# Woo-hyun Shin

**Address** 214, Nowon-ro, Nowon-gu, Seoul, Korea
**Phone** +82 10-9534-8684
**Email** asw9234hi@gmail.com

## Internships

### Process-Aware Data Analytics Lab
Jan 2024 to Present

Undergraduate researcher, in Kwangwoon University

- Conducting research in Data Science based on Process mining and Text mining

## Project Experience

### Embedding Process Models Using Language Models
Feb 2024 to Present

Collaborative Research, in PADA Lab

- Presented a methodology for fine-tuning **Sbert models** to efficiently embed process models and convert tree models to text
- Analyzed similarities between process models based on their semantic and structural differences
- Visualized the results of the embeddings through dimensionality reduction
- Currently writing a paper targeting the EAAI Journal

### Fashion Trend Analysis & Development of Image Search Service
Feb 2024 to Oct 2024

Team Member, for *Capstone design* project for graduation

- Crawled data from three fashion websites and automated the process using **Apache Airflow**
- Developed a website for comparing and analyzing collected data
- Developed image search API using **Grounding Dino** & **SAM** models

### Development of a Language Model for Risk Analysis in Cybercrime Prevention
Mar 2024 to Jun 2024

Team Leader, for Major Courses *Text Mining* project

- Utilized crime glossary data for terms such as drugs and sexual offenses
- Fine-tuned **BERT classification model** for risk category and severity analysis of internet posts
- Provided crime prevention solutions including relevant legal provisions and penalties based on analytical results to authors
- Created a chatbot for counseling and advice using **GPT API** and **RAG** technology

### Prediction of Thyroid Disorders via Patient Biometric Signals with Machine Learning
Sep 2023 to Dec 2023

Team Member, for Major Courses *Data Mining* project

- Collected data including patient test information and thyroid symptoms
- Handled variables through Exploratory Data Analysis (EDA)
- Compared model performance based on various **feature selection** methods and Machine Learning techniques(e.g., **Naive Bayes**, **XGBoost**)

### Development of an Item-Based Collaborative Filtering Recommendation System

Sep 2023 to Dec 2023

Team Member, for Major Courses *Big Data Processing and Applications* project

- Constructed an item matrix using tags and genre information from MovieLens data
- Built a system based on various similarity measures such as **TF-IDF**, **Pearson Similarity**, and **Jaccard Similarity**
- Implemented big data parallel processing using **PySpark**

### Prediction Used Car Prices via Vehicle Information with Machine Learning

Mar 2023 to June 2023

Team Member, for Major Courses *Data Mining* project

- Utilized various **Machine Learning** techniques for prediction, including **Linear Regression**, **Decision Trees**, and **Random Forests**
- Enhanced model performance through **Ensemble Methods**
- Interpreted results using **SHAP** and **LIME** for **Explainable AI (XAI)**

## EDUCATION

### Kwangwoon University

Mar 2019 to Feb 2025

B.S. in Information Convergence, Major in Data Science
- Total GPA : 4.0 / 4.5
- Major GPA : 4.14 / 4.5

## Awards & Honors

### Awards

2023.11 Dean's List, for Academic Excellence *Kwangwoon Univ.*

### Honors

2024 Semester High Honors, Awarded to students with high achievements throughout the semester
2023 Semester High Honors, Awarded to students with high achievements throughout the semester
2023 Semester High Honors, Awarded to students with high achievements throughout the semester

## SKILLS

### Programming Language

**Python** 

**SQL** 

### Machine Learning / Deep Learning

**PyTorch** 
**Pandas** 
**Scikit-learn** 
**Keras** 
**PySpark** 

**TensorFlow** 
**Numpy** 
**Matplotilib** 
**Seaborn** 

### Tools

- **Visual Studio Code**
- **Google Colab**
- **Anaconda**
- **Jupyter**
- **Apache-Airflow**
- **Wandb**

### Cloud

- **GCP**
- **AWS**

### Used at least once

- **HTML/CSS**
- **Javascript**