Raghabendra Kumar Shah

Data Analyst · Data Scientist

Odalarevu, East Godavari District, Andhra Pradesh-533210, India

(+91) 6304690343 | shahraghabendra@gmail.com | ragbendra | raghabendrakumarshah | Portfolio

Summary

Aspiring Data Analyst with skills in Python, SQL, and data visualization. Proficient in data preprocessing and statistical analysis using Pandas, NumPy, and Power BI. Passionate about extracting insights and solving real-world problems. Currently building projects to apply analytical skills and seeking an entry-level role to gain industry experience.

Projects

Customer Segmentation Analysis

AP, India

Quantium(Virtual Internship) *Mar. 2025 - Apr. 2025*

- Cleaned and processed over 200,000 transaction records, removing outliers and irrelevant data (e.g., salsa products) to ensure accurate analysis.
- Identified key customer segments contributing to 70% of total chip sales, with Budget Older Families and Mainstream Young Singles/Couples driving revenue.
- Analyzed purchasing patterns across 10+ product brands and 5 pack sizes, revealing a 30% higher preference for premium brands among Mainstream customers.
- Conducted statistical tests (e.g., t-tests) on 50,000+ customer records, validating significant differences in price sensitivity and brand loyalty across segments.

Sentiment Analysis

AP, India

Personal Project *Feb. 2025 - Feb. 2025*

- Developed a binary text classifier achieving over 94% accuracy on validation, effectively categorizing tweets into positive and negative sentiments using LSTM layers.
- Processed more than 14,000 data samples, transforming text into vector embeddings with a vocabulary size of 5,000 words for efficient sentiment analysis.
- Implemented key NLP preprocessing techniques including tokenization, padding sequences to a maxlen of 200, and applying dropout regularization to prevent overfitting during model training.

Movie Recommendation System

AP, India

Personal Project *Jan. 2025 - Jan. 2025*

- Developed a content-based movie recommendation system achieving personalized suggestions by utilizing metadata such as cast, crew, keywords, and genres, processing over 48,000 movies.
- Engineered a feature extraction pipeline transforming textual metadata into vectorized representations, applying CountVectorizer with English stop words to generate a cosine similarity matrix for accurate recommendations.
- Implemented a recommendation function that outputs the top 10 similar movies based on user input, achieving efficient mapping of movie titles to indices and sorting similarity scores in descending order for precise results.

Fake News Detection

AP, India

Personal Project *Dec. 2024 - Jan. 2025*

- Developed a fake news detection model achieving 95% accuracy by applying Multinomial Naive Bayes to 35,000 preprocessed text samples.
- Processed and vectorized over 35,000 text samples using TF-IDF Vectorizer with 50,000 max features, converting raw text into meaningful numerical representations for machine learning.
- Implemented NLP preprocessing techniques, including lemmatization, stopword removal, and regex cleaning, to transform text data into a clean, analyzable format for training and evaluation.

Skills

- **Programming:** Python (Pandas, NumPy, Scikit-Learn), SQL
- **Data Analysis & Machine Learning:** Data preprocessing, feature engineering, statistical analysis, Model training, Evaluation
- **Tools & Visualization:** Jupyter Notebook, VS Code, Excel, Matplotlib, Seaborn, Power BI
- **Databases:** SQL (PostgreSQL, MySQL)
- **Soft Skills:** Communication skills, attention to detail
- **Version Control:** GitHub, Git

Education

BVCEC (Bonam Venkata Chalamayya Engineering College)

Odalarevu, Andhra Pradesh

B.Tech. in Computer Science and Engineering *Oct. 2022 - Apr. 2026*