# Umang Mehta

**Machine Learning Engineer** & **Data Scientist** with **4.5 years** of active research & development experience in **Machine Learning (ML), Neural Networks, Natural Language Processing (NLP)**, **MLOps, Big Data Engineering, Social Media Mining, Conversational AI**, **Information Extraction (IE), Knowledge Graph**, **Predictive Analytics** with **Python & Scala** and **5 years** of professional experience as **Full-Stack Software Engineer** & **Team Mentor** with expertise in **Database** & **Java**.

## Work Experience

**Audible - An Amazon Company,** Austin, TX — *SDE -Machine Learning Engineering (Full-Time)*
NOVEMBER 2021 - APRIL 2023

- Developed workflows using AWS components like Sagemaker, Lambda, Batch, Glue, EMR & Step Functions as building blocks for creating pipelines with stages for training, validation, scoring, publishing, load testing & deployment to production for Machine Learning Models
- Worked on the Frontend and Backend for the platform used for hosting Machine Learning pipelines using the aforementioned building blocks
- Used AWS Cloud Development Kit (CDK) to deploy and manage Cloud Resources on AWS
- Worked with Data Scientists to make the model code production ready and deploy them as containers to AWS ECS repository for AWS Sagemaker to run Training and Inferencing
- Overall my work on the Machine Learning Platform resulted in reducing the engineering efforts of Data Scientists by 75%

**Capital One,** Chicago, IL — *Data Engineer (Contractor)*
JANUARY 2020 - OCTOBER 2021

- Developed Data Processing Pipeline using Spark-Scala to process Credit Card Requests joining data from multiple sources & APIs and delivering 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 (Full-Time)*
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 text classifiers with ELMo, BERT, LSTM, CNN using TensorFlow, Keras & Sagemaker
- These modules replaced existing disjoint processes to save more than 90% of manual effort.

**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 business model and board leadership structure of companies with Text Classifiers using NLTK & SciKit-Learn with 90% validation accuracy

**AriaData Inc. (Aridat),** Bloomington, IN — *NLP Research Engineer (Volunteer)*
FEB 2018 - MAY 2019

- Developed a Data Preprocessing module using NLTK involving multiple NLU steps & Sentiment Scoring for feature engineering for Sentiment Analysis & Opinion Mining
- Built ensemble Sentiment Classifier using Naïve Bayes & Logistic Regression with 87.89% acc.

## Independent Projects

**Sentiment Classification** — *Using ELMo, BERT, LSTM, CNN in TensorFlow*
- https://github.com/umangrmehta/tf-text-classifier

## 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

**Bachelor's Degree in Computer Engineering** — *K J Somaiya College of Engineering(KJSCE), Mumbai University, India*
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, C/C++, R, GraphQL.

**Database-RDBMS:** MySQL, PostgreSQL.

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

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

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

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

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

**Deployment:** Docker, Kubernetes, AWS ECS, AWS CDK, AWS CloudFormation, AWS EC2, AWS Lambda, AWS S3, AWS Batch.

**Data & ML Ops:** AWS Sagemaker, AWS Step Function, AWS EMR, AWS Glue, Databricks, MLFlow, Airflow.

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

**Project Management:** Git, Jira, SCRUM.

**Certification:** Oracle Certified Professional JavaSE6 Programmer (OCPJP), DeepLearning.AI TensorFlow Developer.

## Extra-Curricular

**President & Treasurer** of the 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 consecutive **Annual Fall Chili Cook-offs** in 2017 & 2018 organized 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.