Benedict Neo

515-388-0996 benedict.neo@outlook.com linkedin.com/in/benedictneo github.com/benthecoder

Education

Iowa State University Expected: Dec. 2023
Bachelor of Science, Major: Statistics, Minor: Computer Science *Ames, IA*

- Cumulative GPA: 4.0
- Coursework: Experimental Design, Probability and Statistics I & II, Data Structures & Algorithms, Discrete Math

Experience

Tesla May 2022 – Aug. 2022
Data Science Intern *Fremont, CA*

- Developed and optimized 10 PowerBI dashboards tracking KPIs for various operations and systems within the factory
- Reduced PowerBI dashboard data refresh time by 80% through data model improvements and optimized DAX queries
- Applied NLP techniques to work orders to create failure codes, enabling more accurate analysis of asset failures
- Investigated work orders and identified opportunities to save $100,000/year in preventive maintenance costs
- Conducted time series analysis on the factory cooling tower systems to identify anomalies and potential problems

Iowa State University Jan. 2022 – May 2022
Undergraduate Research Assistant *Ames, IA*

- Developed and contributed to WEPPR, an R package that emulates the Water Erosion Prediction Project model
- Wrote functional code in R (Tidyverse) to clean and transform 2TB of raw slope, soil, and climate data into tidy format
- Performed exploratory data analysis with principal component analysis (PCA) to identify hidden patterns in the data
- Ran R scripts in a Linux environment to perform complex data transformations with custom functions and packages
- Implemented unit and functional tests and documented code according to best practices with testthat and roxygen2

Projects

GPT-3 Blog Title Optimizer | *Python, Selenium, BeautifulSoup, OpenAI, Weights & Biases, BigQuery, SQL, Streamlit*

- Scraped 200k blog articles from medium.com with Selenium and BeautifulSoup and ingested data into BigQuery
- Utilized OpenAI's API to fine-tune GPT-3 to classify article titles and tracked experiments and dataset with W&B
- Containerized the application with Docker and deployed the model as a web app using Streamlit for users

Readabl | *JavaScript, Python, Svelte, FastAPI, BeautifulSoup, NLTK, Google Cloud Platform (GCP), Docker*

- Led a team of 4 to build a web app that offers readability metrics for google search results with the custom search API
- Implemented concurrent processing in the FastAPI backend to parse webpages in parallel and improve search speed
- Coded Svelte components to call the backend to receive search results and display results in an accessible format

Energytics | *Python, Pandas, Scikit-learn, LightGBM, OpenWeatherMap API, Plotly, Seaborn, Streamlit*

- Led a team of 6 to build a web app that visualizes energy production cost and building energy usage in the US
- Conducted EDA on 60 million records with 18 features to investigate trends, outliers, missing data, and anomalies
- Trained and deployed a LightGBM model to predict building energy usage with user location and building details

LinkedIn Insights | *Python, Numpy, Pandas, thefuzz, NLTK, Plotly, Pyvis, Networkx, Streamlit, wordcloud*

- Developed a web app with 500+ users that allows them to gain insights into their LinkedIn connections
- Applied advanced data cleaning techniques, including fuzzy matching, to improve the accuracy of insights
- Visualized user connections and messages with interactive bar charts, time series plots, and network graphs

Skills

Languages: Python, R, SQL, SAS, JavaScript (React), Java, HTML/CSS, Bash
Libraries: Pandas, NumPy, Matplotlib, Plotly, Tidyverse, ggplot2, scikit-learn, LightGBM, NLTK, PyTorch
Tools: AWS, Google Cloud, Docker, Power BI, Tableau, Looker, Git, Linux, MongoDB, Excel, JIRA, Confluence
Data Science: Data cleaning, Data visualization, Hypothesis Testing, Regression, Classification, NLP

Leadership

President of Google Developer Student Club Aug. 2022 - Present
Iowa State University *Ames, IA*

- Led a team of 12 core team officers to organize and host tech workshops and talks to over 100 students