Sagar Patel

9913758770 • patel.sagar939@gmail.com • GitHub • LinkedIn • Portfolio

SUMMARY

Data Scientist with over 3 years of experience in data analysis, visualization, and process improvement. Proficient in Python, SQL, and Power BI, with a strong track record of leading end-toned data analysis projects and delivering actionable insights. Skilled in creating visually appealing dashboards and optimizing processes to achieve measurable results. Detail-oriented and results driven, with familiarity in machine learning techniques and tools like TensorFlow and Keras. Experienced in leveraging Large Language Models (LLM) and Generative AI technologies, including Gemini Pro, OpenAI, and Llama, to enhance data analysis and drive innovation.

WORK EXPERIENCE

Topia Life sciences ***Data Scientist Ahmadabad*** 01/2024 - PRESENT

Retrosynthesis with GNN

- Implemented Graph Neural Networks (GNN) for retrosynthesis, automating organic compound pathway prediction
- GNN Model: Developed and deployed a GNN model for molecular structure analysis.
- Data Processing: Preprocessed chemical data, engineered features, and optimized model performance.
- Integration: Collaborated with cross-functional teams to seamlessly integrate GNN into retrosynthetic analysis.
- Results: Accelerated synthetic route design, improving efficiency, and reducing time-to market for new compounds.

Predictive Research ***sr. Data Analyst Bangalore*** 02/2021 - 08/2023

LEAD SCORING ENGINE

- Objective: Predict customer propensity to buy products.
- Customer: Auto data Solutions/Hyundai Motor America.
- Tech Stack: Python, Scikit-Learn, MySQL Server, Flask, Docker.
- Role: Data collection, cleaning, feature selection, model building, continuous testing and deployment, report submission, model accuracy improvement.

PREDICT-OPS

- Objective: Develop log monitoring product.
- Tech Stack: Python, Django Rest, Elasticsearch, Prometheus, EC2, MongoDB, PostgreSQL.
- Role: Data ingestion from Elasticsearch, pattern recognition, micro-service design and development, Elasticsearch-alert configuration.

CUSTOMER LIFETIME VALUE

- Objective: Identify the best customers.
- Tech Stack: Python, Django Rest, Elasticsearch, Prometheus, EC2, MongoDB, PostgreSQL.
- Role: Data collection, cleaning, feature selection, model building, continuous testing and deployment in Flask, report submission, model accuracy improvement.

EDUCATION DETAILS

Government Engineering College *Surat CGPA: 6.98*
Diploma in Information Technology, [2014-2017]

Government Engineering College *Gandhinagar CGPA: 7.23*
Bachelor in Information Technology, [2017-2020]

Internship
Tops Technology: python
Applied AI: Artificial Intelligence

PROFESSIONAL SKILLS

- Programming Languages: Python
- Machine Learning: TensorFlow, Keras, Scikit-Learn, pytorch
- Data Analysis: NumPy, Pandas, Matplotlib, Seaborn
- Tools: Power BI, Excel
- Web Development: Django, Flask, Streamlit
- Databases: PostgreSQL, MySQL, MongoDB
- Generative AI: Gemini Pro, OpenAI, Llama, Langchain
- Vector Databases: Pinecone, Vector Databases

PROJECTS

Care Predict

- Developed ML and DL models to diagnose and predict diseases like Alzheimer's, brain tumours, breast cancer, COVID-19, diabetes, heart disease, and pneumonia.
- Utilized CNNs for image recognition and Random Forests for tabular data.
- Built and deployed a Flask application for diagnostic predictions.

Buy Till You Die - Customer Lifetime Value Prediction

- Predicted customer lifetime value (CLTV) using the Beta-Gamma-NBD (BG/NBD) approach.
- Gained valuable insights into customer behaviour.

Movie Recommender System with Streamlit

- Developed a movie recommender system using cosine similarity and Streamlit.
- Analysed movie metadata to calculate similarities and provide recommendations.
- Built a user-friendly interface with Streamlit for personalized movie recommendations.

Dell Image Generation with LLM APIs

- Used LLMs and APIs to generate unique images for Dell's technology.
- Experimented with OpenAI's DALL-E 2 APIs to create Dell-specific imagery.

AI-Powered Vector Database Search Engine

- Developed a search engine leveraging vector databases for enhanced data retrieval.
- Utilized Pinecone and other vector databases for efficient and accurate search capabilities.
- Integrated Lang chain for advanced search functionalities and AI-driven insights.

Objective

A dedicated and results-driven Data Scientist with over 3 years of experience in data analysis, visualization, and process improvement. Seeking a challenging position where I can leverage my expertise in Python, SQL, Power BI, and machine learning techniques to deliver actionable insights and drive innovation. Passionate about optimizing processes, developing scalable solutions, and contributing to the success of a dynamic team. Eager to apply my skills in data analysis and generative AI technologies to solve real-world problems and support organizational growth.