Ravi Kumar

✉ kumarsoniravi705@gmail.com 📞 +91 9608710567 📍 Bengaluru, Karnataka

in linkedin.com/in/ravi-soni123 github.com/RaviSoni804426 k kaggle.com/ravikumarsonis 🌐 ravisoni.vercel.app

Professional Summary

Data Science undergraduate at **Indian Institute of Technology (IIT) Guwahati** (CPI: 7.1/10) with hands-on experience in **machine learning, NLP, LLM integration, and advanced SQL analytics**. Proficient in Python, XGBoost, LSTM, TensorFlow, and Retrieval-Augmented Generation (RAG). Delivered measurable impact: **60% HR effort reduction, 72% prediction accuracy, 30% SQL performance gain**. Hackathon winner and Kaggle-certified SQL practitioner.

Technical Skills

- **Languages:** Python, SQL, C++, JavaScript
- **Machine Learning / Deep Learning:** Regression, Classification, Clustering, Random Forest, XGBoost, Gradient Boosting, LSTM, Neural Networks, Cross-Validation, Hyperparameter Tuning, Feature Engineering, TensorFlow, Keras, Scikit-learn
- **NLP / LLM / Generative AI:** RAG, TF-IDF, Cosine Similarity, Vector Embeddings, Semantic Search, LLM Integration, Prompt Engineering, Transformers
- **Data Analysis:** EDA, Statistical Analysis, Data Wrangling, Time Series, A/B Testing, Pandas, NumPy, Matplotlib, Seaborn
- **Databases:** MySQL, PostgreSQL, Advanced SQL (CTEs, Window Functions, Joins, Query Optimization)
- **Tools:** Git, GitHub, Streamlit, FastAPI, Power BI, Jupyter Notebook, Google Colab

Projects

- **RAG-Based Knowledge Retrieval System** *Python, LLM, RAG, Vector Search, NLP*
  Built a **Retrieval-Augmented Generation (RAG)** pipeline using embedding-based semantic search and vector indexing for contextual document retrieval, improving answer relevance by **40%**. Integrated open-source LLMs for context-aware Q&A over custom corpora. ⊙ GitHub

- **AI Resume Screener** *Python, NLP, TF-IDF, Cosine Similarity, Streamlit*
  Architected an automated resume ranking system using **TF-IDF vectorization** and **cosine similarity**, reducing manual HR shortlisting effort by **60%**. Deployed interactive **Streamlit** dashboard for real-time recruiter evaluation. ⊙ GitHub

- **Voice AI Assistant** *Python, LLM Integration, FastAPI, Speech Recognition*
  Developed a speech-enabled AI assistant integrating real-time **speech recognition** with **LLM response generation** via **FastAPI**, achieving sub-second latency for real-time conversational performance. ⊙ GitHub

- **Walmart Sales Data Analysis** *Advanced SQL, MySQL, CTEs, Window Functions, Power BI*
  Analyzed large-scale retail data using **CTEs, window functions, joins, and aggregations**. Identified peak sales cycles and optimized SQL query execution by **30%** via indexing and query restructuring. ⊙ GitHub

- **Stock Price Prediction System** *Python, XGBoost, LSTM, Time Series, Backtesting*
  Built end-to-end **ML pipeline** on 5+ years of financial data with **15+ technical indicators** (RSI, MACD, Bollinger Bands). **XGBoost + LSTM** achieved **72% accuracy**; backtesting engine evaluated Sharpe ratio and risk-adjusted returns. ⊙ GitHub

Education

B.Sc. (Hons.) Data Science & Artificial Intelligence *Indian Institute of Technology (IIT) Guwahati*
Roll No: 23035010346 — CPI: 7.1/10 — 2023 – 2027

Senior Secondary (Class XII) *Bihar School Examination Board (BSEB)* 85% — 2022

Achievements & Certifications

- **Innovation Hackathon Winner** – Physics Wallah (2024): Secured **1st place** among 500+ participants with an AI-driven solution.
- **Top 20%** – Kaggle Machine Learning Competition: Ranked globally in top quintile demonstrating applied ML and feature engineering skills.
- **Advanced SQL Certification** – Kaggle Learn (2025): CTEs, window functions, joins, and query optimization.
- **Freelance Data Analyst**: Delivered Python & SQL dashboards and automation solutions for external clients.