

Zixi Wang

linkedin.com/in/zixiwang0508
wzx952013@gmail.com • (858) 335-8506 • Los Angeles, CA

EDUCATION

University of Southern California

MS in Electrical Engineering - Data Science & Engineering

January 2018 – December 2019

University of California, San Diego

BS in Electrical Engineering (Provost's Honor)

September 2013 – June 2017

WORK EXPERIENCE

CarParts.com (previously US Auto Parts Network Inc.)

February 2020 - Present

Data Scientist, Full time

Torrance, CA

Machine Learning | XGBoost | Prophet | Linear Regression | Google Analytics | Big Query | SQL

- Analyze sales seasonality and general trending and build time series model to forecast overall units sold. Fine-tune Prophet model and achieve error rate of 0.8% at weekly granularity. Present and explain results to non-technical audience
- Developed linear models to study competitors' price dataset and predict competitor price for 170k SKUs, with R^2 score > 0.99
- Research customer behavior with price change to optimize gross profit with Google Analytics data and Big Query
- Maintain existing XGBoost models' lifecycle in Agile environment to keep up to date with recent changes

Web App Development | Python | Bokeh | Flask | Data Visualization | Web Development | SQL

- Create from scratch an interactive dashboard application using Bokeh and Flask so stakeholders can monitor financial KPI as well as machine learning models' performance
- Write scripts to scrape internal price changes and store them as historical data for pricing strategy analysis

Data Engineering | Apache Airflow | SQL | MongoDB

- Manage and automate workflow with Apache Airflow, including ETL, report generation and machine learning model deployment
- Designed schema and maintain SQL and NoSQL (MongoDB) database to store scraped data and application data

DevOps | Docker Compose | Kubernetes

- Provide support for platform engineers and develop prototype containers using Docker Compose and Kubernetes
- Deploy ML models, dashboard application and report automation in on-premise Kubernetes clusters

Data Analysis | Excel | Hypothesis Test | Python

- Create analytics tools to keep track of shipping operation at warehouses and deliver recommendation on container selection, reducing shipping cost by 15% YOY
- Collect data from multiple sources and communicate with stakeholders to understand customer behavior, validate long-held hypothesis and optimize decision making process

Leadership

- Facilitate adoption of state-of-the-art technology into current production environment; lead weekly brainstorming sessions to address new challenges and create project roadmaps
- Represent company in big data conferences or AI webinars, such as Spark + AI Summit 2020

CarParts.com

Fall 2019

Data Scientist, Intern

Torrance, CA

- Optimized operating cost at 3 distribution centers by analyzing historical sales data and providing weekly predictions of sales and required labor using machine learning techniques (XGBoost) with standard deviation of error of 1.4 labor hours
- Visualized 3-year historical sales data, identified sales pattern through exploratory data analysis and extracted 40+ features from times series data for machine learning model using MySQL and Python

Maryland Transportation Institute

Summer 2019

Data Science Intern

College Park, MD

- Established machine learning pipelines at 10TB scale, including data preprocessing and wrangling by integrating geospatial and time series data, and feature engineering to improve decision tree model performance by 5%
- Alleviated class imbalance using Bootstrap and SMOTE, and converted high-cardinality categorical features by target encoding
- Overcame overfitting by adding regularization to machine learning model and transformed dataset
- Contributed to publication of proposal of enhancing triage decision of high-risk elderly trauma patients via machine learning techniques in Transportation Research Board

Deep Learning Projects

Part-of-Speech Tagging | RNN | Natural Language Processing

- Designed and developed Bidirectional LSTM model to tag part-of-speech of sentences
- Fine-tuned hyperparameters and achieved accuracy of 97% in given test languages, top 20 (out of 240 students) in the school competition