Heesuk Jang

(720) 201-7402 | Irvine, CA
heesuk.jang@gmail.com
https://www.linkedin.com/in/heesukjang/
https://www.heesukj.com/
https://github.com/heesukjang

PROFESSIONAL SUMMARY

Experienced and highly motivated Data Scientist and Analyst with a passion for Natural Language Understanding (NLU), Deep Learning, and researching new technologies with innate curiosity and a love of learning. 5+ years' experience in data analytics, data science, and machine learning. Proven track record of writing quality code and delivering projects on schedule as well as demonstrated ability to learn new tools quickly and develop innovative solutions to energy market analytics. Leadership experience includes providing training and assisting colleagues on new technologies as well as being active in S&P Global's Women In Technology.

SKILLS AND TOOLS

- Natural Language Processing (NLP) – BERT and BERTweet
- Convolutional Neural Networks (CNNs), Transfer Learning with CNNs, Long Short Term Memory Networks (LSTMs), Vision Transformer (ViT)
- Regression, Classification, Gradient Boosting, Tree-based Algorithms, Clustering, Time Series
- Python with TensorFlow and Keras, SQL, R
- AWS, GCP
- Azure Databricks, PySpark, Kafka, Hadoop, Presto, Git
- Tableau, Plotly Dash
- Docker, Airflow

EXPERIENCE

S&P Global Commodity Insights — **Irvine, CA (Remote)**
Data Analyst — Sept. 2021 – Present

- Develop five key accuracy metrics to evaluate oil and gas price forecasts for 23 hedge funds.
- Build Tableau dashboards to visualize price forecasts, saving 5 hours per week in manual reporting.
- Engineer ETL solutions using Databricks, PySpark, Python, SQL, AWS, and Airflow, reducing processing time by at least 70%, from days or weeks to minutes.
- Build a machine learning pipeline to detect anomalies on time-series data across different environments (production vs. development) and database systems (PostgreSQL and SQL Server) using various statistical methods as well as advanced machine learning techniques, improving detection accuracy by 95%
- Train and mentor team members on new technologies such as Databricks, Python, and PySpark, improving data pipeline efficiency from weeks to minutes in a day or few days.
- Active board member and speaker for S&P Global's Women in Technology, sharing expertise in data science and machine learning.

Austin Capital Data — **Austin, TX (Remote)**
Teaching Assistant — Jan. 23, 2023 – Jan. 27, 2023

- Answered technical questions in Python training for Machine Learning (ML) and NLP for 75+ data practitioners at Center For Disease Control and Prevention (CDC) live stream training.
- Guided trainees in breakouts with code exercises in real time and to proceed ML- and NLP-driven work-related projects.

S&P Global Market Intelligence — **Boulder, CO**
Senior Data Researcher — May 2018 - Sept. 2021

- Applied machine learning and data science techniques to generate insights on global power generation across 10 regions, covering 190+ countries.
- Developed interactive Dash Python dashboards to analyze energy market impacts on fossil fuel outlook.
- Engineered high-performance ETL pipelines for power and utility data, reducing processing time from 10 days to under 10 minutes.

PROJECT

Please visit www.heesukj.com for more projects

WasteWizard with Computer Vision, UC Berkeley (Capstone Project) — April, 2024
https://www.ischool.berkeley.edu/projects/2024/wastewizard
Utilized the Vision Transformer (ViT) for waste item image classification with a 90% macro F1 score, and 91% precision, and developed a website using AWS Amplify, AppSync, S3, DynamoDB, and EC2 for users to upload images and review results, while focusing on data wrangling, cleansing, EDA, and building advanced models such as CNNs and leveraging Transfer Learning techniques.

Flight Departure Delay Prediction with MLflow, Databricks, ML/DL Models, and GCP, UC Berkeley — December, 2023
Predicted flight departure delays (15+ minutes) 2 hours in advance using a high-performing XGBoost model (69% F2 score) on 60+ million records of flight and weather data, involving EDA, statistical analysis, data pre-processing, feature engineering, and Bayesian Optimization for hyperparameter tuning.

Climate Change Dashboard using Tableau, D3.JS, & Vega-Altair, UC Berkeley August, 2023
https://drive.google.com/file/d/1b4dWghYWxcTrHJvncjXVlBWemisMCdLA/view
Led the development of key webpages including Home, Overview, and Financial and Risk Indicators using HTML, CSS, Bootstrap, and JavaScript, integrating Tableau charts and graphs. Designed the overall layout and static visual components to unify team's work.

Improve English Essay with Artificial Intelligence, UC Berkeley May, 2023
Automatically evaluated six analytic measures to argumentative essays from grade 8-12 using BERT and BERTweet.

Invasive Ductal Carcinoma (IDC) Breast Cancer Classification, UC Berkeley December, 2022
Classified IDC Breast Cancer histopathology images into cancerous vs. non-cancerous, leveraging XGBoost, CNNs, and CNNs with Transfer Learning including VGG16, VGG19, ResNet50, ResNet152V2, and DenseNet201.

Global Energy Trend with Fossil Fuel Price Prediction, S&P Global Market Intelligence 2020
Built an interactive web interface using Python Dash and Plotly to provide insights into global energy trends and fossil fuel price predictions applying Seasonal ARIMA, STL Decomposition, Exponential Smoothing, and Prophet time series models.

EDUCATION

University of California, Berkeley – Master of Information and Data Science GPA 4.0/4.0 May 2024
Relevant Courses: Research Design and Applications for Data and Analysis, Data Science Programming, Statistics for Data Science, Applied Machine Learning, Data Engineering, NLP with Deep Learning, Data Visualization, ML at Scale, Capstone (WasteWizard with Computer Vision)