

EDUCATION

- Dalhousie University**
Master of Digital Innovation - Data Science
- Halifax, NS
Sept. 2024

EXPERIENCE

- TechVida Labs**
Data Science Intern
- Ahmedabad, India
Jan 2023 - Apr 2023
- Developed **RESTful APIs**: Designed and deployed scalable **Flask-based APIs** for data-driven applications, improving system efficiency.
 - Web Scraping**: Built automated web scrapers using **BeautifulSoup** and **Selenium** to extract real-time data for analysis.
 - Data Analysis**: Performed extensive data cleaning and analysis using **Pandas**, **NumPy**, **Excel**, **Power Bi**, **Tableau**, **SQL** and **SQL Server**, leading to actionable business insights.
 - Team Collaboration**: Collaborated with cross-functional teams, refining development workflows and reducing project delivery time by 10%.
 - Visualization Dashboards**: Developed interactive dashboards in **Tableau** to showcase key trends and predictive insights.
 - Machine Learning Models**: Implemented **machine learning pipelines and ensemble predictive models** using **Scikit-learn**, **Tensorflow**, **Keras** and **PyTorch** and **Pickle** to save and load model to improve decision-making processes.
 - Tech Stack**: Flask, Python, Pandas, NumPy, Power BI, SQL, SQL Server, BeautifulSoup, Selenium, SKlearn, Tensorflow, Keras, PyTorch and Machine Learning Pipelines

SKILLS AND CERTIFICATION

Languages: Python, SQL, R, MATLAB, JavaScript, React
Machine Learning: Scikit-learn, TensorFlow, PyTorch, XGBoost, SMOTE, Random Forest, Logistic Regression, Neural Networks, Transformers, CNN, LSTM
Computer Vision: YOLOv8, OpenCV, TensorFlow Object Detection API, Transfer Learning, SigLIP, ByteTrack
Cloud & Deployment: Amazon SageMaker, AWS S3, SageMaker Endpoints, TensorFlow.js, TFLite, Model Optimization
Data Processing: Pandas, NumPy, UMAP, Feature Engineering, Data Pipelines, ETL
Visualization: Matplotlib, Seaborn, Plotly, Streamlit, Interactive Dashboards
Tools: Git, Docker, Jupyter, VS Code
Certifications: IBM Data Science Specialization, Machine Learning with Python (IBM), Databases and SQL for Data Science (IBM), Data Science Math Skills (Duke University)

PROJECTS

Soccer Player Detection and Match Analysis System

Python, YOLOv8, Amazon SageMaker, PyTorch, OpenCV, UMAP, ByteTrack

- Computer Vision Pipeline**: Developed a comprehensive soccer match analysis system using **YOLOv8** models for player, ball, and referee detection with training datasets stored in **Amazon S3**.
- Team Classification**: Implemented team classification using **SigLIP embeddings** and **UMAP** dimensionality reduction, achieving 94% accuracy in distinguishing teams based on visual features.
- Pitch Mapping**: Created a homography-based view transformer to map camera coordinates to pitch coordinates, enabling spatial analytics for player positioning and team formations.
- Model Training**: Trained custom object detection and keypoint recognition models on **Amazon SageMaker** using GPU instances, optimizing hyperparameters for soccer-specific detection tasks.
- Ball Trajectory Analysis**: Developed algorithms for ball trajectory tracking and filtering using **ByteTrack**, providing insights into possession patterns and play development.

- **Visualization Tools:** Built interactive visualization tools for match analysis including Voronoi diagrams for team control areas and player movement heatmaps.
- **Model Deployment:** Deployed trained models using **SageMaker Endpoints** with auto-scaling for real-time video processing, enabling efficient analysis of match footage.

Predicting Chronic Conditions Using Machine Learning

Python, Scikit-learn, TensorFlow, XGBoost, Streamlit, SMOTE, Pandas

- **Data Processing Pipeline:** Developed a comprehensive data processing pipeline for the Canadian Community Health Survey dataset, implementing categorical transformations and bivariate statistical analysis to prepare features for chronic condition prediction.
- **Multi-Target Modeling:** Designed a sequential prediction framework for three chronic conditions (high blood pressure, diabetes, cardiovascular disease), reflecting clinical progression and preventing data leakage between models.
- **Class Imbalance Handling:** Implemented **SMOTE** and **RandomUnderSampler** techniques to address significant class imbalance, improving model performance for minority health condition classes.
- **Model Evaluation:** Evaluated multiple classification algorithms including **Logistic Regression**, **Random Forest**, **XGBoost**, and **Neural Networks** using 5-fold cross-validation, achieving ROC-AUC scores of 0.81-0.86 across target conditions.
- **Threshold Optimization:** Applied modified **Youden's J statistic** to optimize decision thresholds, increasing recall from 78% to 96% for high blood pressure detection while maintaining acceptable precision.
- **Feature Importance Analysis:** Extracted and visualized logistic regression coefficients to identify key predictors for each chronic condition, enhancing model interpretability for healthcare applications.
- **Interactive Screening Tool:** Deployed the final models as a **Streamlit** application with progressive prediction capability, enabling non-clinical screening for chronic disease risk factors.

Hand Gesture Recognition with TensorFlow Object Detection API

Python, TensorFlow, OpenCV, TensorFlow.js, TFLite, Computer Vision

- **Custom Dataset Creation:** Developed an image collection pipeline using OpenCV to capture and organize training data for four distinct hand gestures, implementing automated file naming and directory structure creation.
- **Data Annotation & Preprocessing:** Utilized labelImg for bounding box annotation and implemented TFRecord conversion scripts to transform raw image data into TensorFlow's optimized training format.
- **Transfer Learning:** Configured and fine-tuned a pre-trained SSD MobileNet V2 FPNLite model using TensorFlow's Object Detection API, adapting the network architecture for the custom gesture recognition task.
- **Model Training & Evaluation:** Executed model training with custom hyperparameters and batch size optimization, achieving high detection accuracy across all gesture classes through systematic evaluation.
- **Multi-Platform Deployment:** Exported the trained model to multiple formats (TensorFlow.js, TFLite) for cross-platform deployment, enabling real-time gesture detection on both web and mobile applications.
- **Real-Time Detection System:** Implemented a webcam-based real-time detection system using OpenCV integration, optimizing inference speed while maintaining detection accuracy for interactive applications.
- **Model Optimization:** Applied model freezing and quantization techniques to reduce model size by over 75% while preserving detection performance, making it suitable for resource-constrained environments.

LEADERSHIP

-
- **Dalhousie Digital Innovation Society** 2024 - Present
Data Science Student Representative
 - **Student Advocacy:** Representing Data Science students, fostering engagement with faculty, society and peers to enhance learning experiences.
 - **Workshop Development:** Proposing and organizing workshops to help students develop essential technical and analytical skills in data science.