Chaitanya Kumar Reddy Goukanapalli

Binghamton, New York | (607)-595-8133 | chaitanya3168@gmail.com |
https://www.linkedin.com/in/ckrg2001/ | github.com/CKRG2001 | ckrg2001.github.io/

Education:

Binghamton University, State Univeristy of New York — **Binghamton, New York**
Master of Science in Computer Science — Expected May 2024
Relevant Coursework: Introduction to Machine Learning, Social Media Data Pipeline

VIT-AP University — **Andhra Pradesh, India**
Bachelor of Technology in Computer Science and Engineering — August 2018 - May 2022
Relevant Coursework: Foundations of Data Analytics, Introduction to Machine Learning, DataBase Management Systems

Technical Skills and Certifications:

- **Programming Languages:** Python, R
- **Data Manipulation and Analytics:** Pandas, NumPy, SQL, PostgreSQL, Excel
- **Data Visualization:** Tableau, PowerBI, Matplotlib, Seaborn
- **Machine Learning:** Feature Engineering, Scikit-Learn, TensorFlow - Keras, XGBoost
- **Certifications:** Machine Learning A-Z™: AI, Python -Udemy Certification

Professional Experience:

Shiash Infotech Solutions — **Chennai, India**
Data Analyst Intern — January 2022 - May 2022

- Established a seamless process for regular data export, import and backup to ensure accessibility.
- Conducted an in-depth analysis of database systems, optimizing the efficiency of integration and tested database servers.
- Implemented trend analysis and statistical monitoring to identify valuable insights from the data in Excel and SQL.

Project Experience:

Delivery Duration Prediction: DoorDash | Independent Project | GitHub — September 2023

- Engineered features and optimized data quality using Pandas and Numpy in Python, managing a dataset of size approximately 200,000 x 100.
- Implemented advanced techniques including feature engineering, redundancy removal, and resolution of multicollinearity using correlation matrix.
- Explored the efficacy of six machine learning regression algorithms, tested four distinct feature set sizes, and assessed three different scalers to develop a robust predictive model for estimating delivery duration.

Analysis of Trends on Twitter, Reddit, HackerNews | Group Project — September 2022 - November 2022

- Implemented Python based data crawler to collect an extensive dataset comprising 37 million entries and utilizing PostgreSQL to ensure seamless accessibility.
- Applied advanced Natural Language Processing techniques, including TF-IDF, K-Means Clustering, to analyze the massive dataset.
- Identified technological trends, integrated sentimental analysis to evaluate emotional tone across platforms. Engineered an interactive dashboard using Dash Module with word clouds and daily user activity graph.

COVID-19 Data Exploration Using SQL and Tableau | Independent Project | GitHub — March 2022

- Utilized SQL to dissect COVID-19 datasets from February 2020 to February 2022, employing advanced methodologies such as joins, CTEs, temp tables, window functions, and more to gain deep insights into pandemic effects across global regions.
- Identified a few interesting patterns which constitute the highest infection rate and highest death percentage by country, highest death percentage by continent and performing visualization of these on Tableau.

Stock Market Trend Analysis using Python and Tableau | Independent Project | GitHub — January 2023

- Utilized pandas to manipulate data and analyze stock market data from Apple, Facebook, Google, Tesla and Twitter spanning the years 2016-2020.
- Designed an interactive dashboard to showcase the dynamic company selection functionality for changes in stock price and trading volumes.

Netflix Movies and TV-Shows using Tableau | Independent Project | GitHub — February 2023

- Designed an interactive dashboard highlighting Netflix Movies and TV shows content insights featuring top 10 genre distribution, description, ratings distribution, along with a world map showcasing global distributing trends.
- Created visualizations of top 10 genres, ratings distribution and annual counts of movies and TV shows.