Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Data Mining Footprints in Academics and Industry in Colombia
A Personal Perspective

Alvaro J. Riascos Villegas
University of los Andes and Quantil

Leveraging Big Data for Sutainable Development
Bogotá

November 29, 2016

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:

  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:

  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:

  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
    1. Crime prediction.
    2. Satellite images illegal mining detection.
    3. Mining medical data.
    4. Text mining.
    5. Data driven marketing.
    6. Scoring and classification.
    7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:

  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
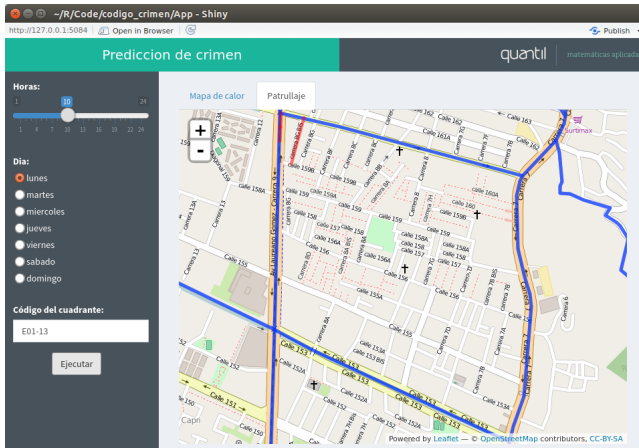Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
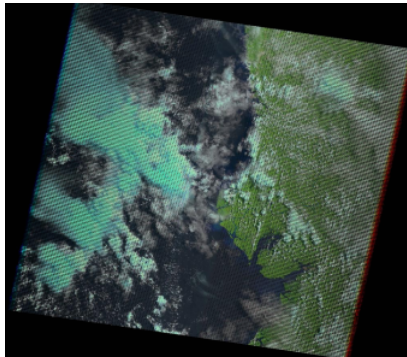Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

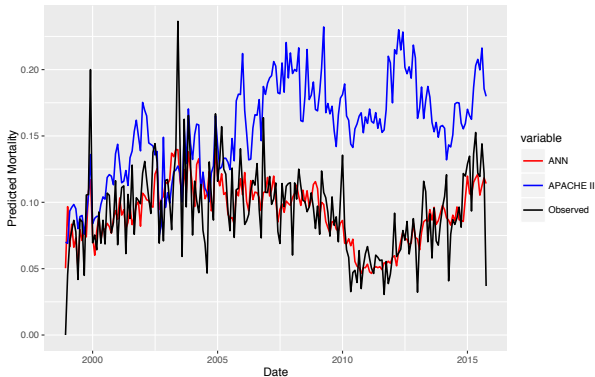**Introduction**
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
  6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Introduction

- I provide the most significant examples of successful data mining application in academics and industry in Colombia: My very personal experience in los Andes and Quantil.

- Focus: Capacities.

- These are applications to:
  1. Crime prediction.
  2. Satellite images illegal mining detection.
  3. Mining medical data.
  4. Text mining.
  5. Data driven marketing.
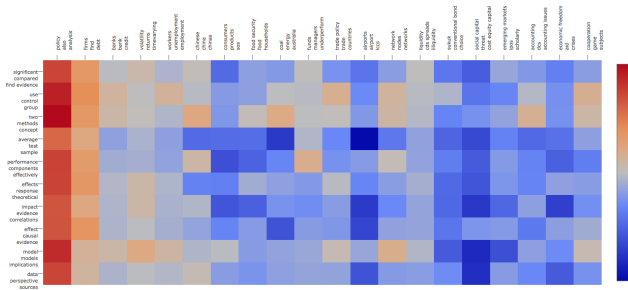 6. Scoring and classification.
  7. Public policy.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

Crime Prediction

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

1. Introduction

2. Crime Prediction

3. Satellite Images Illegal Mining Detection

4. Mining Medical Data

5. Text Mining

6. Data Driven Marketing

7. Scoring and Classification

8. Public Policy

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Satellite Image Illegal Mining Detection

- Courtesy of Santiago Saavedra (PhD Candidate Stanford University)

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Intense care unit data

- Courtesy Fundación Valle del Lili (FVL).

Gráfica: Courtesy of Johns Hopkins University

(a) Spinal Cord   (b) Spinal Cord Injury
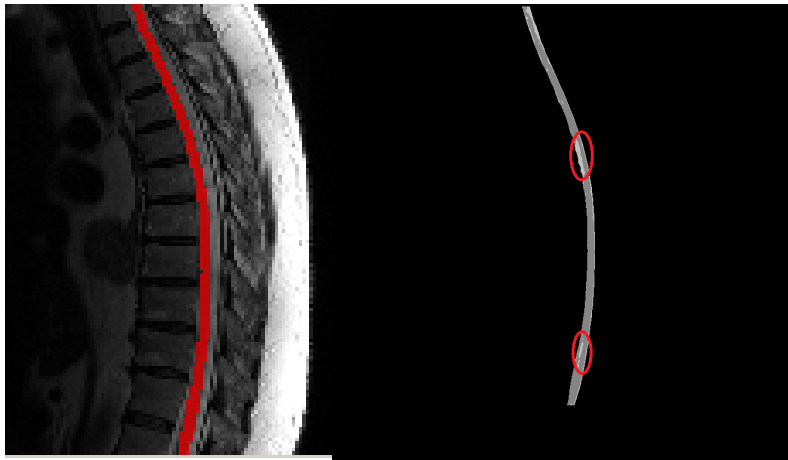
Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
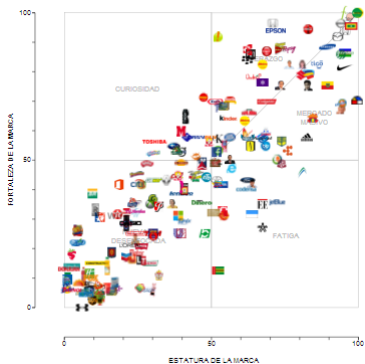Data Driven Marketing
Scoring and Classification
Public Policy

Contents

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

GAP MAPS

- Manual gap maps http://gapmaps.3ieimpact.org/
- Automatic gap maps (work in progress CESED - UNIANDES: crime and drugs gap maps).
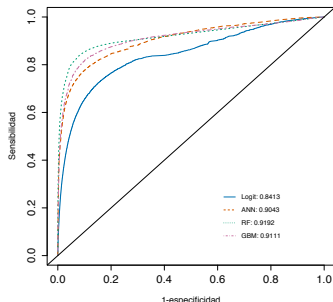
Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

GAP MAPS
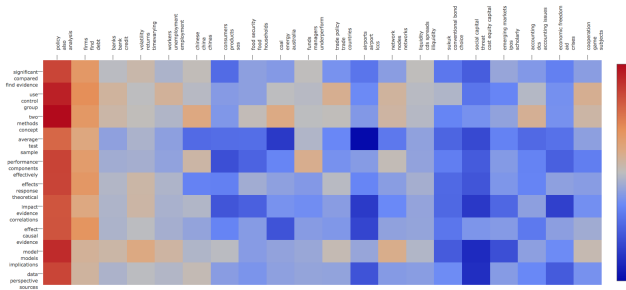
- Manual gap maps http://gapmaps.3ieimpact.org/
- Automatic gap maps (work in progress CESED - UNIANDES: crime and drugs gap maps).

Automatic summarization

- Courtesy of Quantil (independent internal project).

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
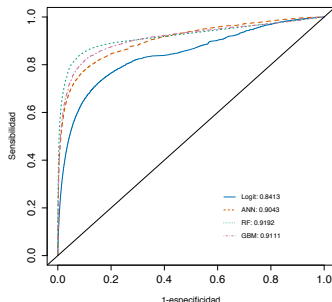**Data Driven Marketing**
Scoring and Classification
Public Policy

Contents

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Data Driven Marketing

- Courtesy: Young & Rubican.
- Mining of social networks data (twitter).

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Data Driven Marketing

- Courtesy: Young & Rubican.
- Mining of social networks data (twitter).

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Scoring and Classification

- Cross industry applications: credit, cross selling, client retention, etc.
- Beating to death the logistic model.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Scoring and Classification

- Cross industry applications: credit, cross selling, client retention, etc.
- Beating to death the logistic model.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Contents

1. **Introduction**

2. **Crime Prediction**

3. **Satellite Images Illegal Mining Detection**

4. **Mining Medical Data**

5. **Text Mining**

6. **Data Driven Marketing**

7. **Scoring and Classification**

8. **Public Policy**

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Health Records

- Colombia has a mandatory competitive health insurance system.

- Payments are determined using information provided by HMOs.

- Health records: 400 million data items per year, 20 million individuals.

- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Health Records

- Colombia has a mandatory competitive health insurance system.

- Payments are determined using information provided by HMOs.

- Health records: 400 million data items per year, 20 million individuals.

- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Health Records

- Colombia has a mandatory competitive health insurance system.

- Payments are determined using information provided by HMOs.

- Health records: 400 million data items per year, 20 million individuals.

- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Introduction
Crime Prediction
Satellite Images Illegal Mining Detection
Mining Medical Data
Text Mining
Data Driven Marketing
Scoring and Classification
Public Policy

Health Records

- Colombia has a mandatory competitive health insurance system.

- Payments are determined using information provided by HMOs.

- Health records: 400 million data items per year, 20 million individuals.

- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

- Courtesy of Ministry of Health and Social Protection.

- Problems: Detect likely cases of tax evasion.
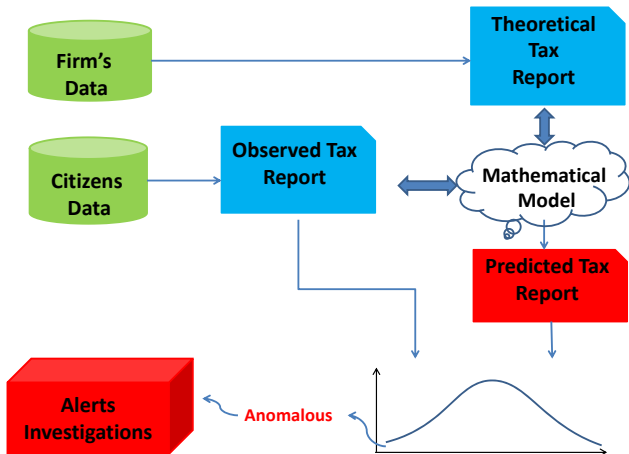
- Problem: Detect anomalous transactions that may be suspicious of money laundry.
- Confidentiality agreements forbids me of explaining further details.

- Problem: Detect anomalous transactions that may be suspicious of money laundry.
- Confidentiality agreements forbids me of explaining further details.

Lawsuits Against the State

- Courtesy of National Agency for Judicial State Defence (ANDJE).

- Courtesy of Ministry of Health and Social Protection.

- Courtesy of National Department of Statistics (DANE).

Histograma Pobreza Subjetiva 2016-1
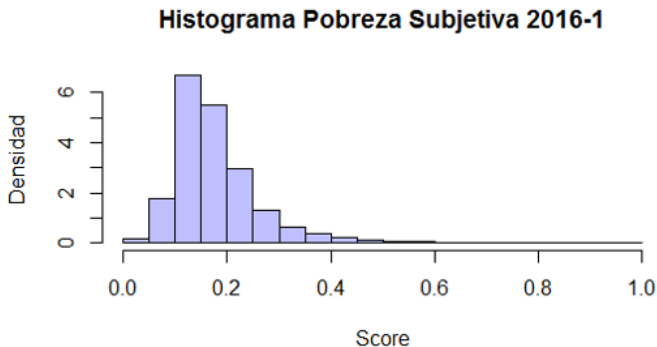
alvaro.riascos@quantil.com.co