BENEDICT NEO

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

EDUCATION

Iowa State University   Ames, IA
Bachelors of Science in Statistics, Minor in Computer Science   *May 2023*
- Cumulative GPA: 4.0
- Relevant Coursework: Probability and Statistics, Discrete Mathematics, Design and Analysis of Algorithms
- Statistics Department Undergraduate Scholarship ($1,000 awarded)

SKILLS

Programming Languages: Python, R, SQL, Java, JavaScript (React), HTML/CSS, Bash
Libraries: Pandas, NumPy, Matplotlib, Plotly, Tidyverse, ggplot2, Scikit-Learn, LightGBM
Tools: SPSS, Excel, Tableau, Power BI, Git, Linux, MongoDB, Google Cloud Platform, AWS, Docker
Data Science: Data cleaning, Data visualization, Hypothesis Testing, Regression, Classification

PROFESSIONAL EXPERIENCE

Incoming Data Science Intern   May 2022 – Aug. 2022
Tesla Inc   *Fremont, CA*
- Interpret and analyze high volume manufacturing data from various sources with statistical techniques

Data Science Intern   Jan. 2022 – May 2022
bitgrit Inc.   *Tokyo, Japan*
- Managed data collection and preprocessing, and formulated solutions for data science competitions
- Published technical data science and machine learning articles 🔗 with over 300,000 views
- Taught 50+ people best practices and common tasks for data cleaning in Python in a workshop

Undergraduate Research Assistant   Jan. 2022 – May 2022
Iowa State University   *Ames, IA*
- Assisted ISU Statistics Professor, Jarad Niemi, in the development of WEPPR 🔗, an R package that emulates the Watershed Erosion Prediction Project (WEPP)
- Wrote functional code in R (Tidyverse) to clean and transform 2TB of raw soil and land data into tidy format
- Utilized Object-oriented programming principles to create classes that validates and preprocess data
- Presented at the 2022 ISU 16th Annual Symposium on Undergraduate Research Creative Expression

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
- Performed feature preprocessing and generation to extract additional features from temporal and weather data
- 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

COVID-19 Dashboard 🔗 | *R, Flexdashboard, Tidyverse, Plotly, Docker*   Feb. 2021
- Informed users of key metrics and trends of COVID-19 worldwide with interactive maps and charts
- Implemented automatic daily updates to dashboard using Github Actions and Docker