Tarun Arora
+1 (773)-236-3163 | tarunarora@uchicago.edu | Jersey City, NJ | LinkedIn | GitHub | Portfolio

EDUCATION

Master of Science in Computer Science (Specialization: Data Science), *University of Chicago, Chicago, IL* ***Mar 2024***
Bachelor of Technology in Engineering Physics (Major: Electronics), *Delhi Technological University, India* ***Aug 2022***

PROFESSIONAL EXPERIENCE

Data Analyst, *EXL, Jersey City, NJ* ***Apr 2024- Present***

- Spearheaded comprehensive data analysis and cleansing initiatives on large-scale financial datasets for a globally renowned healthcare company, resulting in a 20% boost in data quality and accuracy, directly enhancing financial decision-making and operational efficiency.
- Developed a robust ETL pipeline in Alteryx for data extraction, cleaning, and loading from SAP tables, enhancing data integration processes, and ensuring the availability of high-quality data for analysis and reporting.
- Engineered and implemented sophisticated machine learning algorithms, including Isolation Forest, to identify outliers and mitigate FCPA compliance risks, achieving a 25% improvement in fraud detection accuracy and safeguarding the company's financial integrity.
- Authored detailed, actionable reports on FCPA risk findings, empowering auditors and key stakeholders with critical insights and fostering a culture of data-driven decision-making that strengthened overall corporate governance.
- Facilitated quick and efficient ad-hoc changes for diverse audits across multiple countries, reducing audit turnaround time by 30% and improving sample selection accuracy by 40%, thereby streamlining the auditing process and ensuring timely and precise compliance assessments.

Machine Learning Research Assistant, *Pritzker School of Molecular Engineering, Chicago, IL* ***Jun 2023- Mar 2024***

- Directed the development of advanced machine learning and deep learning models to revolutionize electrolyte property prediction for cutting-edge battery applications, accelerating electrolyte discovery by 50%.
- Applied Principal Component Analysis (PCA) to extract 208 critical features from chemical formulas, enhancing computational efficiency by 30% while preserving essential information for accurate predictions.
- Conducted extensive hyper-parameter optimization, rigorously evaluating the Chemprop regressor against 5 baseline ML regressors, and achieved an exceptional R2 score of 0.94, demonstrating a 20% improvement in predictive accuracy over existing models.

Data Science Researcher, *Fermi National Accelerator Laboratory, Chicago, IL* ***Mar 2023- May 2023***

- Collaborated within a team of four in pioneering the development of a state-of-the-art deep learning algorithm, achieving unparalleled accuracy in filtering neutrino events, and substantially advancing scientific research capabilities.
- Engineered a novel PyTorch data-loader to efficiently manage H5 files, transforming events into labeled images and optimizing data processing workflows, resulting in a 40% reduction in preprocessing time.
- Implemented a powerful deep learning classifier, yielding a 90% increase in filtration efficiency for neutrino events, thereby enhancing data quality and enabling more precise and reliable scientific analyses.

KEY ACADEMIC PROJECTS

Generative AI-Powered Data Cleaning and Visualization *(Code | Report | Video)* ***Jan 2024- Mar 2024***

- Developed a Langchain agent with specialized tools for cleaning data in uploaded CSV files, ensuring data privacy and rigorously testing effectiveness using a synthetic dataset generated with GPT-4.
- Built a user-friendly frontend in Streamlit, enabling users to upload files, interact through a Generative AI-powered chatbot interface for data cleaning, and providing an additional page for data visualization with AI-driven plot suggestions and code generation.
- Facilitated seamless and secure data cleaning processes, automating plot code and visualization generation with Generative AI, enhancing user engagement, and ensuring the availability of clean, analyzable datasets.

Reinforcement-Learning Based Cryptocurrency Trading Bot *(Code | Report)* ***Sep 2023- Dec 2023***

- Engineered a Recurrent Proximal Policy Optimization model, trained on Q1 and Q2 2023 OHLCVT data with MACD and RSI indicators, achieving a peak reward of $97 on a $1000 investment over 100 episodes during Q3 2023 testing.
- Seamlessly integrated the model with Kraken API and live websockets to access real-time trading data, executing successful live trades that generated a 0.34% return on investment over a 24-hour period, thereby demonstrating the model's practical viability in live trading scenarios.

Function-as-a-Service (FaaS) Platform Development ***Mar 2023- May 2023***

- Designed and implemented a high-performance FaaS platform, MPCSFaaS, using Python, FastAPI, and ZeroMQ, achieving a 90% reduction in execution time with 7 pull workers compared to a single worker.
- Enhanced platform performance through advanced load balancing strategies and robust monitoring, optimizing task execution and resource utilization.
- Conducted rigorous testing and validation, confirming the system's robustness and achieving up to a 50% reduction in execution times, with comprehensive reports demonstrating its effectiveness.

RESEARCH PUBLICATIONS AND AWARDS

- Acclaimed for remarkable contributions to pioneering big data projects with Fermilab. (Article)
- Presented and defended research paper titled "Semiconductor Wafer Map Defect Classification Using Transfer Learning" at IEEE Conference. (Research Paper) Models used: VGG19, MobileNet, ResNet, DenseNet.
- Runner-up at HackBattle 2020 organized by IEEE Computer Society. Project: AI for Sign Language Detection using CNN.

SKILLS

Programming Languages: Python, Java, Haskell, HTML, CSS, SQL, Git, Docker, C, Bash Scripting
Tools/Libraries: Keras, Tensorflow, PyTorch, Matplotlib, Seaborn, Numpy, Pandas, scikit-learn, API, Hadoop
Relevant Coursework: Functional Programming, Algorithms, Probability, Statisitics, Machine Learning, Distributed Systems, Computer Security, Reinforcement Learning, Generative AI, Operating Systems, Computer Networks, Product Management, Cloud Computing using AWS