Jordan DSouza

Mumbai | +91 8291946071 | jordandd01@gmail.com | linkedin.com/in/jordanddsouza/ | github.com/jordan-dsouza/

EDUCATION

E and ICT Academy, IIT Kanpur Remote
Post Graduate Diploma in Data Science September 2023 – July 2024

St. Francis Institute of Technology Mumbai
BE Electronics and Telecommunication Engineering 8.85 GPA July 2019 – May 2023

WORK HISTORY

Slash Mark Internships Remote
Machine Learning Engineer Intern November 2023 – February 2024
- Built ML models to extract insights from datasets, achieving an R^2 score of 0.6689
- Performed hypothesis testing and feature engineering to improve variable representation
- Conducted statistical analysis to support conclusions and drive model decisions
- Communicated findings with data visualizations (charts and summary tables)

Code Clause HR Remote
Data Science Intern August 2023 – September 2023
- Created a logistic regression model that achieved 91% classification accuracy
- Conducted data cleaning, outlier handling, and validation to improve reliability
- Applied core machine learning algorithms in a business-oriented context
- Documented model assumptions and tested against real world inputs

PROJECTS

Brain Cancer Detection
- Built a deep learning model using **TensorFlow** to classify grayscale scans for brain cancer detection
- Applied **OpenCV** for image preprocessing and implemented early stopping
- Fine-tuned MobileNetV2 architecture, achieving 82% test accuracy
- Visualized model learning using training plots and pixel distributions for interpretability

Hospital Cost Predictor
- Designed an end-to-end machine learning pipeline to predict healthcare costs
- Conducted data cleaning, visualization and dashboarding using **Power BI**
- Built an **XGBoost** model achieving a test accuracy of 89.17%
- Applied real world cases to test the model

Intelligent Q&A System for Research Papers with Gemini 2.0 Flash
- Designed a **RAG** system using **Google Gemini API** to process and query long-context academic papers
- Implemented data parsing, indexing, and retrieval pipelines with **Python** and **LangChain**
- Integrated **LLM-based summarization** and contextual answering for research insights
- Optimized performance for large datasets with focus on scalability and response accuracy

Linguistic Feature Extraction and Text Analysis Pipeline
- Built a Python-based **NLP pipeline** to extract readability and sentiment metrics from 100+ news articles
- Used **Beautiful Soup** for web scraping and applied **NLTK** and regex for polarity, subjectivity, and Fog Index calculation
- Extracted linguistic features (pronouns, sentence length, syllables) via rule-based logic and dictionary lookups
- Consolidated results into structured **Pandas** data frames for reporting and further analysis

SKILLS

Technical Skills: Python, SQL, Excel, Power BI, Tableau
Libraries and Frameworks: TensorFlow, Keras, Scikit-Learn, Pandas, NumPy, PyTorch, OpenCV, NLTK
Machine Learning: Computer Vision, Natural Language Processing, Deep Learning, LLM, RAG
Other Skills: Data Cleaning, Feature Engineering, Data Visualization, and Database Management

CERTIFICATIONS

Machine Learning in Excel – 365 Data Science November 2024
The Data Science Course: **Complete Data Science Bootcamp** – Udemy August 2023
Neural Networks and Deep Learning – Coursera September 2021