SANNIDHYA DAS

https://www.linkedin.com/in/sannidhya-das3/
Andul-Mouri, Howrah, West-Bengal, pin - 711302 , IN
+91 8910049917
sannidhyadas0howrah@gmail.com
GitHub Repository

PROFILE SUMMARY

Dedicated Data Science professional with a strong foundation in statistics and machine learning, experienced in building multimodal models, vision-language systems, and speech processing pipelines. Proficient in Python, R, SQL, Hugging Face, Scikit-learn, and XGBoost, with expertise in predictive modeling, data visualization, and process automation. Passionate about transforming raw data into strategic insights that drive innovation and deliver measurable value.

TECHNICAL AND THEORITICAL SKILLS

- Programming Languages: Python (2 Yrs), R (3+ Yrs), SQL (1.5 Yr)
- Data Visualization: Ggplot2, Power BI (1 Yr), Matplotlib, Seaborn
- Data Analysis & Processing: Pandas, NumPy, Excel, Feature Engineering
- Statistical Methods: Regression, Clustering, Classification, Exploratory Data Analysis
- Machine Learning & Deep Learning: Scikit-Learn, XGBoost, Computer Vision, NLP, Keras, Pytorch
- Databases: Oracle 11g, PostgreSQL
- Big Data: Hadoop, Spark
- Other Tools: Jupyter Notebook, Colab Notebook, Kaggle, Git, GitHub, Hugging Face, AI Tools

EDUCATION

2024 - 2026
ST. XAVIER'S COLLEGE (AUTONOMOUS) KOLKATA
- MSc Data Science

2020 - 2023
CALCUTTA UNIVERSITY
- BSc. Statistics (Honours)
- CGPA: 8.1 / 10.0

INTERNSHIP

19th MAY 2025 - Ongoing
JADAVPUR UNIVERSITY, DEPARTMENT OF CSE
- Project-Based Intern

SOFT SKILLS

- Attention to Detail Nature
- Collaborative Teamwork
- Time Management
- Leadership
- Effective Communication
- Critical Thinking

CERTIFICATIONS

- MS Excel for Business Intelligence
- Complete SQL Bootcamp - Udemy
- Python for DS & ML - Udemy
- Statistical Modelling (in R) - Udemy
- Big Data Bootcamp - NIELIT, Kolkata

PROJECTS

Spam Classification Project

- Implemented a machine learning pipeline for spam detection using Python. Preprocessed text data, engineered features, and applied algorithms like Naive Bayes, Logistic Regression, and ensemble methods. Achieved 95% accuracy and 90% precision, demonstrating expertise in text analytics and classification models.
- Tools: Numpy, Pandas, Scikit-learn, Streamlit.

Movie Recommendation System

- Developed a movie recommendation engine based on content similarity. Processed TMDB movie datasets with over 5000 entries, engineering features such as genres, keywords, and cast. Utilized cosine similarity to recommend movies based on user-selected titles.
- Tools: Python, pandas, scikit-learn, NLP, JSON handling, Streamlit.

Identifying Severity of Depression in Forum Posts using Zero-Shot Classifier

- Implemented a two-stage depression severity classifier using zero-shot labeling (BART-MNLI) and DistilBERT fine-tuning, achieving 92% internal and 28.9% official accuracy without gold-standard training data.
- Tools: Python, Google Colab, Transformers,LLM, NLP, Hugging Face

AI-Driven Emotion Detection from Spoken Language

- Developing a CLIP-based multimodal emotion recognition system using facial images and audio transcripts (via Whisper ASR), achieving 88.9% accuracy; currently extending to multilingual and more complex datasets for broader applicability.
- Tools: Python, Kaggle, VLM, Speech to Text Model, Hugging Face