

联系方式

www.linkedin.com/in/bismarckliu
(LinkedIn)

热门技能

R programming
MySQL
Apache Spark

Languages

Chinese (Native or Bilingual)
English (Professional Working)

Certifications

Managing Big Data with MySQL
Introduction to Big Data (2015)
Apache Pig 101
Accessing Hadoop Data Using Hive
Neural Networks and Deep Learning

Honors-Awards

Kaggle Data Mining Competition
BOSCH Production Line
Performance

Publications

Polyhedral Fe₃O₄ nanoparticles for
lithium ion storage

Wenyue Liu

Data Scientist at Kelly Blue Book
Santa Ana, California

个人简介

Versatile data scientist with 2 years of industrial hands-on experience in predictive modeling, large scale end-to-end machine learning, modeling and reporting data pipeline.
Extensive programming skills in Python, R, Scala, SAS, Hive (SQL), Spark and AWS platform

工作经历

Kelley Blue Book
Data Scientist
2018 年 3 月 - Present (2 年 1 个月)
Greater Los Angeles Area

- Used vehicle pricing model:
 - o Built end-to-end machine learning pipeline in AWS for used car valuation.
 - o Improved valuation MAPE score from 6.2% to 5.5% by combining the decision tree and Neural Network models. This method boosted model coverage from 97% to 99.9%.
 - o Used Vehicle Embedding to reduce vehicle representing sparsity to let the model converge.
 - o Trying to use the Transfer Learning technique to transfer knowledge learned from US data to international business which has much fewer data.
- Vehicle reliability score project:
 - o Developed vehicle service mapping tool by using a fine-tuned BERT model in TensorFlow. The language model achieved 92.4% accuracy in the vehicle maintenance and repair classification task.
 - o Designed and implemented the durability score by predicting vehicle in operation ratio in five years.
 - o Using repair and maintenance data to build other main components to support the reliability score.
- Designed and built a cross-account/platform internal price comparison tool by collecting vehicle price data across IBM and RedShift database and RESTful API. Built Tableau dashboard for interactive display.

Siemens PLM Software

Data Scientist

2017 年 6 月 - 2017 年 12 月 (7 个月)

Charlotte, North Carolina

- Built data cleaning and predictive analysis pipelines on customer's data using Impala, Spark SQL, Spark MLlib.
- Enhanced pipeline efficiency by 1200% by optimizing config/path to handle billion level data within 20 mins.
- Deployed distributed deep learning framework on PySpark through Shell operations and ran a Generative Adversarial Networks (GAN) demo on it.
- Worked with engineer team on to implement research projects such as model deployment, out-of-box functions.

University of North Carolina at Chapel Hill

Research Assistant

2016 年 10 月 - 2017 年 6 月 (9 个月)

Kenan-Flagler Business School

NetEase

Data Analyst Intern

2016 年 5 月 - 2016 年 8 月 (4 个月)

Guangzhou, Guangdong, China

- Analyzed user consumption behaviors via k-means, DBSCAN in R to segment anchors into separate clusters, which helped to revise live gifts launching strategy, and notably increased gift sales income.
- Constructed a new anchor rating system based on anchor's multiple income behaviors; Used random forest algorithm to predict anchor's rating based on their non-income features.

教育经历

University of North Carolina at Chapel Hill

Master's degree, Statistics · (2016 - 2017)

Sun Yat-Sen University

Bachelor's degree, Chemistry · (2010 - 2014)