

ANEESH KRISHNA

+1(716) 303-1623 | aneeshkr@buffalo.edu | LinkedIn | Github | Portfolio
75 Montrose Buffalo NY 14214

EDUCATION

University at Buffalo, The State University of New York

Master of Science in Data Science

Key Courses: Statistical Learning and Data Mining, Machine Learning, Deep Learning, Database Management

Buffalo, NY

Dec 2024

Visvesvaraya Technological University

Bachelor of Engineering in Computer Science

Key Courses: Algorithm Analysis, Big Data Analytics, Cloud Computing, Data Structures and Applications

Mangalore, INDIA

July 2020

SKILLS AND CERTIFICATIONS

- **Programming Languages:** Python, JavaScript, C#, SQL, R, MATLAB, Scala.
- **Data Analysis & Visualization:** pandas, numpy, matplotlib, seaborn, Plotly, Tableau, Power BI, Microsoft Excel.
- **Machine Learning & AI:** TensorFlow, PyTorch, scikit-learn, Keras, Deep Learning, NLP, Transformers, CNN, LLM, Time Series Analysis.
- **Big Data & Cloud Technologies:** AWS, GCP, Azure, Hadoop, Spark, Dask, ETL
- **Certifications and Specializations:** Google Data Analytics Specialization, Data Science and Business Analytics, Deep Learning Specialization.

EXPERIENCE

Graduate Research Assistant, University at Buffalo, NY:

July 2024 - Present

- Achieved 98.55% classification accuracy and an F1 score of 0.984 by engineering a dataset of 18,000+ records and optimizing ML classifiers.
- Enhanced ML model performance through advanced feature engineering techniques, achieving up to 98.55% accuracy and an F1 score of 0.984.
- Designed and optimized an NLP-based IDS with BERT, achieving a near-perfect F1 score of 1.0 and an AUC of 1.0 for sequence classification.

Senior System Engineer, Infosys Ltd, Mysore, India:

Oct 2020 - Apr 2023

- Modernized 15+ outdated web components using SPFX, reducing loading time by 30% and enhancing user experience through improved performance and streamlined functionality.
- Monitored migration of over 10 million files and metadata from 67 sites using the Microsoft ShareGate tool to the latest site, meeting stakeholder expectations.
- Developed PowerBI reports and dashboards, providing critical insights and aiding strategic planning.

PROJECTS

Optimizing Healthcare Outcomes through Data-Driven Insights & Machine Learning [click here](#)

- Accomplished 90% size reduction by data-cleaning, encoding (to JSON), and performing feature selection on a large dataset, leveraging libraries such as polars.
- Extracted healthcare insights, spotlighting gender and age-specific trends in diseases, post-operative complications, and hospital length of stay via Plotly, Matplotlib and Seaborn libraries.
- Built predictive models for patient length of stay, attaining an 82% accuracy (R-squared metrics) through Gradient Boosted Trees after Bayesian hyperparameter optimization.

Loan Risk Prediction Model for Small Businesses [click here](#)

- Performed preprocessing on initial Excel data, employing NoSQL to normalize the data and establish a relational database.
- Conducted EDA by querying data with SQLite, producing informative visualizations on business sector distributions and default rates using seaborn and plotly libraries.
- Implemented an ML model for classifying businesses' loan repayment likelihood, achieving a 93% accuracy rate with Random Forest algorithm.

Sports Data Analytics of Cricket World Cup [click here](#)

- Cleaned and preprocessed over 1 million records from ESPNcricinfo of all T20 matches (2015-2023) using Python and Pandas, enhancing data quality for analysis.
- Extracted and analyzed stadium statistics and team performance data for the 2024 T20 World Cup using SQL, Pandas, Tableau, Matplotlib, and Seaborn, providing strategic insights and detailed team profiles.
- Compiled and stored all data insights into CSV files and generated an interactive Tableau dashboard, ensuring efficient data handling and effective communication of key insights.

Cricket Statistics LLM Chatbot [click here](#)

- Developed an AI chatbot using LLM, converting natural language queries into Pandas codes for cricket data analysis.
- Built data processing pipelines with Pandas, handling up to 1 million rows and generating insights on player performance and match statistics.
- Integrated Google Gemini Pro LLM, achieving 95% accuracy in transforming user prompts into executable code for data manipulation.
- Created an interactive Streamlit UI, improving query response time by 50% and enhancing user experience with intuitive input and visualization.