

YOON TAE PARK

Jersey City, NJ | yp2201@nyu.edu | (551) 280-6244
<https://github.com/yoontaepark> | <https://www.linkedin.com/in/yoontaepark>

EDUCATION

NEW YORK UNIVERSITY

New York, NY

Master of Science in Data Science, GPA: 3.95

Expected Graduation: May 2023

- Relevant coursework: Machine Learning, Big Data, Natural Language Understanding, Time Series Forecasting

SEOUL NATIONAL UNIVERSITY

Seoul, Korea

Bachelor of Arts in Business Administration

Graduated Feb. 2014

TECHNICAL SKILLS

- Python
- SQL
- pyspark
- Microsoft Excel
- GCP
- Tableau
- Git, Github

WORK EXPERIENCE

Walmart Global Tech

Sunnyvale, CA

Data Science Intern

May.2022 – Aug.2022

- Worked in e-commerce team, leveraging rest of market data source to improve Walmart e-commerce platform
- Created a benchmark dataset using rest of market and Walmart item information that enriched the team's research
- Proposed a boosting model to train the benchmark dataset that improved current category mapping performance by 35%
- Built a machine learning pipeline that replaced the team's deliverable which is to predict category for new items

UNDP(United Nations Development Programme)

New York, NY

Research Assistant

Dec. 2021 – Mar.2022

- Developed an earthquake response model with Bilkent University to safely route people after an earthquake
- Evaluated a reference risk of streets, and suggested the safest path by calculating total risk score

SAMSUNG ELECTRONICS

Suwon, Korea

Finance Manager

Feb. 2014 – Jun.2021

- Elected as chair of automation projects for finance division that applies machine learning solutions
- Applied Decision Tree and Naive Bayesian method that enhanced the accuracy of the tasks from 60% to 95%
- Completed a prototype production of a chatbot that provides information on expense management and reimbursement

PROJECTS

A Toxicity Detection Dataset with r/WallStreetBet Comments

Jan.2022 - May.2022

- Created a new toxicity detection dataset using comments from r/WallStreetBets and toxicity labels from Perspective API
- Evaluated the dataset by comparing human baseline to current SOTA models: GPT-3, BERT, RoBERTa, and DeBERTa

Collaborative-Filter Based Modeling for Movie Recommendation

Jan.2022 - May.2022

- Created a recommender system using PySpark's ALS method that provides top 100 movies for each user
- Evaluated the model on Normalized Discounted Cumulative Gain (NDCG), and created a comparison to a single-machine implementation using LensKit

Citadel Fall 2021 Central Regional Datathon (1st place): Tobacco usage data analysis

Nov.2021

- Generated geographic data visualizations that shows the pattern of tobacco usage in relation to MPOWER (WHO Framework Convention on Tobacco Control)
- Built a regression model on tobacco usage using MPOWER and Geographic location

NYU Center for Data Science: Graduate Student Analysis and Visualization

Sep.2021 – Dec.2021

- Analyzed NYU data science past graduate students' academic data to uncover distinct correlations
- Visualized student academic performances and career outcomes that will assist prospective incoming students