Enhancing the Operation of 13 GOOD Market by Ranking 5 Vendors with Worst Performance

Evaluate vendor performance based on revenue growth rate

Team 4

I Chun Chao
Yi Chun Chuang
Sherry Wu
Chia Li Chien

BUSINESS GOAL

- Customer: Market manager of 13 GOOD Market

  We are going to help market manager to know about the performance of each vendor, and then help him take some promotional strategies.

Revenue growth

Decisions on promotional strategies

DATA MINING GOAL

- Ranking Top 5 vendors
  whose revenue growth rate have worst performance
- Do prediction on every Wednesday(3 days before market day)
- Calculation:

$$\text{Revenue Growth Rate} = \frac{Revenue}{Average\ Revenue\ of\ last\ 3\ time} \times 100\%$$

- If Revenue Growth Rate≥3% ➡ No (no promotion needed)

  Revenue Growth Rate<3% ➡ Yes (promotion needed)

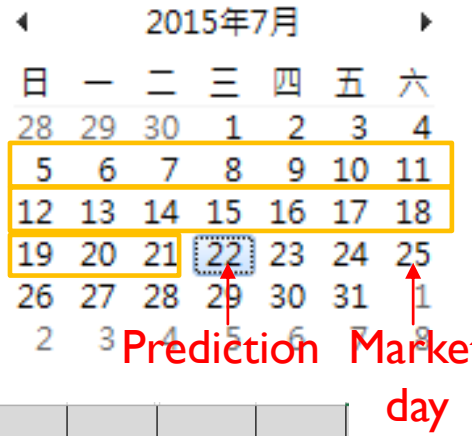
DATA MINING GOAL

Classifying Revenue Growth Rate(RGR)<3%(Yes/No)

Ranking Top5 Vendor(needs promotion)

DATA DESCRIPTION

- (After deleting missing value)There are 33 columns and 1186 rows
- The data includes the vendor information during May 2014 and June 2015
- We include FB data of 17 days ahead prediction.



Date	Vender ID	#fb post	# _ photo	# _ status	# _ link	# _ video link	#post comment	#post like	#post share	New product(1/0)	Activity host (1/0)	Revenue	last 1 time_ revenue	last 2 time_ revenue	last 3 time_ revenue	Temperature of Wed.	Special holiday (1/0)
2014/8/9	12	9	6	1	2	0	12	586	16	0	0	1240	150	150	220	30.9	0
2014/10/18	16	6	6	0	0	0	10	647	16	0	0	1350	100	450	250	25.2	0
2015/9/19	5	9	8	0	1	0	9	408	55	0	0	1500	300	400	240	26.2	0
2014/10/11	16	6	6	0	0	0	19	543	14	0	1	100	450	250	350	24.2	0

Product category _fresh	Product category _cuisine	Product category_ handicraft	Product category_ other	Activity type_ handmade	Activity type_ speech	Activity type_ lunch	F.13-17	F.18-24	F.25-34	F.35-44	F.45-54	F.55-64	F.65+	M.13-17	M.18-24	M.25-34	M.35-44	M.45-54	M.55-64	M.65+	distance from market	season	average last-3-time revenue	Revenue growth rate	RGR<3%
0	0	1	0	0	1	0	19	50	393	657	262	182	22	10	20	74	168	175	116	17	20.5	summer	173.333	615.38%	No
0	1	1	0	0	0	0	12	38	328	550	175	101	13	8	11	69	106	104	50	12	10.7	summer	266.667	406.25%	No
0	0	0	1	1	0	0	23	54	247	546	234	93	19	8	28	69	134	132	46	13	20	summer	313.333	378.72%	No
0	1	1	0	1	0	0	15	32	296	562	182	100	16	9	9	59	104	127	47	12	10.7	summer	350	-71.43%	Yes

DATA DESCRIPTION

Numerical(26)		Catagorical(7)
#fb_post	F.13-17, F.18-24, F.25-34, F.35-44, F.45-54, F.55-64, F.65+	New product(1/0)
#_photo		Activity host(1/0)
#_status		Special holiday(1/0)
#_link	M.13-17, M.18-24, M.25-34, M.35-44, M.45-54, M.55-64, M.65+	Product category(4 categories)
#_video link		Activity category(3 categories)
#_post comment		Season(2 categories)
#_post like	Average last-3-time revenue	RGR<3%(Yes/No) ←**Output**
#_post share	Revenue growth rate	
Temperature of Wed.	Distance from market	

METHOD

- **Data Pre-process:**

Missing value, Create dummy variables

- **Dimension Reduction: using PCA**

- **Logistic Regression**
  - ✓ Partition (Training:Validation:Test)0.5:0.3:0.2
  - ✓ Performance evaluation

- **Naïve Bayes**
  - ✓ Create binned variables
  - ✓ Partition (Training:Validation:Test)0.5:0.3:0.2
  - ✓ Performance evaluation

METHOD

- We need 12 principal components to account for 75% of total variance.
- Our data are not very correlated and we also get worse predictive performance

→PCA doesn't work well

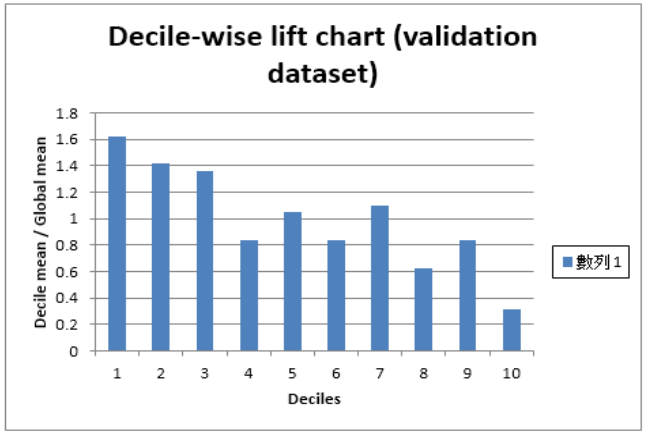
Variances										
	1	2	3	4	5	6	7	8	9	10
Variance	8.505373	4.1102	2.596652	2.401311	2.264598	1.826042	1.530245	1.352465	1.28731	1.2238129
Variance Per	21.80865	10.53897	6.658083	6.157209	5.806662	4.682158	3.923705	3.467858	3.300795	3.1379817
Cumulative V	21.80865	32.34762	39.00571	45.16291	50.96958	55.65173	59.57544	63.0433	66.34409	69.482074

PERFORMANCE EVALUATION

Logistic Regression

Error Report (Training Data Scoring)			
Class	# Cases	# Errors	% Error
Yes	295	84	28.47457627
No	289	89	30.79584775
Overall	584	173	29.62328767

Error Report (Validation Data Scoring)			
Class	# Cases	# Errors	% Error
Yes	191	69	36.12565445
No	159	58	36.47798742
Overall	350	127	36.28571429



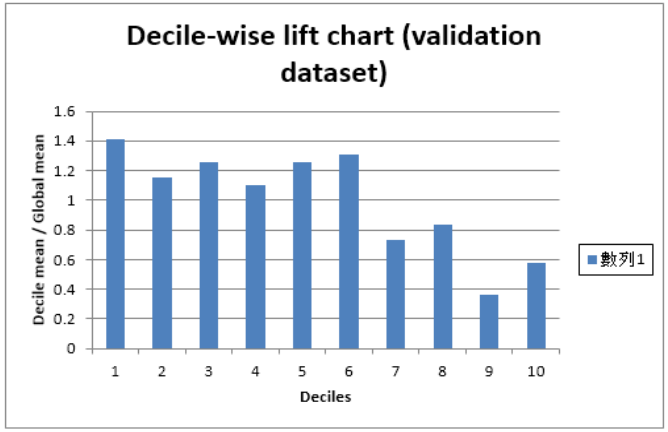
Naïve Rule

Error Report			
Class	# Cases	# Errors	% Error
Yes	616	0	0
No	552	552	100
Overall	1162	552	47.5043

Naïve Bayes

Error Report (Training Data Scoring)			
Class	# Cases	# Errors	% Error
Yes	295	91	30.84746
No	289	81	28.02768
Overall	584	172	29.45205

Error Report (Validation Data Scoring)			
Class	# Cases	# Errors	% Error
Yes	191	65	34.03141
No	159	59	37.10692
Overall	350	124	35.42857



RECOMMENDATIONS

Conclusions

- Why we choose Ranking of vendors, instead of classify every vendor ?
  - Avoiding the condition that all vendors perform bad and all need promotion.
  - Considering the usefulness and costs(time and labor) of promotion.
- The information of new vendors can't be used because of the lack of last_time revenue.

Improvements

- Try different binning(for Naïve Bayes)
- Include more external data to improve accuracy

THANK YOU


YAMAHA


草地講堂
多肉盆栽DIY
十時間：10/3(六)
第一場 15:00-16:00
第二場 16:30-17:30
十地點：新民屋四方草坪
十費用：250元/人


新 埔 鎮
北平社區