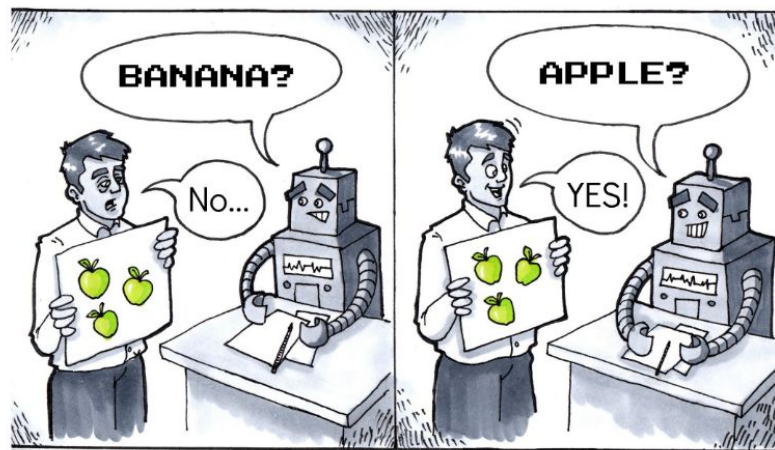


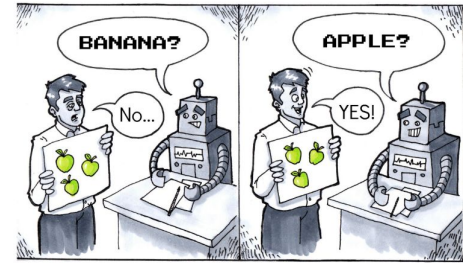
# Supervised Learning



**Supervised Learning**

# Supervised Learning

Feature Engineering, Models, Hyperparameters, Data Leaks

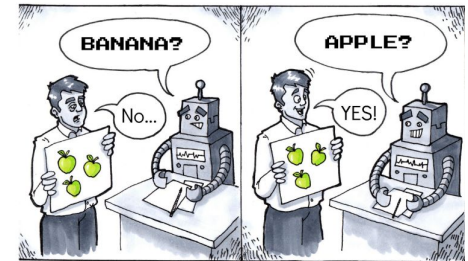


**Supervised Learning**

# Supervised Learning

Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness

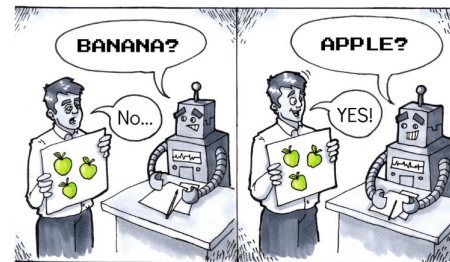


**Supervised Learning**

# Supervised Learning

## Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness
- **Naive-Bayes** - Good for text data, does not work for continuous variables by default, its naive.

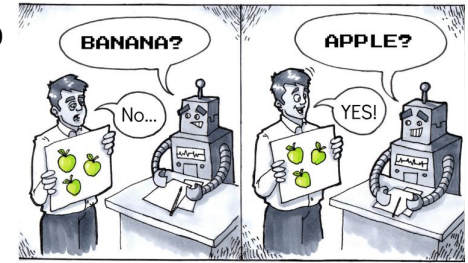


Supervised Learning

# Supervised Learning

## Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness
- **Naive-Bayes** - Good for text data, does not work for continuous variables by default, its naive.
- **Decision Trees** - Interpretable, handle mixed data types, robust to outliers, overfitting, instability, bias towards majority class

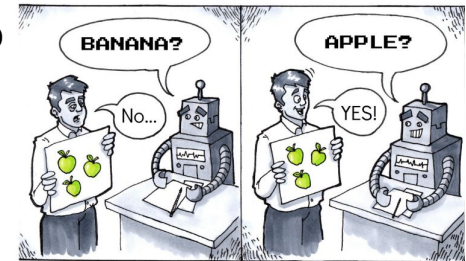


Supervised Learning

# Supervised Learning

## Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness
- **Naive-Bayes** - Good for text data, does not work for continuous variables by default, its naive.
- **Decision Trees** - Interpretable, handle mixed data types, robust to outliers, overfitting, instability, bias towards majority class
- **SVM** - Good in high dimensions, hyperparameter choice, compute

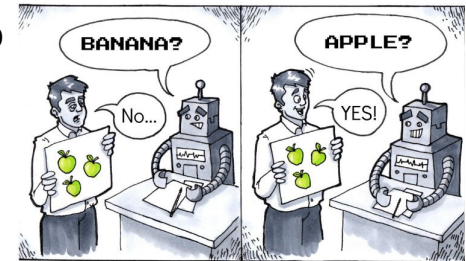


Supervised Learning

# Supervised Learning

## Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness
- **Naive-Bayes** - Good for text data, does not work for continuous variables by default, its naive.
- **Decision Trees** - Interpretable, handle mixed data types, robust to outliers, overfitting, instability, bias towards majority class
- **SVM** - Good in high dimensions, hyperparameter choice, compute
- **KNN** - Simple, no training, adaptable, hyperparameter choice, compute, sensitive to class imbalance, curse of dimensionality

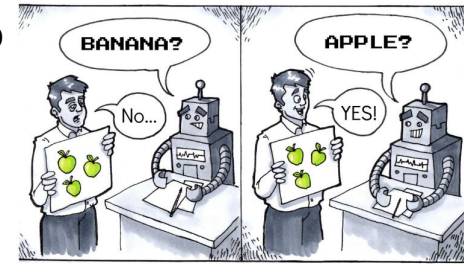


Supervised Learning

# Supervised Learning

## Feature Engineering, Models, Hyperparameters, Data Leaks

- **Logistic Classification** - Simple, efficient, regularization, sensitive to outliers and imbalanced data, limited expressiveness
- **Naive-Bayes** - Good for text data, does not work for continuous variables by default, its naive.
- **Decision Trees** - Interpretable, handle mixed data types, robust to outliers, overfitting, instability, bias towards majority class
- **SVM** - Good in high dimensions, hyperparameter choice, compute
- **KNN** - Simple, no training, adaptable, hyperparameter choice, compute, sensitive to class imbalance, curse of dimensionality
- **Ensembles** - Mix the best of everything, more complex, more compute



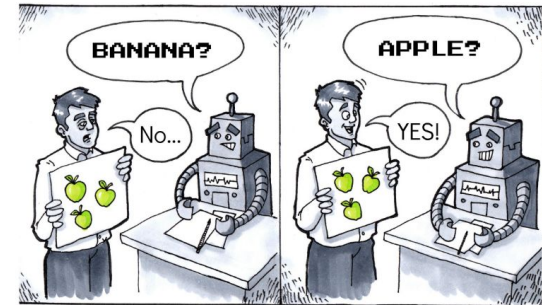
Supervised Learning



# Supervised Learning

There is always more.

- What if there are not enough data points?
- What if there are not enough data points with labels ?
- How to interpret the model?
- Are all features worth the same ?
- Is the model fair ?
- How to monitor the model ?
- What to monitor ?
- How to create feedback loops ?
- How to know if the models are still working after being deployed ?
- How do we retrain the model ?
- How do we deploy the model ?



**Supervised Learning**

# Supervised Learning

Lets compete.

The word "kaggle" is written in a light blue, lowercase, sans-serif font. The letters are slightly shadowed, giving them a 3D appearance as if they are floating above the white background.

# Supervised Learning

Lets compete.

The Kaggle logo, consisting of the word "kaggle" in a lowercase, blue, sans-serif font.

[Sentiment Analysis on Movie Reviews | Kaggle](#)

[Binary Classification with a Software Defects Dataset | Kaggle](#)