HSU-SHENG (JOHNSON) KO

550 Vanderbilt Ave #610, Brooklyn, NY 11238
808 999 9601 | hk3176@columbia.edu | linkedin.com/in/hsushengko |jko0401.github.io

PROFESSIONAL EXPERIENCE

ASML Wilton, CT
Data Scientist Jan 2023 - Present
First data scientist in the factory organization, supporting manufacturing, finance, and operations with advanced analytics and ML solutions.

- Championed ML/AI use cases for factory-wide optimization, implementing an XGBoost regression model and A* search algorithm to streamline the high-speed lens polishing process, accelerating convergence from raw blanks to qualified optical components.
- Led development of foundational data products in sensor integration and equipment effectiveness, enabling predictive maintenance, real-time alerting, and KPI tracking (e.g., MTTR, MTBF, utilization rate).
- Established process mining infrastructure from scratch, building SQL-based pipelines and Celonis dashboards to surface friction points and drive improvements across manufacturing, logistics, and finance.
- Reduced scrap cycle time by 2 weeks and cut holding costs by ~$500K/year by mapping ownership across the scrap process and automating milestone reminders using custom workflows.
- Segmented product families with K-means clustering and statistical profiling, enabling planners to incorporate historical variability into release strategies and buffer calculations.
- Served as cross-functional thought partner to senior leadership, defining technical requirements for analytics solutions, while actively participating in the DS/ML community of practice to elevate site-wide analytics standards and knowledge sharing.

Production Engineering Data Analyst Intern Jun 2022 - Sep 2022

- Implemented a new compilation process of 2148-image datasets using MatLab and reduced inspection time by 53%.

Terex Aerial Work Platforms (Genie) Redmond, WA
Data Analyst/Design Engineer May 2018 - Apr 2020

- Constructed a steel component pricing Random Forest model boosting price prediction accuracy from 70% to 94%.

RELEVANT PROJECTS

GenAI Script Writer & Video Editor
Personal Project Oct 2024

- Developed an AI-driven video content creation application leveraging Google Video Intelligence and GPT-4o for video annotation, and LLMs for script and shot list generation with RAG, reducing content production time for businesses.
- Designed and implemented scalable data pipelines integrating multiple tools (Google Cloud Storage, Firebase, and Qdrant) to preprocess, annotate, and store video data, ensuring high-quality, structured datasets for analysis and AI-driven insights.

NYPD Dispatch Simulation Model
Course Project May 2022

- Constructed a discrete event simulation model of NYPD dispatch using historical crime and response data, providing means to analyze efficiency of the current system.
- Proposed different working and back-up policies across precincts to decrease response time by 69% with the same number of vehicles.

EDUCATION

Columbia University New York, NY
Master of Science, Operations Research Dec 2022

- Relevant coursework: Probability & Statistics, Optimization, Stochastic Models, Simulation, Machine Learning, Deep Learning, Data Analytics, Transportation & Logistics Analytics, Supply Chain Analytics, Sports Analytics, Analytics on the Cloud.

University of Washington Seattle, WA
Bachelor of Science, Mechanical Engineering Mar 2018

- Extra-curricular: Formula SAE Drivetrain Team Lead

SKILLS & TOOLS

Data Analytics, Engineering & Science: Python (NumPy, Matplotlib, Pandas, Sklearn, Tensorflow, Langchain), SQL, MatLab.
Business Intelligence: Spotfire, Grafana, Celonis, Disco (Process Mining).
Cloud Services: AWS EC2, Google Cloud, Qdrant (VectorDB), Azure Databricks