Raghav Upadhyay

raghav-upadhyay2002 | Raghav Upadhyay |
raghavupadhyay@arizona.edu | +1 (520)-343-6432

Summary

Data Science graduate student with strong foundations in data analysis, machine learning, and statistical modeling. Experienced in building predictive models (LSTM, XGBoost, Random Forest), developing data pipelines, and creating interactive dashboards for decision-making.

Work Experience

Data Analyst Intern Jan 2025 – Aug 2025

*Sudhir Mehrotra & Associates, Chartered Accountants · Bareilly, India (Hybrid)*Analyzed financial statements (balance sheet, income statement, cash flow) to support audit and compliance decisions. Developed forecasting models that improved cash flow estimation accuracy by 15%. Automated data entry workflows using Excel macros and Python scripts, increasing efficiency by 30%. Built interactive dashboards for audit and tax operations, enabling faster insights and reporting. Collaborated with auditors to deliver data-driven recommendations for risk assessment and financial reporting.

Projects

Bitcoin Market Behavior: Forecasts Across Timeframes GitHub Repo

Analyzed Bitcoin's historical price behavior and built predictive models for short-term (3M), mid-term (1Y), and long-term (5Y) forecasts using machine learning and statistical methods. Compared performance of LSTM (TensorFlow/Keras) and XGBoost regression models, and evaluated moving averages (7-day, 45-day) for trend detection. LSTM captured long-range cycles effectively, while XGBoost excelled in short- and mid-term forecasts.

Hate Speech Detection Model GitHub Repo

Developed a machine learning model to classify hate speech in online text using a Kaggle dataset, achieving 92%. Implemented preprocessing (tokenization, stop-word removal) and applied algorithms (Logistic Regression, Random Forest) to improve classification accuracy.

Sentiment Analysis Annotation GitHub Repo

Performed sentiment analysis on Amazon reviews to classify feedback as positive/negative. Annotated data with 95% inter-annotator agreement, then trained a neural network in Python to achieve robust binary classification performance.

Commonsense Reasoning with Pre-trained Models GitHub Repo

Benchmarked large language models (OPT-1.3B, RoBERTa-large-MNLI) on commonsense reasoning using the PIQA dataset. Achieved 84% accuracy with RoBERTa-large-MNLI, outperforming baseline models in reasoning tasks.

Education

2024 – Present **University of Arizona**, Tucson, AZ
Master of Science in Data Science (GPA: 3.67/4.0)

2020 – 2024 **SRM University**, Chennai, India
Bachelor of Technology in Computer Science
Specialization: Software Engineering (GPA: 3.4/4.0)

Skills

Programming	Python, R, SQL, C++, HTML, CSS
Data Analysis & Processing	Pandas, NumPy, Tidyverse, Excel (Advanced), Data Cleaning, Feature Engineering
Machine Learning & AI	Scikit-learn, XGBoost, Random Forest, Logistic Regression, LSTM, Neural Networks, Natural Language Processing (NLP)
Deep Learning Frameworks	TensorFlow, Keras, PyTorch
Data Visualization	Matplotlib, Seaborn, ggplot2, Power BI, Tableau
Databases	MySQL, PostgreSQL, NoSQL (MongoDB)
Tools & Platforms	Jupyter Notebook, Git/GitHub, yfinance, Excel Macros, Docker (basic), Linux/Unix