Giang Nguyen

Baldwin Park, CA 91706 | **Phone**: (626) 295-7032 | **Email**: gcnguyen@cpp.edu. LinkedIn

EDUCATION

- **California State Polytechnic University, Pomona (CPP)** August 2021- Present
  - Major: Business Analytic GPA: 3.85
  - Degree: Master of Science
  - Relevant course: Data Mining in Business Analytic | Database Management in Business Analytic | Social Media Text Mining | Statistic Essentials for Data Analytic |
- **Mount San Antonio College (Mt Sac)** February 2019- June 2021
  - Major: Business/Accounting GPA: 4.0

TECHNICAL SKILLS

- **Power BI | Python | Social Media Mining | Data Warehouse | Data Cleaning | Data Visualization | Natural Language Processing | Machine Learning | Scikit-learn | NumPy | Predictive Modeling | Sentiments Analysis | Regression and Decision Analysis | Qualitative Analysis | Quantitative Analysis | A/B Testing | MySQL | Microsoft SQL | RStudio | IBM SPSS | Microsoft Word | Excel | GitHub | Hadoop | Hive | Spark | Map-Reduce.**

RESEACH EXPERIENCES

- **Twitter Analysis Modeling (Big Data)** March 2022 - Present
  Dr. Mohammad Salehan, CIS Department, Cal Poly Pomona
  Research Student Assistant
  - Successfully created and optimized Twitter API to extract **1.3 Millions** of Tweets from multi social media channels in 10 years.
  - Applying NLP to modify the data to improve the data's readiness and availability for further sentiments analysis.
  - Using R to test Polarity Hypothesis by using Negative Binomial and Poisson Regression.
  - Performing quantitative analysis on big data set for hypothesis testing and research paper thesis.

PROJECT

- **Fake Reviews Detection Using Machine Learning**
  - Successfully extracted **250k** data from Yelp platform, JSON files including users, reviews, ratings and products.
  - Created, trained, and tested model to detect fake reviews using Scikit-learn package.
  - Successfully predicted and detected **90%** fake reviews from the Logistic Regression algorithm.
  - Calculated accurate result using Logistic Regression, Naïve Bayes and Random Forest classification algorithms for under-sampling data
- **Data Modeling and Data Warehouse Project, Bookstore**
  - Successfully created the database, implemented the raw data into tables in MS SQL Server and developed the data warehouse.
  - Lead a team of four to design the database by creating the conceptual and logical design.
  - Identified the business domain and business rules for business process.
  - Used Power BI to create the visualization of supplier tracking, customer trends and book sales.
- **Analyzing House Sale at King County, Seattle (Github)**
  - Applied many libraries such as NumPy, Pandas, Seaborn, Matplotlib to analyze the large housing data.
  - Utilized Python to clean, modify, organize, and create visualization on the data.
  - Applied machine learning algorithms with linear regression to predict house sale price.
  - Applied statistic methods to find out correlation between attributes.
- **Improving the Quality of Hotel Banquet Staff Performance**
  - Used the Statistic Package for the Social Sciences (SPSS) of IBM to analyze the quality of staffs basing on the survey, training program and the hiring tools.
  - Applied statistic method (Cronbach's Alpha) to find out the correlation coefficient between the attributes.
  - Created the data visualization to observe the reliability of customers for the banquet staffs' service.

WORK EXPERIENCES

- **Internship- working in Human Resource Department** January 2018 – June 2018
  - Attended training presentations and career workshops in university
  - Collected data and analyzed data for Program Manager during job interview process.
  - Interacted with the other department, especially working with the Food and Beverage department.