Kumar Baibhav

Buffalo, New York, 14214 | (716) 253-5029 | baibhav06june@gmail.com | LinkedIn | GitHub

EDUCATION

University at Buffalo (SUNY), Buffalo, NY Aug 2023 - Dec 2024
Master of Professional Studies, Data Science, **GPA 4.0/4.0**
Coursework: Machine Learning, Data Mining, Probability & Statistics, Database Management Systems

Manipal Institute of Technology (MAHE), Manipal, India Jul 2018 - Jun 2022
Bachelor of Technology, Civil Engineering, **GPA 8.57/10.0**

TECHNICAL SKILLS

Programming Languages: Python, R Programming, SQL, DAX
Data Science Libraries: Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, Scikit-Learn
Machine Learning Techniques: Regression, Classification, Clustering, Time Series Analysis, Natural Language Processing (NLP), LSTM, Hypothesis Testing, ANOVA
Data Engineering Tools: Git, MLFlow, Docker, AWS, PySpark

EXPERIENCE

Data Science Consultant - Baldwin Richardson Foods & Co, Buffalo, NY Sept 2024 – Dec 2024

- Engineered SARIMAX and Prophet models to forecast demand and optimize inventory management with a MAPE below 20%.
- Integrated external factors like promotions, stock prices, and product prices to enhance predictive accuracy and make data-driven decisions.

Data Science Intern - StatSkew, Remote Mar 2023 - May 2023

- Leveraged Python with APIs, Beautiful Soup, and Selenium for automated web scraping, data extraction, and transforming raw data into actionable insights.
- Predicted insurance purchases using models like Logistic Regression, tree-based models, boosting methods, and ANN, achieving a top F1 score of 0.88 on 380k customer records.
- Engineered features with frequency encoding for high-cardinality categories and resolved class imbalance to boost model performance.

Data Science Intern - CodeClause, Remote Feb 2023 - Mar 2023

- Crafted gradient boosting, decision tree, and random forest models to predict customer churn with 83% accuracy using over 7,000 records; implemented rigorous model validation processes ensuring actionable insights for customer retention.

PROJECTS

Bike Sharing Demand Prediction | *Python, Sklearn, Pandas, NumPy, Matplotlib, Seaborn* GitHub

- Designed and optimized multiple models for bike-sharing demand prediction, achieving 92% accuracy with Random Forest through Grid Search CV.
- Conducted hypothesis testing to analyze seasonal and temporal impacts on rental demand, informing feature engineering.

Cardiovascular Risk Prediction | *Python, Sklearn, Pandas, Numpy, Matplotlib, Seaborn* GitHub

- Analyzed 4,000 patient records through EDA while engineering critical variables that enhanced the performance of cardiovascular risk models utilizing Logistic Regression, Decision Tree, Gradient Boosting techniques.
- Addressed class imbalance using SMOTE-Tomek, achieving 70% accuracy and enhancing recall of the minority class by 56% with Random Forest through hyperparameter tuning.

Sentiment Analysis on Kindle Reviews | *Python, Pandas, Numpy, NLTK, Tensorflow, SKlearn, Gensim* GitHub

- Performed sentiment analysis on Amazon Kindle reviews, utilizing CBOW and TFIDF models with thorough text preprocessing, achieving an accuracy of 58%.
- Improved model accuracy by 17% through the implementation of advanced word vector representations using Word2Vec, followed by a further enhancement of 24% utilizing an LSTM architecture for refined predictions.

Invoice Extractor (InvoiceAI) | *Python, Gemini 1.5 Flash, Streamlit* GitHub

- Developed and deployed an invoice extractor app using Streamlit, leveraging Gemini 1.5 Flash to allow users to upload invoice images and extract information based on custom prompts.

PowerPlay Analytics | *PowerBI*

- Delivered actionable insights using advanced DAX formulas within a sophisticated Power BI dashboard built to track key performance indicators throughout each game of the ICC T20 World Cup 2024.