# Kumar Baibhav

Buffalo, NY, 14214

(716) 253-5029 | baibhav06june@gmail.com |linkedin.com/in/kumarbaibhav66/ | github.com/kumarbaibhav6


## EDUCATION

**University at Buffalo (SUNY), Buffalo, NY** Aug 2023 - Dec 2024

Master's of Professional Studies, Data Science, **GPA 4.0/4.0**

**Coursework:** Machine Learning, Data Mining, Probability & Statistics, Database Management Systems

## TECHNICAL SKILLS

**Programming Languages**: Python, R, SQL, DAX, Excel (Pivot Tables, Vlookups)

**Data Science & Machine Learning Frameworks**: Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, Scikit-Learn, Regression, Classification, Clustering, NLP, LSTM, RAG, Hypothesis Testing, A/B Testing, ANOVA, EDA

**Tools**: Azure Data Factory, PySpark, AWS EC2, AWS ECR, Azure Machine Learning, MLFlow, Docker, PowerBI, Git

## EXPERIENCE

**Data Scientist (Co-op) - Baldwin Richardson Foods & Co,** Buffalo, NY Sept 2024 – Dec 2024

- Engineered advanced time series models such as SARIMAX and Prophet to forecast future demand, achieving an average MAPE of 9.8% across all items.
- Executed comprehensive correlation analysis by examining variables like stock prices, inflation rates, holiday schedules, and menu item pricing; uncovered critical trends for improved sales strategies within a 12-month timeframe.
- Developed a user-friendly forecasting prototype that streamlined stakeholder access to predictive analytics, resulting in stakeholders completing forecasts 30% faster and enhancing decision-making efficiency across the organization.

**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.

**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, FastAPI, Docker, Azure*

- Designed and optimized predictive techniques such as Linear Models, Regularization, Tree-based, and Boosting methods, achieving 92% accuracy in bike-sharing demand prediction with Random Forest and Grid Search CV.
- Built an ETL pipeline in Azure Data Factory to fetch and store live weather data in Azure Blob Storage, and developed a FastAPI-based prediction app that accessed this data, leveraging Docker to create images and containers for deployment on Azure App Services.

**Cardiovascular Risk Prediction |** *Python, FastAPI, Docker, AWS, Streamlit*

- 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.
- Created a FastAPI prediction application, containerized it with Docker, deployed the solution using AWS EC2 and Elastic Container Registry, and implemented a user-friendly interface with Streamlit for end-users.

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

- 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.

**SQL Query Generator Using GroqAPI and Llama 8 |** *Python, Groq API, Streamlit*

- Automated the extraction of invoice data from PDFs and stored it in a structured database, ensuring efficient and accurate data management.
- Built a Generative AI system using GroqAPI and Llama 8 to generate SQL queries based on user prompts, streamlining data retrieval processes.