Seungju Han

seungjoohan1@gmail.com | 609-558-8851 | New York, NY | github.com/seungjoohan

PROFESSIONAL EXPERIENCE

ViFiVE, New York, NY — Aug 2022 – Current
Data Scientist

- Boosted **3D pose estimation accuracy by 17%** with a custom self-attention module, improving clinical assessment performance.
- Developed an comprehensive ML pipeline using MySQL, Firestore, and Python, from data parsing to model production, **boosting clinical assessment efficiency by 30%** across multiple tests.
- Revamped data augmentation pipeline for a human pose estimation model, **achieving a 10% increase in computational efficiency**, reducing model training time, and accelerating overall project timeline.
- **Enhanced a Random Forest pose classification model accuracy from 89% to 96%** by applying advanced feature engineering techniques, boosting reliability of clinical assessment predictions.
- Led efforts to validate the proficiency of AI-based models in clinical assessments with researchers at Stanford and Seoul National University Hospital, successfully applying vision AI models to high-stakes assessments like the Short Physical Performance Battery test, **impacting over 300 patients and clinicians** across different regions.

New York Mets, Remote — Oct 2021 – Jul 2022
Data Scientist

- Generated expected distribution model for infield and outfield batted balls with 89% accuracy by leveraging pitcher and batter data, including pitch types, pitcher/batter handedness, and batted ball statistics.
- Applied K-Means and KNN clustering methods to group similar pitchers, simulating more realistic pitcher-batter matchups. This approach led to a **25% improvement in prediction accuracy** in player performance forecasts.
- Leveraged data analysis predictive models to inform defensive alignment strategies, which contributed to the Mets **improving their Defensive Runs Saved (DRS) by 130%**, from +20 DRS in 2021 to +46 DRS in 2022, significantly enhancing overall team defense.

Fastcampus Language, Seoul, Korea — Jun 2020 – Aug 2020
Business Analyst

- Automated KPI dashboard for leadership team, **reducing operating time by 75%** through integrating Google Suite tools (BigQuery, Looker, and Google Sheets).
- Leveraged SQL to analyze data for multiple teams, uncovering key trends and providing actionable insights that improved decision-making and contributed to **13% decrease in operating cost on average**.

PROJECTS

Item Recommendation Engine

- Built and trained a matrix factorization model using a deep neural network in PyTorch to predict users' implicit feedback, achieving 3rd place in the competition.
- Applied 1:2 negative sampling to generate a balanced dataset of negative feedback samples and a dropout layer to improve the model's training process.

Scalable Movie Rating Prediction

- Developed a scalable end-to-end data pipeline to preprocess over 28 million movie and user records using Spark and stored engineered features in MongoDB for efficient retrieval.
- Trained an Alternative Least Square (ALS) collaborative filtering model, achieving r2 score of 0.83 within 2 hours using distributed machine learning, leveraging the scalability of Spark to process the large dataset efficiently.

EDUCATION

University of San Francisco, San Francisco, CA — Jul 2022
MS in Data Science
Relevant Coursework: Machine Learning, Deep Learning, Linear Regression, Time Series Analysis, A/B Testing, Relational Databases, NoSQL, Distributed Computing (Spark), Data Structures & Algorithms

University of California Berkeley, Berkeley, CA — Dec 2019
BA in Statistics

TECHNICAL SKILLS

Python (numpy, pandas), Machine Learning (pytorch, tensorflow, scikit-learn), Deep Learning, SQL (BigQuery), NoSQL (MongoDB, Firestore), Spark, Data Visualization (matplotlib, Seaborn), AWS (S3, EC2, EMR), Databricks, GCP