Shuai Gao

E-mail: leoshuaigao@163.com
Tel: 18811580546 | WeChat: Demon_First
Blog URL: leoshuaigao.com

Education

Royal Melbourne Institute of Technology Melbourne, Australia
MSc Data Science 2017.07-2019.07

Hebei University of Technology City College Tianjin, China
BSc Surveying and Mapping Engineering 2012.09 – 2016.07

Internships

Coles, Data Analysis Depart., Senior Analyst Melbourne, Australia | 2019.03 - 2019.06

- Coles is Australia's second largest retailer. Used unsupervised machine learning algorithms for user information grouping based on the purchase behavior and personal information statistics of users in Australia;
- Responsible for data processing, including extracting data from SQL, cleaning data, feature selection and statistical analysis as well as data visualization;
- Results: Scored the user value according to the user's RFM (recent purchase, purchase frequency, purchase expenses), which was recognized by the team. Grouped the information such as discount tendency, quality requirement, purchase quantity and prices, etc., laying the basis for advertisement placement;

IT MAN, Machine Learning Project, Data Scientist Intern Melbourne, Australia | 2018.06– 2019.02

- Independently researched the Beijing job market and used machine learning algorithms for salary forecasting;
- Grabbed job related data from 51Job, performing statistical analysis and feature selection for data cleaning and processing;
- Used machine learning algorithm (deep learning, Random Forest, SVM, etc.) to predict job salary levels;
- Results: The salary level can be basically predicted given a description of the position information, and the prediction accuracy reached over 80%;

Projects

Cloud Computing - Cloud Platform Data Analysis & Webpage Deployment, Personal Project Melbourne, Australia | 2019.01-2019.02

- Description: Used Google's cloud computing platform for data extraction, cleaning, visualization and other processing and then immigrated to the cloud platform, used multiple third-party APIs to achieve faster and easier access to more statistical information;
- Independently completed cloud platform deployment and webpage design in the project to solve various platform incompatibility issues during the deployment process;
- Used the distributed BigQuery for data extraction, cleaning, storage, visualization and other processing steps and wrote code by Python;
- Highlights: Used the distributed computing technology (BigQuery) to reduce the time consumed in data extraction. Got the praise of the project review professor;

Australian Workforce (Full-time/Part-time) Male-Female Ratio Shinny App Data Visualization Melbourne, Australia | 2018.10– 2018.11

- Description: Visualized the historical trends of the number of employees in various industries/positions in Australia and analyzed trends and proportions of male and female practitioners;
- Completed data selection, cleaning, grouping, and created interactive charts independently;
- Found out through the interactive chart that women still occupied a small proportion in most industries. Although the total number of male and female employees in some industries was almost the same, men were more likely to hold full-time jobs; developed interactive light applications through Shinny app to share visualization results through url;
- Project website: https://leogao.shinyapps.io/assignment3/; ranked top 5% in the class;

Comparison of the Impact of MapReduce Algorithm and Cluster Number on Performance, Team of 2 Melbourne, Australia | 2018.09– 2018.10

- Analyzed the speed difference between Pair and Stripe's literal statistical algorithm with and without the combiner and partitioner; analyzed the effects of the two algorithms (pair and stripe) on different data sizes and different cluster numbers;
- Wrote MapReduce in Java, compared the impact of combiner, partitioner data size and cluster number on the two different algorithms;
- Chose the data, cleaned non-English characters, implemented two different algorithms, and created visualization charts;

- Results: ① Pair performed similarly with Stripe when the data size was small. As the size of data amounted, Stripe performed much better; ② Combiner optimized the processing speed of Pair; ③ The speed of Stripe algorithm was faster than that of Pair when the Cluster increased;
- Analyzed the results, wrote a summary report and attended the defense, and finally got a full mark;

Time Series Analysis - Bitcoin Price Forecast, Team of 4 Melbourne, Australia | 2018.03– 2018.05

- Extracted data of the historical prices of bitcoin, used R language for time series analysis, established a ARIMA+Garch model to predict the price of the next 10 days;
- Completed data co-correlation analysis, extracted the parameters required by arima and garch models, and performed model optimization and prediction;
- Difficulties: The trends and variance changes of Bitcoin prices varied widely;
- Results: The price prediction accuracy was very high (±30) 10 days after the assignment was submitted. Got a full mark;

Competition

EY NEXTWAVE DATA SCIENCE COMPETITION 2019

- ranked 90th in the world and fifth in China among 4,500+ participants from 470+ universities in 15 countries

Skills

Language: Chinese (native), English (proficient);

Computer language: Python (2 yrs), R (2 yrs), Java (2 yrs), SQL (2 yrs); statistical software: SaS.