Abhishek Kumar

Data Scientist

abhishek1atulya@gmail.com +919625135515 New Delhi github.com/abhishek1atulya

linkedin.com/in/abhishek-kumar-760216189

EDUCATION

Integrated B.tech+M.tech (5 year) Electronics and communication engineering
Jawaharlal Nehru University
07/2018 – 07/2023 | New Delhi, India

SKILLS

Python
5 star on Hacker Rank

Data Structure and Algorithm
200+ Questions on LeetCode using python

Machine Learning Algorithms (10+ project done on Kaggle)
ScikitLearn, Regression, Classification, Clustering

Deep Learning
Pytorch, Tensorflow, openCV, ANN, CNN, RNN, LSTM, GRU, Transformers

Generative AI
HuggingFace Transformers, RAG, Langchain, vector databse, FAISS, CromaDB, weaviate, LlamaIndex, LLMs fine Tuning

Database Management System
SQL, NoSQL, MongoDB

Statistical Modeling and Inference
Scipy, Statsmodel, Statistics, sklearn metrics

Data gathering, Cleaning and Preprocessing
web scraping, Selenium, beautifulsoup, numpy, pandas, matplotlib, seaborn, regex, nltk, spacy

Frame work
FastAPI

Other Skills
AWS, git, MLOPS , Pyspark, Azure Databricks

SOFT SKILLS

Communication | Team-Work
adaptability | Confidence

PROFESSIONAL EXPERIENCE

Data Scientist
Weboconnect Technologies Private Ltd
01/2023 – present | New Delhi, India

- Developed a RESTful API using FastAPI and MySQL, implementing CRUD operations and ensuring optimized database queries for high performance.
- Developed an automated question-answer generation system using RAG, Hugging Face Transformers, LangChain, ChromaDB, and Llama2, significantly streamlining content creation workflows and reducing manual effort.
- Designed and deployed a vehicle image classification system with YOLO v8, cv2, and PyTorch, automating vehicle type identification.

PROJECTS

SmartCoach Recommender

- Built a **sentence-based recommendation system** to match users with coach profiles by understanding their problems.
- Utilized the **open-source model Gemini 1**, generated **synthetic data**, and created a **model prototype** for accuracy evaluation.
- Conducted **experiments with different models and embedding techniques** to optimize recommendations.
- Used **vLLM for efficient processing** and **ChromaDB for vector storage** to perform similarity search on user queries and coach profiles.
- Retrieved the **most relevant coach index** based on similarity scores to fetch profiles from the database.

Predictive Analytics (Credit-Risk Modelling)

- **Scoring Model Development**: Built a model to classify merchants for bank loans.
- **Data Analysis**: Analyzed credit scores, RTR reports, and historical data.
- **SQL Data Extraction**: Queried Q1 data (2019-2022) from Salesforce for analysis.
- **BI & Visualization**: Used Tableau to identify trends in merchant data.
- **EDA & Data Cleaning**: Handled missing values, outliers, and performed feature engineering.
- **Model Training**: Trained Logistic Regression and XGBoost with hyperparameter tuning.
- **Risk Assessment**: Used decile analysis to classify reliable and risky merchants.

Facial Expression Recognition (FER)

- **Facial Emotion Recognition**: Developed a deep learning-based facial emotion recognition system using OpenCV and PyTorch.
- **Dataset Management**: Created and structured custom datasets for efficient training.
- **Data Processing**: Applied image preprocessing, data loaders, and augmentation for better model performance.
- **Transfer Learning**: Used a pre-trained ResNet50 model with PyTorch to improve accuracy and reduce training time.
- **High Accuracy**: Achieved 90%+ accuracy on the Kaggle dataset with GPU training.

ACHIEVEMENTS

GATE Qualification

- Qualified GATE 2022 in Electronics and Communication Engineering (ECE), demonstrating a strong foundation in engineering principles and analytical skills.
- Successfully qualified GATE 2024 in Data Science and Artificial Intelligence, highlighting expertise and commitment to advancing knowledge in cutting-edge technologies and methodologies.