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 – Expected Dec 2025*
Bachelor of Science in Computer Science   *Jan 2022 – May 2024*

Experience

Graduate Research Assistant, SocWeb Lab - Atlanta, Georgia   *Feb 2025 – Present*

Supervisors: Dr. Munmun De Choudhury

- Contributing interdisciplinary research leveraging computational techniques, natural language processing, and social computing to analyze digital trace data and improve mental health outcome
- Engaging in the development of a methodology to identify how bias against mental health entities is generated by LLMs using the toxicity rabbit hole framework

Data Analysis Intern, PTKOREA - Seoul, South Korea   *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 ensure data quality and system consistency using Excel and SQL
- Collaborated with cross-functional teams to test market dashboard functionality, identifying and resolving operational issues to enhance performance and user experience

Undergraduate Research Assistant, C21U Lab - Atlanta, Georgia   *Aug 2022 – Dec 2023*

Supervisors: Dr. Jeonghyun Lee, Meryem Yilmaz Soylu

- 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 25% in prediction precision through preprocessing raw data and designing a custom training dataset

Skills

- **Programming Languages:** Python (Pandas, NumPy), SQL, R, PySpark, Scala, Java, JavaScript
- **Data Analysis and Visualization:** Tableau, Scikit-learn, Excel, Matplotlib, Seaborn, 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 utilizing a data preprocessing techniques (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 emotion 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
- Accomplished a model accuracy increase from 64% to 88%, measured by evaluation metrics, by developing Know-BERT and integrating BERT with a domain-specific knowledge graph

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%, evaluated through model performance, by deploying a GRU model and enhancing audio data with augmentation techniques
- Managed dataset using MongoDB and Amazon S3, integrating the database with Django to improve data accessibility efficiently