Prakul Sharma

Vancouver, BC (Willing to relocate)

+1 (236) 863-4883 • prakuls13@gmail.com • linkedin/sharmaprakul • github/prakulsharma • prakulsharma.github.io

---

Profile

Experienced Data Scientist with **three years** in **advanced data analytics** and **AI-driven** solutions. Applied machine learning to transform complex data into strategic insights and developed scalable software solutions to **drive data-centric initiatives**. Seeking a challenging role to improve **informed decision-making** and promote **collaborative success**.

Work Experience

Graduate Research Assistant — Vancouver, Canada
Department of Sociology, UBC — December 2023 – Present
- Analyzed semantic evolution of the word **"community"** to identify **5+ significant shifts** in its meaning over 10 decades.
- Employed topic modeling using **BERTopic**, coreference resolution, and NER using **spaCy** on over **50,000 news articles**.

Data Scientist — Bengaluru, India
SymphonyAI Industrial — October 2021 – July 2023
- Spearheaded **anomaly detection**, **time-series analysis**, and **predictive modeling** projects for industrial setups.
- Developed semi-supervised models for predicting motor equipment failure with **90%+** accuracy, yielding **$200k+** profit.
- Implemented EHS incident forecasting models using **PyTorch**, achieving **80%+** accuracy using leading safety indicators.
- Created a no-code **analytics interface** using **Dash**, saving over **1200 man-hours** annually.
- Formulated algorithms to extract **100+** features from high-frequency vibration data for anomaly detection modeling.

DevOps/Software Engineer — Bengaluru, India
Infosys — November 2020 – October 2021
- Built web application to automate VM life-cycles, saving **500+ man-hours** yearly in manual resource management.
- Automated user-specific software installation via UNIX scripts during VM creation, reducing setup time by **60%**.
- Built scalable CI/CD pipelines using Azure DevOps Toolkit, ensuring efficient workflows in a cross-functional environment.
- **Supervised and mentored 2 junior developers.**

Projects & Research

BCI: Quantitative Topic Modeling for Enhanced Stock Analysis — May 2024 – June 2024
- Automated parsing, transformation of 10-K filings into JSON files using regex, reducing data extraction time by **half**.
- Implemented topic data models using **LDA**, **BERTopic** and **TopicGPT**, achieving a **30%** higher coherence score.
- Integrated Yahoo Finance API to add stock performance data, improving topic identification accuracy by **10%**.

TextDetox — March 2024 – April 2024
- Developed multilingual detox models for **9 languages** to enforce content moderation, using English as base.
- Built a translation-detoxify pipeline with **NLLB-200** and used **BART-base, T5-base** as evaluation baselines.
- Fine-tuned **Flan-T5** and **TinyLlama** on the Paradetox dataset, achieving **86%+** detoxification in texts.

ARCOT: Abstraction and Reasoning Corpus — February 2024 – March 2024
- Annotated **400** tasks for logical reasoning, enhancing LLM problem-solving with **chain-of-thought** prompting.
- Created Streamlit web app to visualize and search annotated tasks using cosine similarity.

Education

Master of Data Science (Computational Linguistics) — Vancouver, Canada
The University of British Columbia — September 2023 – June 2024
- **GPA: 9.24/10**
- **Coursework:** Supervised & Unsupervised Learning, Deep Learning, Natural Language Processing, Data Structures & Algorithms, Databases & Data Retrieval, Data Visualization, Descriptive & Inferential Statistics, Feature Engineering

Bachelor of Technology (Information Technology) — Delhi, India
Guru Gobind Singh Indraprastha University — August 2016 – July 2020
- **GPA: 8.91/10**

Skills

Data Science	EDA, Statistical analysis, Power BI, Tableau, Data Analysis, Data Processing, Data Visualization
Machine Learning	Classification, Regression, Clustering, Reinforcement Learning, Bagging, Boosting, Ensemble Models
Deep Learning	Neural Networks, Transformers, Autoencoders, CNNs, RNNs, LSTM, GANs, GNNs, OpenCV, CUDA
NLP	GPT, BERT, spaCy, NLTK, Word Embeddings, OpenAI API, LangChain, Prompt Engineering, RAG
Development	Software Life Cycle, Docker, CI/CD, FastAPI, Jenkins, Kubernetes, Agile, RESTful APIs, Git, Testing
Database	SQL (PostgreSQL, MySQL), NoSQL (MongoDB), Cloud Computing, Azure, AWS, Google Cloud (GCP)
Programming	Python (Pandas, NumPy, scikit-learn, PyTorch, TensorFlow), R (Tidyverse), Java, C++, JavaScript
Essential traits	Collaboration, Adaptability, Problem-solving skills, Communication skills, Mentorship, Eye for detail