Quanticate
The Clinical Data Experts

Global Solutions from the World's Largest Data-Focused CRO

Data Science in Pharmaceutical Industry - Introduction

Ganpatlal Suthar and Shveta Natu

Disclaimer

This presentation does not represent views of Quanticate International and belongs solely to the author. Authors take full responsibility for the work and errors if any.

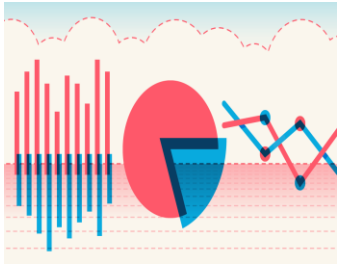
Agenda

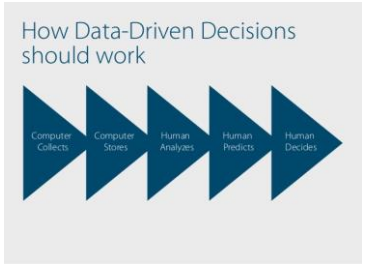
- What is Data Science
- What is Machine Learning
- Types of Machine Learning (ML)
- Unsupervised Learning
  - Example of Unsupervised Learning
- Supervised Learning
  - Example of Supervised learning
- Reinforcement Learning
  - Example of Supervised and Reinforcement learning
- SAS Data Mapper Tool : ML Algorithm
- Challenges of Data Science in Pharma Industry
- Regulatory Processes for Approvals of AI/ML Products
- Summary
- Q/A

What is Data Science

- Data science is a big umbrella covering each and every aspect of data processing and not only statistical or algorithmic aspects.

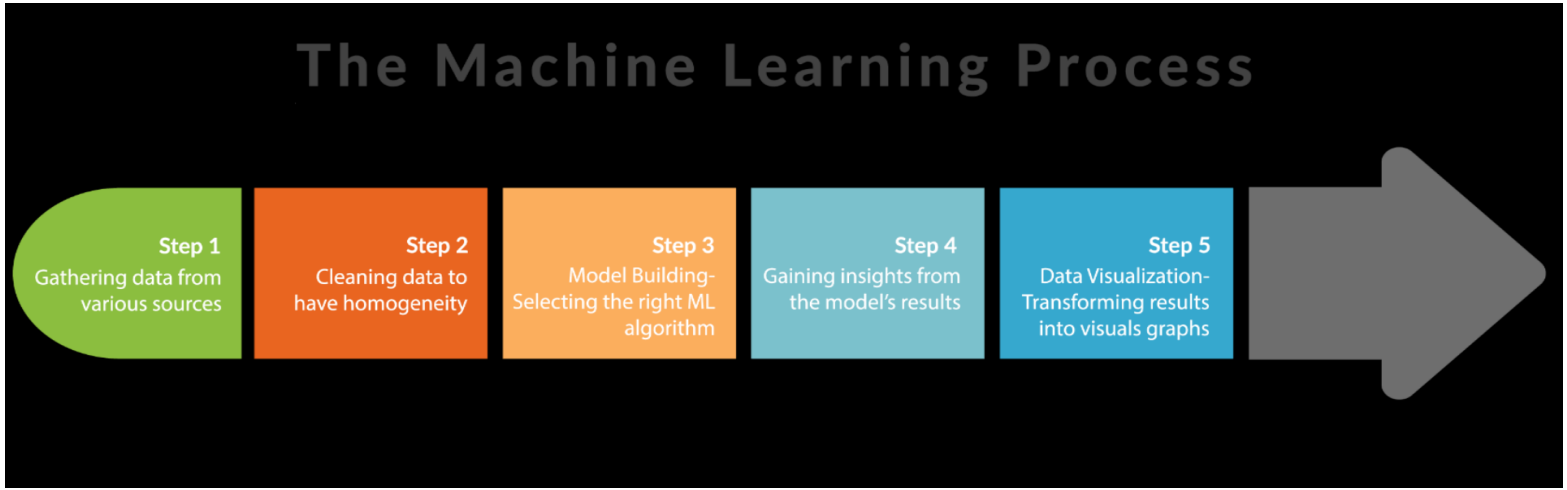
- To mention, data science includes
  - data visualization
  - data integration
  - dashboards and BI
  - distributed architecture
  - data-driven decisions
  - **Automation using ML**
  - data engineering





What is Machine Learning

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
- It focuses on the development of computer programs that can access data and use it learn for themselves.



Types of Machine Learning

Quanticate
The Clinical Data Experts

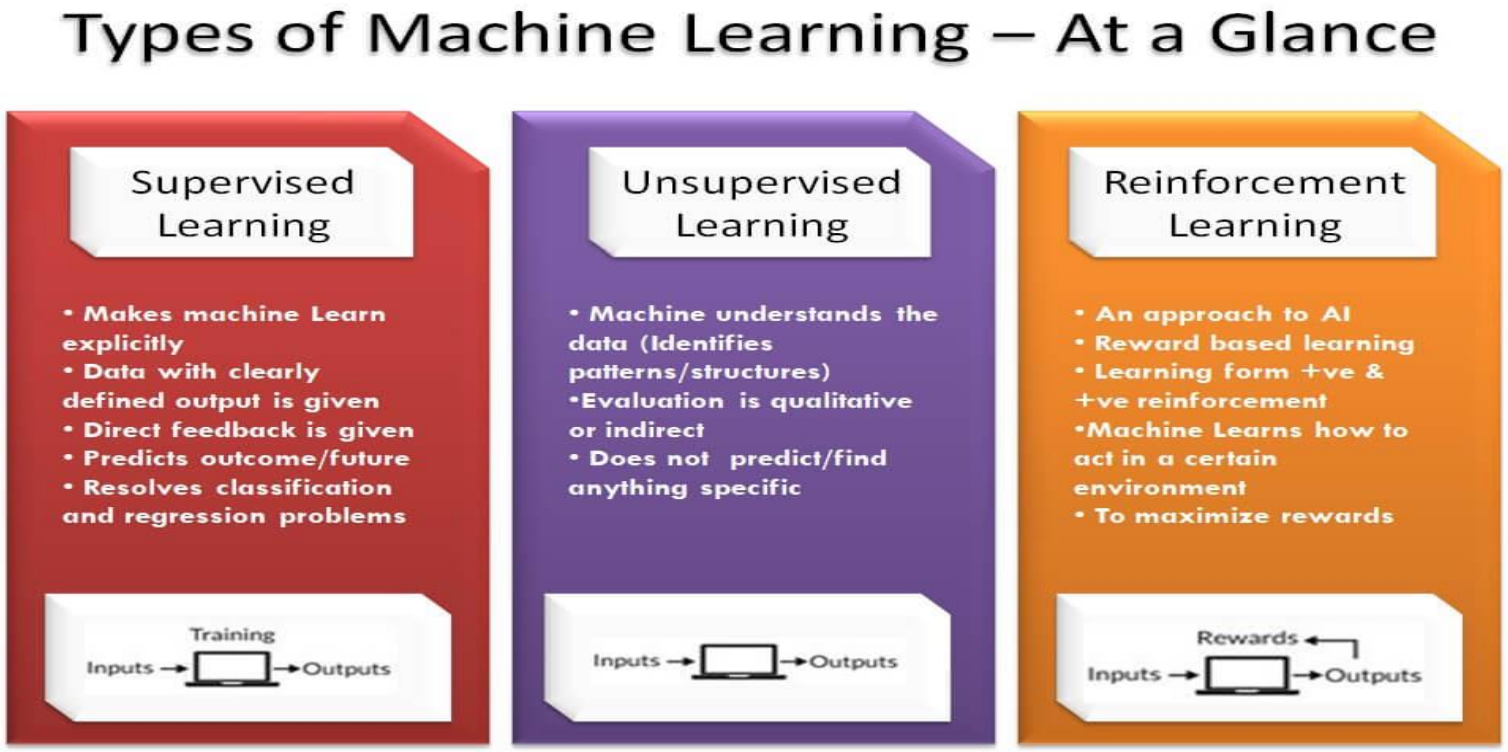
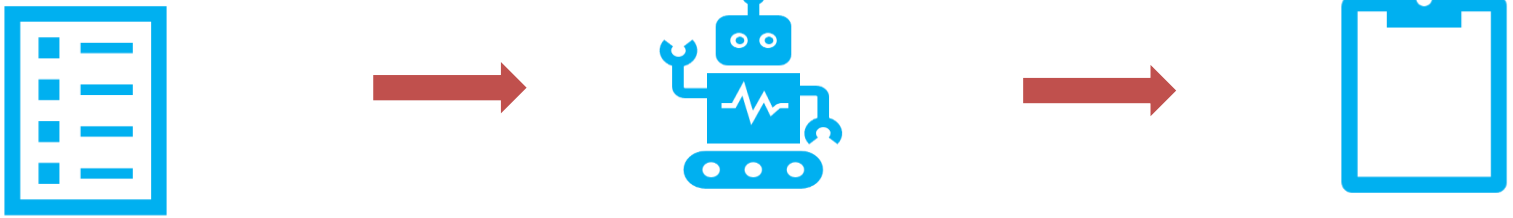


Illustration: Unsupervised Learning

- Spread of Zika Virus



Step1
Data is entered of patients suffering from Zika Virus from various locations in India

Step2
Machine algorithm analyzes the data and clusters the data based on coastal region patients and inland region patients

Step3
Based on the clustering density, awareness campaigns can be launched in the concerned regions

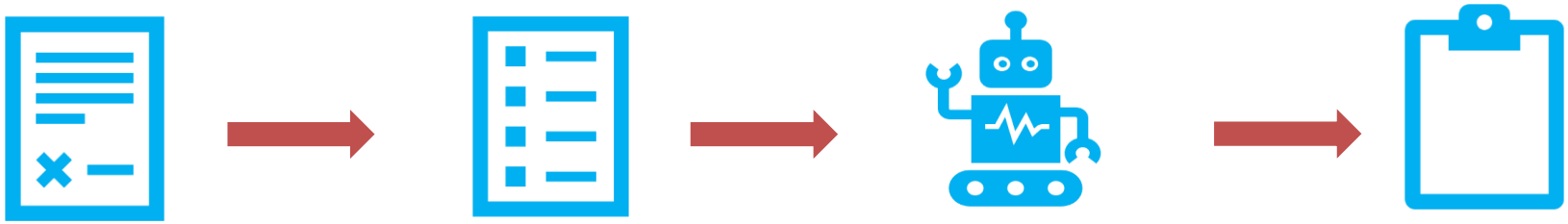
Example: Unsupervised Learning

Quanticate
The Clinical Data Experts

- Test A (95% Accurate, 10x costlier than blood tests)
- Problem: To find an alternative lab tests, which will help us in reducing the patients going for Test A.
- Process:
  - Step 1: We will feed the past data (Test A and laboratory results) into the platform
  - Step 2: We run the algorithms, to form the clusters.
- The Clusters formed will help in identifying abnormal results, which further helps in selecting which patients will go for Test A. This will help reduce the diagnostic cost.

Illustration: Supervised learning

- Lab tests for Anemia



Step1
Machine is trained about Hb level and corresponding output of either anemic or non anemic based on labelled data

Step2
When data for patients is fed into machine for Hb

Step3
Machine algorithm analyzes the patient's data with step1 Inputs

Step4
Machine recognizes the Hb level and generates report if patient is suffering from anemia or not

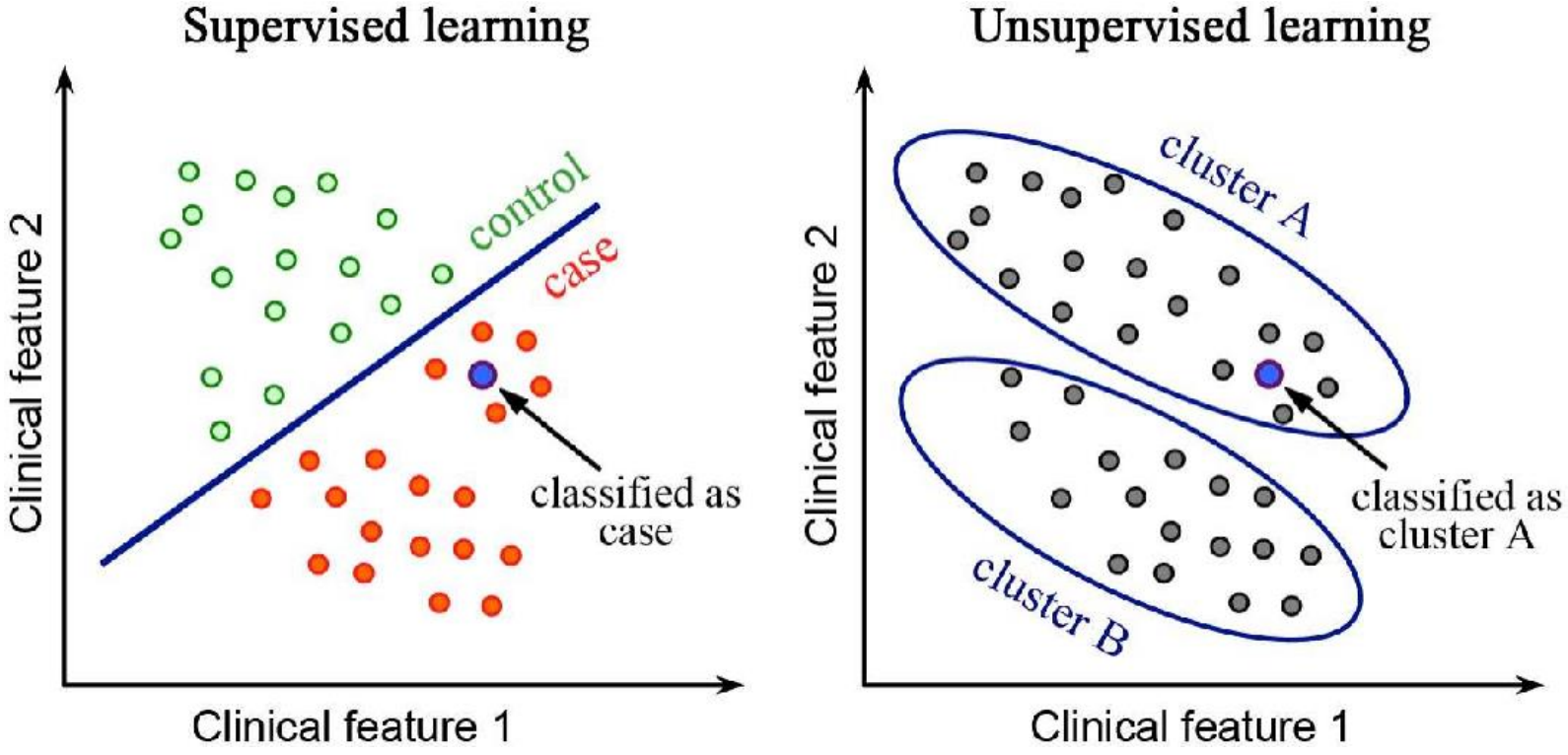
Example: Supervised Learning

Quanticate
The Clinical Data Experts

- Company X claims it leverages NLP to help researchers more efficiently manage clinical trial workflows.
- Objective: To identify risk factors and provide recommendations for clinical trial optimization.
- Study Optimizer platform are trained on “billions of data points from past clinical trials, medical journals, and real-world sources”.
- Process:
  - Step 1: A user uploading his research protocol documents to the platform.
  - Step 2: Risk factors are identified by analysing the protocol text, and any potential barriers are reported.
  - Step 3: The platform provides recommendations for mitigating risk to optimize the research protocol.

Supervised / Unsupervised Learning

Quanticate
The Clinical Data Experts



Reinforcement Learning

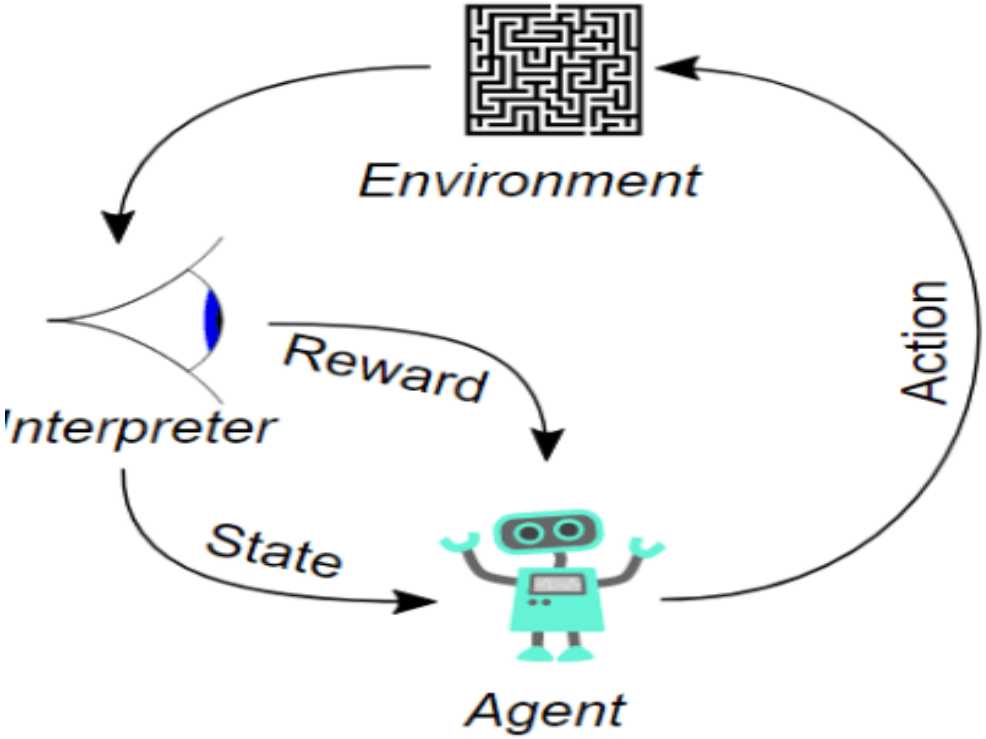


Illustration: Reinforcement Learning

- Diagnosis based on X-ray

Step1
Images of x-rays correctly labelled as “Diagnosed with Disease X” and “Normal” are loaded in the system. Machine Learning builds an algorithm on which it is trained.

Step2
Now new x-ray image is loaded on the system and based on past learning, model predicts the condition of the patient.

Step3
Simultaneously doctor also diagnosis the patient condition by taking a look at same x ray and giving a feedback on “Correctly diagnosed by ML” or “Incorrectly diagnosed by ML”.

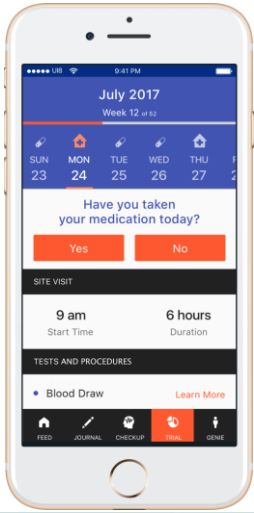
Step4
This feedback makes the algorithm better for the future diagnosis.

Example: Supervised and Reinforcement Learning

- Company Y claims that it leverages machine learning to better manage patient engagement in clinical trials.

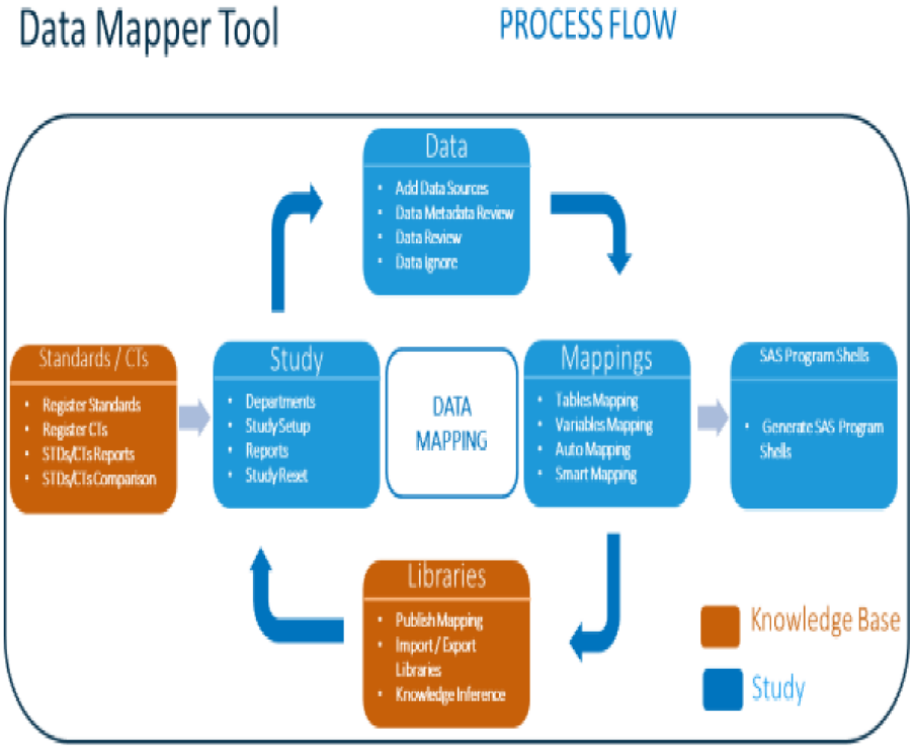
- The Company's patient app and site dashboard are trained on millions of clinical data points.

- Advantages:
  - The recommendation engine is able to identify key markers that tend to correlate with patient disengagement from research studies.
  - The system sends notifications to the user related to their scheduled tasks and site visits.
  - Engagement with the app through curated content and personalized communication.
  - A conversational chatbot customized with approved study materials is available to patients.



SAS Data Mapper Tool: ML Algorithms

- Standards / Controlled Terminologies – Provides ability to register Standards like – SDTM, ADaM or company specific standards and CTs
- Studies – Provides ability to register different studies and control permissions.
- Data – Provides ability to capture Data for studies from different sources
- Mapping – Provides ability to map source to destination data
- Generate SAS Programs – Provides ability to generate SAS programs based on mapping metadata
- Libraries – Provides ability to capture mapping metadata into different libraries.



SAS Data Mapper Tool: ML Algorithms

- Since mapping metadata is captured into Libraries, different type of Machine Learning algorithms can be applied to learn information about existing mapping and these algorithms can be trained to help predict mapping for new source data.
- Machine algorithms can be applied to different type of metadata captured at Dataset, Variable and Value level. Below screenshot represent model Similarity vs NGram similarity for Tables Mapping

```
Predictions: ['ae' 'cm' 'lb' 'fa' 'eg' 'ie'] Expected: ['AE', 'CM', 'LB', 'FA', 'EG', 'IE']
             Model_Matched_Term  Model_Similarity NGram_Matched_Term  NGram_Similarity
Search_Term
adverse2                     ae          0.717276                 ae          0.583333
chemistry                    lb          0.515414                 lb          1.000000
conmed                       cm          0.703553                 cm          1.000000
electrocardiac               eg          0.699650                 eg          0.521739
follow                       fa          0.683542                 fa          1.000000
inclusion                    ie          0.537428                 ie          1.000000
```

Challenges in Pharma Industry to Implement Data Science

Quanticate
The Clinical Data Experts

- Lack of Data Standards
  - Disparity of Data Sources
  - Ambiguity Around Accuracy along with time based relevance
  - Time-consuming Analytics Process
- Anxiety over change holds back progress
  - Employee concern over potential job loss
- Skills shortage hits data science
  - Shortage of adequately qualified workers with the right technical skills

Regulatory Processes for Approvals of AI/ML Products

- The emerging use of AI in drug discovery is prompting important questions about how AI, and AI-derived innovations should be regulated. Another concern is the patentability of AI and AI-derived innovations.

- European Patent Office's (EPO) has published a draft of its updated guidelines on patenting, which include a new section devoted to AI.

- U.S. Food and Drug Administration is actively developing regulatory framework to promote innovation in artificial intelligence for healthcare.
  - Apr2018 - FDA permitted marketing of artificial intelligence-based device to detect diabetes-related eye problems
  - May2018 - FDA permitted marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures
  - Jan2019 - FDA started PreCert version 1.0 pilot program for more streamline and efficient regulatory oversight of Software as a Medical Devices (SaMD) under FDA's Digital Health Innovation Action Plan.

Summary

For any queries you can reach out to us

Ganpatlal.Suthar@quanticate.com

Shveta.Natu@quanticate.com