Vijay Gandhi

Machine Learning Engineer / Data Scientist

Chicago, USA • vijayg.ece.4@gmail.com • (551) 260-1912

Machine Learning Engineer with 5+ years of professional experience in AI/ML, cloud platforms, data analysis, visualization, ETL data pipelines. Highly skilled in end-to-end data lifecycle encompassing various phases from requirements to delivery of solutions. Recognized for navigating diverse projects to success, including sentiment analysis and nutrition table extraction. Proven ability to quickly adapt and master new tools, technologies, and languages. **Excellent communication skills and proven expertise working in a team and individual settings.**

TECHNICAL SKILLS

Languages:	Python, R, SQL, Java, PySpark, SQL Server
Data Science:	XGBoost, Random Forest, Deep Learning (BERT, RNN, CNN), Statistical tests
Cloud Platform:	GCP, AWS, Azure
ML Libraries:	TensorFlow, PyTorch, Scikit-Learn, XGBoost, Keras, AutoML, Spark, Kafka, Pandas
Visualization:	Tableau, Power BI, Microsoft Excel, Data Modelling, DAX, ETL

PROFESSIONAL EXPERIENCE

Machine Learning Engineer *May 2021– Present*
Syndigo, USA

- Built, trained, and deployed **nutrition table extraction deep learning model (CNN)** from product images with **83% accuracy** using **Python, SQL, Cognitive Services** that directly resulted in reducing the business cost by 80%, time by 60%, and increased client retention**.** Check out an example of model's cropped and text detected image here.
- Trained, tested, and deployed a deep learning-based **sentiment analysis NLP model (BERT)** using **Python, PyTorch** to analyze the sentiment of customer feedback with **92% F1-score**.
- Built, trained, and deployed dollars left on table **regression ML model** to quantify the missed sales opportunity in dollars with **0.73 R-squared** value using AutoML in Databricks. Performed hyperparameter tuning to optimize model performance.
- Designed and developed a **microservice application** using **Flask** to host the ML models, containerized the application using **Docker**, built a docker image, and deployed the container in **Kubernetes cluster**
- Utilized **MLflow** to implement **MLOps framework** to periodically retrain the ML models with the new data coming into the system and also learn from its own mistakes.
- Developed **interactive reports, dashboards to analyse and visualize** the TB-scale e-commerce datasets and identified trends, patterns to make data-driven decisions using **Excel and Tableau**
- Ingested **structured and unstructured e-commerce data** from multiple sources such as Databricks, SQL Server, Sharepoint, Blob Storage and standardized it using ETL for BI reporting.
- Performed **data modelling in Tableau** to combine multiple tables for analysis with relationships and published it for reuse
- Setup functions, triggers, stored procedures in **SQL Server relational database** and utilized TSQL to query, manage, and update product and product metadata JSON

- Developed an end-to-end **ETL cloud data pipeline in python, sql, and databricks** to collect, organize, transform e-commerce products data, identified and resolved the data quality issues in products increasing the revenue by 30%
- **Orchestrated the ETL data pipelines** using **Azure Data Factory** to schedule or trigger the pipelines
- Proficient in **agile development methodology** and project management tools like **JIRA**
- **Collaborated cross-functionally** with diverse stakeholders, product managers, business analysts and communicated the results of analysis in a clear and effective manner.
- **Tech Stack**: **Python, SQL, Tensorflow, PyTorch, Docker, Kubernetes, MLflow, Flask, PySpark, Microsoft Azure (Cognitive Services, Databricks, Data Factory, ADLS), SQL Server, AutoML**

Data Scientist *Jan 2019–Dec 2020*
NYU, USA

- Cross-functional collaboration with internal and external teams, followed agile development methodologies, communicated the results in an effective manner with technical and non-technical stakeholders.
- Built and trained **predictive and statistical models** to remotely diagnose cardiovascular heart diseases like sudden cardiac arrest, hypertension using wearables like Fitbit and Garmin in real-time. Published the work and results in IEEE conferences. **Check out my paper publications** here
- Performed exploratory data analysis and developed **machine, deep learning models** using **Python, SQL, AWS, PyTorch** and deep learning models like **Recurrent Neural Networks**, **Long Short-Term Memory (LSTM)** achieving **93.7% best F1-score.** Integrated the LSTM model with the application to continuously monitor user's heart rate and send alerts if a disease is diagnosed.
- Containerized the ML model microservice using **docker**, built the docker image and pushed it to docker hub, and deployed it using **AWS Sagemaker** for model serving.
- Setup **MLOps framework** using **Airflow** to continuously retrain the models with curated proportion of old and new data to keep the model relevant to the behavioural changes over time
- Implemented and deployed **ETL data pipeline** to connect to multiple sources, ingest data, transform and feed it to the **ML model** to make predictions. Stored the results of the prediction in **RedShift Data Warehouse** for BI and reporting applications.
- **Visualized, analysed, and presented the findings** of cardiovascular dataset as an interactive report to the business audience using **Power BI.**
- Loaded the aggregated data into **RedShift data warehouse** for business intelligence and analytics applications/use-cases.
- Utilized Python, Spark SQL, TSQL, Excel to query, explore, process, and transform healthcare datasets
- Create, grow, and manage Vertica database instances in all environments such as Dev/QA/Prod in different regions
- Responsible for **performance tuning of the database** to improve the efficiency of SQL queries
- Stored the code repository in **version control system** such as **Github** for active collaboration with developers
- **Tech Stack: Python, SQL, Tensorflow, PyTorch, AWS (Sagemaker, EC2, S3, RedShift), Flask, Docker, Kubernetes, Pandas, Keras, Scikit-Learn, Seaborn, Matplotlib**

Machine Learning Engineer *June 2017–Dec 2018*
Pluto7, India

- Built, trained, and tested classification ML models for customer churn prediction with **92% F1-score** using Tensorflow and XGBoost.
- Performed K-means clustering algorithm on customer data to identify similar customers based on various parameters like user behaviour, products subscribed using Tensorflow

- Developed a microservice using Flask to deploy the ML models, containerized the application using Docker and deployed it on a Kubernetes cluster.
- Experience collecting the customer dataset using SQL from various databases and visualize them using BI tools like Tableau to identify the trends and patterns to make data-driven decisions
- Presented the interactive reports and dashboards to the C-suite executives and communicated the results of the analysis in clear and effective manner
- Utilized custom charts such as Sankey diagram, drill-down charts to summarize and present the customer sales data across years
- Setup stored procedures to update and manage the customer data in SQL Server, tuned the performance of the database to retrieve the query results faster by 5x.
- Built ETL pipeline to collect data from multiple sources, cleaned, transformed, and stored results in Azure Synapse Data Warehouse to derive metrics and KPIs for analytics.
- Optimized SQL queries by partitioning the data and implemented the best practices around keys which improved the performance by 20% to query Terabyte scale databases.
- Migrated on-prem data to Azure cloud helping client with their digital/cloud transformation strategy
- Cross-functional collaboration with different organizations and stakeholders from collecting the business requirements to delivering the project
- **Tech Stack: Python, Tensorflow, PyTorch, R, Excel, SQL, Tableau, Microsoft Azure, Databricks, SQL Server, Pandas, Keras, Scikit-Learn, XGBoost**

EDUCATION & CERTIFICATION

Master of Science in Computer Science
New York University, USA

Bachelor of Technology in Computer Engineering
National Institute of Technology - Trichy, India