Benedict Neo

📞 515-388-0996 ✉ benedict.neo@outlook.com in/benedictneo @benthecoder

Education

Iowa State University   Grad Date: Dec. 2023

Bachelor of Science in Statistics, Minor: Computer Science   *GPA: 3.95*
- Coursework: Experimental Design, Bayesian Statistics, Design & Analysis of Algorithms, Large-scale Data Analysis

Experience

Tesla   May 2023 – Aug 2023

Data Science Intern   *Fremont, CA*
- Engineered a time series ML forecasting pipeline predicting factory utilities consumption ($1+ Million/month)
- Architected and deployed Airflow ETL pipelines with Docker on Linux, supporting 10+ PowerBI dashboards
- Designed and optimized MySQL database schemas, incorporating unique indexes for efficient data upserts
- Developed Python packages interfacing with various APIs, streamlining backend operations and data ingestion
- Built an interactive web app with Streamlit for failure analysis using text embeddings and NLP techniques

Tesla   May 2022 – Aug 2022

Data Analyst Intern   *Fremont, CA*
- Leveraged NLP and machine learning algorithms to analyze work orders text to categorize failures
- Developed 10 PowerBI dashboards, enhancing real-time KPI tracking across operations at the factory
- Wrote DAX queries to optimize data models, improving PowerBI dashboard refresh time by 80%
- Analyzed shift hours and badging data to provide insights into optimizing during peak hours

Iowa State University   Jan 2022 – May 2023

Undergraduate Research Assistant   *Ames, IA*
- Led the development of WEPPR, an R package that emulates the Water Erosion Prediction Project model
- Cleaned and transformed 2TB of slope, soil, environment, and climate data into tidy format using R (Tidyverse)
- Conducted exploratory data analysis with principal component analysis (PCA) to uncover patterns in the data
- Wrote unit tests and documented code with testthat and roxygen2 to ensure quality and correctness of code

Projects

 ClassGPT | *Python, LlamaIndex, LangChain, AWS S3, OpenAI, Docker, Streamlit*
- Developed a ChatGPT-like interface for uploading, categorizing, and talking to class lecture slides
- Utilized pypdf, LlamaIndex, OpenAI, and AWS S3 for efficient indexing, embedding, and storage
- Designed an interactive and user-friendly UI with Streamlit, enabling students to learn effectively

 GPT-3 Blog Title Optimizer | *Python, Selenium, BeautifulSoup, OpenAI, Weights & Biases, BigQuery, SQL, Streamlit*
- Scraped 200k blog articles from medium.com using Selenium and BeautifulSoup and loaded data into BigQuery
- Fine-tuned OpenAI's GPT-3 API 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*
- 3rd place winner of HackPrinceton 2022, a web app that augments Google search with readability metrics
- Spearheaded a 4-person team to develop a Svelte and FastAPI web app, hosted with Netlify and Cloud Run
- Improved search speed by implementing concurrent processing to parse webpages in parallel in FastAPI
- Built Svelte components interacting with the backend, rendering search results in a user-friendly design

Skills

Languages: Python, R, SQL, SAS, JavaScript (React), Java, HTML/CSS, Bash
Libraries: Pandas, NumPy, Matplotlib, Plotly, Tidyverse, Scikit-Learn, NLTK, PyTorch, PySpark
Tools: AWS, Google Cloud, Docker, Power BI, Tableau, Git, Linux, Hadoop, Spark, Airflow
Data Science: Data visualization, Hypothesis Testing, Regression, Classification, NLP, Time Series Analysis

Leadership

President of Google Developer Student Club   Aug. 2022 - May 2023

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