JOONJE WOO

Los Angeles, CA • woojoonje@gmail.com • 213-676-6176 • joonjewoo • joonjewoo.github.io

EDUCATION

UNIVERSITY OF CALIFORNIA, LOS ANGELES — Los Angeles, CA

Bachelor of Science in Statistics & Data Science, Cumulative GPA: 3.618/4.00 — Sep. 2022 - Jun. 2026

- **Relevant Coursework**: Mathematical Statistics, Intro to data mining and statistical method, Monte Carlo Method, Intro to Python, Design and Experimental Statistics.

PASADENA CITY COLLEGE — Pasadena, CA

Associate of Science in Computational Engineering, Cumulative GPA: 3.86/4.00 — Sep. 2020 - May. 2022

- Dean's list with distinction
- **Relevant Coursework:** Intro to stats, Singular and Multivariable calculus, Linear Algebra, Differential equation

TECHNICAL CAPABILITIES

- **Programming**: R, Python
- **Database**: SQL
- **Visualization**: Power BI, Tableau
- **Version Control:** Git, GitHub

EXPERIENCES

Republic of Korea Army — Yangju, South Korea

Sergeant - Squad Leader — Oct. 2023 - Apr. 2025

Leading 81mm Mortar Squad and Managing Tactical Operations

- Led an 7-member mortar squad in the Republic of Korea Army's 81mm mortar Platoon, overseeing fire missions, ammunition logistics, and field coordination during intensive combat training and joint operations.
- Trained and mentored junior soldiers in weapon handling, firing procedures, and safety protocols, achieving zero safety incidents and maintaining top performance ranking within the platoon.
- Executed tactical decisions under high-pressure environments by synchronizing with infantry units and command posts, improving mission response efficiency by 20% and strengthening overall team discipline and cohesion.

PROJECT EXPERIENCES

Optimizing Supply Chain and Forecasting Delivery using R and ML on Amazon Dataset — Oct. 2025

- Performed comprehensive exploratory data analysis (EDA) and feature engineering on 12K Amazon supply chain records using R (tidyverse and data.table) to identify inefficiencies in delivery flow, uncovering three major delay patterns related to regional warehouse congestion.
- Built and validated predictive models using random forest and logistic regression, improving delivery time forecasting accuracy by 18% and revealing key predictors such as carrier type and fulfillment center distance.
- Implemented model evaluation and hyperparameter tuning through 5-fold cross-validation to ensure generalizability, reducing overfitting by 12% and enhancing model robustness for unseen data.
- Created automated scripts for data ingestion and preprocessing in R, reducing manual data preparation time by 30% and standardizing analysis workflows across multiple datasets.
- Developed interactive visualizations with ggplot2 to monitor fulfillment rates, shipping delays, and demand variability, enabling data-driven logistics optimization and actionable insights for process streamlining.

Predicting Skin Cancer Risk using Supervised Classfication Models on Kaggle Dataset — Sep. 2025

- Processed and analyzed 70,000+ samples (50,000 training / 20,000 testing) using R (Tidyverse, Caret, XGBoost) to identify key demographic, behavioral and environmental predictors of skin cancer risk.
- Engineered and tuned 50 predictors through data cleaning, scaling, and cross-validation, then trained and compared classification models (Logistic Regression, Random Forest, XGBoost), achieving 92% accuracy and high ROC-AUC performance.
- Implemented model evaluation metrics including precision, recall, and F1-score to assess classification balance, identifying false negatives as a key risk area for early detection improvement.
- Optimized XGBoost hyperparameters through grid search and early stopping, enhancing predictive stability and improving training efficiency by 25%.
- Visualized and interpreted model results with vip and SHAPforxgboost, revealing UV exposure, age, and skin type as top contributors to malignancy prediction, supporting early detection and data-driven prevention strategies.

Optimizing 100m Sprint Performance through Adrenaline and Age Interaction Analysis with R — Jul. 2023

- Designed and executed a Two-Way Randomized Block Design experiment with 240 participants, using R to automate randomization, simulate treatment assignments, and preprocess timing data for analysis.
- Applied Two-Way ANOVA to quantify the main and interaction effects of adrenaline injections and age brackets on sprint times, followed by Tukey's HSD to identify statistically significant group differences.
- Quantified the performance impact of adrenaline across age groups, revealing up to an 8.4% improvement in the 20-29 age cohort, while older groups exhibited non-significant gains, suggesting age-dependent physiological response patterns.
- Discovered that adrenaline improved sprint performance by up to 8.4% in the 20-29 age group while older brackets showed diminishing returns, generating insights on targeted performance enhancement strategies.