**Seohee Yoon**

|  |  |
| --- | --- |
| 470-838-2401 | [akheyn029@gmail.com](mailto:akheyn029@gmail.com) | Atlanta, GA | <https://syoon029.github.io/> |  |
|  |  |

**Education**

|  |  |
| --- | --- |
| **Georgia Institute of Technology** | **Atlanta, Georgia** |
| Master of Science in Computer Science  Bachelor of Science in Computer Science | *Aug 2024 – Dec 2025*  *Jan 2022 – May 2024* |

**Experience**

|  |  |
| --- | --- |
| **PTKOREA** | **Seoul, South Korea** |
| *Big Data Analysis and Operation Assistant Intern* | *Jun 2024 – Aug 2024* |

* Organized and maintained data for over 1,000 customer accounts, using Jira and Tableau to track and resolve account-related issues efficiently
* Managed the periodic modification and mapping of databases to cater to the needs of professionals, ensuring system consistency and reducing errors in Excel
* Participated in team meetings with international clients, enhancing communication and resolving complex customer issues

|  |  |
| --- | --- |
| **Data Driven Education, Georgia Institute of Technology** | **Atlanta, Georgia** |
| *Research Assistant - Data Scientist* | *Aug 2022 – Dec 2023* |

* Preprocessed raw data and designed a training dataset for a regressor model using a customized data pipeline
* Leveraged the TF-IDF technique to evaluate term importance based on frequency and identify key terms critical for determining problem difficulty
* Applied machine learning techniques (Random Forest, Decision Trees, Support Vector Machines) to predict difficulty levels of assessments based on the Depth of Knowledge (DOK) framework
* Augmented the dataset to improve the random forest model’s accuracy by 50% compared to the baseline

**Projects**

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

* Designed and executed a data preprocessing pipeline, incorporating tokenization, stop word removal, and language filtering to clean and prepare 3.8 million Reddit posts
* Conducted sentiment analysis on Reddit posts using a fine-tuned BERT model to classify emotions into seven distinct categories
* Developed interactive visualizations including choropleth maps and word clouds to present sentiment trends and keyword relations dynamically

|  |  |  |
| --- | --- | --- |
| **Respiratory Diagnosis Assistant** | *Feb 2024 – May 2024* | *September 2023* |

* 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* |

* Predict stock price changes using several machine learning algorithms such as Support Vector Regression, Linear Regression, and LSTM
* 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
* Enhanced pretrained Resnet using Pytorch by modifying layer of model to get a testing accuracy of 85%

**Skills**

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