Pratham Aggarwal

San Diego, CA | 619-953-7147 | praggarwal@ucsd.edu
pratham-aggr.github.io | linkedin.com/in/pratham-agg | github.com/pratham-aggr

Education

Bachelor of Science in Data Science, Minor in Probability & Statistics Expected Dec 2026
University of California, San Diego (School of Computing, Information and Data Sciences)
Relevant Coursework: Machine Learning, Probability & Statistics, Exploratory Data Analysis, A/B Testing, Time Series Analysis, Statistical Inference, Linear Algebra, Optimization, Data Structures, Object-Oriented Programming, Databases

Skills

Programming Languages: Python, SQL, Java, JavaScript, MATLAB, C++
Frameworks & Libraries: PyTorch, TensorFlow, Pandas, Scikit-Learn, SciPy, SymPy, React, Node.js, FastAPI
Tools & Platforms: Google Cloud Platform (GCP), Docker, Git/GitHub, Linux/Unix, Excel, Terminal, Hugging Face
Data Visualization & Analysis: Tableau, Matplotlib, Seaborn, Experiment Design, Research Ethics, Documentation

Experience

Quantitative Trader Intern Mar 2025 – Jun 2025
Student Foundation Investment Committee, Quantitative Technologies *San Diego, CA*

- Increased options trading returns by 12% by developing a Deep Reinforcement Learning agent for a $1.3M student-run investment fund, leveraging adaptive policy learning and optimized early exercise strategies.
- Enhanced trading realism and decision quality by 18% by modeling a sequential decision-making process that dynamically selects option type, expiration, and exercise timing from time-series patterns.
- Reduced Mean Squared Error to 0.96 with an LSTM-based stock price prediction model using 5 years of historical time-series data and extensive hyperparameter tuning.

Data Science Researcher Jun 2025 – Present
Climate Analytics Lab, Scripps Institution of Oceanography *San Diego, CA*

- Awarded a $4,500 research scholarship under PhD Duncan Watson-Parris; processed 243+ observational and projected climate datasets (e.g., CMIP6) on Google Cloud to assess machine learning models for physical plausibility.
- Developed 20+ custom evaluation metrics and an ML pipeline to benchmark 50+ climate model outputs, creating interactive time series plots that expose black-box limitations.
- Designed an interactive React/Node.js web application to visualize model benchmarks and trends, enhancing stakeholder understanding and accessibility of climate insights.

Program Manager & Consultant Mar 2025 – Jun 2025
Solana Center & Data Science Student Society (DS3) Consulting *San Diego, CA*

- Evaluated program success by analyzing 2,000+ composting records, revealing a 6.7% increase in waste diversion.
- Improved environmental data accuracy by 9% by imputing missing values, validating entries, and performing statistical analysis on large-scale datasets to enable precise, scalable insights.
- Improved stakeholder decision-making and ensured on-time project delivery by creating structured, data-informed workflows and presenting actionable insights.

Projects

HackFrontier Winner: Geospatial ML & CV tool for Homeless Services *Learn more*

- Won a hackathon as the youngest among 100+ professionals by developing a ML forecasting system with 67% accuracy, leveraging 35+ transit, demographic, and geographic features to strategically place homeless service centers in San Diego.
- Deployed a real-time computer vision system using Oxen.ai and EyePop.ai to address the challenge of tracking a transient homeless population, providing live monitoring and precise demand insights to enable accurate, data-driven responses.

Simulating Black Hole Evolution: Comparative Analysis of Light and Heavy Seeds *Learn more*

- Demonstrated that seed mass impacts black hole growth rates by up to 40% through research visualizations developed over 20 weeks of PhD-supervised analysis using 10+ visualization techniques.
- Simulated growth trajectories for over 100 black hole seed scenarios using Eddington accretion models, analyzing evolutionary differences between light and heavy seeds over 20 million years of data.

Predictive Modeling of Building Energy Loads *Learn more*

- Achieved 91% accuracy in predicting building heating and cooling loads by implementing a multiple linear regression model using 8 key architectural and environmental features on 750+ samples.
- Optimized operational costs by 9% by applying k-means clustering to identify distinct consumption patterns, driving a 12.3% behavioral shift in load forecasting.