



NOVA
IMS
Information
Management
School


Data Mining

Association Rules

9/12/2021
NOVA-IMS
Fernando Lucas Bação
bacao@isegi.unl.pt
<http://www.isegi.unl.pt/fbacao>

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa


1



NOVA
IMS
Information
Management
School

Association Rules

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa



2

NOVA

IMS

Information
Management
School

Association Rules

- **Association Rules:**
 - Aims at the extraction of compact patterns that describe subsets of data
 - Events that occur together (market basket analysis);
 - The main purpose is to establish relationships between fields;
 - Are rules of the form if X then Y;
 - Association rules provide information about things that tend to happen together.

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

3

NOVA

IMS

Information
Management
School

Association Rules

- **Association Rules:**
 - Table with a set of purchases from a supermarket, with 5 purchases and 5 items.

Cliente	Itens
1	Orange juice, soda
2	Milk, Orange juice, Glass cleaner
3	Orange juice
4	Orange juice, detergent, soda
5	Glass cleaner, soda

- With these data we can create a table of co-occurrences with the number of times that any pair of products was purchased together.

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

4

NOVA
IMS
Information Management School

Association Rules

	Orange Juice	Glass Cleaner	Milk	Soda	Detergent
Orange Juice	4	1	1	2	1
Glass Cleaner	1	2	1	1	0
Milk	1	1	1	0	0
Soda	2	1	0	3	1
Detergent	1	0	0	1	2

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

5

NOVA
IMS
Information Management School

Association Rules

- From the analysis, one can conclude that:
 - Orange juice and soda are more likely purchased together than any other two items;
 - Detergent is never purchased with milk or window cleaner;
 - Milk is never purchased with soda or detergent.

	OJ	GC	Milk	Soda	Det.
OJ	4	1	1	2	1
GC	1	2	1	1	0
Milk	1	1	1	0	0
Soda	2	1	0	3	1
Det.	1	0	0	1	2

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

6

NOVA IMS Information Management School

Association Rules

- Association Rules:**

The Apriori Algorithm -- Example

Database D

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

Scan D → C_1

item set	sup.
{1}	2
{2}	3
{3}	3
{4}	1
{5}	3

Scan D → L_1

item set	sup.
{1}	2
{2}	3
{3}	3
{5}	3

Scan D → C_2

item set	sup.
{1 2}	1
{1 3}	2
{1 5}	1
{2 3}	2
{2 5}	3
{3 5}	2

Scan D → L_2

item set	sup.
{1 3}	2
{2 3}	2
{2 5}	3
{3 5}	2

Scan D → C_3

itemset	sup.
{2 3 5}	2

Scan D → L_3

itemset	sup.
{2 3 5}	2

Note: {1,2,3}, {1,2,5} and {1,3,5} not in C_3

© D. Chaves R. Zahedi, 2001 Web Technologies and Applications University of Alberta 54

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

7

NOVA IMS Information Management School

Association Rules

- Association Rules:**
 - The rules are expressed in the form of:

"if the item A is part of an event, then the item B will also be part of the event X percent of the time"
 - The rules should not be interpreted as a direct causation, but only as an association;
 - It is not legitimate to infer rules of causality.

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

8

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**

*If a customer buys **shoes**, then 10% of the time also buys **socks***

*If a customer buys **plastic paint**, then 85% of the time also buys **brushes***

 - All rules have an antecedent and a consequent

*Shoes and plastic paint – **antecedents***

*Socks and brushes – **consequents***

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

9

NOVA
IMS
Information Management School

Association Rules

- **Venn Diagram:**

Antecedent

Consequent

Database

Combination A and C

U

A

B

C

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

10

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*

Confidence

Support

Expected Confidence

Lift
 - *The most important measures to assess the rules*

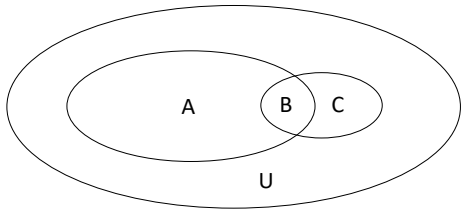
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

11

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*
 - **Confidence** – the **strenght of an association**, the percentage of a consequent appears given that the antecedent has occurred



$Confidence = \frac{B}{A}$

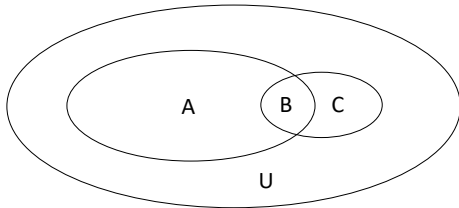
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

12

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*
 - **Support** – shows **how frequently** the combination occurs in the database.



$$\text{Support} = B/U$$

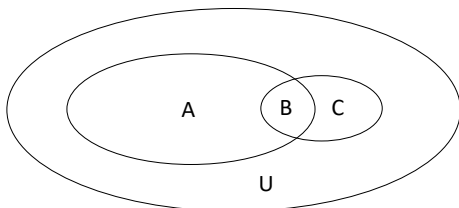
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

13

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*
 - **Expected confidence** – equal to the number of consequent transactions, divided by the total number of transactions



$$\text{ExpectedConfidence} = C/U$$

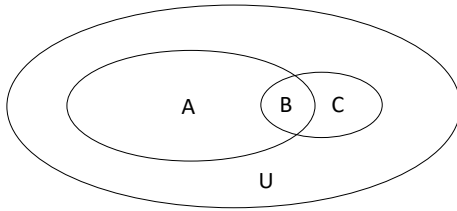
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

14

NOVA
IMS
Information Management School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*
 - **Lift** – equal to the confidence factor divided by the expected confidence. Lift is a factor by which the **likelihood** of consequent increases given an antecedent



$$\text{Lift} = \text{Confidence} / \text{Expected Confidence}$$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

15

NOVA
IMS
Information Management School

Association Rules

1,000,000	Total de Transações
200,000	Sapatos
50,000	Meias
20,000	Sapatos e Meias

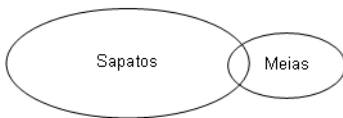


Diagrama de Venn

$$\text{Confidence} = B/A$$

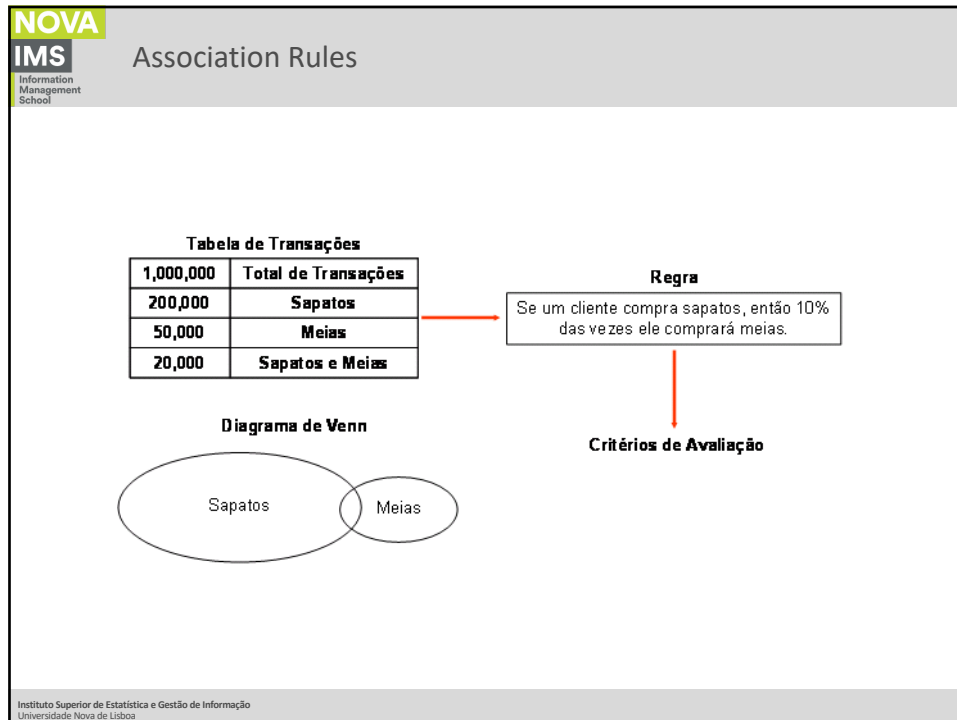
$$\text{Support} = B/U$$

$$\text{Expected Confidence} = C/U$$

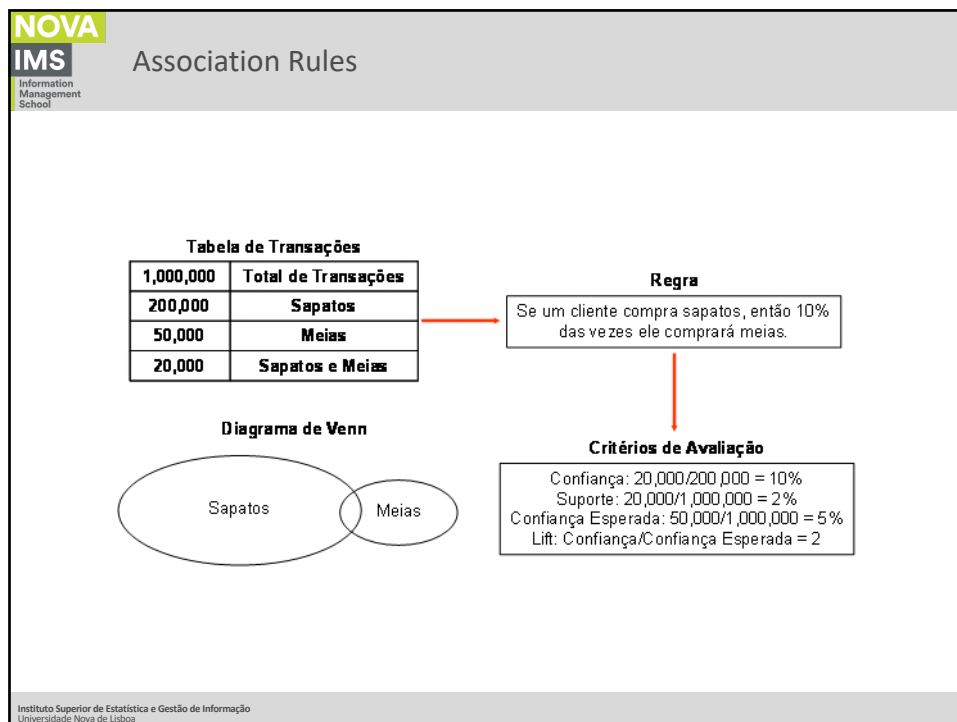
$$\text{Lift} = \text{Confidence} / \text{Expected Confidence}$$

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

16



17



18

NOVA

IMS

Information
Management
School

Association Rules

- **How association rules work:**
 - *Evaluating the quality of Association Rules:*

Lift – a value of 2 means that, if one bought shoes, one is twice as likely to buy socks than those who have not bought shoes

A **credible rule** must have a good confidence factor, a high level of support and lift higher than 1

Rules with a **high level of confidence** but with a **low support** should be interpreted with caution, as it can result in **idiosyncrasies** that rise due to the small number of cases to support the rule

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

19

NOVA


IMS

Information
Management
School

Association Rules

- **Use:**

The Wine Bible
by [Karen MacNeil](#)



List Price: \$49.95

Price: \$13.97 & eligible for **FREE Super Saver Shipping** on orders over \$25. [See details.](#)

You Save: \$5.98 (30%)

Availability: Usually ships within 24 hours

58 used & new from \$13.00


Edition: Paperback

[Look inside this book](#)


▶ [See more product details](#)

Better Together

Buy this book with [Wine for Dummies](#) by Ed McCarthy, et al today!



+



Total List Price: \$44.94
Buy Together Today: \$29.36

[Buy both now!](#)

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

20

NOVA

IMS

Information
Management
School

Association Rules

• **Use:**

Customers who bought this book also bought:

- [How to Taste : A Guide to Enjoying Wine](#) by Jancis Robinson ([Rate it](#))
- [The World Atlas of Wine](#) by Hugh Johnson, Jancis Robinson ([Rate it](#))
- [The Wall Street Journal Guide to Wine: New and Improved: How to Buy, Drink, and Enjoy Wine](#) by Dorothy J. Gaiter, John Brecher ([Rate it](#))
- [The University Wine Course: A Comprehensive Text and Tutorial](#) by Marian W. Baldy Ph.D. ([Rate it](#))
- [Parker's Wine Buyer's Guide 6th Edition : The Complete, Easy-to-Use Reference on Recent Vintages, Prices, and Ratings for More Than 8,000 Wines from All the Major Wine Regions](#) by Robert M. Parker ([Rate it](#))

► **Explore Similar Items:** [20 in Books](#), [20 in DVD](#), and [16 in Magazine Subscriptions](#)

You could win a
\$50 Amazon gift certificate
or the **Grand Prize**. Start by
adding five items
to your Wish List.
[Learn more.](#)

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

21

NOVA

IMS

Information
Management
School

Association Rules

• **Types of rules**

- Trivial rules;
- Inexplicable Rules;
- Actionable Rules;
- Rules that result of promotions made.

Instituto Superior de Estatística e Gestão de Informação
 Universidade Nova de Lisboa

22

NOVA
IMS
Information Management School

Association Rules

- **Additional aspects**
 - Available data are essential to success;
 - It takes large amounts of data;
 - What is an item
 - Frozen pizza, or frozen pizza 4 seasons;
 - Setting the appropriate level of detail;
 - Classifications.

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

23

NOVA
IMS
Information Management School

Association Rules

- **Additional aspects**

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

24

NOVA
IMS
Information
Management
School

Association Rules

- **Additional aspects**
 - Use of virtual items;
 - Allow you to get information from data beyond the information expressed in the products;
 - May include information about the purchase (paid in cash, credit card...), about the customer (new, old...) day of the week, hour of the day.

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa