

Kartikay Kaul

+91-7550170675 | kartikaykaul13@gmail.com | [LinkedIn](#) | [Github](#)

Recent CS graduate with hands-on experience in ML lifecycle, data analytics, and automation. Skilled in Python, SQL, Spark, Databricks and Tableau with a strong foundation in data structures, algorithms, and cloud-based data engineering.

Education

University of California, Riverside

Master of Science in Computer Science, GPA: 3.33/4

Riverside, US

09/22 – 06/24

Vellore Institute of Technology

Bachelors of Technology in Computer Science and Engineering, GPA: 8.35/10

Chennai, India

07/16 – 09/20

Experience

Tech Mahindra

Python Developer Intern

04/25 – 07/25

Remote/Plano, Texas

- Developed an **automation bot** to extract PDF reports from Outlook, saving them locally and mapping structured table and text data into Excel sheets using Python (**PyMuPDF**, **Openpyxl**)
- Streamlined **manual reporting workflows** for internal business units, improving operational efficiency and reducing human error
- Gained hands-on experience with **Apache Spark** and **Databricks**, actively preparing for Databricks Data Engineer certification
- Collaborated with **data engineering team** to explore **pipeline orchestration** for downstream data transformation

Pro-Tek Consulting

Computer Programmer (Data & Test Automation)

11/24 – 05/25

Woodland Hills, California

- Developed **Project Vertex**, a Python-based ETL pipeline to transform raw XML data into clean, SQL-ready datasets, applying **data quality (DQ) rules** and validation checks to improve data reliability for analytics and modeling, boosting data accuracy by 30%.
- Created and maintained **automated Pytest frameworks** for testing REST APIs, XML data parsers, and database insertions, improving test coverage by 40% and reducing manual QA effort.
- Integrated test suites into **GitHub Actions CI/CD** workflows, enabling early bug detection and continuous deployment readiness.
- Leveraged **Python scripting and data engineering techniques** to optimize data workflows, ensuring data integrity and supporting predictive analytics initiatives
- Completed **12 weeks of internal technical training**, gaining in-depth exposure to **software development best practices, version control, and system architecture**.

Radiance Technologies

Data Science Intern

08/24 – 10/24

Sheridan, Wyoming

- Built a **machine learning model** for **customer churn prediction**, using **Random Forest** and **Logistic Regression**, which identified high-risk clients with **88% accuracy**.
- Implemented a **time series forecasting pipeline** to predict **monthly revenue trends**, using **ARIMA** models and visualizing forecasts in **Tableau** dashboards.
- Applied **NLP techniques** to analyze customer feedback, creating **sentiment scores** that informed product development and customer engagement strategies.
- Identified, analyzed, and interpreted **trends and patterns** in complex data sets to uncover insights and business opportunities.

Zummit Infolabs

Data Science Intern

01/22 – 06/22

Bangalore, India

- Developed a **Recommendation Engine** using customer behavioral analytics, incorporating **data quality checks** and **governance rules** to ensure clean and actionable data.

- Designed and implemented a **data pre-processing pipeline** for facial feature segmentation using a 2D U-Net model, improving accuracy by 25%
- Collaborated with stakeholders to define and implement Data Quality (DQ) rules, monitored output quality, and refined transformation logic accordingly.
- Built a **GAN-based architecture** for hairstyle recommendation using computer vision, enhancing product recommendation precision.
- Documented experiments and visualized results using **Seaborn** and **Matplotlib**, ensuring reproducibility and stakeholder clarity.

Darwinbox

05/19 – 07/19

Software Development Intern

Hyderabad, India

- **Built a Chrome Extension using JavaScript and Chrome APIs** to automate LinkedIn profile scraping, reducing candidate data collection time by 40%.
- **Implemented a Node.js backend following MVC architecture**, separating concerns for data models (MongoDB), controller logic, and Chrome extension views.
- **Integrated MongoDB as the primary data store**, enabling structured storage and efficient querying of candidate profiles
- Created an **equivalent Python backend** pipeline for data enrichment and batch processing, ensuring modularity and cross-platform compatibility
- Designed and styled interactive UI components using **HTML/CSS (material.io)**, enhancing the recruiter experience.

Projects

Brain MRI Analysis using 2D U-Net Model

05/20

[\[Paper\]](#)

- Implemented a **2D U-Net architecture** in **Keras** to automate **tissue segmentation** in MRI scans, facilitating downstream **medical image analytics**.
- Achieved improved **diagnostic support** by accurately delineating brain structures, aiding in the detection of neurological anomalies
- Employed **data preprocessing techniques** to enhance image quality and model performance

DDOS Detection

12/2018

[\[Github\]](#)

- Engineered features and developed **Random Forest, KNN, and MLP models** to classify **DDOS attacks** on SDN dataset
- Achieved a 15% improvement in detection accuracy through **data-driven security insights**.
- Utilized **data visualization tools** to interpret model results and identify key indicators of network anomalies

Matrix – Linear Algebra Library in C++

01/24 - Present

[\[Github\]](#)

- Developed a **high-performance C++ library** for **matrix operations**, implementing custom **BLAS routines** for optimization.
- Supports **computational analytics workflows** in scientific computing by providing efficient linear algebra computations.
- Integrated functionalities for matrix manipulation, including addition, multiplication, and inversion

Autonomous Driving in GTA5

04/23 – 05/23

[\[Github\]](#)

- Created a **deep learning model** using **Inception v3** to predict keystrokes in real-time, simulating **autonomous driving behavior** in the GTA5 environment.
- Collected and pre-processed large-scale gameplay data to train the model, enhancing **predictive accuracy by 25%**.
- Demonstrated the application of **computer vision techniques** in dynamic, real-world-like scenarios.

Yale3 – LinkedIn Profile Scraping Chrome extension

11/21- Present

[\[Github\]](#)

- Developed a **Chrome extension** to scrape **LinkedIn profiles**, automating **candidate data collection** for **recruitment analytics**.
- Integrated scraped data into a **MongoDB-based pipeline** for **candidate profiling** and **engagement metrics analysis**
- Enhanced the extension with features like **PDF export** and real-time **data enrichment workflows**.

Dirasa – Student-Centric File Sharing Platform

10/18-12/18

[\[Github\]](#)

- Developed a **web-based platform** enabling students to upload, share, and access academic resources, fostering a collaborative learning environment
- Implemented **user authentication**, **file upload/download functionalities**, and a **discussion forum** using **PHP** and **MySQL**.
- Designed responsive **front-end interfaces** with HTML, CSS, and JavaScript to enhance user experience.
- Employed **data management techniques** to organize and retrieve educational materials efficiently

Technical Skills

Languages

Python, SQL, C++, Javascript, Java, Julia, PHP, HTML, CSS, CUDA, GNU Octave, MATLAB

Data Analytics & Visualization

Pandas, NumPy, Tableau, Matplotlib, Seaborn, Excel, Plotly, Data Cleaning & Transformation, Feature Engineering, Feature Engineering, Jupyter Notebook

Databases

MongoDB, NoSQL Databases, SQLAlchemy, PostgreSQL, MySQL

Machine Learning & Deep Learning

Scikit-Learn, Tensorflow, Keras, PyTorch, CNN, U-Net, RNN, Random Forest, KNN, MLP, Model Evaluation (AUC, Confusion Matrix, etc), LSTMs, Transformers, VAEs, BERT, Bi-RNNs, LLMs (learning)

Data Engineering & Cloud

ETL Pipeline Development, AWS (S3, Lambda, Redshift, Glue), Docker, Git, Apache Spark, Databricks, XML Parsing, Apache Airflow

Data Governance & Quality

Data Profiling, Data Quality Rules Translation, Data Lineage, Data Cleansing, SQL-based DQ Checks, Data Stewardship, ETL QA, Great Expectations (Python)

Development & Automation Tools

Flask, RESTful API Development, Chrome API, Github Actions, Jenkins, Pytest, Postman, Selenium

Web Development and Data Scraping

Node.js, Express.js, Scrapy, BeautifulSoup, Material.io, jQuery, Chrome Extension Development, PyMuPDF

Online Certifications

- Databricks Certified Data Engineer Associate [\[Link\]](#)
- Google Chrome Extension Development [\[Link\]](#)
- Deep Learning [\[Link\]](#),
- Machine Learning [\[Link\]](#)
- Julia Programming [\[Link\]](#)
- Full Stack Web Dev [\[Link\]](#)
- Databricks Fundamentals Accreditation [\[Link\]](#)