy measured with 10-Fold Cross-Validation

AriaData Inc. (Aridat), Bloomington, IN — *NLP Research Engineer (Volunteer)*

FEB 2018 - MAY 2019

- Built an analytics engine to determine the critical reception of an artists work based on social media using Sentiment Analysis and Opinion Mining in Python
- Developed a Data Preprocessing module using NLTK involving multiple NLU steps & Sentiment Scoring for feature engineering
- Built a Tweet Sentiment Classifier using an ensemble of Naïve Bayes & Logistic Regression with 87.89% accuracy

Vitruvian Technologies Pvt. Ltd. (Realty Redefined), Mumbai, India — *Senior Developer & Team Mentor*

JUNE 2014 - JULY 2017

- Collaborated in a team of 12 & mentored a sub-team of 5 for full-stack development including my contribution of 100+Java & Scala functionalities, 20+ new Data Structures, Database Design in MySQL with 50+ stored procedures & 200+ SQL queries and Web Design of 30+ pages for web-based ERP & CRM systems in Real Estate domain
- Overall my work translated to a 50% increase in the business efficiency of the end-users

EDUCATION

Master's Degree in Data Science — *Indiana University Bloomington, USA*

AUGUST 2017 - MAY 2019

Major Courses: Advanced NLP, Machine Learning, Data Mining, Social Media Mining, Elements of AI, Deep Learning Systems, Introduction to Statistical Modeling. GPA: 3.6

Bachelor's Degree in Computer Engineering — *K J Somaiya College of Engineering(KJSCE), Mumbai University, India (GPA: Not on a 4.0 Scale)*

AUGUST 2009 - MAY 2013

PERSONAL INFORMATION

Cell: +1 (812) 955-9507
Email: umangrmehta@gmail.com
umangrmehta.github.io
linkedin.com/in/umangrmehta
github.com/umangrmehta

TECHNICAL SKILLS & CERTIFICATION

Programming Languages: Java, Python, Scala, Javascript, Groovy, C/C++, R, GraphQL, Gremlin.

Database-RDBMS: MySQL, PostgreSQL, Oracle.

Database-NoSQL: MongoDB, Dgraph, Neo4J, JanusGraph, Apache Solr, DynamoDB.

Data, ML & AI: TensorFlow, Keras, Spark ML, DL4J, ND4J, Scikit-Learn, NumPy, SciPy, Pandas, Plotly.

NLP: Spark NLP, BERT, ELMo, Stanford CoreNLP, Rasa, NLTK, Spacy, OpenNLP, Polyglot, Gensim, Flair.

Big Data: Apache Hadoop, Apache Spark, Apache Kafka, Snowflake.

Middleware Frameworks: Spring, Hibernate, Flask, Django, NodeJS, Netty.

Deployment: Docker, Kubernetes, AWS EC2, AWS Lambda, AWS Sagemaker, AWS S3, AWS EMR, Google Compute Engine.

OS & Cloud Platforms: Linux, Windows, Amazon Web Services (AWS), Google Cloud Platform.

Project Management: Git, Jira, SCRUM.

Certification: Oracle Certified Professional JavaSE6 Programmer (OCPJP).

EXTRA-CURRICULAR

President & Treasurer of Data Science Club at Indiana University in 2018-19.

Organized a hackathon in collaboration with SPEA at Indiana University for analyzing the opioid crisis.

Creative Head of Students' Council 2011-12 in KJSCE for organizing Technical and Cultural Festivals.

Won 2 consecutive **Annual Fall Chili Cook-offs** in 2017 and 2018 organized by SICE at Indiana University.

Author & Editor of a review blog The Minimalist Critic on Medium.

Published my poetic verses in Kshitij, the Annual College Magazine of KJSCE.