ROCKET
SCIENCE

Necessary profiles to build your Data Science team

Now that we know which the priority projects are. We have scoped our projects, and we can move on to building the team and get talents. The data science world is complex and there are more than just data scientists out there in the market. Many different profiles are available, and a proper team is composed of complementary profiles from different backgrounds.

Your ability to identify the necessary profiles and understand what skills are needed to complete the projects you have is very important. An R&D project doesn't require the same profiles as business-driven projects. Additionally, deploying models and managing machine learning life cycle require special skills. We will start by identifying the different roles available in the market.

>>>>>>>>

The different roles possible in data science

Title	Role	Skills	Tools
Data analyst	Analyze data and deal with data requests from internal customers.	Business analysis / Data engineering (SQL mainly)/ Reporting/ Statistics	SQL/ Tableau / Excel / PowerBI / R / SPSS
Business analyst	Understand the business requirement for every project. Deal with requests from internal customers.	Business analysis / Reporting / Communication	SQL/ Powerpoint / Excel / PowerBI / Tableau
Data Scientist	Develop machine learning models. Perform data analysis if required.	Data Science / Statistics / Machine Learning / Mathematics / Reporting	Python / R / SQL / Jupyter Notebooks / Tensorflow / Pytorch / Git
Research Scientist	Search for an innovative approach to solve machine learning problems.	Data Science / Machine Learning / Mathematics / Statistics	Python / C++ / C / R / SQL /Notebooks / Tensorflow / Pytorch
Machine Learning Engineer	Create machine learning solutions, from the development of models to deployment.	Data Science / Machine Learning / Statistics / Data engineering / Model deployment / Microservices / Containers	Python / R / SQL / Tensorflow / Pytorch / Git / Cloud providers / Docker / Kubernetes / Go / C# / C ++
MLOps Engineer/ ML Software Engineer	Manage machine learning life cycle and create solutions and maintain them in production.	Data engineering / Model deployment / Microservices / Containers / Software engineering / Infrastructure / ML pipelines	Python / Git / Cloud providers / Docker / Kubernetes / Go / C# / C ++ / CI-CD / Microservices architecture / SQL
Data Engineer	Make data available to data scientists/Machine Learning engineers. Make data ready for production	Data engineering / ETL / Data Lake / Big data / Reporting	Python / ETL Tools / Hadoop

The tasks in data science life cycle and associated roles

Task	Typical day-to-day	Associated role
Understanding customers need	- Meet customers and discuss their projects - Conduct workshops and design sessions to understand where the valuable projects are.	Business analyst Lead data scientist Director data science
Scope a project and define requirements	- Evaluate project feasibility - Understand necessary resources - Validate the potential of the project - Evaluate effort	Business analyst Lead data scientist Director data science
Collecting data	- Create pipelines for ingestion of the data - Create a data warehouse - Create databases (data marts) - Create a data lake	Data engineer Software engineer
Cleaning data	- ETL scripts/flow - Scripting - Validate transformation	Data engineer Data scientist Data analyst
Labeling data	- Create a training dataset - Use statistics to estimate labels - Use open data	Data scientist Research scientist
Exploring the data	- Visualize the data - Understand the distribution of the data - Correlation - Univariate/Multivariate analysis	Data analyst Data scientist Research scientist
Reporting on data and model performance	- Reports and dashboards creation about data quality - Reports and dashboards creation about data sources - Reports and dashboards creation about model performance	Data analyst MLOps engineer (for guidance about model performance)
Training machine learning models	- Train models - Data preprocessing - Model selection - Model evaluation	Data scientist Research scientist Machine Learning Engineer

Task	Typical day-to-day	Associated role
Optimizing training phase	- Running parallel computing - Optimizing hardware - Running on GPU	MLOps engineer Machine Learning Engineer
Creating new machine learning algorithms	- Develop a new machine learning algorithm - Research for new ML approach	Research scientist
Researching new technics in machine learning	- Reading research paper - Attending conferences	Research scientist Data scientist
Refactoring machine learning code	- Making machine learning code ready for production - Optimizing code - Rewriting code	Machine Learning Engineer MLOps engineer
Deploying machine learning models	- Deploy ML models as microservice - Deploying ML models to model registry - Creating docker images - Deploying scripts and applications	Machine Learning Engineer MLOps engineer
Creating retraining pipeline	- Create automated pipelines for retraining models	Machine Learning Engineer MLOps engineer
Managing machine learning environment	- Creating development environment - Managing deployment environment - Training and coaching team on how to use tools (MLOps tools)	MLOps engineer
Defining production requirements for machine learning models	- Defining best practices in terms of deployment - Meeting data scientists to understand what requirements for ML models	MLOps engineer Machine Learning Engineer
Setup environments (development or production)	- Create environment for development and for production	MLOps engineer
Create a CI/CD environment Define retraining strategies	- Creating Continuous integration pipelines and tools - Creating Continuous delivery pipelines and tools - Define best practices for retraining	MLOps engineer Machine Learning Engineer
Setup ML infrastructure	- Setup hardware (resources) for ML projects	MLOps engineer

Looking for support in your journey?

GET IN TOUCH NOW