**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 hands-on experience applying statistical modeling, machine learning, and Python-based workflows to real-world datasets. Built scalable pipelines using Jupyter, PySpark, and REST APIs, with a focus on model interpretability and actionable insights. Proficient in Python, SQL, and visualization tools including Kibana and Power BI (learning). Adept at correlation analysis, data wrangling, and cross-functional collaboration to deliver market-aligned analytics.

**SKILLS**

* **Languages & Frameworks:** Python, SQL, PySpark, Scikit-learn, TensorFlow, Flask
* **Statistical Methods:** Regression, Classification, Correlation Analysis, A/B Evaluation
* **Tools:** Jupyter Notebooks, Power BI (learning), Kibana, GitHub
* **ML Techniques:** Model Tuning, Forecasting, Feature Engineering, Semantic Retrieval
* **Deployment:** REST APIs, MLflow (basic), CI/CD Concepts
* **Collaboration:** Agile Projects, Stakeholder Communication, Cross-functional Alignment
* **Interests:** Market Analytics, BI-Driven Insights, Automated Statistical Reporting

**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** (GCP, PySpark, ELK)

* Built predictive models using PySpark + Random Forest to classify NYC congestion patterns from 94M+ rows.
* Conducted feature correlation analysis, time-based EDA, and deployed results to Kibana dashboards.
* Used CI/CD logic to simulate automated retraining pipelines for statistical monitoring and insight automation.

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

* Designed a semantic search pipeline using Transformers, FAISS, and modular APIs for querying structured documents.
* Processed large unstructured datasets into searchable formats, aligning with BI-style knowledge workflows.
* Focused on retrieval accuracy, relevance ranking, and statistical consistency across document types.

**Facial Attribute Classification**

* Built a CNN model with OpenCV preprocessing and VGG16 transfer learning for multi-class facial attribute prediction.
* Conducted hyperparameter tuning, batch normalization, and exploratory data evaluation for training optimization.
* Deployed the model as a REST API, simulating business-facing model access for low-latency inference.