

Benedict Neo

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

Education

Iowa State University

Expected: May 2023

Bachelors of Science, Major: Statistics, Minor: Computer Science

Ames, IA

- Cumulative GPA: 4.0
- Merit Scholarship, Statistics Department Undergraduate Scholarship

Skills

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

Libraries: Pandas, NumPy, Matplotlib, Plotly, Tidyverse, ggplot2, Scikit-Learn, LightGBM, NLTK

Tools: Tableau, Power BI, Git, Linux, MongoDB, Google Cloud Platform, AWS, Docker, Excel

Data Science: Data cleaning, Data visualization, Hypothesis Testing, Regression, Classification, NLP

Experience

Data Science Intern

May 2022 – Aug. 2022

Tesla

Fremont, CA

- Developed and revamped 10 PowerBI dashboards tracking KPIs for various operations and systems within the factory
- Reduced PowerBI dashboard data refresh time by **80%** by rebuilding the data model and writing efficient DAX queries
- Utilized **NLP** techniques on work orders which led to creation of failure codes enabling better insights for asset failures
- Investigated work orders and identified assets where **\$100,000** per year can be saved in preventive maintenance costs
- Cleaned and analyzed shift hours and badging data to provide insights into optimizing parking space during peak hours
- Detected anomalies and trends by performing time-series analysis for the cooling tower systems at the factory

Data Science Intern

Jan. 2022 – May 2022

bitgrit Inc.

Remote

- Led data collection and cleaning, and authored comprehensive solutions, in Python for data science competitions
- Managed the company data science publication and published 20+ technical articles with over 300,000 views 🔗
- Taught 50+ people in a workshop best practices and common tasks for data cleaning in Python

Undergraduate Research Assistant

Jan. 2022 – May 2022

Iowa State University

Ames, IA

- Assisted Dr. Jarad Niemi in the development of **WEPPR**, an R package that predicts water erosion
- Wrote functional code in R (Tidyverse) to clean and transform 100GB of raw soil and land data into tidy format
- Utilized Object-oriented programming principles to create classes that validates and preprocess data
- Authored clear documentation for functions and classes in the package with the roxygen2 package
- Developed unit and functional tests to ensure proper functioning of code with the testthat package

Projects

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

Mar. 2022

- Built a web app that presents insights on energy production cost and building energy usage in the US
- Conducted EDA on 60M records with 18 features to investigate trends, outliers, missing data, and anomalies
- Trained and deployed a LightGBM model to predict building energy usage from user location and building details

🔗 Ensurance | Python, Streamlit, Pandas, Google Cloud, Plotly

Feb. 2022

- Awarded Best Healthcare Hack Powered by Anthem for Georgia Tech's 36 hour datathon
- Led a team of 4 to build an interactive web app that evaluates health risks based on user demographics
- Aggregated data from multiple sources and performed data cleansing for analysis and visualization

🔗 LinkedIn Insights | Python, Streamlit, Pandas, thefuzz, Plotly, Pyvis, Networkx

Jan. 2022

- Developed a web app with **500+ users** that allows them to gain insights into their LinkedIn connections
- Utilized fuzzy matching to clean and manipulate raw user data for more accurate insights
- Visualized user connections with interactive bar charts, time series plots, and network graphs

🔗 Next Word Prediction | R, Shiny, Tidytext, Tidyverse, ggplot2, dplyr

Aug. 2021

- Created a Shiny web app for users to input text and obtain multiple next word predictions
- Analyzed and cleaned over 4 million lines of text corpus data sourced from news, tweets, and blogs
- Utilized the Katz Back-Off (KBO) language model and Markov Chains to generate next word predictions