



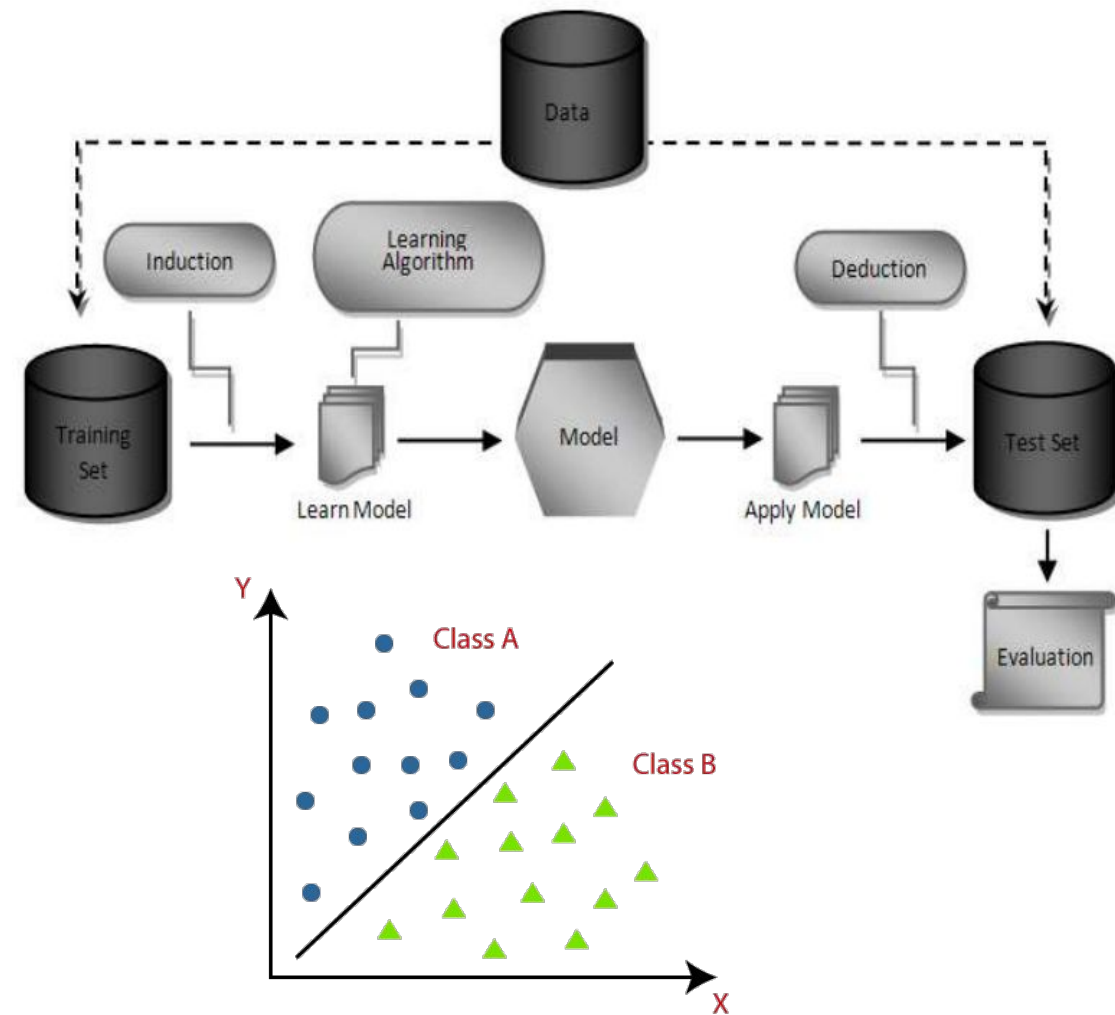
ARTIFICIAL INTELLIGENCE

2nd assignment

Juan Bellon - up210908142
Luísa Araújo - up201904996
Nuno Castro - up202003324

Problem Description

- On a daily basis, a large amount of transactions are made, causing it hard to identify which ones are fraudulent.
- Our problem is a Classification Supervised Learning problem.
- The given dataset is represented as a record data matrix.



Task	Attribute set (x)	Class Label (y)
Identify if a transaction is fraud or not	Distance from home (where the transaction happened), Distance from the last transaction, Ratio of purchased price to median purchased price, Repeat Retailer (Bought from that retailer before), Used chip (used credit card), Used pin (used pin number), Online order	Fraud or not

- **There are eight attributes in this dataset:**

Attribute	Type	Properties	Discrete vs Continuous
Distance from home	Ratio	Distinctness, Order, Meaningful differences and Meaningful ratios	Continuous
Distance from last transaction	Ratio	Distinctness, Order, Meaningful differences and Meaningful ratios	Continuous
Ratio of purchased price to median purchased price	Ratio	Distinctness, Order, Meaningful differences and Meaningful ratios	Continuous
Repeat Retailer	Nominal (binary)	Distinctness	Discrete
Used chip (used credit card)	Nominal (binary)	Distinctness	Discrete
Used pin (used pin number)	Nominal (binary)	Distinctness	Discrete
Online Order	Nominal (binary)	Distinctness	Discrete
Fraud	Nominal (binary)	Distinctness	Discrete

Packages and algorithms

Packages	Algorithms
<ul style="list-style-type: none">• NumPy• SciPy• Matplotlib• IPython• Pandas• Scikit learn• OpenAI Gym• TensorFlow• Keras• Machine Learning Agents• Seaborn	<ul style="list-style-type: none">• Decision Tree• Neural Networks• K-Nearest Neighbour Classifier• Support Vector Machine• Logistic Regression• Random Forest• Naive Bayes• K-Means

Implementation

Data analysis and preprocessing:

- Verify the existence of missing or null data to the proceed with plotting.

```
movieFrame.isna().sum()
```

```
movieFrame.isnull().sum()
```

```
import matplotlib.pyplot as plt  
import seaborn as sb  
  
sb.pairplot(movieFrame.dropna(), hue='fraud')
```

Related work

- <https://towardsdatascience.com/credit-card-fraud-detection-using-machine-learning-python-5b098d4a8edc>
- <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.675.408&rep=rep1&type=pdf>
- http://www.iaeng.org/publication/IMECS2011/IMECS2011_pp442-447.pdf
- B. N. V. Madhubabu¹, T. Vyshnavi², K. Ashok², “Credit Card Fraud Detection Algorithm using Decision Trees-based Random Forest Classifier”, Turkish Journal of Computer and Mathematics Education

Data Pre-processing

1. Check and remove duplicate values;
2. Look for outliers and remove them;
3. Remove NA values;
4. Remove NULL values;

Developed Models

1. Support Vector Machine
2. Random Forest
3. Decision Tree
4. Logistic Regression

Evaluation

The algorithms employed in this study were:

- Logistic regression
- SVM
- Decision Tree
- Random Forest

Regarding evaluation we opted to calculate the following metrics:

- Runtime execution
- Accuracy
- F-Measure
- Recall
- Support
- Precision

Comparison

