Vishal Sharma, Senior Data Scientist

Bengaluru, India, +91-8171292932, vishalsharma2204@gmail.com

PROFILE

Data Solution Expert with 7 years of experience delivering innovative Zero to One offerings and datasets for tech enterprises

Specialized in Custom Solution Design, Recommender Systems, Machine Learning, Natural Language Processing and Large Language Models with proven success in driving revenue growth, targeted marketing, customer retention and generating valuable insights

Recognized for enhancing data quality, developing features, embeddings, and scalable custom datasets using CRISP-DM methodology and leveraging expertise in Research, Statistics, Python, Pyspark, SQL, AI Agents, RAGs, Large Language Models, Cross-functional Collaboration, MLOps

---

EMPLOYMENT HISTORY

Aug 2020 — Jun 2024

Senior Data Scientist, Demandbase
Remote (Hyderabad)

- Owned a Fleet Product involving account identification(classification), prediction(regression), and multi-class classification, resulting in a dataset of 1.4 million accounts valued at over $2.1M, actively utilized by 9 clients for targeted marketing
  - Spearheaded end-to-end solution, from source identification, hypothesis formulation, data scraping, statistical testing, data cleaning, imbalance treatment, data mining, feature engineering, data visualization, modeling, deployment and monitoring. Also creating a streamlit app for stakeholder presentation and later automation of the processes.
  - Utilized Python, SQL, NLP, PyCaret, AWS Sagemaker, Streamlit, Flask, Docker
  - Quarterly refresh with ensembles techniques lead to increase of accounts identified from 436k to 1.4M and F1 accuracy from 0.6 to 0.85 via Recall increase.
- Led the full development cycle of Premium Intent(Next Technology Purchase + Intent), from use case breakdown, textual data processing, feature engineering, modeling to custom dataset creation to stream-lit deployment.
  - The project involved ensemble of Neural Collaborative Filtering, Association Rule Mining and NLP based Topic Modeling to categorize technology products, embedding accounts and products using firmo-graphics and descriptions, and utilizing intent scores. The model leveraged account–embedding and product–embedding via cosine similarity, outputting a purchase probability score for each account and product combination.
  - Popularity model to Market basket to Deep Learning made the NDCG increment from 0.54 to 0.88 for the top 15 product recommendation for the account
- Implemented LLM based tech product description generator using mistral and mxbai embeddings of tech pool keywords to expand historically mapped keywords and use prompt template including key features, benefits, and usage of the product
  - Langchain, Chromadb, Mistral, Transformers, SageMaker
- Created products for the techno graphic space that help in generating Ideal Customer Profile for an account for its custom use cases
  - IT Spend (Regression) for budget allocation, Product Renewal(Purchase Timeline), Trending Skills and Adoption to provide valuable insights on technology adoption and de-adoption trends
  - Executed comprehensive text analysis on job postings, project descriptions, and tech usage. Created scoring models to categorize technologies as emerging, stable, or declining, with quarterly updates to refine predictive accuracy

Aug 2019 — Aug 2020

Data Scientist, G2
Bengaluru

- Developed automated framework for mapping textual transaction feeds to products using text processing techniques, including fuzzy matching, cosine similarity matrices, and TF-IDF analysis.
- Conducted descriptive and inquisitive trend analysis to generate insights related to product overlaps, spending trends, contracts, and comparisons among products, overlapping products on dashboards
- Worked with contracts data to extract relevant information like the term of contract, entities involved and start date-end date

Jun 2017 — Aug 2019

Data Analyst, Virtusa
Chennai

- Designed and developed solution Impact Analysis Tool, identifies areas of system where a Change request will cause Impact using Hierarchial Clustering, Market Basket Analysis and Collaborative Filtering using NLTK. Tool being used internally by Virtusa in 36 Teams with total of 900 developers
- Applied in-depth debugging and code refactoring in Java for a code analyzer tool for ProC

| Jun 2013 — Jun 2017 | **B.Tech, Indian Institute of Technology** | Roorkee |

CGPA - 7/10

Course- Biotechnology

Major Project: Delivery of drugs for cancer via nanoparticles and A/B testing

---

SKILLS

Python	Recommender Systems
SQL	Retrieval Augmented Generation
Natural Language Processing	AI Agents
Large Language Models (LLM)	Docker
Machine Learning	Git
Deep Learning	Solution Design

TECHNICAL AND BUSINESS SKILLS

Technical Skills

- **Hypothesis Testing & A/B Testing**: Hypothesis formulation, T-Test, Z-Test, Mann-Whitney U, Chi-Sq test, ANOVA
- **Statistics & Visualization**: Distribution, Descriptive stats, EDA, outlier detection, correlation, covariance; word cloud, n-grams for text
- **Text Processing**: Stopwords removal, lemmatization, stemming, spacy NER
- **Feature Engineering**: Text vectorization (TF-IDF, BOW, Word2Vec, GloVe, FastText); Sentence embeddings (SBERT), BERT, Mxbai
- **Dimensionality Reduction & Scaling**: PCA, LDA
- **Clustering & Topic Modeling**: K-means, BERTopic, LDA
- **Classical ML Algorithms**: Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, Random Forest
- **Deep Learning & Sequence Models**: NCF, LSTM for sequential data; Transformers (BERT, RoBERTa, GPT) for contextual NLP tasks
- **Model Selection & Validation**: Cross-validation, hyperparameter tuning, grid search
- **Similarity Search**: FAISS, Chromadb
- **Model Performance Metrics**: MSE, RMSE, adjusted R2, Accuracy, precision, recall, F1-score, BLEU, ROUGE, perplexity
- **Web Frameworks**: Flask, streamlit
- **Cloud & Storage**: AWS S3, Athena, EC2, SageMaker, Azure AI Services
- **Deployment**: Docker, Git
- **Libraries**: pandas, langchain, numpy, matplotlib, seaborm, scikit-learn, statsmodels, scipy, nltk, spacy, gensim, transformers, sentence-transformers, textblob, tensorflow, pytorch, xgboost, lightgbm, faiss, mlflow, dtale, wordcloud, beautifulsoup4, requests, fake-useragent, boto3, gcfs, surprise

Business Skills

- **Communication**: Stakeholder management, cross-functional team collaboration
- **Problem-Solving**: Innovative solutions, data-driven decision making
- **Leadership**: Mentorship, project leadership, managing junior data scientists
- **Stakeholder Communication & Education**: Simplifying complex DS insights for non-technical audiences

---

CERTIFICATIONS

Microsoft Certified: Azure AI Engineer Associate

As a Microsoft Azure AI engineer, built, managed, and deployed AI solutions that leverage Azure AI.

LANGUAGES

| English | Hindi |