

# Parth Gupta

+1 858-319-6082 | U.S. Citizen | p5gupta@ucsd.edu | [LinkedIn](#) | [GitHub](#)

EDUCATION

# University of California San Diego

*B.Sc in Mathematics - Computer Science*

San Diego, CA

Sep. 2021 – Mar 2025

## EXPERIENCE

## Data Annotator

Toloka AI | *Data Labeling, Quality Assurance, Model Evaluation*

May 2025 – Present

Remote

- Annotated and validated large-scale datasets for machine learning models across domains such as natural language processing and image recognition.
  - Ensured data quality through inter-annotator agreement checks and outlier identification, improving dataset consistency by 20%.
  - Collaborated with data scientists to refine annotation guidelines, supporting model training pipelines and downstream analytics.
  - Utilized analytical dashboards to monitor project metrics and optimize labeling efficiency.

## Software Engineer Intern

Jan 2024 – May 2024

More Seconds | *MongoDB, React, TypeScript, Node.js, Express.js, Docker, OpenAI, Python*      Remote, USA

- Containerized development environment with Docker, ensuring consistency and streamlined deployment.
  - Performed data analysis through bar and line charts for implementing AI agent across various dashboards, improving developer task allocation accuracy by 30%.
  - Developed and validated APIs to support **data pipelines**, ensuring accuracy and reproducibility.
  - Developed full-stack features for developer dashboards like displaying mutual followers and billable toggle, optimizing projects and sub-projects management processes by 60%.
  - Used Git and Agile methodologies for iterative development and on-time feature delivery.

Full Stack Intern

Aug. 2022 – Sep 2022

ThatNeedle.com | *JavaScript, Flask, HTML, CSS*

Remote, India

- Engineered backend pipelines to process and serve multi-year stock trend data and enabling dynamic sector-wise comparisons via **Chart.js** visualizations, driving a 30% increase in platform engagement.
  - Built an interactive geo heat map with **Leaflet** and **D3.js** to analyze global population disparities, allowing stakeholders to identify high-risk regions for targeted resource allocation.
  - Optimized visual layout and color schemes to improve readability and user engagement.

PROJECTS

[Credit Card Fraud Detection](#) | *Python, NumPy, Pandas, SciKit Learn*

- Processed 6M+ transactional records; applied SMOTE to handle imbalance, and engineered features with **MinMaxScaler** and **LabelEncoder**, predicting frauds using **Logistic Regression** and **Random Forest Classifier**.
  - Improved classification performance by replacing Logistic Regression with Random Forest, boosting accuracy from **92.4%** to **99.9%** and **F1 score from 6.1 to 88.7**.

[E-commerce Sales Dashboard](#) | *Tableau, Excel, Financial Data*

- Used provided **ERDs** to transform multi-segment e-commerce sales data with **SQL** (JOINs, CTEs, aggregates, CASE statements), delivering clean, analysis-ready datasets for reporting and dashboards.
  - Designed an interactive Tableau dashboard with **KPI cards, geospatial heat maps, dual-axis charts, calculated fields, parameter controls, and dashboard actions** to deliver insights by segment, region, and manufacturing origin, optimized with data source filters.

## TECHNICAL SKILLS

## Programming Languages: JavaScript, Python, Java, C/C++, R, HTML/CSS

**Web Development Skills:** React Native, React NextJS, Flask, Express.js, Django, React Redux, PyGame, pandas, NumPy, Matplotlib, Scikit-learn, Socket.io, Tableau, NoSQL, MySQL, MongoDB, Supabase, REST APIs Docker, AWS, Heroku, Vercel

**Relevant Coursework:** Data Structures, Computer Organization, Discrete Mathematics, Data Mining, Numerical Optimization and Linear Algebra

**Online Courses:** GitLab CI/CD: Pipelines, CI/CD and DevOps for Beginners, Google Data Analytics, Tableau Fundamentals