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 **Expected Graduation: May 2023**
- Relevant coursework: Machine Learning, Big Data, Natural Language Understanding

SEOUL NATIONAL UNIVERSITY **Seoul, Korea**
Bachelor of Arts in Business Administration **Graduated Feb. 2014**

WORK EXPERIENCE

Walmart Global Tech **Sunnyvale, CA**
Incoming Data Science Intern **Starting May 2022**

UNDP(United Nations Development Programme) **New York, NY**
Research Assistant **Dec. 2021 – Mar.2022**
- Developed a model with Bilkent University to safely route a person through the streets after an earthquake
- Evaluated a reference risk of streets, and suggested the safetest 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
- Identified problems on data preprocessing and proposed categorization of input variables to prevent data overfitting
- Applied Decision Tree and Naive Bayesian method that enhanced the accuracy of the tasks from 60% to 95%
- Built an automation scenario from invoice to accounting system and proposed learning methodologies for each field with a developer
- Developed a chatbot that provides information on expense management and reimbursement in company intranet; used node.js and natural language processing and completed prototype production

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 attribute 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 **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

TECHNICAL SKILLS

- Python
- SQL
- Spark
- Microsoft Excel
- Tableau
- Git, Github