ANH PHAM

duyanh.phamhoang96@gmail.com • 312-805-0858 • Chicago IL 60605 • Portfolio • LinkedIn

PROFESSIONAL SKILLS

- Computer skills: Excel, Access, Word, PowerPoint, Azure Data Studio, Azure Machine Learning Studio Bloomberg Terminal, SAP
- Programming: Python (NumPy, Pandas, Scipy, Scikit-learn, TensorFlow, PyTorch, BeautifulSoup, NLTK), R (Tidyverse, RANN), Git, PostgreSQL, SAS
- Visualization and Statistical Software: Tableau, Power BI, Python (Matplotlib, Seaborn, Plotly, Wordcloud), R (Ggplot2), SAS
- Big Data: Hadoop, Hive, Spark, SQL Server Management Studio, Azure Synapse
- Analytics, Statistical Analysis, Machine Learning, Financial Modeling, Quantitative Analysis, Security Analysis, Problem Solving, Valuation, Research

PROFESSIONAL EXPERIENCE

MIDTRONICS INC. — Chicago, IL
Data Scientist — *Apr 2021 - Present*

- Collaborated with Engineering team to understand electrical physic and to detect fraud and outlier values that point out the core features that impact on vehicle battery decisions
- Applied physical math and derivative function for creating new features that increase the model performance
- Built and tunned XGBoost algorithm independently (97% accuracy score with 5% cross entropy score) on Azure Machine Learning Studio to detect vehicle battery decision that helps utilize Midtronics battery charger and tester tools, which is one of the most important projects of the company in 2021
- Deployed model to production with Software Engineering team through Azure Web App to utilize Midtronics' Battery Management Information System
- Build visualization dashboards with Power BI to help the marketing team understand the trending of the market and the sale of the company

FIVERR — Chicago, IL
Freelance Data Analyst — *Jun 2020 – Apr 2021*

- Provided data processing, data visualization, and data extracting insights from customer requests using Python, R and SAS
- Applied time series analysis, predictive analysis, and supervised learning (model preparation, classification problems) by applying machine learning algorithms (Logistic and Linear Regression, Decision Trees, Random Forest, SVM) for utilizing dataset
- Strengthened predictions by developing hypermeter optimizations for the dataset and established suggestions for clients to get better predictions in the future

ANALYTICS PROJECT EXPERIENCE

AMAZON FINE FOOD REVIEW — *Jan 2021*

- Utilized Sparse Matrix by using Python packages (NumPy, Pandas, Scipy) to design recommendation systems based on food item popularity and users' ratings
- Helped merchandises and distributors to understand how the systems impact on buyers' experience with RMSE scores
- Applied text processing with NLTK package and Scikit-learn packages (TfidfVectorizer, CountVectorizer, Gensim) to develop the sentiment analysis that predicts positive and negative reviews by utilizing machine learning technique (Logistic Regression and Bernoulli Naïve Bayes)
- Created clustering model with K-mean to extract top words that impact the sentiment analysis and applied t-SNE to plot those words with Plotly package
- Developed deep learning model to learn the dataset better with ANN and RNN – LSTM by applying TensorFlow.Kera and tunning model to gain higher accuracy scores, which are 3% accuracy improvement compared to traditional models

AIRPLANE CRASHED — *Nov 2020*

- Utilized Python packages (Pandas, NumPy, Searborn, Matplotlib) to develop a story of the airplane crashed trending based on time, regions, operators, and aircraft
- Built clustering models (K-Mean, Hierarchical, DBSCAN) to cluster airplane crashed and texting on the dataset by utilizing machine learning technique (Scikit-learn) and evaluated the best model based on a loop function of Silhouette Coefficient scores
- Applied dimension reduction (PCA, t-SNE, UMAP) to visualize accuracy results

LENDING CLUB'S LOAN DATA FROM 2007 TO 2011 — *Oct 2020*

- Developed data story efficiently and effectively by cleaning out the unnecessary features and pointing out the core variables that have impacts on loan prediction models
- Built models in Python to improve a loan prediction repayment and analysis to the next level by utilizing machine learning technique (Scikit-learn) and by incorporating a strategy for hyperparameter tuning (GridSearchCV) to Radom Forest Classifier and Gradient Boosting Classifier, which gained almost 12% higher accuracy scores and ROC scores than traditional models

EDUCATION

DEPAUL UNIVERSITY, KELLSTADT GRADUATE SCHOOL OF BUSINESS — Chicago, IL
Master of Science in Business Analytics: Data Science — *Jun 2021*

- Dean's Scholarship
- GPA 3.78 / 4.00
- Vice President of Kellstadt Business Analytics Organization

THINKFUL, INC. — Chicago, IL
Data Science Program Certificate — *Feb 2021*

LIBERTY UNIVERSITY, SCHOOL OF BUSINESS — Lynchburg, VA
Bachelor of Science in Business Administration: Finance — *May 2018*