PentaAnalytics

Inteligencia de Negocios

Predicting Out of Stock using Transactional Data for a Supermarket Chain

Luis Aburto, Departamento de Ingeniería Industrial. Universidad de Chile

Trinidad Figeroa, Departamento de Ingeniería Industrial. Universidad de Chile

ALIO – INFORMS

June 7th, 2010

Agenda

1. Business Problem
2. Objectives
3. Methodology
4. Available data & creation of input variables to predict OOS
5. Model evaluation
6. Conclusions

Supermarket Industry

PentaAnalytics
Inteligencia de Negocios

- Supermarket industry in constant growing.
  - August 2008 : 817 stores → April 2009 : 972 stores (19%).
- Suppliers release new products increasing the numbers of SKY to be managed in store.
  - 10,000 – 100,000 sku depending on the size of the store.

Supermarket Industry

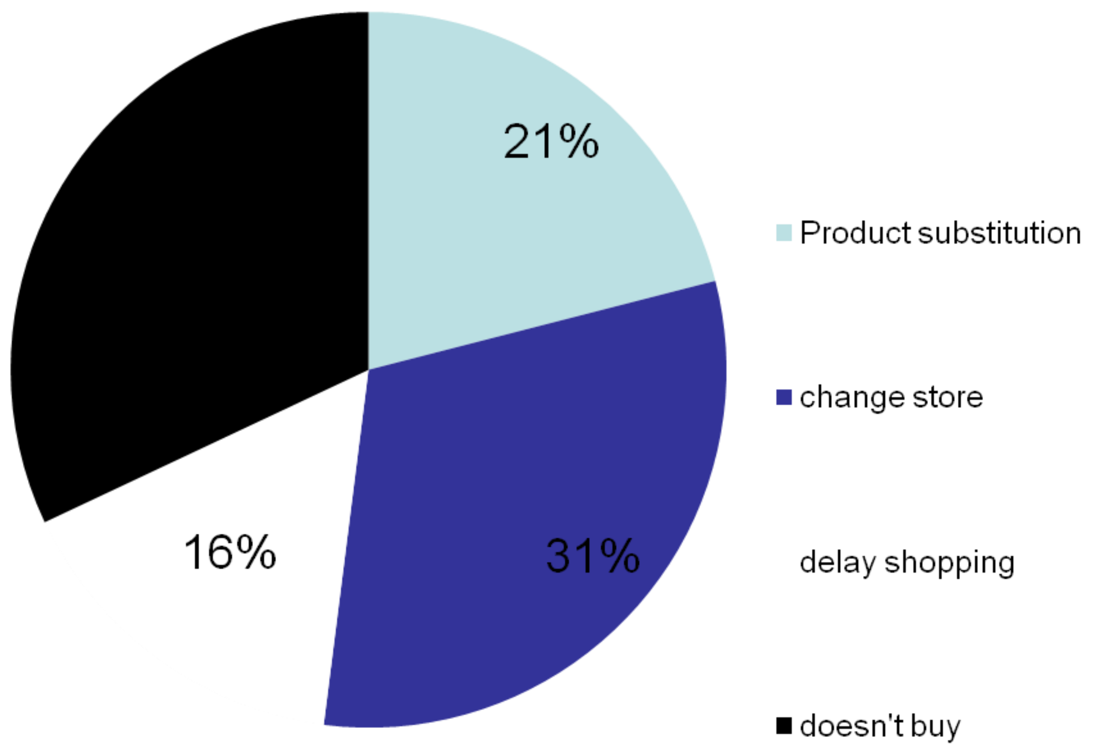
- Out of Stock (OOS) or Out of Shelf:
  - KOM 2005 estimate OOS= 15% in Chile.
  - Cost to industry of 10% of revenue.
  - 78% of OOS are produced due to problems in store replenishment.

Causes of the OOS

PentaAnalytics
Inteligencia de Negocios

- Difference between theoric and

Consumer reaction when OOS happen



Objectives

PentaAnalytics

General:

Detect OOS using predictive model based on POS transactional data

In other words:

- Compare model performance in different categories.
- Identify variables related to detect OOS.
- Evaluate different supervised models to predict OOS.
- Evaluate generalization of predictive models to extend to other categories.

Methodology

PentaAnalytics
Inteligencia de Negocios

1. Analyze POS transactional data and define categories to analyze.
2. Inspect category shelves and detect OOS to define target variable to be predicted
3. Define input variables from POS related.
4. Build and evaluate supervised models: logistic regression and decision trees.
5. Apply predictive model to other categories and evaluate generalization capacity

Available data

- Two categories: Liquid Milk (fast product) and Diapers (slow product).
- OOS measures for 19 days.
- Daily POS data for two years.

Variables used in predictive model

PentaAnalytics
Inteligencia de Negocios

- **Dependent Variable**:
  - OOS= { 1 If out of stock exists; 0 if don't }
- **Independent variables:**

Sales behavior
- Units sold previous day
- Unit sold 7 days ago
- Avg unit sales last week
- Avg unit sales last month
- Sales growth
- Daily sales standard deviation for last week
- Daily sales standard deviation for last month

Coeficiente de variación
- Coefficient of variation daily (last week) $\frac{\sigma_{semana}}{\mu_{semana}}$
- Coefficient of variation daily (last month) $\frac{\sigma_{mes}}{\mu_{mes}}$

Record of OOS
- % days with sales= 0 last week
- % days with sales =0 last month
- Dummy variables for each supplier.
- Dummy variables for each day of the week.

Standarized variables.

Predictive models

PentaAnalytics
Inteligencia de Negocios

- **Logistic Regression:**

  - OOS probability for sku i for category j:

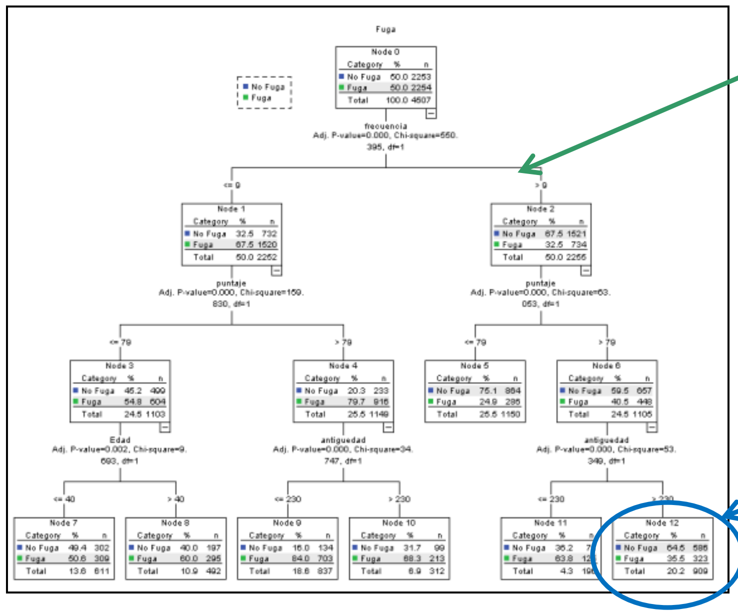
$$P_{ij}(z) = \frac{1}{1 + e^{-z}}$$

  - where :

- $P_{ijk}(z) = probabilidad\ de\ de\ que\ el\ sku\ i\ de\ la\ categoría\ j\ quiebre\ el\ día\ k.$
- $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_K X_K$
- $X_1, X_2 \ldots X_k\ \ variables\ predictivas\ del\ modelo$

Predictive models

- **Decision trees:**
  - Mathematical model for classification
  - Rules



Cada rama es una regla del árbol

Each leaf has a probability of OOS

- Algorithms used: C5, CHAID, CART

Model Evaluation

- Errors:

		PREDICTION		
		No Quiebre	OOS	Efficiency
REAL	No quiebre	a	Error type 1	$\frac{a}{\text{Error tipo 1} + a}$
	OOS	Error type 2	b	$\frac{b}{\text{Error tipo 2} + b}$
	Efficacy	$\frac{a}{\text{Error tipo 2} + a}$	$\frac{b}{\text{Error tipo 1} + b}$	

- **Efficiency:** % of the real OOS detected by the model
- **Efficacy:** predictive capacity. % of OOS predicted that are correct.

Model Evaluation

•Cost Curve:

$$\boxed{Costo = (Costo_{error1}) * (Error\ tipo\ _1) + (Costo_{error2}) * Error\ tipo\ _2}$$

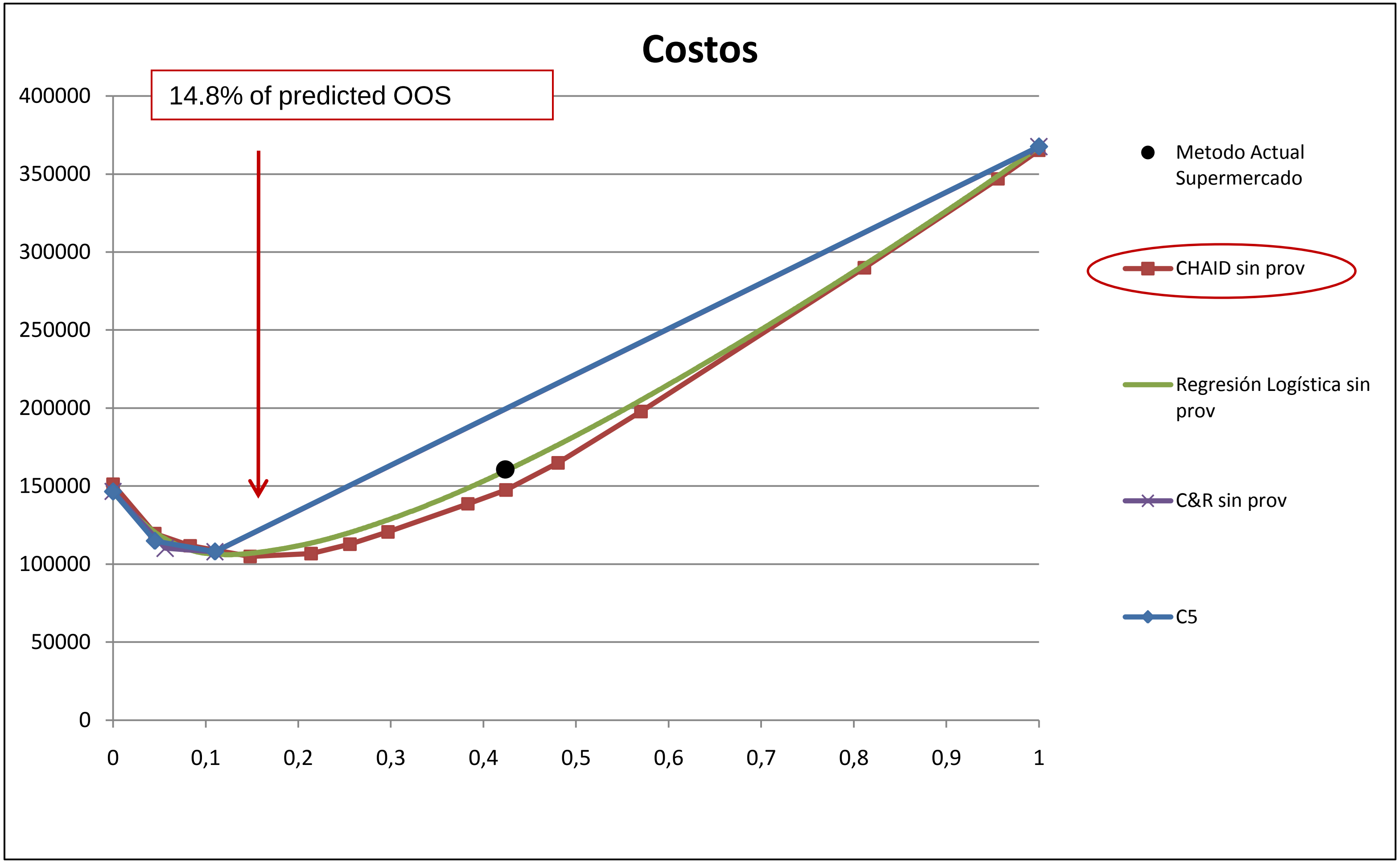
Minimize total cost.

- Cost types:
  - type 1Error cost : Missing time for replenishment people to check stock for a category with products. ($0,5)
  - type 2Error cost : Missing sales ($1)
  - .

Categoría	Valor venta perdida
Leche Liquida	$ 717
Pañales Desechable	$ 294
Margarina	$ 410
Leche en polvo	$ 179

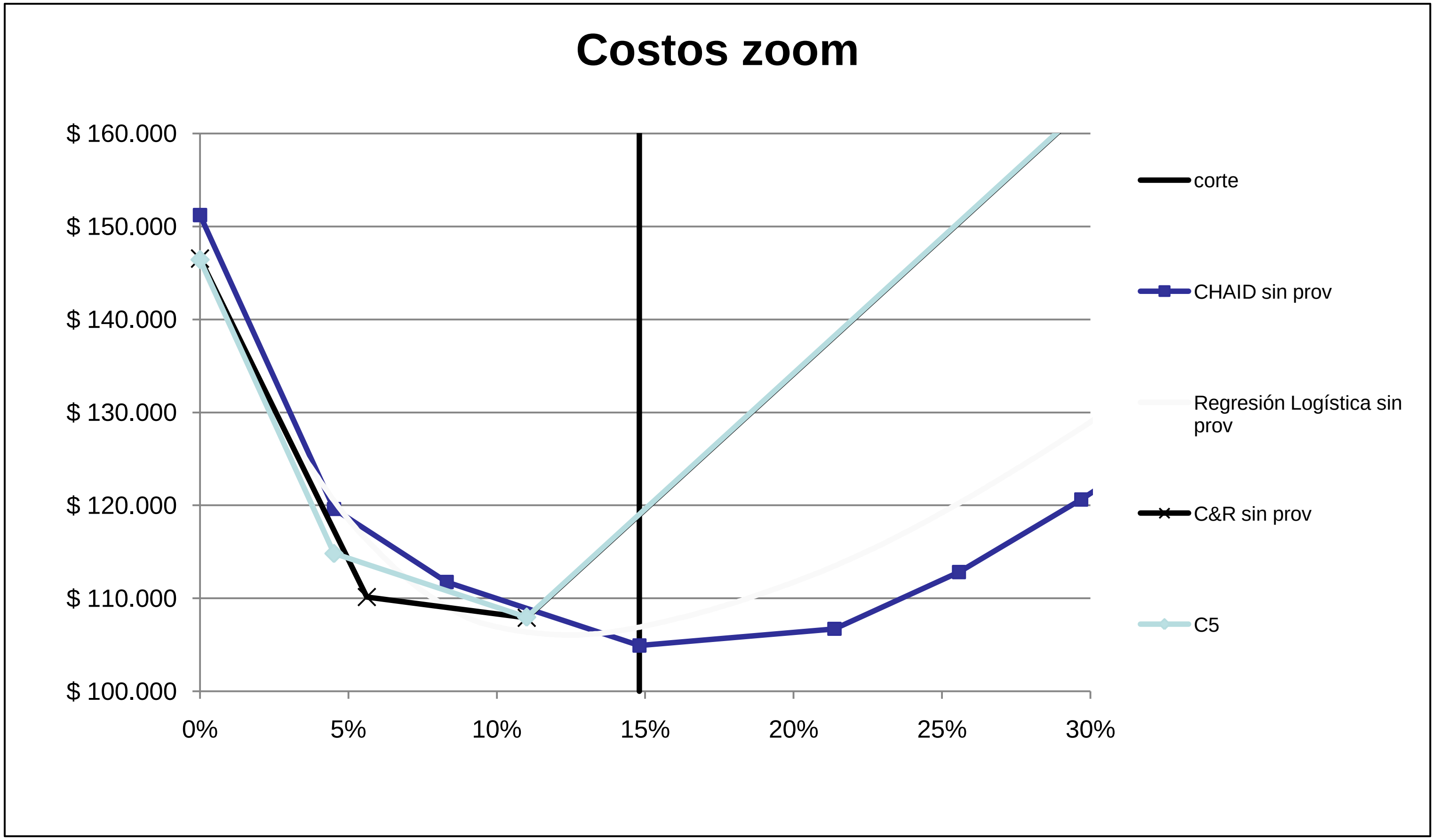
Liquid Milk Results

PentaAnalytics
Inteligencia de Negocios



Liquid Milk Results - zoom

PentaAnalytics
Inteligencia de Negocios



Chaid Model for Liquid Milk

PentaAnalytics
Inteligencia de Negocios

- Confusion Matrix

Train Set

Real	Prediction		Eficiencia
	0	1	
0	977	82	
1	99	105	51%
Efficacy		56%	

% OOS predicted= 15%
% OOS detected= 51%

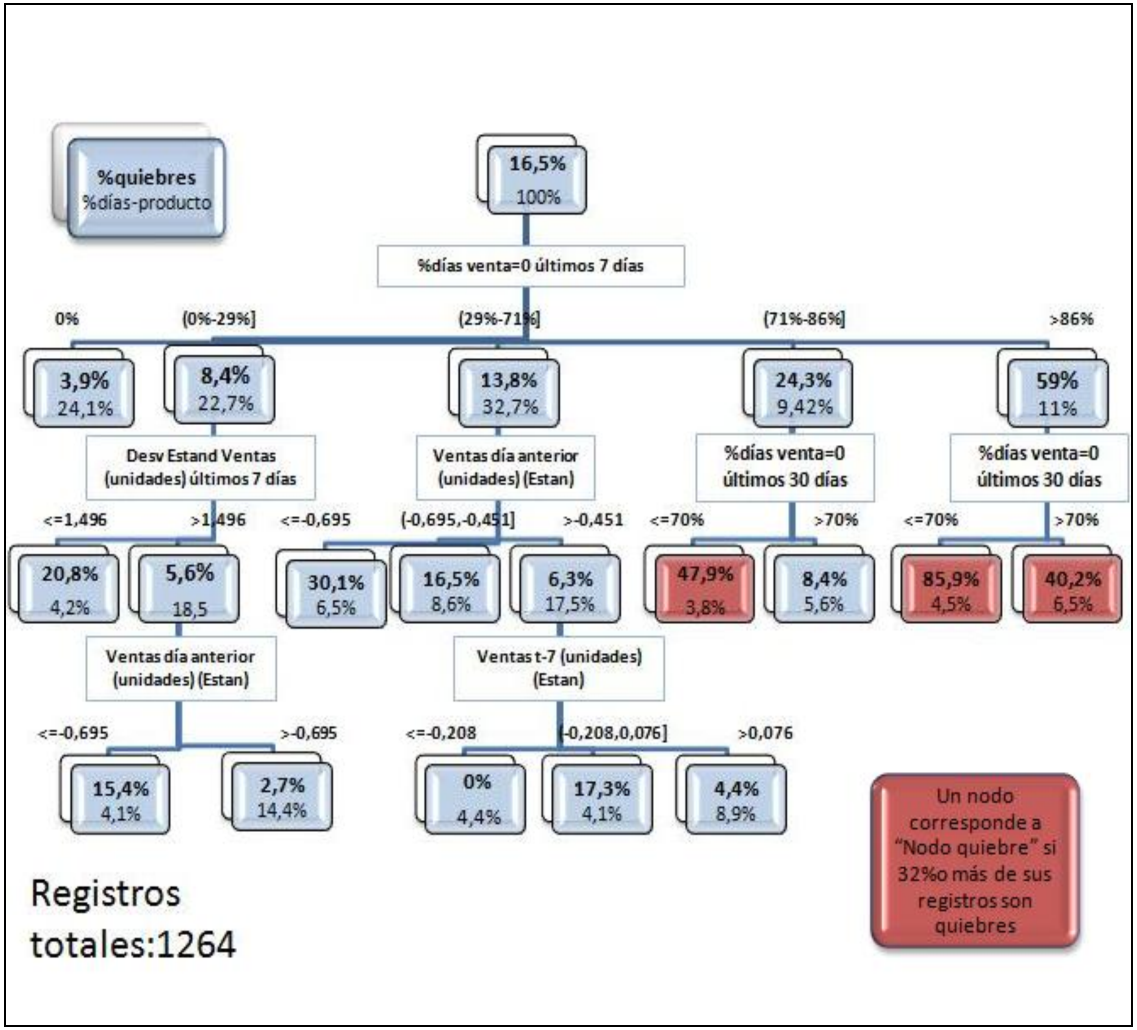
Evaluation Set

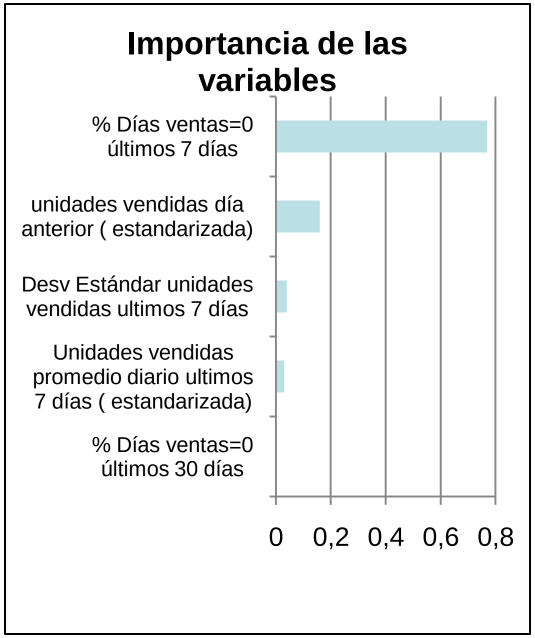
Real	Prediction		Eficiencia
	0	1	
0	245	30	
1	26	29	53%
Efficacy		49%	

% OOS predicted= 18%
% OOS detected= 53%

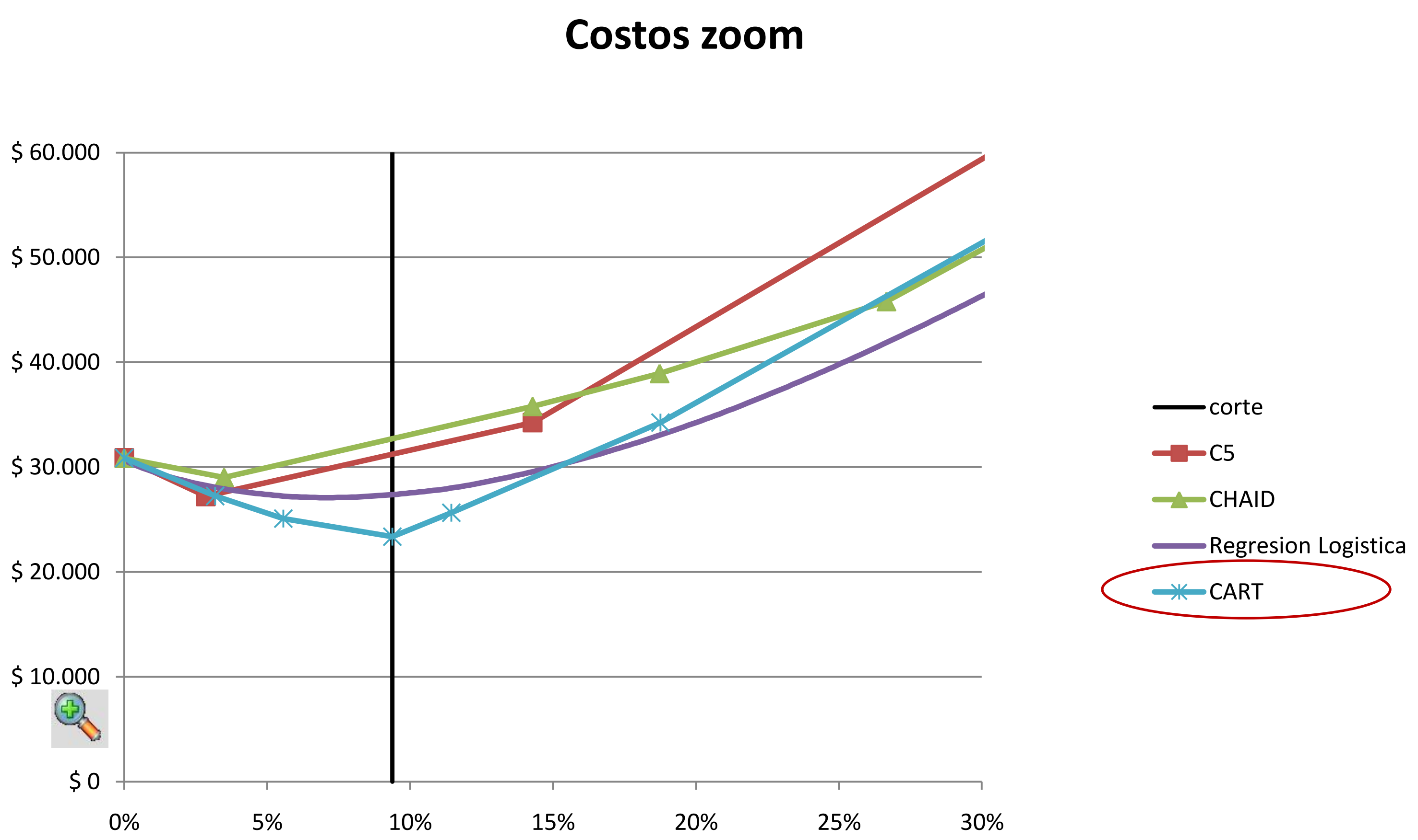
OOS Chaid Model

PentaAnalytics
Inteligencia de Negocios





Diapers category Results



Fuente: Elaboración Propia

Cart model for Diapers OOS

PentaAnalytics
Inteligencia de Negocios

- Confusion Matrix

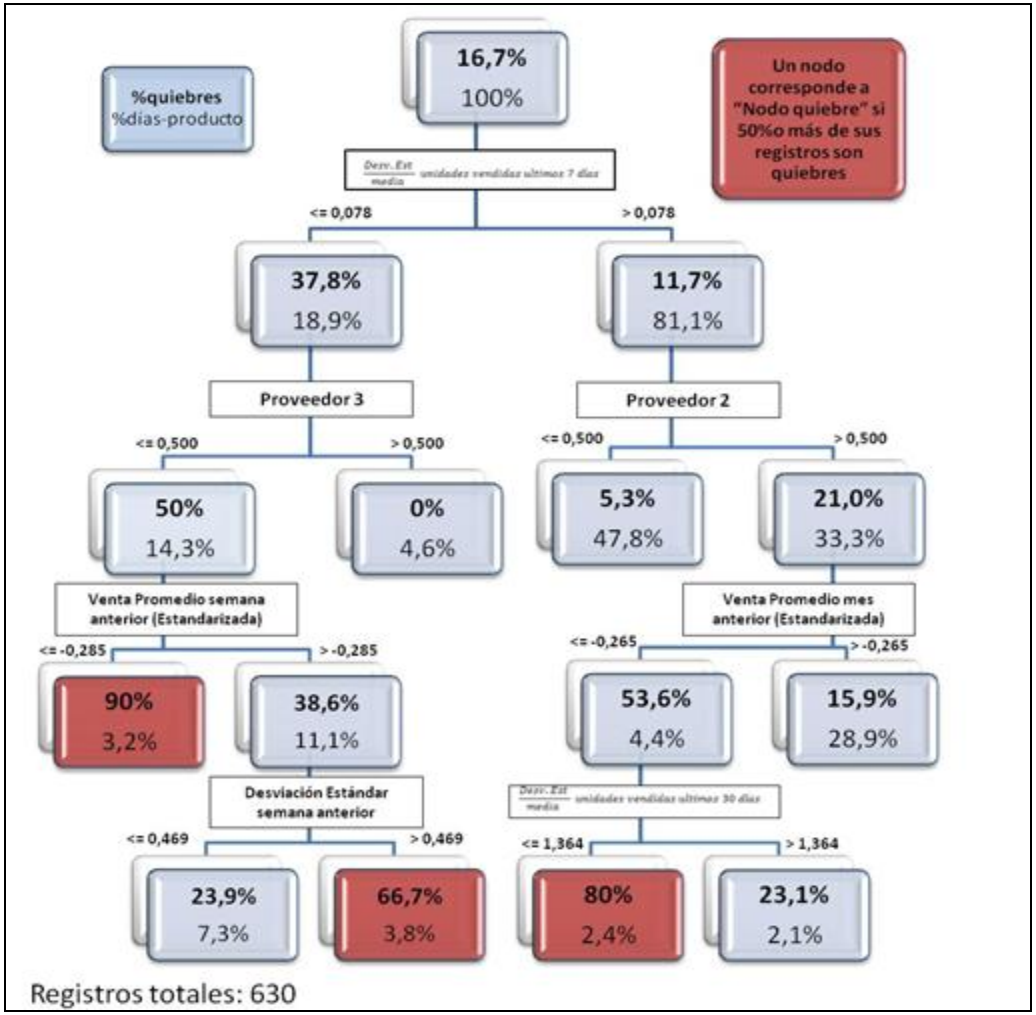
Train Set

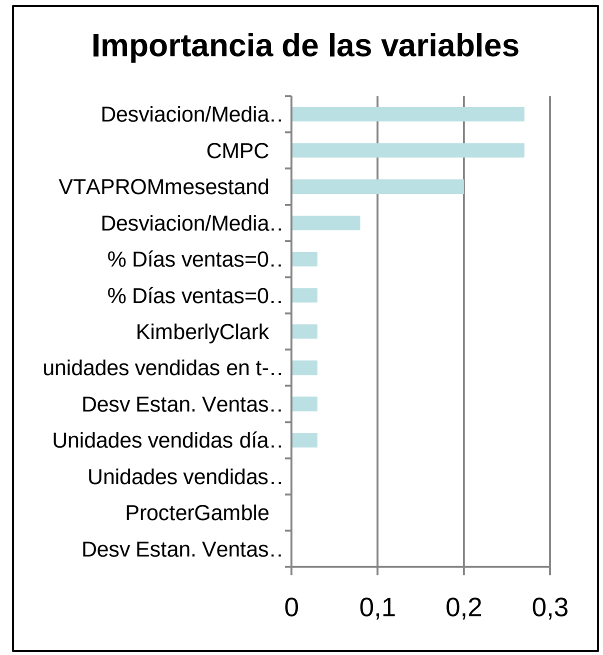
Real	Prediction		Efficiency
	0	1	
0	512	13	
1	59	46	44%
Efficacy		78%	
		% OOS predicted=	9%
		% total OOS=	44%

Validation set

Real	Prediction		Efficiency
	0	1	
0	128	3	
1	14	12	46%
Efficacy		80%	
		% OOS predicted=	10%
		% total OOS=	46%

Cart model for predicting diapers OOS





Model Generalization – high demand

PentaAnalytics
Inteligencia de Negocios

Category Margarine

Train Set

Models	% OOS pred	Cost	Efficiency	Efficacy
Liquid Milk	7%	$ 39,363	36%	48%
Diapers	3%	$ 50,811	3%	8%
Model margarine	10%	$ 23,562	73%	67%
Actual model Sales=0	24%	$ 84,696	61%	23%

Validation Set

Models	%OOS pred	Cost	Efficiency	Efficacy
Liquid milk	10%	$ 10,125	60%	55%
Diapers	3%	$ 15,423	0%	0%
Model margarine	11%	$ 8,895	70%	58%
Actual model Sales=0	30%	$ 27,255	90%	26%

Model Generalization – low demand

PentaAnalytics
Inteligencia de Negocios

Powdered milk category

Train

Modelos	% OOS pred	Costo	Eficiencia	Eficacia
Liquid milk	24%	$ 73,671	48%	21%
diapers	8%	$ 31,659	21%	29%
Powdered milk	12%	$ 23,100	64%	57%
Actual model Sales=0	56%	$ 160,338	55%	23%

Validation

Modelos	%OOS pred	Costo	Eficiencia	Eficacia
Liquid milk	19%	$ 14,673	40%	27%
diapers	7%	$ 6,882	30%	50%
Powdered milk	14%	$ 8,931	50%	45%
Actual model Sales=0	56%	$ 38,649	60%	27%

Conclusions

PentaAnalytics
Inteligencia de Negocios

- **Differences in prediction models for a slow and fast product.**

- **Actual supermarket model tends to over estimate real OOS.**

- **High rotation categories should have a specific model due to cost of OOS. Low rotation categories can be predicted with one model**

- **Extend the work to analyze other categories and stores**

PentaAnalytics
Inteligencia de Negocios

Predicting Out of Stock using Transactional Data for a Supermarket Chain

Luis Aburto, Departamento de Ingeniería Industrial. Universidad de Chile

Trinidad Figeroa, Departamento de Ingeniería Industrial. Universidad de Chile

ALIO – INFORMS

June 7th, 2010