

# Benedict Neo

☎ 515-388-0996 ✉ [benedict.neo@outlook.com](mailto:benedict.neo@outlook.com)  [linkedin.com/in/benedictneo](https://www.linkedin.com/in/benedictneo)  [github.com/benthecoder](https://github.com/benthecoder)

## Education

### Iowa State University

Ames, IA

*Bachelors of Science in Statistics, Minor in Computer Science (CGPA: 4.0)*

May 2023

- Relevant Coursework: Probability and Statistics, Object-Oriented Programming, Data Structures, 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, Scikit-Learn, Optuna, TensorFlow

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

**Data Science:** Data cleaning, Data visualization, Hypothesis Testing, Regression, Classification, AB testing


## Professional Experience

### Data Science Intern

Jan. 2022 – Present

*bitgrit Inc.*

Tokyo, Japan

- Managed data preparation and cleaning, and formulated solutions for data science competitions
- Wrote technical data science and machine learning articles  with over **400,000 views**
- Taught 50+ people best practices and common tasks for data cleaning in Python in a workshop

### Undergraduate Research Assistant


Jan. 2022 – Present

*Iowa State University*

Ames, IA

- Undergraduate statistics research with Dr. Jarad Niemi, Department of Statistics
- Analyzed over 2TB of soil and land data from the Watershed Erosion Prediction Project (WEPP)
- Wrote functions in R to clean and transform raw data into tidy format to build a statistical model

## Projects

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


Feb. 2022

- Developed a web app that presents insight on energy production cost by state and energy source in the US
- Conducted EDA on 60M records with 18 features and performed feature preprocessing and generation
- 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

- Built 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

## Leadership

### Vice President of American University Program Society

Kuala Lumpur, MY

*INTI Subang College*

Aug. 2019 - Sep. 2020

- Spearheaded a team of 12 to organize a social event with over 100 participants