

Seohee Yoon

akheyn029@gmail.com | Atlanta, GA | <https://syoon029.github.io/>

Education

Georgia Institute of Technology

Master of Science in Computer Science

Bachelor of Science in Computer Science

Atlanta, Georgia

Aug 2024 – Dec 2025

Jan 2022 – May 2024

Experience

Big Data Analysis and Operation Assistant Intern

Seoul, South Korea

PTKOREA

Jun 2024 – Aug 2024

- Streamlined data tracking for 1,000+ customer accounts by leveraging Jira and Tableau, reducing account-related issues by 20%
- Led periodic database modifications by collaborating with cross-functional teams, decreasing data-entry errors by 30% and maintaining system consistency
- Facilitated international client meetings by clarifying technical requirements and resolving complex issues

Research Assistant - Data Scientist

Atlanta, Georgia

Data Driven Education, Georgia Institute of Technology

Aug 2022 – Dec 2023

- Enhanced a regressor model's accuracy by preprocessing raw data and designing a custom training dataset, resulting in a 25% improvement in prediction precision
- Identified key terms critical for determining problem difficulty by leveraging TF-IDF to evaluate term frequency, leading to more targeted feature engineering
- Improved assessment difficulty predictions by applying Random Forest, Decision Trees, and SVM, increasing classification accuracy by 40%
- Elevated the Random Forest model's performance by 50% over baseline through data augmentation, resulting in more reliable difficulty-level predictions

Projects

Sentiment Analysis of University Student Discussion on Reddit

Aug 2024 – Dec 2024

- Cleaned and prepared 3.8 million Reddit posts by implementing advanced preprocessing (tokenization, stop word removal, language filtering), resulting in higher-quality input data
- Boosted sentiment classification performance by fine-tuning a BERT model to categorize emotions into seven classes, increasing accuracy by 35%
- Developed interactive visualizations (choropleth maps, word clouds) to explore sentiment trends and keyword relations, providing real-time insights for stakeholders

Respiratory Diagnosis Assistant

Feb 2024 – May 2024

- Created a machine learning-powered web application for lung sound classification to assist in medical diagnostics
- Improved diagnostic accuracy from 70% to 83% by implementing a GRU model and applying data augmentation techniques
- Managed data using MongoDB and Amazon S3, integrating the database with Django to improve data accessibility

Stock Market Prediction Project

Feb 2024 – May 2024

- Increased forecast accuracy by applying SVM, Linear Regression, and LSTM models, achieving a 20% improvement over baseline predictions
- Decomposed price trend data into low and high frequency components using Discrete Wavelet Transform (DWT) to enhance model generalization
- Visualize prediction results from various models and compared each accuracy using Tableau

Scene Recognition with Deep Learning Project

Nov 2023

- Implemented a CNN model (SimpleNet) to classify natural images with 2 convolutional layers, improving accuracy by 30% through data augmentation and regularization techniques
- Achieved 85% testing accuracy by enhancing a pretrained ResNet in PyTorch and selectively modifying its layers, leading to higher scene recognition precision

Skills

Python, SQL, Tableau, Pyspark, Scala, Git, Scikit-learn, Java, Excel, JavaScript, C, Assembly, React Native, C#, Docker