Oscin Jaypal

Data Scientist

Jersey City, NJ| +1 (551)-339-6397 |oscin.jaypal@myemailjob.com| GitHub | LinkedIn

SUMMARY

- Highly skilled Data Scientist with 4+ years of experience in data extraction, modeling, wrangling, statistical analysis, and machine learning.
- Skilled in machine learning, deep learning, statistical modeling, NLP, and data engineering, with a strong track record of building scalable models and driving actionable insights.
- Proficient in developing robust pipelines using Python, SQL, and Spark, and deploying models using cloud platforms such as AWS and GCP.
- Proficient in SDLC, Agile, and Waterfall methodologies for effective project management.
- Experienced in engineering and deploying Generative AI models on multi-GPU infrastructure, utilizing expertise in model fine-tuning, optimization, and distributed computing to drive AI innovation and scalability.
- Strong in data visualization tools like Tableau, Power BI, and Excel, creating insightful visualizations for effective communication.

SKILLS

Language: Python, R, SQL, SAS
Methodologies: SDLC, Agile, Waterfall
IDEs: Visual Studio Code, PyCharm, Jupyter Notebook
Statistical Methods: Hypothesis Testing, ANOVA, Time Series Analysis
Machine Learning: Regression analysis, Bayesian Method, Decision Tree, Random Forests, Support Vector Machine, Neural Network, Sentiment Analysis, K-Means, KNN, Classification, Naive Bayes, Natural Language Processing (NLP), LLM, CNN, XGBoost
Frameworks: LangChain, Retrieval-Augmented Generation (RAG)
Packages: NumPy, Pandas, Matplotlib, SciPy, ggplot2, scikit-learn, PyTorch, TensorFlow, Keras, Spark
Visualization Tools: Tableau, Power BI, Microsoft Excel
Cloud Technologies: AWS, GCP, Azure
Database: MySQL, SQL Server, Oracle, MongoDB, Vector Database (Pinecone, ChromaDB), Graph Database (Neo4j)
Software/Other Skills: Jira, Data Cleaning, Data Wrangling, Critical Thinking, Communication Skills, Problem-solving, Decision-Making, EDA, Databricks, Data Visualization, Predictive Analytics, Pattern Recognition, JMP, Data Integrity, Quantitative Data, Data Science, Statistics, Statistical Analysis, Data Analytics, Data Modeling, BigQuery, Snowflake
Operating System: Windows, Linux

EXPERIENCE

S&P Global, USA | Data Scientist | Jun 2024 - Present

- Employed Agile methodologies to adapt to evolving project scopes, reducing iteration cycles by 20% and delivering solutions ahead of schedule.
- Conducted hypothesis tests to evaluate variable relationships, improving decision-making accuracy by 15% in high-impact models.
- Led cross-functional collaboration across product, engineering, and compliance teams, aligning data science strategy with regulatory standards and reducing risk exposure by 30%.
- Delivered actionable insights to executives via dynamic Power BI dashboards, increasing stakeholder reporting efficiency by 40% and establishing a new benchmark for credit portfolio monitoring.
- Optimized MySQL database schema, leading to a 25% reduction in query response times and improved system scalability.
- Spearheaded the development of credit risk and fraud detection models using XGBoost and LightGBM, resulting in a 22% increase in prediction accuracy and $4.3M+ in annual fraud loss prevention.
- Designed and deployed real-time inference pipelines on AWS (S3, EC2, Lambda, SageMaker), reducing batch latency by 60%.
- Created elegant visualizations using ggplot2 in R, based on the grammar of graphics principles.
- Implemented a Retrieval-Augmented Generation (RAG) pipeline to enhance the accuracy and relevance of automated responses by combining generative and retrieval-based AI techniques, increasing solution accuracy by 30%.
- Architected and automated ML model lifecycle management using MLflow, Git, and CI/CD pipelines, reducing deployment time by 45% and increasing experiment traceability.

Cybage Software, India | Data Scientist | Jan 2019 - Jul 2022

- Implemented time series and supervised/unsupervised learning algorithms, enabling predictive analytics that improved operational efficiency by 23%.
- Integrated Databricks solutions into pipelines, decreasing processing time by 35% through streamlined data workflows.
- Collaborated with data engineering to migrate workflows to Google BigQuery, improving query speed by 45% and reducing ETL costs by 20%.
- Built and orchestrated ETL pipelines with Azure Data Factory, automating ingestion and transformation across multiple sources, increasing data freshness by 50%.
- Implemented backup and recovery for SQL Server databases, achieving zero data loss incidents and 100% uptime during scheduled failovers.
- Developed predictive models using logistic regression, random forest, and gradient boosting to forecast medication adherence, leading to a 22% reduction in non-compliance and a measurable increase in clinical outcomes.
- Engineered NLP pipelines using NLTK to analyze 500K+ support call transcripts and prescription notes resulting in a 17% decrease in

repeated support calls by surfacing actionable trends.

- Built end-to-end data pipelines in PySpark and SQL to automate ingestion, validation, and aggregation of pharmacy and patient data from multiple sources, enhancing data readiness by 65%.
- Conducted root cause analyses for high-value churn segments using unsupervised learning (K-Means, PCA), influencing a new patient retention strategy that increased LTV by 10%.
- Created robust Tableau dashboards to visualize prescription trends, care gaps, and cost-saving opportunities viewed weekly by over 50 stakeholders.
- Managed secure data storage and access through Azure Data Lake Storage Gen2, applying RBAC and data encryption for compliance with enterprise standards.

EDUCATION

Master of Science in Computer and Information Science
New York Institute of Technology, Manhattan, NY

Bachelor of Engineering in Computer Engineering
Gujarat Technological University, India

CERTIFICATIONS

- **Microsoft Certified: Azure Data Fundamentals (DP-900)**
- **Microsoft Certified: Azure Fundamentals (AZ-900)**
- **Microsoft Certified: Azure AI Fundamentals (AI-900)**
- **Microsoft Certified: Azure Data Engineer Associate (DP-600)**

PROJECTS

Software Defect Prediction Analysis

- Engineered a machine learning model to predict software defects, improving release quality and reducing post-deployment bugs by 35%.
- Conducted statistical evaluations and trend analysis, enhancing test coverage and cutting defect discovery time by 20%.

Twitter Dataset Analysis

- Executed sentiment analysis and user segmentation on Twitter data, surfacing key behavioral clusters that improved campaign targeting precision by 18%.
- Employed Python (pandas, NLTK, scikit-learn) for data preprocessing and visualization, accelerating reporting cycles by 30%.

Image Caption Generator

- Developed an AI-driven image captioning system integrating CNNs and RNNs, increasing caption accuracy by 28% on validation datasets.
- Utilized TensorFlow, Keras, and LSTM to generate descriptive captions, improving contextual relevance by 20% based on BLEU score evaluations.

Temperature Screening System with Facial Recognition

- Built a temperature screening framework using OpenCV and deep learning, achieving 94% facial recognition accuracy in real-time conditions.
- Integrated automated email alerts for elevated temperature detections, reducing manual monitoring workload by 70% and enhancing compliance.