Justin Kim

Seattle, WA | kim.j14@northeastern.edu | 425-495-6599
jkim-portfolio.dev | linkedin.com/in/jkim555/ | github.com/iluvcoding123

SUMMARY

Graduate student in Data Analytics Engineering with hands-on experience in Python, SQL, and machine learning. Experienced in designing end-to-end data workflows, from schema design and data ingestion to analytics and model deployment. Strong interest in finance-driven analytics and applied machine learning.

EDUCATION

Northeastern University, Seattle, WA *2025-2027 Expected Graduation*
Master of Science in Data Analytics Engineering

University of Washington, Seattle, WA *2018-2022*
Bachelor of Science in Electrical Engineering

PROJECTS

End-to-End ML Pipeline: Housing Price Prediction

- Developed a full machine learning pipeline in Python (Pandas, scikit-learn, XGBoost) to predict home sale prices, performing EDA, feature engineering, and preprocessing to optimize model inputs
- Trained, tuned, and evaluated multiple regression models, achieving $R^2 = 0.85$ and persisting the best model with joblib for deployment
- Built a FastAPI REST API for real-time predictions, containerized with Docker, and documented end-to-end workflow for reproducibility and cloud deployment

Real-Time ML System: Stock Sentiment Analyzer

- Built a real-time stock sentiment analysis pipeline using NewsAPI and FinBERT to classify financial headlines by ticker, outputting structured sentiment data in partitioned Parquet format.
- Implemented a Kafka-compatible streaming loop with Redpanda to automate ingestion, sentiment scoring, and live dataset updates.
- Developed a Streamlit dashboard to visualize sentiment trends over time, display label distributions, and surface the latest headlines with filtering support.

SmartRefi – GPA-Based Student Loan Refinance Data Platform

- Designed and implemented a fully normalized relational database in MySQL to model a GPA-based student loan refinancing system, enforcing referential integrity across applications, offers, loans, and refinance contracts.
- Generated and populated realistic, scalable sample data, then wrote analytical SQL queries to evaluate lender pricing, refinance volume, GPA-driven rate adjustments, and multi-loan consolidation behavior.
- Explored a denormalized MongoDB representation for exploratory analytics and built a Python data access layer to query the relational database and support downstream analysis.

PROFESSIONAL EXPERIENCE

Seattle Husky Robotics - Founding President
Northeastern University, Seattle WA *Sept 2025 – Present*

- Founded interdisciplinary robotics student organization in collaboration with faculty advisors and Student Affairs, recruiting graduate students across engineering, computer science, and analytics disciplines
- Designed and delivered technical workshops on CAD design, 3D printing, and microcontroller programming to develop member skills and launch hands-on robotics projects

SKILLS

Programming: Python, C, C++, Java, JavaScript, HTML, CSS, SystemVerilog

Data Analytics: SQL, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tableau, PyArrow/Parquet, Kafka (Redpanda)

Cloud: FastAPI, Docker, Streamlit

Machine Learning / NLP: Regression models, Feature Engineering, Transformers, Sentiment Analysis

CERTIFICATES & CERTIFICATIONS

AWS
- Certified AI Practitioner

DataCamp
- Associate Data Analyst (SQL)
- Data Analyst (Python)

Northeastern University
- Foundations of Responsible AI Learning
- Foundations of Business Learning