



Equipa de Trabalho

Uma imagem com pessoa, vestuário, interior, sorriso

Descrição gerada automaticamente

Maria Inês Cerqueira Couto

(PG50602)

pg50602@alunos.uminho.pt

MEGSI-SIOS

Mariana Araújo Peixoto

(A92917)

[a92917@alunos.uminho.pt](mailto:a92917@alunos.uminho.pt)

MiEGSI

Joana Gonçalves de Sousa

(PG50454)

[pg50454@alunos.uminho.pt](mailto:pg50454@alunos.uminho.pt)

MEGSI-SIOS



Uma imagem com pessoa

Descrição gerada automaticamente

Uma imagem com pessoa, parede, vestuário

Descrição gerada automaticamente

Dong Xuyong

(PG50343)

[pg50343@alunos.uminho.pt](mailto:pg50343@alunos.uminho.pt)

MEGSI-ESI

Indice

Índice de Figuras

# Introdução

No âmbito do projeto proposto na unidade curricular de Sistemas para a Inteligência do Negócio e da Organização, o grupo irá desenvolver um sistema de apoio à decisão para a empresa Bimbo.

Este sistema procura auxiliar a tomada de decisão da referida instituição no que toca às vendas de produtos da Bimbo, localizada no México, durante 9 semanas. O grupo irá utilizar o CRISP-DM como base para o projeto, sendo esperado que todas as suas fases sejam realizadas.

Para a realização deste projeto foram fornecidos quatro datasets por parte da equipa docente contendo dados sobre os clientes, os diferentes produtos fornecidos pela Bimbo, assim como também outros fatores que possam afetar a decisão do mesmo.

# Business Understanding

## Determinar objetivos de negócio frita

## Background

## Objetivos de negócio

## Critérios de sucesso de negócio

# Avaliação da Situação Atual zi

## Inventário dos recursos

## Requisitos, pressupostos e restrições

## Riscos e contingências

## Terminologia

## Custos e benefícios

# Definição dos Objetivos do Data Mining

## Objetivos do Data Mining zi

## Critérios de sucesso do Data Mining frita

Produzir Plano do Projeto

## Plano do Projeto zizi

## Pressuposto Inicial das Ferramentas e Técnicas frita

# Data Understand

# Aquisição inicial de dados

Para a realização do projeto serão utilizados 4 *datasets*, fornecidos pelo docente da unidade curricular.

### Ficheiro: train.csv

Este dataset possui informações relativas … havendo 964000 exemplos e … atributos.

### Ficheiro: producto\_tabla.csv

Este dataset contém informações relativas aos produtos havendo 2592 exemplos e 2 atributos. Possui dados referentes ao produto, nomeadamente Producto\_ID e NombreProducto.

### Ficheiro: town\_state.csv

Este dataset possui informações relativas ao estado e cidade onde foram feitas compras, isto é, Agencia\_ID, Town e State havendo 790 exemplos e 3 atributos.

Ficheiro: cliente\_tabla.csv

Este dataset contém informações relativas aos clientes havendo 935362 exemplos e 2 atributos. Possui dados referentes aos clientes, nomeadamente Cliente\_ID e NombreCliente.

VER COMO COLOCAR ISTO

Cliente filtrado: se é um individuo, empresa,

Produto filtrado: passamos o peso dos produtos todos a gramas (pq havia alguns a kg, …)e fizemos nova coluna chamda descrição , que é a descrição do produto

Adress: junção do town com o state para termos a localização/ endereço. Fizemos um inerjoin com train para obter as empresas que estão na lista dos produtos comprados. ????

Adress\_filtered- manualmente inserimos as coordenadas de forma a ser mais fácil a previsão.

4.df\_adress\_coord- Junção do adress com train

5.1.df\_clean\_droped.csv- fizemos lateração do nome das colunas para ingles e drop /rejeitamos as colunas agencia\_id(mesma razão do cliente, é so um meio para fazer join), canal\_id, cliente\_id (uma vez que não influencia na previsão, ma vez que os clientes estão identificados na previsão), producto\_id, town (porque já temos o adress), state (pela mesma razão), nombre do produto (pq já está identificado na desgricaão, g, marca)

5.df\_clean.csv- junção de tudo numa só tabela, sem correção de dados

# Descrição dos dados

Os dados estão relacionados com transações de vendas de produtos Bimbo no México, em 9 semanas.

Nas tabelas que se seguem são apresentadas informações sobre os atributos dos datasets referidos anteriormente.

|  |  |  |  |
| --- | --- | --- | --- |
| Coluna | Descrição | Tipo | Valores esperados |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Exploração dos dados

# Relatório da Qualidade dos dados

### Conclusão das Análises

# Avaliação da Situação Atual

## Inventário dos recursos

Os recursos necessários para a elaboração deste projeto encontram-se divididos em 4 categorias:

* **Pessoal:** Dispõe-se de quatro alunos do 4º ano do Mestrado Integrado em Engenharia e Gestão de Sistemas de Informação inscritos na cadeira Aprendizagem Automática em Sistemas Empresariais.
* **Dados:** Os dados são fornecidos pelo docente da unidade curricular num ficheiro.csv separado por vírgulas.
* **Hardware:** O grupo tem ao seu dispor quatro computadores portáteis para o processamento de dados.
* **Software:** A equipa de trabalho utiliza a aplicação RapidMiner Studio para a o Data Mininge a criação de modelos, o Talend para a análise e tratamento dados dos, Microsoft Word e Excel para a elaboração da documentação e o Tableau para gerar as dashboards para facilitar a visualização da informação

## Requisitos, pressupostos e restrições

## Riscos e contingências

## Terminologia

## Custos e benefícios