Vatsal Parsaniya

Data Scientist

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

Data Scientist with over 3 years of experience in Data Science, Machine Learning, and Microservice development. I have hands-on experience of various search components. Additionally, I have experience working with various NLP technologies like Transformers for tasks such as semantic textual similarity, classification, entity detection, etc.

SKILLS

- **Data Stores :** 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, LLM
  - **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

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

- My role revolves around enhancing the user's search experience by optimizing multilingual search outcomes across a wide range of customer products and internal tools.
- **Multilingual Semantic Search**: Developed Semantic Similar entity based Search
  - Finetuned MiniLM (SBERT Sentence Transformer Model) for semantic similarity for our academic/Indian domain entities, which helped us improve search retrieval and relevancy for user queries by 8%.
  - Implemented results re-ranking utilizing the LambdaMART algorithm, which is a gradient boosting-based ranking algorithm. This significantly improved results ranking on our golden dataset compared to using embedding similarity for ranking.
  - Integrated embedding model into the NVIDIA Triton Inference Server and incorporated a vector database into the hybrid search pipeline.
  - Developed search capabilities in 11 Indic languages, and used the LABSE (Language Agnostic BERT) Model for semantic similarity.
  - Established a search feedback pipeline using the Gradio Interface for SME validation and for the evaluation of various search algorithms on golden dataset.
  - Worked on query understanding (QU) and query expansion (QE) modules, enabling users to search in mixed languages (English+Indic).
- **Entity Extraction (NER)**: For Search – Developed Solr based entity detection to facilitate search.
  - Implemented end-to-end Solr-based Entity Detection Service, using Solr's tagger request handler.
  - Used various natural language analyzers and filters like stemmer, stopwords, synonyms, lemmatizer, and possessive in our analyzer chain of Solr's tagger request handler.
  - The algorithm serves as a core service in various client products, tasked with highlighting academic entities and retrieving associated academic content.

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 Innovation and Incubation Centre
- 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