Brandon Lee Concepcion

(650)-892-4233 | brandon_concepcion@berkeley.edu | linkedin.com/in/brandonconcepcion | brandonconcepcion.github.io

Education

UC Berkeley Berkeley, CA

***Majors**: B.A. Data Science, B.A. Computer Science* *Class of 2026 | GPA: **3.8**/4.0*

Organizations: College of Computing, Data Science, and Society, Data Science Society at Berkeley, Berkeley Unboxing Data Science, Data C8 (Foundations of Data Science) Undergraduate Course Staff, Data Scholars, Chinese Student Association, Intramural Sports

Skills

Languages: Python, Java, SQL, RegEx, LaTeX, HTML | Learning Ruby, Javascript

Libraries: Pandas, NumPy, TensorFlow, PyTorch, CV2, SciPy, StatsModels, SkLearn, Seaborn, Plotly, MatplotLib, NLTK

Technical Skills/Coursework: Deep Learning, Neural Networks, Computer Vision, Natural Language Processing, Machine Learning Theory, Git, Data Structures, Advanced Calculus, Linear Algebra, Discrete Math, Advanced Statistics and Probability, Clustering, Classification, Linear and Logistic Regression, Data Visualization, Data Analytics, Exploratory Data Analysis

Soft Skills: Efficient Communication, Adaptability & Flexibility, Teamwork and Collaboration, Organizational Ability, Simple and Creative Problem Solving, Leadership, Critical-Thinking, Effective Time Management, Self-Starter, Attention to Detail, Curiosity

Experience

UC Berkeley College of Computing, Data Science, and Society Berkeley, CA

Undergraduate Student Instructor - Foundations of Data Science *Jun 2024 - Present*

- Developed and delivered educational content to a student population of over 1,800, providing comprehensive and effective instruction through office hours, review sessions, and management of 3 tutors to answer student questions and address concerns
- Enhanced data8.org/su24 using HTML, CSS, and Javascript for front-end web development; adding 19 dynamic tabs to filter 250+ past exam problems by topic, leading my students to achieve the highest median final exam scores (90th percentile), as well as the highest overall median cumulative course grade among all student instructors (93rd percentile)

Data Science Society at Berkeley Berkeley, CA

Lecturer, Course Director *Aug 2023 - Present*

- Managing 8 Teaching Assistants and 6 tutors to operate the "An Introduction to Real World Data Science" course, promoting accessibility by hosting educational workshops from industry and academic leaders for the 70+ diverse students in the course
- Teaching and facilitating 70+ students towards the completion of an 8-week-long introductory data science project, lecturing on industry-level topics such as Long Short Term Memories, Random Forests, Classification, and Multiple Linear Regression

University of Washington Remote

Data Scientist *Jan 2024 - Jun 2024*

- Utilized natural language processing (NLP) techniques, including Latent Dirichlet Allocation (LDA) and Bidirectional Encoder Representations from Transformers (BERT), to uncover hidden themes in student health and wellness survey responses
- Applied supervised learning models to classify topics in student feedback, enhancing the analysis of student engagement patterns and the effectiveness of self-care initiatives, providing insights for professors at the School of Pharmacy

Research/Projects

To Spam, or Ham? 🔗 | *Python, Scikit Learn, Principal Component Analysis* *Apr 2024*

- Engineered a logistic regression model for classifying emails as Spam (1) or Ham (0), using a dataset of over 7,500 points and resulting in a classifier that achieved 87% testing accuracy and an Area Under the ROC Curve (AUC) of 0.91
- Applied advanced feature engineering and Principal Component Analysis (PCA) to reduce dimensionality from 18 to 5 features, mitigating overfitting and ensuring 92% of legitimate (ham) emails in the test set of 1065 were correctly identified
- Enhanced model performance through GridSearch optimization across 4 hyperparameters, resulting in a 5% increase in accuracy when classifying a validation set of 1250 emails

Prophecy of Properties 🔗 | *Python, Pandas, Sklearn, Regularization* *Mar 2024*

- Trained an 85-feature machine learning model on 500,000+ housing records to predict housing prices in Cook County, Illinois
- Achieved a Root Mean Squared Error (RMSE) of $103k across a testing set of 30,000+ housing records, leveraging techniques such as feature engineering and k-fold cross-validation to tune hyperparameters and optimize model performance

Gen AI Retinal Video Sequences 🔗 | *PyTorch, CV2, Deep Neural Networks, Long Short Term Memories* *Nov 2023*

- Coded a Variational Auto-Encoder (VAE) neural network to convert retinal movement numerical data into generative AI video simulations of retinal movement afflicted by one of three different diseases
- Implemented a data preprocessing pipeline that converts .avi files into sets of 300 individual frames
- Ran training data through a Long-Short Term Memory (LSTM) network to encode data into latent space, then decoded data using a Gated Recurrent Unit (GRU), producing video simulations in 512x512 resolution

Interests

Academic: Computer Vision, Deep NNs, Teaching, Probability, Data Analytics, NLP, Pedagogy, Astronomy

Personal: Watching Movies, Hiking, Photography, Basketball, Gym, Swimming, Web Development, Music, Marvel, Ping Pong