Vatsal Parsaniya

Data Scientist

(+91) 84604-22610
vatsalparsaniya@gmail.com
vatsalparsaniya.com
github.com/Vatsalparsaniya
linkedin.com/in/vatsal-parsaniya

Data Science professional with over 2.5 years of expertise in Information Retrieval, NLP, and microservice development. I have hands-on expertise in innovating new products using Data Science and Machine Learning. I have experience collaborating with cross-functional project teams, SMEs, and leading teams with multiple data scientists and engineers. I am passionate about multilingual search systems and uncovering valuable insights from complex data.

SKILLS

- **Data Strores :** Elasticsearch, Milvus, Solr, MongoDB, Redis, PostgreSQL
- **Development Tools :** Git, Curl, Jupyter Notebook, PyCharm, Postman
- **Backend Tools :** FastAPI, Airflow, Docker, Azure, Jenkins(CI/CD)
- **Observability :** Newrelic, Loggly, Pyinstrument
- **Statistics/Machine Learning / Deep Learning / NLP :**
  - **Frameworks for NLP :** NLTK, Spacy, PyTorch, Pandas, Scikit-Learn, Text Blob
  - **NLP Models Used :** BERT, RoBERTa, Elastic-ELSER, ALBERT, T5
  - **Model Deployment & Lifecycle :** NVIDIA Triton, MLflow
  - **ML Algorithms Implemented :** Linear Regression, Logistic Regression, XGBoost, KNN, KMeans, PCA, TSNE, TF-IDF, Word2Vec, Ensemble Algorithms, Topic Modeling
  - **Visualization :** Elastic-Kibana, Metabase, Matplotlib, Seaborn, Plotly
  - **Application Demo :** Gradio, Streamlit

PROFESSIONAL EXPERIENCE

Embibe **Sep 2021 — Present**
Data Scientist ***Bangalore***

- Currently working on optimizing multilingual hybrid search outcomes (lexical + semantic) through the implementation of microservices across a wide range of customer products and internal tools. This optimization includes comprehensive support for 11 Indic languages.
- Currently involved in Retrieval Augmented Generation (RAG) for search applications. This entails retrieving academic entities using ontologies and dynamically selecting prompts to generate contextually relevant responses from a generative model.
- Designed and implemented a scalable image(OCR) and text entity extraction algorithm along with a microservice. This system efficiently identifies academic entities by leveraging multiple ontology datasets, synonyms, spellchecking, and the solr analyzer for enhanced accuracy.
- Developed an intent and entity-based ranking module that optimizes the trade-off between latency and business logic demands, leading to an impressive 8% boost in Click-Through Rate (CTR).
- Hybrid search (lexical + semantic) :
  - Conducted benchmark evaluations on various vector databases, including Milvus, Qdrant, Elasticsearch, and Solr, with a primary focus on retrieval latency and vector index types.
  - Integrated a fine-tuned MiniLM model into the inference server and incorporated a vector database into the hybrid search pipeline, enabling semantic search capabilities.
  - Established a search feedback pipeline using the Gradio Interface for SME validation and the evaluation of various search algorithms.
  - Developed search capabilities in 11 Indic languages, including query understanding (QU) and query expansion (QE) modules, enabling users to seamlessly search in any language.
- Setup search utilization dashboard by consuming user event logs to monitor and measure search performance metrics.
- Implemented dockerized ingestion pipeline comprising an extensive collection of 400+ Airflow DAGs, significantly enhancing system observability.

Intellica.ai **Nov 2020 — Sep 2021**
Machine Learning Engineer (Intern + Full Time) ***Ahmedabad***

- Collaborated on the development of a real-time telephonic conversational AI system aimed at effectively streamlining the interview pre-screening process.
- Enhanced the speech-to-text pipeline for Indian English accents through the application of transfer learning on a deep-speech (STT) model.
- Designed a microservice that includes an evaluation system for the Montreal Cognitive Assessment (MoCA), integrating both computer vision-based drawing evaluation and context-based answer evaluation.

Education

- **Pandit Deendayal Energy University (PDEU), Gandhinager** 2017 - 2021
  - Bachelor of Engineering (B.E.) in **Information and Communication Technology** with CGPA: 9.2/10.

ACHIEVEMENTS

- 1st Runner up & Best Pitch Award in
- Kaggle 3X Expert
  - Classifies drugs based on their biological activity, Mechanisms of Action (MOA), 208/4373 (Silver Medal).
  - Multi-label classification, SIIM-ISIC Melanoma Classification : 127/3314 (Silver Medal)

COURSES

- Generative AI with Large Language Models, Coursera
- DeepLearning Specialization, Coursera
- Coursera Machine Learning, Coursera
- Time Series with Python (SKILL TRACK), Data Camp