VAIBHAV VAGHELA

249-989-0761 | vaibhavvaghela2010@gmail.com | [LinkedInProfile](#) | [GitHub](#)

Winnipeg, MB

PROFILE

- AI & Machine Learning Engineer with hands-on experience designing ML pipelines, building backend Python automation workflows, and deploying AI models in production-like environments.
- Built end-to-end ML systems for the *Bruce County Smart Beach* project, including image preprocessing, YOLOv8/Faster R-CNN/Detectron2 pipelines, real-time crowd-detection, and automated CSV + heatmap reporting.
- Developed NLP and transformer-based models (Longformer, XGBoost, TF-IDF) for long-document classification and sentiment analysis, integrating data cleaning, feature engineering, and hyperparameter tuning to achieve 85% accuracy and reduce loss by 17% compared to RoBERTa.
- Strong foundation in Python-based backend development, ML model training, and workflow automation using Airflow, LangChain, and cloud services (GCP, Azure, AWS).
- Hold a Bachelor's Degree in Physics (India) and currently pursuing a Post-Degree Diploma in Artificial Intelligence at UWinnipeg (PACE), with coursework in ML, Deep Learning, Big Data Platforms, and Applied Statistics.

TECHNICAL SKILLS

- **Programming & Backend:** Python (Pandas, NumPy, Scikit-Learn, PyTorch), SQL (advanced querying & optimization), Bash, REST APIs, backend automation workflows.
- **Data Engineering & Pipelines:** ETL/ELT (batch + streaming), data ingestion & transformation pipelines, data modeling, warehousing (Snowflake, BigQuery), Airflow workflow orchestration, automated data-quality checks, large-scale dataset processing.
- **Cloud & Big Data:** Google Cloud (BigQuery, Vertex AI, Looker Studio), Azure (Azure ML Studio, Data Factory, Databricks, Synapse), AWS (ML Foundations, Cloud Foundations), PySpark..
- **Machine Learning & AI:** LLMs, NLP, RAG pipelines (LangChain, embeddings), Longformer, transformer-based models, classical ML (Random Forest, XGBoost), intelligent automation systems, document classification & processing.
- **MLOps & Deployment:** CI/CD pipelines, model packaging & deployment, Docker, Kubernetes (container orchestration), Git/GitHub, experiment tracking.
- **Computer Vision:** YOLOv8, Detectron2, OpenCV for detection, tracking, and real-time video analytics.
- **Data Visualization:** Power BI, Looker Studio, Matplotlib, Seaborn.

EDUCATION

Post-Degree Diploma in Artificial Intelligence   Jan 2025 - Dec 2025

The University of Winnipeg Professional, Applied and Continuing Education   *Winnipeg, MB*

- Designed and optimized databases using SQL and Python to acquire, clean, and manipulate large datasets, improving efficiency of analytics workflows.
- Developed and tuned machine learning models including regression, classification, and clustering, achieving higher accuracy through feature engineering and hyperparameter optimization.
- Explored advanced AI techniques such as deep learning, reinforcement learning, and generative AI, applying TensorFlow and Keras to prototype solutions for real-world problems.

Post-Degree Diploma in Artificial Intelligence - Architecture, Design, and Implementation   May 2024 - Dec 2024

Georgian College   *Barrie, ON*

Bachelor of Science in Physics   Jun 2018 - Dec 2021

Sardar Patel University   *Gujarat, India*

ACADEMIC PROJECTS

Longformer – Transformer-based Text Classification
- Engineered a long-document NLP classification pipeline to address Transformer limitations on long input sequences, improving contextual understanding for multi-page documents.
- Fine-tuned the Longformer architecture using Hugging Face Transformers; performed hyperparameter optimization (batch size, LR, epochs) and configured global attention for efficient long-range dependency handling.
- Reduced evaluation loss from $1.18 \rightarrow 0.98$, outperforming RoBERTa baselines by 17%, while implementing unit + integration tests to ensure pipeline reliability and reproducibility.

IMDB Sentiment Analysis – Binary Classification
- Developed an end-to-end sentiment analysis system on a dataset of 25,000 IMDB reviews, creating a balanced 6,000-sample training set for efficient modeling.
- Preprocessed text using tokenization, stopword removal, and TF-IDF; trained and compared Random Forest and XGBoost classifiers using GridSearchCV for hyperparameter tuning
- Achieved 85% accuracy and 0.853 F1-score, demonstrating XGBoost's effectiveness for text analytics while performing end-to-end pipeline validation for production-like stability..

RELEVANT EXPERIENCE

Machine Learning Developer (Volunteer)   **Jun 2024 – Dec 2024**
Bruce County Smart Beach   Kincardine, ON
- Collaborated with a team of 4 to develop a computer-vision pipeline predicting real-time beach crowd density using YOLOv8, Faster R-CNN, and Detectron2 models
- Built data-processing workflows including image preprocessing (cropping, augmentation) and region-focused detection pipelines to improve model precision.
- Automated CSV exports and heatmap generation to support operational decision-making for public-safety teams..
- Tested and validated model performance under varying environmental conditions to ensure stability and reliability.
- Presented technical updates and deliverables during biweekly meetings, aligning model outputs with client safety requirements and feedback.

ADDITIONAL TRAINING

- *Snowflake Hands-On Essentials: Data Warehouse* – Snowflake | Jun 2025
- *AWS Academy Graduate – Machine Learning Foundations* – Amazon Web Services (AWS) | Dec 2024.
- *AWS Academy Graduate – Cloud Foundations* – Amazon Web Services (AWS) | Nov 2024.