**Shubham Patel**

[**shubhu.p2000@gmail.com**](mailto:shubhu.p2000@gmail.com) **| (647) 717-4553 | Scarborough ON**

[**LinkedIn**](https://www.linkedin.com/in/shubham1026/) **|** [**GitHub**](https://github.com/shubhu1026) **|** [**Portfolio**](https://shubhu1026.github.io/)

**SUMMARY**

Data Scientist with expertise in Python, PyTorch, and TensorFlow, and experience in deploying real-time analytics solutions and optimizing models for performance and interpretability. Proven ability to work in fast-paced environments and collaborate cross-functionally.

**SKILLS**

* Programming Languages: Python, PyTorch, TensorFlow
* Machine Learning: Supervised & Unsupervised Learning, Model Optimization, Time Series, NLP, Transfer Learning
* Data Engineering: PySpark, Apache Spark, Real-Time Stream Processing, ETL Pipelines, GCP
* Mathematics & Statistics: Hypothesis Testing, Correlation Analysis, Regression, Probability Models
* Tools & Platforms: Git, Docker, Jupyter, Google Cloud Platform, MLflow, Kibana, Elasticsearch
* Soft Skills: Cross-functional collaboration, Technical Communication, Initiative, Continuous Learning

**EDUCATION**

|  |  |
| --- | --- |
| **Humber College, Toronto, ON** | September 2024 – April 2025 |
| Graduate Certificate in Artificial Intelligence with Machine Learning |  |

* Gained hands-on experience in supervised/unsupervised learning, deep learning, and reinforcement learning.
* Built and optimized models using scikit-learn, TensorFlow, Keras, and PyTorch.
* Worked on real-world problems in NLP, computer vision, and predictive analytics.
* Designed scalable ML pipelines, including data preprocessing, feature engineering, and hyperparameter tuning.
* Explored big data and distributed ML using Apache Spark, Google Cloud Dataproc, and Spark MLlib.
* Learned ML deployment techniques with Flask APIs, Docker, and basic Kubernetes for containerized inference.

|  |  |
| --- | --- |
| **Centennial College, Scarborough, ON** | September 2023 – April 2024 |
| Graduate Certificate in Mobile Application Development |  |

* Built native and cross-platform apps using Kotlin, Swift, JavaScript, and Flutter, applying MVVM and clean architecture principle.
* Completed capstone project (SwiftTrend) with full-stack features: product search, PayPal integration, admin CRM, and deployment via Android Studio & Xcode.

|  |  |
| --- | --- |
| **Maharashtra Institute of Technology ADT, Pune, India** | July 2018 – July 2022 |
| Bachelor of Technology in Computer Science & Engineering |  |

* Gained strong foundation in C++, Python, OOP, DBMS, assembly, and basic machine learning using scikit-learn.
* Developed Leafio (major project): a CNN-based plant disease detection app and published a research paper comparing VGG16, Xception, and MobileNet architectures.

**PROFESSIONAL EXPERIENCE**

|  |  |
| --- | --- |
| **Machine Learning Engineer** | January 2025 – April 2025 |
| Kevares (Capstone project via Humber College AI Program) |  |

* Developed a real-time trash detection system using a YOLO Nano-based model, leveraging transfer learning on the TACO dataset and a custom robot-collected dataset.
* Led the data annotation process to create high-quality labeled training data, improving generalization and robustness across varied conditions.
* Applied targeted data augmentation to enhance dataset diversity and improve detection under environmental noise.
* Tuned model performance through extensive hyperparameter optimization, achieving mAP@50: 82, mAP@50–95: 0.51, precision: 0.85, and recall: 0.71 on validation sets.
* Designed and tested multiple training strategies to maximize real-world performance, enabling reliable deployment in resource-constrained robotic environments.
* Worked closely with team members and faculty sponsor to align system performance with business needs.

**PROJECTS**

\*\*Traffic Congestion Prediction using Big Data\*\*

* Built a big data pipeline to process 94.5M+ rows of NYC DOT Traffic Speeds data using Apache Spark on Google Cloud Dataproc. Engineered features like hour, weekday, and speed delta to train a PySpark Random Forest classifier for congestion level prediction.
* Trained a PySpark Random Forest classifier for congestion level prediction
* Deployed the model via Flask REST API, enabling real-time inference integrated with Logstash and visualized results using Kibana

Achieved ~85% accuracy in classifying congestion levels (low/medium/high) and displayed key metrics using interactive dashboards

* \*\*Facial Attribute Classification and Enhancement\*\*
* Developed a deep learning pipeline to classify facial attributes such as age group, gender, and emotion using convolutional neural networks (CNNs) and OpenCV for preprocessing. Integrated facial landmark detection to align and crop facial regions, improving feature localization and enhancing model robustness to pose and lighting variations.
* Classified facial attributes using convolutional neural networks (CNNs)

Integrated facial landmark detection to improve feature localization and model robustness

* Achieved high accuracy across multiple attribute classes on benchmark datasets and deployed a prototype capable of real-time predictions on webcam feeds