sm educamos

# Grupo SM – Avoiding Non-Use of Textbooks

Group Datathon

# Our approach to generate value with analytics for Grupo SM

## The different developing phases on a glance

**Preprocessing phase**

Business and Data Under-standing

**Exploring phase**

Data Modeling & Clustering

**Assessment phase**

Validation

**Value phase**

Business application

Insights

**Avoiding Non-Use of Text-books**

## Description

### Preprocessing phase

- Problem and Business Understanding
- Checking for relevant features
- Using different features from different years
- Merging data from different years

### Exploring phase

- Dealing with very unbalanced classes
- Explorative data analysis

### Assessment phase

- Cross-fold validation (5 –folds) of our results
- Balancing the training dataset to achieve a better Model-Score (F1-Score)

### Value phase

- Generation of insights
- Prediction of “Non-Use”-Cases
- Developing our sales approach based on our predictions and the data itself

# Our Random-Forest predicition model build the fundament for our business strategy

## Understanding the Random Forest Algorithm

*Illustrative example*

| ID | Curso | TipoMaterial | Asignatura | Prediction |
|---|---|---|---|---|
| 1 | 24 | 1 | 79 | Use (0) |

**Input**

**All Tree Predicitons:**

Tree 1: Use
Tree 2: Use
Tree 3: N-Use

**Random Forest Predicition:**
Use

## Description

### How it works

- The algorithm checks each record
- Many different decisions are made on the basis of different decision trees
- The trees act as an "ensemble", the Random Forest Algortihmus delivers the result with the highest frequency

### Advantages

- Returns uncorrelated outcomes
- Algortihm delivers very valid and accurate results
- Easy to understand
- Very fast model

### Special highlight of our algorithm

- The model was optimized with regard to a very high recall score
- The "non-use" prediction is therefore very likely to be correct
- F1-Score: 0.83

# The deep analysis of the previous data help to understand the environment

## Non-use subjects

**Ca. 70% of all „non-use-cases“ are related to only 10 subjects**

*(Read: 10% of all “non-use cases” are related to "Musica”)*

**7076 analyzed clients**

**975.76 € av. value (per class)**

## Non-use by curso

# Our market research shows that Grupo SM are in a dynamic and competitive environment

## Industry trends and influences

- **High spending in private education**
- **Majority of children in Spain attend state schools**
- **Autonomous regions control their own education systems**
- **Slow trend towards digitalization in education**

## Market development

- The recent years show a strong growth in the production of educational books
- Group „infantil“ and „secondary education“ showed the highest growth rates in the previous years
- Grupo SMs main competitors are: Editorial Planeta, Peguin Random House, Editorial Santillana

**Publishing Production(Spain) - Titels by subject**

| | 2014 | 2015 | 2016 | 2017 | % Variación 2017/2016 |
|---|---|---|---|---|---|
| **TOTAL** | **78.508** | **80.181** | **81.496** | **87.262** | **7,1** |
| Texto no universitario | **17.203** | **18.109** | **18.145** | **19.108** | **5,3** |
| Educación infantil | 3.094 | 2.743 | 2.703 | 2.921 | 8,1 |
| Educación primaria | 6.618 | 7.237 | 7.033 | 7.016 | -0,2 |
| E.S.O. | 3.587 | 4.581 | 4.805 | 5.104 | 6,2 |
| Bachillerato | 639 | 789 | 893 | 941 | 5,4 |
| F.P. | 318 | 251 | 256 | 248 | -3,3 |
| Libros y materiales complementarios | 2.947 | 2.509 | 2.455 | 2.878 | 17,2 |

*Fuente: Comercio Interior del Libro en España 2017*

## Potential challenges

- **Digital books are cheaper**
- **Competition against cheap online offerings**

## Success factors

- **Long customer relationships**
- **“Fear” of digital theft**
- **High costs for schools to switch to digital platforms**

# The extraordinarily high accuracy of our algorithm builds the basis for a targeted customer scoring approach to detect the most important clients

## Focus group

**What we do:**
- Focus on "non-use" customers
- Target on school level

**How we score:**
- -

**Why we do it:**
- The high accuracy of our algorithm shows us which business would most likely be lost if we do not take action

## Client value

**What we do:**
- Focus on customers with high „ValorClase“

**How we score:**
- Range from 1 to 4 based on the quantile
- The higher the more important

**Why we do it:**
- Focus on the most important customers in terms of financial value

## Loss exposure

**What we do:**
- Focus on customers where we expect a high loss due to „non-use“

**How we score:**
- Range from 1 to 4
- GrupoEdit=1 x ValorClasse

**Why we do it:**
- Avoid loss
- Easier to convince existing cusotmers

## Age group

**What we do:**
- Focus on young ages

**How we score:**
- 1-3 depending on ratio of classes which are taught to younger age groups

**Why we do it:**
- To achieve strategic influence
- Children should associate our brand with school materials

## School type

**What we do:**
- Focus on private schools

**How we score:**
- 0 or 3

**Why we do it:**
- Market research indicate that parents tend to spend more money for education material, if children attend a private school

**Result: Client rating based on the total score, which serves as a priority list for our sales team**

# We plan to be present in every region with a dedicated focus on the most valuable customers

## Our sales approach on a glance

Presence in every region

Focus on valuable cusotmers

Sales goal: 3 Schools per day

Sales expect.: 75% success rate

| Area | Clients | Avg. Client_Value (summed ratio) | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cataluna | 677 | 990.08 | 3 | 34 | 95 | 259 | 147 | 57 | 56 | 17 | 7 | 1 |
| Valencia | 487 | 1203.13 | 0 | 1 | 26 | 119 | 136 | 103 | 81 | 14 | 4 | 3 |
| Islas Canarias | 193 | 993.22 | 0 | 14 | 34 | 34 | 51 | 32 | 22 | 4 | 2 | 0 |
| La Rioja | 12 | 1047.20 | 0 | 0 | 2 | 2 | 5 | 1 | 1 | 1 | 0 | 0 |
| Madrid | 708 | 1227.99 | 0 | 3 | 40 | 140 | 234 | 94 | 123 | 50 | 15 | 9 |
| Navarra | 14 | 1093.50 | 0 | 0 | 4 | 1 | 2 | 1 | 5 | 1 | 0 | 0 |
| Pais Vasco | 41 | 1263.42 | 0 | 0 | 5 | 5 | 9 | 10 | 10 | 1 | 0 | 1 |

## 3 Sales takeaways

1. **Focus on all regions, for strategic market penetration**
2. **Main focus on Cataluna, Valencia, Islas Canarias and Madrid**
3. **A low number of sales people should also focus the small regions, due to their high importance**

# Additional strategic guidlines and recommendations

- Future investments in digital services to meet the challenges of the 21st century
- Offering of E-textbooks and possible ‘Book-subscription‘ models to generate further cash flows and recuce cost basis
- Trend towards second hand: The establishment of a book second hand market (operated by Grupo SM) can generate further cash flows
- „Mass-customizing-approach”: Offering of further (paid) material (i.e. scanning a QR-Code in a book lead to additional exercices or videos)