# SHABANA BEGUM

Data Scientist | Python Developer

shabanads14@gmail.com +91 9669090098 www.linkedin.com

## SUMMARY

As a recent graduate with a strong foundation in data science, I am eager to apply my skills in real-world scenarios. My academic projects and internships have given me hands-on experience in data analysis, machine learning, and data visualization. I enjoy working closely with team members to understand project requirements and develop data-driven solutions that make a difference.

## TECHNICAL SKILLS

- **Programming Languages/Frameworks**: Python, C, SQL, PySpark, Java (Basics), Flask, Streamlit.
- **Platforms:** GCP, AWS, MongoDB, Snowflake, Databricks, Microsoft Azure.
- **Business Intelligence/Visualization:** Power BI, Seaborn, Matplotlib, Plotly, Power Query, Excel.
- **Algorithms and Techniques:** Data Mining, Data Cleaning, Text Mining, Data Manipulation, Data Processing, ETL, DSA, EDA, NLP, Regression, Classification, Clustering algorithm, ANN, CNN, RNN, LLMs, GAN, OpenAI, Gemini.

## PROJECT(S)

### MRI-Based Brain Tumor Detection — CNN | Deep Learning

**Problem:** Develop a system that accurately classifies brain tumor using MRI images.

- Developed a multi-class brain tumor classification model using CNNs, achieving accurate classification across 4 categories.
- Enhanced image variability and reduced model overfitting through advanced data augmentation techniques.
- Utilized OpenCV, TensorFlow/Keras, and Scikit-learn to build and train the model, improving diagnostic accuracy.

### Chat With Multiple Documents — GenAI | Google Generative AI

**Problem:** Develop a Streamlit web application for querying information from PDF documents using NLP

- Developed a Streamlit app using the Gemini model to query PDF content with NLP techniques, processing over 1,000 documents with high efficiency.
- Achieved 95% accuracy in PDF text extraction and querying by integrating the Gemini model for NLP
- Delivered a user-friendly interface, improving response time by 40% and enabling seamless interactions.

### Medi-Quality Article Clustering — NLP | Machine Learning

**Problem:** Developing a system to effectively and accurately cluster medical articles based on quality and relevance

- Utilized Python-based NLP and ML techniques to cluster 1,000+ medical articles efficiently.
- Solved challenges in optimal clustering and preprocessing, achieving 90% clustering accuracy.
- Improved article organization and navigation by 25%, enhancing accessibility and structured

## EDUCATION

Bachelor of Computer Application (BCA), Computer Programming — Aug 2021 - Jun 2024

Osmania University Hyderabad — CGPA: 8.4/10

## RELEVANT COURSEWORK

- Advanced Machine Learning
- Business Intelligence
- Visualizing Big Data Trends
- Linear Algebra
- Predictive Modeling
- Data Mining Strategies
- Natural Language Processing

## ACHIEVEMENTS

- Top 5 percentile in a Kaggle machine learning and deep learning competition.
- Completed a Data Science course at NareshIT, Hyderabad (2025).
- Awarded gold distinction in a NASSCOM Data Science exam (2025).