

SONALIKA SINGH

8434756189 | singhsonalika5@gmail.com | [LinkedIn](#) | [GitHub](#)

Education

Master of Mathematics, Indian Institute of Technology Madras	7.43 CGPA (2023-2025)
Bachelor of Science (Mathematics), Patliputra University, Patna	62.2% (2018–2021)
Senior Secondary, Bihar School Examination Board	68% (2016–2018)
Secondary, Central Board of Secondary Education	9.4 CGPA (2015–2016)

Experience

Data Analyst Intern, ERNET India, IIT Madras Research Park July 2024 – June 2025

- Developed groundwater prediction models (MLR, RFR, ANN, GRU, LSTM) on **2.5K+** data points for sustainable planning.
- Automated **FSOC** outage analysis, detecting 4 major causes from **180+** days of signal data in Kohima, Nagaland.
- Created **10+** Python visualizations and KPI reports to enhance internet service in remote hilly regions.

Winter Intern – Applications of ML Techniques, E&ICT Academy, IIT Guwahati Dec 2024 – Jan 2025

- Built and tuned **5+** deep learning models (FFNNs, CNNs) on MNIST, CIFAR-10 using TensorFlow/Keras; improved accuracy by **15%**.
- Fine-tuned VGG and ResNet models; evaluated using **6+** metrics (ROC, PR curves, AUC, confusion matrix).

Projects

Multi-label EHR Classification (DA5401 ML Challenge): Built a deep learning model (4 dense layers, LeakyReLU, Dropout) to perform multi-label classification on **200K+** EHR embeddings, achieving a Micro F_2 score of **0.8183** on validation data and mastering threshold tuning, data engineering, and evaluation techniques.

Sequence Modeling for Language Decoding: Built and optimized Seq2Seq models (RNN, LSTM, GRU) for Roman-to-Devanagari transliteration, enhancing validation accuracy by **8%** through Bayesian hyperparameter sweeps, attention mechanisms, and beam search decoding on **10,000+** samples.

Multilingual Language Classifier: Processed **27** roman-script languages and built a Naive Bayes classifier; achieved high language prediction accuracy after RDA-based feature reduction and hyperparameter tuning.

Music Recommender System: Preprocessed **1M+** song lyrics using NLTK (cleaning, tokenization, stopword removal) and engineered a TF-IDF + cosine similarity model (**5K** features), deployed via Streamlit with Spotify API integration for interactive **top-5** song recommendations.

Relevant Courses and Skills

Core Courses: Data Structures and Algorithm, Foundation of Machine Learning, Introduction to Deep Learning, Programming Lab, Data Analytics Laboratory

Mathematics: Linear Algebra, Probability, Numerical Analysis, Multivariable Calculus, ODEs, PDEs, Applied Statistics, Numerical Optimization

Online Courses: **Data Science** Job Simulation (Lloyds Banking Group), **Data Analytics** Job Simulation (Deloitte), Intro to AI & MathLab (IITM Summer School), Master in Data Science (Udemy), Communication Skills (TCS iON), Advance Diploma in Computer Applications (MICT Wizard-Tech), Power BI Workshop

Skills: Python, LaTeX, SQL, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Machine Learning, Deep Learning, Power BI, RDBMS, MS Excel

Leadership, Activities, and Achievements

Scholastic: AIR **417** in IIT-JAM (Mathematics); Qualified **JEE** (2018, 2019) and **GATE** (Data Science, 2025).

Leadership: Volunteer at Forays 2023 (IIT Madras), managing event operations and participant coordination.

Sports: Demonstrated teamwork, leadership, and competitive spirit by representing hostel in **Football** and **Cricket**; secured four positions in inter-hostel Cricket tournaments.