Faizan shaikh

Email: mdfaizan3299@gmail.com

Mob No.: 7219419796

About Me: -

Immediate Joiner| Ready to relocate | Data Scientist with 3+ Years of Experience in Python, Statistical Analysis, ML, Genai, LangChain, Prompt Engineering, LLMs, Fine tuning LLMs, Model building, NLP and Deep Learning, AWS and Data Visualization.

Career Objective: -

To achieve a challenging position in an organization where I can gain professional guidance and expertise to exhibit my technical skills for personal and professional growth.

Academic Qualifications: -

Education	Passing Year	University/Board	Specialization
B.Sc	2021	University of Pune	Mathematics
H.S.C	2018	Maharashtra State Board	Science
S.S.C	2016	Maharashtra State Board	-

Professional Experience: -

HAPPYTECH SOFTWARE SOLUTIONS PVT. LTD.

Period: 3 Year 5 months. (1st Sep 2021 – till date)

Designation: Associate sociate Data Scientist

Projects : -

1.Vehicle Insurance Fraud Detection
Domain: Insurance
Description

- The goal is to develop a Machine Learning model that can effectively detect instances of fraud in the context of vehicle insurance. Insurance fraud is a significant issue, costing companies and customers billions of dollars each year. By leveraging machine learning techniques, we aim to build a system that can identify fraudulent claims and help insurance companies take appropriate action.

Roles and Responsibilities:

- Understand, analyze, and interpret large datasets.

- Develop advanced programs to extract the data needed, prepare data for further analysis.

- Discover, design, and develop analytical methods to support novel approaches of data and information processing

- Perform analysis to assess the quality of the data, determine the meaning of the data, and provide data facts and insights

2.Handwritten to Text Data Extraction using Deep learning (OCR)
Domain: Banking/Healthcare Description

- Handwritten checks, forms, and financial documents can be converted to digital text for automated processing, data extraction, and record-keeping. To develop a system that can convert handwritten text into digital text using deep learning. It allows the recognition and extraction of characters from scanned images or handwritten documents, enabling the conversion of analog content into searchable and editable digital formats.

Roles and Responsibilities

- Work on Machine Learning/Deep Learning

- Write quality code while following best practices

- Analyze and preprocess the handwritten text data.

- Clean and enhance the images for better performance.

- Used libraries: Tensorflow, Keras, OpenCV, etc.

3.Extract Topic modeling from video using LLMs
Description

- This project uses language models to analyze and summarize text transcripts, identifying key topics and providing detailed descriptions. It processes the text to extract and expand on these topics using vector stores for contextual retrieval. The final result is a structured summary of topics, making it useful for organizing podcast content, summarizing meeting notes, and extracting information from documents.

Roles and Responsibilities:

- Gather and preprocess podcast transcripts or other relevant text data to ensure it's suitable for analysis.

- Assess the accuracy and relevance of the extracted topics, ensuring they align with the project goals and provide meaningful insights.

- Collaborate with the team to integrate topic modeling results into applications or platforms, optimizing processes for efficiency and cost-effectiveness.

Technical Skills: -

- Strong Python Programming skills using OOPS, Functions, Modularization & decorators.

- Sound Knowledge of Python's Data Analysis and Machine Learning Libraries.

- Implementation of regularization techniques like Lasso and Ridge in regression.

- Data Modelling: Using algorithms- Linear Regression, Logistic Regression,

- Support Vector Machine, Random Forest, XG Boost, Gradient Descent, PCA.

- Data mining algorithm experience in the families of predictive algorithms (Regression, KNN, Decision Trees)

- Data extraction, Data Manipulation

- Data Clustering: Clustering Algorithms

- Data Analytics on large data sets to solve business problems.

- Ability to process Text Processing using NLTK library.

- Thorough understanding of Probability and Statistics, Hypothesis Testing.
- Exposure to Large Language models (LLM): GPT (OPEN AI), BERT.
- Experience in data management tools –SQL databases.
- Source code management and Version Control system using Git and GitHub.
- Cloud platform for deployment- AWS.
- Data modelling.

Technologies

- Python/ML Packages: Scikit Learn, Pandas, NumPy, Reg Ex, Matplotlib, Plotly, Seaborn for Visualization.

- Genai: Exposure to LLMs, Finetuning LLMs like Llama 2, Llama 3, GPT 3.5 Turbo, GPT 4, Hugging Face, Vector Databases, ChromaDB, Pinecone, RAG.

- Experience in LangChain.

- Deep Learning: Neural Networks, ANN, CNN, RNN, Transfer Learning, Back Propagation, Activation & Loss Functions, Optimizers, Tensor flow2.x, Keras.

- NLP: Text Understanding, Representation, Classification & Text Clustering Skills, nltk, spacy, gensim,langdetect, googletrans, unidecode, BOW, TFIDF, word2vec, doc2vec, sent2vec, OpenCV, Image Detection, Image Classification, Transformer, LSTM.

- AWS: Elastic Compute 2, Notebook instance, AWS container, Simple Storage Services S3, AWS Lambda.

- Databases: MySQL, Command, Constrains, Clauses, CRUD operations, Subqueries, Joins.

- Math & Stats: Filter, Wrapper, Embedded Methods, P-Value, T-Test, Z-Test, ANNOVA test, Chi-Square Test, Info-Gain Test, Hypothesis Testing. Probability, Statistics, Linear Algebra, Gradient Descent.

Hobbies: -

- Solving Problem & Drawing

Personal Information: -

- **Date of Birth** : 12/05/2001
- **Gender** : Male
- **Marital Status** : Unmarried
- **Languages known** : English, Hindi, Marathi, Urdu
- **Current Address** : 40/63, Gujar Nagar, Maharashtra Colony, Sai Colony, Thergaon, Pimpri-Chinchwad, Maharashtra 411033.

Declaration: -

I hereby declare that the above mentioned information is correct up to my knowledge.

Place: - Pune.

Date: - **(FAIZAN SHAIKH)**