WONHA SHIN

Phone: +1) 201-566-0147 | Email: wshin7@ur.rochester.edu
Linkedin: https://www.linkedin.com/in/wshin7 | Project Portfolio: https://leahnote01.github.io/

EDUCATION

University of Rochester   **Rochester, New York**
Master of Data Science | GPA 3.43 / 4.0   *Aug. 2023 – Dec. 2024(EGD)*

Pusan National University   **Busan, South Korea**
Bachelor of Arts in International Business and Economics | GPA 3.73 / 4. 0   *Mar. 2011 – Aug. 2018*

PROFESSIONAL EXPERIENCE

VOESH New York   **South Plainfield, New Jersey**
Administrative Accounting Specialist, Admin/Accounting   *Oct. 2018 – Mar. 2023*

● Spearheaded the development of the company's sales analytics framework, conducting advanced statistical analysis on over 10,000+ clients to forecast trends and predict customer behavior. Utilized Oracle DB for robust data management, generating insightful weekly reports that guided strategic decision-making and drove sales strategies across 60+ in-house and outsourced sales representatives.

● Led the design and implementation of new CRM and ERP systems, focusing on scalability & data integration actively collaborating with cross-functional teams to engineer databases that improved operational efficiency and data quality. Successfully managed the transition, refining the company's dataset by 70%, which enhanced decision-making processes and prepared the organization for future adoption of ML technologies.

EASTERN AMERICAN CDC (Non-Profit Federal Community Bank)   **Englewood, New Jersey**
Credit Analyst Intern, Credit Assessment   *May. 2018 – Oct. 2018*

● Conducted comprehensive analysis and processing of SBA 504 loan and credit applications for over 100 community business owners, meticulously assessing financial information and performing in-depth risk analysis to determine credit risk levels and support informed decision-making and enhancing the bank's financial stability.

● Generated detailed credit reports and assessments proactively contributed to business development efforts, securing three new clients within the first four months of the internship through strategic marketing initiatives.

PROJECTS & INTERESTS

● **Real-Time Tweet Sentiment Analysis Pipeline :** Designed and implemented using PySpark, Databricks, and AWS, enhancing data flow efficiency from raw to refined stages, leveraging cloud services for robust data hosting and management, which supported dynamic data analysis and decision-making through applied MLops techniques.

● **Data Mining: Cross-Cultural NLP Analysis of Luxury Hotel Reviews in Europe- LDA Topic Modeling :** Led a pioneering NLP project analyzing over 515,000 luxury hotel reviews, leveraging Python and advanced Scikit-learn to uncover intricate customer behavior patterns across diverse global regions. Utilized Latent Dirichlet Allocation (LDA) for topic modeling, integrating deep learning techniques to enhance model accuracy and predictive power.

● **Rstudio Statistical Analysis: Multidimensional Analysis of Video Game Sales and Global Market Trends:** Directed a comprehensive data analysis project using RStudio, scrutinizing over 16,000 video game titles to extract significant market trends and sales patterns. Implemented advanced statistical methods guiding strategic market analysis

● **ClinicOps: Integrated DBMS Solution**: Engineered a sophisticated DBMS for a clinical setting, developing a high-performance platform with PHP & SQL. Implemented role-specific access and utilized data normalization, ensuring data integrity and efficiency. The project was acclaimed a top project for superior technical prowess of the semester.

● **Tri-States Area Quality of Life Analysis:** Conducted a detailed statistical analysis of quality of life indicators across NY-NJ-CT, employing Tableau for advanced data visualization. Uncovered socio-economic factors affecting crime rates, providing valuable insights for public policy and regional planning.

SKIIS

● **Big Data Technologies & Database:** Apache Spark, AWS, DataBricks, SQL
● **Machine Learning & Deep Learning**: NLTK, NLP, Tensorflow, Keras, PyTorch, Scikit-Learn, RandomForest, GVM