

# YIYANG JIANG

Phone: +1(413)800-9028, E-mail: yiyangjiang0714@gmail.com

Address: 1443 Whitewood CT, San Jose, CA, 95131

## EDUCATION

---

Northeastern University  
Master of Science in Data Science

San Jose campus, CA  
09/2022

- Major course: Database Management Systems, NLP, Unsupervised Machine Learning
- GPA: 3.8

University of Massachusetts Amherst  
Bachelor of Science in Mathematics  
Minor in Computer Science

Amherst, MA  
09/2018 – 05/2022

Skills: Java, LaTeX, Matlab, Python(Numpy, Pandas, Matplotlib, Sklearn, Tensorflow, PySpark), SQL(MySQL, SQLite), SAS, Javascript, AWS, XML, Tableau

## WORKING EXPERIENCE

---

Data Science Intern, Hilton Worldwide Holdings Inc.

06/2022 – 09/2022

- Developed a Naive Bayes classifier for sentiment analysis of customer reviews. The model was trained on a balanced dataset of positive and negative reviews, with preprocessing steps including cleaning, tokenization, stopword removal, and stemming.
- The model was used to predict the sentiment of a given review, providing valuable insights into customer satisfaction and areas for improvement in the hotel's services.
- Created a user-friendly interface for the model, allowing team members to input a review and receive an instant sentiment prediction. The model parameters were saved using pickle for easy deployment and future use.
- This model significantly improved the efficiency of processing and understanding customer feedback. As a result, the hotel was able to make targeted improvements in areas identified by the model. This initiative received positive feedback from the hotel management team, recognizing the value of data-driven decision making in enhancing customer service.

Amazon Co., Ltd (Intern)

09/2021

- Introduced SQL primary language (where, having, group by, etc.) and SQL high-level language (window function, subquery, date function) for people without SQL foundations with PowerPoint.
- Transformed the Excel formula into Python language, built a blank Excel table to facilitate the subsequent filling of the calculated values, and employed the For loop to fill the values in the table of over 25,000 lines.
- Achieved multi-table consolidation and target columns filtering based on Amazon Financial Operations Account Receivable financial source data, as well as generated pivot tables in excel.

## PROJECT

---

Build a Data Warehouse and Mine Data (Project)

Supervised by Schedlbauer Martin, Professor

01/2023 – 05/2023

- Managed a 30MB XML document, created a corresponding DTD, and extracted data using R Studio to store it relationally in a SQLite database, effectively constructing a transactional database.
- Leveraged data from the transactional database to create an analytical database using a star schema in MySQL.
- Executed queries to extract facts from the MySQL analytical database, using the results for data visualization.

Predicting Movie Revenues (Project)

09/2022 – 12/2022

- Using Linear Regression and Random Forest with a sample of 45,466 movies, compare their prediction performance, and investigate the important features that have the most effect on movie box office revenue.
- Using cross validation to search for best parameters on normalized data increases the accuracy to 0.77.
- Examining the relationship between the number of famous actors and the movie revenues, and use the polynomial regression to predict hiring 6 famous actors are the most cost-effective.