

Data Scientist | Location: Bangalore

Position Summary:

Responsible for building algorithms based on statistical modelling procedures and build, deploy, and maintain scalable machine learning algorithms in production, while implementing prototypes and proofs-of-concept. Evaluating and using various data modelling and evaluation strategies for training and applying machine learning algorithms and libraries models for finding patterns to predict instances. Responsible for leading software engineering and design process improvements, while conducting performance analysis, troubleshooting, and corrective/perfective maintenance.

Responsibilities:

- Design, develop, and deploy machine learning models to production
- Develop and support a platform that enables data scientists to rapidly develop, train, and experiment with different machine learning models
- Expand and optimize data pipelines, data flow, and collection for cross-functional teams
- Create and maintain optimal data pipeline architecture by assembling large, complex data sets to meet functional and non-functional business requirements
- Create and implement process improvements for automating manual processes, optimizing data and infrastructure for scalability and performance
- Support building of machine learning, data platforms, and infrastructure required for optimal data extraction, transformations, and loading of data from a wide variety of data sources
- Work with other stakeholders to support data infrastructure needs
- Implement Machine Learning and Big Data platforms in Hybrid and multi-cloud environment (AWS experience preferred)
- Contribute to technical discussions to improve overall software processes and practices, software quality, cadence and delivery lifecycle through automation and CI/CD pipelines using DevOps methodologies
- Prototyping and coding to evaluate multiple approaches
- Deploy and troubleshooting in dev, testing, and production environments.
- Create functional test cases, build test automation and test harness for ML functions and algorithms

Qualifications & Requirements:

- Bachelor's degree (Master's preferred) in Computer Science, Engineering, Mathematics, or a related field with strong fundamentals on algorithms, data structures, multithreading, object-oriented design & development, distributed applications, client-server architecture
- Strong coding skills in Python using scientific libraries like numpy or scipy.
- 2-3 years of experience required with one or more deep learning frameworks such as PyTorch, Tensorflow, or Keras
- 2-3 years of experience in building and deploying Machine Learning Models to production
- Experience with training deep neural networks on large-scale datasets
- Hands-on experience with Time Series Forecasting, and beyond statistical algorithms like ARIMA, preferred
- Hands-on experience with Sequence Models in Deep Learning, like LSTMs, etc. especially in the context of Time Series Forecasting, preferred
- Deep knowledge of machine learning concepts and their application, including traditional approaches to both supervised and unsupervised learning as well as deep learning networks and reinforcement learning
- Deep understanding of demand forecasting techniques including forecast errors, time-series, statistics, ML models, clustering, building and refining training datasets
- Good understanding of mathematics, statistics, and algorithms.
- Ability to present technical details of machine learning algorithms to stakeholders
- Ability to develop and implement novel approaches to machine learning
- Ability to process and manipulate structured and unstructured data before ML modeling
- Experience with database technologies, data warehousing, and hands-on experience writing SQL queries
- Experience with cloud service, specifically AWS
- Extensive knowledge of ML frameworks, libraries, data structures, data modeling, and software architecture
- Experience in container, streaming, and messaging technologies
- Excellent analytical, problem-solving, and interpersonal skills