

KUMAR BAIBHAV

716-253-5029 | baibhav06june@gmail.com | [linkedin.com/in/kumarbaibhav66](https://www.linkedin.com/in/kumarbaibhav66) | Buffalo, NY | github.com/kumarbaibhav6

EDUCATION

Master of Professional Studies: Data Science, University at Buffalo, The State University of New York,
GPA: 4.0/4.0 **Aug 2023 - Expected Dec 2024**

Courses: Python Programming, Probability & Statistics, Machine Learning, Statistical Data Mining, Database Management Systems

Bachelor of Technology: Civil Engineering, Manipal Institute of Technology, Manipal, India
GPA: 3.54/4.0 **Jul 2018 – Jun 2022**

SKILLS & TOOLS

Programming Languages: Python, R, SQL, DAX

ML Libraries: Pandas, Numpy, Matplotlib, Seaborn, Sklearn, Scipy, TensorFlow, Keras, Xgboost

ML Algorithms: Classification, Regression, Clustering, Dimensionality Reduction, Time Series Forecasting

Natural Language Processing & AI: NLTK, Word2Vec, RNN, LSTM, Recommendation Systems

Tools & Environments: Jupyter Notebook, Visual Studio Code, AWS, Google Collab, MLFlow, PowerBI

WORK EXPERIENCE

Data Science Consultant, **Baldwin Richardson Foods Co.** **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 and stock prices to enhance predictive accuracy and drive data-driven decisions.

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

- Leveraged Python and REST APIs for web scraping online sources, applying hypothesis testing to reveal insights.
- Implemented predictive analytics using Decision Trees and ensemble methods like Random Forest; achieved a 30% increase in accuracy of advertisement success prediction by fine-tuning model parameters based on historical performance data.

Data Science Intern, **CodeClause** **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, Jupyter, Streamlit

- Developed and deployed a Random Forest bike-sharing demand prediction model with 92% accuracy, enabling data-driven improvements in resource allocation and operational planning to meet customer demand.

Cardiovascular Risk Prediction: Python, Jupyter

- Engineered and trained patient classification models, including Logistic Regression and Random Forest, achieving approximately 70% accuracy; enhanced recall by 56% through targeted data oversampling to improve patient identification precision.

Sentiment Analysis on Amazon Kindle Reviews: Python, Jupyter, NLTK, Tensorflow

- Created a sentiment analysis model with 82% accuracy using BOW, TF-IDF, Word2Vec, and LSTM techniques to extract insights from unstructured text, driving informed decision-making, and targeted engagement strategies.

Gemini Invoice Extractor: Python, VS Code, Gemini 1.5 Flash, Streamlit

- Developed and launched an invoice extraction application utilizing Streamlit, allowing users to seamlessly upload images for detailed information retrieval, resulting in reduced manual entry errors across financial operations.

CinemAround Database Project: Oracle SQL, PowerBI

- Led a team of four to design a Netflix-like cinema database, leveraging Oracle SQL for insights and building PowerBI dashboards for real-time analytics.