

SHUBHAM MURTADAK

Data Scientist

☎ +91-9322191338

✉ shubhammurtadak022@gmail.com

in [linkedin](#)

🐙 [github](#)

ABOUT ME

Entry-level Data Scientist with proficiency in Python, Machine Learning ,Generative Ai, eager to make a meaningful impact through contribution. I'm passionate, energetic, and geeky individual whose desire to learn is endless.

EDUCATION

B.E. in Artificial Intelligence & Data Science

2021-2025

PES Modern College of Engineering

CGPA: **9.09**

Senior Secondary (XII), Science

2020

Shree Ganesh Jr College, HSC Board

Grades: **83.38%**

Secondary (X)

2018

New English School Korhale, India

Grades: **89%**

SKILLS

Primary Skills:

- **Programming Languages:** Python, R, C++, SQL
- **Python/ML/CV Packages:** Keras, OpenCV, Pandas, Scikit-Learn, Numpy, Matplotlib, Seaborn, Plotly
- **Knowledge of Machine Learning**
- **Deep Learning:** Neural Network, ANN, CNN, DNN, Transfer Learning, Back Propagation
- **TensorFlow**
- **NLP :** Text understanding, representation & classification techniques, Text clustering
- **Libraries:** BOW, TFIDF, word2vec, doc2vec, sent2vec, keyphrase extraction
- LSTM RNN, Transformers
- **Generative AI,** Large Language Models, LangChain Framework,LlamaIndex Framework
- **Agentic AI,** Autogen Framework
- Statistics
- DSA, Problem Solving

Secondary Skills:

- **WEB TECHNOLOGIES:** HTML, CSS, JavaScript, Django, Flask, Streamlit
- **DATABASES:** MySQL, MongoDB
- **Visualization Tools:** PowerBI, Microsoft Excel
- **Web Scraping:** Scrapy, BeautifulSoup, Selenium
- **Cloud Platform Services:** AWS
- Docker

EXPERIENCE

Data Science Intern
DataNnovite Solutions LLP

May 2024 - present
Pune, Maharashtra (Onsite)

- Worked on a diverse range of Data Science technologies including Machine Learning, Deep Learning, and Generative AI .
- Developed a comprehensive solution for analyzing and processing clients' transaction and campaign data, incorporating predictive models to optimize campaign performance.
- Designed and deployed a question-answering module for *dataRobo*, DataNnovite's AI-driven platform, allowing users to query and retrieve document content via a conversational interface.
- Leveraged Generative AI and Large Language Models to deliver contextually accurate responses, enhancing data accessibility and automating information retrieval from documents, contributing to significant efficiency improvements.
- Contributed to the creation of automated recommendation systems and predictive analytics tools to enhance customer targeting and campaign optimization.

PROJECTS

1. Automated Business Insights and Campaign Optimization

- Developed an LLM-powered solution using RAG architecture to analyze client transaction and campaign data, enabling data-driven decision-making through a custom dashboard.
- Segmented customers using customer lifetime value (CLV) and predicted next purchases using LSTM based on historical data.
- Targeted specific customer segments with personalized texts generated by an LLM for future campaigns.
- Containerized the entire application using Docker, simplifying deployment, ensuring consistency across environments, and enhancing scalability.
- **Tech Stack:** Python | Pandas | Numpy | LLM | LangChain | Milvus VectorDb | Flask | Docker

2. dataRobo

- Developed a Generative AI module for DataNnovite's dataRobo platform to enable question answering on unstructured documents, including KYC forms, Aadhaar cards, PAN cards, driving licenses, invoices, proof of concept documents, and bank statements.
- Implemented multi-document type querying capabilities, allowing users to ask questions across multiple document types such as invoices and purchase orders, enhancing the flexibility and utility of the platform.
- Managed multi-user chat history in MongoDB on a session-wise basis, enabling seamless tracking of user interactions and contextual responses for an enhanced, personalized user experience.
- Utilized user query history to maintain context across sessions, allowing for accurate and contextually relevant responses in follow-up questions.
- Achieved over 95% accuracy in extracting relevant information and providing context-aware responses, significantly improving data accessibility and operational efficiency.
- **Tech Stack:** Python | LangChain | RAG | ChromaDB | Gemini Embeddings | Gemini 1.5-pro Model | Flask | MongoDB

3. DocEase - Chat with PDF

- Developed an interactive web application that enables users to upload and engage with PDF, PowerPoint, and Word documents via a chat interface, leveraging generative AI models to deliver advanced document processing features.

- Implemented a React frontend to streamline document uploads, chat-based interactions, and content visualization, ensuring a responsive and intuitive user experience.
- Designed a Flask backend to handle secure document storage, process user queries, and integrate with generative AI services via LangChain, optimizing LLM utilization.
- Integrated AI capabilities such as question answering, summarization, and multilingual translation using the Gemini-1.5-pro model, enabling actionable insights directly from user documents.
- Established a MongoDB database for efficient storage of user data, chat history, and processed document content, facilitating easy retrieval and interaction analysis.
- Employed LlamaParser for accurate parsing of complex PDF structures, enhancing the app's versatility in handling diverse document layouts.
- **Tech Stack:** Python | React | Flask | MongoDB | LangChain | Gemini-1.5-pro | ChromaDB | LlamaParser

4. Anuwad - English to Marathi Translator [click here](#)

- Developed a neural machine translation model using TensorFlow, leveraging the Bahdanau Attention mechanism to translate sentences from English to Marathi.
- Implemented an encoder-decoder architecture, where the encoder processes input sequences and the decoder generates target sequences based on context.
- Integrated the Bahdanau Attention algorithm to enhance translation quality by allowing the decoder to focus on relevant parts of the input sequence during generation.
- Created a user-friendly Flask-based UI, enabling users to input English sentences and receive Marathi translations with ease.
- Demonstrated model effectiveness through various sample translations, showcasing high accuracy and contextual relevance in translations.
- **Tech Stack:** TensorFlow | Python | Flask | Bahdanau Attention

5. D-predicto [click here](#)

- Developed an end-to-end application predicting **diabetes**, **heart disease**, and **Parkinson's disease** using **XGBoost** and **RandomForest** models.
- Integrated an NLP-driven chatbot into the website to improve user interaction and access to information.
- Included a feature for users to book appointments with doctors if needed.
- **Tech Stack:** Python | SK Learn | ML | Numpy | Pandas | NLP | Flask | GitHub

6. RetainIQ [click here](#)

- Developed a system for predicting employee churn using **RandomForest** and **XGBoost** models to analyze workforce data.
- Implemented features for model training, batch prediction, and single employee prediction to determine retention likelihood.
- Built using Flask for web-based interaction
- **Tech Stack:** Python | SK Learn | ML | Numpy | Pandas | NLP | Flask | GitHub

7. GestureFlow [click here](#)

- Innovatively utilized Computer Vision and Gesture Recognition to redefine presentation interaction by creating a dynamic, hands-free platform for slide navigation and annotation.

- **Tech Stack:** Python | OpenCV | cvzone | OS Module | Numpy | Hand Tracking Module

CERTIFICATES

- Quatium Data Analytics
- Cognizant Artificial Intelligence
- Google Cloud Jam

ACHIEVEMENTS

- Winners Of Prostart 2023 -Team Matrix First Rank In College
- Academic topper throughout the graduation
- Earned Golden Badge in Python On HackerRank
- Solved over 300+ problems across diverse platforms