

Daniel T. Liao

Dallas TX | dtliaomr@gmail.com | (469) 388-8097

LinkedIn: <https://linkedin.com/in/danieltliao>

GitHub: <https://github.com/dtliao>

Personal Website: <https://dtliao.github.io>

Tableau Public: <https://public.tableau.com/profile/daniel.tliao>

TECHNICAL SKILLS & CERTIFICATIONS

Big Data Ecosystems:	Hadoop, MapReduce, HDFS, Flume, Hive, Impala, Pig, Sqoop, Spark
Programming and Databases:	Python, R, SAS, SQL, Bash
Software and Analysis Tools:	MS Excel, Tableau, Microsoft Power BI, SSIS, Git, Selenium, AWS
Web Development:	HTML5, CSS3, JavaScript, React, NodeJS
Certifications:	Tableau Desktop Specialist, Google Analytics, Microsoft Office Specialist
Achievement:	Covid-19 Data and Technology Contest Finalist

EDUCATION

The University of Texas at Dallas , Dallas, TX	December 2020
Master of Science in Information Technology and Management concentration in Data Analytics	GPA 3.52
<ul style="list-style-type: none">Coursework: Applied Machine Learning, Big Data, Programming for Data Science, Statistics and Data Analysis, Cloud Computing, Data Management, Business Data Warehousing, Business Analytics with R, Objected-Oriented Programming in Java, System Analysis and Project Management, Predictive Analytics Using SAS	

PROJECT EXPERIENCES

King County House Sales Data - Regression | Python - Pandas, NumPy, Matplotlib, Seaborn, ScikitLearn, Keras

- Employed a Gradient Boosting Regression Model to predict the house sale prices sold between May 2014 and May 2015, and enhanced the R-squared score from 65.12% to 90.29% through feature scaling techniques
- Applied machine learning algorithms including Regularized regression model, Polynomial Regression, KNN, SVM Regression and other ensemble learning techniques using Python Scikit-Learn
- Tuned and optimized hyperparameter selections using grid search and k-fold cross validation
- Compared the performance of the model with that of Pandas Keras ANN

University Library Management System | SQL, Oracle APEX, MS Visio

- Designed and implemented a library book lending software using Oracle SQL Database
- Developed a client-server application to add, borrow, remove books and calculate late fees using Oracle APEX
- Designed a library system ER Diagram with 8 tables following normalization rules using Microsoft Visio
- Optimized SQL select speed by over 90% by creating composite index on complex lookup queries

Truck Fleet Risk Factor Analysis | Hadoop, Hive, Pig, R, Shell, Tableau

- Created process flow in Hadoop ecosystem to calculate risk factors using Pig script. Then imported, created, and loaded truck fleet data into Hive tables
- Implemented a Logistic Regression model using R to identify risky drivers in California
- Created interactive dashboards through Tableau that pinpoints dangerous commercial truck drivers identified in regression model

Face Recognition App | HTML, CSS, JavaScript, React, Node.js, Express, PostgreSQL

- Designed a web app that identifies faces in images using React, NodeJS, PostgreSQL and the Clarifai API
- Designed a login system and enhanced security by hashing passwords and using knex to avoid SQL injections

Analysis of Customer Churn in Telecommunication Sector - Classification | SAS

- Implemented Logistic Regression Model to predict the likelihood of customer left the company in the previous month, and enhanced the performance with an 85% accuracy through stepwise model selection approaches
- Analyzed the behavior of telecom customers, and provided marketing suggestions for current and potential customers to maximize the profit of the company

ADDITIONAL INFORMATION

- Eligibility:* authorized to work for any employer