

Jiaxin “Tammy” Tang

tammytongcn@gmail.com | 209-305-9802 | [linkedin.com/in/jiaxints](https://www.linkedin.com/in/jiaxints) | github.com/tammysilva

Summary

Seeking a Machine Learning Engineer position

An incoming full-stack data scientist with robust STEAM background and 4+ years of sufficient project experience in programming, data analysis and machine learning. Involves in **building data ETL pipeline, data cleaning, feature engineering, data visualization, predictive modeling and analytical reporting** with big data warehouse. Has a wide variety of working experience in e-commerce, education and finance industry. Passionate about data mining, self-motivated and able to collaborate in a diverse team environment.

Education and Certificates

Master of Science in Engineering/Data Science @UC Riverside | Riverside, CA Expected 12/2023

Related Courses: Machine Learning, Data Mining, Systems Engineering, Info Retrieval & Web Search, Database Management **current GPA: 3.9/4**

Database and Data Analytics Certificate @UC Santa Cruz Extension 06/2021

Related Courses: Data Analysis with R, Big Data Tools and Use Cases, Internet of Things, Big Data Applications, DB Design and SQL Programming

Programming in Python @De Anza College 06/2021

Related Courses: Python Programming, Advance Python Programming, Java Programming **GPA: 3.5/4**

Bachelor of Science in Science Education/Physics @Guangxi Normal University | Guilin, China 07/2013

Related Courses: Probability and Statistics, Linear Algebra, Calculus, Physics Experiments, Literature Retrieval, C++ Programming **GPA: 3.54/4**

Publication: Jiaxin Tang, Zizhen Zhao, *A New Determination of the Fusion Heat of Ice*, Journal of Experimental Science and Technology, 2012

Skills

Programming: Python (Sk-Learn, Pandas, Numpy, Beautifulsoup, TensorFlow, Pytorch), R (dplyr, pastecs, rcompanion), Spark, Java, Linux, LaTeX

Visualization: Power BI, Tableau, MATLAB, Matplotlib, Seaborn, ggplot2, Advanced Excel, PowerPoint, Adobe Creative Suite, GSuite

Database management: ManageBac, AWS, MySQL, PostgreSQL, dbt, Terraform, BigQuery

Machine Learning Techniques:

- Linear Regression, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, K-Nearest Neighbors, Principal Component Analysis (PCA), K-Means Clustering, Ensemble Learning (Random Forest, AdaBoost, Gradient Boosting, XGBoost)
- Exploratory Data Analysis, Regularization, Feature Engineering, Model Evaluation, Experimental Design, A/B Testing

Project Highlights

Movie Recommendation Engine Development in Apache Spark

- Built a data ETL pipeline to analyze movie rating dataset and conducted online analytical processing (OLAP) with Spark SQL.
- Implemented the Alternative Least Square model to provide personalized movie recommendations and developed user-based and item-based approaches to handle system cold-start problems.
- Conducted model hyper-parameters tuning with Spark ML cross-evaluation toolbox.
- Evaluated model by RMSE and compared to the models with K-Nearest Neighbors and Neural Network algorithms.

Financial Anomaly Detection and Risk Analysis

- Developed an algorithm in Python to predict fraudulent transactions and built an alert system to prevent fraudulent activities.
- Applied exploratory data analysis on 138K+ transactions, preprocessed data by removing duplicates, encoding categorical features and handling imbalanced labeled data by SMOTE.
- Trained supervised machine learning models including Logistic Regression, Random Forest and XGBoost via k-fold cross validation and found optimal model parameters via grid search.
- Selected the best model based on AUC and designed a real-time alert system to prevent fraudulent activities.

Natural Language Processing and Topic Modeling on User Review Dataset

- Clustered customer reviews into groups and discovered the latent semantic structures using Python.
- Preprocessed review text by tokenization, stemming, removing stop words and extracted features by Term Frequency – Inverse Document Frequency (TFIDF).
- Identified latent topics and keywords of each review for clustering.
- Trained unsupervised learning models of K-means clustering and Latent Dirichlet Allocation.

Relevant Experience

Licensed Financial Planner

04/2019 - Present

@Dream Builder Financial | San Jose, CA

- Interviewed and followed up with 100+ clients, provided professional recommendations and strategies on investment opportunities and risk mitigation, managed \$14 Millions in client investment portfolios and the current ROI is approximately 7%.
- Satisfied each client’s individual needs by providing financial products and services based on customer-centric and data-driven use cases, over 60% of my clients are through referral.
- Conducted data mining with various financial models from different financial providers to align each client’s financial objectives, risk tolerance, return on investment, developed overall personalized financial plan and devised presentation deliverables.
- Led a team of 10 to conduct marketing campaigns, contributed over 50% of the copywriting for the external advertisements, improved metrics gathering, drove rebranding strategy for 80% of the product offerings by collecting, analyzing and presenting data with MS Office Suite/GSuite.
- Provided trainings to the team on fraud detection and anti-money laundering, stay up-to-date on the latest fraud prevention strategies and techniques as well as suspected fraudulent activities reporting procedures in accordance with state/federal laws and regulations.
- Collaborated with case manager, data analyst and underwriter to bring analytical rigor and statistical methods to monitor clients’ long-term and short-term investment goals, as well as developed innovative solutions with marketing, engineering and UX personnel through agile framework meetings.