

Vince(Wensi) Pan

A data scientist digging into the probabilistic model and machine learning. Experienced Data Analyst with a demonstrated history of working in the retail industry.

EXPERIENCE

Senior Data Analyst 2015 - 2017
Biostime Group Guangzhou China
- Analyzed key commercial investments including marketing campaigns, media investments, etc.
- Implemented and augmented statistical analyses and modeling techniques with SQL and R to support business decision making.
Grows: Database queries, Project Management

Marketing Analyst 2014 - 2015
IKEA Guangzhou China
- Discovered local market insight, including customer segmentation, RFM analysis, and MMM model.
Grows: Business Modelling, Business planning

Marketing Researcher 2012 - 2014
MSR-China Beijing China
- Provided data research plans and data analysis results to clients, including Microsoft, Suning Group, etc.
Grows: Marketing Research, Coordination

PROJECTS

Don't talk, emoji it
An unsupervised Word2Vec model exploring the usage similarity between emoji and English words on twitter. Provided a real-time APP on AWS cloud and collected user feedback to server database.
Model: Ensembled Word2Vec CBOW-Skip-Gram model
Tools: Python(BeautifulSoup, NLTK, Gensim, Pydash), AWS EC2, PostgreSQL

Analysis of Customer Lifetime Value(LTV)
A comprehensive probabilistic model for evaluating customer lifetime value using Spark and python, providing access to individual customer spending and churn forecast.
Model: Pareto/NBD Gamma-Gamma Model
Tools: Python(PySpark, Lifetime, SKlearn-linear regression, random forest, support vector machine), Spark, SQL

Local tracker of COVID-19
Web-scraped the latest figures from daily press releases of the LA public health department. Provided an interactive map showing the density and trend of COVID19 development on local communities.
Model: Data Visualization with GIS mapping
Tools: Python(BeautifulSoup, Folium, Altair)

Media Mix Model(MMM)
Provided evaluation of media expense efficiency in terms of ROI and offered media distribution optimization suggestions.
Model: Multivariate linear regression
Tools: R, Oracle SQL, Excel

Phone: (626) 545-8466
Email: wensi.pan@gmail.com
Github: <https://github.com/constiny>
Linkedin: [linkedin.com/in/wensi-pan](https://www.linkedin.com/in/wensi-pan)
Portfolio: <https://constiny.github.io/>
Location: Los Angeles, CA

SKILLS

Database
Postgres SQL, MongoDB, Spark, AWS S3, Oracle SQL

Machine Learning
Linear/Logistic Regression, Regularized Regression, Random Forests, Boosting, NLP, Probabilistic Models, Time Series Analysis, Neural Networks, CNN, RNN, Autoencoder

Others
Git, Docker, R

PYTHON PACKAGES
Data/Text Processing
Numpy, Pandas, NLTK, BeautifulSoup

Modeling
Statmodel, TensorFlow, Keras, SciPy, Sci-Kit Learn, Lifetime, Gensim, Pymc3

Presenting
Matplotlib, Seaborn, Plotly, Flask, Dash, Folium, Altair

EDUCATION
Data Science Cert Spring 2020
Galvanize / HackReactor Los Angeles

MS in Statistics 2012
Uppsala Universitet Sweden

BS in Statistics 2010
BS in Economics 2010
South China Normal U China

LANGUAGES
English, Chinese (Mandarin/ Cantonese)