MIDS W207
Applied Machine Learning

Spring 2024
Week 01

History and Timeline of Artificial Intelligence

Artificial Intelligence

Machine Learning

Deep Learning

Data Science
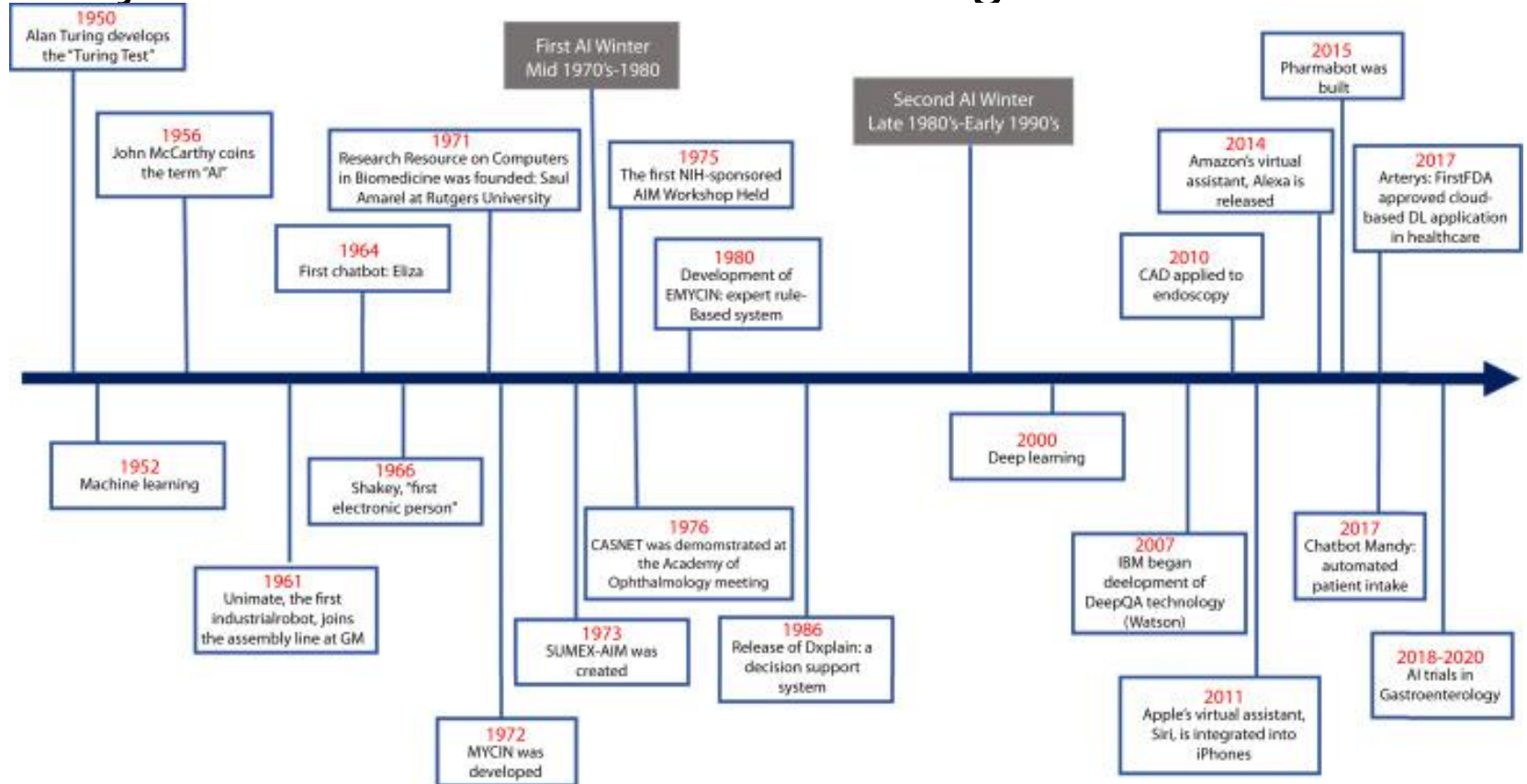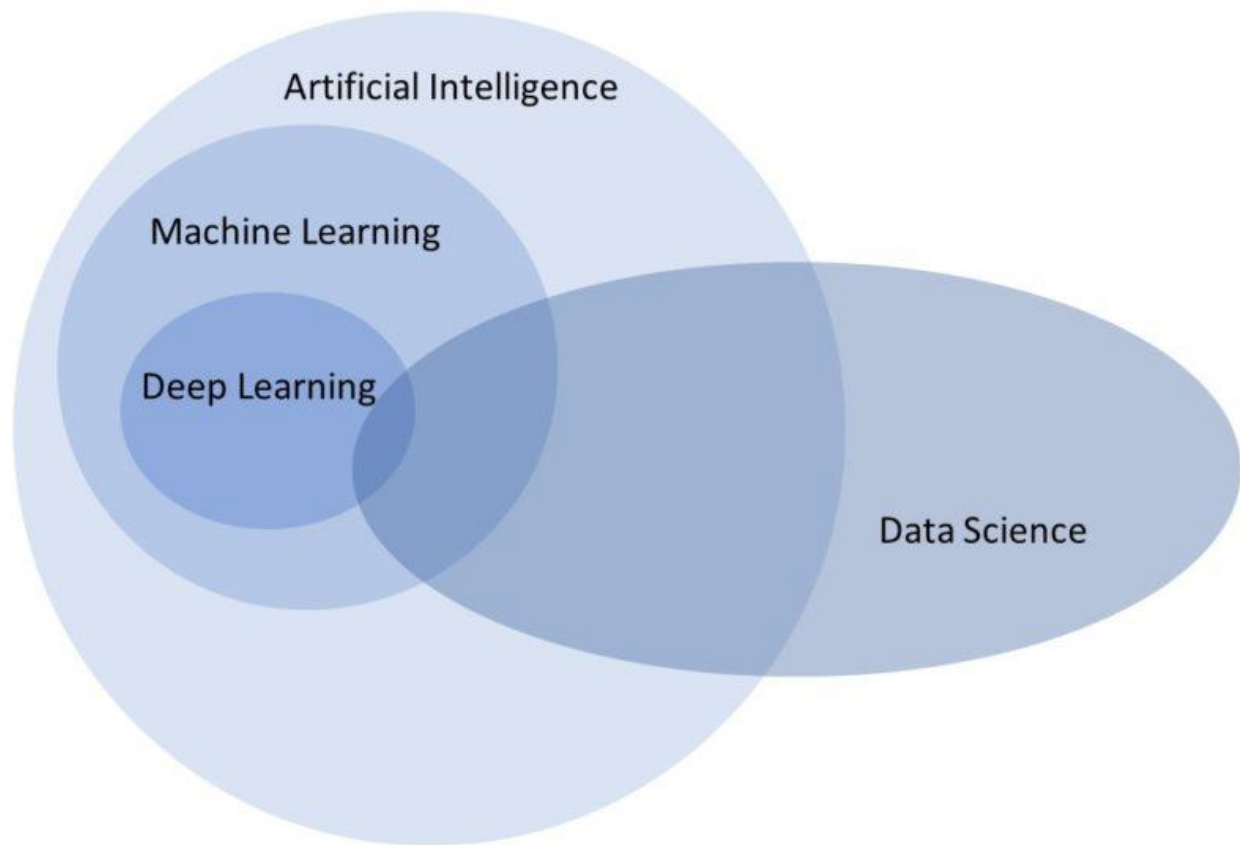
Who is my ML system for?

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**
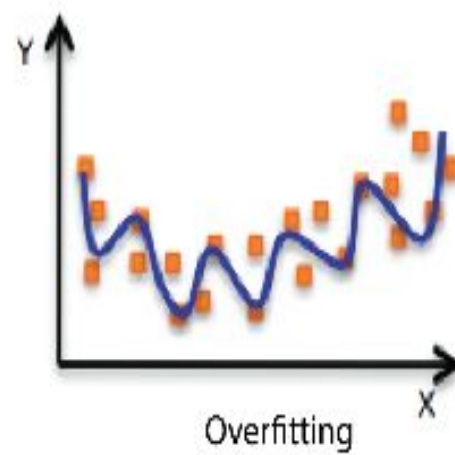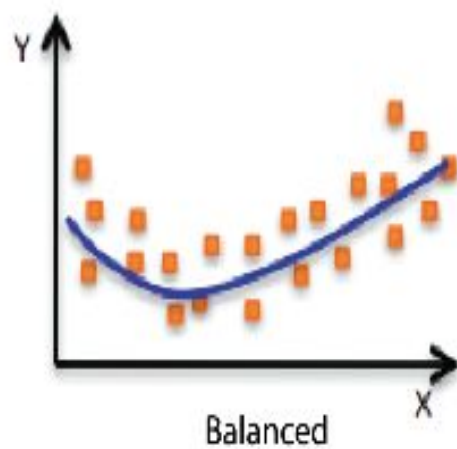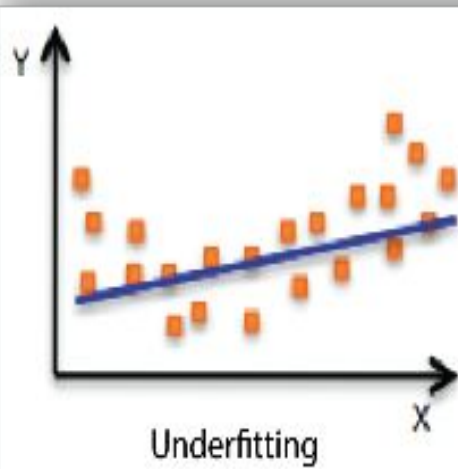
Are there any privacy considerations?

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

Applications of Machine learning

- Automatic Language Translation
- Medical Diagnosis
- Stock Market trading
- Online Fraud Detection
- Virtual Personal Assistant
- Email Spam and Malware Filtering
- Self driving cars
- Product recommend-ations
- Traffic Prediction
- Speech Recognition
- Image Recognition

Machine Learning

Supervised Learning
- Regression
  - Linear Regression
  - Polynomial Regression
  - Ridge/Lasso Regression
- Classification
  - Logistic Regression
  - SVM
  - Decision Trees
  - Naive Bayes
  - K-NN

Unsupervised Learning
- Clustering
  - K-Means
  - DBSCAN
  - Agglomerative
  - Mean Shift
  - Fuzzy C-Means
- Association Rule Learning
  - FP Growth
  - Euclat
  - Apriori
- Dimensionality Reduction
  - t-SNE
  - PCA
  - LSA
  - SVD
  - LDA

Reinforcement Learning
- Q-Learning
- DQN
- SARSA
- Genetic Algorithm
- A3C

Ensemble Learning
- Stacking
- Bagging
  - Random Forest
- Boosting
  - XGBoost
  - LightGBM
  - AdaBoost
  - CatBoost

Neural Networks and Deep Learning
- Convolutional Neural Networks (CNN)
  - DCNN
- Recurrent Neural Networks (RNN)
  - LSM
  - LSTM
  - GRU
- Generative Adversarial Networks (GAN)
- Autoencoders
  - seq2seq
- Perceptrons

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

Training
Inputs → 💻 → Outputs

Unsupervised Learning

• Machine understands the data (Identifies patterns/structures)
•Evaluation is qualitative or indirect
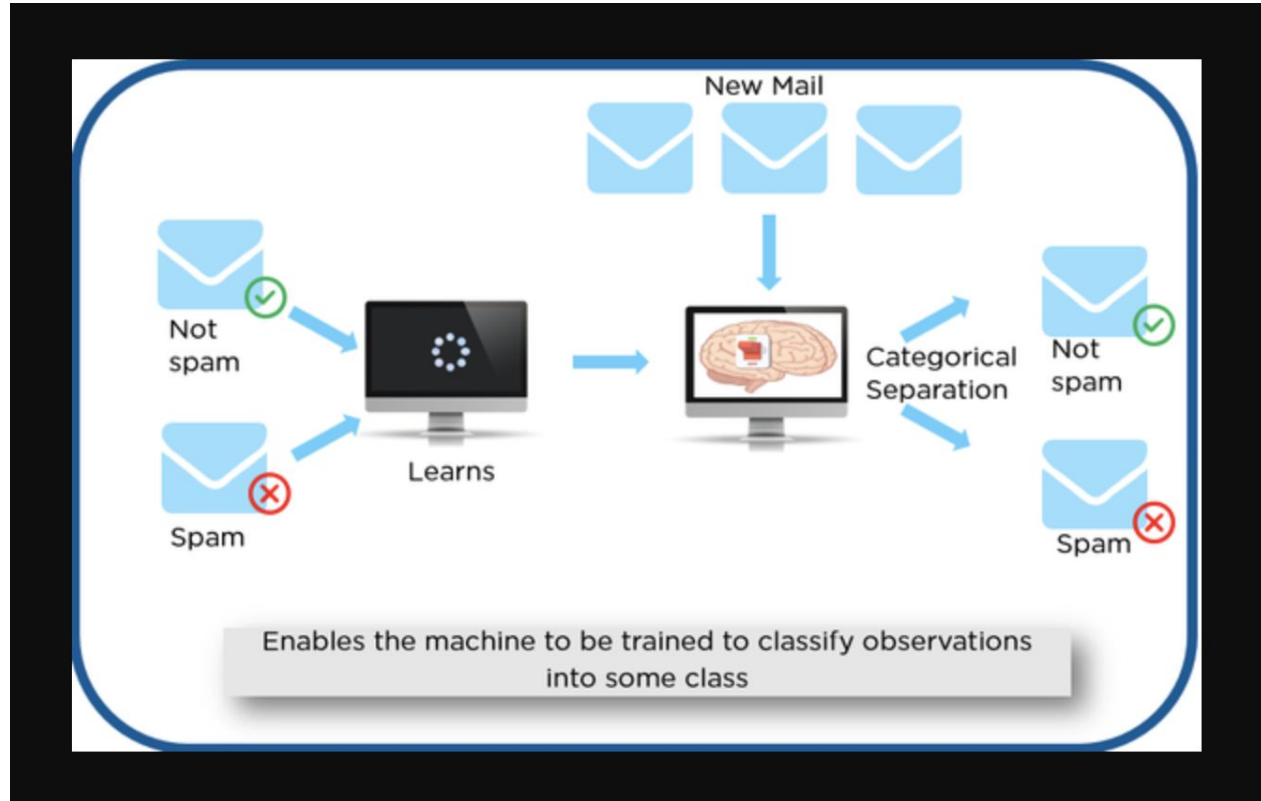• Does not predict/find anything specific
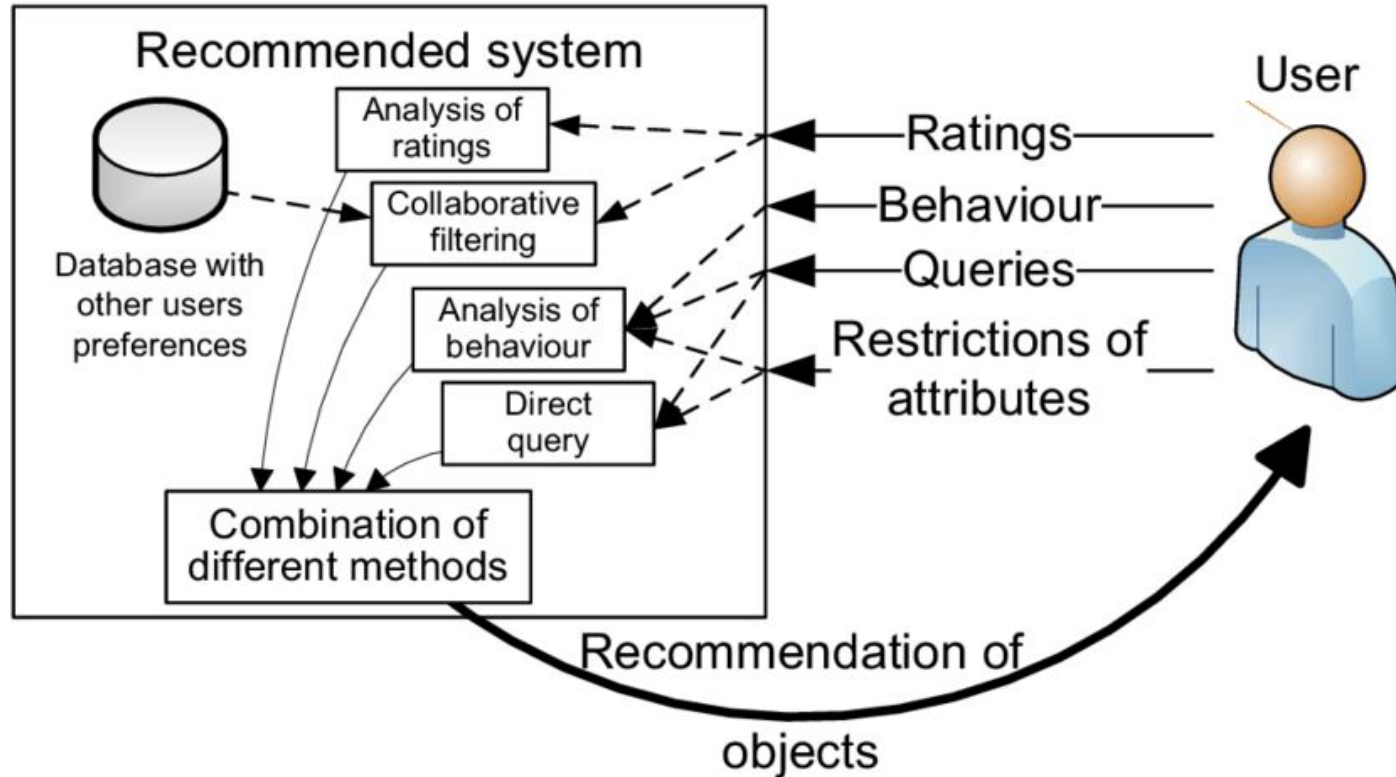
Inputs → 💻 → Outputs

Reinforcement Learning

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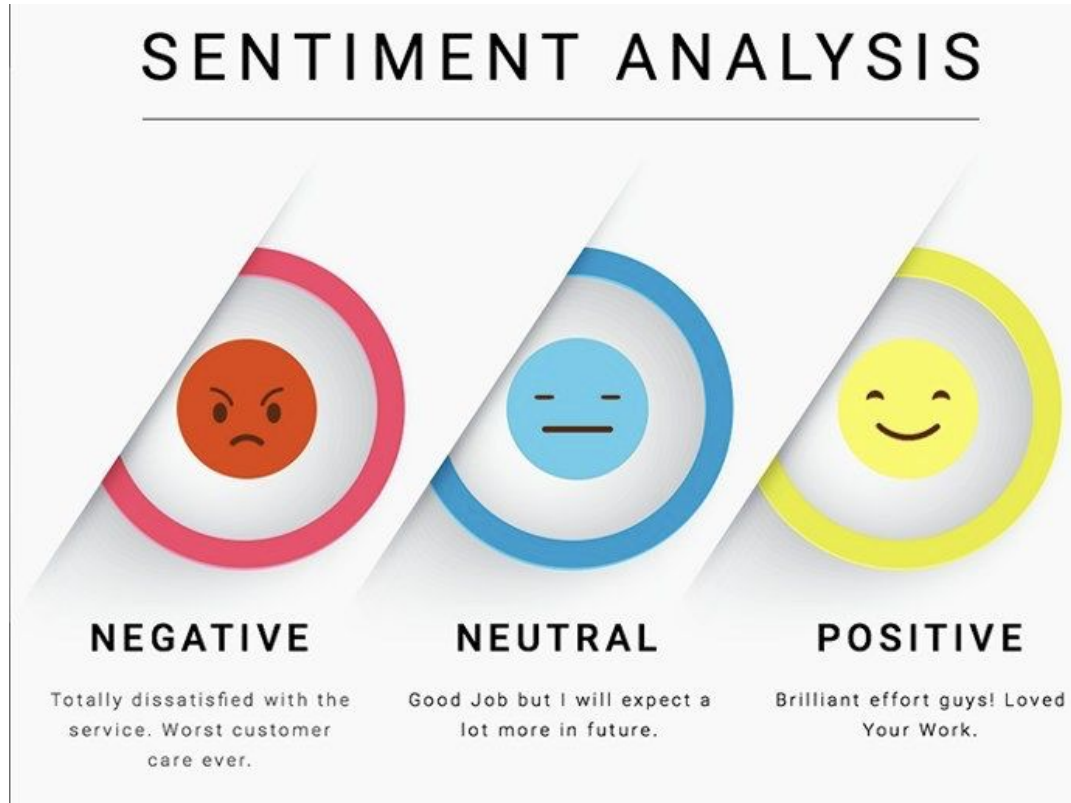•Machine Learns how to act in a certain environment
• To maximize rewards

Rewards
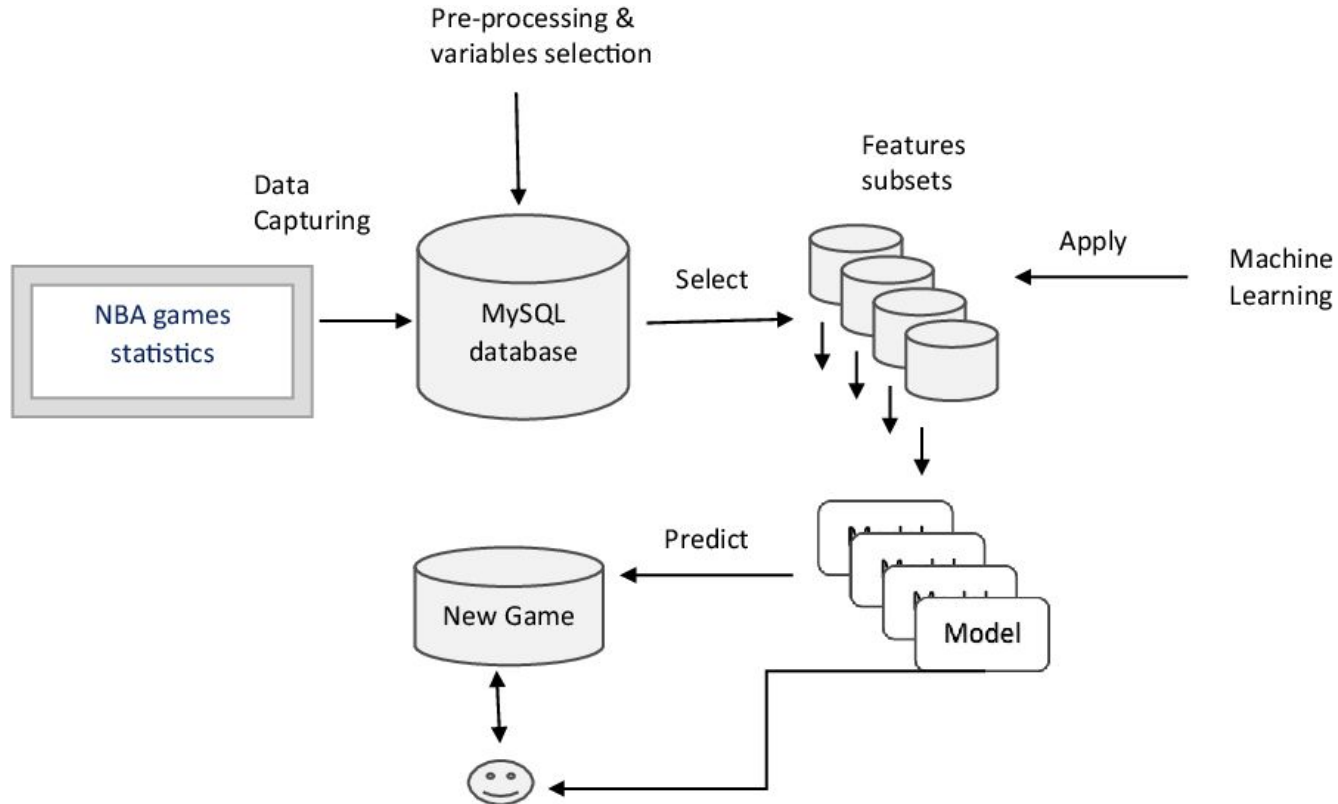Inputs → 💻 → Outputs

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 identifed
  - **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**: $2 \cdot (P \cdot R) \div (P + R)$