

Pranav Kumarsubha

(424)-358-0515 | pkumarsubha@gmail.com | github.com/pranavks12345 | linkedin.com/in/pranav-kumarsubha

EDUCATION

University of California, San Diego

Sep 2022 - Jun 2026

B.S. in Data Science & B.S. in Math-Economics, Minor in Cognitive Science

Relevant Organizations: IMPACT Lab Research, Data Science Society, UCSD ACM AI, UCSD Chess Club

Relevant Coursework: Data Structures & Algorithms, Principles of Data Science, The Application of Data Science, Theoretical

Foundations: ML, Systems for Scalable Analytics, Applied Machine Learning, Data Mining Optimization, Computational Linear Algebra,

TECHNICAL SKILLS

Programming Languages: Python • SQL • Java • R • HTML • JavaScript • MATLAB

Libraries & Frameworks: Numpy • Pandas • Seaborn • SciKit-Learn • PyTorch • Cytoscape.js • Tensorflow • Langchain • Hugging Face

Tools and Platforms: Jupyter • Power BI • AWS • Snowflake • Git • VSCode • Tableau • Apache Spark • OpenAI • Flask

WORK EXPERIENCE

ADP, Machine Learning Intern - Atlanta, GA

May 2025 - Aug 2025

- Designed and deployed an AI-driven knowledge graph platform for ADP Assist leveraging **LangChain**, **LLM tool calling**, **Pydantic**, and **Amazon Neptune**, enabling ingestion, normalization, and intelligent linking of heterogeneous datasets for **150,000+ clients** with a **Cytoscape.js** interface for real-time exploration

- Developed and integrated **ML-based anomaly detection models** within advanced knowledge graph pipelines to identify **fraud**, **payroll inaccuracies**, and **compliance risks**, achieving **100% detection accuracy** and delivering context & risk-sensitive insights

Solely, AI Intern - San Diego, CA

Feb 2025 - Jun 2025

- Orchestrated the integration of **Cassidy.AI**, **Salesforce**, and **Copilot** in collaboration with the sales team, developing custom **Python** scripts and **AI agents** to enable instant query resolution and boost task efficiency by **30%** for employees

- Directed **ML-driven modules** to analyze sales data across **10,000+ stores and products**, leveraging **machine learning models** in partnership with the business intelligence team to uncover key pricing/display patterns and accelerate **data-driven decision-making**

Impact Lab, Data Science Undergraduate Researcher - San Diego, CA

Jan 2025 - May 2025

- Engineered **neural network** and **graph theory-based** solutions in **Python** to analyze **5,000+ datasets**, uncovering **linguistic patterns** tied to social perception and emerging behavioral trends

- Optimized **sentiment extraction pipelines** by implementing **LSTM deep learning models**, delivering higher comparative accuracy than prior baselines for **complex, multi-variable datasets**

Ally Financial, Data Science & AI Intern - Charlotte, NC

May 2024 - Aug 2024

- Led the end-to-end development of a **GenAI-powered Q&A platform** leveraging **Python**, **Snowflake**, and **NLP**, streamlining access to **5,000+ Confluence documents** and enhancing information retrieval efficiency for **1,000+ employees** across data teams

- Applied **statistical modeling**, **advanced prompt engineering**, and **scenario-based analysis** to evaluate **400+ prompt-response combinations**, optimizing **model accuracy to 82%** and earning direct recognition from the **CIO**

Write On Inc., Data Science Intern - Cypress CA

Jun 2023-Aug 2023

- Evaluated the success of **150+ students** by employing **Support Vector Machine (SVM) algorithms** to identify key performance predictors, optimizing learning efficiency and driving a **27% improvement** in student performance through **data-driven analysis**

- Collaborated with **instructors** to assess performance metrics, leveraging **Seaborn** to create advanced **statistical visualizations**

PROJECT EXPERIENCE

Financial News Knowledge Graph System

July 2025 - Present

- Developed a **multi-stage NLP pipeline** to process financial news, combining a **fine-tuned FinBERT model** with **LLM-based validation**. Achieved **95%+ accuracy** and an **F1-score of 0.91** in extracting critical entities, events, and relationships

- Designed and deployed an **AI-driven financial knowledge graph in Neo4j**, integrated with an interactive **React dashboard**, enabling analysts to uncover company relationships, monitor executive activity, and visualize market trends with **87% precision**

MyEdMaster

Nov 2024 - Apr 2025

- Built an **AI-powered medical chatbot** using **Python**, **NLP**, **LangChain**, and **CrewAI**, processing **1,000+ medical prompts** for **5,000+ users** and contributing to a **published research paper** on the methodology and performance

- Spearheaded development of a **Retrieval-Augmented Generation (RAG) pipeline** for SAT question creation, incorporating adaptive retrieval strategies and automated validation to improve question accuracy by **22%**

Doordash Market Analysis

Jun 2024

- Developed a **regression model** on **18K+ DoorDash orders** to predict delivery time with **<12.5% error** (vs. 48.7-min average), uncovering patterns such as higher tips for slower deliveries, minimal tip difference, and peak delay times

- Presented findings with **Seaborn statistical visualizations** and actionable insights, highlighting opportunities to **optimize delivery efficiency** and improve **driver incentive strategies**