Sannidhya Das

+91 8910049917 | sannidhyadas0howrah@gmail.com | linkedin.com/in/sannidhya-das3 | github.com/SannidhyaDas

Education

St. Xavier's College (Autonomous) — Kolkata, WB
M.Sc. in Data Science (Current CGPA: 6.93) — *2024 – 2026*
Relevant Coursework: Machine Learning, Deep Learning, Statistical Inference, Linear Algebra, Big Data Analytics.

University of Calcutta — Kolkata, WB
B.Sc. in Statistics (Hons) (CGPA: 8.1/10) — *2020 – 2023*

Technical Skills

Languages: Python, R, SQL (Advanced), HTML/CSS
ML & Deep Learning: Scikit-Learn, XGBoost, PySpark, PyTorch, Keras, Transformers, Pandas, NumPy, OpenCV
GenAI & NLP: LLM Fine-tuning, Prompt Engineering, CLIP, Whisper ASR, RAG Pipelines, HuggingFace
Developer Tools: Git/GitHub, Docker, FastAPI, Streamlit, Power BI, PostgreSQL, MySQL, n8n

Experience

Research Intern — May 2025 – Aug 2025
Jadavpur University, Dept. of CSE — *Kolkata, WB*

- Authored two research papers on **Multimodal NLP** and depression analysis (accepted at SPELLL & RANLP).
- Engineered a **Zero-Shot Depression Severity pipeline**; optimized **BART-MNLI** for pseudo-labeling and fine-tuned **DistilBERT**, boosting classification accuracy to **92%** over baseline models.
- Pioneered a visual approach to Speech Emotion Recognition (SER) by converting audio to **Spectrograms**, employing CNNs to outperform traditional statistical feature extraction methods.

Publications

From Voice to Vision: A Comprehensive Approach to SER | *SPELLL 2025 (Accepted)* — 2025

- Proposed a novel SER framework treating spectrograms as visual surrogates for emotion; achieved **0.4975 Macro F1** using class-wise feature selection on text-audio pairs.
- Designed a multimodal fusion architecture using **CLIP and Whisper** embeddings to decode the inaudible spectrum of human emotion, benchmarking against state-of-the-art implementations.

Identifying Severity of Depression in Forum Posts | *RANLP 2025 (Link)* — 2025

- Implemented a robust two-stage classifier using **Zero-Shot labeling (BART-MNLI)** and **DistilBERT** fine-tuning.
- Addressed extreme data scarcity in the shared task environment, achieving **significant accuracy gains** (28.9% Official Acc) without gold-standard training data.

Projects

Text-to-SQL GenAI App | *Gemini Pro, SQLAlchemy, MySQL, Streamlit* — 2025

- Developed an NLP-to-SQL engine using **Gemini Pro**, enabling non-technical users to query databases via natural language.
- Integrated **NeonDB** for cloud-scale storage and optimized prompt engineering to handle complex Joins/Aggregations.
- Deployed the full inference pipeline on Streamlit, reducing query formulation time by approximately **80%**.

redBus Data Decode - Demand Forecasting | *XGBoost, LightGBM* — 2025

- Secured **Rank 21 out of 694** (Top 3%) in a national predictive modeling hackathon by optimizing Mean Squared Error.
- Conducted feature engineering on temporal route data to identify seasonality, improving model robustness.
- Built an ensemble regressor using XGBoost and LightGBM, automating the training-to-inference pipeline.

SolMate-AI Assistant | *Gemini Pro, APIs, Python* — 2025

- Architected a personal assistant agent orchestrating **Gemini API**, OpenWeather, and SerpAPI for real-time context.
- Implemented context-aware prompt chaining to generate dynamic travel itineraries and daily summaries.

Achievements & Extracurriculars

Rank 17: National Data Visualization Hackathon (MoSPI, Govt. of India).
Rank 21: redBus Data Decode Hackathon 2025 (Top 3%).
Leadership: Logistics Head for DatAspire 2025; Student Volunteer for ICDMAI 2025.
Public Speaking: State-Level Elocution Competitor representing the PG Data Science Department.