

Colin Bradley

Data Scientist

(916) 865-8345 | c.bradley.do@gmail.com | [Website](#) | [LinkedIn](#) | [GitHub](#)

Highly skilled Data Scientist with over 3 years of experience in designing, building, and deploying machine learning models and data-driven solutions. Proficient in statistical modeling, data analysis, and advanced machine learning techniques, including deep learning and natural language processing. Adept at transforming large, complex datasets into actionable insights to drive business outcomes across diverse domains, including healthcare, public safety, and eCommerce. Experienced in leading end-to-end projects from data ingestion to model deployment, with a proven track record of automating workflows, improving data processing efficiency, and creating innovative data applications. Strong expertise in Python, SQL, Spark, and cloud platforms like AWS and Azure. Passionate about delivering high-quality, scalable solutions, and driving data strategy to optimize decision-making. Holds a Master's degree in Physics with published research in advanced statistical methods.

TECHNICAL SKILLS

Libraries: scikit-learn, Tensorflow, XGBoost, Keras, statsmodels, pandas, NumPy, matplotlib, SciPy, Flask, BeautifulSoup, selenium, pyspark, seaborn, plotly, regex, geopandas, shapely

Languages: Python, SQL, C++, SAS

Developer Tools: Git, Github, Snowflake, AWS, Azure Cloud Services, Oracle, MS SQL Server, Databricks, Apache Spark, PostgreSQL, Excel, PowerBI, Jupyter, Linux/UNIX, GIS (ESRI & geopandas), Azure DevOps

Data Analysis: statistics, data visualization, machine learning, deep learning (CNN), time-series, LSTM, ARIMA, GenAI

PROJECTS

[Fantasy Football Analysis](#) (In Progress) | *Python, XGBoost, Scikit-Learn, Plotly Dash*

- Achieved a 10% improvement over expert predictions by training and optimizing multiple models (linear regression, random forest, gradient boosted random forest), resulting in a mean absolute error (MAE) of 5.6 and empowering players to make informed decisions and win more games.
- Built a complex data pipeline that seamlessly integrates data from diverse sources, generates new features, and facilitates smooth and accurate model building and data analysis.

[Book Recommender](#) | *Python, Flask, PostgreSQL, AWS RDS & EC2*

- Designed and deployed an end-to-end book recommender system on AWS using collaborative filtering and content-based algorithms, enhancing user experience with personalized recommendations.
- Utilized Python, pandas, and scikit-learn to build and evaluate recommendation models, demonstrating proficiency in machine learning, data analysis, and cloud-based deployment.

[Data Jobs Web Scraper](#) | *Python, Selenium, Pandas, Regex, Matplotlib, Seaborn*

- Engineered a web scraper using Selenium to extract and analyze over 6,000 Data Science job postings, identifying trends in job locations, salary ranges, and required skills. Designed more than 10 visualizations to communicate key findings effectively.
- Optimized data pipelines for large-scale web scraping and processing, ensuring the efficient collection, cleaning, and transformation of both structured and unstructured data leading to actionable insights on pursuing a career in data science.

EXPERIENCE

California State Government

Sacramento, CA

Data Analyst - California Department of Public Health

Feb 2024 – Current

- Developed an end-to-end data application using Flask to help analysts query and process data from several healthcare related databases, reducing response time on key stakeholder inquiries from weeks to hours.
- Automated Extract, Transform, and Load (ETL) processes using Databricks and Spark to enhance data processing efficiency and scalability.
- Led a high-impact project in response to a federal audit, delivering critical insights to both internal and external executive stakeholders and senior leaders. Collaborated with federal partners to develop, clean, and analyze over 20 unique datasets, ensuring policy compliance and enhancing data transparency.
- Responded to over 100 complex data requests from diverse internal and external stakeholders, ensuring data integrity and accuracy. These data analyses frequently informed state-wide policy decisions.

Data Analyst - CAL FIRE

Feb 2023 – Feb 2024

- Created a scalable web scraping pipeline using Python, Selenium, and Azure Cloud Services, automating manual data processes and saving thousands of hours of labor.
- Collaborated with a cross functional team to create a cloud-based data validation REST API in Azure Function Apps, leveraging geographic information systems APIs and internal techniques to ensure the quality of incoming data to a database.
- Designed a comprehensive PowerBI dashboard, providing invaluable business insights through text analysis (using both classical and GenAI techniques) and data visualization of thousands of structured and unstructured survey responses.

- Rapidly extracted wildfire data from hundreds of PDFs for a federal partner, ensuring timely and accurate delivery under tight deadlines.

Data Analyst - California Department of Public Health

Sep 2021 - Feb 2023

- Streamlined data validation and outlier detection processes on healthcare datasets by leveraging statistical, machine learning, and deep learning techniques, resulting in a time savings of over 20 hours per month.
- Demonstrated expertise in feature engineering (new variable creation), statistical analysis, and programming to visualize, merge, analyze, and validate intricate datasets, effectively delivering actionable insights to stakeholders to facilitate data-driven decision making.
- Automated routine manual data processes, enabling staff to concentrate on innovative and strategic initiatives.

Graduate Research Assistant

Jan 2019 – Sep 2020

University of Nevada, Reno

Reno, NV

- Employed advanced statistical analysis techniques, including confidence intervals, hypothesis testing (A/B Testing), Markov chain Monte Carlo simulation, Bayesian Statistics, and Gaussian Mixture Modeling, to explore exotic physical phenomena.
- Collaborated with a cross functional team to prepare and publish research papers in esteemed academic journals, making notable contributions to the field of study (refer to publications [here](#) and [here](#)).

EDUCATION

University of Nevada, Reno

Reno, NV

Master of Science in Physics

Aug 2018 – Aug 2020

University of Nevada, Reno

Reno, NV

Bachelor of Science in Physics

Aug 2015 – May 2018