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 **applied machine learning**, **statistics** and **data visualization**. Skilled in solving complex problems and translating data into actionable strategies by developing **scalable** and **end-to-end solutions** to drive revenue growth. Seeking a challenging role to advance innovation and optimize data utilization through collaborative efforts.

Work Experience

NLP Research Assistant, *Department of Sociology, UBC*   *December 2023 – Present, Canada*

- Analyzed the semantic and sentiment evolution of **"Community"** across 50,000+ news articles, identifying key shifts over **10 decades**.
- Used **AllenNLP** for coreference resolution, **BERT** for contextual analysis, and NER to extract key entities, enhancing semantic accuracy.
- Designed interactive **t-SNE clustering** and timeline visualizations in **Dash**, providing comprehensive view of semantic shifts over time.

Data Scientist, *British Columbia Investment Management Corporation (Co-op)*   *May 2024 – June 2024, Canada*

- Implemented **LDA**, **BERTopic**, **TopicGPT** models on 1M+ **SEC 10-K filings**, increasing topic coherence scores by **30%** over baseline.
- Engineered **NLP pipeline** for sentence segmentation, tokenization, lemmatization, and stop word removal, reducing prep time by **half**.
- Developed **20+ visualizations** using **Dash** and an evaluation framework using **c_v**, **c_npmi** metrics, improving overall interpretability.

Data Scientist, *SymphonyAI Industrial*   *October 2021 – July 2023, India*

- Spearheaded **time-series analysis**, **anomaly detection**, and **predictive modeling** for industrial setups, resulting in **$200k+** cost savings.
- Developed semi-supervised **LSTM**, **XGBoost**, and **Random Forest** models, significantly improving motor-failure prediction accuracy.
- 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, automating workflows and reducing deployment cycles by **40%**.

DevOps Engineer, *Infosys*   *November 2020 – October 2021, India*

- Developed automated **VM** lifecycle management using **Azure DevOps**, **Jenkins**, and **UNIX** scripts, saving **500+** man-hours annually.
- Built an intra-organisation app with **Angular**, **Node.js**, and **MongoDB**, deployed on Azure Kubernetes Service, for **2,000+** employees.

Projects & Research

TextDetox, *Multilingual Text Detoxification*   *March 2024 – April 2024*

- Developed **multilingual** detox models for 9 languages using **transformers**, moderating social media content while preserving meaning.
- Built translation-detox pipeline with **NLLB-200** and **BART-base** & **T5-base** as baselines for content neutrality and semantic accuracy.
- Fine-tuned **Flan-T5** & **TinyLlama** on ParaDetox, achieving **86%** detoxification with high style transfer accuracy, content preservation.

Smart Storyteller, *Image Captioning and Story Generation*   *January 2020 – May 2020*

- Built image captioning model using **InceptionV3** pretrained on **ImageNet**, with **95%** accuracy in creating captions from input images.
- Fine-tuned **GPT-2** on **Gutenberg** dataset for story generation, enhancing narrative coherence by **20%** based on BLEU score evaluations.
- Optimized self-attention-based captioning pipeline for real-time processing on Google **Colab**, reducing inference time by **30%**.

Education

Master of Data Science (Computational Linguistics), *University of British Columbia*   *September 2023 – June 2024, Canada*

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

Bachelor of Technology (Information Technology), *Guru Gobind Singh Indraprastha University*   *August 2016 – July 2020, India*

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

Skills & Frameworks

Data Science	Statistical & Stochastic Modeling, Hypothesis Testing, Data Distributions, Regression, SAS, SPSS, Excel
Programming	Python (Pandas, NumPy, scikit-learn, PyTorch, TensorFlow, Keras, OOP), R, PySpark, C++, JavaScript, Linux
Machine Learning	Classification, Regression, Clustering, Forecasting, Reinforcement Learning, Recommender Engine, Boosting
Deep Learning	Neural Networks, Transformers, Autoencoders, CNNs, RNNs, LSTM, GANs, GNNs, XGBoost, GBM, OpenCV
NLP	GPT, RoBERTa, BERT, spaCy, NLTK, SHAP, LangChain, HuggingFace, RAG, Gensim, LLMs, GenAI, OpenAI
Development	RESTful APIs, A/B Testing, Agile, OpenAPI, Django, Flask, Jupyter, FastAPI, JIRA, CUDA, Seaborn, Matplotlib
Database	SQL, PostgreSQL, MySQL, MongoDB, BigQuery, Spark, Databricks, Snowflake, Hadoop, ETL/ELT, Kafka
Tools	Docker, Jenkins, Kubernetes, CI/CD, MLFlow, Git, Azure, AWS, GCP, Tableau, PowerBI, Looker, Airflow
Essential Traits	Communication skills, Collaborative, Adaptable, Critical thinking, Problem-solving, Proactive, Mentorship