JENNIE TRAM LE

35 W 33rd St., New York, NY| (347) 449-2616| **tle35@fordham.edu**
Github: https://github.com/bigforehead| **Tableau** : https://public.tableau.com/profile/tram.ngoc.le#!/

EDUCATION

Fordham University – Gabelli School of Business Aug.2019 – Expected Dec.2020
Master of Science in **Business Analytics, Data Science Focus**
GPA: 3.81/4.00; Activities: March Madness Data Crunch, Business Analytics Society
Relevant Coursework: Database Management, Web Analytics, Applied Regression Analysis, Data Mining, Business Decision for Data Manager, Big Data Analytics, Text Analytics, Statistical Methods and Computation I, Business Performance & Risk Management

Elizabethtown College Jan.2014 – Dec.2017
Bachelor of Science in **Business Administration, Accounting and Finance Focus**
GPA: 3.50/4.00; Honors & Activities: Dean's List, International Club's Representative, The National Society of Leadership and Success, Delta Mu Delta (Top 20% student International Business Honor Society)

SKILLS

Data Science: Python (BeautifulSoup, matplotlib, Tweepy, NumPy, Pandas, nltk, seaborn), R (ggplot2, R markdown), Tableau, SQL, Hive, Cloudera, TensorFlow, Spark, Batch Programming, SPSS, Alteryx, Qlikview, Spotfire
Business: Jira, Excel (Microsoft Office Specialist Certificate), Advanced Level IRS Certificated Tax Preparer, QuickBooks
Language: Vietnamese, English

PROFESSIONAL EXPERIENCE

FORDHAM UNIVERSITY, GABELLI SCHOOL OF BUSINESS New York, NY
Graduate Research Assistant Jan.2020 – Present
- Assist professors with multiple projects (Fashion AI & Public Health) under supervision of Yilu Zhou and G. O'Connor
- Develop a crawling algorithm to gather metadata of +200 influencer's accounts and +7 Million media from Instagram utilizing virtual machines of Google Cloud Platform (GCP)
- Preprocess, collate and clean Health survey and Instagram data with sample size of +500,000 high dimensional records with application of python pandas for both projects
- Visualize Instagram caption, hashtags, and comments using matplotlib and seaborn to examine feature importance on fashion trend for Fashion AI project

HAVER ANALYTICS INC. New York, NY
Economic Research Assistant Mar.2018 – Feb.2019
- Maintained and updated extensive time-series database utilizing Data Link Express (DLX) as well as other proprietary computer software (EViews)
- Developed automation program with DOS batch to collect data from government sources of European and Southeast Asian countries
- Performed database cleansing by fixing calculated fields with error to ensure validity and of updates processed
- Designed, revised and tested update procedures to enhance operation efficiency by reaching an approximately 5 or less seconds of process per update

PROJECT

Music Recommendation System, Big Data Analytics Project Jan.2020 – April.2020
- Built a personalized recommendation system with Spark and MLib using a 3GB song dataset published by Million Song Dataset
- Instantiated an Ubuntu Virtual Machine to set up the Jupyter and PySpark environment on Google Cloud Platform
- Implemented the alternating least squares recommender algorithm, tuned parameters with grid-search and cross-validated on the prediction results
- Performed features creation of TF-IDF, Word2vec and LDA on lyrics to build a content-based song recommender
- Utilized NLTK and Latent Dirichlet Allocation to visualize and model 20 topics from the lyrics.

Google Q&A Types & Topics Classification, Text Analytics Project Jan.2020 – April.2020
- Utilized Keras to train recurrent neural network, Long Short-Term Memory, to perform multi-class classification on Q&A types based on the Google QUEST Q&A Labeling dataset published by Google Research team, CrowdSource
- Applied scikit learn to train Support Vector Machine, Naïve Bayes, Decision Tree and Random Forest for Q&A topics classification then performed cross-validation to decide the model with the highest accuracy, reaching 95% of accuracy with SVM
- Implemented feature engineering with Natural Language Toolkit to preprocess and extract TFIDF, Word2Vec, and Bag-of-Word from the Google's Q&A text data.
- Used Synthetic Minority Oversampling Tecnique (SMOTE) to random oversample the training data to fix the unbalanced classes problem

Literate Rate Examination, Applied Regression Analysis Project Jan.2020 – April. 2020
- Built a multi-level regression model from scratch to predict which factors might affect literacy rate with the Word Bank data of +10 variables across 195 countries in R Markdown
- Conducted initial data exploration, filtered outliers, and visualized one-variable summary and linear relationships between variables using GGally

Twitter Malicious Bots Classification, Web Analytics Project Aug.2019 – Dec.2019
- Crawled information for +2,000 Twitter accounts and +200,000 of tweets using Tweeter API and python
- Implemented analytics exploratory through Tableau charts and text visualization of Tweets to display significant pattern of malicious bot-like behaviors
- Applied Random Forest technique and Tweet Semantics to predict and classify malicious bots based on behavior patterns, reaching 91.68% accuracy rate on testing set