Kartik **Sirwani**

LEAD DATA SCIENTIST

(+1) 604 966 1181 | kartiksirwani@gmail.com | about.me/kartiksirwani | kartikwar | kartik-sirwani

Summary

Lead Data Scientist with 8+ years of experience in Machine Learning, Deep Learning, NLP, and MLOps. Specializing in LLMs, Computer Vision, and cloud-based model deployment, with expertise in model acceleration, optimization, and inference efficiency using AWS, GCP, TensorFlow, and PyTorch. Proven track record of building and deploying large-scale AI solutions that drive business impact. Passionate about low-latency AI, hardware-aware ML, and mentoring teams to innovate through A/B testing, experimentation, and deep learning advancements.

Skills

Languages and Technologies

- Programming & Tools: Python, NLTK, PyTorch, OpenCv, Pillow, TensorFlow, Pandas, scikit-learn, C++, CUDA, ONNX, TensorRT
- Plotting & Visualization: Seaborn, Altair, Matplotlib (for AI analysis)
- Machine Learning & AI: Deep Learning, Transformers, LLMs (GPT-4, LLaMA), GenAI, NLP, Computer Vision, Pruning, Model Optimization, Quantization, Distillation, RAG
- Cloud : ECS, Lambda, CloudWatch, Autoscaling, Route53, EC2, SQS, S3, LoadBalancer, Docker
- Experimentation & Analytics: A/B Testing, Hypothesis Testing
- Databases : MongoDB, SQL, Snowflake, BI tool (Looker)
- MLOps & Deployment: CI/CD (GitHub Actions), AWS Lambda, ECS, Docker, Cloud Deployment (AWS, GCP)
- Soft Skills: Cross-functional Collaboration, Mentorship, Technical Leadership, Business Impact Optimization

Work Experience

Cimpress

LEAD DATA SCIENTIST *Sept 2020 - July 2024*

- **Background Removal API** :
  Directed the development of a high-volume background removal API for **backgroundly.io**, handling 200,000+ daily requests and saving $3M annually.
  Innovated and trained state-of-the-art **FBA-Matting** and U-Net models, achieving an 87 percent visual acceptance rate in saliency detection.
  Technologies: python, pytorch, opencv
- **Create Portrait API** :
  Designed an API to transform regular photos into portrait-style images by maintaining subject focus and subtly blurring the background.
  Collected data for training the model using snowflake and looker.
  Partnered with cross-functional teams to integrate the **MegaDepth model** with in-house background removal services.
  Optimized model inference speed by 20% using quantization and pruning techniques. Technologies: Python, PyTorch, OpenCV, Pillow.
- **AI Business-Card Generator** :
  Supervised the development of an AI-powered business card generator leveraging GPT-3.5 to generate layout recommendations and Stable Diffusion for AI-driven background creation, and Retrieval-Augmented Generation (RAG) to utilize a database of business card backgrounds, enabling real-time, print-ready customization and reducing design iteration time by 60%.
  Deployed ML pipelines with AWS Lambda & CI/CD using GitHub Actions, automating model updates.
  Technologies: Python, OpenAI API, Hugging Face, Stable Diffusion, FastAPI, AWS Lambda, Docker.

Foghorn Systems

DATA SCIENTIST *Feb 2020 - Sept 2020*

- **Mask Detection Model**:
  Pioneered a real-time mask detection system to identify individuals not wearing masks in factories during the COVID-19 pandemic, triggering real-time alerts when counts exceeded thresholds via IIoT.
  Trained an **SSD MobileNet V2** object-detection model, achieving 85% precision.
  Technologies: Python, TensorFlow, OpenCV.
- **Connector Detection Model**:
  Architected a pipeline for identifying unsafe pipe connectors in oil factories, preventing hazardous drilling operations.
  Elevated model precision to 88% using a **YOLO v3** object-detection framework.
  Technologies: Python, TensorFlow, OpenCV.

Razorthink

SENIOR ARTIFICIAL INTELLIGENCE ENGINEER *Sept 2018 - Dec 2019*

- **Table Detection Model**: Formulated a deep learning model to detect table-like structures in PDF documents.
  Trained a Faster R-CNN (VGG16) network using curriculum learning, achieving 84% precision.
  Technologies: Python, TensorFlow, OpenCV.
- **Template Detection Service**: Engineered an AI-powered service to compare the layout and structure of PDF documents, classifying similar documents under the same template.
  Oversaw the development and deployment of a Siamese network using a pre-trained VGG16 model.
  Technologies: Python, TensorFlow, OpenCV, MongoDB.

Nowfloats

BACKEND DEVELOPER *June 2016 - Sept 2018*

- **Update Synchronize API**: Orchestrated an API to synchronize merchant updates and reviews across social platforms like Facebook, LinkedIn, Twitter, and Quikr, serving over 19,000 customers and processing 50,000 weekly updates.
  Implemented REST APIs, services, Lambda functions, cron jobs, and created deployment pipelines on ECS.
  Technologies: Python, NodeJS, ECS, Docker, Lambda, Express, MongoDB, Route 53, Ubuntu, SQS.
- **Update Categorization Service**: Collaborated on an NLP-based service that fetched and categorized customer product updates into offers, discounts, or sale prices using natural language processing algorithms such as 'bag of words' and an SVC (Support Vector Classifier).
  Handled class imbalance between updates by using appropriate metrics like f1-score instead of accuracy. Technologies: Python, Scikit-Learn, Pandas, Matplotlib, MongoDB, MySQL.
- **Purchase Probability Model**: Conceptualized a predictive model to analyze sales data and predict purchase probabilities based on customer characteristics, reducing acquisition costs by 50% and increasing conversion rates from under 2% to 20%.
  Performed feature engineering and exploratory data analysis to find the right features for classification.
  Developed classifiers using logistic regression and decision trees.
  Technologies: Python, Scikit-Learn, Pandas, Matplotlib, MongoDB.

Education

University of British Columbia, Vancouver *Vancouver, Canada*

MASTER OF DATA SCIENCE IN COMPUTATIONAL LINGUISTICS *Aug 2024 - June 2025*

Courses:- Computational Semantics, Advanced Corpus Linguistics, Sentiment Analysis, Machine Translation

Birla Institute of Technology and Science, Pilani *Pilani, India*

B.PHARM. PRE-PHARMACY STUDIES *2012 - 2016*

Professional Courses

Nov 2020 **Deep Neural Networks with PyTorch**, IBM
Nov 2017 **Applied Machine Learning in Python**, University of Michigan on Coursera
Oct 2017 **Applied Plotting, Charting and Data Representation in Python**, University of Michigan on Coursera
Sept 2017 **Introduction to Data Science in Python**, University of Michigan on Coursera

Certificates

Nov 2020 **Communicating about Culturally Sensitive Issues** , Linkedin
Nov 2020 **Confronting Bias: Thriving Across Our Differences**, Linkedin
Oct 2020 **Unconscious Bias**, Linkedin