Nithin Sameer Yerramilli

Nithin Sameer Yerramilli | Nithinsameer | ynsameer@gmail.com | Nithin Sameer | +1 240 927 7584

Education

Masters Data Analytics Engineering, George Mason University (GPA: 3.96/4.0) **Aug 2022 - Dec 2024**
Computer Science Engineering, Dayanada Sagar University (GPA: 3.50/4.0) **Aug 2018 - Jun 2022**

Skills

Languages and Tools Python, R, MySQL, Exploratory.io, Tableau, HTML/CSS/JavaScript, Latex, Databricks
Packages Selenium, Scikit-learn, TensorFlow, PyTorch, Seaborn, Scipy, Statsmodel, Keras, NLTK, Langchain

Experience

Graduate Research Assistant - Costello College of Business (GMU) Aug 2023 - Present

- Employed R for comprehensive data manipulation, analysis, and modeling. Performed KNN imputation, variable matching, and regression with the help of powerful R packages to uncover demographic trends.
- Performed statistical analysis on CMIE household data using R, revealing a 12% decrease in spending patterns after app bans by utilizing CEM , regression with survey weighting, and difference-in-difference analysis to assess spending patterns.
- Scraped, formatted & analyzed web data with Selenium; validated and recreated academic research methodologies.

Graduate Teaching Assistant(AI/ML) - Volganeau School of Engineering (GMU) Aug 2023 - Present

- Improved students' comprehension of data mining algorithms, as evidenced by enhanced assignment scores, by providing R programming guidance and support in implementing these concepts.
- Expanded understanding of ethical AI practices, as demonstrated by successfully evaluating over 150+ student essays, by providing in-depth feedback on the implications of AI utilization.

Programmer Analyst - Cognizant Feb 2022 - Jul 2022

- Supported a well-established data science team of 25 as a data science business analyst, documenting project progress, analyzing datasets, and assisting with operations.
- Collaborated with seasoned data scientists on data mining and basic modeling projects, developed skills in exploratory data analysis, statistics, and machine learning, and presented findings through presentations and prototypes.

Data Analyst intern - The Sparks Foundation Jan 2021 - Feb 2021

- Conducted in-depth data analysis and exploration using Python, resolving analysis discrepancies in pattern recognition, leading to enhanced data visualization techniques and a 13% increase in accurate predictions.

Projects

RecipeMate Jan 2024 - May 2024
*(Streamlit, Python, **NLP**)*

- Developed a robust recipe recommendation system leveraging TF-IDF vectorization, OpenAI's GPT-3.5-turbo model for personalized recipe matching and Indianization, and Streamlit for an intuitive frontend, achieving high accuracy in ingredient-based suggestions.

Classifying Celestial Objects with Spectral Data from the Sloan Digital Sky Survey Aug 2023 - Dec 2023
*(**Databricks**, Python, Keras)*

- Optimized Logistic Regression, K-Means, and Decision Trees in Databricks/Python/Keras to classify Sloan Survey celestial objects with 93.15% accuracy, validating redshift's significance via EDA and Deep Learning.

Predicting Recidivism Based on Demographical Bias and Criminal History Jan 2023 - May 2023
*(**R**, Chi-Square Analysis, Stepwise Selection)*

- Executed R-based Stepwise Selection and Chi-Square analysis in logistic regression, achieving AUC values of 0.765 and 0.77 respectively, to identify age at release, gang affiliation, and employment status as significant predictors of recidivism.
- Implemented Decision Tree analysis using R, identifying critical employment thresholds that differentiate recidivism probabilities, with employment status influencing recidivism predictions at a key threshold of 64.3% days employed.

Forecasting Air Quality Index for Mumbai Aug 2022 - Dec 2022
*(Python, **Prophet**, Flask, Streamlit)*

- Developed PollutionPulse web app using Streamlit to predict Mumbai's air quality index (AQI) in real-time utilizing the Prophet model; achieved MAPE consistently under 20%, with enhancements driving it below 10% for validation data.
- Incorporated an hourly updating responsive feedback loop, enhancing prediction accuracy by reducing model error to 13-15%, showcasing strong short-term forecasting performance.

Publications

College Exam Allocation Using MongoDB and Python3 **(GCAT)**

Automatic Exam Answer Checker using OCR and Sentence Embedding **(CENTCON)**