|  |  |
| --- | --- |
| Faculty of Computer & Information Sciences  Ain Shams University  Subject: (SCC 234)  Statistical Analysis and Applications  Year: (2nd year) undergraduate  Academic year: 2nd term 2019-2020 |  |

**Research Topic (C)**

**Title: Using Probability to Build Decision Trees for Classification**

**Introduction**

Life is all about making decisions. We keep on making decisions in both voluntary and involuntary state. This phenomenon has influenced a wide area of machine learning, covering both classification and regression. But in this research I’ll talk in detail about Decision Trees and Classification and explain examples with real datasets. In decision analysis, a decision tree can be utilized to outwardly and expressly speak to choices and decisions . As the name goes, it utilizes a tree-like model of decisions.

**What ‘s a Decision Tree ?** (Ref. 1)

A decision tree is a graph or outline that individuals use to decide a game-plan or show a statistical probability. It forms the outline of the namesake woody plant, usually upright but sometimes lying on its side. Each branch or part of the decision tree speaks to a possible decision, result, outcome or response .The farthest branches on the tree represent the end results.

Individuals use decision trees to clarify and find an answer to a complex problem. Decision trees are frequently employed in determining a course of action in finance, investing, or business.

By displaying a sequence of steps, decision trees give people an effective and easy way to visualize and understand the potential options of a decision and its range of possible outcomes. The decision tree also helps people identify every potential option and weigh each course of action against the risks and rewards each option can yield.

An organization may use decision trees as a kind of decision support system. The structured model allows the reader of the chart to see how and why one choice may lead to the next, with the use of the branches indicating mutually exclusive options. The structure allows users to take a problem with multiple possible solutions and to display those solutions in a simple, easy-to-understand format that also shows the relationship between different events or decisions.

If a person uses a decision tree to make a decision, they look at each final outcome and assess the benefits and drawbacks. The tree itself can span as long or as short as needed in order to come to a proper conclusion.

**Mathematical formulation of Decision Trees.** (Ref. 2)

**Information gain** IG(A) is the measure of the difference in entropy from before to after the set S is split on an attribute A. In other words, how much uncertainty in S was reduced after splitting set S on attribute A.

IG(A,S) = H(S) -

Where,

• H(S) – Entropy of set S

• T – The subsets made from parting set S by attribute A.

• P(t) – The proportion of the number of elements in t to the number of elements in set S

• H(t) – Entropy of subset t.

In ID3, information gain can be calculated (instead of entropy) for each remaining attribute. The attribute with the largest information gain is used to split the set S o()n this iteration.

**Entropy** H(S) is a measure of the amount of uncertainty in the (data) set S (i.e. entropy characterizes the (data) set S).

H(S) =

Where,

* S – is the current (data) set for which entropy is being calculated (changes every iteration of the ID3 algorithm)
* c – Set of classes in S c = { yes, no}
* p(c) – The proportion of the number of elements in class c to the number of elements in set S.

When H(S) = 0, the set S is classified (i.e. all elements in S are of the same class ).

In ID3, entropy is determined for each remaining attribute. The quality with the littlest entropy is utilized to part the set S on this emphasis.

The higher the entropy, the higher the possibility to improve the classification here.

**What are the terms in the Decision Trees mean and the intuitions behind them?** (Ref. 2)

Important Terminology related to Decision Trees

* **Root Node:** It represents the entire population or sample and this further gets divided into two or more homogeneous sets.
* **Splitting:** It is a process of dividing a node into two or more sub-nodes.
* **Decision Node:** At the point when a sub-hub splits into further sub-nodes, then it is known as the decision node.
* **Leaf / Terminal Node:** Nodes do not part is called Leaf or Terminal node.
* **Pruning:** When we remove sub-nodes of a decision node, this procedure is called pruning. You can say the contrary procedure of parting.
* **Branch / Sub-Tree:** A subsection of the whole tree is called branch or sub-tree.
* **Parent and Child Node:** A node, which is partitioned into sub-nodes is known as a parent node of sub-nodes though sub-nodes are the child of a parent node.

**Explain the ID3 method to construct the decision tress how it is related to probability?**

ID3 represents Iterative Dichotomiser 3

It is an arrangement calculation that follows an avaricious methodology by choosing a best trait that yields most extreme Information Gain(IG) or least Entropy(H).

## What are the steps in ID3 algorithm?

1. Calculate entropy for dataset.
2. For each attribute/feature

* Calculate entropy for all its categorical values.
* Calculate information gain for the feature.

1. Find the feature with maximum information gain.
2. Repeat it until we get the desired tree.

**The Relation between ID3 and Probability**

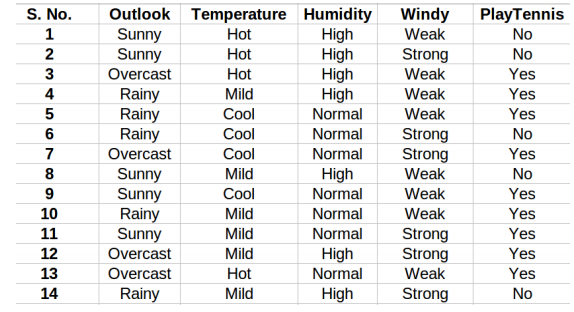
By Using Probability, we can calculate the entropy and Information Gain that are considered as the mathematical formulas of ID3.

## Example shows Use of ID3 algorithm on a data

We'll discuss it mathematically in the next part.

**Working a numerical example of how Decision Trees are built.**

**Ex. ;** Make a decision tree that predicts whether tennis will be played on the day?



**Solution in detail:**

**Step 1 : We’ll create a Root Node**

* How to choose the root node?

The attribute that best classifies the training data, use this

attribute at the root of the tree.

* How to choose the best attribute?

So from here, ID3 algorithm begins

- Calculate **Entropy** (The amount of uncertainty in dataset):

Entropy =

- Calculate **Average Information :**

I (Attribute) = Entropy (A)

- Calculate **Information Gain:** (Difference in Entropy before and after splitting dataset on attribute A)

**Gain = Entropy (S) – I (Attribute)**

1.compute the entropy for data-set Entropy(s)

2.for every attribute/feature:

* calculate entropy for all other values Entropy(a)
* take average information entropy for the current attribute
* calculate gain for the current attribute

3. pick the highest gain attribute.

4. Repeat until we get the tree we desired.

P = 9 , N = 5 , Total = 14

* Calculate **Entropy (S)**

Entropy =

Entropy(S) =

Entropy(S) = = **0.940**

For Each Attribute: (let say **Outlook**)

* Calculate Entropy for each values, i.e. for ‘Sunny’, ‘Rainy’, ‘Overcast’

|  |  |
| --- | --- |
| **Outlook** | **Play Tennis** |
| Sunny | No |
| Sunny | No |
| Sunny | No |
| Sunny | Yes |
| Sunny | Yes |

|  |  |
| --- | --- |
| **Outlook** | **Play Tennis** |
| Rainy | Yes |
| Rainy | Yes |
| Rainy | No |
| Rainy | Yes |
| Rainy | Yes |

|  |  |
| --- | --- |
| **Outlook** | **Play Tennis** |
| Overcast | Yes |
| Overcast | Yes |
| Overcast | Yes |
| Overcast | Yes |

* Calculate **Entropy (Outlook = ‘Value’):**

E (Outlook = sunny) = = **0.971**

E (Outlook = overcast)=

E (Outlook = rainy) = = **0.971**

* Calculate **Average Information Entropy**:

I (Outlook) = Entropy (Outlook = Sunny) +

Entropy (Outlook = Rainy) +

Entropy (Outlook = Overcast)

I (Outlook) = \* 0.971 + \* 0.971 + \* 0 = 0.693

* Calculate **Gain**: attribute is Outlook

Gain = Entropy (S) – I (Attribute)

Entropy (S) = 0.940

Gain (Outlook) = 0.940 – 0.693 = 0.247

|  |  |  |  |
| --- | --- | --- | --- |
| **Outlook** | **P** | **n** | **Entropy** |
| Sunny | 2 | 3 | 0.971 |
| Rainy | 3 | 2 | 0.971 |
| Overcast | 4 | 0 | 0 |

For Each Attribute: (let say **Temperature**)

* Calculate Entropy for each Temp, i.e. for ‘Hot’, ‘Mild, ‘Cool’

|  |  |
| --- | --- |
| **Temperature** | **Play Tennis** |
| Mild | Yes |
| Mild | No |
| Mild | Yes |
| Mild | Yes |
| Mild | Yes |
| Mild | No |

|  |  |
| --- | --- |
| **Temperature** | **Play Tennis** |
| Hot | No |
| Hot | No |
| Hot | Yes |
| Hot | Yes |

|  |  |
| --- | --- |
| **Temperature** | **Play Tennis** |
| Cool | Yes |
| Cool | No |
| Cool | Yes |
| Cool | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Temperature** | **p** | **n** | **Entropy** |
| Hot | 2 | 2 | 1 |
| Mild | 4 | 2 | 0.918 |
| Cool | 3 | 1 | 0.811 |

* Calculate **Average Information Entropy**:

I (Temperature) = Entropy (Temperature = Hot) +

Entropy (Temperature = Mild) +

Entropy (Temperature = Cool)

I (Temperature) = \* 0.811 => 0.911

* Calculate **Gain**: attribute is Temperature

Gain = Entropy (S) – I (Attribute)

Entropy (S) = 0.940

Gain (Temperature) = 0.940 – 0.911 = 0.029

For Each Attribute: (let say **Humidity**)

|  |  |
| --- | --- |
| **Humidity** | **Play Tennis** |
| Normal | Yes |
| Normal | No |
| Normal | Yes |
| Normal | Yes |
| Normal | Yes |
| Normal | Yes |
| Normal | Yes |

* Calculate Entropy for each Temp, i.e. for ‘High’, ‘Normal’

|  |  |
| --- | --- |
| **Humidity** | **Play Tennis** |
| High | No |
| High | No |
| High | Yes |
| High | Yes |
| High | No |
| High | Yes |
| High | No |

|  |  |  |  |
| --- | --- | --- | --- |
| **Humidity** | **p** | **n** | **Entropy** |
| High | 3 | 4 | 0.985 |
| Normal | 6 | 1 | 0.591 |

* Calculate **Average Information Entropy**:

I (**Humidity**) = Entropy (Humidity = High) +

Entropy (Humidity = Normal)

I (Humidity) = \* 0.811 => 0.788

* Calculate **Gain**: attribute is Humidity

Gain = Entropy (S) – I (Attribute)

Entropy (S) = 0.940

Gain (Humidity) = 0.940 – 0.788 = 0.152

For Each Attribute: (let say **Windy**)

* Calculate Entropy for each Windy, i.e. for ‘Strong’, ‘Weak’

|  |  |
| --- | --- |
| **Windy** | **Play Tennis** |
| Weak | No |
| Weak | Yes |
| Weak | Yes |
| Weak | Yes |
| Weak | No |
| Weak | Yes |
| Weak | Yes |
| Weak | Yes |

|  |  |
| --- | --- |
| **Windy** | **Play Tennis** |
| Strong | No |
| Strong | No |
| Strong | Yes |
| Strong | Yes |
| Strong | Yes |
| Strong | No |

|  |  |  |  |
| --- | --- | --- | --- |
| **Windy** | p | n | **Entropy** |
| Strong | 3 | 3 | 1 |
| Weak | 6 | 2 | 0.811 |

* Calculate **Average Information Entropy**:

I (**Windy**) = Entropy (Windy = Strong) +

Entropy (Windy = )

I (Windy) = => 0.892

* Calculate **Gain**: attribute is Windy

Gain = Entropy (S) – I (Attribute)

Entropy (S) = 0.940

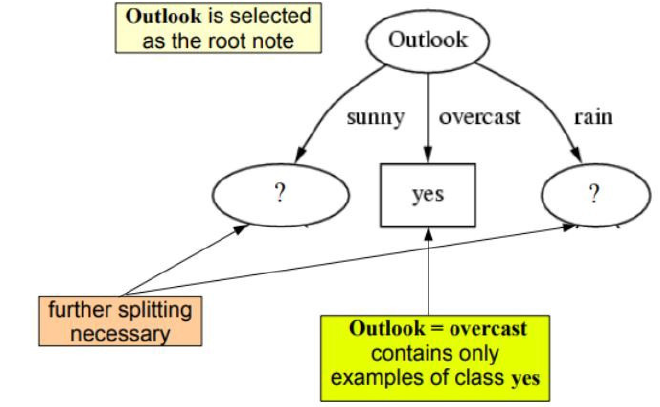
Gain (Windy) = 0.940 – 0.892 = 0.048

**Pick The Highest Gain Attribute.**

|  |  |
| --- | --- |
| **Attributes** | **Gain** |
| Outlook | 0.247 |
| Temperature | 0.029 |
| Humidity | 0.152 |
| Windy | 0.048 |

So 🡺  **Root Node: OUTLOOK 🡸**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play Tennis** |
| Overcast | Hot | High | Weak | Yes |
| Overcast | Cool | Normal | Strong | Yes |
| Overcast | Mild | High | Strong | Yes |
| Overcast | Hot | Normal | Weak | Yes |

****

* **We’ll repeat the same thing for Sub-Trees till we get the tree.**

Outlook: Sunny **P = 2 , N = 3, Total = 5**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play Tennis** |
| Sunny | Hot | High | Weak | No |
| Sunny | Hot | High | Strong | No |
| Sunny | Mild | High | Weak | No |
| Sunny | Cool | Normal | Weak | Yes |
| Sunny | Mild | Normal | Strong | Yes |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play Tennis** |
| Rainy | Mild | High | Weak | Yes |
| Rainy | Cool | Normal | Weak | Yes |
| Rainy | Cool | Normal | Strong | No |
| Rainy | Mild | Normal | Weak | Yes |
| Rainy | Mild | High | Strong | No |

Outlook: Rainy **P = 3 , N = 2, Total = 5**

* **Entropy:**

Entropy =

Entropy (Sunny) = => **0.971**

For Each Attribute: (let say **Humidity**)

* Calculate Entropy for each Humidity, i.e. for ‘High’, ‘Normal’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Humidity** | **Play Tennis** |
| Sunny | High | No |
| Sunny | High | No |
| Sunny | High | No |
| Sunny | Normal | Yes |
| Sunny | Normal | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Humidity** | **p** | **n** | **Entropy** |
| high | 0 | 3 | 0 |
| normal | 2 | 0 | 0 |

* Calculate **Average Information Entropy**: I(Humidity) = 0
* Calculate **Gain:** Gain = 0.971

For Each Attribute: (let say **Windy**)

* Calculate Entropy for each Windy, i.e. for ‘Strong’, ‘Weak’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Windy** | **Play Tennis** |
| Sunny | Strong | No |
| Sunny | Strong | Yes |
| Sunny | Weak | No |
| Sunny | Weak | No |
| Sunny | Weak | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Windy** | **p** | **n** | **Entropy** |
| Strong | 1 | 1 | 1 |
| Weak | 1 | 2 | 0.918 |

* Calculate **Average Information Entropy**: I(Windy) = 0.951
* Calculate **Gain:** Gain = 0.020

For Each Attribute: (let say **Temperature**)

* Calculate Entropy for each Temp, i.e. for ‘Cool’ ,’Hot’, ‘Mild’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Temperature** | **Play Tennis** |
| Sunny | Cool | Yes |
| Sunny | Hot | No |
| Sunny | Hot | No |
| Sunny | Mild | No |
| Sunny | Mild | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Temperature** | **p** | **n** | **Entropy** |
| Cool | 1 | 0 | 0 |
| Hot | 0 | 2 | 0 |
| Mild | 1 | 1 |  |

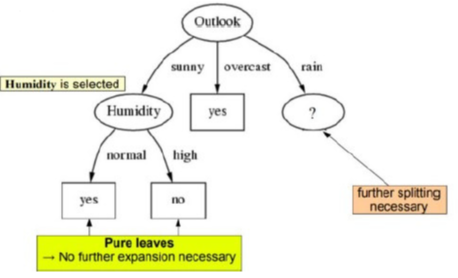
* Calculate **Average Information Entropy**: I(Temp) = 0.4
* Calculate **Gain:** Gain = 0.571

**Pick the highest gain attribute.**

|  |  |
| --- | --- |
| **Attributes** | **Gain** |
| Temperature | 0.571 |
| Humidity | 0.971 |
| Windy | 0.02 |

So 🡺  **Next Node in Sunny: Humidity 🡸**

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Humidity** | **Play Tennis** |
| Sunny | High | No |
| Sunny | High | No |
| Sunny | High | No |
| Sunny | Normal | Yes |
| Sunny | Normal | Yes |

****

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Windy** | **Play Tennis** |
| Rainy | Mild | High | Weak | Yes |
| Rainy | Cool | Normal | Weak | Yes |
| Rainy | Cool | Normal | Strong | No |
| Rainy | Mild | Normal | Weak | Yes |
| Rainy | Mild | High | Strong | No |

* **Entropy:**

Entropy =

Entropy (S Rainy) = => **0.971**

For Each Attribute: (let say **Humidity**)

* Calculate Entropy for each Humidity, i.e. for ‘High’, ‘Normal’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Humidity** | **Play Tennis** |
| Rainy | High | Yes |
| Rainy | High | No |
| Rainy | Normal | Yes |
| Rainy | Normal | No |
| Rainy | Normal | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **p** | **n** | **Entropy** |
| high | 1 | 1 | 1 |
| normal | 2 | 1 | 0.918 |

* Calculate **Average Information Entropy**: I(Humidity) = 0.951
* Calculate **Gain:** Gain = 0.020

For Each Attribute: (let say **Windy**)

* Calculate Entropy for each Windy, i.e. for ‘Strong’, ‘Weak’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Windy** | **Play Tennis** |
| Rainy | Strong | No |
| Rainy | Strong | No |
| Rainy | Weak | Yes |
| Rainy | Weak | Yes |
| Rainy | Weak | Yes |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **p** | **n** | **Entropy** |
| Strong | 0 | 2 | 0 |
| Weak | 3 | 0 | 0 |

* Calculate **Average Information Entropy**: I(Windy) = 0
* Calculate **Gain:** Gain = 0.971

For Each Attribute: (let say **Temperature**)

* Calculate Entropy for each Temp, i.e. for ‘Cool’ ,’Hot’, ‘Mild’

|  |  |  |
| --- | --- | --- |
| **Outlook** | **Temperature** | **Play Tennis** |
| Rainy | Mild | Yes |
| Rainy | Cool | Yes |
| Rainy | Cool | No |
| Rainy | Mild | Yes |
| Rainy | Mild | No |

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **p** | **n** | **Entropy** |
| Cool | 1 | 1 | 1 |
| Mild | 2 | 1 | 0.918 |

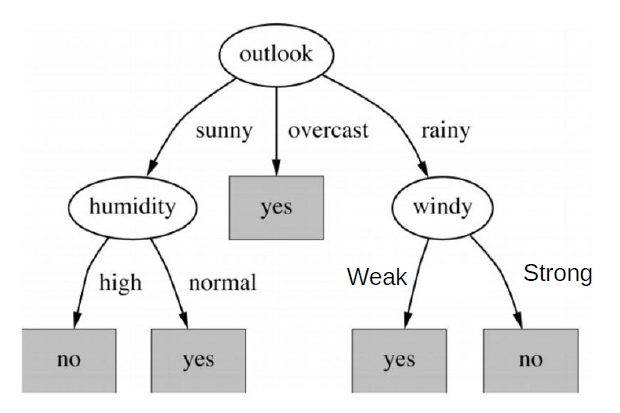
* Calculate **Average Information Entropy**: I(Temp) = 0.951
* Calculate **Gain:** Gain = 0.020

**Pick the highest gain attribute.**

|  |  |
| --- | --- |
| **Attributes** | **Gain** |
| Temperature | 0.02 |
| Humidity | 0.02 |
| Windy | 0.971 |

So 🡺  **Next Node in Rainy: Windy 🡸**

**Final Decision Tree**

****

**Working numerical examples of how Decision Trees are used in classification.**

In the previous part, I explained how the decision tree is built but here I’ll go on explaining how we can use it in classification for machine learning .

When we trace the first row of the previous dataset over the decision tree that we got previously :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S. No.** | **Outlook** | **Temperature** | **Humidity** | **Windy** |
| 1 | Sunny | Hot | High | Weak |

**First Step:**In the most left part of the tree, we will get into branch of “sunny”.

**Second Step:** When we come to Humidity node, I’ll get into branch of “high” as mentioned in the row.

**Final Step:** We’ll get the answer “**NO**” and It’s the correct result.

So NO one’ll play tennis on that day.

**What’re the Applications of Decision Trees in Real Life?** (Ref. 3)

1. **Selecting a flight to travel**  
   Assume you have to choose a trip for your next movement. How would we go about it? We check first if the flight is accessible on that day or not. In the event that it isn't accessible, we will search for some other date yet on the off chance that it is accessible, at that point we search for might be the span of the flight. On the off chance that we need to have just non-stop flights, at that point we look whether the cost of that flight is in your pre-characterized spending plan or not. On the off chance that it is excessively costly, we take a gander at some different flights else we book it!
2. **Choosing a best friend**  
   Well, this example can be very subjective but nevertheless, we will proceed with it. We all have many friends but a best friend is always one or two maybe but not all for sure. In this case, we may go about first whether the other person understands you or not? If another person is not pretty much understanding, we look for someone else. When we find people who have a good understanding level, we look for people who have similar habits or hobbies as we have and the selection criteria go on till we find people whom we can call “close friends”!
3. **Handling late night cravings**  
   I guess the following picture will explain it all to you.

**Applications in Business**

1. **Alternatives**  
   Decision trees help organizations to view alternatives to different events that can happen. For example, an organization is planning to build a new product. A lot of resources and money will be invested in it. Decision trees will help the organization to have a look at all other alternatives if they do not want to build the product from scratch. Alternatives may include not building the new product at all or bringing modifications to existing products to incorporate new features in them
2. **Events**  
   An event is an occurrence that happens outside your direct control and as a result of your actions. In the example of whether to change the technical stack of the platform, the possible results from the decision to proceed would be the development of a successful product or a product failure.
3. **Outcomes**  
   The outcomes are the results of making a decision, weighted with the probabilities of particular events occurring. For the example of product development, you might have to invest $50,000 to develop the new product, expecting it to generate total profits of $100,000. The probability of developing a successful product might be 80 %. As a result, you have an 80 % chance of earning $100,000, a 20 % chance of earning nothing if the product fails, and a 100 % chance of earning no additional profit if you don’t develop the product.

* 1. **Exploratory data analysis for A real dataset:**
  2. [**The Attached Dataset**](https://www.kaggle.com/nsaravana/bank-data-set-using-decision-tree)contains a list of Clients’ data of A Bank.

Fields include:

* Age – Age of the person
* Job
* Marital – The Marital of the person
* Education
* Balance – The Balance of the person who going to the bank
* Housing
* Loan - If the person took a loan or not
* Contact
* Day – The Day of going to the bank
* Month - The Month of going to the bank
* Duration
* Campaign
* Deposit – If the person will deposit or not

**I’ll discuss the column of**  **Balance.**

**Features:**

* **Median**

550.0

* **Mean**

1528.5385235620856

* **Mode**

97

* **Variance**

10403291.123191012

* **Standard** **Deviation**

3225.413325946151

* **Interquartile Range**

1586.0

* **Min**

-6847.000000

* **Max**

81204.000000

* **25%**

122.000000

* **50%**

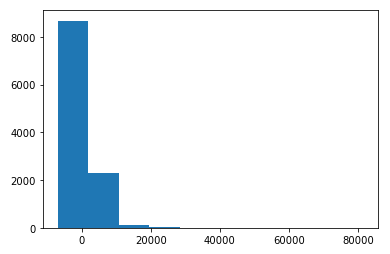
550.000000

* **75%**

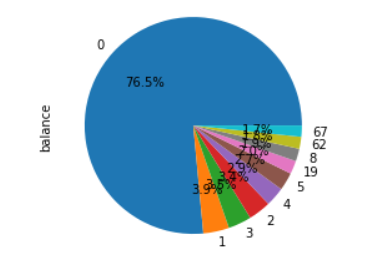
1708.000000

---------------------------------- **Graphs**  ------------------------------

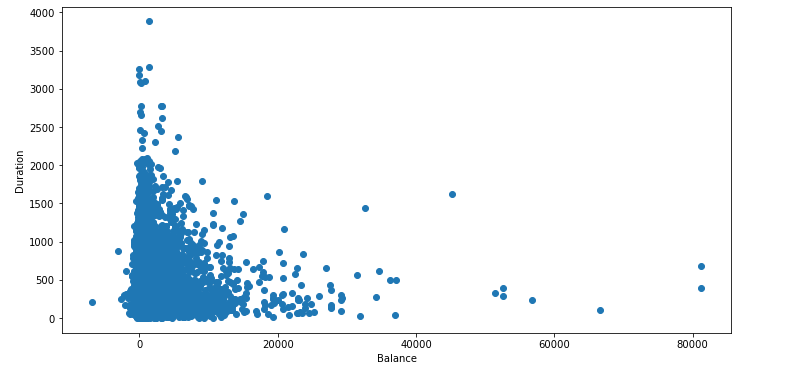
* **Histogram**

****

* **Pie Chart**

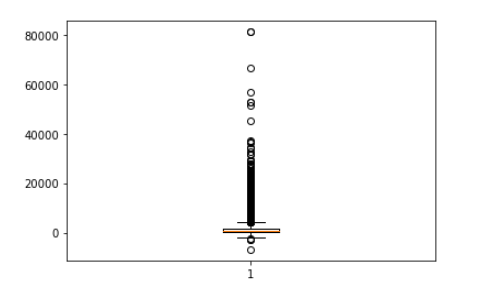
****

* **Scatter Plot**

****

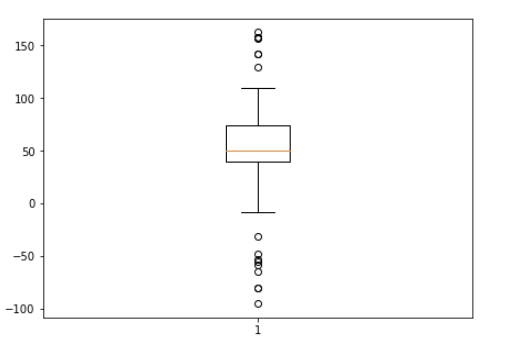
* **Box Plot**

**(**For all the data**)**



* **Box Plot**

**(**For a subset of the data**)**



**References and Books:**

**[1] Investopedia** (Website): [Link](investopedia.com/terms/d/decision-tree.asp)

**[2] Medium** (Website): [Link](https://medium.com/deep-math-machine-learning-ai/chapter-4-decision-trees-algorithms-b93975f7a1f1)

**[3] Quora** (Website): [Link](https://www.quora.com/In-what-real-world-applications-is-the-decision-tree-classifier-used)

**[4] Kaggle** (Website): [Link](https://www.kaggle.com/nsaravana/bank-data-set-using-decision-tree)

**[5]** “**Probability and Statistics for Engineers and Scientists**” by Ronald E. Walpole et al.,, 9th Edition, Pearson Education International. (Book): [Link](https://www.amazon.com/Probability-Statistics-Engineers-Scientists-Update/dp/0134115856)