

Aditya Singh

Award winning Data Scientist with enterprise/startup experience and a research background

☎ 878-999-6570 ✉ adiisingh1099@gmail.com 🌐 in/adityas5 🧑 [personal website](#)

EDUCATION

Carnegie Mellon University

Master of Information Systems Management with Highest Distinction (ML & NLP)

Pittsburgh, PA

Aug 2020 - Dec 2022

Coursework: Machine Learning, ML in Production, Deep Learning, Distributed Systems, Databases, Statistics, Data Mining

Ashoka University

Bachelor of Science (Honors) in Computer Science

New Delhi, India

Aug 2016 - May 2020

EXPERIENCE

Thermo Fisher Scientific

Senior Data Scientist, Intelligence Generation, Corporate [*Pyspark, Machine Learning*]

Pittsburgh, PA

Jan 2023 - Present

B2B Recommendation Engine

- Generated ~\$8MM in revenue and improved conversion rate by **45%** by developing a contact ranking model to help the sales representatives target the right customers for B2B account-product recommendations
- Achieved > **90%** coverage for recommendations by mining contact attribute data to develop a collaborative filtering approach
- Pioneered a paradigm shift in model development by developing a package to backtest model performance, promoting benchmarking models. Conducted backtesting to establish baselines and retro studies for feature selection

GenAI and MLOps

- Developed a RAG-based AI agent leveraging open source LLM (Llama 3) to answer questions about divisional models from Confluence pages, improving the response time for stakeholders by ~**90%**
- Managed junior data scientists, delivered division-wide symposiums, and supported hypothesis testing for 2 product teams
- Worked with stakeholders from multiple divisions, optimized jobs for compute time by over **50%**, set up MLflow tracking, anomaly detection, and measurement dashboards, and responded to stakeholder requests for bug fixes & enhancements

Verusen.AI

Data Scientist Intern [*Pytorch, scikit-learn, nltk*]

Atlanta, GA

Jun 2022 - Sep 2022

- Developed a model to enhance the quality of MRO material descriptions by detecting and removing truncated entries
- Drove a **32%** reduction in error rate for data harmonization services, saving an estimated ~\$2MM per client annually
- Trained & Regularized LSTM language model and engineered token features into a classifier, achieving **84%** precision
- Designed NLP pipeline, conducted EDA, error analysis, and production modeling. The project has been applied for a **patent**

PROJECTS, PUBLICATIONS & TEACHING

3CAI4H Health Bowl First Prize, [AIqitas](#)

Jun 2022 - May 2023

- Developed an embedding-based algorithm to match patient data with clinical trial eligibility criteria, improving efficiency in recruitment process by **90%**. Represented CMU in the [3CAI4 Health Bowl](#) competition winning 1st prize of **\$35k**
- Fine-tuned BioBert model on publicly available clinical trial data, devised a matching algorithm to compare phrase embeddings and compute a compatibility score, achieving nDGC of **0.84** on the test set
- Presented research at [AMIA CIC 2023](#) conference

Teaching Assistant for Deep Learning ([11-785](#)) and Machine Learning ([10-601](#)), CMU SCS

May 2022 - Jan 2023

- Delivered lecture on [GANs](#) and [Contrastive Losses](#), held office hours, mentored project groups for **400+** students
- Created new datasets, ran ablations to set benchmarks for CNNs and Transformers. Taught by [Dr. Bhiksha Raj](#)

Intelligent Radar Lab, Ashoka University - Deep Learning Team Lead

Jan 2020 - Jun 2021

- Achieved classification accuracy of **98.3%** through occlusions by training a CNN for object, material, and human detection
- Presented preliminary results in [ASPAI'20](#) & further work in IEEE Radar Conference 2020. Conference - <https://bit.ly/Waldo-IEEE>

SKILLS

Programming Languages & Databases: Python, Java, MySQL, MSSQL, MongoDB, Redshift, Amazon S3

Framework & Libraries: Pytorch, Tensorflow, Pandas, Matplotlib, Numpy, Scikit-learn, Hugging Face(deep learning NLP/NLU library), PySpark, LangChain, MLFlow | **Training:** [UCSD CIML Summer Institute 2020](#)

Tools & Services: Git, Github, Databricks, GCP, AWS, Microsoft Azure, Visual Studio, Agile Development Process (SCRUM)