Tarun Arora

773-236-3163 | tarun.arora2k@gmail.com | linkedin.com/in/tarun-arora-dtu | github.com/tarun2k | tarun2k.github.io

Education

University of Chicago Chicago, IL

Master of Science in Computer Science; Specialization: Data Analytics *Sep 2022 – Mar 2024*

***Course Grader**: Mentored 100+ students in Discrete Mathematics and Machine Learning promoting collaboration and problem-solving*

Delhi Technological University New Delhi, India

Bachelor of Technology in Engineering Physics; Major: Electronics; Minor: Robotics and Intelligent Systems *Aug 2018 – Aug 2022*

Experience

Data Scientist Apr 2024 – Present

EXL *Jersey City, NJ*

- Led automation and analytics for a Fortune 50 client's compliance audits, designing Alteryx-SQL ETL pipelines and 200+ interactive Tableau dashboards to deliver real-time insights, assisting 30+ audit projects annually.
- Combined interactive Tableau dashboards with ML models (Isolation Forest, KNN, SVM) to surface data anomalies and guide auditors, applying hypothesis testing for validation and reducing manual effort in high-risk area selection and sampling by 70%.
- Built and deployed a GenAI audit automation pipeline using AWS Lambda, GPT-4, and RAG agents to standardize and generate compliance-ready reports, integrating UiPath and S3 for seamless data flow, reducing auditors' manual report creation effort by over 60% and accelerating senior leadership review and feedback through a streamlined process.
- Designed a cloud-native system storing AI-generated JSON reports in DynamoDB, enabling real-time editing and dynamic visualization for auditors on the front end, facilitating faster iterations, cutting review cycles by 40%, and improving collaboration efficiency between auditors and senior leadership.
- Conducted product demos and hands-on workshops to upskill auditors on automation tools, ensuring user adoption, proficiency, and smooth integration into audit workflows.

Machine Learning Engineer Jun 2023 – Mar 2024

Pritzker School of Molecular Engineering, The University of Chicago *Chicago, IL*

- Led development of ML models predicting electrolyte properties from SMILES, using RDKit descriptors to improve features and provide new formula candidates to PhD researchers, accelerating their battery testing process by 50%.
- Applied RDKit molecular descriptor extraction combined with Principal Component Analysis (PCA) and feature scaling to select key features, improving model training efficiency by 30% without sacrificing accuracy in predicting electrolyte properties.
- Performed extensive hyper-parameter optimization and conducted comparative evaluation of Chemprop against five baseline machine learning models (XGBoost, Random Forest, LightGBM, Lasso Regression, Support Vector Machine), achieving a high R^2 score of 0.94 and improving predictive accuracy by 20% on regression tasks for molecular properties.

Data Scientist - Student Researcher Mar 2023 – May 2023

Fermi National Accelerator Laboratory *Chicago, IL*

- 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 Convolutional Neural Network (CNN)-based deep learning classifier, increasing filtration efficiency for neutrino events by 45%, thus enhancing data quality and enabling more precise and reliable scientific analyses.

Projects

Generative AI-Powered Data Cleaning and Visualization Jan 2024 - Mar 2024

- Developed a LLM based Langchain AI 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 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.

Skills, Research Publications & Awards

Technical Skills: Python, Java, Haskell, C/C++, HTML, CSS, SQL, Git, Docker, Bash Scripting, Alteryx Tableau, Jenkins, Bitbucket, Snowflake, AWS, Google Cloud, RAG, Keras, Tensorflow, PyTorch, Matplotlib, Seaborn, Numpy, Pandas, scikit-learn, PySpark, Excel
Key Coursework Completed: Algorithms, Probability, Statisitics, Machine Learning, Distributed Systems, Computer Security, Reinforcement Learning, Generative AI, Operating Systems, Computer Networks, Product Management, Natural Language Processing
Research Paper & Article: Semiconductor Wafer Map Defect Classification Using Transfer Learning, Big Data projects with Fermilab
Awards: Runner-up at HackBattle 2020 organized by IEEE Computer Society. Project: AI for Sign Language Detection using CNN