Seohee Sunny Yoon

akheyn029@gmail.com | 470-838-2401 | Atlanta, GA,30308 | https://syoon029.github.io/

Education

Georgia Institute of Technology **Atlanta, Georgia**

Master of Science in Computer Science *Aug 2024 – Dec 2025*

Bachelor of Science in Computer Science *Jan 2022 – May 2024*

Experience

Data Analysis Intern **Seoul, South Korea**

PTKOREA *Jun 2024 – Aug 2024*

- Resolved approximately 20 customer account issues for data dashboards using JIRA and Tableau daily, minimizing the company's profit losses
- Monitored anomalous data, regularly updating the database to meet customers' needs and ensuring data quality and system consistency using Excel and MySQL
- 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*

- Suggested a new project direction and applied the Depth of Knowledge (DOK) framework to evaluate assessment quality efficiently, translating problem-solving approaches into data-driven solutions
- Identified key terms critical for determining assessment difficulty by leveraging TF-IDF to evaluate term frequency, leading to more targeted feature engineering
- Enhanced a regressor model's accuracy by preprocessing raw data and designing a custom training dataset, resulting in a 25% improvement in prediction precision

Skills

- **Programming Languages:** Python (Pandas, NumPy), MySQL, R, PySpark, Scala, Java, JavaScript, C#, C
- **Data Analysis and Visualization:** Tableau, Jupyter Notebook, Scikit-learn, Excel, Seaborn, Matplotlib, D3, Adobe Analytics
- **Relevant Coursework:** Database Technologies | Machine Learning | Data Visualization | Natural Language

Projects

Sentiment Analysis of University Student Discussion on Reddit *Aug 2024 – Dec 2024*

- Prepared 3.8 million Reddit posts by implementing a data preprocessing pipeline (tokenization, stop word removal, language filtering)
- Developed interactive visualizations (choropleth maps, word clouds) using D3 and Tableau, providing overall sentiment trends on specific topics and keyword relations

Implicit Emotion Classification *Aug 2024 – Dec 2024*

- Processed dataset using regular expression and Python libraries, reducing training time by 20%
- Evaluated various combinations of language models and knowledge graphs to identify the optimal solution for implicit emotion classification
- Led the development of the Know-BERT language model by integrating the BERT model with the domain-specific knowledge graph, resulting in the improvement of the model's accuracy from 64% to 88%

Respiratory Diagnosis Assistant *Feb 2024 – May 2024*

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

Stock Market Prediction Project *Feb 2024 – May 2024*

- Decomposed price trend data into low and high frequency components using Discrete Wavelet Transform (DWT), improving model accuracy by 5%
- Increased forecast accuracy by analyzing SVM, Linear Regression, and LSTM models, achieving a 20% improvement over baseline predictions