Seohee Yoon

Objective

I am pursuing a Master’s in Computer Science at Georgia Tech, integrating industry experience and research as a Data Scientist. Passionate about solving challenges posed by complex real-world datasets using machine learning techniques and data analytics skills.

|  |  |  |  |
| --- | --- | --- | --- |
| 470-838-2401 | [akheyn029@gmail.com](mailto:akheyn029@gmail.com) | Atlanta, 30308 | <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
* Managing the periodic modification and mapping of databases, ensuring consistency with data dashboard using Excel and Adobe Analytics
* 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* |

* Processed a raw data and designed a dataset for training Regressor mode by using customize data pipeline
* Leveraged the TF-IDF technique to evaluates 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
* Enhanced the dataset to improve the random forest model’s accuracy by 50% compared to the baseline

Projects

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

* Developed a machine learning-powered web application for lung sound classification, enhancing diagnostic accuracy from 70% to 83% by implementing a GRU model and 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
* Evaluate linear regression models using K-fold cross-validation, resulting in 95% accuracy in prediction
* Visualize prediction results from various models and compared each accuracy using Tableau

|  |  |
| --- | --- |
| **Scene Recognition with Deep Learning Project** | *Nov 2023* |

* Developed a convolutional neural network (SimpleNet) with 2 convolution layers which aligns with given training
* Implemented data augmentation techniques, normalization, and regularization to improve model accuracy by 30%
* Enhanced a pretrained ResNet model using PyTorch by modifying specific layers to optimize its architecture, resulting in an improved testing accuracy of 85%

Skills

Python, SQL, Tableau, Pyspark, Git, Scikit-learn, JavaScript, D3, Excel, Java, C, Assembly, React Native, C#, Docker