# Sanya Shourya

📞 +91 9263483611   ✉ sanyashourya1803@gmail.com   📍 Bihar, India   in LinkedIn   ◯ GitHub

## Work Experience

**Data Science Intern, Cognifyz Technologies** *Jun 2024 – Jul 2024*
Remote Internship
- Executed comprehensive data cleaning and preprocessing on real-world datasets to ensure quality and consistency.
- Performed exploratory data analysis (EDA) to uncover hidden patterns and support data-driven business insights.
- Contributed to the design, training, and evaluation of machine learning models, improving performance and interpretability.
- Delivered deployable end-to-end data science solutions independently, showcasing initiative, critical thinking, and effective remote collaboration.

## Education

**VIT Bhopal University**, Bhopal, India *2022 – Present*
B.Tech in Computer Science and Engineering (AI & ML), CGPA: 8.85

**Carmel High School**, Patna, India *2020 – 2022*
Indian School Certificate (ISC), Percentage: 81%

**May Flower School**, Patna, India *2020*
Matriculation, Central Board of Secondary Education (CBSE), Percentage: 95.6%

## Projects

**Sales Data Analysis using SQL and Python** *2024*
- Designed and implemented a pipeline to clean and preprocess **10,000+ sales records** using Python (Pandas) in Jupyter Notebook.
- Loaded transformed data into a MySQL database with **5+ linked tables**, enabling structured analysis and storage.
- Executed **20+ advanced SQL queries** to extract insights such as **top 3 categories** and **peak revenue months**.
- Used `dotenv` and SQLAlchemy to securely manage credentials and build a modular codebase with **100% script automation**.

**Next Word Prediction using LSTM** *2025*
- Built a word-level next word prediction model using TensorFlow and LSTM, trained on a dataset of **5,000+ motivational and philosophical quotes**.
- Applied custom tokenization and generated **10,000+ n-gram sequences**; limited vocabulary to **3,000 tokens** to optimize memory usage.
- Designed a 3-layer model (Embedding, LSTM, Dense with Softmax) and trained it for **30 epochs** using the Adam optimizer and categorical cross-entropy loss.
- Achieved **66% prediction accuracy**; model can generate up to **15 semantically relevant words** based on a user-provided input phrase.

## Certifications

- Applied Machine Learning with Python – Coursera *Dec 2023*
- Privacy and Security in Online Social Media – NPTEL *May 2024*
- GEN AI Using IBM Watsonx – IBM *June 2025*
- Intoduction to Deep Learning – Infosys *June 2025*

## Technical Skills

**Programming Languages:** Python, Java, SQL
**Libraries & Tools:** Scikit-learn, TensorFlow, OpenCV, Pandas, Matplotlib, Streamlit, Excel, Git
**Concepts:** Machine Learning, EDA, Transfer Learning, NLP, Data Visualization, Dashboards

## Languages

- **English:** Proficient
- **Hindi:** Native/Bilingual
- **Japanese:** Basic