AJAY SHARMA

Phone: (510) 575-8812 Email: ajay.sharma@berkeley.edu Website: linkedin.com/in/asharma2718

Summary

Senior at UC Berkeley, double majoring in Applied Mathematics (Data Science) and Statistics, with a strong passion for Machine Learning and Data Science. Skilled in leveraging advanced statistical methods and data tools, with proven strengths in problem-solving, organization, and project leadership.

Education

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

Technical Projects

Housing Price Predictor (Python-Based)

October 2024

- **Objective**: Developed a predictive model for housing prices in Cook County, Illinois, optimizing for low root mean square error (RMSE) and achieving a ranking in the **top 6% out of 1200 participants**. This model provided accurate price forecasts to determine market trends.
- **Exploratory Data Analysis (EDA)**: Analyzed **100,000+** household records using `pandas`, identifying key trends and patterns in the data. Used `matplotlib` and `seaborn` to visualize relationships between property features and prices, uncovering important predictors such as location, property size, and number of bedrooms.
- **Feature Engineering**: Enhanced model performance through strategic data transformations, including:
  - **One-Hot Encoding**: Built a custom `ohe_column` function using `OneHotEncoder` to encode categorical variables, ensuring compatibility between training and test datasets by handling unseen categories.
  - **Data Cleaning**: Designed a `clean_data` function to handle missing values (using median imputation for numerical features and mode for categorical features), replace infinite values, and ensure all inputs were positive, resulting in a more robust dataset.
- **Modeling & Error Analysis**: Trained a multiple linear regression model, optimizing hyperparameters to improve accuracy. Performed cross-validation and residual analysis to validate model stability and detect systematic errors across different property types, ensuring reliable predictions.

S&P 500 Investment Simulator (R Shiny App)

October 2024

- Developed an interactive web application using **R Shiny** to model and simulate long-term portfolio growth for the **S&P 500 index**, allowing users to experiment with customizable financial scenarios.
- Implemented **Monte Carlo simulations** to project potential portfolio trajectories under varying market conditions, incorporating adjustable inputs for initial investment, periodic contributions, average annual return, volatility, and investment duration.
- Integrated intuitive user interfaces for real-time input adjustments, enabling users to explore how financial decisions influence portfolio performance over time.
- Designed dynamic, interactive visualizations to depict portfolio growth, including multiple simulated trajectories and percentile bands (10th, 25th, median, 75th, and 90th percentiles) for clearer risk analysis.

Climate Change Impact Analysis (Python-Based)

August 2024

- Preprocessed datasets by handling **25,000+ missing values** and cleaning **20,000+ rows** of data, converting it into a **time-series format** to improve data quality and ensure accurate analysis.
- Developed a predictive model using **machine learning techniques** such as **Stochastic Gradient Descent (SGD)** to estimate model coefficients, achieving an **89% accuracy** in predicting temperature changes.
- Explored the **log-normal distribution rigorously** to assess the underlying data patterns and applied a **regression model** to quantify the impact of CO2 emissions on temperature anomalies.
- Conducted hypothesis testing to evaluate the relationship between CO2 emissions and temperature anomalies, yielding a statistically **significant p-value (< 0.05)** that validated the model's findings.
- Achieved a **strong correlation ($R = 0.94$)** between CO2 emissions and temperature anomalies, highlighting the model's effectiveness in identifying global warming trends.
- Utilized **Python libraries** (e.g., Pandas, NumPy, Matplotlib) to analyze data and create **10+ visualizations**, illustrating key insights such as the direct relationship between temperature changes and CO2 emissions.

NGordnet (Java-Based Lexical Analysis)

April 2023

- Designed a **Java-based application** to analyze historical word usage trends using data from **WordNet** and **Google Ngrams**.
- Utilized **dynamic data structures** such as `HashMaps` and `ArrayLists` for efficient storage and retrieval of word data, and integrated WordNet to perform **graph traversals**, enhancing semantic and linguistic analysis.
- Developed efficient algorithms to compute **relative word frequencies**, enabling the generation of **insightful visualizations** showcasing how word usage evolved over time.
- Extended functionality to support **interactive user queries** for custom word comparisons, demonstrating proficiency in **object-oriented programming** and algorithm design.

Harry Potter Text Analysis App (R Shiny App)

October 2024

- Built an **interactive R Shiny app** to analyze all seven Harry Potter books using **natural language processing (NLP)** techniques, focusing on **sentiment analysis**, **word frequency**, and **tf-idf** rankings.
- Integrated multiple sentiment lexicons (**Bing**, **NRC**, **Afinn**, and **Loughran**) to compute **sentiment scores** and classify emotions (e.g., joy, sadness, anger), visualizing trends across books and chapters.
- Designed interactive visualizations for key metrics, including **tf-idf rankings**, sentiment distributions, and chapter-level sentiment comparisons, allowing users to uncover unique text patterns.
- Enabled detailed data exploration through interactive tables and filtering options, providing users with a flexible tool for understanding linguistic and emotional dynamics in the series.

Work Experience

Beats by Dre (Consumer Insights & Data Analytics Externship)

June 2024 - August 2024

- Applied natural language processing (NLP) techniques in Python to analyze 500+ customer reviews, uncovering key trends in sentiment and preferences that guided product and marketing strategies.
- Extracted insights on consumer pain points and valued features, leading to a 15% improvement in targeted marketing by aligning campaigns with refined customer profiles.
- Prepared and presented data-driven insights to stakeholders, offering actionable recommendations that enhanced understanding of customer needs and informed strategic decisions.

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.
- Developed a structured curriculum with tactical exercises and personalized feedback, which led 30% of students to reach finalist positions in tournaments.
- Organized 10+ tournaments, contributing to a 20% year-over-year increase in registration by creating a competitive and 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, Judge)

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** and served as a **judge** during the hackathon, evaluating projects on their creativity, technical implementation, and real-world applicability, while providing mentorship and guidance to participants to foster a **collaborative and supportive environment**.