ANGELA SUPERNA A

DATA SCIENCE | MACHINE LEARNING | AI

PROFILE

Motivated MSc Data Science graduate with hands-on experience in Machine Learning, Deep Learning, NLP, and Big Data analytics. Skilled in developing and deploying end-to-end AI solutions using Python, TensorFlow, PyTorch, and AWS. Demonstrated ability to build impactful projects including AI-powered transcription, resume screening bots, and NLP toxicity filters. Strong analytical and problem-solving skills with a passion for transforming data into actionable insights and deploying scalable ML models for real-world applications.

PROJECTS

M**indScribe – AI-Powered Audio/Video Transcriber**
Technologies: Streamlit, SpeechRecognition, AWS EC2, Jenkins, Python

- Developed an end-to-end AI transcription system capable of converting audio/video inputs into structured text.
- Integrated SpeechRecognition for speech-to-text conversion and Streamlit for an interactive UI.
- Deployed the application on AWS EC2, configuring Jenkins CI/CD pipelines for automated updates and continuous integration.
- Enhanced usability by adding real-time transcription, file upload support, and accuracy optimization techniques.

Agentic Resume Bot – AI-Powered RAG Assistant
Technologies: LangChain, FAISS, GPT Models, Python

- Built an AI-driven resume screening assistant leveraging Retrieval-Augmented Generation (RAG) to match resumes with job descriptions.
- Used LangChain for prompt orchestration and FAISS for efficient vector-based candidate retrieval.
- Integrated GPT-based summarization for generating concise candidate summaries and insights.
- Improved HR screening efficiency by automating shortlisting based on skill relevance and contextual analysis

.

Toxicity Shield – AI-Powered Toxic Comment Filter & Rewriter
Technologies: Flask, BERT, T5, Hugging Face Transformers, Google Translate API

- Created a Flask-based NLP web app to detect, analyze, and rewrite toxic or offensive comments in real time.
- Implemented BERT for sentiment and toxicity classification and T5 for generating polite rephrased alternatives.
- Added multilingual detection and translation to handle regional and global languages.
- Improved community engagement by introducing user-friendly suggestions for positive communication.

CONTACT

+91 9363088296
superna2004@gmail.com
Chennai
https://github.com/AngelaSuperna29
https://www.linkedin.com/in/angela-superna-5b992b35b/

EDUCATION

2024 - 2026
MASTER OF DATA SCIENCE:
Loyola College , Chennai.

2021 - 2024
BACHELORS OF MATHEMATICS:
Stella Maris College, Chennai.

SKILLS

- Programming Languages: Python, SQL, NoSQL
- Machine Learning & AI: TensorFlow, Keras, PyTorch, Scikit-learn, CNNs, Transformers (BERT/T5),
- Generative AI
- Mathematics
- Big Data & Cloud: PySpark, AWS EC2, Jenkins, Docker
- Data Analysis & Visualization: PoweBI, Pandas, NumPy, Matplotlib, Seaborn
- Deployment & Tools: Streamlit, Flask, Jupyter Notebook, Google Colab, Git/GitHub
- Soft Skills: Problem Solving, Critical Thinking, Time Management, Communication

PROJECTS

Urban Sound Classification using Deep Learning

Technologies: CNN, Librosa, TensorFlow, Python

- Designed a Convolutional Neural Network (CNN) model to classify urban sound categories using MFCC audio features.
- Preprocessed and augmented datasets to reduce overfitting and improve generalization.
- Achieved high model accuracy through feature scaling, dropout layers, and hyperparameter tuning.
- Visualized classification results and performance metrics to evaluate model behavior.

Sales Prediction Web App using Machine Learning

Technologies: Streamlit, Scikit-learn, Pandas, NumPy, Matplotlib

- Developed a machine learning-based sales forecasting web app using regression models.
- Implemented data preprocessing, encoding, and scaling pipelines to improve model performance.
- Enabled real-time prediction and interactive visualization of sales trends within Streamlit.
- Deployed the app for accessible business analytics and data-driven decision-making.

Uber Ride-Sharing Data Analysis using Big Data

Technologies: PySpark, Apache Spark, Python, Pandas

- Analyzed large-scale Uber ride datasets to extract key insights on demand patterns, pricing, and temporal trends.
- Used Apache Spark for distributed data processing and MLlib for regression-based price prediction.
- Optimized query performance for large datasets and visualized surge trends for business insights.

Customer Response Prediction for Marketing Campaigns

Technologies: Random Forest, Scikit-learn, Pandas, Seaborn, Matplotlib

- Built a Random Forest classifier to predict customer responses to marketing campaigns.
- Addressed class imbalance using oversampling and applied feature selection to improve prediction accuracy.
- Conducted exploratory data analysis (EDA) and visualized campaign performance trends.
- Delivered actionable insights for optimizing customer targeting and marketing strategies.

INTERNSHIP EXPERIENCE

Data Science Intern

Thaimozhi Kalvi Private Limited (Nackl)

May 2025 – June 2025

- Developed Nack DNA scoring system to track user engagement and performance.
- Built “Daily Zinga Dare” gamified feature to increase Gen Z participation.
- Created an AI chatbot for user interaction and behavior-based challenges.
- Analyzed user engagement data and contributed to interactive Power BI dashboards.

CERTIFICATIONS

Data Science Using Python – Department of Computer Science, Stella Maris College, Chennai.

LANGUAGES

- English: Fluent
- Tamil : Fluent