



SHIYI TAN

EDUCATION

- University of San Francisco, San Francisco, CA** 08/2015 – 06/2017
Master | Computer Science
- Yantai University, Shandong, CN** 09/2011 – 07/2015
Bachelor | Electrical information science and technology

WORKING EXPERIENCE

- Algorithm Expert - Yunquna Inc., Shanghai** 04/2022 - 01/2023
- Led the design and technical development of AI Hub products for digital marketing scenarios and managed the team;
 - Built up the team through a combination of mentoring and technical leadership, making others better and raise the bar for those around them.
 - Participated in all aspects of designing, researching, building and delivering data-focused products for our clients.

- Algorithm Expert - Yunquna Inc., Shanghai** 06/2020 - 04/2022
- Responsible for the AI algorithm of logistics big data, perform real-time data processing and prediction using logistics information of user orders on aliyun dataworks;
 - Responsible for the algorithm support of the shipping industry, predicting shipping container price for all lines, then use the price forecast to book containers effectively;
 - Develop recommendation algorithms based on user behavior data to recommend different types of products to users;
 - Lead the recommendation/forecasting team to explore the application scenarios of artificial intelligence in the shipping container industry and coordinate with the business department.

- Algorithm Engineer - Atzuche Inc., Shanghai** 08/2019 - 05/2020
- Responsible for the construction and implementation of the core recommendation system, as well as optimizing the comprehensive ranking algorithm of the search list to improve CTR and conversion rate;
 - Implemented automation of model training and prediction on AWS using SDK for python.
 - Continuously develop and optimize core models, which are applied to different business scenarios such as risk control and marketing (such as forecasting monthly KPI indicators(monthly order quantity)under the influence of the COVID-19);
 - Analyze operation/marketing-related business data and user data, form relevant reports, and provide data support to other departments;

- Machine Learning Engineer(AI) - Actiontec Inc., Shanghai** 11/2017 - 06/2019
- Built a medical AI system, helping doctors to diagnose lung cancer, using nodule detection and segmentation models (using the public datasets of NDSB and LUNA16)
 - Trained different models to predict the degree of malignancy of patient nodules on real lung CT images.
 - Participated in and be responsible for the development of distributed and embedded AI systems (Face Recognition), which provides deep learning models and APIs.

- Data Scientist Intern - CRTCL, San Jose, CA** 06/2017 - 10/2017
- Collect virtual currency data, perform visual analysis and pre-processing to obtain pattern of price fluctuations, and use different machine learning algorithms (Logistic Regression, SVM, etc.) to train/evaluate models and predict new data.
 - Used Twitter and Weibo's API to obtain user data, use Python for data processing and NLP, extract relevant content that affects the price of virtual currency, and conduct a second round of model training.

- Data Scientist Intern - SnapLogic, San Mateo, CA** 01/2017 - 06/2017
- Solved ticket prioritizing issues for customer team to improve their efficiency.
 - Contributed software engineering expertise in the development of data exploration and data processing.

- Built intelligent agent for snaplogic cloud platform, providing recommended snaps for users.
- Provided different trained models to predict future data in a weekly-updated way through machine learning.

PROJECT EXPERIENCE

Intelligent Customer Service System

05/2023

This system is to generate intelligent response for customers' questions, we provided large language model(LLaMA) running on the cloud.

- Clean and preprocess customer historic chatting data.
- Developed and implemented training plans for LLaMA models to ensure they meet specific requirements.
- Collaborated with backend developers to build customer service chatting system.

Marketing AI Hub ConvertLab Inc., Shanghai

10/2022 - 01/2023

- Led the development of an AI platform product based on a big data framework, be responsible of architecture and technology selection;
- Analyzed various marketing scenarios and developed generic machine learning models, as well as adapting models to fit different scenarios' data;
- Worked and collaborated with different cross-functional teams end-to-end.

E-commerce Recommendation System ConvertLab Inc., Shanghai

04/2022 - 01/2023

- Migrated the recommendation system for an oversea e-commerce platform (from Alibaba Cloud to AWS) ;
- The project involved migrating data storage, AI architecture, and algorithm models, improving performance, and scheduling tasks;
- Adjusted the click and conversion rates for different categories while ensuring consistency, and redeveloped some of the platform's proprietary algorithms originally developed by Alibaba Cloud team.

Intelligent Cargo System Yunquna Inc., Shanghai

07/2021 - 04/2022

This system is to predict container vessels estimated time of arrival to a destination and provide valuable logistics information.

- Clean and preprocess AIS data of container ships on aliyun dataworks, then find all port and berth location using clustering algorithm and calculated mooring duration/historical berth list.
- According to the point-to-point shipping schedule data, matching similar ships in AIS data, so as to predict the number of ships (by type) calling at all ports in the future.
- Train the prediction model to effectively predict the future congestion of the port using behavior and static features of all ships

Email Pushing Project Yunquna Inc., Shanghai

01/2021 - 08/2021

A recommendation system which ranks all cargo container items based on users behaviors and items features.

- Responsible for the product recommendation content on the homepage and search page of the official website, as well as different types of email pushing content for corporate users and individual users.
- Clean and analyze the user's behavior data on aliyun dataworks, and give different weights to different types of behaviors (click/order/transaction) to get the user's score for each item.
- Then train the model based on the user's static data (company address/age/gender, etc.)/behavior data and product labels (port of departure/port of destination/airline, etc.), and then predict the scores of all recalling products, and rank them according to the scores;

CargoWare System Yunquna Inc., Shanghai

06/2020 -03/2021

This system is designed to rank all cars in the database for various users displaying by a particular order on search and find-a-car pages.

- Analyze and process the e-commerce price data of different carriers, including releasing time/price deviation/space type, etc.

- Provide the crawler team with the optimal port combination of each carriers and the perfect time period to collect freight rate data.
- According to the historical transaction price of users and the crawled freight rate, as well as related factors (SCFI index, historical diesel price, historical weather data, port congestion data, etc.), train a time series model to predict future container prices.

KPI Predicting Atzuche Inc., Shanghai

01/2020 - 04/2020

This project is to predict KPI for marketing department under the influence of COVID-19.

- Collect historical data (monthly order quantity) and finished data cleaning and processing.
- Collected predicted COVID-19 data based on medical models and processed as features.
- Trained statistical models and machine learning models to predict the future monthly order quantity of next year.

Car Recommendation System Atzuche Inc., Shanghai

08/2019 - 01/2020

This recommendation system is used to recommend favorite car renting product to customers based on their behaviors.

- Collect and clean user click/favorite/order data and define a formula to calculate the score of cars for different users.
- Generated recall dataset according to user's company information and chatting text data.
- Trained a model to predict the score for a car list based on user's static/behavior data and product labels.

Distributed and embedded AI system Actiontec Inc., Shanghai

05/2018 - 06/2019

This AI system is used to provide non-AI developers with deep learning models and APIs to easily train a model and do predictions on our cloud platform.

- Designed the technical architecture for the system.
- Using AutoML to train the best model for face recognition/detection.
- Applied first-class paper into a real project, making it works in real scenarios.
- Lead the team to develop more and more Apps and web services for end users.

Recommendation System for Story Sticker App Actiontec Inc., Shanghai

09/2018 - 12/2018

The Project is designed for the development of a social application called story sticker. this app is divided into four modules - home page, explore, message and user management. People can easily post/follow/share/like/comment others and their posts.

- Established a recommendation system (Collaborative Filtering)for story sticker's users.
- Analyze user data and post data collected by story sticker app.
- Trained a model to predict user's interested keywords and push related posts and users for them (CF).
- Produced some AI bots as virtual users trying to interact with actual users.

Medical AI - lung cancer prediction Actiontec Inc., Shanghai

11/2017 - 04/2018

A medical AI system which assist doctors to predict lung cancer malignancy level, doctors can use web service to upload CT images and get feedbacks

- Independently responsible for the development of medical AI system - trained nodule detection and segmentation models (based on the public datasets of NDSB and LUNA16)
- Preprocessed various kinds of lung CT images to fit the model, data cleaning and data normalization.
- Applied AutoML into model selection and finetuning parts, to get the best models and scores.

Intelligent Assistant System SnapLogic, San Mateo

01/2017 - 05/2017

Intelligent Assistant System is a recommendation system for end users using SnapLogic cloud data integration tools. This system will predict next snaps according to users' previous behavior, to Improve user experience.

- Contributed to data preprocessing and integration part of recommendation system
- Cleaned data and normalized data for model training. (DT/Random Forest)
- Integrated model training and preprocessing part into the whole system.

SKILLS

Programming Languages: Python, SQL, Java, R, Nodejs

Others: Linux, MacOS, Hadoop, AWS, Docker, Deep Learning, Machine Learning