# Inteligência no Negócio para aplicação em Telecomunicações

**Estágio do Mestrado em Engenharia de Software**
**Universidade de Coimbra**

André Miguel Pereira Pinho
Orientador DEI: Prof. Doutor Pedro Furtado
Orientadores Altice Labs: Eng. Helena Margarida, Eng. Ricardo Ângelo

09 de janeiro, 2018

# Conteúdos

1. Enquadramento
2. Problema
3. Objetivo
4. Solução
5. Experiências e resultados
6. Testes
7. Conclusão
8. Metodologia e planeamento

# Enquadramento

- Ecossistema Next Generation Intelligent Network - Policy, Charging and Control (NGIN PCC)
  - Solução de controlo de tráfego e cobrança em tempo real
  - Pretende dar resposta aos desafios colocados
    - Aumento do consumo de serviços de dados
    - Concorrência de outros operadores e de aplicações de dados por internet
  - Duas aplicações
    - Business Intelligence Tools (BIT)
    - Active Campaign Manager (ACM)

- Oportunidade: analisar dados com recurso às tecnologias atuais de análise de dados

# Problema

- Concorrência provoca quebra nos lucros
- Taxa de adesão de cerca de 5% nas atividades de marketing provoca desperdício de recursos financeiros
- Sobrecarga ou desperdício de recursos de rede

# Objetivo

- Desenvolvimento de uma solução de suporte à decisão para alcançar três objetivos:
  - Fornecer insights durante as campanhas para identificar com maior precisão o público-alvo
  - Prever consumos para ajudar no planeamento e gestão de recursos de rede e melhorar a qualidade de serviço
  - Extrair relações entre serviços subscritos pelos clientes

# Solução: Plataforma com modelos preditivos e descritivos

# Requisitos [1/3]

## Funcionais

**Perfil de clientes**
- Extrair insights do perfil de clientes que aderiu ou não a uma campanha
- Cálculo da taxa de adesão e do lift over grupo de controlo

**Previsão de consumo**
- Número de recargas
- Número de eventos SMS
- Volume de dados de internet

**Relações entre serviços subscritos**
Extrair regras do tipo A -> B com o respectivo valor de confiança

# Requisitos [2/3]

## Não funcionais

- Desempenho
  - Realizar o trabalho no tempo especificado
- Interoperabilidade
  - Permitir integração e comunicação com o exterior
  - Formato de dados universal
- Robustez
  - Presença de parâmetros inválidos de entrada
  - Falhas no acesso à fonte de dados

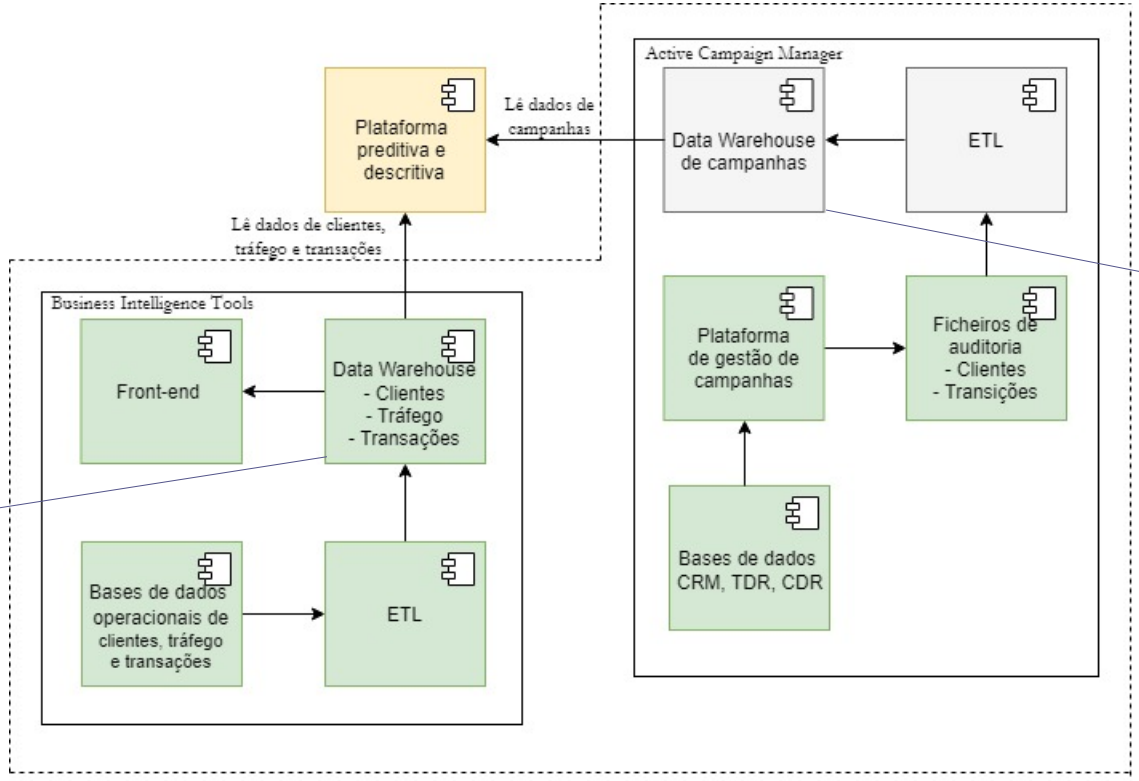
# Requisitos [3/3]

## Restrições técnicas

- Tecnologias open source
- Estilo arquitetural REST

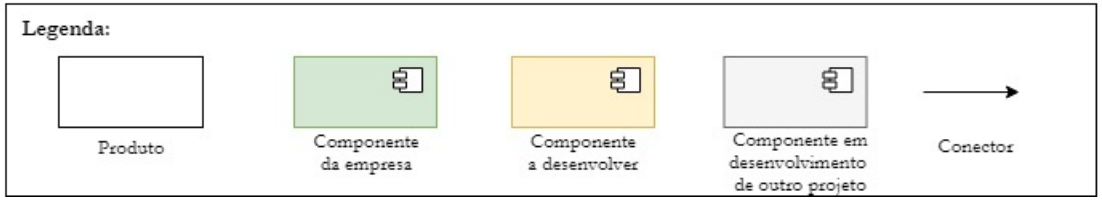
# Desenho de alto nível [1/3]

## Vista de componentes



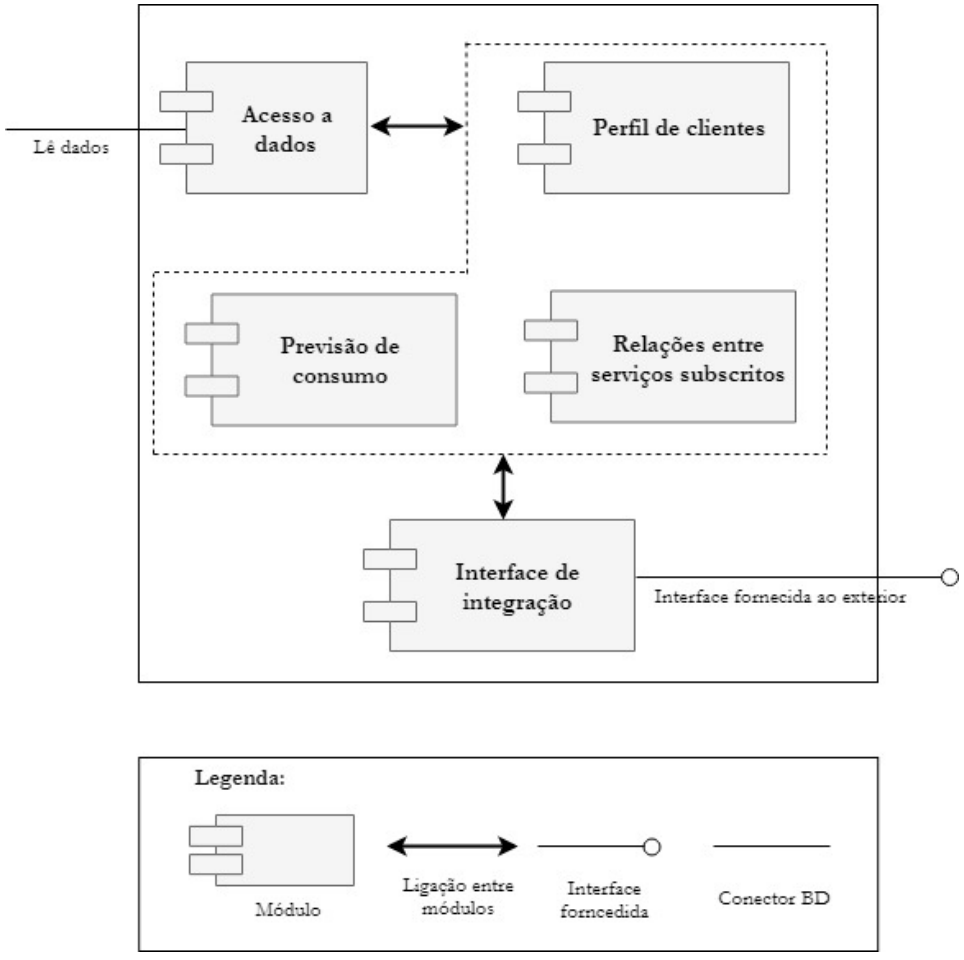

- Clientes
- Tráfego
- Transações

- Clientes
- Campanhas
- Transições de estado



# Desenho de alto nível [2/3]

## Vista lógica da plataforma




Acesso a dados: responsável por conectar e recolher dados das fontes existentes

Interface de integração:

- Fornece endpoints com parâmetros configuráveis
- Comunica com os módulos dos requisitos funcionais

Restantes: responsáveis pelo processo especifico de cada requisito funcional

# Desenho de alto nível [3/3]

## Tecnologias

- Python
  - Flexível para integração e otimização do modelo numa aplicação de suporte à decisão
  - Bibliotecas
    - scikit-learn (clustering e seleção de atributos)
    - mlxtend (regras de associação)
    - statsmodels (séries temporais)
  - Flask usado na Interface de integração
    - API REST com endpoints disponibilizados por HTTP
    - JSON
    - JWT para proteger os recursos da API

# Desenho detalhado [1/6]

## Módulo de perfil de clientes [1/3]

Aplicou-se o método de clustering K-Means

- Agrupa  os elementos descritos pelos seus atributos (consumo, comportamento) baseado na similaridade

Divide-se em duas fases

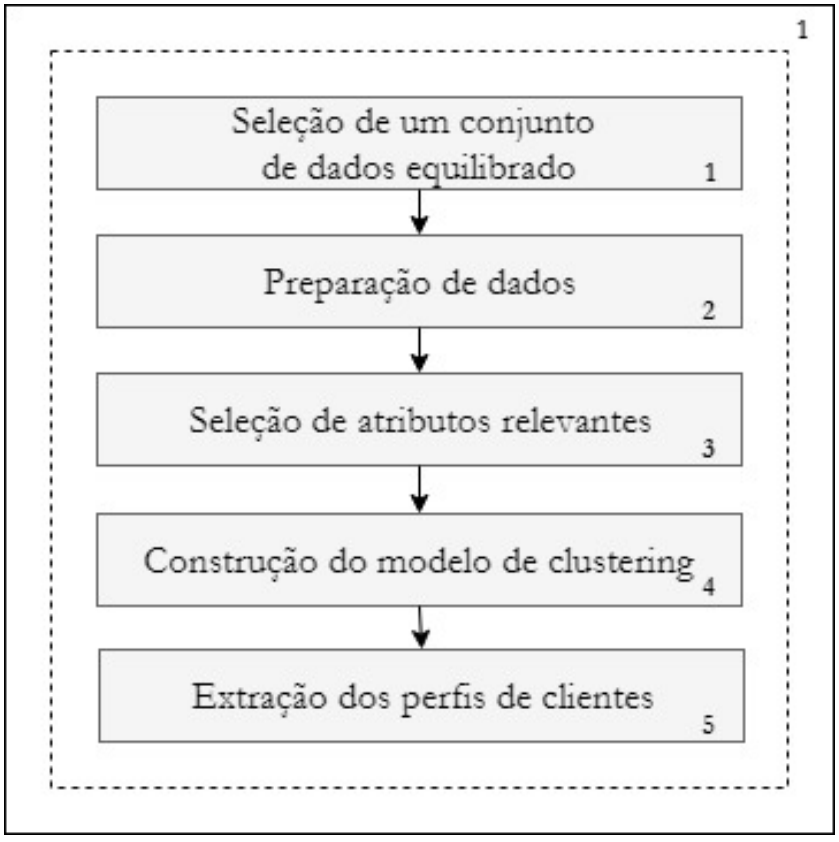
1. Perfil de clientes aderentes ou não a uma campanha - extraí insights
2. Cálculo do lift over grupo de controlo (GC) - mede e avalia o sucesso da campanha. Relaciona a taxa de adesão do público-alvo (PA) com a do grupo de controlo (GC). Fórmula:

$$\text{Lift over GC (\%)} = \frac{\text{Taxa de adesão do PA} - \text{Taxa de adesão do GC}}{\text{Taxa de adesão do GC}}$$

# Desenho detalhado [2/6]

## Módulo de perfil de clientes [2/3]

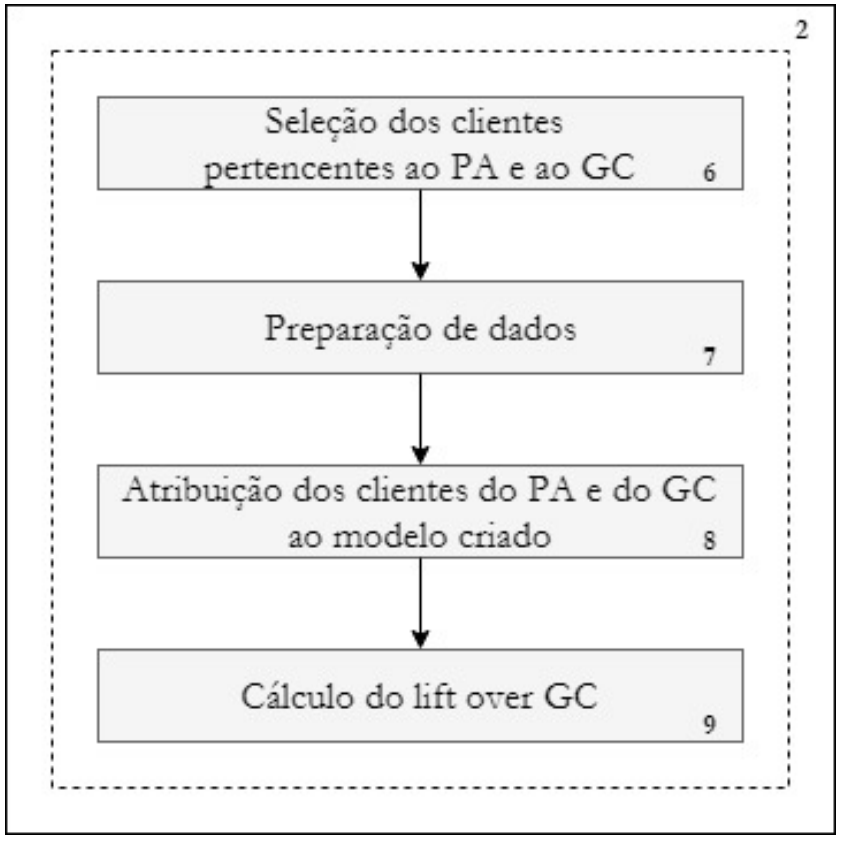
Fase 1 - Perfil de clientes aderentes ou não a uma campanha



# Desenho detalhado [3/6]

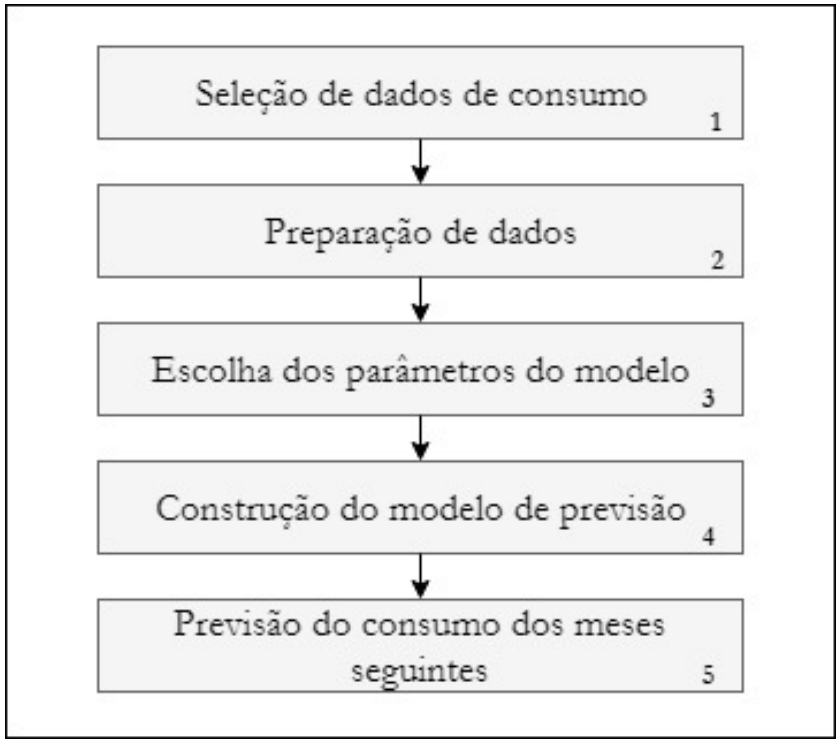
## Módulo de perfil de clientes [3/3]

Fase 2 - Cálculo do lift over grupo de controlo por grupo



# Desenho detalhado [4/6]

## Módulo de previsão de consumo

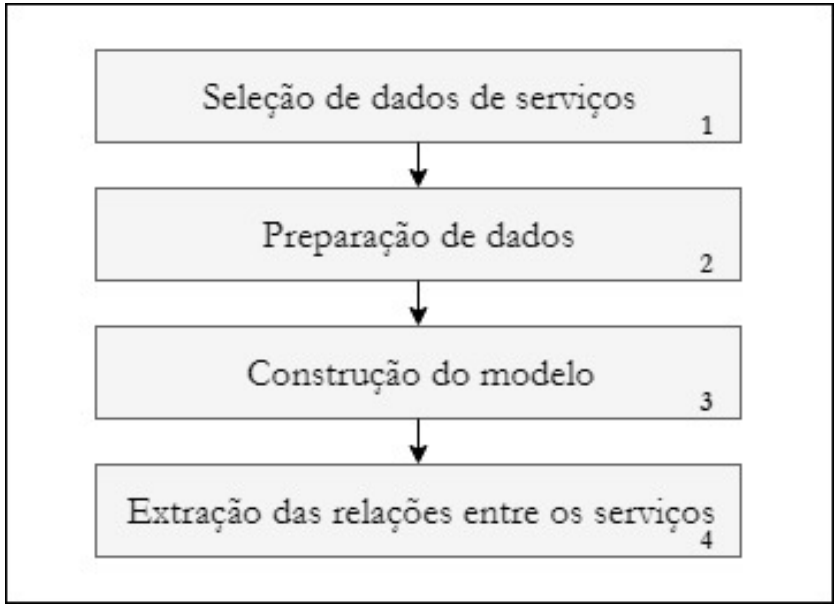



Aplicou-se o método de séries temporais ARIMA

- Dependente do tempo
- Sazonalidade
- Tendência

# Desenho detalhado [5/6]

## Módulo de relação entre serviços subscritos




Aplicou-se o algoritmo de associação Apriori

- Modelo construído com o valor de suporte e confiança recebido por parâmetro
- Valor mínimo de suporte e confiança das regras parametrizável

# Desenho detalhado [6/6]

## Interface de integração

| Método | Endpoint | Parâmetro | Descrição |
|---|---|---|---|
| POST | /login | Login e password de administrador definidos | Devolve o token que permite o acesso a recursos |
| GET | /perfil | Campanha<br>Número de grupos | Devolve informação do perfil de clientes que aderiu e não a uma campanha |
| GET | /previsao | Tipo de consumo | Devolve os valores previstos |
| GET | /relacoesServicos | Ano<br>Mês<br>Suporte<br>Confiança | Devolve um conjunto de regras com o seu grau de confiança |

# Experiências e resultados

# Perfil de clientes [1/5]

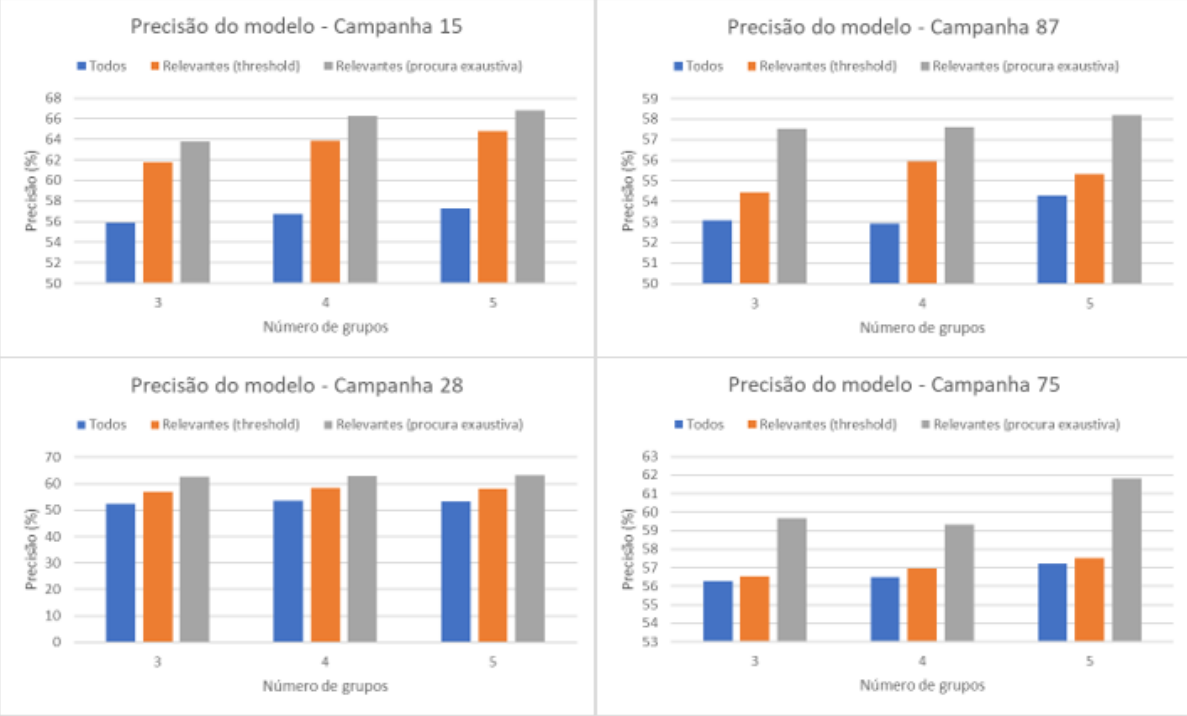
## Experiências com quatro campanhas

Objetivo: separar tanto quanto possível os clientes aderentes ou não em diferentes grupos (ou clusters). Intervalo de 3 a 5 decidido internamente

1. Seleção de atributos com maior importância
   - Threshold definido no método SelectFromModel
   - Procura exaustiva - teste de subconjuntos de atributos para encontrar o ponto ótimo (ou de corte)
2. Métodos de clustering
   - Particionamento K-Means
   - Hierárquico aglomerativo
3. Funções de distância (ou proximidade)
   - Euclidean
   - Manhattan

# Perfil de clientes [2/5]

## Validação da seleção de atributos relevantes

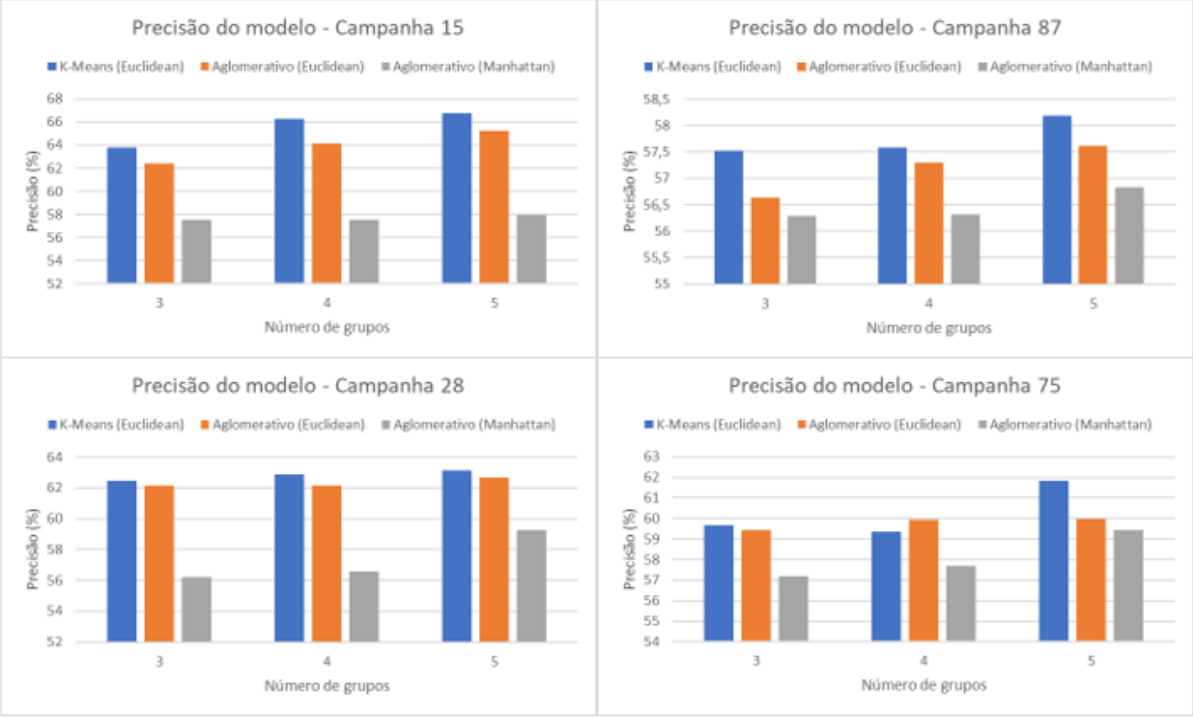



Conclusão

- Modelos com todos os atributos foram os menos discriminatórios
- Modelos por procura exaustiva dos atributos foram mais discriminatórios

# Perfil de clientes [3/5]

## Comparação da precisão de dois métodos e duas funções de distância

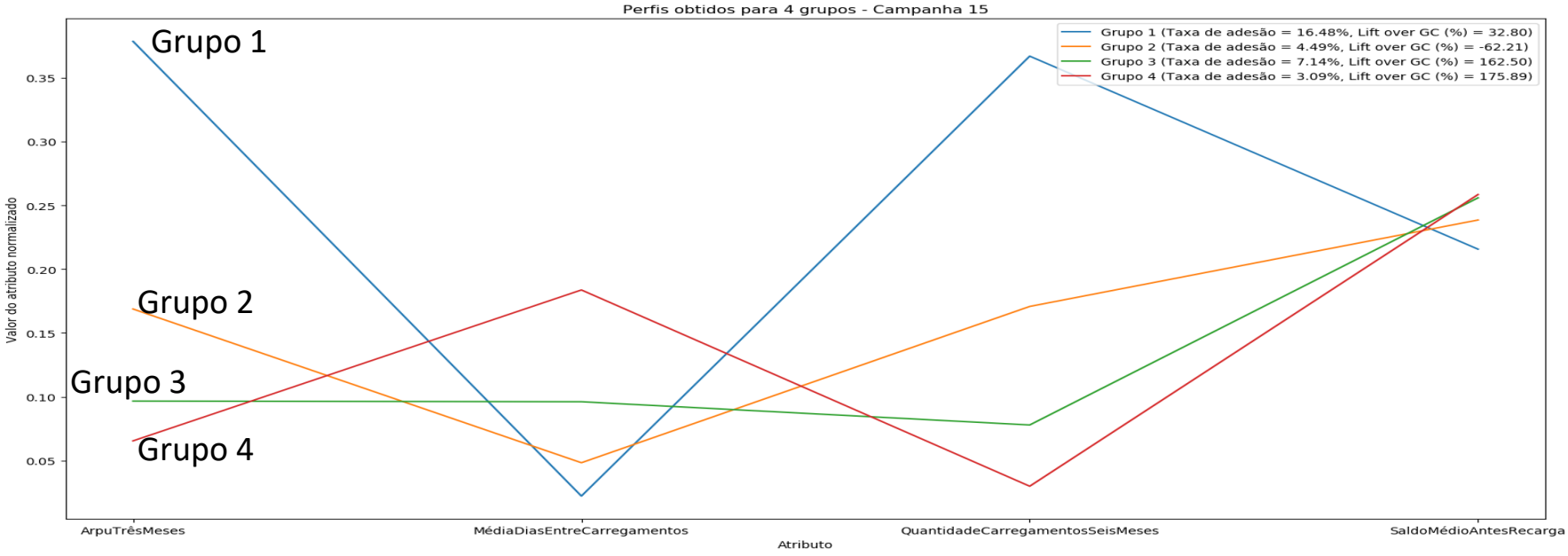



Conclusão

- Modelo K-Means superior
- Modelo com a função de distância Euclidean superior

# Perfil de clientes [4/5]

## Resultado [1/2]




<u>Conclusão</u>

- Bem separados entre si

# Perfil de clientes [5/5]

## Resultado [2/2]

| Grupo | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **Atributos de perfil de cliente** | | | | |
| ARPU a três meses | 62.58 | 24.51 | 11.41 | 5.76 |
| Média de dias entre carregamentos | 5.02 | 9.74 | 18.39 | 34.25 |
| Quantidade de carregamentos a seis meses | 41.0 | 19.61 | 9.49 | 4.26 |
| Saldo médio antes da recarga | 1.68 | 2.35 | 3.01 | 3.08 |
| **Métricas da campanha** | | | | |
| Taxa de adesão (universo de teste) | 88.08% | 75.14% | 51.66% | 35.26% |
| Taxa de adesão (universo total) | 16.48% | 4.49% | 7.14% | 3.09% |
| Lift over GC no universo total (global) = 45.82% | | | | |
| Lift over GC | 32.8% | -62.21% | 162.5% | 175.89% |

Conclusão - taxa de adesão

- Grupo 1: mais aderente
- Grupo 4: menos aderente

Conclusão - lift over GC

- Grupo 3 e 4: clientes que mudaram mais o comportamento. Dependentes do incentivo
- Grupos restantes: independentes do incentivo

# Previsão de consumo [1/2]

## Experiências com três tipos de consumo

Objetivo: minimizar o erro de previsão

1. Métodos de séries temporais
   - Auto-Regressive Integrated Moving Average (ARIMA)
   - Prophet
2. Quatro testes variando o trimestre do ano previsto com histórico de três anos

# Previsão de consumo [2/2]

## Comparação do erro de previsão a 3 meses

Root Mean Squared Error

Mean Absolute Percentage Error

| Método | RMSE | | MAPE | |
|---|---|---|---|---|
| | Média | Máximo | Média | Máximo |
| ARIMA - Recargas | 3.69e+4± 1.31e+4 | 4.97e+4 | 3.71 ± 1.25 | 5.01 |
| Prophet - Recargas | 7.63e+4 ± 5.93e+4 | 1.05e+5 | 7.82 ± 2.54 | 11.08 |
| ARIMA - SMS | 7.95e+5 ± 5.5e+5 | 1.66e+6 | 4.67 ± 3.09 | 9.43 |
| Prophet - SMS | 2.7e+6 ± 2.82e+6 | 4.54e+6 | 14.25 ± 8.76 | 27.4 |
| ARIMA - Dados | 2.59e+13 ±1.48e+13 | 5.12e+13 | 6.99 ± 5.79 | 17.0 |
| Prophet - Dados | 4.67e+13± 4.18e+13 | 6.35e+13 | 8.11 ± 3.39 | 13.26 |

Conclusão

- Modelo ARIMA superior nas três séries temporais

# Relações entre serviços subscritos

## Resultado

| Antecedente | Consequente | Suporte | Confiança |
|---|---|---|---|
| Serviço de carregamento em qualquer lugar | Serviço de dados móveis | 0.68 | 0.7 |
| Serviço de carregamento em qualquer lugar | Serviço de chamadas | 0.83 | 0.85 |
| Serviço de saldo para chamadas urgentes | Serviço de chamadas | 0.83 | 0.84 |
| Serviço de notificação de saldo, Serviço de carregamento em qualquer lugar | Serviço de dados móveis | 0.68 | 0.7 |
| Serviço de notificação de saldo, Serviço de carregamento em qualquer lugar | Serviço de chamadas | 0.82 | 0.85 |
| Serviço de carregamento em qualquer lugar, serviço de saldo para chamadas urgentes | Serviço de dados móveis | 0.67 | 0.7 |
| Serviço de carregamento em qualquer lugar, serviço de saldo para chamadas urgentes | Serviço de chamadas | 0.83 | 0.86 |
| Serviço de carregamento em qualquer lugar, serviço de saldo para chamadas urgentes | Serviço de chamadas | 0.82 | 0.85 |
| Serviço de notificação de saldo | Serviço de dados móveis | 0.67 | 0.96 |
| Serviço de saldo para chamadas urgentes | Serviço de dados móveis | 0.66 | 0.96 |

Suporte (A -> B) = P (A U B)

Confiança (A -> B) = P (B | A)

# Testes

# Testes [1/3]

## Aceitação

- Apresentação dos resultados aos Orientadores da empresa
- Validar se estes cumprem os objetivos de negócio

# Testes [2/3]

## Requisitos funcionais

- Testes unitários - black-box à API e módulos da aplicação
  - Parâmetros inválidos de entrada
  - Acessos não autorizados
  - Validar o resultado com asserções

- Mecanismo de parametrização automático do modelo de previsão
  - Validar se este encontra o modelo com menor erro de previsão
  - Comparar os resultados com as experiências de configuração manual

# Testes [3/3]

## Requisitos não funcionais

- Desempenho
  - Tempo médio de processamento de cada pedido
- Robustez
  - Parâmetros de entrada inválidos realizados juntamente com os testes unitários
  - Falhas de rede e de ligação à base de dados
- Interoperabilidade
  - Ainda não foi possível testar, no entanto suportada:
    - Tecnologia REST
    - Formato de dados JSON

# Conclusão

# Trabalho realizado

- Plataforma base desenhada com foco na interoperabilidade
  - Funcionalidades especificadas
  - Automatização dos modelos
- Empresa
  - Previsão de consumo no BIT
  - Sugestões de público-alvo a uma campanha no ACM - melhorar a sua assertividade, garantir a sua fidelização e poupar recursos no incentivo
- Ponto de vista científico
  - Vantagem de usar alguns métodos em relação a outros, no contexto de telecomunicações
  - Artigo “Experimental Comparison and Tuning of Time Series  Prediction for Telecom Analysis” submetido na International Conference on Time Series and Forecasting (ITISE 2018) a realizar a 19-21 de setembro, Granada, Espanha

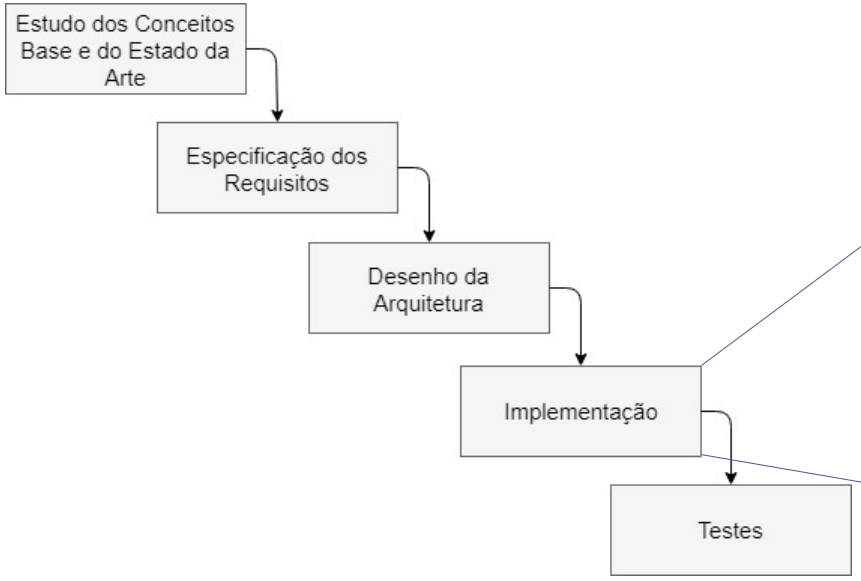
# Trabalho futuro

- Integrar com as ferramentas de visualização de resultados
- Tornar mais eficiente o mecanismo de parametrização do modelo de previsão com recurso a heurísticas
- Requisito de deteção de anomalias de consumo

# Metodologia e planeamento

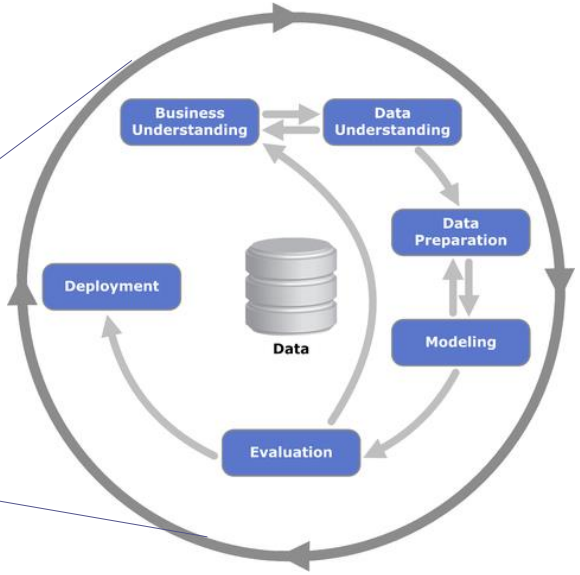
# Metodologia

## Waterfall




- Natureza das tarefas seguirem uma ordem sequencial de execução
- Reduzida probabilidade dos requisitos sofrerem modificações

## Cross-Industry Standard Process for Data Mining (CRISP-DM)




- Desenvolvimento de modelos serem realizados e melhorados iterativamente

# 1º semestre

| Tarefa | Data de início | Data de fim | Estimativa |
|---|---|---|---|
| Gestão do projeto | 18-09-2017 | 19-01-2018 | - |
| Conhecimento base | 18-09-2017 | 13-10-2017 | 20 |
| • Telecomunicações | 18-09-2017 | 22-09-2017 | 5 |
| • Data mining | 25-09-2017 | 06-10-2017 | 10 |
| • Preparação dos dados | 09-10-2017 | 13-10-2017 | 5 |
| Estado da arte | 16-10-2017 | 10-11-2017 | 20 |
| • Sistemas de suporte à decisão concorrentes | 16-10-2017 | 20-10-2017 | 5 |
| • Abordagens analíticas em telecomunicações | 23-10-2017 | 07-11-2017 | 12 |
| • Ferramentas de data mining | 08-11-2017 | 10-11-2017 | 3 |
| Especificação de requisitos | 13-11-2017 | 24-11-2017 | 10 |
| • Requisitos funcionais | 13-11-2017 | 17-11-2017 | 5 |
| • Requisitos não funcionais | 20-11-2017 | 22-11-2017 | 3 |
| • Restrições técnicas e de negócio | 23-11-2017 | 24-11-2017 | 2 |
| Desenho da arquitetura | 27-11-2017 | 15-12-2017 | 15 |
| Familiarização com as tecnologias | 18-12-2017 | 22-12-2017 | 5 |
| Início do desenvolvimento de modelos | 26-12-2017 | 12-01-2018 | 13 |
| Detalhes | 15-01-2018 | 19-01-2018 | 5 |
| Preparação da defesa | 22-01-2018 | 26-01-2018 | 5 |
| | | | |
| Total | 18-09-2017 | 26-01-2018 | 93 |

# 2º semestre

| Tarefa | Data de início | Data de fim | Estimativa |
|---|---|---|---|
| Gestão do projeto | 30-01-2018 | 29-06-2018 | - |
| Revisão do projeto | 30-01-2018 | 02-02-2018 | 4 |
| Estudo de séries temporais | 05-02-2018 | 09-02-2018 | 5 |
| Implementação | 12-02-2018 | 18-05-2018 | 69 |
| • Codificação do módulo | 12-02-2018 | 18-05-2018 | - |
| • Experimentação e desenvolvimento de modelos | 12-02-2018 | 18-05-2018 | - |
| • Integração na plataforma | 12-02-2018 | 18-05-2018 | - |
| Testes | 21-05-2018 | 15-06-2018 | 20 |
| • Requisitos funcionais | 21-05-2018 | 05-06-2018 | 12 |
| • Requisitos não funcionais | 06-06-2018 | 15-06-2018 | 8 |
| Detalhes | 18-06-2018 | 29-06-2018 | 10 |
| Preparação da defesa | 02-07-2018 | 06-07-2018 | 5 |
| | | | |
| Total | 30-01-2018 | 06-07-2018 | 113 |

- Desvios
  - Estudo de séries temporais não antecipado por desconhecimento
  - Desenvolvimento do primeiro requisito funcional devido a inexperiência
  - Não impediram de alcançar os objetivos propostos

# Questões