

AJAY SHARMA

PHONE: (510) 575-8812 EMAIL: ajay.sharma@berkeley.edu WEBSITE: [linkedin.com/in/asharma2718](https://www.linkedin.com/in/asharma2718)

SUMMARY

Senior at UC Berkeley, double majoring in Applied Mathematics (Data Science) and Statistics, driven to bridge analytical methods with robust software solutions. Skilled in data modeling, machine learning, and optimization. Proven ability to lead projects, consistently delivering impactful and reliable results.

EDUCATION

MAY 2025 **University of California, Berkeley**
B.A. Applied Mathematics (Data Science), B.A. Statistics
Major GPA: 3.576/4.000

TECHNICAL PROJECTS

If you would like to see the source code for any of the projects, send me an email (see above).

Spam & Ham Email Classifier

December 2024

- Conducted comprehensive **exploratory data analysis** on a **dataset containing thousands of emails**, uncovering critical patterns and insights for effective classification.
- Engineered **high-quality features** such as **email length, punctuation usage, token frequency, and the presence of stop words**, ensuring a robust model design while minimizing multicollinearity and redundancy among features.
- Incorporated **advanced text preprocessing techniques**, including **filtering out common stop words** and **emphasizing unique terms**, to enhance the **precision and relevance** of the model's predictions.
- Implemented a **logistic regression model** to classify spam emails, achieving an **accuracy of over 92%**, placing me in the **top 2% of my class of 1200 students**.
- Evaluated the model's performance using the **Receiver Operating Characteristic (ROC) curve**, achieving an **Area Under the Curve (AUC) value of 0.980**, underscoring the model's **superior classification capabilities**.
- Refined and **optimized feature selection iteratively**, leveraging **statistical insights** to create a **scalable and reliable model** suitable for **real-world applications**.

Index Fund Investment Simulator (R Shiny App)

November 2024

- Built an interactive investment simulator in **R Shiny** to model and visualize long-term portfolio growth for **Total Stock Market Index Funds**, including popular mutual funds and ETFs like **VTSAX, FSKAX, SWTSX, and ITOT**.
- Integrated options for users to specify **initial investment, periodic contributions, and time horizon**, allowing them to explore how varying inputs impact investment outcomes.
- Simulated market returns with **variable rates and volatility**, using historical data to create a **realistic projection** of portfolio performance over time.
- Provided users with **visual insights on compounding interest** through interactive charts, empowering them to make informed decisions about index fund investments.

Housing Price Predictor

October 2024

- **Objective:** Developed a **predictive model** for housing prices in Cook County, Illinois, achieving **low RMSE (< 240K)** and ranking in the **top 70 out of a class of 1200**.
- **EDA:** Analyzed **100,000+ household records** using **pandas**, identifying **key trends** and visualizing relationships between features and prices with **matplotlib** and **seaborn**.
- **Feature Engineering:** Implemented **data transformations**, including custom **OneHotEncoder** to handle **unseen categories** and ensure dataset compatibility.
- **Data Cleaning:** Designed a **clean_data** function to **impute missing values**, replace infinite values, and validate inputs, creating a **robust dataset**.
- **Modeling:** Trained a **multiple linear regression model**, optimized hyperparameters, and validated performance with **cross-validation** for stability.
- **Results:** Delivered **accurate price predictions**, leveraging insights to uncover important predictors like **location, size, and number of bedrooms**.

Harry Potter Sentiment Analysis (R Shiny App)

October 2024

- **Objective:** Conducted **sentiment analysis** on the Harry Potter book series using **natural language processing (NLP)** techniques to uncover **emotional trends** and character sentiment over the storyline.
- **EDA:** Processed **100K+ lines of text data** from the books, identifying **word frequencies**, sentiment scores, and **emotional patterns** across characters and chapters using **pandas** and **nlTK**.
- **Feature Engineering:** Extracted sentiment features using the **VADER** sentiment analysis tool and tokenized text data with **spaCy** to quantify **emotional intensity and polarity**.
- **Data Cleaning:** Preprocessed text data by **removing stop words**, handling punctuation, and normalizing tokens to improve **sentiment classification accuracy**.
- **Modeling:** Built a **text-based sentiment classification pipeline**, utilizing **logistic regression** and **decision trees** to classify chapters by dominant sentiment categories.
- **Results:** Discovered **key sentiment trends**, such as the gradual increase in negative sentiment during climactic events, and visualized findings using **matplotlib** and **seaborn**.

NGordNet (Wordnet): Graph-Based Data Processing (Java)

May 2023

- **Objective:** Developed **NgordNet**, a graph-based system to model relationships and compute shortest paths in a network, using **Agile methodology** to iteratively refine requirements and deliverables.
- **Data Structures & Algorithms:** Designed **adjacency lists** for graph representation and implemented **Dijkstra's** and **breadth-first search (BFS)** algorithms for efficient pathfinding.
- **Optimization:** Improved graph traversal performance by analyzing **runtime complexity** and reducing query response times for large-scale networks.
- **Error Handling:** Implemented **robust error-checking mechanisms** for edge cases, including disconnected nodes and invalid inputs, ensuring system reliability.
- **Results:** Delivered an **efficient and scalable solution** that accurately modeled real-world networks, validated through interactive testing and benchmarks.

WORK EXPERIENCE

BLCK UNICRN · CDSS (AI Student Researcher Internship)

January 2025 - May 2025

- Collaborate with a team to define the Ideal Customer Profile (ICP) for BLCK UNICRN by leveraging predictive analytics, clustering models, and AI-driven web scraping to identify high-value prospects.
- Contribute to the development and refinement of advanced AI functionalities, including retrieval-augmented generation (RAG), intelligent agent integration, and natural language processing (NLP) models to enhance platform capabilities.
- Engage with real-world AI development and integration within a fast-growing startup environment to provide insights to enhance the platform's performance, and support a 1000x growth in the sales pipeline.
- Utilize programming languages (Python, R), data visualization tools (Tableau), and cloud platforms (AWS) to analyze and present data.
- Apply machine learning techniques such as classification and regression to extract actionable insights and communicate findings clearly to drive strategic decisions and product innovation.

Beats by Dre (Consumer Insights & Data Analytics Externship)

June 2024 - August 2024

- Leveraged Natural Language Processing (NLP) techniques in Python to analyze over 500+ customer reviews from platforms such as Amazon, utilizing sentiment analysis to uncover key trends in consumer preferences and pain points.
- Conducted comprehensive data preprocessing (of large data base), including text cleaning and tokenization to extract meaningful insights from unstructured data.
- Developed a customer segmentation strategy based on sentiment clusters, enabling the identification of 10+ distinct visualizations, leading to tailored marketing campaigns and personalized product recommendations.
- Identified actionable insights on consumer priorities (e.g., sound quality, design, price sensitivity), contributing to a 15% improvement in targeted marketing by refining campaign messaging and alignment with audience profiles.
- Analyzed customer sentiment trends, defined satisfaction metrics, and identified key feature preferences, providing stakeholders with intuitive and interactive tools for decision-making.
- Delivered a final presentation to cross-functional stakeholders, including product managers and marketing teams, translating technical findings into strategic recommendations that informed product development.

Success Chess School (Chess Instructor)

September 2021 - May 2023

- Mentored 150+ students over the course of 2 years, emphasizing strategic thinking and sportsmanship, resulting in consistent performance improvement and higher student retention.
- Authored an instructional chess book tailored for beginners, integrating puzzles and detailed analyses of 10 fundamental openings to reinforce strategic learning and critical thinking.
- Organized 10+ tournaments, leveraging a custom-built Excel-based matchmaking system that automatically paired participants by analyzing ELO ratings and skill levels (beginner, intermediate, advanced), resulting in a 20% year-over-year increase in registrations and fostering a competitive yet supportive environment.

AS Mathematics (Mathematics Instructor)

August 2018 - Present

- Taught statistics, calculus, and algebra to over 100 students, achieving an average grade improvement of two letter grades through individualized instruction.
- Created engaging course materials tailored to diverse learning styles, enhancing comprehension and student engagement in both in-person and virtual settings.
- Conducted outreach to 100+ parents and students, increasing program visibility and driving a steady growth in enrollments.

ADDITIONAL ACTIVITIES

American Allegiance of Education (Co-Founder, Volunteer)

May - October 2024

- Co-founded the [American Allegiance of Education](#), a non-profit organization dedicated to promoting education and innovation through technology-focused initiatives, including organizing international events for aspiring developers and students.
- Volunteered as a lead organizer for a global hackathon, coordinating efforts across teams to create an inclusive and engaging event that attracted 155 participants from multiple countries.
- Successfully raised \$16,419 in sponsorships and prizes by engaging with corporate sponsors and educational institutions, securing funding and resources that significantly increased the event's impact and reach.
- Took the initiative to recruit and organize a panel of expert judges, creating volunteer opportunities for industry professionals to participate and evaluate innovative projects.
- Provided mentorship and guidance to participants throughout the hackathon, fostering a collaborative and supportive environment to help participants develop their skills and showcase their projects.