Ankush Jain

Cupertino, CA • ankushjain2001@yahoo.co.in • ankushjain2001.github.io

EDUCATION

New York University, NY, USA   **Sep 2019 – May 2021**

M.S. Computer Science • **GPA**: 3.8 • **Coursework**: Algorithms, Big Data, Cloud Computing (AWS), Deep Learning, Machine Learning & NLP

Dr. APJ Abdul Kalam Technical University, Lucknow, IND   **Aug 2013 – Jun 2017**

B.Tech. Computer Science • **Aggregate**: 74.5% • **Coursework**: Data Structures, DBMS, Distributed Systems, Networks, OS, Software Engg & Testing

TECHNICAL SKILLS

- **Languages**: Python, JavaScript, TypeScript, Java, C/C++, HTML, CSS
- **Frameworks & Libraries**: Node, FastAPI, Flask, React, Angular, Express, JQuery, Spark, PyTorch, TensorFlow/Keras, Boto3, Bootstrap, D3.JS
- **Databases:** Oracle, MS SQL Server, MongoDB, Elasticsearch, DynamoDB
- **Cloud Platforms**: Amazon Web Services (S3, Lambda, EC2 etc.)
- **Platforms**: Linux, Unix, Windows
- **Tools**: Git, Docker, Jupyter, Spyder, Anaconda, Travis CI, NPM

EXPERIENCE

Graduate Research Assistant • NYU Tandon, NY [LINK] *PyTorch, Tensorflow, Jupyter, Python (Pandas, SciKit-Learn)*   **Jan 2021 – May 2021**

- Created data pipelines to process, visualize, and model the vast amounts of data produced by the NYU Tandon Bridge Program
- Built a BERT based topic model for network analysis of the courses offered in the program and conceptualizing its use-cases

Product Lead (Part-Time) • Vyorius, Delhi, IND [LINK] [LIVE] *Node.js, React, D3.JS, Amazon Web Services (EC2, S3, DynamoDB)* **Mar 2020 – Feb 2021**

- Deployed Vyorius (a platform for operating unmanned robots) on AWS EC2 and built AI dashboards for robot data analytics
- Converted the monolith app to microservices architecture using DynamoDB, Cognito, SNS, SQS, Code Pipeline etc.

Graduate Research Assistant • NYU Centre for Data Science, NY [LINK] *Python (IBM AIF360, LIME, SHAP)*   **Jan 2020 – May 2020**

- Developed an open-source Python library called FairPrep to design and evaluate machine learning models for social fairness and enable the users to model human demographics data without any bias (based on race, gender, orientation etc.)

System Engineer • Infosys Limited, Mysore, IND   **Oct 2017 – May 2019**

- **▪ *Infosys AI-Led Quality Assurance Platform*** [LINK] *Node.js, Angular, MongoDB, D3.JS, Python (NumPy, Pandas, NLTK, SciKit-Learn, SpaCy, Matplotlib)*
  - Deployed ML solutions to full-stack application (MEAN) and designed D3.JS visualizations for SQA processes such as test-suite optimization, model evaluation tool, requirement-testcase traceability, and defect analysis
  - Built the NLP & ML model for the Test Suite Optimizer to identify up to 30% redundancies and similarities in regression tests
  - Platform being used by leading Infosys projects in Financial Services and Energy domain
- **▪ *Infosys DevOps Testing Platform*** [LINK] *Node.js, Express.js, D3.JS, Elasticsearch, Kibana, Python (NumPy, Pandas, SciKit-Learn, Matplotlib)*
  - Built a DevOps software quality dashboard to analyze the data from test cycles and enable project managers to make data driven decisions
  - Engineered the data pipelines to redirect the test logs from data source to Elasticsearch via Express RESTful APIs
  - Blended data with ML regression models and Kibana for analysis and visualization

PROJECTS

SecuroDrive [LINK] [DEMO] *AWS (Lambda, API Gateway, SageMaker, Cognito, Rekognition, Kinesis), Java, JQuery, Elasticsearch, Kibana*   **Sep 2020 – Dec 2020**

- Built a driving assistant to avoid road accidents by applying AI on live video stream and issuing alerts when the driver is drowsy
- Developed serverless API backend with Lambdas, implemented user auth with Cognito and Rekognition, trained and deployed the AI model with SageMaker, and produced real-time data analytics visualizations with Elasticsearch and Kibana

Multi-hop Question & Answering [LINK] *Python (PyTorch, Transformers, NumPy, Pandas, SciKit-Learn)*   **Sep 2020 – Dec 2020**

- Implemented a question and answering model for multi-hop questions that requires logical inference or aggregation of information from various parts of the information text (like referring multiple wikis to answer a question)
- Trained a transformer-based Longformer model as it is better suited for long-text inputs (unlike BERT)

AI Email Search Engine (Ascend) [DEMO] *FastAPI, React, MongoDB, Python (Tensorflow, Transformers, NLTK, SciKit-Learn)*   **Jun 2020 – Aug 2020**

- Developed an end-to-end natural text search engine for business teams to mine business information, documents, contacts, and meetings from the mailbox data of all the project team member (solo project inspired by COVID-19 remote work policies)
- Deployed BERT model for question and answering on FastAPI Python backend server and built the frontend with React

Big Data Profiling & Quality Analysis [LINK] *Apache Spark, Hadoop, SparkSQL, MapReduce, Python (Matplotlib, NLTK, SpaCy)*   **Sep 2019 – Dec 2019**

- Built a high-performance Spark pipeline to ingest 1900 New York City open big data sets in order to emulate real-world data volumes, handle data noise, and perform semantic profiling of the columns
- Classified column types like address, name, organization, vehicle, color etc. with 72.4% avg. accuracy and up to 95% precision

ACCOMPLISHMENTS & VOLUNTEER EXPERIENCE

- Among top 10% business ideas at Y Combinator Winter'20 (for project Ascend)   **Sep 2020**
- Taught underprivileged kids at 'Teach for Mysuru', an initiative by Infosys Mysore   **Oct 2018 – May 2019**
- Rated 'Outstanding' (top 5%) at Infosys in every annual appraisal from fiscal years 2017-19

linkedin.com/in/ankushjain2001   github.com/ankushjain2001