

# Data Mining on Promotional Sales

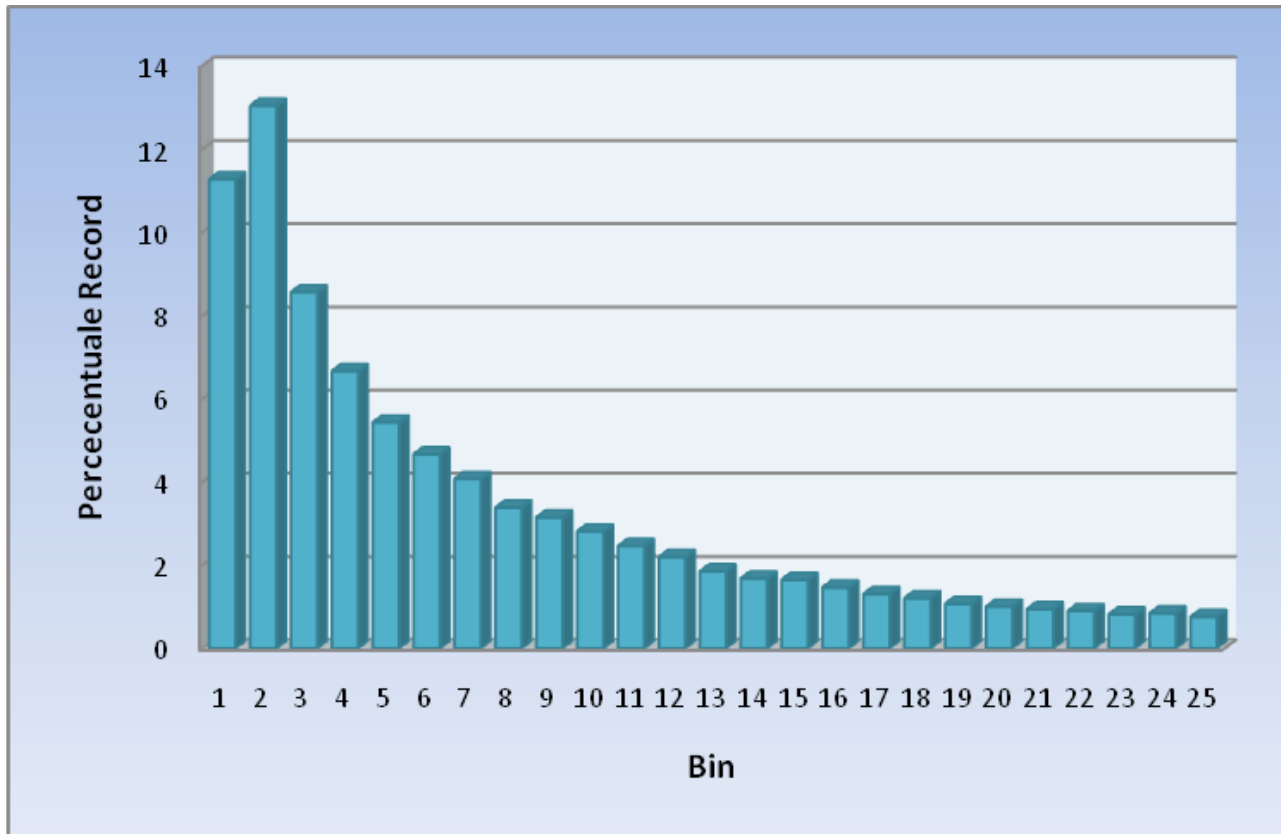
# Goals

Goal: focusing on the effects of promotions on the sales of a single product, mainly aimed at optimizing its stocking:

- **Forecasting sales of promoted products**
- **Forecasting “out-of-stock”**

Case study on product category = Food,  
Two months April 2006 & April 2007.

# Sales volume distribution of promotions with at least 5 items sold



- 11.5% of the promotions in a single store sold between 5 and 24 items (the leftmost bar in the figure),
- the 13.79% sold between 25 and 49 items,
- The tail is less flat

# Data Preparation

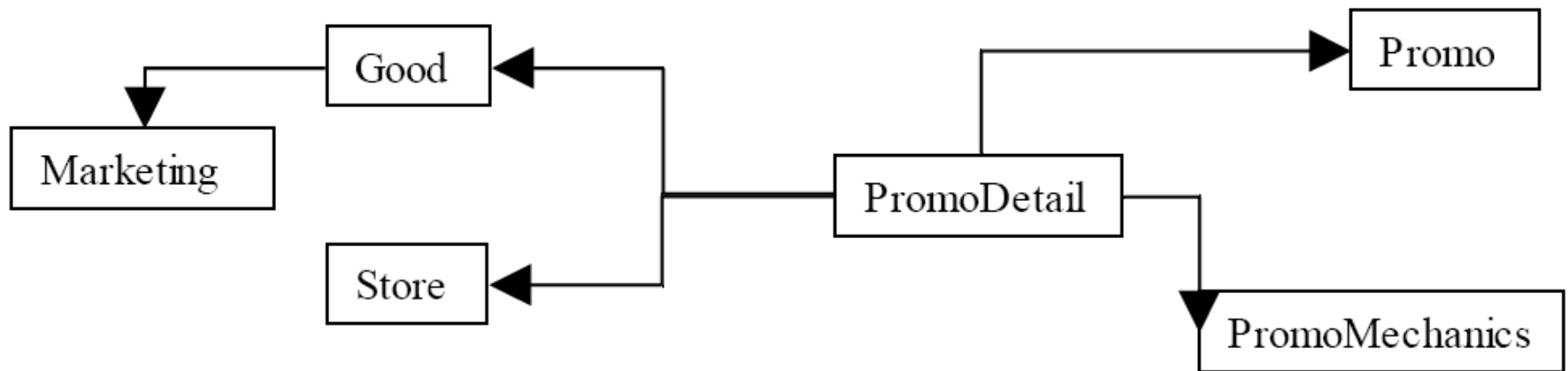


Figure 3: structure of relevant portion of the data warehouse

# Model building

## Predictors:

- Product details
- Promo details
- Volume of sales in the periods before the promotion

## Target Variable:

- Number of sales for the promoted item
- **Variation w.r.t the month before the promotion**

# Mining Table

Field nme	Description
Vend_Art_3_1	Sales of the article from 3 months to 1 month before the promotion
Vend_Seg_3_1	Sales of the segment from 3 months to 1 month before the promotion
Vend_Art_1_0	Sales of the article in the last month before the promotion
Vend_Seg_1_0	Sales of the segment in the last month before the promotion
Giorni_Promozione	

data sales of 16 months in 134 stores (522,541,764 records).

# Target variables

**Case1** - The **sales amount** of the promoted item and

**Case2** - The **sales variations** of the promoted item and

**Case3** - The number of “out-of-stock” that occurred during the promotion

## MINING TASK:

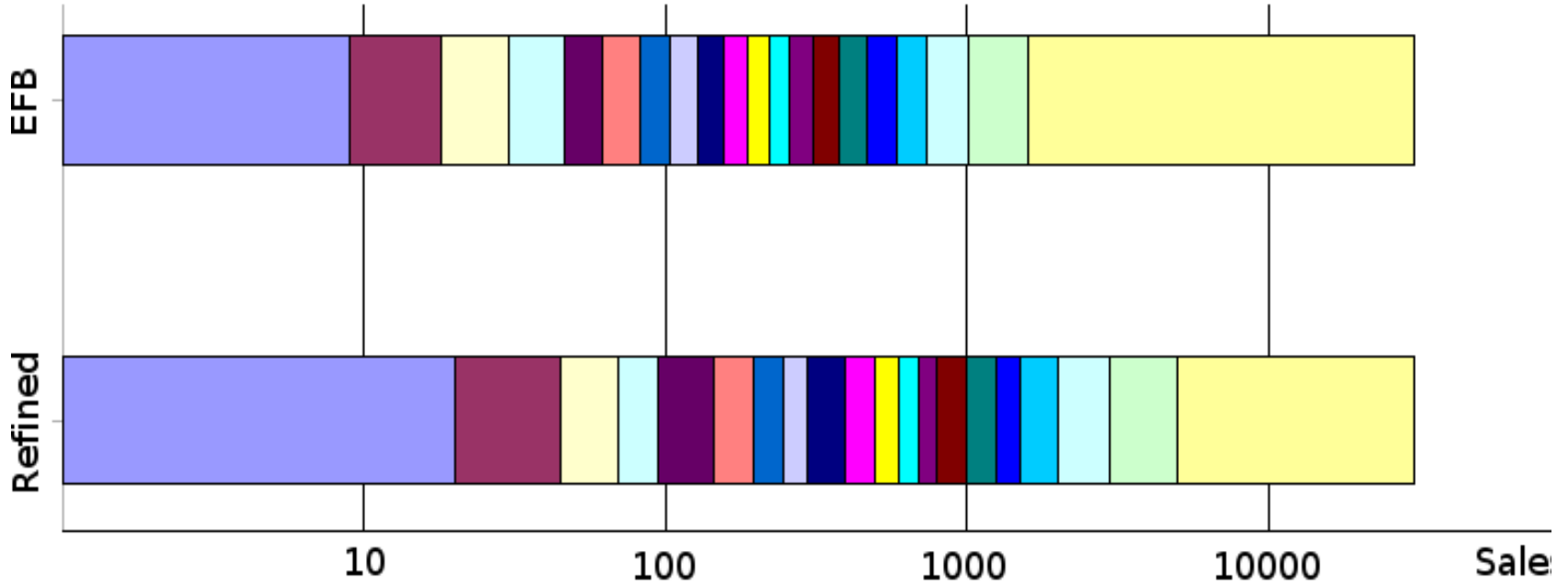
- MULTICLASS classification
- ORDINAL CLASSES i.e. order among classes

# Case1: Predicting **sales volume**

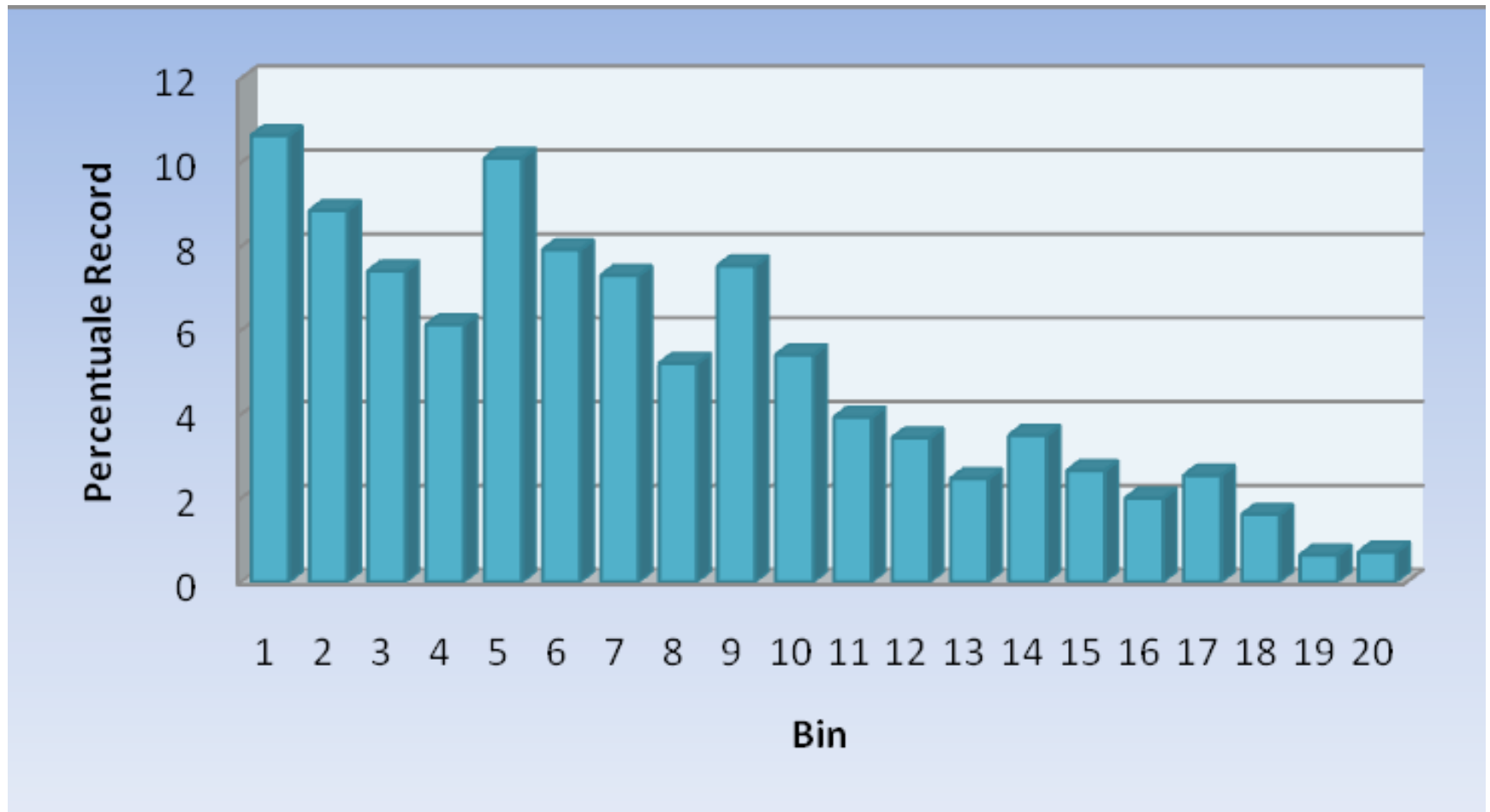
- Volume is continuous, range 0 -105.650, 80% less than 500 item
  - Discretize (how many classes?)
  - Multiclass predictor
- Equal size binning:
  - bin width = 10=> 965 bins, 18% in first 3 classes
  - bin width = 100 => 249 bins, 64% in first 3 classes
- Equal frequency binning
  - 20 bins => refined afterward



# Manual discretization



# Distribution of sales volume discretized in 20 bins refined discretization



# Results evaluation

- Accuracy of 55,1% on the training set, which drops to 22,45% over test set

Risultati per campo di output VEND\_ART\_PROMO\_TILEN\_String

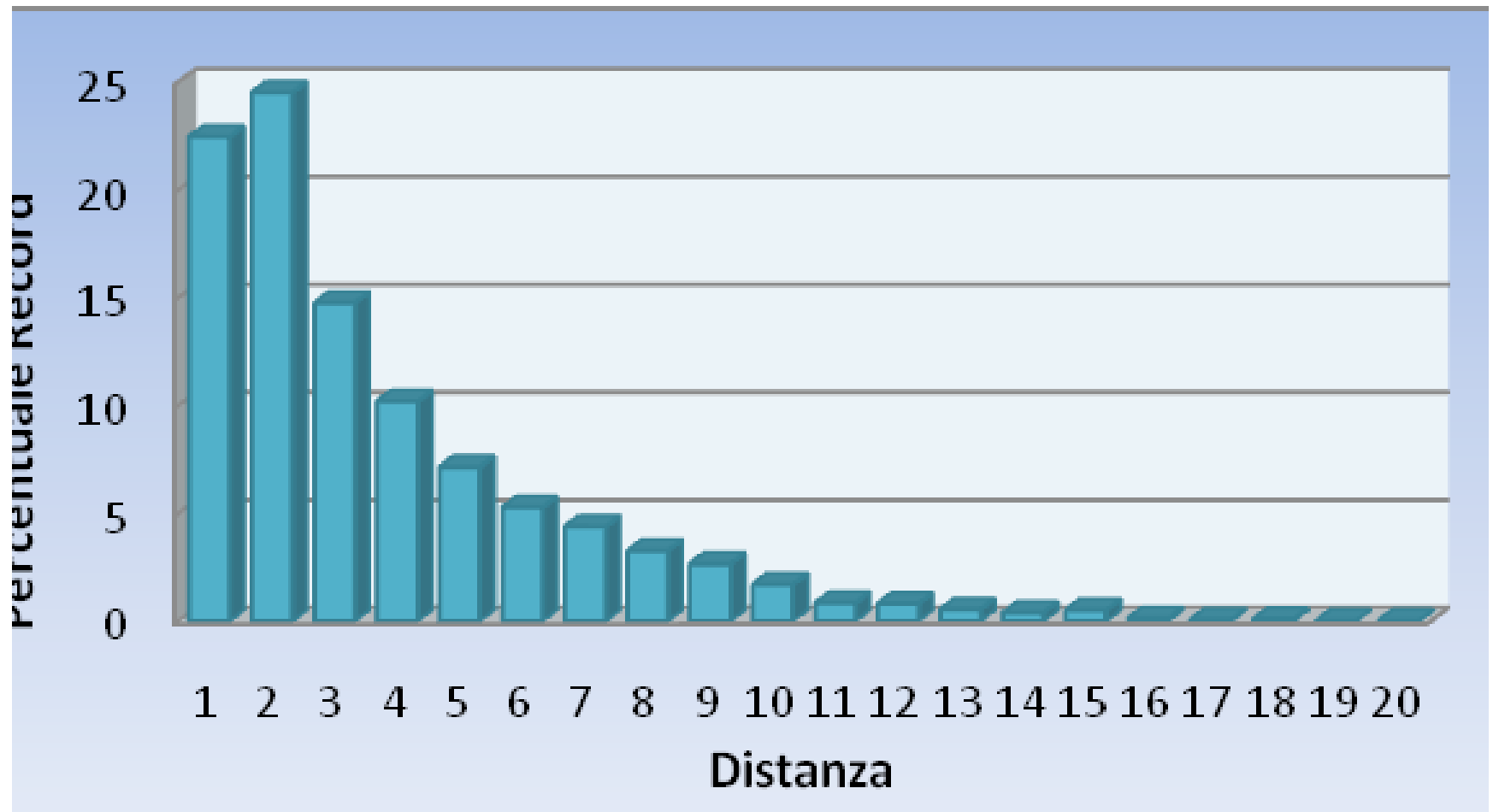
Confronto di \$C-VEND\_ART\_PROMO\_TILEN\_String con VEND\_ART\_PROMO\_TILEN\_String

'Partizione'	1_Addestramento		
Corretto		3.386	55,1%
Sbagliato		2.759	44,9%
Totale		6.145	

Matrice coincidenza per \$C-VEND\_ART\_PROMO\_TILEN\_String (le righe mostrano i valori effettivi)

'Partizione'= 1_Addestramento	1	2	3	4	5	6	7	8	9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20
1	545	44	25	6	3	6	5	1	5	4	2	0	0	2	0	0	0	0	0	0
2	111	318	47	12	21	13	6	1	16	5	1	0	0	0	1	0	1	0	1	0
3	47	51	270	16	23	9	5	0	9	9	2	0	0	0	2	1	0	0	0	1
4	50	32	44	179	40	6	9	4	3	1	2	3	1	1	0	0	0	0	0	0
5	45	22	43	26	401	32	17	11	16	7	8	1	1	4	2	0	1	0	0	0
6	34	18	22	19	62	243	24	11	15	6	3	1	1	3	0	1	0	1	0	0
7	25	10	15	12	38	43	255	11	20	6	3	5	1	2	2	0	3	0	0	0
8	19	11	12	16	21	12	39	141	20	7	5	3	0	4	0	1	1	0	0	0
9	27	9	12	9	29	34	32	38	247	15	6	2	3	3	2	0	1	0	0	0
_10	14	11	8	8	17	20	18	14	40	146	3	15	4	6	3	1	1	1	0	1
_11	9	3	5	2	11	11	14	12	25	25	94	12	2	9	2	0	5	1	0	0
_12	5	5	6	6	10	9	11	7	16	20	10	91	5	7	1	2	2	1	0	1
_13	6	1	3	2	6	6	5	7	6	7	9	12	47	8	7	4	2	1	0	1
_14	7	2	4	5	5	2	7	8	9	13	12	9	6	119	3	2	3	2	0	0
_15	3	3	4	5	8	2	8	8	7	10	8	6	1	18	78	4	6	0	0	1
_16	3	0	2	1	6	3	4	4	1	5	3	5	7	13	11	43	11	2	1	0
_17	3	0	3	4	6	3	4	2	5	4	6	7	4	11	9	1	80	2	0	0
_18	1	0	2	1	4	3	0	3	2	0	2	2	3	7	11	7	12	36	2	1
_19	1	0	1	1	1	0	1	2	0	1	1	0	1	0	3	2	5	0	14	4
_20	0	0	0	0	0	0	0	0	0	3	2	0	0	0	1	1	1	0	1	39

# Class displacement of predicted class vs. real class



Rule	Support	Confidence	Confidence with error $\leq 1$	Confidence with error $\leq 2$
<b>if</b> CATEGORIA = ZUCCHERO E DOLCIFICANTI e FL_VOLANTINO = No e VEND_ART_1_0 > 37 <b>then class = 2</b>	47	23%	82%	93%
<b>if</b> CATEGORIA = 'ALIMENTI INFANZIA' e VEND_ART_1_0 > 275 <b>then class = 3</b>	138	50%	82%	97%
<b>if</b> CATEGORIA = CONSERVE DI FRUTTA e MESE = 8 <b>then class = 5</b>	113	24%	65%	86%
<b>if</b> CATEGORIA = YOGURT e DESCRIZ. = TAGLIO PREZZO e MESE = 9 e VEND_ART_1_0 > 54 e VEND_SEG_1_0 <= 4487 <b>then class = 6</b>	110	35%	65%	82%
<b>if</b> CATEGORIA = 'PASTA FRESCA' e MESE = 10 e VEND_ART_1_0 > 51 <b>then class = 7</b>	42	38%	57%	78%
<b>if</b> FL_COOP = Si e CATEGORIA = BISCOTTI e FL_VOLANTINO = Si e VEND_ART_1_0 <= 275 <b>then class = 8</b>	52	25%	61%	78%

Table 6 - Classification rules with support and confidence, including limited tolerance to errors

## Rule 1:

- **if** more than 37 articles were sold in the last month before the promotion (vent\_art\_1\_037) in the category “sugar” (categoria = zucchero e dolcificanti),
- **and** the promotion was not advertised in the advertising leaflets,
- **then** the promoted item will sell the same or just a slightly higher amount than before the promotion (class = 2).

# Case2: New Target Variable

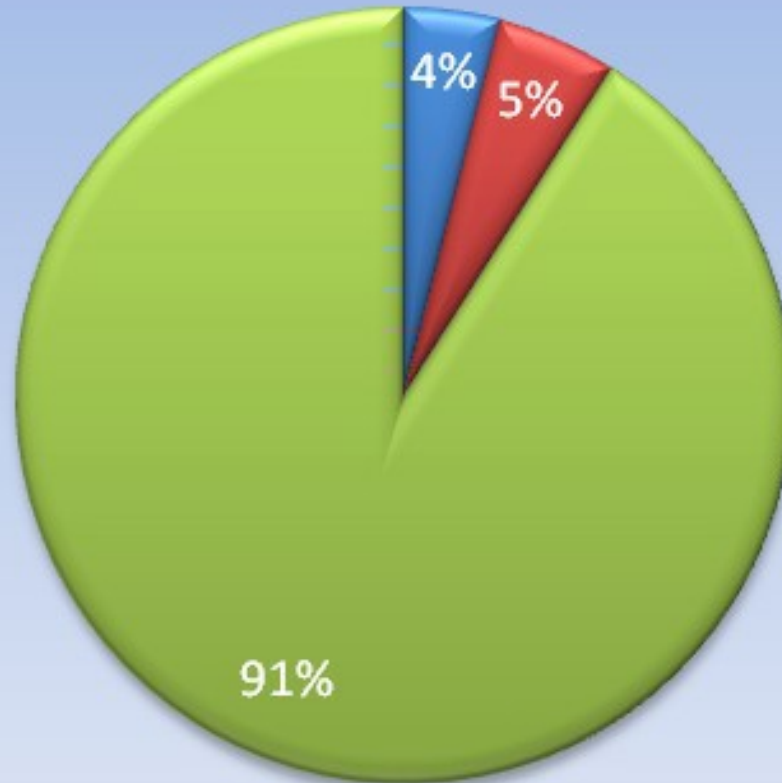


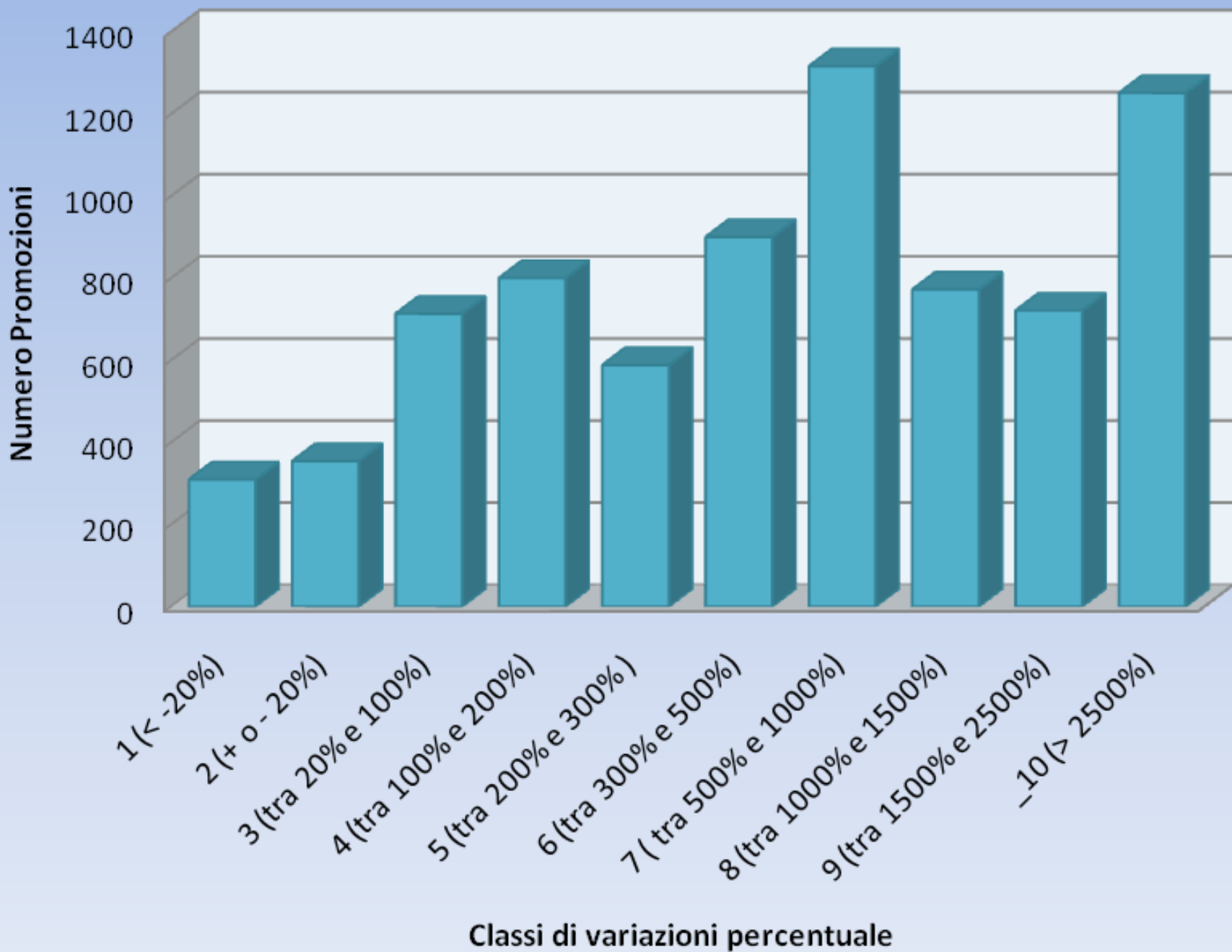
Figure 7: Percentage variation of sales under promotion

■ Minore (< -20%)    ■ Uguale (+ o - 20 %)    ■ Maggiore (> +20%)

# Case2: Variation of sales

Class	Meaning
1	Drop of sales (sales variation $\leq -20\%$ )
2	Drop of sales (sales variation $\leq -20\%$ )
3	Small increase 1 ( variation between + 20% and +100%)
4	Small increase 2 ( variation between +100% and 200%)
5	Small increase 3 ( variation between +200% and 300%)
6	Large increase 1 ( variation between +300% and 500%)
7	Large increase 2 ( variation between +500% and 1000%)
8	Large increase 3 ( variation between +1000% and 1500%)
9	Extreme increase 1 ( variation between +1500% and 2500%)
10	Extreme increase 2 (variation $\geq 2500\%$ )





# Results evaluation

- Accuracy reaches the 49.99% on the training set and 32.67% on the test set

[-] Risultati per campo di output VariazionePercentualiS

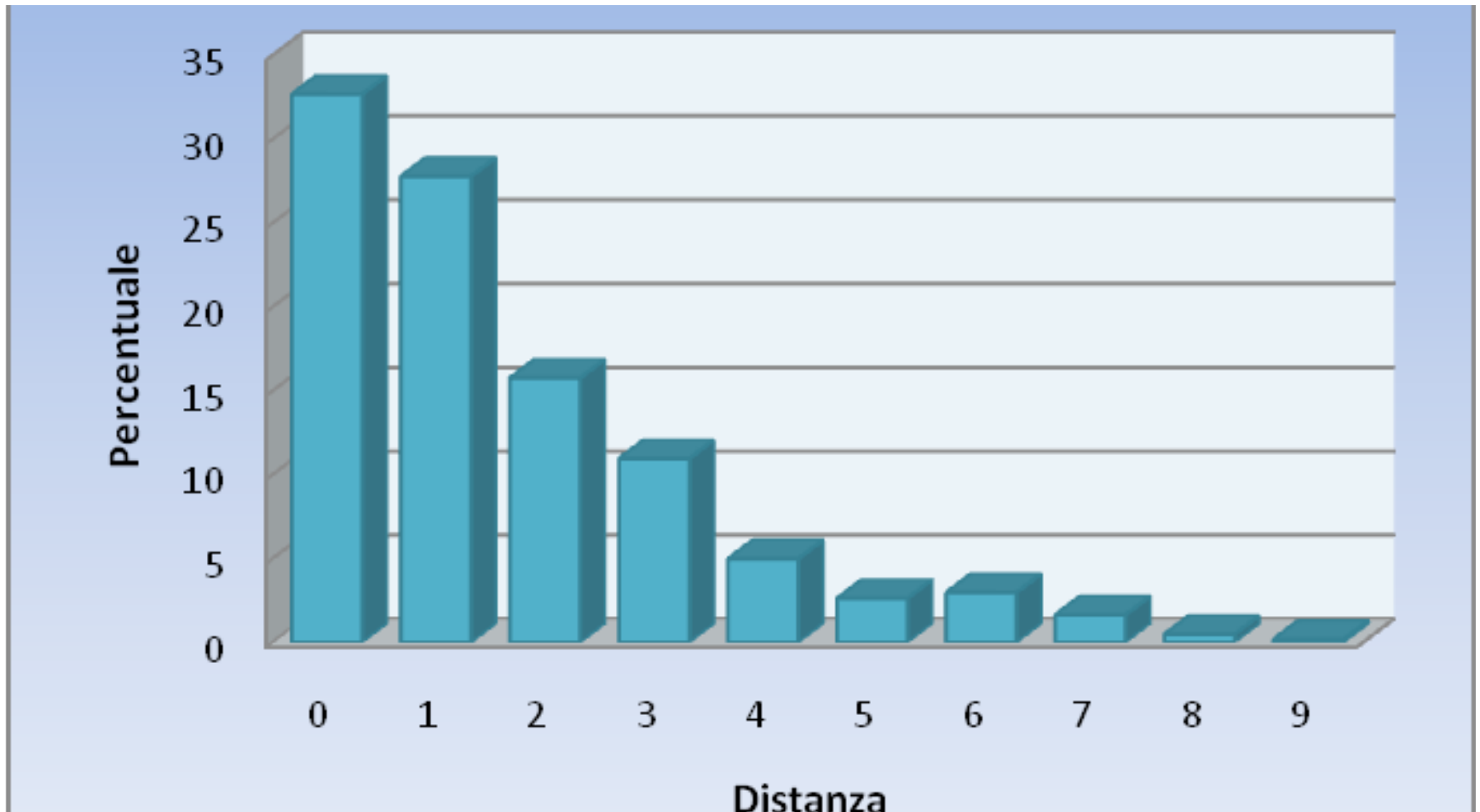
[-] Confronto di \$C-VariazionePercentualiS con VariazionePercentualiS

Corretto	2.702	49,99%
Sbagliato	2.703	50,01%
Totale	5.405	

[-] Matrice coincidenza per \$C-VariazionePercentualiS (le righe mostrano i valori effettivi)

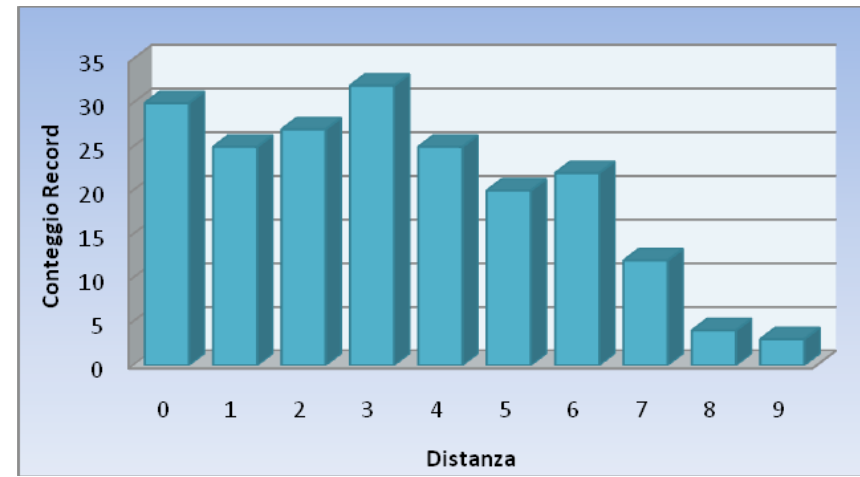
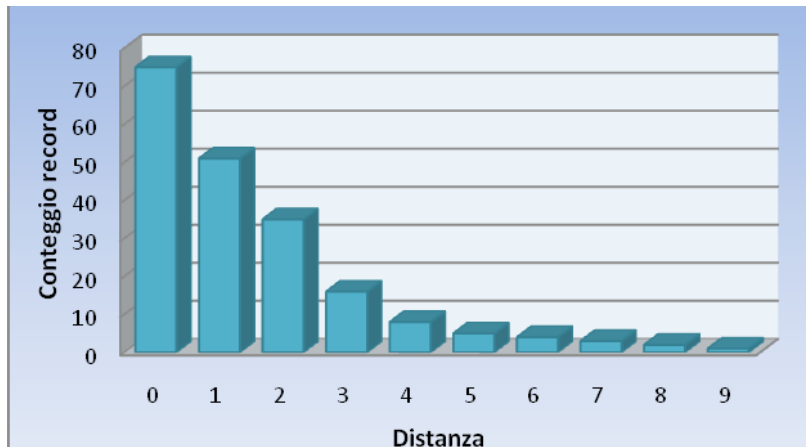
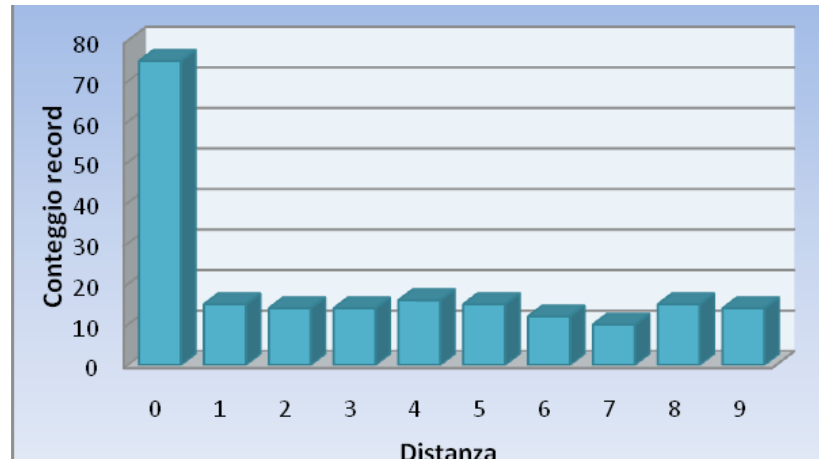
	1	2	3	4	5	6	7	8	9	_10
1	75	16	53	14	7	9	25	0	7	10
2	11	83	69	25	3	9	28	4	7	16
3	6	17	281	63	11	23	57	1	10	35
4	2	9	88	265	21	32	77	5	9	55
5	2	3	61	57	100	51	80	6	8	34
6	2	5	64	39	16	250	147	8	10	59
7	6	5	64	41	16	53	572	29	28	92
8	2	8	32	10	3	13	164	180	27	103
9	1	1	19	14	0	11	132	26	159	162
_10	0	2	24	5	1	4	71	9	39	737

# Class displacement of predicted class vs. real class



# Evaluating ordinal (multiclass) classifiers

# Evaluating Ordinal Classifiers from their distance matrix

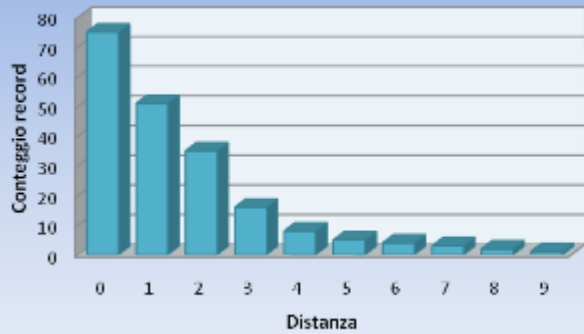


# Weights Vector-based approach

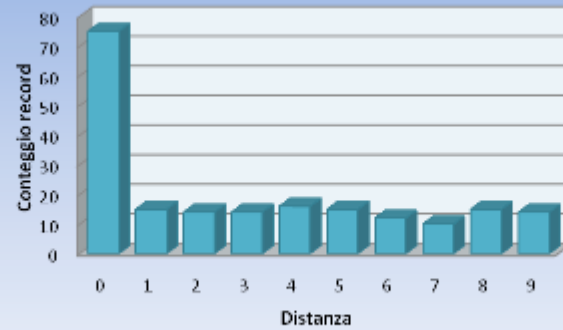
<i>Distance</i>	<i>Weights 1</i>	<i>Weights 2</i>	<i>Weights 3</i>
<b>0</b>	1	1	1
<b>1</b>	0	0,7	0,7
<b>2</b>	0	0,5	0,5
<b>3</b>	0	0,2	0,2
<b>4</b>	0	0	0
<b>5</b>	0	0	0
<b>6</b>	0	0	0
<b>7</b>	0	0	-0,2
<b>8</b>	0	0	-0,3
<b>9</b>	0	0	-0,5

$$Accuracy^{vector} = \frac{\sum_{i=1}^N (freq[i] \cdot weights[i])}{\sum_{i=1}^N freq[i]}$$

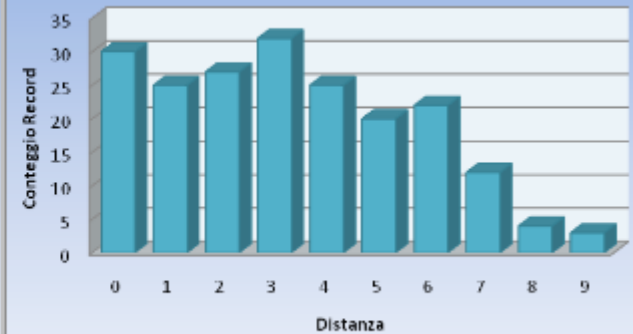
# Weights Vector-based approach



(a) a good classifier



(b) a not so good classifier



(c) a bad classifier

	Distribution (a)	Distribution (b)	Distribution (c)
<b>Weights 1</b>	37,5%	37,5%	15,0%
<b>Weights 2</b>	65,7%	47,6%	33,7%
<b>Weights 3</b>	64,8%	40,9%	31,1%

# Variant: Matrix of Weights

$$Accuracy^{matrix} = \frac{\sum_{i=1}^N \sum_{j=1}^N (mat\_confusion[i,j] \cdot mat\_weights[i,j])}{\sum_{i=1}^N \sum_{j=1}^N mat\_confusion[i,j]}$$

		Predicted Class									
		1	2	3	4	5	6	7	8	9	10
	1	1,00	0,85	0,70	0,50	0,40	0,00	0,00	-0,50	-0,75	-1,00
	2	0,85	1,00	0,85	0,70	0,50	0,30	0,00	0,00	-0,50	-0,75
	3	0,70	0,85	1,00	0,80	0,65	0,40	0,20	0,00	0,00	-0,50
	4	0,50	0,70	0,80	1,00	0,80	0,65	0,30	0,10	0,00	0,00
	5	0,40	0,50	0,65	0,80	1,00	0,80	0,65	0,20	0,00	0,00
	6	0,00	0,30	0,40	0,65	0,80	1,00	0,75	0,60	0,20	0,00
	7	0,00	0,00	0,20	0,30	0,65	0,75	1,00	0,75	0,60	0,15
	8	-0,50	0,00	0,00	0,10	0,20	0,60	0,75	1,00	0,70	0,55
	9	-0,75	-0,50	0,00	0,00	0,00	0,20	0,60	0,70	1,00	0,70
	10	-1,00	-0,75	-0,50	0,00	0,00	0,00	0,15	0,55	0,70	1,00

Table 14: Matrix of weights 1



# Case 3: Out of stock

# Two levels of out-of-stock

- In the warehouse: shortage of products during the promotion, preventing refill of shops
- In the shop: shortage for the single shop, which requires transferring from the warehouse

# Capturing out-of stock in a day

- Divide the day into 4 segments
  - Morning (“Mattina”)
  - Lunch time (“Pranzo”)
  - Afternoon (“Pomeriggio”)
  - Evening (“Sera”)
- Out-of-stock = sudden and huge drop of sells of the product (e.g. -90%)

Mattina	Pranzo	Pomeriggio	Sera
40	30	2	1

# Examples

- No out-of-stock if sells restart

Mattina	Pranzo	Pomeriggio	Sera
40	2	10	10

- No out-of-stock if sales were too low

Mattina	Pranzo	Pomeriggio	Sera
2	1	1	0

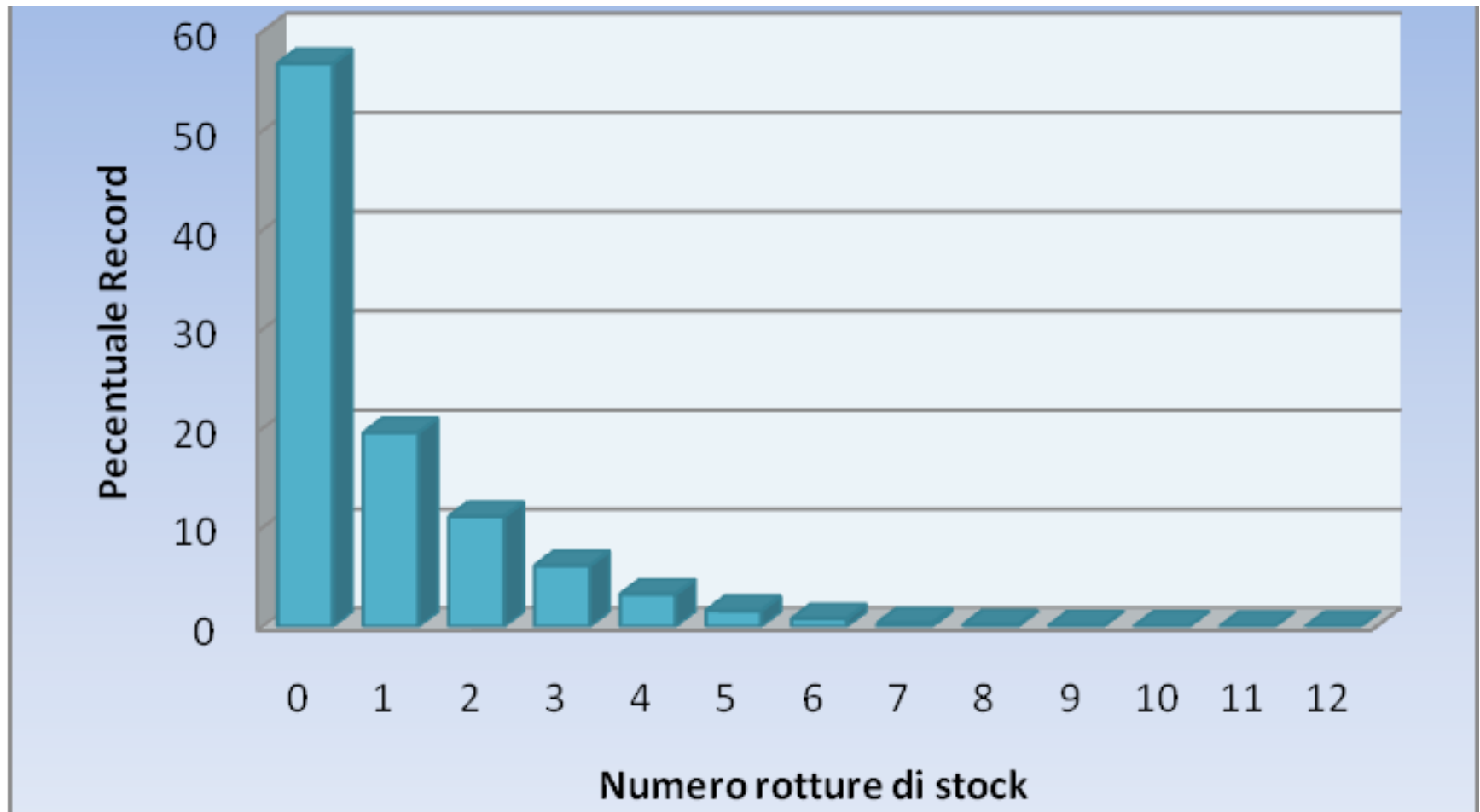
- Ok if drops are gradual

Mattina	Pranzo	Pomeriggio	Sera
25	4	0	0

# Rules for out-of-stock detection

















- Cond 1: decrease of adjacent sales 90%
- Cond 2: no further (no adjacent) increase
- Cond 3: minimum number BEFORE the out-stock
- Cond 4 (to allow gradual drops): if no number after out-of-stock, reduce threshold to 75%

# Distribution of out of stocks in the Super stores

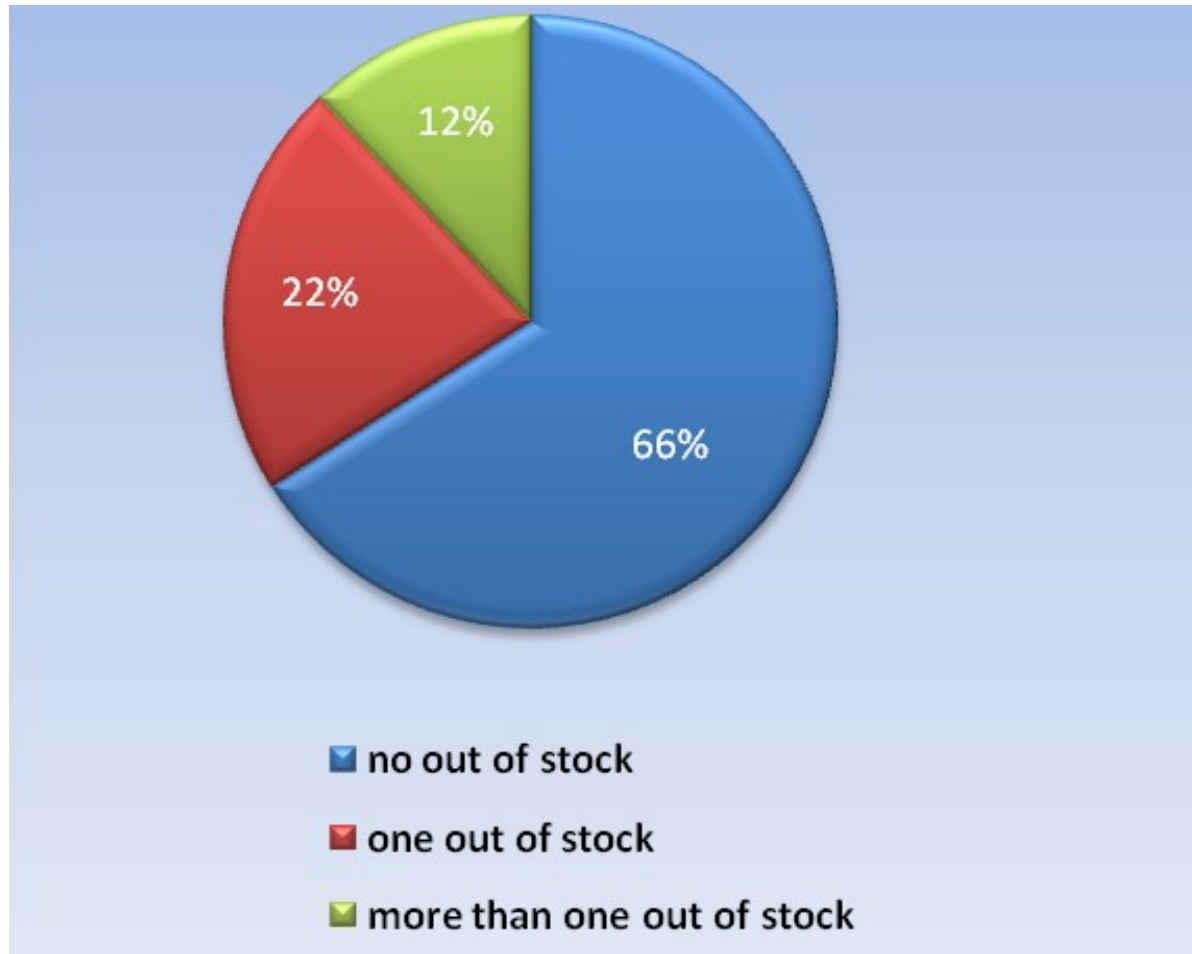


# Out-of-stocks in “Super”

Number of days an out-of-stock happens

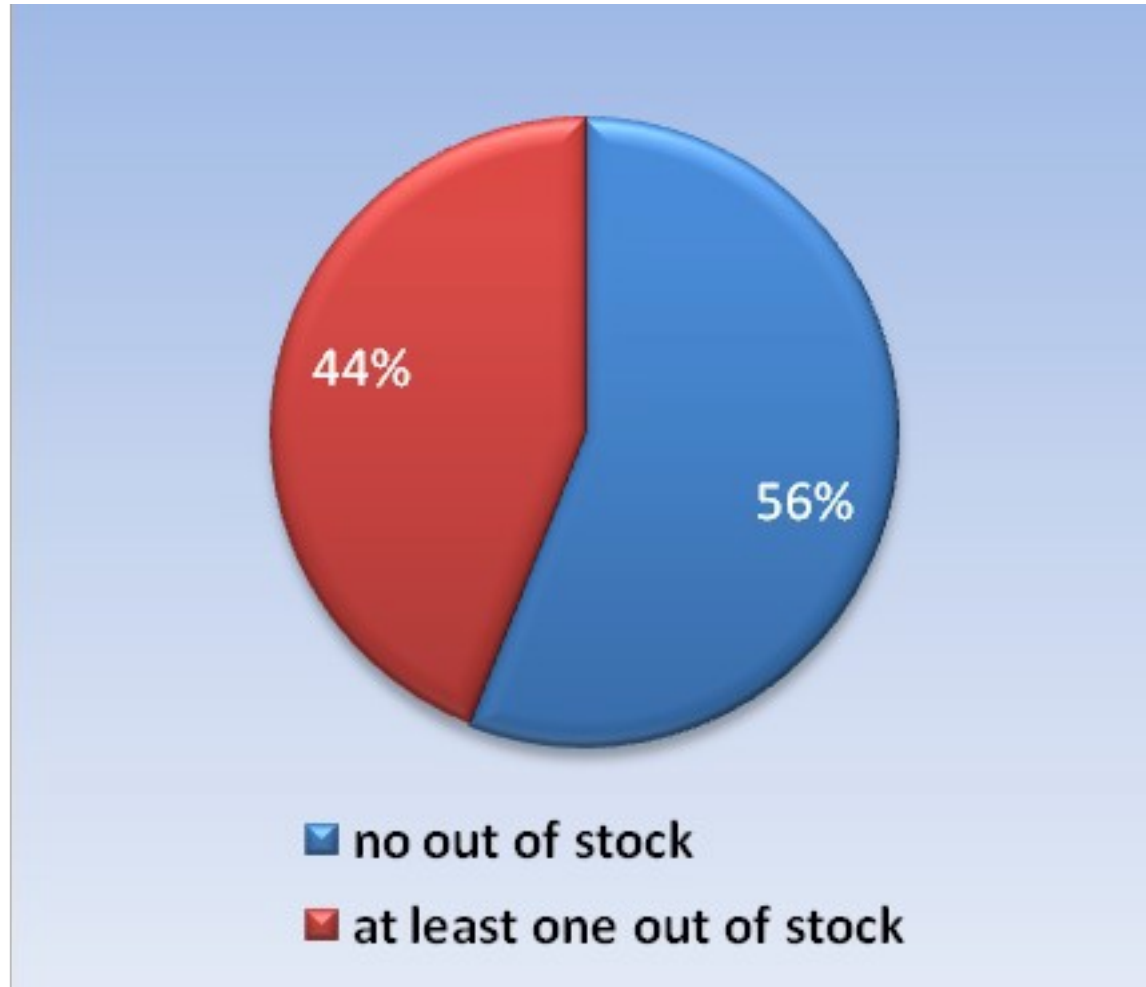
Valore ▲	Proporzione	%	Conteggio
0		56,78	147736
1		19,56	50891
2		11,12	28924
3		6,16	16019
4		3,27	8520
5		1,62	4211
6		0,8	2076
7		0,35	915
8		0,17	441
9		0,08	211
10		0,04	105
11		0,02	59
12		0,01	27
13		0,01	29
14		0,01	17
15		0,0	3

# Target variable: out-stocks (3 values)





# Target variable out-stocks (2 values)



# Es. Rule

if FL\_VOLANTINO = Si  
and CATEGORIA = ALIMENTI INFANZIA  
and VEND\_ART\_3\_1 > 142  
and VEND\_ART\_1\_0 > 96  
then class = 1

support = 677 confidence= 65%

# Rule 2

if MESE = 12  
and CATEGORIA =  
GELATI then class = 0  
supp= 379 conf= 95%

# Accuracy

☒ Risultati per campo di output Rottura\_Stok\_Binario

☒ Confronto di \$C-Rottura\_Stok\_Binario con Rottura\_Stok\_Binario



Corretto	177.209	71,61%
Sbagliato	70.238	28,39%
Totale	247.447	

☒ Matrice coincidenza per \$C-Rottura\_Stok\_Binario (le righe mostrano i valori effettivi)

	0	1
0	109.412	31.424
1	38.814	67.797



Only ~16% of missed out-of-stocks (false negatives)

# Example of rules: Coffee

Valore ▲	Proporzione	%	Conteggio
0		56,78	147736
1		43,22	112461

Overall out-of-stock distribution

se PRES\_MKT =  
LEADER  
e VendSeg\_1\_0 > 479  
e **CATEGORIA = CAFFE'**  
allora 1

Valore ▲	Proporzione	%	Conteggio
0		41,3	1156
1		58,7	1643

Filtered out-of-stock distribution



Support: 1643  
Conf.: 58,7%

# Example of rules: Coffee

se **PRES\_MKT = LEADER**  
e **VendSeg\_1\_0 > 479**  
e **CATEGORIA = CAFFE'**  
allora 1

Supp.: 718


Conf.: 67,93%

Valore ▲	Proporzione	%	Conteggio
0		32,07	339
1		67,93	718

se **PRES\_MKT = LEADER**  
e **VendSeg\_1\_0 > 479**  
e **CATEGORIA = CAFFE'**  
allora 1

Supp.: 560

Conf. 86%


Valore ▲	Proporzione	%	Conteggio
0		13,93	78
1		86,07	482


# Deployment

Google - Impostazioni previsione x

← → ↻ ☆ http://localhost:1983/Pages/Coop.aspx ▶ 📄 🔧

## Previsione volumi di vendita in promozione



Articolo	[18384] - Yogurt Coop Biologico Agrumi Conf. 2 Pezzi		Cerca...
Negozi	Viterbo (40) ▼		
Attivazione Promozione	inizio: 01 agosto 2008 	durata: 15 ▼	
Meccanica Promozione	tipologia: Sconto percentuale ▼	valore: 20% ▼	

[Visualizza statistiche promozioni passate](#)

[Torna al menu principale](#)

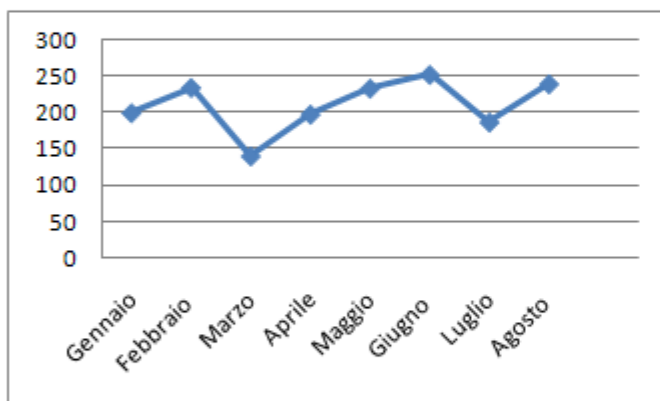
[Effettua previsione promozionale](#)



## Previsione volumi di vendita in promozione

**Dettagli previsione**  
 Articolo: [18384] - Yogurt Coop Biologico Agrumi Conf. 2 Pezzi  
 Negozio: Viterbo (40)  
 Data inizio: 01 Settembre 2008 - Durata: 15 giorni  
 Meccanica: Sconto 10%

### Andamento vendite



Gennaio	200
Febbraio	235
Marzo	140
Aprile	198
Maggio	234
Giugno	253
Luglio	187
Agosto	240

### Provisioni

Vendita	Variazione percentuale	Rischio rottura di stock
da 250 a 300 pezzi	+100% a +200% (da 240 a 360 pezzi)	SI

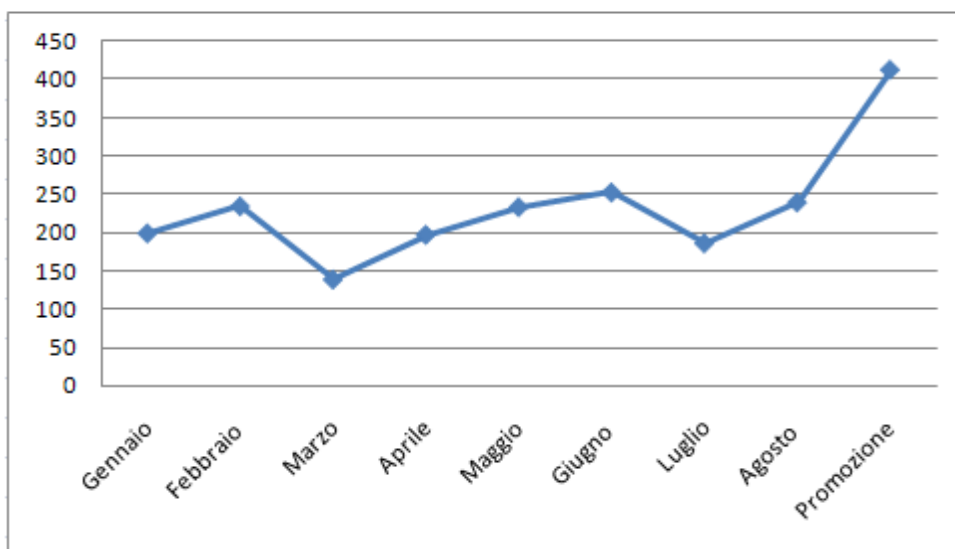
Indietro



## Statistiche promozioni



Negozi	Viterbo (40)
Periodo ricerca	da: 01 agosto 2008 a: 30 settembre 2008
Articolo in promozione	[18384] - Yogurt Coop Biologico Agrumi Conf. 2 Pezzi
Dettagli promozione	Codice promodettaglio - 18384 Codice promold - 235 Data inizio promozione - 01 settembre 2008 Data inizio promozione - 15 settembre 2008 Negozi - Viterbo Codice Negozi - 40



### Confronta con segmento

#### Rotture di stock in promozione

2

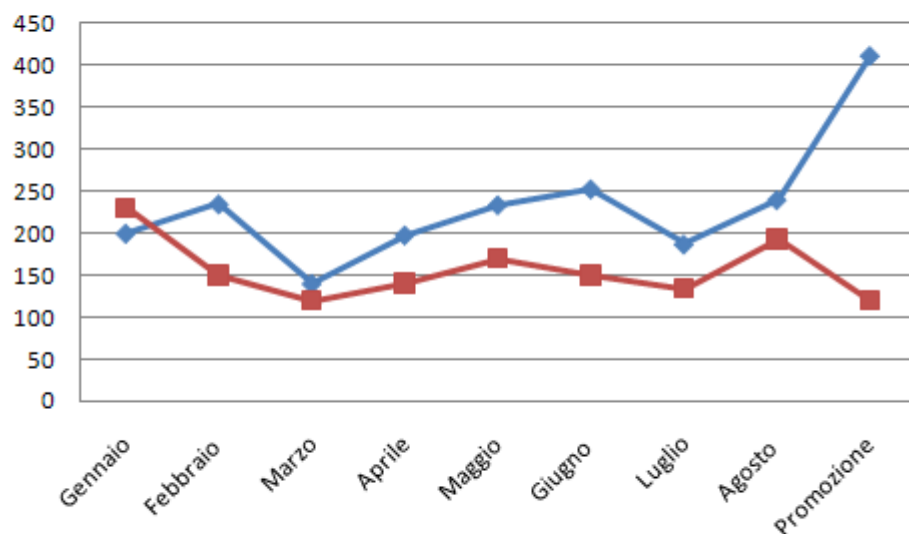
Gennaio	200
Febbraio	235
Marzo	140
Aprile	198
Maggio	234
Giugno	253
Luglio	187
Agosto	240
Promozione	412

[Torna al menu principale](#)

## Confronto con segmento



Negozio	Viterbo (40)
Periodo ricerca	da: 01 agosto 2008 a: 30 settembre 2008
Dettagli promozione selezionata	Codice promodettaglio - 18384 Codice promold - 235 Data inizio promozione - 01 settembre 2008 Data inizio promozione - 15 settembre 2008 Negozio - Viterbo Codice Negozio - 40
Confronta con	[19472] - Yogurt Yomo Agrumi Conf. 2 Pezzi



Gennaio	200	230
Febbraio	235	150
Marzo	140	120
Aprile	198	140
Maggio	234	170
Giugno	253	150
Luglio	187	124
Agosto	240	193
Promozione	412	120

[Indietro](#)