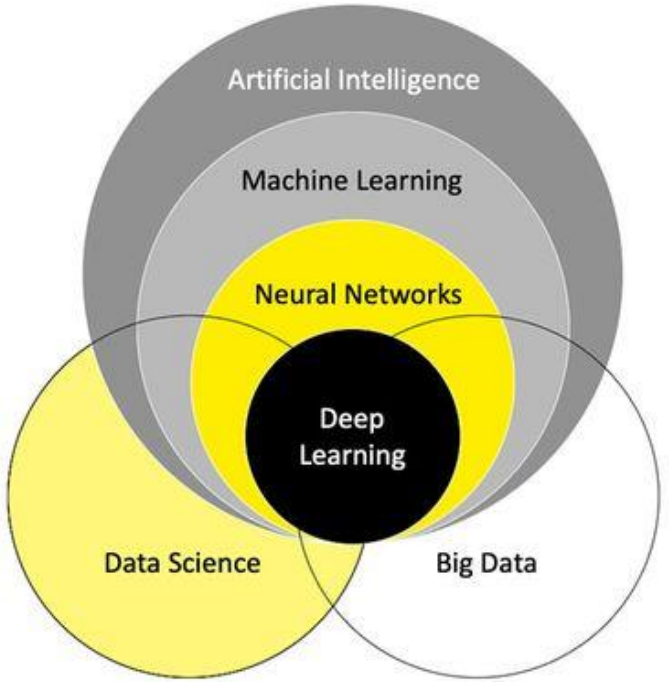
Data Science Process Lifecycle

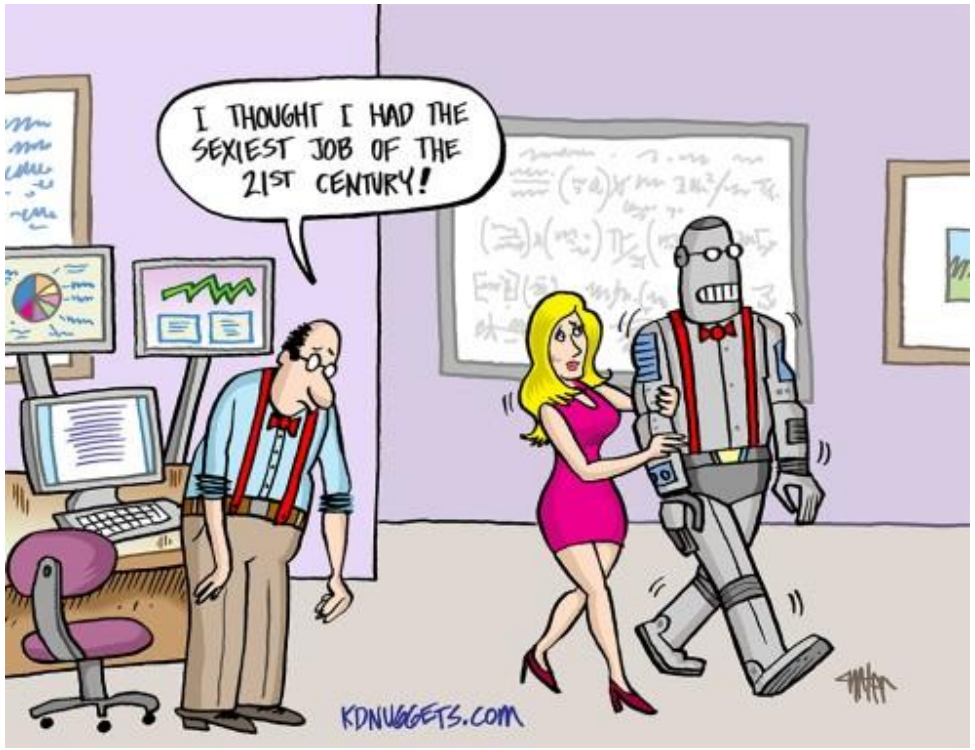
Role of Data in AI

Data science is a field that incorporates some areas of AI, machine learning and deep learning, while having a specific focus of gaining insight from data.

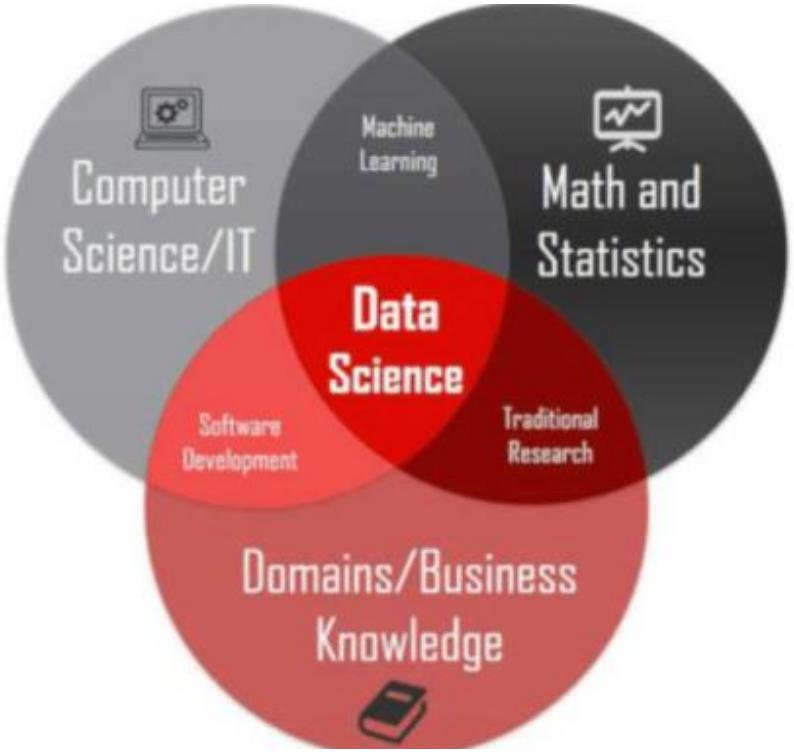


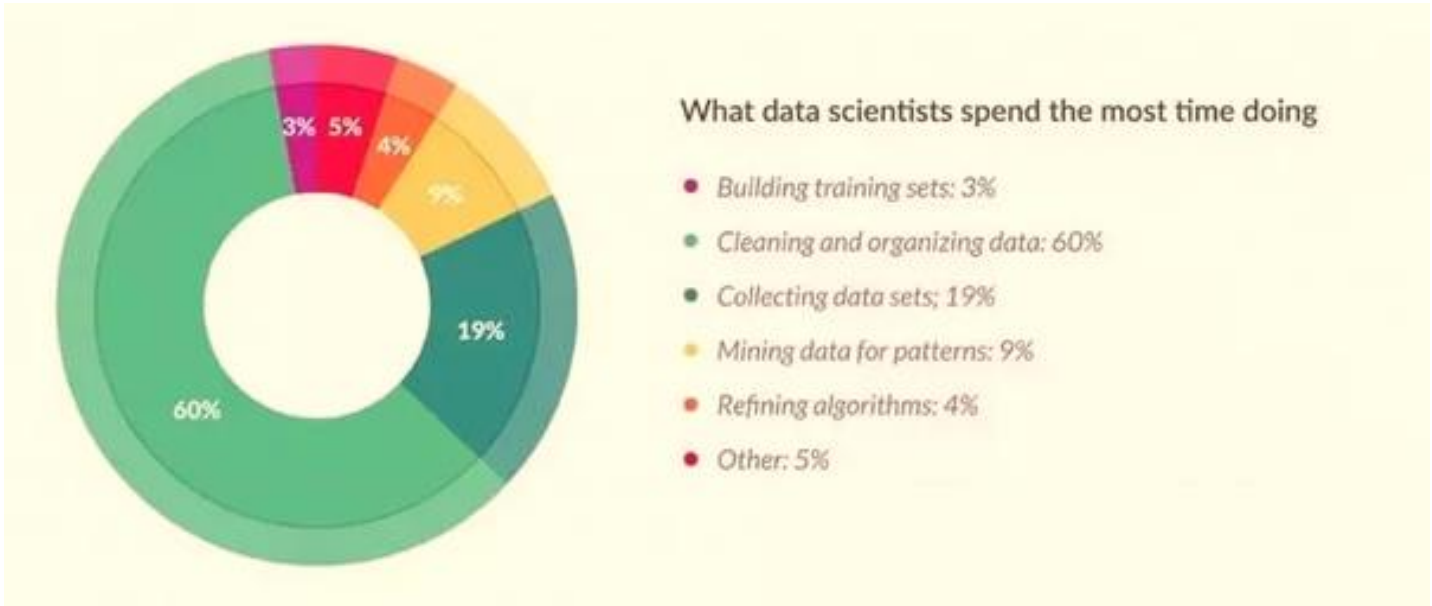
Note: For the videos related to this part I have used Data Science as a general term to avoid repetition of various areas of AI, ML, and Deep Learning

What is Data Science?

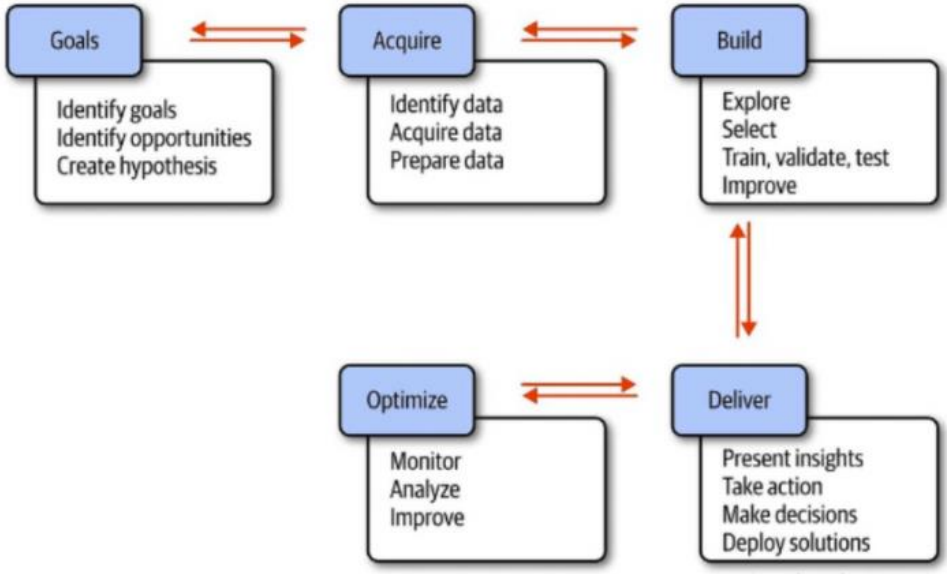


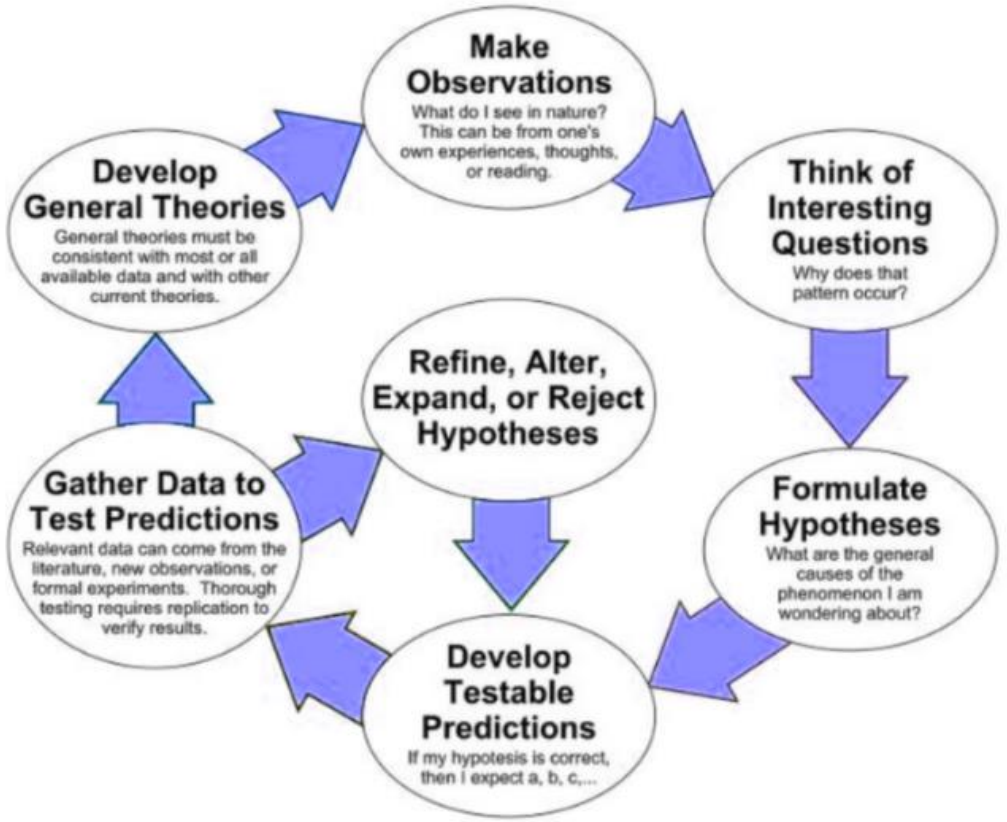
Data scientists are experts that use a wide range of skills — including statistics, computer science, and business knowledge — to analyze data collected from the multiple sources such as web, smartphones, customers, sensors, and a bunch of other sources.


Computer Science/IT
Machine Learning
Math and Statistics
Data Science
Software Development
Traditional Research
Domains/Business Knowledge

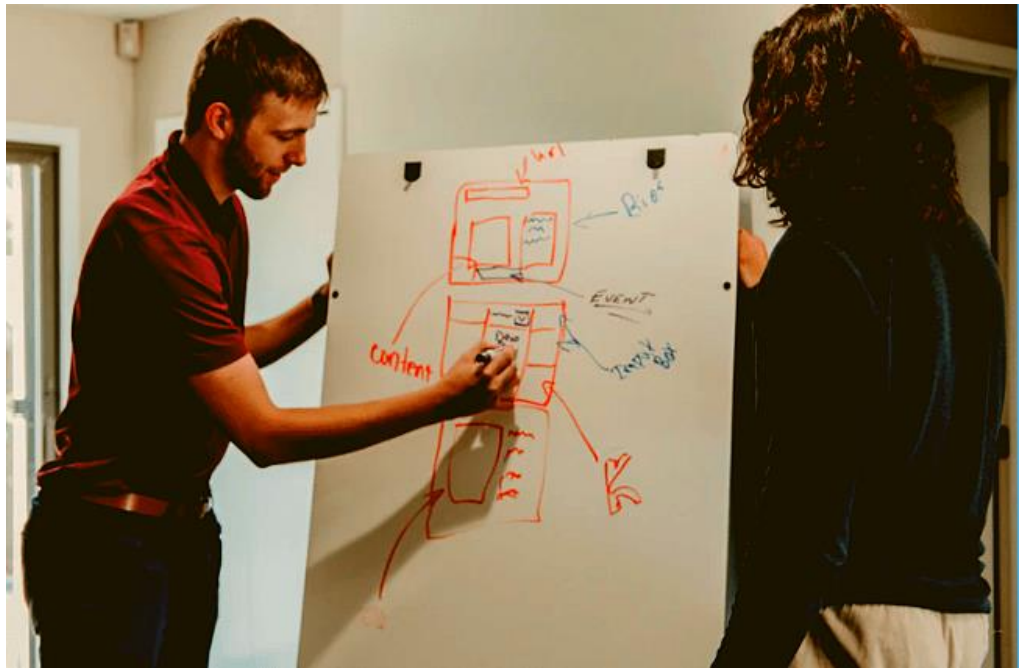

3%
5%
4%
9%
19%
60%
What data scientists spend the most time doing
Building training sets: 3%
Cleaning and organizing data: 60%
Collecting data sets; 19%
Mining data for patterns: 9%
Refining algorithms: 4%
Other: 5%

Data Science Process Lifecycle





Business Understanding



In Business Understanding we:

- Define goals
- We Specify the key variables
- We Identify the relevant data sources

How to do define goals?

- Define objectives: Work with your customer and other stakeholders to understand and identify the business problems.
- Define SMART success metrics

SMART Framework:

- Specific
- Measurable
- Achievable
- Relevant
- Time-bound

Business Understanding

Deliverables at this stage

- Charter Document: You update the template throughout the project as you make new discoveries and as business requirements change.

- Data sources Report: The Raw data sources report specifies the original and destination locations for the raw data.

- Data Dictionaries: This document provides descriptions of the data that's provided by the client.

Analytic Approach

The analytical options may include:-

- Statistical analysis applies to problems that require counts
- If the question is to determine probabilities of an action, then a predictive model might be used.
- If the question is to show relationships, a descriptive approach maybe be required.
- if the question requires a yes/ no answer, then a classification approach to predicting a response would be suitable.

ANALYTICS FRAMEWORK

DESCRIPTIVE	DIAGNOSTIC	PREDICTIVE	PRESCRIPTIVE
			
What happened?	Why it happened?	What will happen?	What action to take?

Data Requirements

Data Requirements analysis process employs a top-down approach that emphasizes business-driven needs, so the analysis is conducted to ensure the identified requirements are relevant and feasible

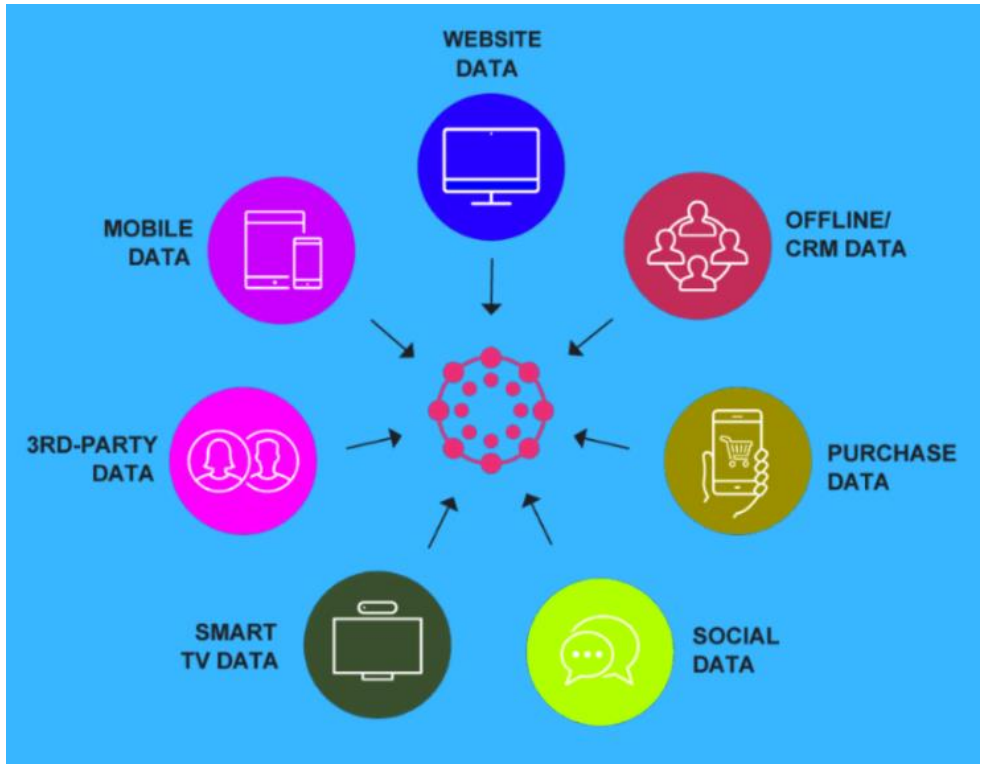
Identifying the business contexts is a 5-step process:

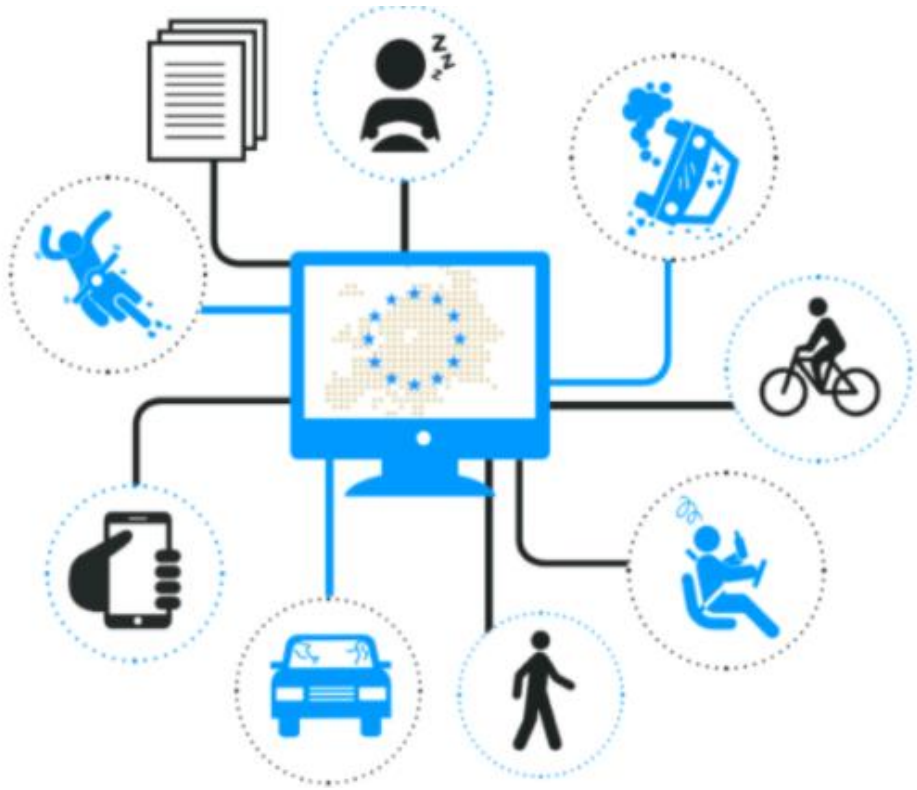
- Identifying relevant stakeholders
- Acquiring documentation
- Documenting goals and objectives
- Summarizing the scope of capabilities
- Documenting impacts and constraints

Data Requirements Analysis Process Phases

- Identifying the business contexts
- Conducting stakeholder interviews
- Synthesizing expectations and requirements
- Developing source-to-target mappings

Data Collection

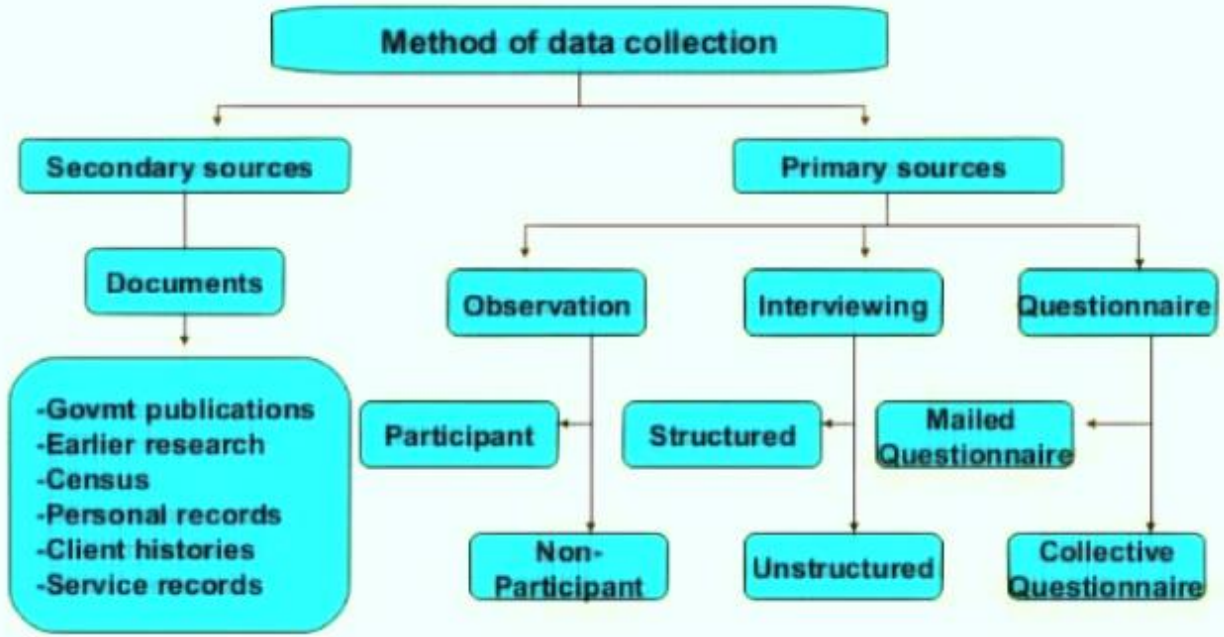




Options for Data Collection:

- observe the data from existing sources
- generate the data via surveys, simulations, and experiments

Method of data collection



Data Understanding

Is the data that you collected representative of the problem to be solved?

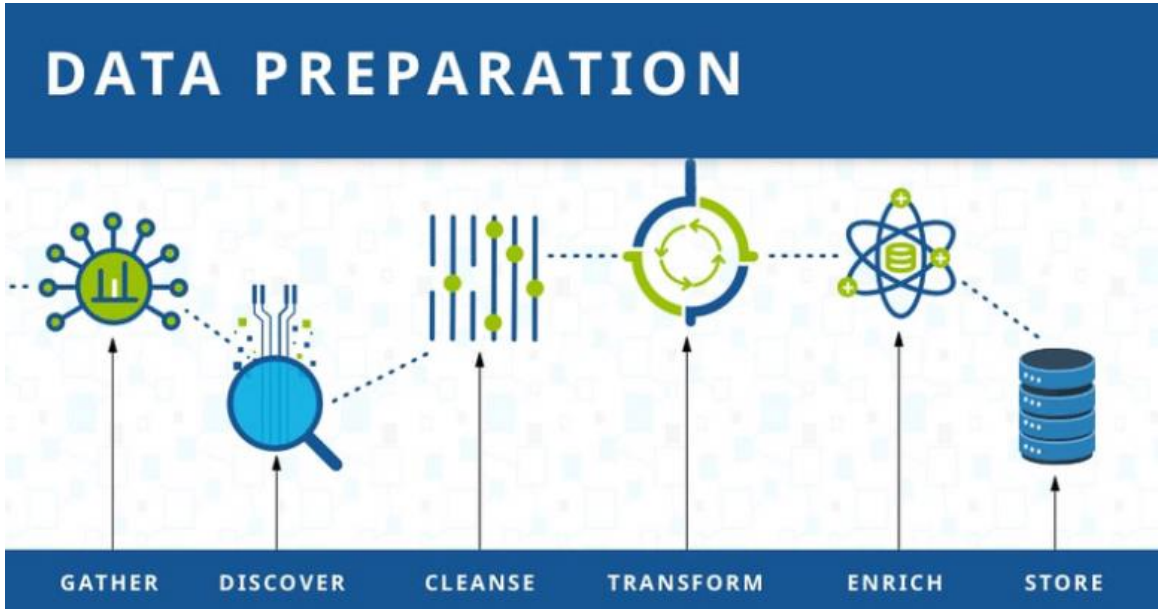
Reports that helps us to know our data better:

- Initial data collection report - list of data sources acquired, their locations, the methods used to acquire them, and any problems encountered.

- Data description report - description of data that has been acquired, including its format, quantity, identities of the fields, etc.

- Data exploration report - address data mining questions using querying, visualization, and reporting techniques.

Data Preparation



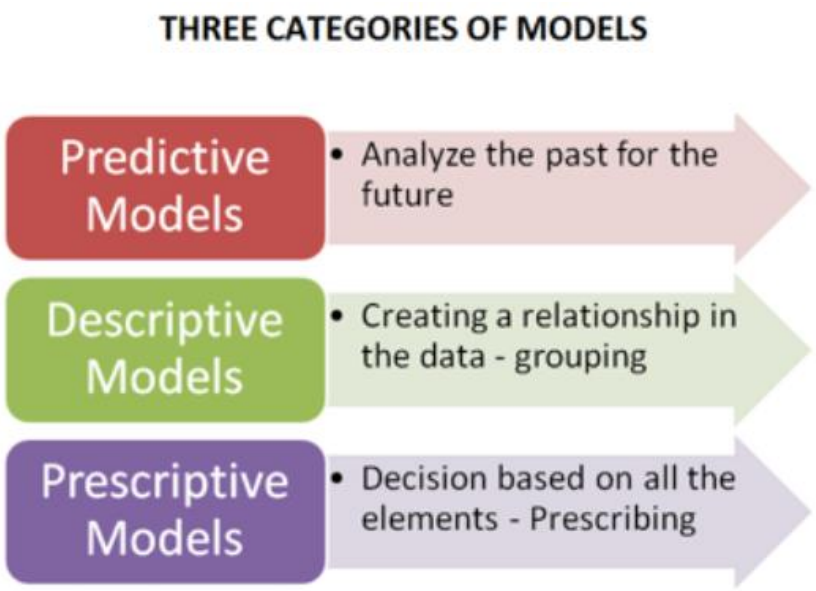
Steps in Data Preparation

- Deleting Unnecessary Data
- Consolidating/Separating Fields
- Transforming Data
- Making Corrections to Data
- Changing Formats

Modeling

Modeling is geared towards answering two key questions

- *What is the purpose of data modeling?*
- *What are the characteristics of the process?*

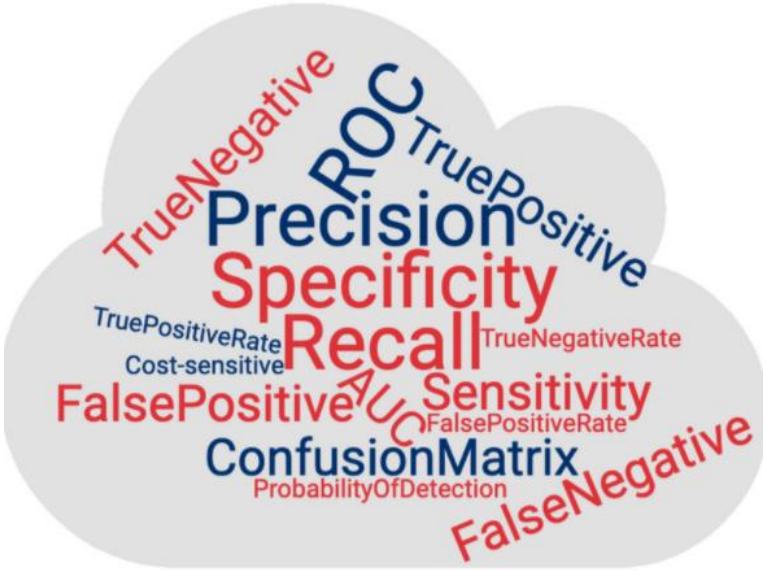


Model Evaluation

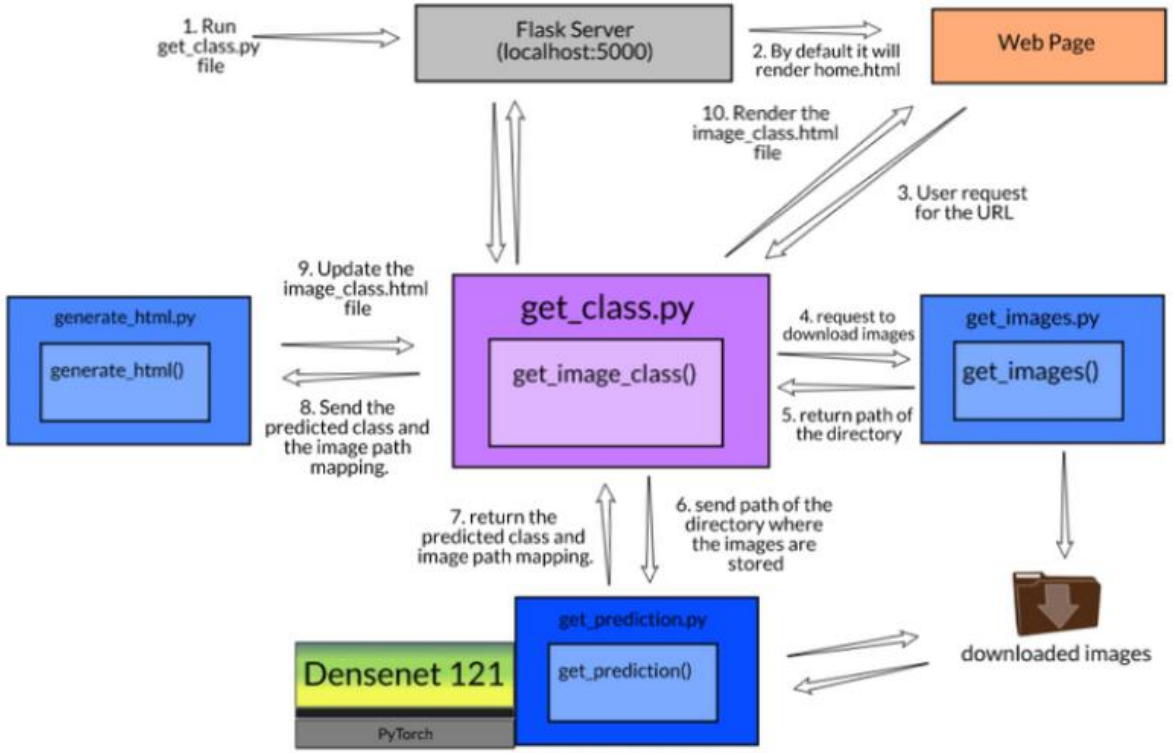
Model evaluation is performed during model development and before the model is deployed

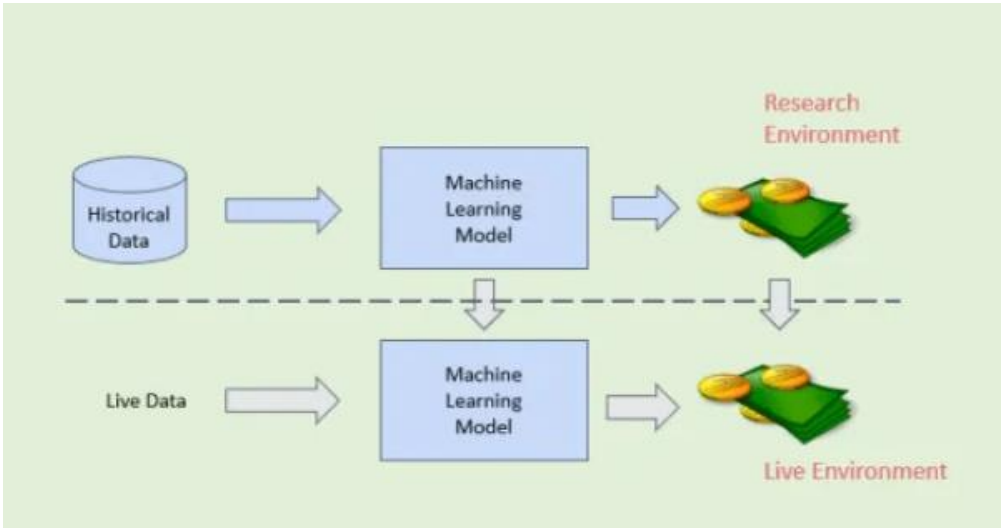
Evaluation of the model

- Hold-Out
- Cross-Validation



Deployment


1. Run get_class.py file
Flask Server (localhost:5000)
2. By default it will render home.html
Web Page
10. Render the image_class.html file
3. User request for the URL
9. Update the image_class.html file
generate_html.py
generate_html()
get_class.py
get_image_class()
4. request to download images
get_images.py
get_images()
8. Send the predicted class and the image path mapping.
5. return path of the directory
7. return the predicted class and image path mapping.
6. send path of the directory where the images are stored
get_prediction.py
Densenet 121
PyTorch
get_prediction()
downloaded images


Research Environment
Historical Data
Machine Learning Model
Live Data
Machine Learning Model
Live Environment

Feedback

Once the model is deployed, feedback from the users will be used to refine the model and assess it for performance and impact