Nikhil Gajghate

nikhilgajghate4@gmail.com | Milwaukee, WI

Experience

Data Scientist at Direct Supply, Milwaukee, WI **July 2022 – Present**

Project: Built a price evaluation pipeline for Service Provider bidded work in senior-care facilities.

- Partnered with Product Managers, Business Partners and Engineers to frame problems.
- Applied Python graphing libraries to perform EDA to gain a deeper understanding of the domain.
- Used Regex, standard Python libraries, to preprocess in-house and third-party data for model consumption.
- Constructed and fit an XGBoost model for margin prediction.
- Collaborated with engineering to implement algorithms in production; validate and monitor model logging.
- Generated $1.2M in revenue through predicted margin.

Project: Rearchitected an in-house legacy product similarity algorithm with using state-of-the-art technologies.

- Legacy solution involved using a Keras Model to perform weighted sum on fuzzy sub-similarities.
- Replaced a word2vec model with a HuggingFace embedding model finetuned for Semantic Text Similarity.
- Created a pipeline to embed millions of products on a weekly cadence.
- Used FAISS index to retrieve similar products.

Project: Scrape Google SERP for product information to provide product pricing insights.

- Referenced third-party APIs to scrape product pricing information from Google Shopping Page.
- Applied Python graphing libraries to perform EDA to gain a deeper understanding of the domain.
- Applied standard Python libraries to preprocess and clean the data for model consumption.
- Leveraged the above product similar algorithm to find similar products to provide insights on pricing.

Data Science Intern at Direct Supply, Milwaukee, WI **April 2021–July 2022**

Project: Create a report on the monthly cost of medical waste disposed per disposal site.

- Performed OCR on invoices using Tesseract OCR engine with R.
- Performed EDA and calculated Summary Statistics by aggregating extracted data.
- Incorporated parallel computing in pipeline, given the high number of invoices.

Research & Development Intern at Dedicated Computing, Waukesha, WI **June 2020–April 2021**

Project: Build a binary classifier that performs anomaly detection and condition monitoring for computer hardware.

- Preprocessed data by performing Data quality assessment, Data cleaning, and Data Transformation.
- Performed Dimensionality reduction using Principal Component Analysis.
- Implemented Multivariate Anomaly Detection and Mahalanobis distance to classify incoming data.

Projects

Breast Cancer Classification

- Cleaned the Wisconsin Breast Cancer dataset, analyzed patterns; fit and compared four models that classified a tumor as malignant or benign.

AI Research Mentor

- Mentored a group of 4 undergraduate students to create a chatbot for Milwaukee School of Engineering using RNNs and the SQuAD dataset.
- Lead a group of 6 undergraduate students that finetuned a Stable Diffusion model to generate images of themselves,

Education

Milwaukee School of Engineering, Milwaukee, WI ***GPA: 3.7***

Bachelor of Science, Computer Science, May 2022

Undergraduate Research: Explore the effectiveness of Leaky ReLU on AlexNet with ImageNet.

Technical Skills

Python, Databricks, Spark, SQL, Pandas, NumPy, Scikit-Learn, TensorFlow, Keras, PyTorch, Matplotlib, Seaborn, Tableau, Apache Spark, MySQL, PostgreSQL, Git, Docker, Jupyter Notebook.