RajDeepak Beniwal

Data Scientist

rajdeepakbeniwal@gmail.com | www.linkedin.com/in/rajdeepak-beniwal

Location: Sydney | PH. No. : +61497100174 | Full Working Rights

Senior Data Scientist & ML Engineer with **8+ years** of expertise in **AI, GenAI, Machine Learning, NLP, LLMs, MLOps, and Big Data** across **Banking, Telecom, Finance, and Manufacturing**. Skilled in **AI/ML solution design, end-to-end MLOps, and cloud data technologies (Azure/AWS)**.

Led **AI/ML initiatives** as a **Solution Architect** at **TCS Microsoft Business Unit**, developing **automated CI/CD pipelines, data drift detection, and real-time model monitoring** with **Azure MLOps, MLFlow, and Synapse Analytics**.

Proficient in **ETL pipelines, scalable ML pipelines, and cloud data management (Azure Data Lake, Synapse, Delta Lake)**. Strong background in **containerization (Docker, Kubernetes)** and **full-stack development (Python, Flask, FastAPI)**.Passionate about leveraging **AI and scalable data architectures** to solve complex business challenges and drive innovation.

Skills Summary

- ETL & Data Pipelines: **Azure Data Factory, AWS Glue, Snowflake**
- Cloud Data Platforms: **Azure Data Lake, Azure Synapse, Databricks, AWS S3**
- Big Data Technologies: **Apache Spark, Kafka, Hadoop, Delta Lake**
- SQL & Database Management: **SQL, NoSQL, Azure Cosmos DB**
- Languages: **Python, SQL, PySpark**
- Model Monitoring & Drift Detection: **Azure Monitor, MLFlow, A/B Testing**
- Version Control: **Git, GitHub, GitLab**
- APIs & Web Frameworks: **Flask, FastAPI**
- Supervised & Unsupervised Learning: **Decision Trees, Random Forest, XGBoost, LGBM, Naïve Bayes, SVM, Clustering**
- Deep Learning: **PyTorch, TensorFlow, Keras, RNN**
- Natural Language Processing (NLP) & Generative AI: **Azure OpenAI, Transformers, LangChain**
- Model Optimization: **Hyperparameter Tuning, Feature Engineering, Data Drift Detection**
- MLOps Frameworks: **MLFlow, Azure MLOps, AWS SageMaker**
- CI/CD & Automation: **Azure DevOps, Git, Jenkins, Docker, Kubernetes**

Work Experience

Senior Data Scientist | ING Bank Sydney via TCS (*Jan 2023 – Present*)

Tech Stack: Azure Synapse, MLFlow, Azure DevOps, Azure Data Lake, Azure MLStudio

- Pioneered the Early Warning System (EWS), a credit risk model identifying clients with a high probability of overdue payments, resulting in a 15% reduction in non-performing loan rates.
- Analyzed and optimized risk models across personal loans, credit cards, and residential mortgages (2 models for each portfolio), enhancing portfolio performance by 24%.
- Implemented scalable ETL pipelines in Azure Synapse Analytics, processing 3TB+ data with 99.9% accuracy.
- Optimized model performance (92% accuracy) via feature selection, hyperparameter tuning, and advanced ML techniques.
- Integrated 6 statistical models into existing banking production systems using Azure Machine Learning Service, improving fraud detection rates by 15% and reducing false positives using explainable AI techniques.
- Introduced an end-to-end MLOps pipeline for automating model retraining, deployment (CI/CD pipelines), and monitoring on Azure, reducing downtime by 40%.
- Tracked and stored model metadata with MLflow, logged key metrics, and created Power BI dashboards, improving model monitoring efficiency by 30%.

AI Engineer | Telstra via TCS (July 2023 – Dec 2023)

Tech Stack: LangChain Framework, Azure OpenAI LLMs, Azure SQL Database, Azure Function App, Power Virtual Agents, Python, SQL,deep learning, generative AI

- Introduced an AI-powered SQL query system using natural language, improving query efficiency by **30%**.
- Engineered system architecture integrating LangChain with Azure SQL, fine-tuned NLP models using Azure OpenAI for accurate SQL queries.
- Championed the development of a natural language chatbot interface, allowing users to effortlessly query the database; tool is now used by over 50 employees across data science teams.

- Led cross-functional teams to integrate ML models into production environments.
- Achieved a 40% reduction in query response time by optimizing NLP models and enhancing database indexing.
- Enhanced user satisfaction by delivering accurate and relevant query results, which significantly boosted user engagement.

Data Scientist | Telstra via TCS (August 2022 – June 2023)
Tech Stack: Azure MLOps, Azure Machine Learning Studio, Azure IOT Hub, Azure Synapse Analytics, Python, Azure DevOps.

- Designed and led MLOps-Cockpit, an end-to-end ML lifecycle platform optimizing data ingestion, model deployment, and real-time monitoring.
- Implemented automated CI/CD/CT frameworks, reducing ML deployment time by 40%.
- Integrated data drift detection and mitigation techniques to ensure model stability and performance over time.
- Redesigned centralized Power BI/Tableau dashboards for real-time model performance monitoring
- Simplified A/B testing framework to compare model variations, optimizing decisions based on performance metrics.
- Led and implemented MLOps solutions, streamlining model operationalization and improving deployment and maintenance efficiency by **30%**.
- Link: mbu-dataai.mlops_cockpit

Data Scientist | TCS (Sep 2021 – July 2022)
Tech Stack: Azure Cognitive Services - OCR, Azure Machine Learning Studio, Azure Databricks, Python, Text extraction, Random Forest, Decision Tree, GridSearchCV.

- Pioneered an ML model to forecast sales by analyzing historical data and uncovering key patterns.
- Conducted analysis on 1 million data points to uncover key insights.
- Executed EDA and produced time series models to project future sales for stores.

Data Scientist | TCS (July 2020 – August 2021)
Tech Stack: Azure Synapse Analytics, Azure Logic Apps, Azure Form Recognizer, Azure Cosmos DB, Power BI, Azure App Service, Azure Machine Learning, Azure Data Lake Storage Gen2, Azure Function App, Azure Cognitive Search.

- Developed an intelligent underwriting workbench using ML/AI to improve business efficiency, predicting anomalies and analyzing root causes, boosting performance by **25%**.
- Developed a predictive model to assess the need for physical exams before insurance policies, improving decision accuracy by using correlation analysis.
- Ingested Azure data, built Power BI dashboards, and trained and tested models in the production environment.
- Used Logistic Regression, Random Forest, Decision Tree, XGBoost, GridSearchCV.
- Link: dataai.modern_underwriting_on_azure

Python Developer | TCS (Jan 2017 – June 2020)
Tech Stack: Python, SQL, Flask, Pandas, NumPy, Git.

- Assisted in the development of scalable web applications using Python and Flask, ensuring adherence to best practices and clean code standards.
- Worked with SQL databases to create efficient queries, improve database performance, and manage data pipelines.
- Participated in the design and implementation of a data processing pipeline for analyzing large datasets, improving the speed of data extraction.
- Collaborated with senior developers to troubleshoot and debug issues.

Education

Bachelor of Technology (Mechanical Engineering) | Kurukshetra University, India (*2016*)

Certificates and Training

- Microsoft Certified: Azure Data Scientist Associate (DP-100)
- Completed Data Scientist Nanodegree on Udacity.
- Certified in Python 3 Essentials on Udemy.
- Certified in Machine Learning on Udemy.
- Completed Agile methodology in TCS.