Shiban Siddiqui

✆ 7020787796 | shibansiddiqui@gmail.com Nagpur, India

SKILLS

• **DATA SCIENCE & MACHINE LEARNING:** NLP, OpenCV, YOLO, Scikit-learn, TensorFlow,
• **MACHINE LEARNING ALGORITHMS:** Linear Regression, Logistic Regression, KNN, SVM, Random Forest, DBSCAN
• **DATA ANALYSIS & VISUALIZATION:** Pandas, NumPy, Matplotlib, Seaborn
• **BIG DATA TECHNOLOGIES:** Apache Spark
• **CLOUD & DEVOPS:** AWS (Elastic Beanstalk), Docker, CI/CD
• **DATA ENGINEERING & PIPELINES:** Luigi Pipeline for ETL
• **DATABASE MANAGEMENT:** SQLite
• **FRONTEND DEVELOPMENT:** HTML, CSS, Bootstrap
• **BACKEND DEVELOPMENT:** Python, Flask, SQLite
• **VERSION CONTROL & OS:** Git, Linux (Terminal Commands, Bash/Shell)
• **NLP & NLU**: Large Language Models, Transformers, Retrieval-Augmented Generation Applications
• **COMPUTER VISION:** YOLO Object Detection Algorithm, OpenCV, Data Annotation (LabelImg)
• **NATURAL LANGUAGE PROCESSING (NLP):** NLTK for Text Processing and Analysis
• **GENERATIVE AI & LLM FINE-TUNING**: OpenAI, Hugging Face, LangChain
• **DATA SCIENCE TASKS:** Data Cleaning, Preprocessing, Wrangling, Data Exploration, Data Augmentation
• **AI AGENT**: Experience in building goal-driven AI agents
• **DEPLOYMENT & TRAINING:** Model deployment on AWS services like Elastic Beanstalk, utilizing Google Colab for training heavy datasets and implementing object detection models
• **GENERATIVE AI & LLMs:** Knowledge of Generative AI, latest LLM models like GPT, Falcon, and their applications
• **SOFT SKILLS:** Strong problem-solving abilities and excellent communication skills

EXPERIENCE / PROJECT

FINAL YEAR THESIS PROJECT
Legal Clause Prediction Using Deep Learning

- Led a team under Prof. Sadia Patka to develop an NLP-based system using the CUAD dataset for legal document analysis.
- Implemented tokenization, B&I tagging, and deep learning models to streamline contract review.

COMPUTER VISION & OBJECT DETECTION
Helmet and Number Plate Detection using YOLO

- Built a YOLOv8-based system to detect helmets and recognize number plates with the help of computer vision.

MEDICINAL PLANT ASSISTANT USING RAG & FALCON LLM

- Built a Retrieval-Augmented Generation (RAG) application using Falcon-RW-1B to answer.
- Used LangChain for document management and retrieval; generated accurate responses via Falcon using contextual prompts.

DATA SCIENCE & MACHINE LEARNING PROJECTS

- **Diabetes Predictor:** Developed a logistic regression model optimized with Grid Search.
- **Movie Recommender System:** Used NLP (stemming, vectorization) and cosine similarity for recommendations.
- **Fandango Data Analysis:** Conducted exploratory data analysis on movie ticket sales and reviews.
- **Customer Segmentation using Unsupervised Learning :** Implemented K-Means Clustering to group customers based on behavioral data, using the Elbow Method to determine the optimal number of clusters (K).

RASA CHATBOT ASSISTANT

- Developing an AI-powered chatbot using Rasa to providesmartly responds to all queries related to real-time weather updates using OpenWeatherMap API, with custom actions, entity extraction, and multi-turn conversation handling.

EDUCATION

BACHELOR OF ENGINEERING
Anjuman College of Engineering and Technology
[2019 – 2023]

- Field(s) of study: Computer Science and Engineering
- Final Grade: 70.3%

PUBLICATION

Published a paper on Legal Clause Prediction Using Deep Learning, focusing on NLP techniques for automating legal document review in a national conference.

COURSEWORK

- Data Science
- Machine Learning
- Python Full Stack Development
- Computer Vision

LANGUAGES

English • Hindi • Marathi • German • Urdu