

Gaurav Mudbhatkal

gmudbhat@gmail.com | [Linkedin](#) | [Website](#)

TECHNICAL SKILLS

Languages and Tools: Python, SQL, Cypher, Git, Docker, Kubernetes, ArgoCD

Data Science and ML: PyTorch, Scikit-Learn, TensorFlow, OpenAI, LangChain, Metaflow, MLFlow, Airflow

Big Data and Cloud: PostgreSQL, Neo4j, Elastic, AWS, Spark, Azure

Project Management: JIRA, Trello, Confluence, Notion, ServiceNow

EXPERIENCE

Data Scientist

June 2022 – Present

Limbik

Kitchener, ON

- Drove 500% increase in Limbik's market valuation by developing an AI system combining two predictive models: 'Believability' (message credibility) and 'Virality' (engagement forecast), enabling precise PFI evaluation
- Boosted model performance by 20% by implementing a two-tower embedding model, leveraging SentenceTransformers for content features and PyTorch for user representation
- Architected a scalable, hierarchical PostgreSQL database to increase data retrieval efficiency by 30%, handled ingestion of 200K+ monthly survey responses through a Python-based data pipeline
- Streamlined ML service deployment process using GitHub Actions and ArgoCD, reducing manual workload by 50% for weekly releases
- Enabled data science experiment tracking and reproducibility, orchestrating model training and monitoring workloads by leveraging Metaflow, Kubernetes and Jupyter
- Spearheaded API development of generative AI-based suggestion feature using Langchain and FastAPI
- Enabled NLU-based data retrieval and analysis for SQL database using Google Gemini and LangChain
- Developed internal data exploration tool enabling vector search, embedding projection, and survey dashboards

Machine Learning Intern

Sep 2021 – Dec 2021

Thomson Reuters - Legal Tracker

Toronto, ON

- Tuned and evaluated document classification models in AzureML after conducting surveys across departments
- Implemented model deployment pipelines, integrating machine learning services with Azure DevOps

Data Engineer Co-op

Jan 2021 – Apr 2021

Manulife - Canada Data Office

Waterloo, ON

- Streamlined ETL workflows in Azure Data Lake with Spark, enhancing data warehousing capabilities
- Coordinated the migration of 30 repositories to GitHub with ServiceNow, JIRA, and Confluence

Data Engineer Co-op

Jul 2020 – Aug 2020

KPMG - Lighthouse

Waterloo, ON

- Calculated upwards of \$2,000,000 in eligible subsidies by wrangling client payroll data using workflows in Alteryx
- Reduced manual reporting time by 70% by automating report creation using VBA in MS Excel

Data Scientist Co-op

Sep 2019 – Dec 2019

RBC - Anti-Money Laundering

Toronto, ON

- Engineered NLP pipeline for 10,000+ documents with Doc2Vec embeddings, entity recognition, and topic modeling
- Achieved 96% accuracy and 82% precision through a neural network-based multi-label text classifier with Keras

PROJECTS

BESST (Battery Energy Storage System Toolkit) | *Flask, Gurobi, Plotly, Heroku*

Apr 2022

- Web-based decision support system to optimize residential power usage, leveraging Battery Energy Storage
- Developed a mathematical optimization model using Gurobi to minimize energy costs and GHG emissions
- Visualized cost savings and reduction in carbon footprint based on user inputs dynamically with Plotly

Enterprise Co-op Student | *Research, Entrepreneurship*

May 2020

- Conducted and pitched research on the storage and analysis of digital evidence data by Law Enforcement for a chance to win \$15,000 in funding through the Problem Pitch Competition

EDUCATION

University of Waterloo

Aug. 2017 – April 2022

Honors, Bachelor of Applied Sciences in Management Engineering, Computing Option

Waterloo, ON