WONJAE LEE

DATA SCIENTIST

- ✉ wonjae.lee09@gmail.com
- ☎ 201-981-6788
- 📍 New York, New York 10004
- in wonjae-lee-8a8b1b209/
- ⌂ wonjaelee09

Skills

LANGUAGES/TOOLS

SQL
Python
Git
NLP
NumPy
pandas
scikit-learn
Beautiful Soup

MACHINE LEARNING

classification
clustering
dimensionality reduction
natural language processing
regression
models
web scraping

DATA MANAGEMENT

AWS
PostgreSQL

VISUALIZATION

matplotlib
seaborn
Tableau

Education

SUNY Binghamton	2009 to 2013
BS Accounting	

Experience

Metis
Data Science

New York, NY
Jan. 2021 to Mar. 2021

- Metis is an intensive 12 week immersive data science bootcamp focused on project oriented learning
- The core curriculum is centered around Python, statistics, supervised and unsupervised machine learning, exploratory data analysis, databases, and visualization techniques
- Completed five self designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations

Globe Tax Services
Business Analyst

New York, NY
May 2014 to Current

- Reduce customer churn by identifying KPIs and acting proactively to build a solicitation system
- Design layouts, interactions, and transitions across modules of ESP for seamless claim submissions and tracking of applications
- Collaborate with stakeholders to source feedbacks and implement systemic enhancements addressing FAQs
- Create and maintain JIRA stories to implement market specific modules and documentation generation rules for the ESP
- Manage teams of five associates to retrieve unclaimed withholdings for Depositary Receipts (foreign equities) on a wide variety of securities by negotiating and advising Corporate and Investment banks on the reclamation process

Projects

Korean Movie Recommender

- Completed collaborative filtering utilizing both cosine distances between the doc topics as well as ratings for different age groups
- Scraped and parsed IMDBPro pages with Beautiful Soup to compile texts corresponding to specific movies
- Created a Flask to visualize insights and to recommend a Korean movie

Topic Modeling Liverpool FC's Press Conferences

- Performed topic modeling with natural language processing techniques, to yield the most frequently mentioned topics during Liverpool's match press conferences
- Scraped 4 seasons worth of press conferences from transcriptions on Youtube
- Devised an unsupervised topic model based on Latent Drichlet Allocation and CountVectorizer

NYC ABNB Classification

- Performed classification models on 48,000+ NYC listings to predict if an Airbnb listing will have less than 200 days of availabilities in a given year
- Carried out cross-validation and hyperparameter to tune and assess multiple classification models, including Naive Bayes, Logistic Regression, k-Nearest Neighbors, and Random Forest to find optimal model
- Compared Accuracy, precision, recall, F1, and ROC-AUC curve to select the best model

Predicting Goals Per Minute Based on FIFA 20 Stats

- Built linear regression models to predict goals per minute based on the game FIFA 20 data to identify players that are being undervalued
- Scraped soccer player's perceived stats data using Beautiful Soup from SoFifa and Infogol