

# SIA PATODIA

+1(858) 319-6366 | [spatodia@ucsd.edu](mailto:spatodia@ucsd.edu) | [linkedin.com/in/sia-patodia/](https://www.linkedin.com/in/sia-patodia/) | [github.com/siapatodia8](https://github.com/siapatodia8) | San Diego, CA

## EDUCATION

---

University of California – San Diego

La Jolla, CA

*B.S. in Data Science* | *Minor in Mathematics* – Major GPA: 3.886

Sept 2022 – Expected June 2026

**Relevant Coursework:** Scalable Analytics Systems, Machine Learning, Database Management, Business Analytics, Statistics, Probabilistic Modelling, Data Mining, Computer Vision, Data Structures and Algorithms, Cryptography.

**Association for Computing Machinery:** Led a 5-person team developing an NLP urgency classifier, managed a 200+ attendee showcase, enhanced site performance by 35%, and evaluated 40+ applications.

## WORK EXPERIENCE

---

Hey Social Good

San Diego, CA

*Data Analyst Intern*

June 2025 – Present

- Analyzed sustainability disclosures from **500+ retail companies**, applying ESG taxonomy to classify and benchmark firms across three tiers; produced structured datasets and visual summaries used in **2** quarterly **client-facing reports**.
- Identified and resolved **100+ scraper-ETL pipeline issues** (malformed URLs, 404s, HTML parsing errors) via regex-based validation logic; improved data reliability by **25%** and stabilized pipeline uptime.
- Automated preprocessing layer to normalize tabular inputs, remove duplicates, and flag data inconsistencies; reduced preprocessing time by **20%** and manual review by **30%**, improving downstream model performance.

Able Credit

Mumbai, India

*Data Science Intern*

July 2024 – Sept 2024

- Built **multimodal ML workflows** integrating speech transcriptions, product images, and transactional price data to automate credit scoring for MSMEs; improved **ETL throughput by 40%**, enabling faster data refresh cycles.
- Designed **Tableau** dashboards visualizing transaction trends and pricing anomalies for business stakeholders; implemented **SQL-based validation** logic to ensure consistency between source and processed datasets.
- Fine-tuned Whisper ASR model on regional datasets achieving **31.5% lower Word Error Rate**; engineered **PyTorch** object detection modules processing **10K+ images** for financial document classification.

Netlynx Technologies Pvt. Ltd.

Mumbai, India

*Software Engineer Intern*

June 2023 – Aug 2023

- Performed backend and API testing for cloud deployments in an agile team; executed **200+ API tests** via BrowserStack assessing responsiveness, latency, and multi-OS security; **resolved 15+ vulnerabilities**.
- Improved system response time by **30%** (**5s** → **3.5s**) by redesigning UI testing workflow and refining request handling.

## PROJECTS

---

Personalized Supplement Recommender (Deloitte-Mentored Capstone)(GitHub)

Sept 2025 – March 2026

- Developed a **knowledge graph-driven AI system** using **Neo4j** and **LLMs** to deliver personalized, evidence-based supplement recommendations across **330K nodes** and **3.4M relationships**; achieved **81% accuracy**.
- Engineered a modular **multi-agent** pipeline for extraction, normalization, and retrieval with **Cypher query validation**; deployed via **React + Vercel** with JSON logging for full auditability - enabling clinically responsible insights.

Credit Card Fraud Detection (GitHub)

Jan 2025 – March 2025

- Processed **284K** anonymized transactions (**0.172% fraud rate**); addressed class imbalance using **SMOTE** and stratified sampling; trained and cross-validated Logistic Regression, Random Forest, XGBoost, and Isolation Forest models.
- Achieved **0.99 ROC-AUC** and **0.93 PR-AUC**; optimized thresholds to reduce false negatives and boost recall by **18%**.

MBTI Personality Classification (GitHub)

Sept 2023 – Jan 2024

- Fine-tuned **BERT-based NLP classifier** with Hugging Face and TensorFlow; improved **F1-score by 12%** through tokenization optimization and hyperparameter tuning.
- Developed an interactive Streamlit app for real-time personality predictions, achieving **94.8% training** and **87% validation accuracy** on **1K+ user inputs**.

## SKILLS

---

- **Languages:** Python, SQL, R, Java, JavaScript, HTML/CSS
- **Data Science & Analytics:** Predictive Modeling, Big Data Analytics (Pandas, NumPy, Spark), NLP, Power BI, Tableau, TensorFlow, PyTorch, Scikit-Learn, Deep Learning, Transfer Learning, Statistical Analysis
- **Platforms & Tools:** SQL/Database Management, Apache Spark, AWS, Neo4j, LangChain, Git/GitHub, Excel