

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension

Production

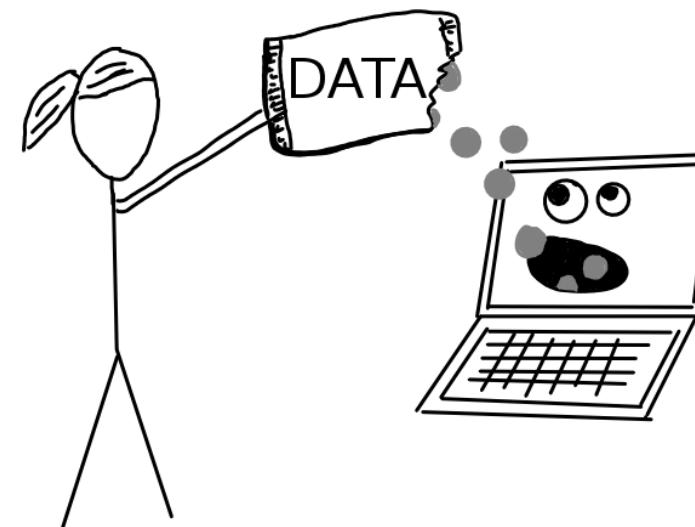
# Giới thiệu chung về Machine Learning

## 1. Tư tưởng của Machine Learning và Lập trình truyền thống

### Without Machine Learning



### With Machine Learning



## 2. Lịch sử của Machine Learning

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống
2. Lịch sử của Machine Learning
3. Ứng dụng của Machine Learning trong thực tế
  - 3.1. Autopilot - Tesla
  - 3.2. Siri - Apple
  - 3.3. AlphaGo - Google Deepmind
  - 3.4. Social credit scores - China government
4. Workflow để xây dựng được mô hình Machine Learning

### 5. Các nhóm mô hình Machine Learning

#### 5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

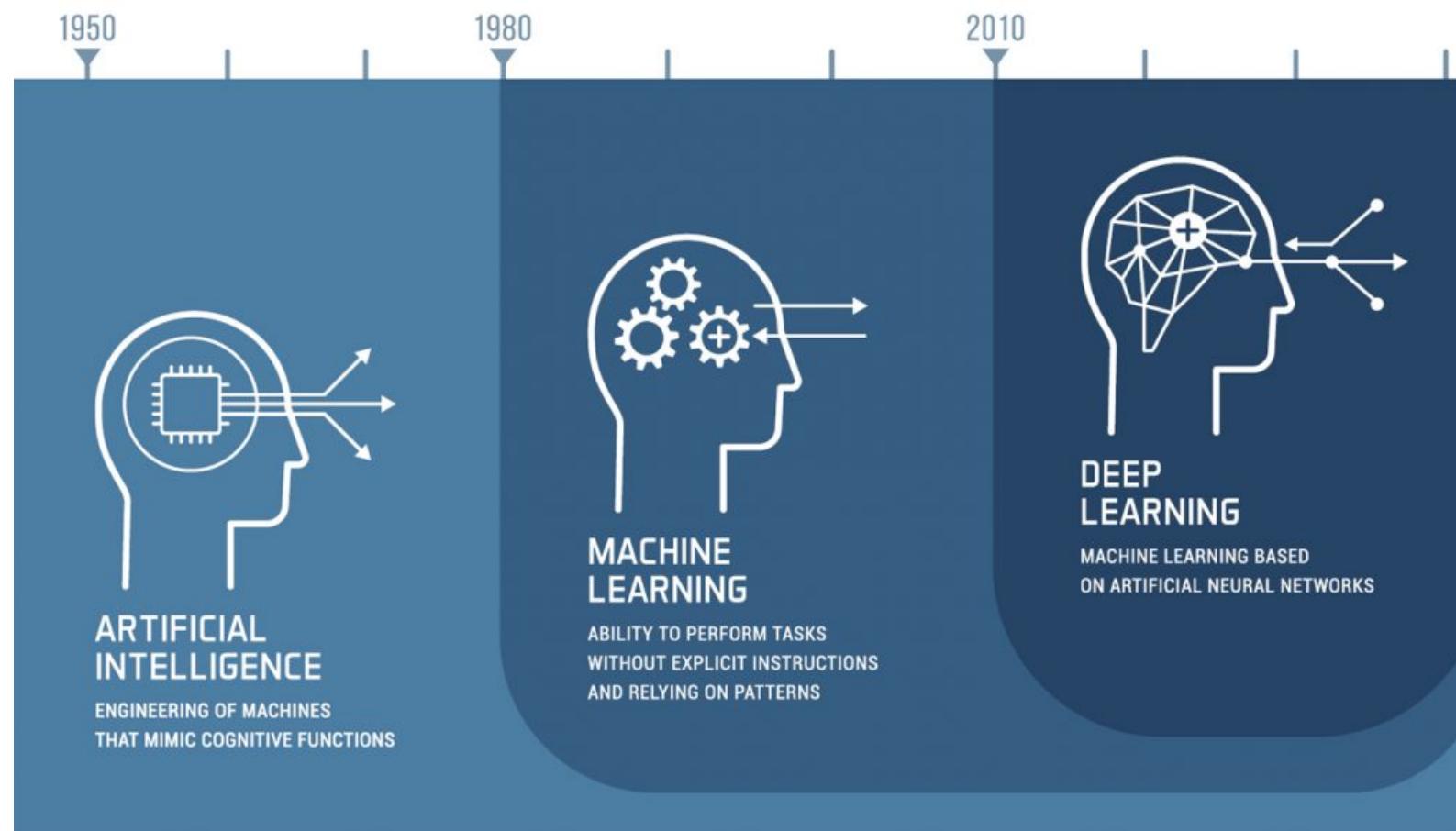
5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

#### 5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension

Production



## 3. Ứng dụng của Machine Learning trong thực tế

### 3.1. Autopilot - Tesla

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



### 3.2. Siri - Apple



# Hey Siri

### 3.3. AlphaGo - Google Deepmind

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

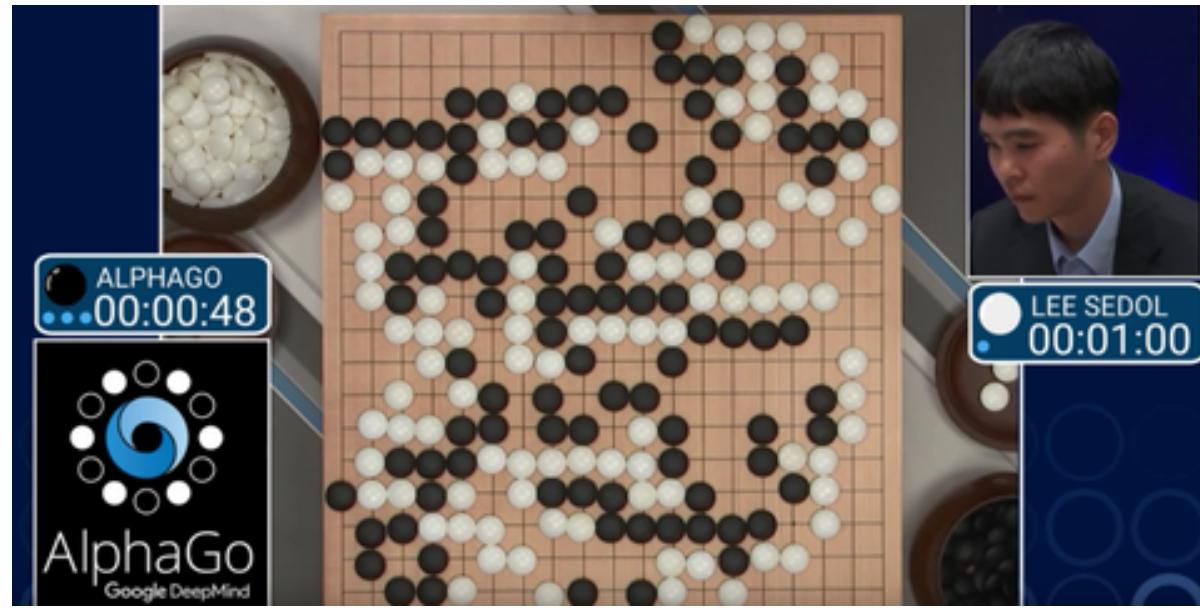
5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

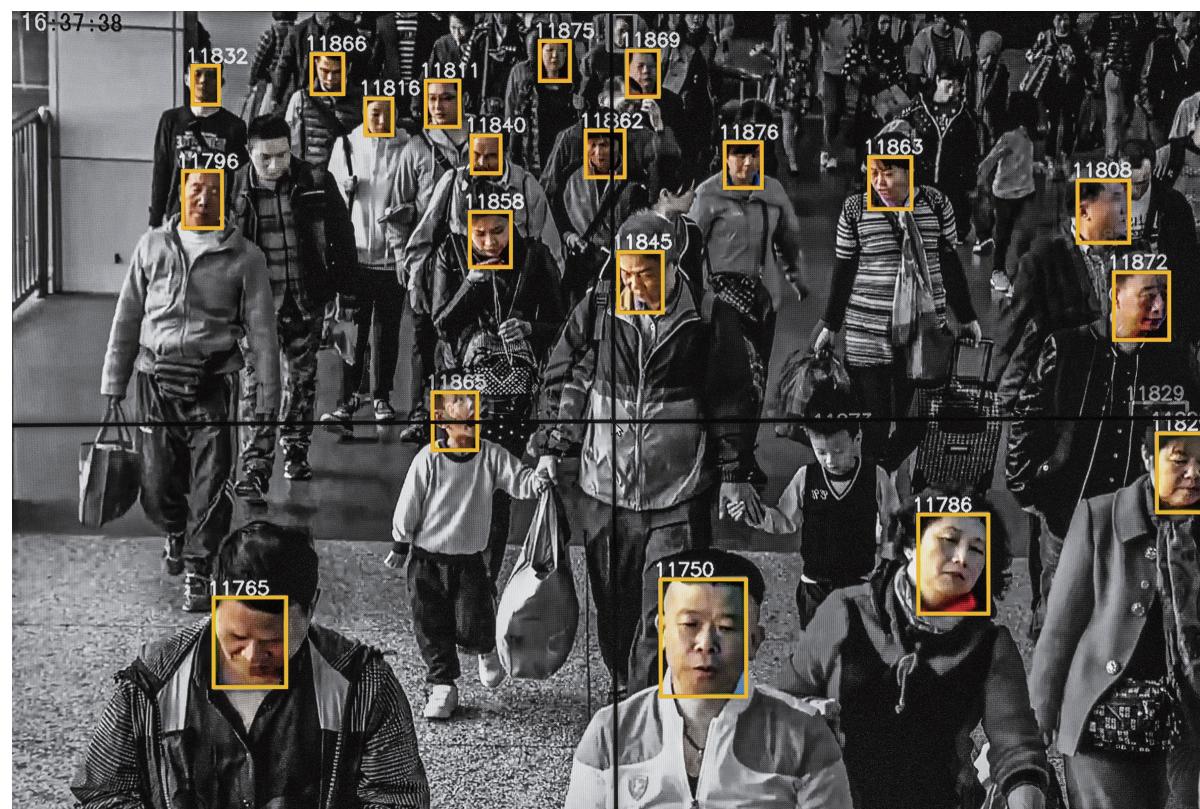
5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



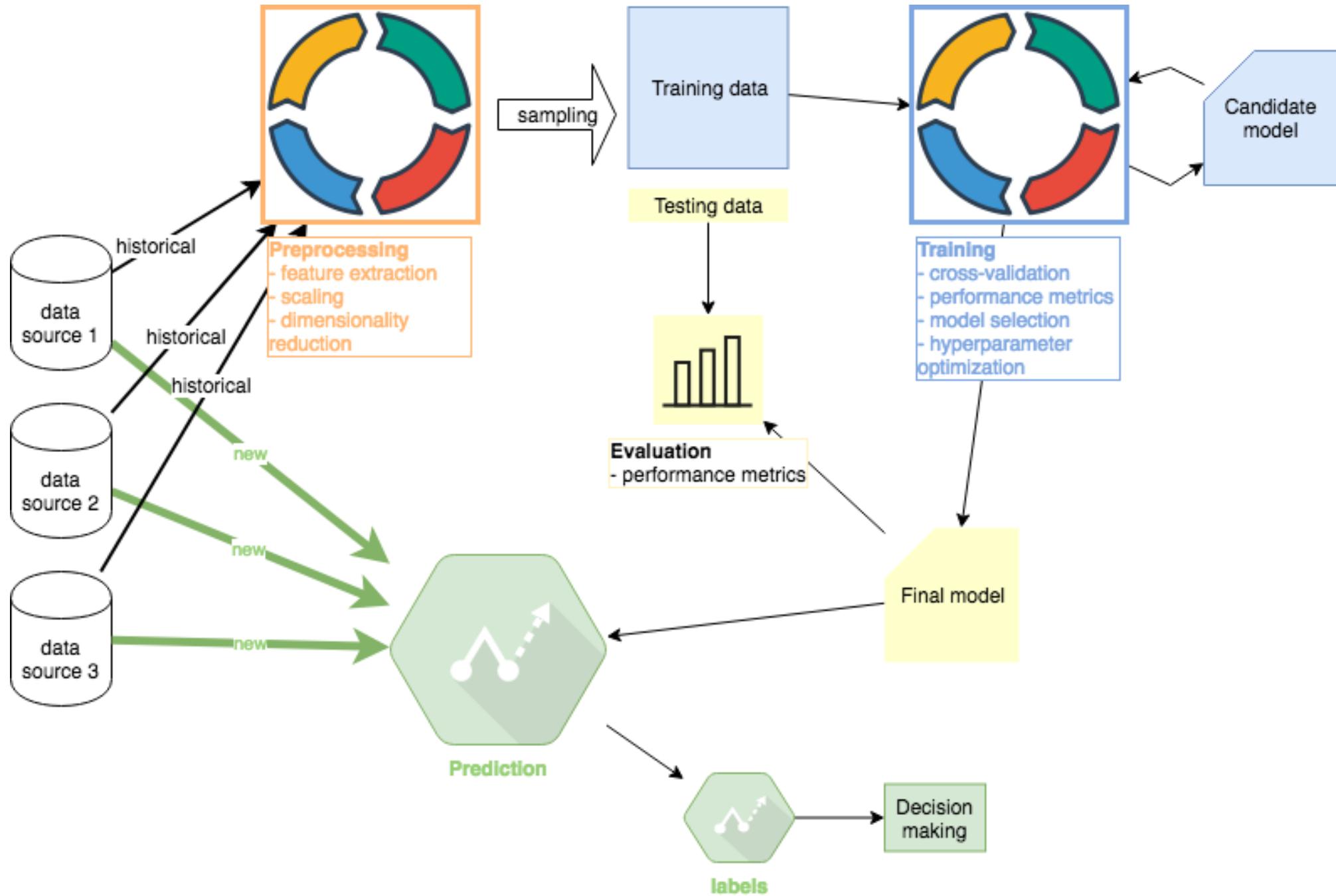
## 3.4. Social credit scores - China government



## 4. Workflow để xây dựng được mô hình Machine Learning

# Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống
  2. Lịch sử của Machine Learning
  3. Ứng dụng của Machine Learning trong thực tế
    - 3.1. Autopilot - Tesla
    - 3.2. Siri - Apple
    - 3.3. AlphaGo - Google Deepmind
    - 3.4. Social credit scores - China government
  4. Workflow để xây dựng được mô hình Machine Learning
  5. Các nhóm mô hình Machine Learning
    - 5.1. Nhóm mô hình Học có giám sát - Supervised learning
      - 5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification
      - 5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression
    - 5.2. Nhóm mô hình Học không giám sát - Unsupervised learning
      - 5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering
      - 5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



## 5. Các nhóm mô hình Machine Learning

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

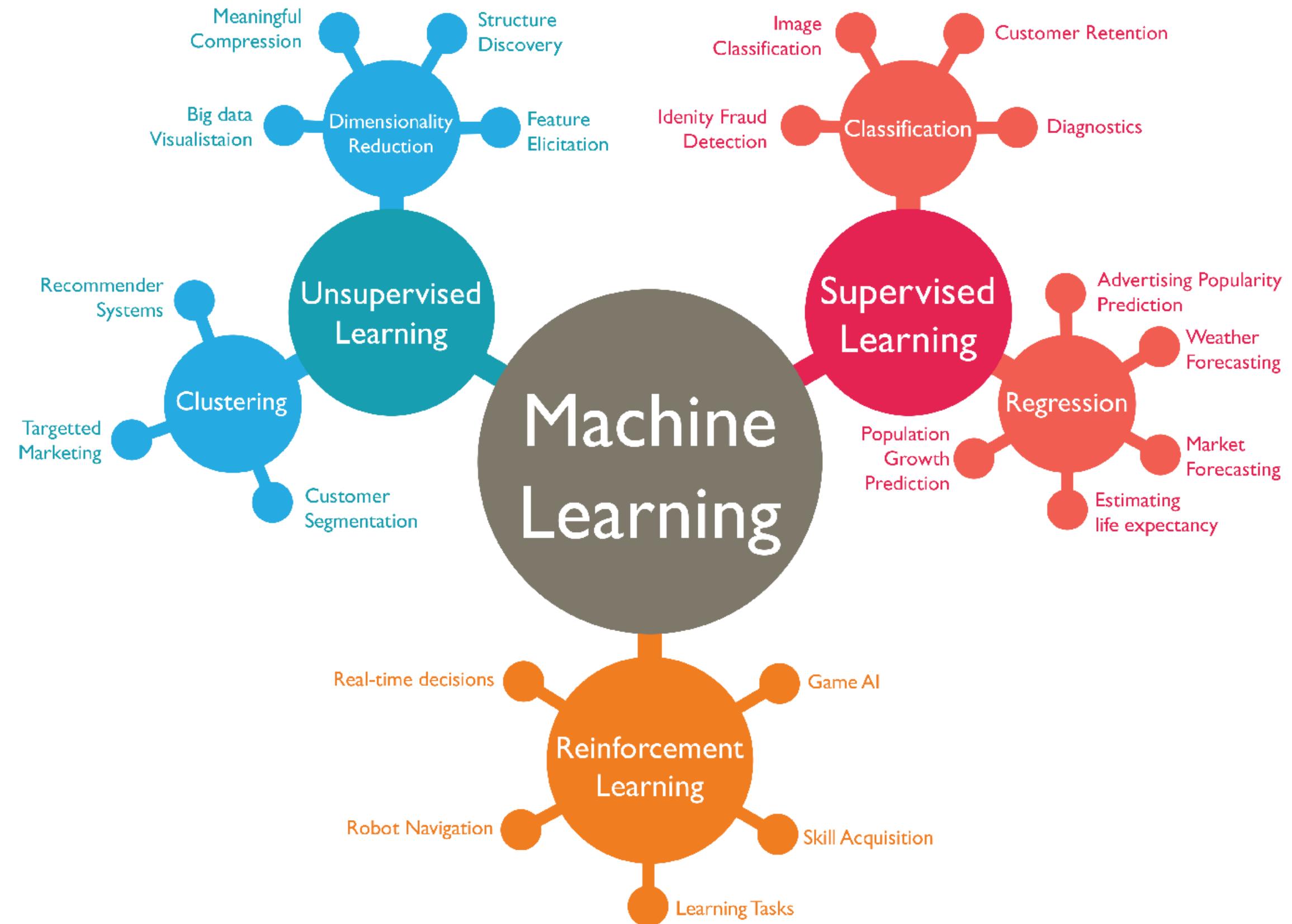
5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



### 5.1. Nhóm mô hình Học có giám sát - Supervised learning

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

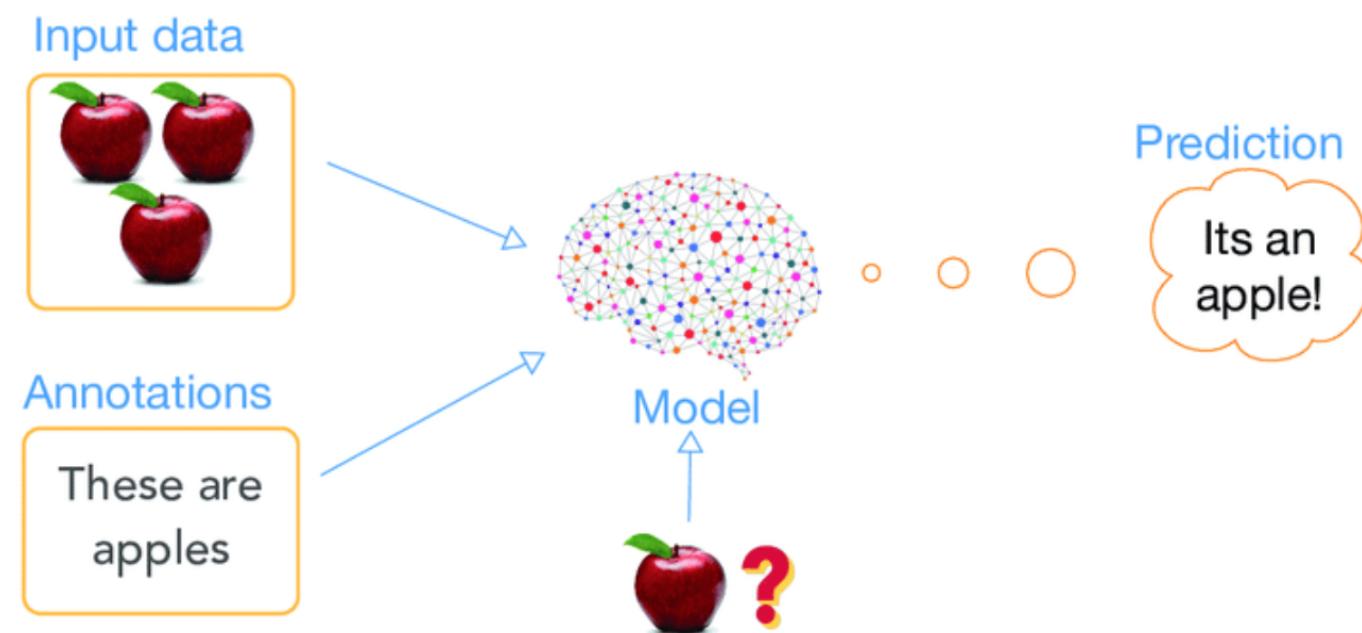
5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension

Production

## supervised learning



### 5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

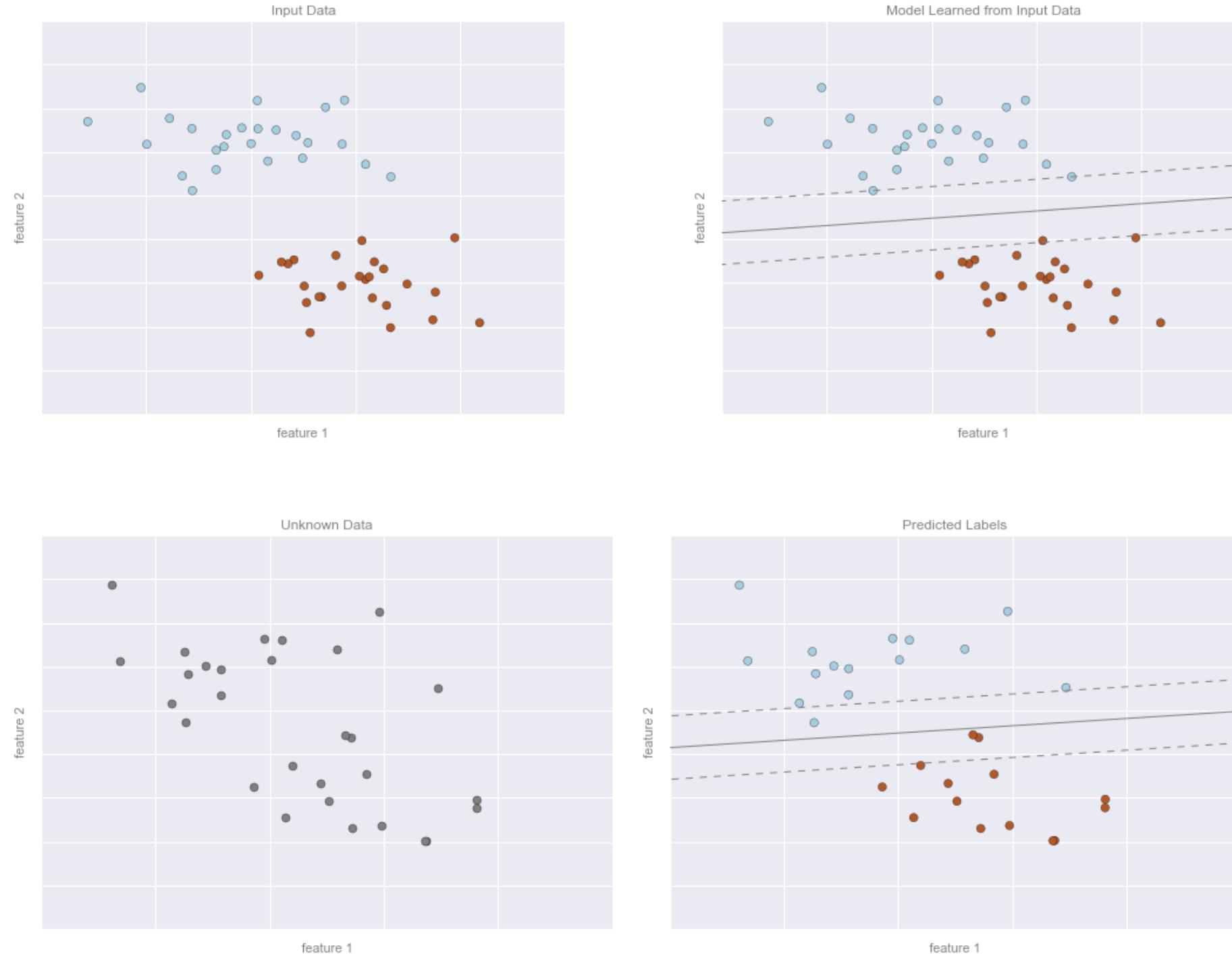
5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



### 5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

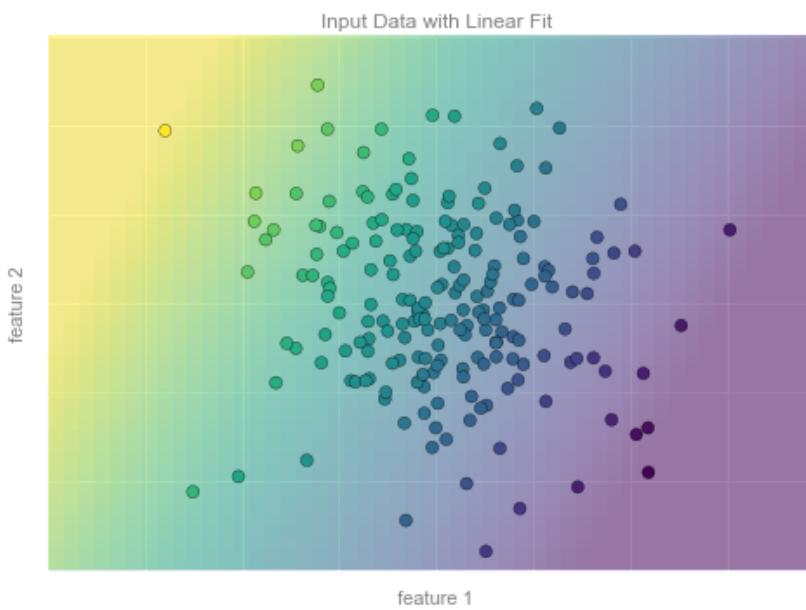
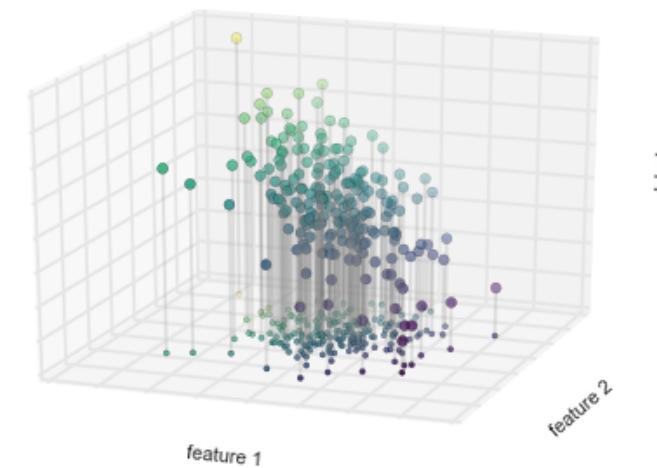
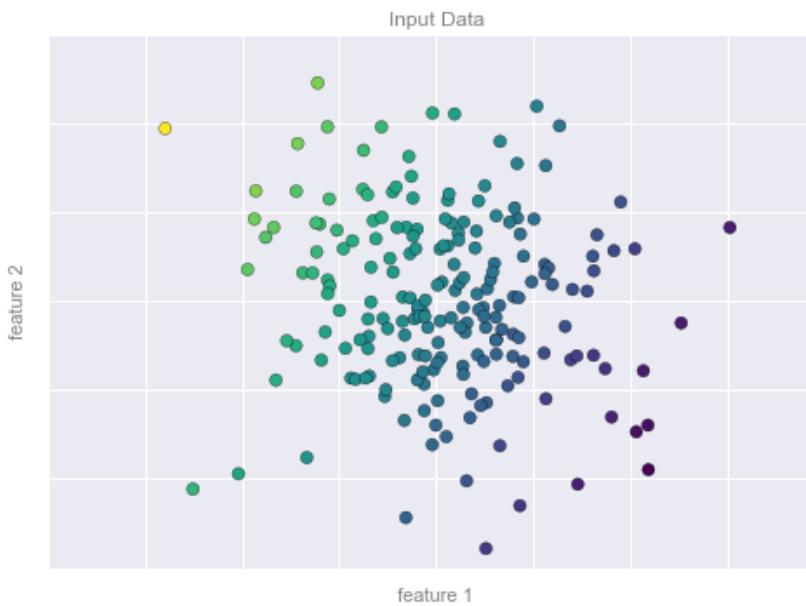
5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

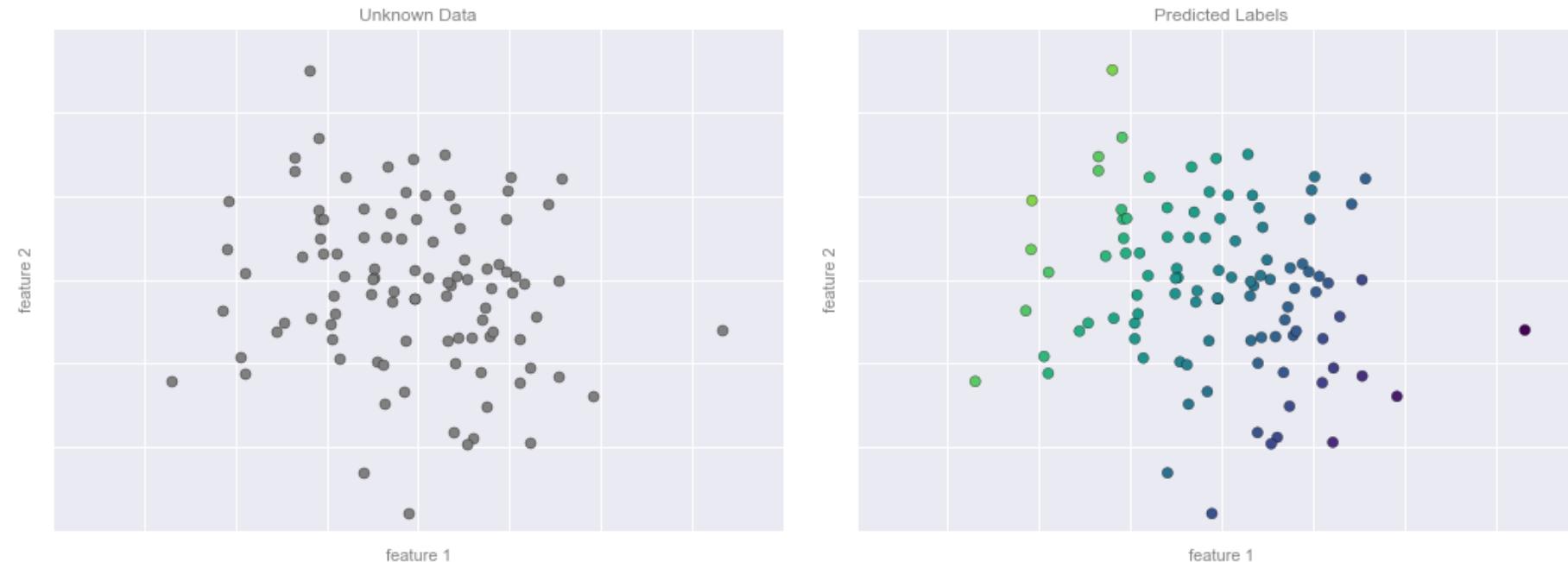
5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

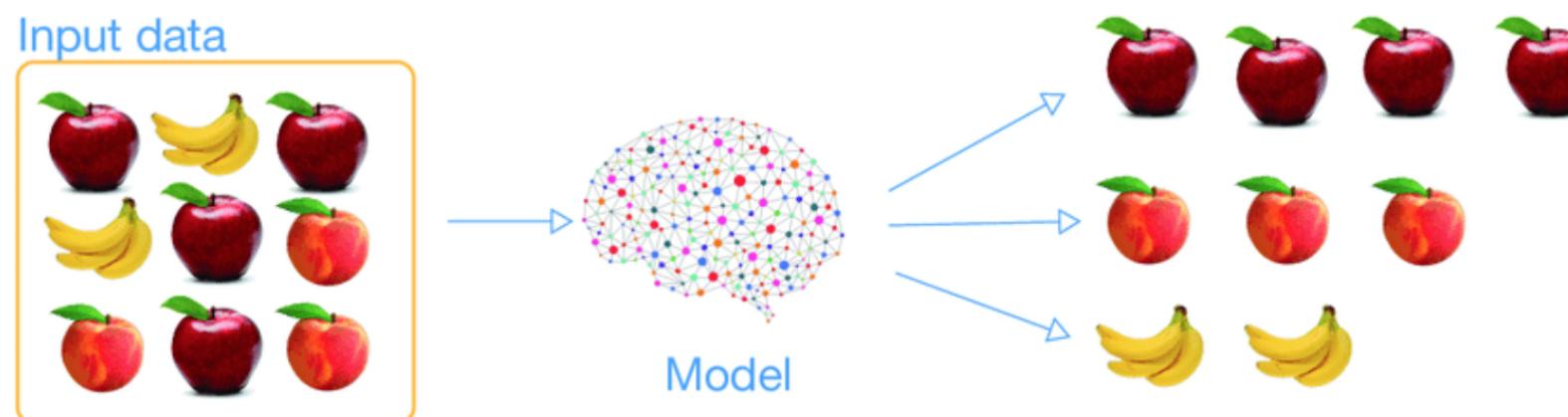
5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension

Production



## 5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

### unsupervised learning



### 5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

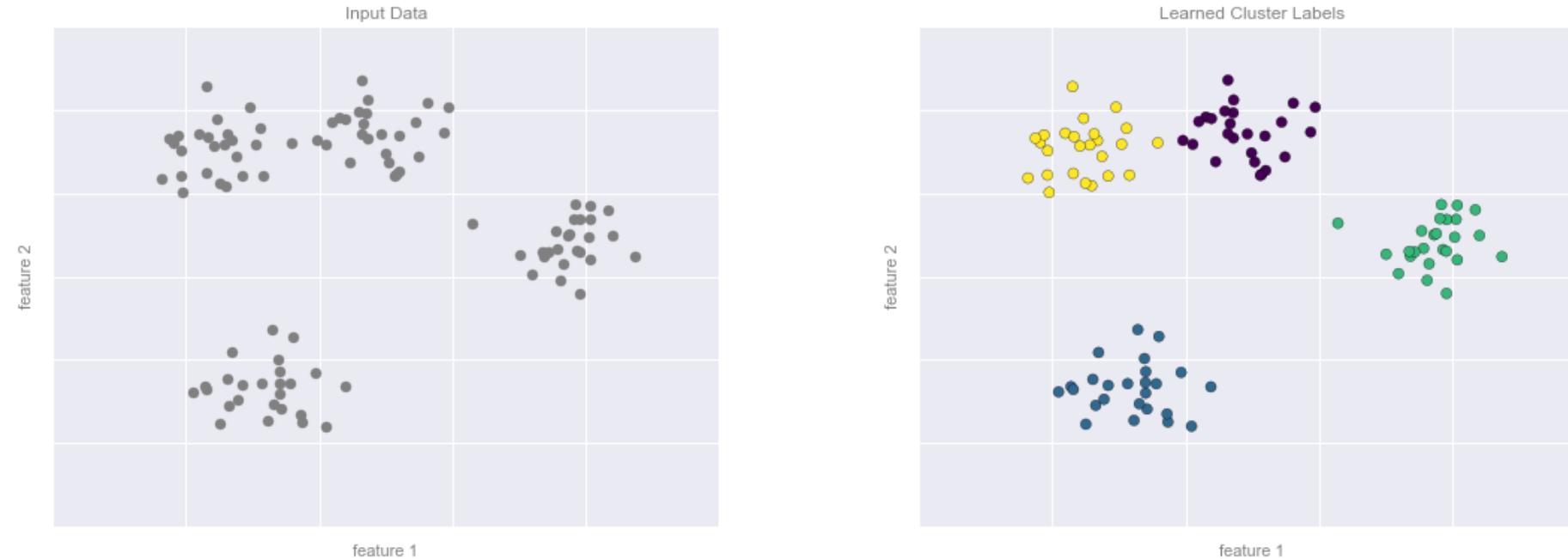
5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

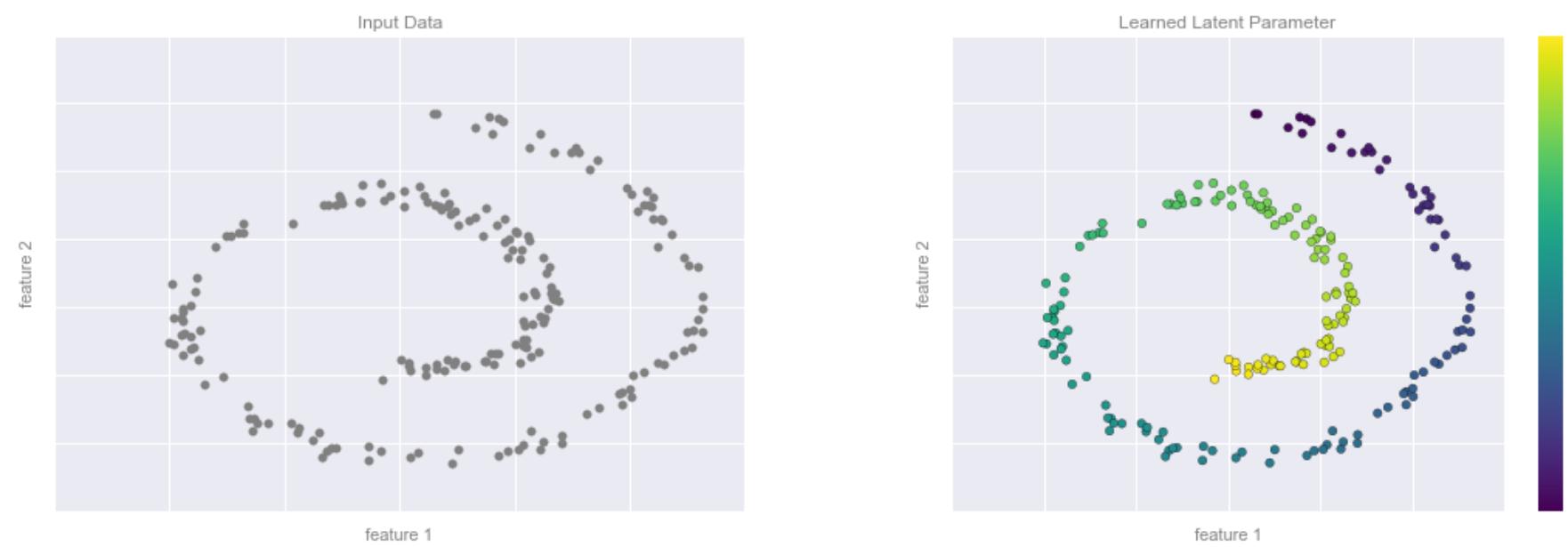
5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



### 5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



### 5.3. Nhóm mô hình Học tăng cường - Reinforcement learning

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

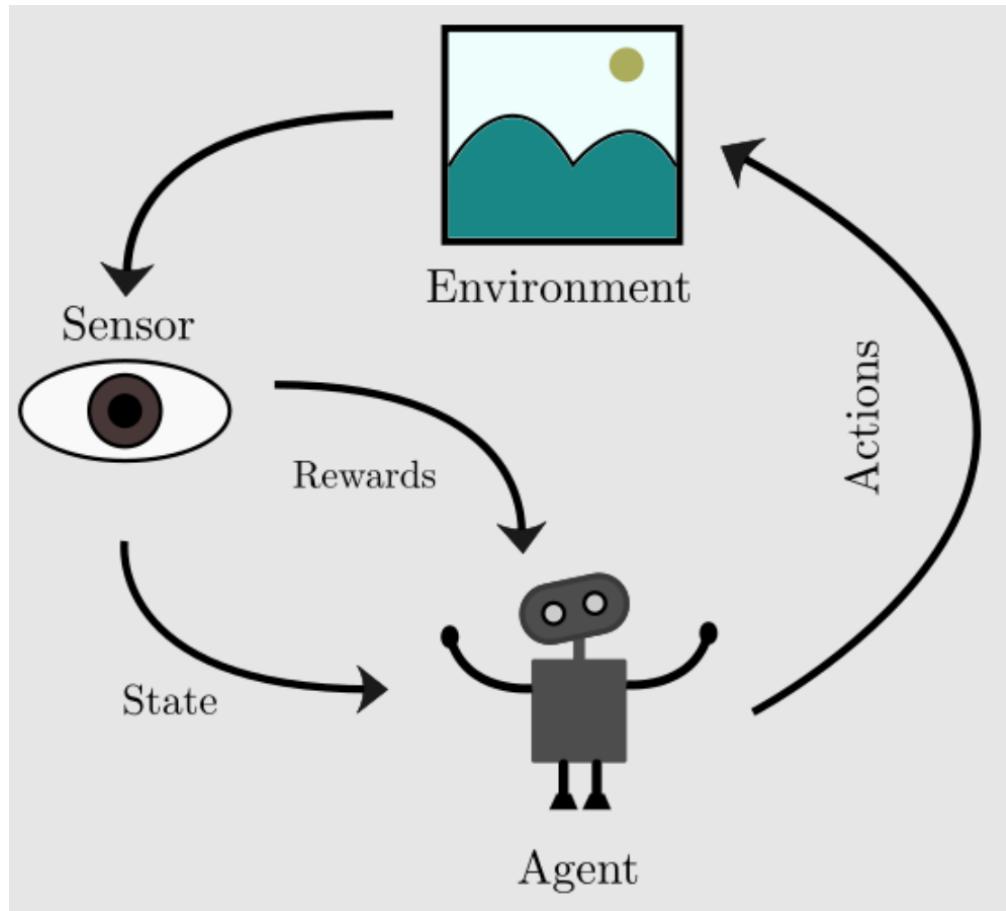
5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension

Production



## 6. Thư viện Sklearn

Ta có thể cài thư viện Sklearn vào môi trường conda thông qua câu lệnh

```
conda install -c anaconda scikit-learn -y
```

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction



The image shows the front page of the Scikit-Learn Cheat Sheet. It features the DataCamp logo at the top left. The title "Python For Data Science" and subtitle "Scikit-Learn Cheat Sheet" are prominently displayed in the center. Below the title, there's a call to action "Learn Scikit-Learn online at [www.DataCamp.com](http://www.DataCamp.com)". The background is orange.

## Scikit-learn

Scikit-learn is an open source Python library that implements a range of machine learning, preprocessing, cross-validation and visualization algorithms using a unified interface.

### A Basic Example

```
>>> from sklearn import neighbors, datasets, preprocessing
>>> from sklearn.model_selection import train_test_split
>>> from sklearn.metrics import accuracy_score
>>> iris = datasets.load_iris()
>>> X, y = iris.data[:, :2], iris.target
>>> X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=33)
>>> scaled = preprocessing.StandardScaler().fit(X_train)
>>> X_train = scaled.transform(X_train)
>>> X_test = scaled.transform(X_test)
>>> knn = neighbors.KNeighborsClassifier(n_neighbors=5)
>>> knn.fit(X_train, y_train)
>>> y_pred = knn.predict(X_test)
>>> accuracy_score(y_test, y_pred)
```

### Loading The Data

Also see NumPy & Pandas

Your data needs to be numeric and stored as NumPy arrays or SciPy sparse matrices. Other types that are convertible to numeric arrays, such as Pandas DataFrame, are also acceptable.

```
>>> import numpy as np
>>> X = np.random.random((10,5))
>>> y = np.array(['M','M','F','F','M','M','M','F','F','F'])
>>> X[X < 0.7] = 0
```

### Training And Test Data

```
>>> from sklearn.model_selection import train_test_split
>>> X_train, X_test, y_train, y_test = train_test_split(X,
...                                                    y,
...                                                    random_state=0)
```

### Model Fitting

**Supervised learning**

```
>>> lr.fit(X, y) #Fit the model to the data
>>> knn.fit(X_train, y_train)
>>> svc.fit(X_train, y_train)
```

**Unsupervised Learning**

```
>>> kmeans.fit(X_train) #Fit the model to the data
>>> pca_model = pca.fit_transform(X_train) #Fit to data, then transform it
```

### Prediction

**Supervised Estimators**

```
>>> y_pred = svc.predict(np.random.random((2,5))) #Predict labels
>>> y_pred = lr.predict(X_test) #Predict labels
>>> y_pred = knn.predict_proba(X_test) #Estimate probability of a label
```

**Unsupervised Estimators**

```
>>> y_pred = kmeans.predict(X_test) #Predict labels in clustering algos
```

## Preprocessing The Data

### Standardization

```
>>> from sklearn.preprocessing import StandardScaler
>>> scaler = StandardScaler().fit(X_train)
>>> standardized_X = scaler.transform(X_train)
>>> standardized_X_test = scaler.transform(X_test)
```

### Normalization

```
>>> from sklearn.preprocessing import Normalizer
>>> scalar = Normalizer().fit(X_train)
>>> normalized_X = scalar.transform(X_train)
>>> normalized_X_test = scalar.transform(X_test)
```

### Binarization

```
>>> from sklearn.preprocessing import Binarizer
>>> binaryc = Binarizer(threshold=0.8).fit(X)
>>> binary_X = binaryc.transform(X)
```

### Encoding Categorical Features

```
>>> from sklearn.preprocessing import LabelEncoder
>>> enc = LabelEncoder()
>>> y = enc.fit_transform(y)
```

### Imputing Missing Values

```
>>> from sklearn.preprocessing import Imputer
>>> imp = Imputer(missing_values=0, strategy='mean', axis=0)
>>> imp.fit_transform(X_train)
```

### Generating Polynomial Features

```
>>> from sklearn.preprocessing import PolynomialFeatures
>>> poly = PolynomialFeatures(5)
>>> poly.fit_transform(X)
```

## Create Your Model

### Supervised Learning Estimators

#### Linear Regression

```
>>> from sklearn.linear_model import LinearRegression
>>> lr = LinearRegression(normalize=True)
```

#### Support Vector Machines (SVM)

```
>>> from sklearn.svm import SVC
>>> svc = SVC(kernel='linear')
```

#### Naive Bayes

```
>>> from sklearn.naive_bayes import GaussianNB
>>> gnb = GaussianNB()
```

#### KNN

```
>>> from sklearn import neighbors
>>> knn = neighbors.KNeighborsClassifier(n_neighbors=5)
```

### Unsupervised Learning Estimators

#### Principal Component Analysis (PCA)

```
>>> from sklearn.decomposition import PCA
>>> pca = PCA(n_components=0.95)
```

#### K Means

```
>>> from sklearn.cluster import KMeans
>>> kmeans = KMeans(n_clusters=3, random_state=0)
```

## Evaluate Your Model's Performance

### Classification Metrics

**Accuracy Score**

```
>>> knn.score(X_test, y_test) #Estimator score method
>>> from sklearn.metrics import accuracy_score #Metric scoring functions
>>> accuracy_score(y_test, y_pred)
```

#### Classification Report

```
>>> from sklearn.metrics import classification_report #Precision, recall, f1-score and support
>>> print(classification_report(y_test, y_pred))
```

#### Confusion Matrix

```
>>> from sklearn.metrics import confusion_matrix
>>> print(confusion_matrix(y_test, y_pred))
```

### Regression Metrics

#### Mean Absolute Error

```
>>> from sklearn.metrics import mean_absolute_error
>>> y_true = [3, -0.5, 2]
>>> mean_absolute_error(y_true, y_pred)
```

#### Mean Squared Error

```
>>> from sklearn.metrics import mean_squared_error
>>> mean_squared_error(y_true, y_pred)
```

#### R<sup>2</sup> Score

```
>>> from sklearn.metrics import r2_score
>>> r2_score(y_true, y_pred)
```

### Clustering Metrics

#### Adjusted Rand Index

```
>>> from sklearn.metrics import adjusted_rand_score
>>> adjusted_rand_score(y_true, y_pred)
```

#### Homogeneity

```
>>> from sklearn.metrics import homogeneity_score
>>> homogeneity_score(y_true, y_pred)
```

#### V-measure

```
>>> from sklearn.metrics import v_measure_score
>>> metrics.v_measure_score(y_true, y_pred)
```

### Cross-Validation

```
>>> from sklearn.cross_validation import cross_val_score
>>> print(cross_val_score(knn, X_train, y_train, cv=4))
>>> print(cross_val_score(lr, X, y, cv=2))
```

## Tune Your Model

### Grid Search

```
>>> from sklearn.grid_search import GridSearchCV
>>> params = {"n_neighbors": np.arange(1,3),
...            "metric": ["euclidean", "cityblock"]}
>>> grid = GridSearchCV(Cestimator=knn,
...                      param_grid=params)
>>> grid.fit(X_train, y_train)
>>> print(grid.best_score_)
>>> print(grid.best_estimator_.n_neighbors)
```

### Randomized Parameter Optimization

```
>>> from sklearn.grid_search import RandomizedSearchCV
>>> params = {"n_neighbors": range(1,5), "weights": ["uniform", "distance"]}
>>> rsearch = RandomizedSearchCV(Cestimator=knn, param_distributions=params,
...                               cv=4, n_iter=8, random_state=5)
>>> rsearch.fit(X_train, y_train)
>>> print(rsearch.best_score_)
```

 Learn Data Skills Online at [www.DataCamp.com](http://www.DataCamp.com)

## Giới thiệu chung về Machine Learning

1. Tư tưởng của Machine Learning và Lập trình truyền thống

2. Lịch sử của Machine Learning

3. Ứng dụng của Machine Learning trong thực tế

3.1. Autopilot - Tesla

3.2. Siri - Apple

3.3. AlphaGo - Google Deepmind

3.4. Social credit scores - China government

4. Workflow để xây dựng được mô hình Machine Learning

5. Các nhóm mô hình Machine Learning

5.1. Nhóm mô hình Học có giám sát - Supervised learning

5.1.1. Nhóm mô hình giải quyết bài toán Phân lớp - Classification

5.1.2. Nhóm mô hình giải quyết bài toán Hồi quy - Regression

5.2. Nhóm mô hình Học không giám sát - Unsupervised learning

5.2.1. Nhóm mô hình giải quyết bài toán Phân cụm - Clustering

5.2.2. Nhóm mô hình giải quyết bài toán Giảm chiều dữ liệu - Dimension Reduction

