

Prakul Sharma

Vancouver, BC, Canada

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

Profile

Dedicated Data Scientist with three years of experience in **machine learning**, **anomaly detection**, and **predictive modeling**. Skilled in solving complex problems and transforming data insights into actionable strategies by developing **scalable, end-to-end AI solutions** to drive **revenue growth**. Seeking a challenging role to **advance innovation** and **maximize data utilization** through collaborative efforts.

Work Experience

NLP Research Assistant, University of British Columbia

Dec 2023 – Present | Canada

- Analyzed 50,000+ news articles to identify 5+ shifts in meaning of “**community**” over 10 decades, attaining **85%** semantic accuracy.
- Enhanced entity extraction accuracy by **25%** by employing **BERT based topic, coreference** and **NER** models for contextual analysis.
- Formulated **t-SNE clustering** and visualizations in **Dash**, enhancing user comprehension by decreasing data exploration time by **30%**.

Data Scientist, SymphonyAI Industrial

Oct 2021 – Jul 2023 | India

- Developed semi-supervised **LSTM** models for predicting motor equipment failures with **90%+** accuracy, yielding **\$200k** cost savings.
- Engineered EHS incident forecasting models using **XGBoost**, achieving an **F1 score of 0.72** by utilizing key industry safety indicators.
- Architected no-code **analytics interface** in Dash for data analysis, model building and visualization, saving **1200+** man-hours annually.
- Streamlined **MLOps** pipelines using **Databricks, MLflow, and Docker**, lowering deployment cycles by **40%** by automating workflows.
- **Led a team** of 4 to design product features and implement solutions, contributing to **15%** of annual revenue; earned **client recognition**.
- Coded **XML parser** & algorithms for advanced **feature engineering** of vibration data, improving anomaly detection accuracy by **10%**.

DevOps Engineer, Infosys

Nov 2020 – Oct 2021 | India

- Built an intra-organisation app with **Angular, Node.js, and MongoDB**, deployed on **Azure Kubernetes Service**, for **2,000+** employees.
- Automated **VM lifecycle** management & software installation employing **Azure DevOps, & UNIX** scripts, cutting setup time by **60%**.

Projects

BCI: Quantitative Topic Modeling, Capstone Project

May 2024 – Jun 2024

- Implemented **LDA** as baseline for SEC 10-K filings, accomplishing a peak **coherence score (c_v) of 0.65** for reliable topic generation.
- Created **BERTopic** using BERT embedding to capture nuanced, context-rich topics, increasing coherence score by **12%** against LDA.
- Crafted a **preprocessing** pipeline with **spaCy** for segmentation, tokenization, & stop-word removal, reducing preparation time by **50%**.

TextDetox, Multilingual Text Detoxification

Mar 2024 – Apr 2024

- Constructed detox models for **9 languages** with **transformers**, reaching **86%** content moderation accuracy while preserving meaning.
- Fine-tuned **Flan-T5** on the **ParaDetox** dataset, retaining original meaning while advancing semantic similarity by **15%** over **BART**.
- Augmented style transfer accuracy by **28%** as opposed to baseline, by **fine-tuning TinyLlama** using **LoRA** for efficient detoxification.

ARCOT, Abstraction and Reasoning Corpus

Feb 2024 – Mar 2024

- Annotated reasoning tasks for **LLMs** problem-solving using **chain-of-thought** prompting, strengthening reasoning accuracy by **20%**.
- Designed **Streamlit** application to visualize & search ARC tasks using **cosine similarity** for phrases, with **80%** reduced retrieval time.

Education

Master of Data Science (Computational Linguistics), University of British Columbia

Sep 2023 – Jun 2024 | Canada

- GPA **9.24/10**. **Coursework:** Statistics, Supervised Learning, Unsupervised Learning, Natural Language Generation, Data Visualization

Bachelor of Technology (Information Technology), Guru Gobind Singh Indraprastha University

Aug 2016 – Jul 2020 | India

- GPA **8.91/10**. **Coursework:** Data Structures, Algorithms, Database Management Systems, Cloud Computing, Software Engineering

Skills

ML	Classification, Regression, Clustering, Reinforcement Learning, Recommender Engine, Boosting, Bagging, SHAP
Data Science	Time series, Statistical & Stochastic Modeling, Hypothesis Testing, Data Distributions, SAS, SPSS, Excel, Selenium
Programming	Python (Pandas, NumPy, scikit-learn, Keras, OOP, PyTest), R, C++, JavaScript, Linux, Shell Scripting (Bash)
Deep Learning	Neural Networks, TensorFlow, Transformers, PyTorch, Autoencoders, CNNs, RNNs, GANs, GNNs, GBM, OpenCV
NLP	Word2Vec, GPT, RoBERTa, NLTK, LangChain, HuggingFace, RAG, Gensim, Large Language Models, GenAI
Development	RESTful APIs, A/B Testing, Agile, OpenAPI, Django, Flask, Jupyter, FastAPI, JIRA, CUDA, Seaborn, Matplotlib
Database	SQL, PostgreSQL, MySQL, BigQuery, Apache Spark, Snowflake, Hadoop, ETL/ELT, Kafka, Kubeflow, PySpark
Tools	Jenkins, CI/CD, Git, AWS Sagemaker, GCP, Tableau, PowerBI, Looker, Airflow, Grafana, Kibana, Plotly, OpenAI
Essential Traits	Communication skills, Collaborative, Adaptable, Critical thinking, Problem-solving, Mentorship, Leadership