

Yejin Hwang

Greater Austin Area | yejincc99@gmail.com | [LinkedIn.com/in/yejin-data/](https://www.linkedin.com/in/yejin-data/) | [GitHub.com/Yejin-Hwang](https://github.com/Yejin-Hwang) | medium.com/@yejincc99

EDUCATION

Texas A&M University, Corpus Christi

Expected Graduation: Dec 2025

Master of Data Science | 4.0 GPA

SKILLS

Languages Proficient Python(3yrs) Intermediate R(2yrs) Beginner SQL(1yrs) ·

Technology Scikit-learn · TensorFlow · PyTorch · Streamlit · Pandas · Numpy · Matplotlib · Yfinance · Statsmodels

EXPERIENCE

ML Engineering Research Fellow | [LAB](#)

Sep 2024 - Present

Texas A&M University-Corpus Christi

Corpus Christi, TX

- Engineered time series datasets with 10+ features and 50K+ records for model development
- Deployed multi-source ingestion pipelines processing 10K+ records/day via dynamic API automation
- Architected scalable forecasting pipelines, reducing RMSE by 18% in validation tests
- Built NLP labeling and spike detection workflows, boosting classification accuracy by 25%

Clinical Data Specialist | [CCE](#)

Aug 2022 - Mar 2024

Samsung Medical Center-CCE(Clinical Center for Epidemiology)

Seoul, South Korea

- Built 100+ patient data pipelines integrating CRF and EMS systems, reducing manual data entry time by 40%
- Managed and cleaned 1,000+ clinical datasets and coordinated data workflows and protocols
- Supported data collection and statistical analysis for studies on rare and oncological diseases
- Collaborated with leading epidemiology researchers, including Prof. Juhee Cho (SAHIST/Johns Hopkins)

Undergraduate Research Assistant | [LAB](#)

Mar 2022 - Jul 2024

SKKU(Sungkyunkwan University)

Seoul, South Korea

- Conducted literature reviews and basic data handling for lab-scale experiments under Prof. Jinhee Hur
- Designed and prepared experimental setups following academic research protocols
- Analyzed datasets using Python, Excel, and statistical methods (e.g., regression analysis, hypothesis testing)
- Interpreted results to support conclusions and inform further experimental design

PROJECTS

Tesla Sentiment Analytics & ETL Pipeline | Individual Project (~50 hours) - [Github](#)

Jul 2025 - Sep 2025

- Built 5+ ML forecasting pipelines using PyTorch integrating Reddit sentiment with TSLA stock prices
- Enhanced TFT model with sentiment features, reducing RMSE by 50.8%(8.00→6.47) over no-sentiment baseline
- Benchmarked FinBERT, VADER, and DistilBERT models, improving classification accuracy by 12%
- Automated SQL-based ETL from finance APIs and Reddit to support daily model retraining and evaluation

Bayesian Healthcare Cost Prediction | Individual Project (~25 hours) - [Github](#)

May 2025

- Built Bayesian linear regression in R and compared with OLS and Ridge baselines
- Achieved RMSE ↓2.7%(4930→4796), MAE ↓2.2%(2952→2888), and R^2 ↑0.002 over frequentist model
- Visualized posterior distribution and credible intervals to explain uncertainty and cost drivers

Stock Forecasting with Transformer Models | Individual Project (~50 hours) - [Github](#)

Feb 2025 - Apr 2025

- For 90-day prediction, reduced RMSE by 49% (ARIMA 83.22 → Chronos 42.24) and MAE by 46%(63.93→34.45)
- For 5-day prediction, reduced RMSE by 20%(ARIMA 11.27→TimesFM 8.99) and MAE by 24%(9.64→7.37)
- Presented at Joint NMSU/UTEP Conference on Math, Computer Science & Computational Sciences
- Presented at Coastal Bend Mathematics and Statistics Conference ([LinkedIn post](#))

Activities

AI Blogs - Authored 5 articles on AI experiments, focusing on performance trade-off in model inference

Aug 2025 - Present

DSA Blogs - Launched blog series on data structures, algorithms, and coding interview prep

Aug 2025 - Present

SKKU President's Award(Seoul, South Korea) - Awarded ₩2M for biotechnology R&D project([LinkedIn post](#))

Nov 2024