Neetesh Dadwariya

Machine Learning Engineer | Data Scientist
Certified AWS Machine Learning Specialist

neeteshdadwariya@gmail.com **|+1-469-927-3826 | Seattle, WA**
https://www.linkedin.com/in/neeteshdadwariya/
https://neeteshdadwariya.github.io/ | https://github.com/NeeteshDadwariya

SUMMARY

Professional Machine Learning Engineer with 6 years of broad-based experience in building data-intensive applications, overcoming complex machine learning architectural, and scalability issues in diverse industries. Proven track record in predictive modeling, data processing, machine learning algorithms as well as computer vision techniques, including PyTorch, OpenCV and PySpark. Adept at coordinating with cross-functional stakeholders to execute revenue generating technology projects.

PROFESSIONAL EXPERIENCE

BlackBerry
Machine Learning Engineer [Jan 2023 – Present]

- Steganography detection in images downloaded by various processes to prevent hidden code in images. Built a PyTorch based image pipeline to calculate embedding for malicious images.
- Converted received image bytes to tensors in PyTorch, performed embedding to identify malicious code injection in images. Using the embedding, created feature vectors and performed image classification in 'safe' or 'malicious' image using DNN models.
- **Architected** and implemented *Machine Learning Classification Pipeline* using **XGBoost** for the detection of malicious processes running in Windows systems.
- Incorporated TF-IDF, Sent2Vec embeddings and performed *Feature Engineering* on PySpark. Processed Gigabytes of streaming data to identify anomalies in executed commands. Verified the binary classification **accuracy as 97%** with **95% *Confidence Interval***.
- Utilized **PyTorch** to create **Deep Neural Network (DNN)** model that successfully detected anomalies in streaming event data. Reduced false anomalies by 10%. Performed *Cyclical Encoding* on the day-of-week and hour-of-day features.
- **Conducted** SHAP analysis in AWS SageMaker on deployed DNN model, with the goal of explaining its outcomes and providing stakeholders with insights into feature contribution.

BlackBerry
Data Scientist - Intern [May 2022 – Aug 2022]

- **Leveraged** statistical data science techniques including **Hypothesis Testing** and Confidence Intervals in Pandas-on-PySpark to perform Network Anomaly Detection on user login and logoff events from approximately 40GB of Windows Event Sensor Data, successfully flagging anomalies with 88% accuracy.
- Conducted Tenant-wise anomaly detection using the ***Time Series Prophet*** *Model*.
- Performed **PCA** on all features to extract relevant transformed features using dimensionality reduction.
- Applied ***Winsorization*** to reduce the effect of outliers by scaling the data. Identified service accounts using K-Means unsupervised clustering algorithm to improvise overall detections improving detections by 20%.

Make My Trip
Senior Software Engineer-II [Jun 2016 – Jul 2021]

- Implemented Machine Learning driven travel chat bot to answer customer travel queries. Used slot filling, entity extraction, intent identification using Custom Embedding for Travelling and Dialog WorkFlows and respond according to business use case.
- **Experienced** in the area of Software Design & Development of applications using **Springboot** Microservices automating 60% of the manual processes.
- **Spearheaded** development and maintenance of backends for all post-booking **microservices**, such as booking cancellation, change in travel dates, and purchasing flight seats and baggage reducing manual overhead by 80%.
- **Contributed** in entire SDLC process of Software development by adopting Agile and Test Driven Development (TDD) methodologies as Individual Contributor.
- **Successfully delivered** in-house projects utilizing SpringBoot, ReactJS, **Kafka**, MySQL, MongoDB, and Docker containers. These projects included various automated features, resulting in a 25% reduction in ticket resolution time.

TECHNICAL SKILLS

Machine Learning	Prefect, MLFlow, PyTorch, PySpark, TensorFlow, Spark MLLib, Hive, Scikit-learn, SageMaker, Data Wrangler
Computer Vision	OpenCV, Pillow, Image Processing, Edge and Contour Detection, Image Enhancement, Object Tracking
Data Science	Predictive Modeling, Statistical Analysis, Forecasting, Random Forest, Recommendation System, Confidence Interval
Big Data	Databricks, Azure ML, AWS Kinesis, AWS Firehose, AWS Lambda, PySpark, AWS EMR, AWS S3, Hive, Kafka
Software Development	Python**,** Java, SpringBoot, Maven, NodeJS, Elasticsearch, ReactJS, Microservices, MySQL, MongoDB, Architecture Design
DevOps and Add-On	AWS ECS, Docker, Kubernetes, Edge, Git, Jenkins, JIRA

EDUCATION

The University of Texas at Dallas — **Richardson, TX**
Master of Science, Computer Science **(GPA: 3.973/4.0)** — Aug 2021 – Aug 2023

Motilal Nehru National Institute of Technology — **Allahabad, India**
Bachelor of Technology, Electronics & Communication Engineering **(CPI: 8.43/10)** — Jul 2012 – May 2016

PROJECTS

Malware detection in Process Command Line on Windows Event Sensor Data | *XGBoost, DNN, PyTorch, Databricks* **[BlackBerry]**

- Successfully architected PyTorch DNN malware classification model pipeline to flag out malicious events in windows event data.
- Used TF-IDF, Sent2Vec embedding for feature creation, Performed XGBoost classification with hyperparameter tuning and cyclical encoding for day and hour features obtaining model accuracy of 98%.

Shoe brand Image Classification on Nike-Adidas Dataset | *CNN, PySpark, OpenCV, Seaborn* **[Link]**

- Constructed shoe brand image classification model on Nike-Addidas dataset using MapReduce based Convolution Neural Network.
- Performed training and inference on Spark Distributed Computing Environment.
- Incorporated Data Augmentation to enhance training data, rescaling. Experimented and logged various Hyperparameters, obtained model accuracy of 80%.
- Additionally, partially retrained model using transfer learning on 'InceptionV3' model to achieve model accuracy as 94%.

Transfer Learning based Flower Classification | *TensorFlow, OpenCV , Image Augmentation, RESNet-V3* **[Link]**

- Developed a Flower classification model using transfer learning and fine-tuned it by retraining the back layers after removing them. By moderating hyper-parameters. Achieved an optimal accuracy of 95%.

Scene Recognition – Convolutional Neural Network bases classification Model | *CNN, OpenCV, Python, Numpy* **[Link]**

- Designed and implemented a 12-layered architecture in Python that incorporated different convolutional network layers, such as Conv2D, MAX Pooling, Dense, and Flatten, all without relying on existing deep learning libraries.
- Conducted scene recognition on various outdoor environments, including sea, buildings, forests, and glaciers, performed hyperparameter tuning and achieving an accuracy of 94.8% in distinguishing different scenes.

Automatic Road Lane Detection | *OpenCV, Python* **[Link]**

- Constructed a static computer vision pipeline capable of detecting white, yellow, and red lane lines using Gaussian Blurring, Canny Edge Detection, Color Masking, and Hough Line Transform. Additionally, incorporated stop sign detection through Contour Detection.

Movie recommendation system using ALS Matrix Factorization | *PySpark, Collaborative Filtering* **[Link]**

- Performed collaborative filtering based on user's past inputs to predict movies. Implemented PySpark + Docker-driven separate train and inference pipeline. Analyzed model performance with RegressionEvaluator using RMSE. Achieved accuracy of 98.3%.

Myra – Travel Chatbot | *NLP, Entity Extraction, Slot Filling, Microservices* **[Link]**

- Bot implementation using NLP for solving customer travel related queries. Responsible for API's powering model for travel-data. Recognized with Tripper of The Month and Tripper of The Quarter for deliverables.

CERTIFICATIONS AND AWARDS

- **AWS Certified Machine Learning Specialty MLS-C01** [*Amazon Web Services Training and Certification*]
- **Python for Computer Vision with OpenCV and Deep Learning** [*Udemy*, Feb-22]
- **Advanced Software Engineering** [CodePath.org, Aug-21]
- **AI-ML Career Track** [*Springboard*, Dec-20]
- Green Hat Award for Exemplary Innovation at MakeMyTrip India Pvt Ltd. [Dec-20]
- *Tripper of the Quarter* & *Tripper of the Quarter (Thrice)*, Exemplary contribution Award at MakeMyTrip India Pvt Ltd. [Dec-19]