**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**

Detail-oriented Data Scientist with strong expertise in Python, machine learning (PyTorch, TensorFlow), and data pipeline automation. Proven ability to deploy real-time analytics solutions and optimize models for performance and interpretability. Experienced in statistical modeling, data engineering with Spark, and transforming raw data into actionable business insights. Passionate about scalable AI systems and cross-functional collaboration in fast-paced environments.

**SKILLS**

* **Languages & Frameworks:** Python (NumPy, Pandas, Scikit-learn, Matplotlib), PyTorch, TensorFlow, SQL, Spark
* **Machine Learning:** Supervised & Unsupervised Learning, Model Optimization, Time Series, NLP, Transfer Learning
* **Data Engineering:** PySpark, Apache Spark, Real-Time Stream Processing, ETL Pipelines, GCP, REST APIs
* **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.
* Collaborated on a capstone project with Kevares, developing a real-time YOLO-based object detection system for robotic environments.

|  |  |
| --- | --- |
| **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.

**PROJECTS**

**Traffic Congestion Prediction using Big Data** (GCP, PySpark, ELK)

* 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.
* 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.

**RAG-Based Q&A System with LLM Integration**

* Developed a Retrieval-Augmented Generation (RAG) pipeline to enhance document-based Q&A with LLMs and private knowledge sources.
* Used document chunking and vector embeddings (FAISS) for efficient storage and retrieval.
* Applied k-nearest neighbors (KNN) search to fetch relevant context and improve response accuracy.
* Optimized query retrieval with multi-query rewriting, semantic routing, and HYDE for better document alignment.
* Integrated sub-question decomposition and hierarchical indexing to refine retrieval and improve context relevance.
* Implemented a Flask-based API for real-time interactions.

**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.
* Applied transfer learning using pretrained CNN backbones (e.g., VGG16, ResNet) with custom classification heads, and incorporated batch normalization to stabilize training and accelerate convergence.
* Achieved high accuracy across multiple attribute classes on benchmark datasets and deployed a prototype capable of real-time predictions on webcam feeds.

**E-Commerce Price Prediction**

* Built a machine learning pipeline to predict future product prices using historical sales data, seasonal discount trends, and promotional activity metadata.
* Cleaned, transformed, and feature-engineered tabular datasets to prepare for supervised learning using regression models such as Random Forests and Gradient Boosting.
* Conducted model evaluation using RMSE and cross-validation, tuning hyperparameters for optimized performance.
* Delivered actionable insights into pricing fluctuations to aid marketing and inventory strategy through visualization dashboards.

**Leafio** – Real-Time Plant Leaf Disease Detection App

* Led the development of an Android app using CNNs for plant disease detection, integrated with a Flask API for real-time predictions, achieving over 85% accuracy. The model, created by our team, was hosted on a server and accessed via API calls.
* Designed and trained a custom CNN model using a dataset of diseased and healthy plant leaves, implementing data augmentation techniques to improve model robustness.
* Developed a Flask API for model deployment, hosting the trained model on a server and enabling seamless communication with the mobile app via API calls.
* Published a research paper on the comparative study of CNN-based disease detection, contributing to AI-driven advancements in precision agriculture.

**TeSteps** – Social Networking Platform with Automated Grading

* Developed a dynamic full-stack social networking platform using PHP, Apache server, and MySQL, featuring user registration, authentication, and interactive profiles.
* Designed custom post and comment modules allowing real-time interactions and post visibility based on user roles and privacy settings.
* Integrated an NLP-based automated test evaluation system via external API to assess user-submitted responses by comparing them with reference answers, utilizing cosine similarity and keyword overlap techniques.
* Achieved ~80% grading accuracy, reducing manual evaluation efforts and providing personalized feedback to users through dashboards.

**Face-it** – Real-Time Facial Recognition Web App

* Built a responsive web application using React for frontend and Node.js with Express for backend REST APIs, enabling secure user authentication and activity tracking.
* Integrated Clarifai's deep learning-powered face detection API to enable real-time recognition from uploaded images and webcam streams.
* Implemented PostgreSQL to persist user data, login activity, image count, and engagement history, with a ranking system to gamify usage.
* Optimized UI/UX with dynamic feedback and asynchronous requests, resulting in a smooth, interactive user experience across devices.

**SwiftTrend** – Cross-Platform E-commerce App with Admin CRM

* Developed a scalable e-commerce mobile app using Flutter for Android and iOS, providing a seamless cross-platform shopping experience.
* Implemented backend services in Node.js with Express, using MongoDB for data storage, covering user accounts, product listings, and order history.
* Integrated PayPal SDK for secure checkout and payments, and enabled search filtering, cart management, and product reviews.
* Built an internal CRM dashboard for admins to add/edit products, manage inventory, view order analytics, and fulfill orders—enhancing operational efficiency.

**PUBLICATIONS**

**A Comparative Study on Convolutional Neural Network Based Plant Leaf Disease Detection**

*Shubham Patel, Sujay Patil, Sourabh Zadbuke, Som Kamthe*

*MIT School of Engineering, Maharashtra, India – 2022*

* Analyzed CNN architectures (Xception, VGG16, MobileNet) for plant leaf disease classification using Keras and TensorFlow.
* Applied transfer learning and fine-tuning; evaluated performance on the DrPlantify dataset using training, validation, and testing splits.

**SELF-DIRECTED LEARNING**

* **Machine Learning Specialization**
  + Supervised Machine Learning: Regression and Classification
  + Advanced Learning Algorithm
  + Unsupervised Learning: Recommenders, Reinforcement Learning