MINH PHAM

Garden Grove, CA (714)-925-9026
https://github.com/minhpham123
mtp.tan.pham@gmail.com

Seeking to obtain an entry-level position to apply acquired skills, knowledge and education in real world projects

I am very eager to learn new techniques and skills with data and is currently in pursuit of the IBM Data Science certificate within Coursera.

SKILLS

Programming: Proficient: RStudio, Python(Pandas,Numpy)
Programming: Beginner/Intermediate: SAS, SQL(MySQL), Latex, Excel, HTML, CSS, JavaScript
Statistics and Probability: Hypothesis Testing, Bayesian/Frequentist Data Analysis, Time Series Analysis, Data Visualization, Mathematics, and Nonparametric Statistics
Machine Learning (In R):
Supervised Learning : Linear/Logistic/Lasso/Ridge Regression, K-nn, Decision Trees, QDA, LDA, Random Forest
Unsupervised Learning : Support Vector Machine, Neural Network
Languages: English, Vietnamese

EDUCATION

University of California Santa Barbara *September 2018 - June 2022*
B.S. in Statistics and Data Science, Cumulative GPA : 3.33
Applicable Classes : Machine Learning, Statistical Data Science, Probablity and Statistics (A, B, and C), Regression Analysis, SAS Base Programming, C++, Python, Bayesian Data Analysis, Time Series Analysis, Nonparametric Statistics, Design of Experiments

WORK EXPERIENCE

University Of California Santa Barbara *September 2021 - March 2022*
Technical Support
- Upheld maximum organization by properly sorting orders and inventory on a database containing over 4000 Professors and TAs
- Responsible for day-to-day front desk operations for providing media and classroom equipment to UCSB faculty and organizations, as well as technical support for computers and other equipment for 100+ classrooms
- Set up and broke down equipment for client events and supervised functionality of equipment during events

PROJECTS

Machine Learning Project : Model building and Analysis of Poverty within the US *October 2021*
Data Cleaning + Model Building
- Lead a team to clean and analyze two data sets : Census (3142 rows with 31 features) and Education (3143 rows with 42 features) in R markdown
- Conducted PCA / clustering with complete linkage for dimensionality reduction of datasets and created dendrograms for visualization
- Performed binary classification of Poverty feature and compared the results from multiple machine learning models and concluded that employment was the most important factor in determining the state of poverty within a given area followed by skin color

Time Series Project : Data Forecasting of Beer Production in Australia *November 2021*
Data Forecasting
- Performed transformations and decomposition of a dataset containing 154 observations in order to achieve stationarity and invertibility of time series data
- Found candidate SARIMA models and performed diagnostic testing of candidate SARIMA models by using various graphs (QQ-plots, ACF, PACF) and tests (Shapiro-Wilk, Box-Pierce, Box-Ljung, Mcleod-li / checking for roots) and concluded that beer production will continue to play a huge role in Australia's economy due to the forecast showing that beer production in Australia inherently increases each year