

VIVEK GUPTA

Lead Data Scientist with 8 years of experience | IIT Kanpur

@ gvivek2692@gmail.com

📞 9815619602

📍 Bangalore

EXPERIENCE

Lead Data Scientist

Cars24

📅 Nov'21 – Present

📍 Bangalore

- Developed **Dynamic Pricing Module** for selling pre-owned cars by training Catboost model on integrated dataset of web analytics, vehicle conditions, and competitor marketplace insights leading to **4% increase in gross margins**
- Developed in-house job allocation model using **Linear Sum Assignment** algorithm achieving **8% reduction** in travel distances for car inspection mechanics

Senior Associate, Data Science

Bain and Co

📅 Mar'20 – Oct'21

📍 Bangalore

- Developed demand **forecasting** model and price optimizer, using **LightGBM** and **Hyperopt**, achieving a **5% reduction in promotions cost** for a global retail giant working alongside strategy consultants and retail business experts
- Implemented **NLP** techniques like **Topic Modeling** and **Sentiment Analysis** to **analyze customer feedback**, identifying key complaint areas and actionable improvements for a Telecom company

Lead Data Scientist

IndMoney

📅 Aug'18 – Feb'20

📍 Gurgaon

- Developed Mutual funds **Recommendation and Ranking Engine** using **XGBoost** to predict and compare mutual funds returns and risk working alongside financial analysts
- Developed an **expense tracking system** from transaction SMS data using **NER** and **supervised text classification**, accurately categorizing expenses into various categories, also built a data labeling pipeline for enhancing training data

Data Scientist

Cogneau Systems

📅 Oct'17 – Jul'18

📍 Gurgaon

- Reduced operations cost by **30%** for a logistics client in Malaysia by developing an optimized DC network using Density Based **Clustering** Algorithm **DBSCAN**

Data Scientist

Simpli5d Technologies

📅 Nov'16 – Sep'17

📍 Gurgaon

- Developed a **Bot Detection Model** using **Random Forest** in **Spark** for more than **10 million** user sessions per day on more than **120 publishers**

GEN-AI/LLM PROJECTS

Candidate Recommendation System

PyjamaHR

📅 Jan'23 – Dec'23

📍 Bangalore

- Developed a matching system for resumes and job descriptions, utilizing **semantic similarity** on embeddings of JDs and resume summaries generated using **LLMs** finetuned by **LoRA** technique

COURSES



Building Advanced RAG

DeepLearning.AI



Finetuning Large Language Models

DeepLearning.AI



Deep Learning Specialization

Andrew NG, Coursera

TECHNICAL SKILLS

- **ML Techniques:** XGBoost, LightGBM, Random Forest, Catboost, DBSCAN, Hyperopt
- **ML Libraries:** Pytorch, Tensorflow, Keras, Scikit Learn, Pandas, Numpy, Matplotlib
- **Gen AI and LLMs:** PeFT, LoRA, RAG, LangChain, LlamaIndex, BERT, Flan-T5
- **MLOps:** KubeFlow, Docker, GCP, AWS, Azure
- **Languages:** Python, SQL, PySpark

EDUCATION

B.Tech. in Bio Science

IIT Kanpur

📅 Jul'10 – May'14

- CGPA - 7.0

EXPERIENCE

Trainee Decision Scientist

Mu Sigma

📅 Aug'14 – Feb'15

📍 Bangalore

- Analysed search advertising business KPIs for Asia Pacific for Bing Ads