

# 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 to track KPIs for various operations within the factory
- Improved data refresh time by 80% in PowerBI dashboards through optimized data models and DAX queries
- Applied NLP techniques to analyze work orders and created failure codes for improved asset failure analysis
- 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 issues

### Iowa State University

Jan. 2022 – May 2022

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
- Executed R scripts in a Linux environment to perform complex data transformations with custom functions
- Wrote 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 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

- Led a team of 4 to build a web app that provides readability metrics for Google search results
- Improved search speed by implementing concurrent processing to parse webpages in parallel
- Built Svelte components to call the backend and display search results in an accessible format

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

- Led a team of 6 to develop a web app that visualizes energy usage and production costs in the US
- Conducted EDA on 60 million records with 18 features to explore trends, outliers, and missing data
- Trained and deployed a LightGBM model to predict building energy usage given custom user input

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

- Built a web app with over 500 users that provides insights into LinkedIn connections and chats
- Improved accuracy of results with advanced data cleaning techniques, including fuzzy matching
- Visualized user connections and messages with interactive bar and line charts and network graphs

## Skills

**Languages:** Python, R, SQL, SAS, JavaScript (React), Java, HTML/CSS, Bash

**Libraries:** Pandas, NumPy, Matplotlib, Plotly, Tidyverse, scikit-learn, NLTK, PyTorch

**Tools:** AWS, Google Cloud, Docker, Power BI, Tableau, Git, Linux, Excel, JIRA, Confluence

**Data Science:** 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