Saksham Jain

Data Scientist

saksham.j@worksemails.com | 201-644-6274 | Lawrence Township, NJ | LinkedIn | GitHub

SUMMARY

- Data Scientist with around 4 years of experience in Data Extraction, Data Modeling, Data Wrangling, Data Mining, Machine Learning, Statistical Modeling, Natural Language Processing and Data Visualization.
- Experienced the full software life cycle in SDLC, Agile, and Scrum methodologies, including creating requirements and test plans.
- Experienced with machine learning algorithms such as Linear and Logistic regression, Ensemble methods (Random Forests), XGBoost, KNN, SVM, Deep Neural Network, Lasso Regression, K-Means, Decision Trees, Naive Bayes, K-NN and more.
- Capable translate complex data insights into actionable business recommendations using tools like Tableau, Power BI, or similar.
- Expertise in designing and analyzing A/B tests to validate hypotheses and improve products or processes.
- Familiarity with distributed computing technologies like Hadoop and Spark and cloud platforms like AWS, GCP, or Azure.
- Hands-on working experience with multiple data sources, relational and non-relational, such as Oracle, MySQL, SQLite, SQL Server, Snowflake, MongoDB.

SKILLS

Methodologies: SDLC, Agile, Waterfall
Language: Python, R, SQL, SAS
IDEs: Visual Studio Code, PyCharm, Jupiter Notebook
Statistical Methods: Hypothetical Testing, ANOVA, Time Series
Machine Learning: Regression analysis, Bayesian Method, Decision Tree, Random Forests, Support Vector Machine, Neural Network, Sentiment Analysis, K-Means Clustering, KNN, Classification, SVM, Naive Bayes, NLP, LLM, CNN, XGBoost, Deep Learning
Packages: NumPy, Pandas, Matplotlib, SciPy, ggplot2, Scikit-Learn, PyTorch, TensorFlow, Keras, Spark
Visualization/Other Tools: Tableau, Power BI, Jira, Microsoft Excel
Cloud Technologies/Database: AWS, GCP, Azure, MySQL, SQL Server, Oracle, MongoDB
Software/Other Skills: Data Cleaning, Data Wrangling, Critical Thinking, Presentation Skills, Problem-solving, Decision-Making, EDA, Communication Skills, Databricks, Data Visualization, Predictive Analytics, Pattern Recognition, JMP, Data Integrity, Quantitative Data, Data Science, Statistics, Statistical Analysis, Data Analytics, Data Modeling, Big Query, Snowflake, Data Analysis, Data Mining, SAP, Mathematics, Computer Science, Programming
Operating System: Windows, Linux

EXPERIENCE

Molina Healthcare, USA | Data Scientist **Jan 2024 - Current**

- Contributed to Agile ceremonies, including sprint planning, reviews, and retrospectives to drive iterative development.
- Conducted A/B tests for product strategies, leading to a 20% increase in conversions and a 10% reduction in costs.
- Implemented predictive analytics solutions using regression models, decision trees, and SVMs, optimizing classifiers through grid search and k-fold cross-validation, enhancing model accuracy and performance.
- Automated predictive modeling processes using AWS, reducing manual effort by 20% and enhancing data infrastructure.
- Applied supervised and unsupervised learning algorithms, including LLM, to uncover hidden patterns within large datasets.
- Crafted complex SQL queries in Teradata to extract and prepare datasets for Tableau dashboard visualizations.
- Built advanced machine learning classification models like XGBoost, and clustering algorithms, such as hierarchical clustering.
- Collaborated with database administrators to optimize MySQL database performance, tuning query execution plans.

Groovy Web, India | Data Scientist **Jan 2020 - Dec 2022**

- Managed testing and validation of data science models, ensuring quality outputs through the Waterfall methodology.
- Conducted ANOVA tests with Python to analyze differences among hotel clusters, driving data-driven decision-making.
- Built NLP models to analyze customer feedback, revealing key themes and sentiment trends for product improvements.
- Deployed AI-driven tools for automated EDA, providing insights from raw data in minutes and cutting analysis time by 40%.
- Created statistical algorithms in Python and SQL to assess welfare dependency risks using Multivariate Regression, Logistic Regression, Decision Trees, and Support Vector Machines.
- Worked with cross-functional teams, increasing data-driven solution delivery by 25% and reducing misalignment by 30%.
- Designed and trained deep learning models with PyTorch and TensorFlow, optimizing architectures for better performance.
- Implemented predictive pricing algorithms, boosting revenue by 12% and enhancing customer satisfaction by 15%.
- Integrated Power BI with MS Excel and other data sources to automate real-time report generation and updates.
- Utilized AWS EC2 instances for scalable data processing, model training, and inference, efficiently managing variable workloads.
- Established backup and recovery strategies for SQL Server databases to ensure data availability during system failures.

EDUCATION

Master in Data Science | New Jersey Institute of Technology, Newark, NJ
Bachelor in Information and Communication Technology | Gujarat Technological University, India

CERTIFICATIONS

MATLAB Onramp, Machine Learning Onramp & Deep Learning Onramp by MathWorks Company.
Python 3 Specialization and Python for Everybody Specialization by Michigan University.
SQL for Data Science by the University of California.