

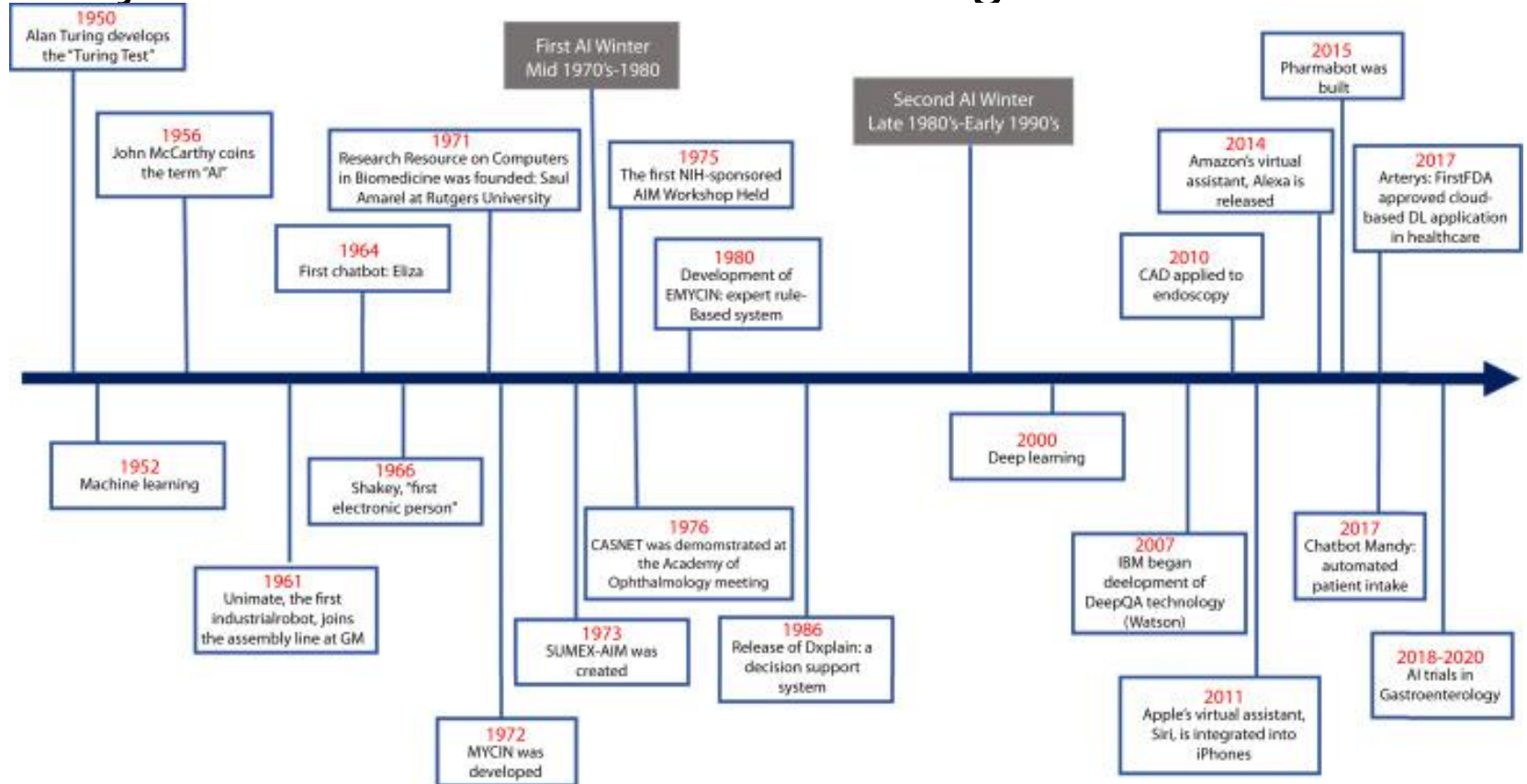
MIDS W207

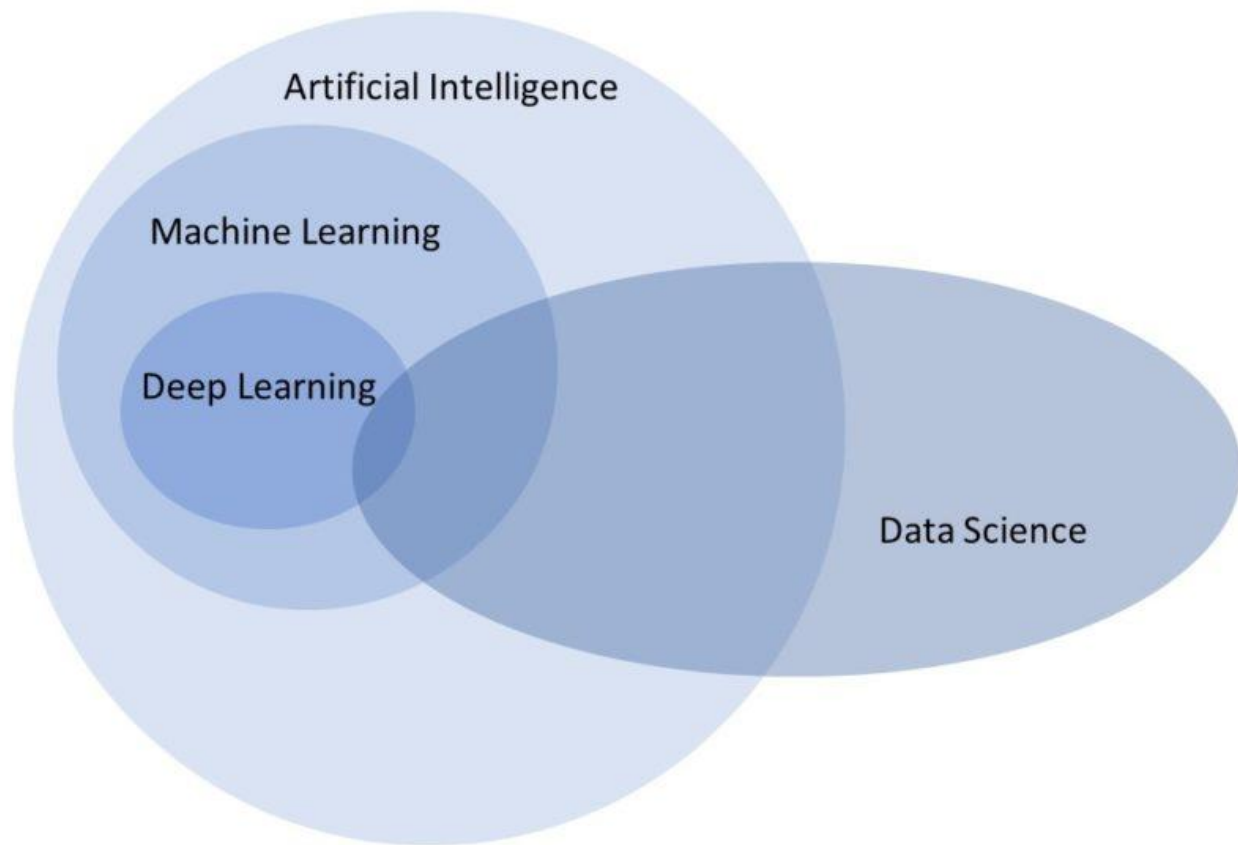
# Applied Machine Learning

Spring 2023

Week 1

# History and Timeline of Artificial Intelligence







Who is my ML system for?

Am I using a representative dataset?

Is there real-world / human bias in my data?

How is my model performing?

What can I do to improve the model?

Define Problem

Construct and Prepare Data

Build and Train Model

Deploy

Iterate

Are there any privacy considerations?

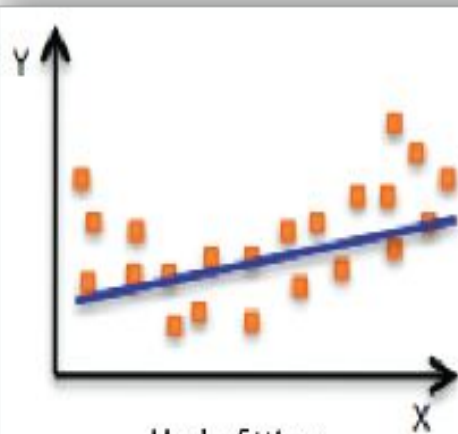
Where do I get relevant features in a privacy preserving way?

Are test users diverse?

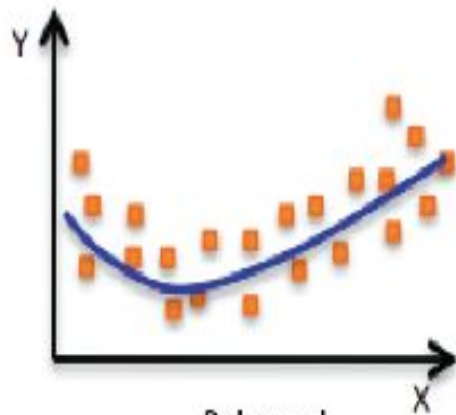
How does my data affect model performance?

Should I deploy my model?

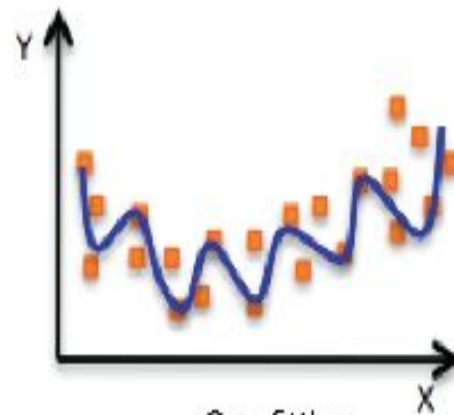
Are there complex feedback loops?



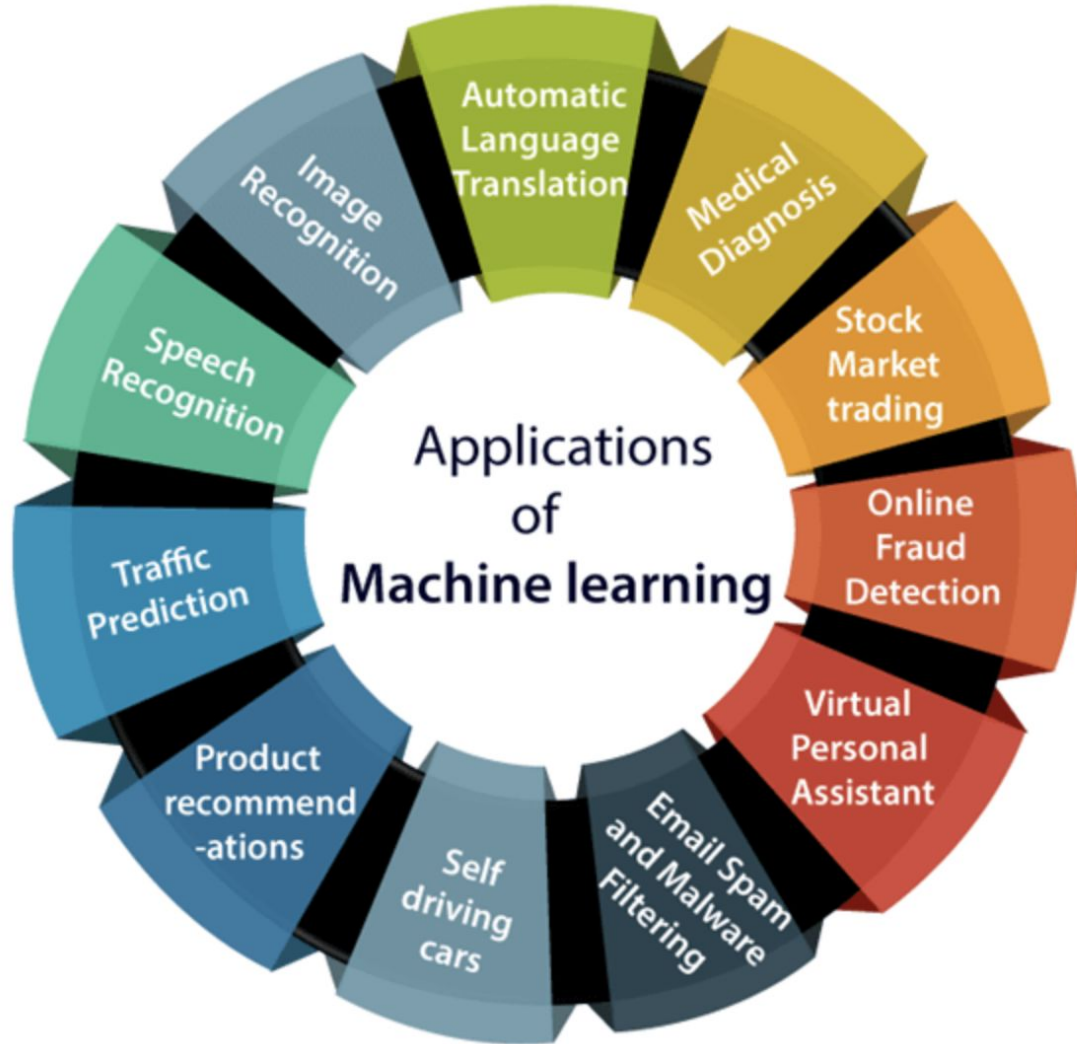
Underfitting

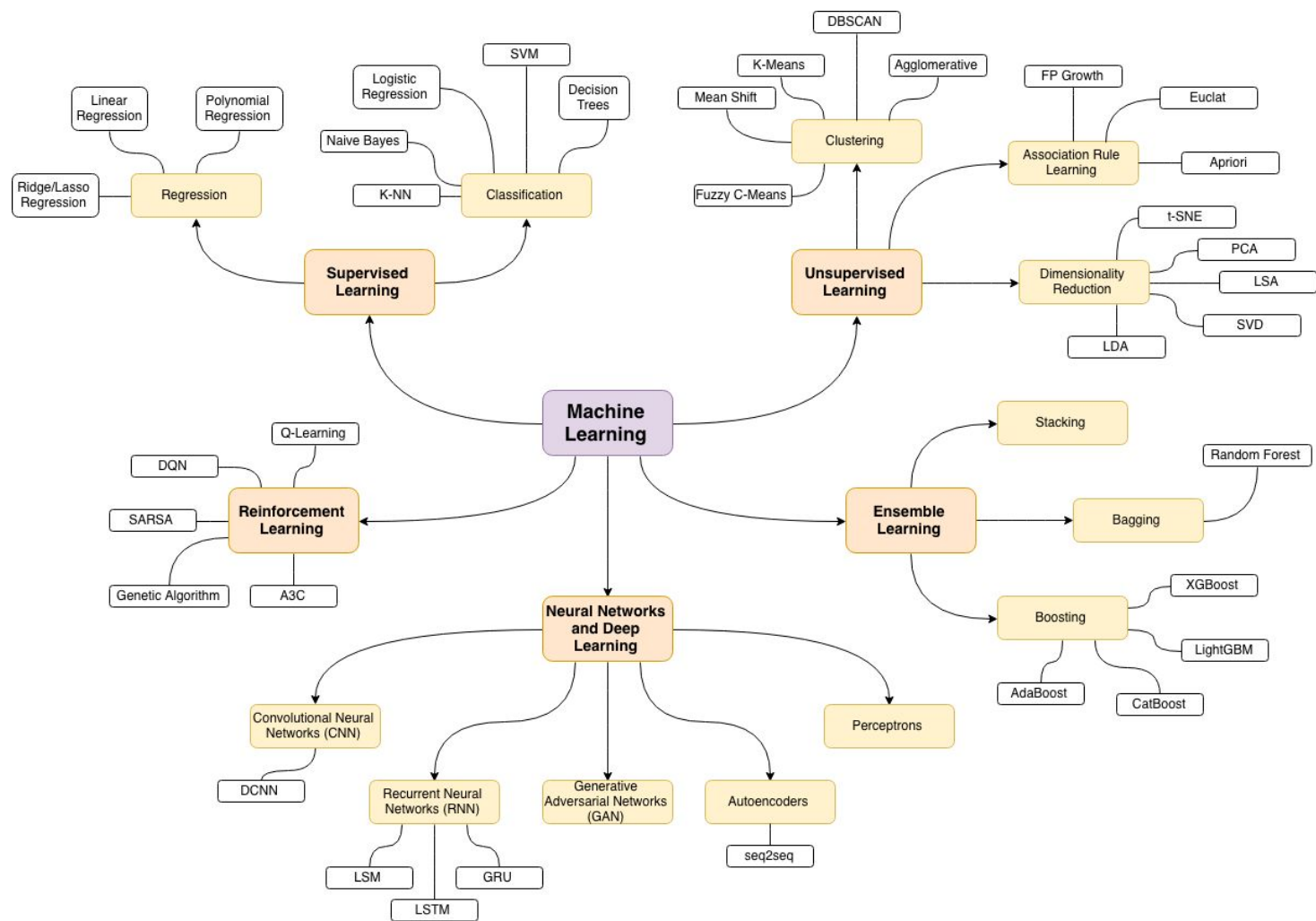


Balanced



Overfitting





# Fundamentals

## Supervised Learning

- Makes machine Learn explicitly
- Data with clearly defined output is given
- Direct feedback is given
- Predicts outcome/future
- Resolves classification and regression problems



## Unsupervised Learning

- Machine understands the data (Identifies patterns/structures)
- Evaluation is qualitative or indirect
- Does not predict/find anything specific



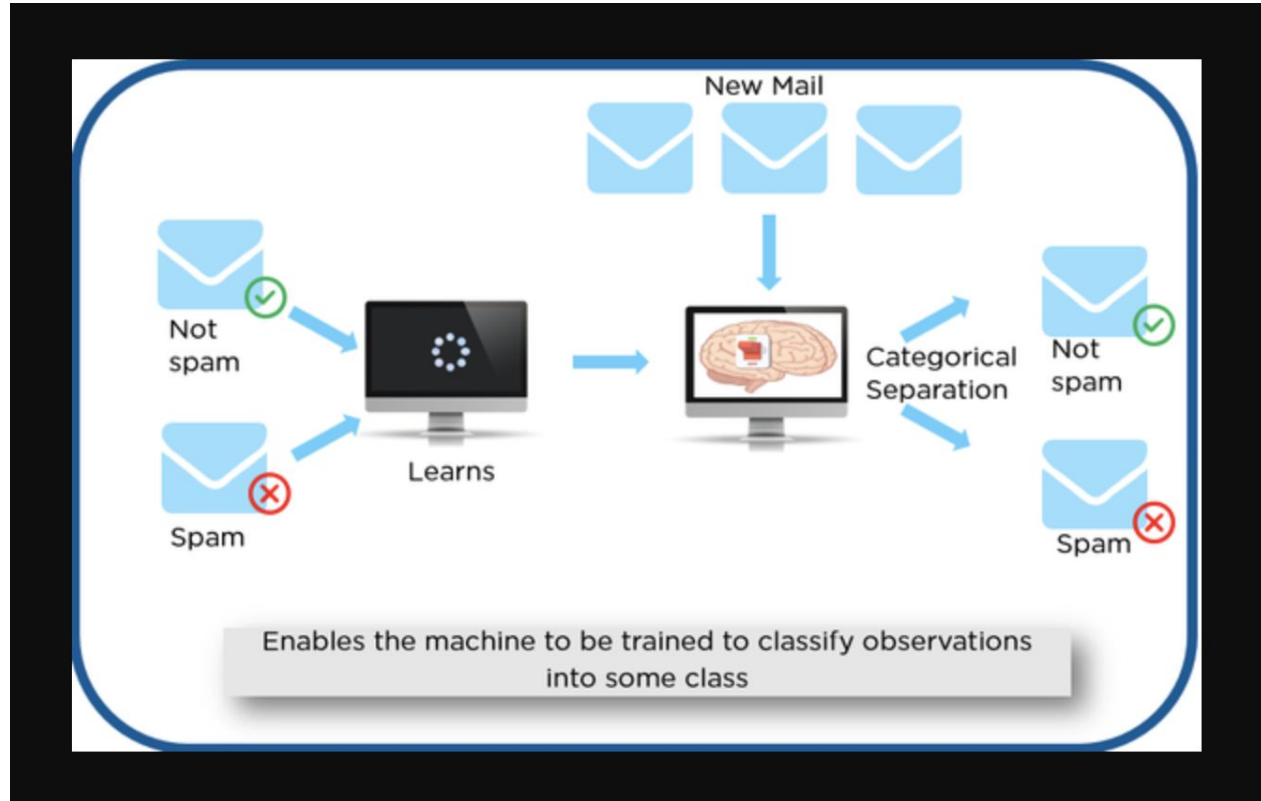
## Reinforcement Learning

- An approach to AI
- Reward based learning
- Learning form +ve & +ve reinforcement
- Machine Learns how to act in a certain environment
- To maximize rewards

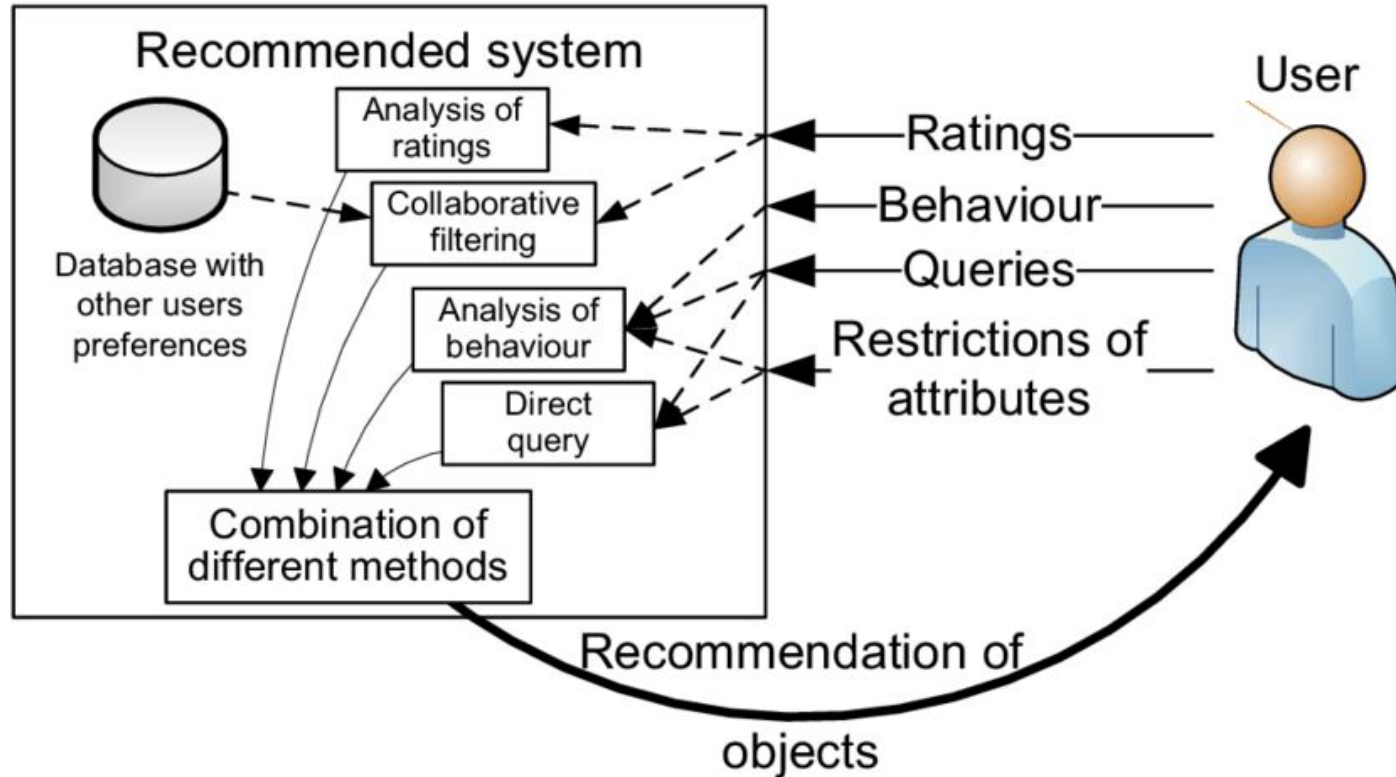




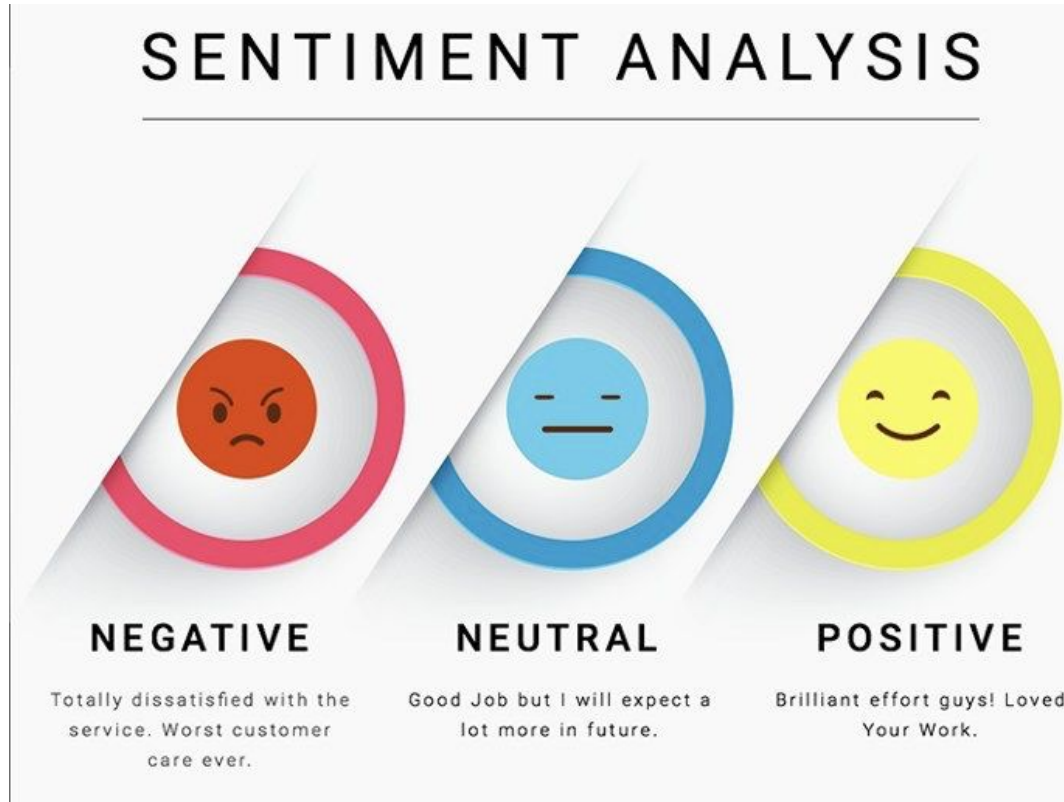
# Applications: Spam and Non Spam



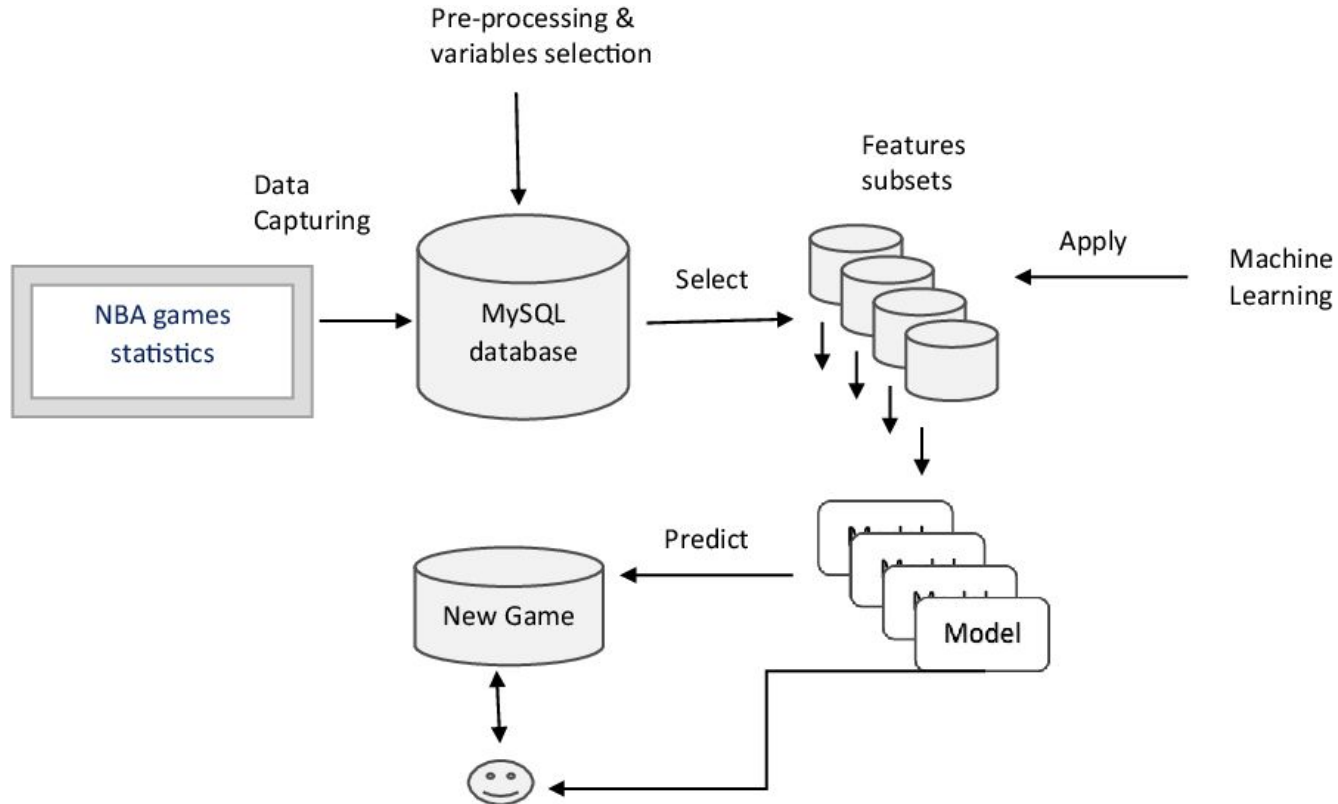
# Applications: Recommendation Systems



# Applications: Sentiment Analysis



# Applications: Sports Prediction



# Evaluating the Models

## Confusion Matrix

		Actual Value	
		Yes (1)	No (0)
Predicted Value	Yes (1)	TP	FP
	No (0)	FN	TN

TP= True Positive

FP= False Positive

FN= False Negative

TN= True Negative

- If you have supervised data, you will want to maximize an objective function.
  - **Precision:**  $TP \div (TP + FP)$  % positives correctly identified
  - **Recall:**  $TP \div (TP + FN)$  % existing positives identified
  - **Optimal point** on ROC (precision/recall) curve
  - **Accuracy:**  $(TP + TN) \div (TP + TN + FP + FN)$
  - **F-test:**  $2 \cdot (P \cdot R) \div (P + R)$