Naveen Kumar Tripathi

Data Scientist

navinkumartripathi88@gmail.com +91 9174773484

Career Objective

Driven Data Scientist with 3.3 years of experience in NLP, Deep Learning, LLMs, Computer Vision and ML. Proficient in Python, RAG, OpenAI, and leveraging AI to deliver intelligent, data-driven solutions that transform business operations. Passionate about applying cutting-edge technologies to optimize processes, enhance decision-making, and fuel growth in a forward-thinking environment."

Experience

Tiger Analytics September 2021-Present

- Developed ML/AI /NLP models using LLM, RAG, and LangChain frameworks.
- Implemented OpenAI and prompt engineering for optimized model performance.
- Utilized Keras and TensorFlow for building and training deep learning models.
- Performed data cleaning and feature engineering for model preparation.
- Integrated ML/AI solutions into existing products, enhancing functionality.
- Monitored and optimized NLP and ML models for improved accuracy.
- Utilized Python and SQL for model development and data manipulation.

Project 1

AI-Driven RAG and LLM Customer Support Chatbot

Description

Created an GEN AI-driven chatbot using RAG and LLMs to automate customer service. Optimized text processing and retrieval for accurate responses. Enabled personalized user sessions with seamless interactions.

Roles & Responsibilities:

- Developed a RAG and LLM-powered chatbot for customer service automation.
- AI skills for chatbots include LangChain, GPT, PyPDFLoader, and ChromaDB for LLM workflows, text extraction, and vector search.
- Optimized performance through experimentation with document loaders and LLM
- models.
- Implemented a Retrieval QA chain for accurate responses from FAQs and PDFs.
- Managed session data with UUIDs and stored chat histories in MongoDB. Monitored performance metrics for continuous improvement used AWS.

Project 2

Speech-to-Text Conversion and Automated Summary Generation Using Whisper Model

Description

Developed a solution using the Whisper model for speech-to-text conversion to automatically generate minutes of meetings, summaries, key points, and action items for sales and marketing teams.

Roles & Responsibilities:

- Developed an automated meeting minutes system using Whisper for speech-to-text.
- Used the Gemini model for generating summaries, key points, and action items.
- The OpenAI Whisper model, Gemini, and Hugging Face enable automatic speech recognition and advanced
- language understanding through efficient prompting methods.
- Collected and prepared audio for model validation.
- Evaluated model accuracy and performance.
- Conducted testing to ensure output relevance.

Project 3

Multiclass Classification of Consumer Complaints by Department Using Deep Learning

Description

Developed a deep learning LSTM(NLP) model to classify consumer complaints and automate routing, reducing response time by 90%.

Roles & Responsibilities:

- Conducted data loading, cleaning, EDA, and handled imbalanced data.
- Tokenized and padded text data for LSTM model training in Python.
- Developed LSTM model with dropout, early stopping, and achieved high accuracy.
- Automated email system to route complaints, cutting response time by 90%.

Project 4

Address Quality Classification Using Machine Learning

Description

Classified customer addresses into Good, Medium, and Bad using Random Forest, applying data cleaning, feature engineering, and SMOTE for balanced accuracy.

Roles & Responsibilities:

- Cleaned and preprocessed address and pincode data.
- Integrated pincode data to enhance address validation and used python.
- Engineered features for address completeness assessment.
- Handled class imbalance using SMOTE for balanced classification.
- Implemented Random Forest ML model for classifying addresses.
- Optimized model performance and saved output for further analysis.

Technical Skills

- Generative AI: LLMs, Langchain,Transformer,Encoder-Decoder,RAG,CrewAi, Agents,Tools,LCEL,Langraph,Output Paraser,Memory,Hyperparameters of LLM, LLM chains, Document Loaders,Map,Reduce,Staff,Refine,Conversational Memory,TextSplitter,RAGA.
- GAN,Parameter Efficient Fine Tuning,LORA,qLORA,Gradio,HNSW,Retrivals,Prompt-Engineering,VectorDatabase,HuggingFace.
- GPT Models,OPENAI,Whisper,Llama2,GEMINI,Grok.
- Programming Languages: Python
- Libraries/Frameworks: TensorFlow, Keras, OpenCV,NLTK, Pandas, NumPy, Matplotlib, Seaborn,Transformer,Accelrate.
- Machine Learning & Deep Learning:BERT, LSTM, CNN, RNN, Linear & Logistic
- Regression, K-NN, K-Means, Decision Trees, Random Forests, Boosting
- Word Embeddings: Word2Vec, GloVe, FastText
- Natural Language Processing: Text Processing, Tokenization, Stemming,
- Lemmatization
- Data Processing: Cleaning, Preprocessing, Regex.
- Cloud Services: AWS
- Tools: Jupyter, Git, SQL, Google Sheets
- Specializations: Flask,Python, Machine Learning, NLP, Deep Learning, Time Series
- Analysis

EDUCATION

2017 Jabalpur Engineering College BE in Industrial Production