ADITYA JAIN

Sunnyvale, CA, USA – 94086 |+1 (323) 986-9260 | adityajn105@gmail.com | LinkedIn**adityajn105** | GitHub**adityajn105 | https://adityajain.me**

TECHNICAL SKILLS

Languages: Python, Scala, Java
Machine Learning: Tensorflow, PyTorch, NumPy, Scikit Learn, Pandas, NLTK
Databases: MySQL, MongoDb, Redis, Apache Iceberg
Data Engineering: Spark, Airflow, Orchestration, Tableau, MapReduce
Other Tools & Libraries: Git, Flask, Docker, Salesforce, Jenkins, Argus

WORK EXPERIENCE

Salesforce.com, San Francisco, CA - Software Engineer (MTS) — May 2023 - Present

Search Analytics - **Independently designed** and **led the initial development** of Search Analytics, a paid feature enabling the flow of Search queries, results, and interactions (**~10m rows per org**) from internal Apache Iceberg to customers' Data Cloud. Utilized technologies such as **Java, Python, Airflow, orchestration, and Spark** to implement this feature. Secured adoption by **over 200+ customers** during the pilot phase. Provided **guidance and onboarding support** to **4 team members** for contributing to the project.

Cavalry - **Developed and enhanced multiple applications** within Cavalry, a Search data and metric platform. Designed the App Schema Manager to automatically update Iceberg schemas before app execution, improving platform efficiency by **deprecating 1 step from app flow**. Built apps to compute performance and adoption metrics and created leadership reports. Developed a partition expiration job using Iceberg table metadata to delete older partitions in dev environments, **reducing Cavalry's AWS costs by 80%**. **Mentored and onboarded 2 interns** for various Cavalry projects. Utilized technologies such as **Scala, Java, Spark, SQL, and Airflow**.

Entity Prediction - Enhanced **Salesforce open-source ml4ir** by improving the performance of classification tasks and enabling support for larger training/validation datasets. Trained an entity prediction model using **Tensorflow** to map queries to Salesforce entities on a dataset **30 times larger, increasing ep accuracy @3** on test datasets from **91.7% in baseline to 94.5%**. Developed a detailed model card and delivered multiple presentations to senior leadership, advocating for **AB testing** of the model.

Trust - As an **Epic owner for Trust** across multiple releases, **onboarded the team** to Slack and PagerDuty **alerts** for failed pipelines, significantly enhancing on-call productivity. Developed **2 FIT tests** to ensure pipelines reliability and smooth operations. Served as a **subject matter expert for Search Analytics**, creating runbooks to assist the team in efficiently managing on-call responsibilities.

Salesforce.com, San Francisco, CA - Software Engineer Intern — May 2022 - Aug 2022

Anomaly Detection - Designed and implemented an automated monitoring system to track Einstein Search metrics and detect anomalies with a **0.98 F1 Score**. Utilized technologies such as **PySpark, Python, Pandas,** etc to build the framework. Developed a Salesforce Hawking orchestration pipeline to schedule framework runs for over **1k organizations in production**.

USC Institute of Creative Technologies, Los Angeles, CA - Research Assistant — Feb 2021 – May 2022

Opentutor - Enhanced **grading accuracy to 85%** in Opentutor, an interactive dialog-based tutoring platform, by implementing an ML pipeline with a clustering-based solution to address cold-start issues. Developed multiple **unit tests** to prevent regression problems. Utilized technologies such as **Python, scikit-learn, numpy, pandas, Tensorflow, and NLTK** to build the grading system.

Cognizant Technology Solutions, Bengaluru, India - Associate Projects (Data Science) — Sep 2018 – Jan 2021

Search-ad click prediction **- Increased click-through rate by 10%** (from a baseline CTR of 67% with BM25) for an ad-search engine by implementing an NLP pipeline for text preprocessing and deploying a CLSM model via TensorFlow Serving with a gRPC interface, achieving a **5 ms query response time**.

Medicare Star Analytics - Built **analytical and ML models** achieving a **0.75 AUC score** to predict patient health decline, enabling targeted outreach programs to improve Medicare plan STAR ratings. Utilized technologies such as **scikit-learn, pandas, and numpy**.

EDUCATION

University of Southern California, Los Angeles, USA — Jan 2021 – Dec 2022

Master of Science in Computer Science — **GPA– 4.0 / 4.0 (Honors)**
Relevant Coursework – Design and Analysis of Algorithms, AI, ML for games, Applied NLP, Data Mining
Teaching Assistant - CSCI - 544 (Applied Natural Language Processing) in Fall 2022

Maharashtra Institute of Technology, Pune, India — Aug 2014 – May 2018

Bachelor of Engineering in Computer Engineering — **GPA – 3.6 /4.0**
Relevant Coursework – Data Structures, Operating System, Computer Networks, Data Warehousing, Cloud Computing

HONORS AND AWARDS

Selected in **MS CS Honors program** at USC| **First Runner Up** in 'Smart India Hackathon 2017' organized by Government of India among 250 teams | **Organizer and Lecturer** in a national level event called "Linuxication" at MIT, Pune during 2016 and 2017 | Among **top 8 finalists** from 100 participants in Infosys Techzooka '16.