Daniel T. Liao

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

LinkedIn: https://linkedin.com/in/danieltliao *Portfolio:* https://dtliao.github.io
GitHub: https://github.com/dtliao *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
Cloud Services:	Azure Data Factory, Azure Storage Solutions, Azure HDInsight & Databricks
Software and Analysis Tools:	MS Excel, Tableau, Microsoft Power BI, SSIS, Git, Selenium
Web Development:	HTML5, CSS3, JavaScript, React, NodeJS
Certifications:	Tableau Desktop Specialist, Google Analytics, Microsoft Office Specialist

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

- 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

Covid19 World Data | Azure - ***Data Factory, Storage Solutions, HDInsight, Databricks, MS PowerBI***

- Built a data engineering solution architecture using Azure technologies stack for data analysts and data scientists to create reports of covid-19 trends and prediction of the spread of virus
- Integrated data from HTTP clients, Azure Blob Storage, and Azure Data Lake Gen2 using Azure Data Factory. Then created pipelines using control flow activities such as Lookup, Validation, ForEach, Delete, IfCondition, Get Metadata
- Created and executed transformation logic and copy data from Data Lake Gen2 into Azure SQL Database using Data Flows, Azure HDInsight, and Azure Databricks Notebook Activities in the ADF pipelines
- Implemented the Orchestration, Monitoring tool and Log Analytics in the ADF pipelines using triggers, alerts, and reporting of metrics

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 the state of 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

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

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