Thomas Kim

714-833-0362 | thomasekim10@gmail.com

EDUCATION

University of California, Berkeley | Berkeley, CA **August 2019 - May 2023**

- B.S. in **Electrical Engineering and Computer Sciences**
- Minor in Data Science
- Relevant Coursework: Data Structures, Algorithms, Artificial Intelligence, Discrete Mathematics /Probability Theory, Optimization Models, Machine Structures, Designing Information Systems, Multivariable Calculus, Principles of Data Science, Databases

SKILLS

Languages/Tools: Java, Python, SQL, R, C, Jupyter, Numpy, Pandas, Matplotlib, Scikit-learn, AWS S3, Jira, Tableau

Data Structures and Algorithms

- Object-oriented programming and functional programming
- Advanced data structures, search algorithms, asymptotic analysis and time/space complexity analysis

Git Version Control, SSH, Agile Methodology

PROJECTS

NBA Draft Analysis | *Python* **June 2023**

- Implemented Python program using Beautiful Soup to pull data from HTML object to create a 1,000-line CSV file
- Aggregated and cleansed 20 years of NBA/NCAA basketball data with SQL into Pandas DataFrame and determined important features
- Trained a Machine Learning (ML) model with decision tree classification to predict NBA draft position for each college player
- Drafted and published an article containing embedded code blocks and visualizations using Matplotlib from DeepNote notebook

Walmart Sales Forecasting | *SQL* **July 2023**

- Developed custom SQL queries to view, aggregate, and perform data cleansing on Kaggle dataset with over 150,000 entries
- Constructed four Machine Learning Models to generate sales predictions with an accuracy greater than **95%** on test set
- Enhanced performance by using ensemble averaging that outperformed the root mean squared error of the top performing model by **6%**
- Leveraged Tableau to create meaningful visualizations about important features and how the model performed over time

MOOCbase | *Java* **May 2022**

- Implemented Java-based SQL Relational Database Management System (RDBMS)
- Added support for join algorithms and query optimization
- Included multi granularity locking to support concurrent execution of transactions
- Implemented write-ahead logging and support for savepoints, rollbacks, and ACID compliant restart recovery

PROFESSIONAL EXPERIENCE

Strolleta Inc. | Software Engineer Volunteer **July 2021 - September 2021**

- Developed voice cloning app built with Python/Django where users record voice samples and generate text-to-speech in their own voice
- Improved existing voice clone program to effectively function for sentences longer than 100 words
- Utilized AWS Simple Storage Service (S3) bucket into heroku application to work around storage restraints
- Designed intuitive and generalized UI using HTML/CSS to make voice clone application more user friendly

LEADERSHIP AND EXTRACURRICULARS

Data Science Society | *Python/Scikit-learn* **August 2022 - December 2022**

- Created LogisticRegression classifier to distinguish between spam and non-spam emails
- Selected impactful heuristics to achieve accuracy of **89%** on test set
- Provided an example for new members to inspire creative project concepts

Data Science Society | *Jupyter/Python* **August 2020 - December 2020**

- Dedicated 10hrs/week to explore and process financial market datasets to generate a semester-long research project
- Collaborated with a diverse team of 6 to create a 25-slide final presentation deliverable for a bi-annual Symposium
- Utilized Jupyter Notebook to efficiently combine technical data to determine the performance of 10 NASDAQ stocks
- Created visualizations using matplotlib to effectively communicate results