

ISTA - Escola de Tecnologias e Arquitetura  
Departamento de Ciências e Tecnologias de Informação

# Previsão de consumo de Energia Elétrica em contexto de *Smart Grids*

Dissertação apresentada como requisito para obtenção do grau de  
**Mestre em Engenharia Informática**

Carlos Manuel Nunes Lima

## **Orientadora**

Professora Doutora Elsa Alexandra Cabral da Rocha Cardoso, Professor Auxiliar

ISCTE – Instituto Universitário de Lisboa,

Escola de Tecnologias e Arquitetura

Departamento de Ciências e Tecnologias de Informação

## **Co-Orientador**

Professor Doutor Pedro Joaquim Amaro Sebastião, Professor Auxiliar

ISCTE – Instituto Universitário de Lisboa,

Escola de Tecnologias e Arquitetura

Departamento de Ciências e Tecnologias de Informação

Outubro 2015

*“Our greatest weakness lies in giving up. The most certain way to succeed is always to try just one more time.”*

*Thomas A. Edison*

# Resumo

---

## Português

Num contexto de modernização e acelerados avanços tecnológicos no sector das *Energy & Utilities*, o conceito das *Smart Grids* surge associado às novas capacidades de gestão de redes energéticas inteligentes, através da integração de mecanismos automatizados de comunicação que possibilitam alcançar uma visão global do estado da rede energética e da totalidade dos seus componentes. As *Utilities* passam assim a ter disponíveis volumes massivos de informação, passíveis de análise e capazes de suportar a estratégia de otimização de todo o ciclo de produção, distribuição e comercialização de energia.

Dessa forma, e para responder aos objetivos deste trabalho, foram utilizados dados provenientes dos contadores inteligentes, colocados em cada local de consumo, de forma a perceber perfis de consumo de clientes e criar modelos de previsão de consumo de energia no curto prazo. Os perfis de consumo dos clientes foram estudados e através de uma análise de *Clustering* foi possível identificar clientes com padrões de consumo semelhantes ao longo do tempo e agrupa-los em grupos de semelhança. O consumo de cada um dos grupos identificados foi seguidamente agregado e foram utilizadas redes neuronais para prever o seu consumo para um horizonte temporal de curto prazo (próximas 24 horas).

Em termos de resultados obtidos foi possível perceber que a metodologia utilizada conseguiu alcançar resultados muito positivos quando comparados os valores reais de consumo com os valores que o algoritmo de previsão teve a capacidade de prever, tendo taxas de acerto superiores a 90%.

As principais contribuições deste trabalho prendem-se com o desenho que uma metodologia de previsão utilizando várias técnicas de aprendizagem automática bem como um contributo ligado ao meio profissional, na medida em que foi ganho conhecimento que será útil no desenvolvimento da oferta da Novabase na área de *Energy & Utilities*.

**Palavras-chave:** Consumo Energia, Análise Preditiva, K-Means, Clustering, Aprendizagem Automática, ANN

# Abstract

---

**English**

In a context of modernization and accelerated technological advances in the Energy & Utilities sector, the concept of Smart Grids appear associated with new intelligent energy network management capabilities through the integration of automated communication mechanisms that enable achieving an overview of network status and of all energy network components. The Utilities will have available massive amounts of information to analysis and to give support the optimization strategy of the entire production cycle, distribution and commercialization of energy.

Thus, and to achieve the objectives of this research were used data from smart meters, placed at each place of consumption, in order to understand customer consumption profiles and create power consumption forecasting models in the short term. Customer consumption profiles were studied and by a clustering analysis was possible to identify customers with similar consumption patterns over time and bundles them into similarity groups. The consumption of each of the identified groups was then added and neural networks are used to predict the consumption of each group for a short-term time horizon (next 24 hours).

In terms of results it was observed that the methodology has achieved very positive results when comparing the actual consumption values with the values that the prediction algorithm has the ability to predict, with accuracy rates of over 90 %.

The main contributions of this work are related to the design a prediction methodology using various machine learning techniques and a contribution linked to the professional world, allowing gain knowledge that will be useful in the Development of Novabase Area Energy & Utilities

**Keywords:** Energy Consumption, Predictive Analytics, K-Means, Clustering, Machine Learning, ANN

# Agradecimentos

---

## Obrigado

Um sincero agradecimento aos meus pais, Alcino e Maria, e irmã Daniela pela preocupação, apoio e energia transmitida nesta fase tão relevante.

Queria deixar uma palavra especial também para os meus orientadores Professora Elsa Cardoso e Professor Pedro Sebastião, pela motivação e interesse e acompanhamento ao longo deste trabalho.

Queria ainda agradecer à Novabase, especialmente ao Pedro Abreu e ao Hugo Matos, pela disponibilidade, apoio e suporte no decorrer desde trabalho.

Deixo também uma palavra de apreço a todos os meus amigos que me foram apoiando ao longo desta jornada difícil e em especial, ao João Lima e ao Miguel Duarte.

Por último, mas claramente não menos importante, queria agradecer à minha namorada, por estar sempre presente e me dar a força necessária a cada dia.



# Índice

## Conteúdo

|   |           |
|---|-----------|
| <b>1. INTRODUÇÃO</b> .....                              | <b>1</b>  |
| 1.1. MOTIVAÇÃO .....                                    | 3         |
| 1.2. OBJETIVOS.....                                     | 4         |
| 1.3. CONTRIBUIÇÕES.....                                 | 5         |
| 1.4. ESTRUTURA DO DOCUMENTO.....                        | 6         |
| <b>2. ESTADO DA ARTE</b> .....                          | <b>7</b>  |
| 2.1. <i>SMART GRID</i> .....                            | 7         |
| 2.2. PREVISÃO DE CONSUMOS DE ENERGIA .....              | 11        |
| 2.2.1. <i>Short-Term</i> .....                          | 12        |
| 2.2.2. <i>Medium-Term</i> .....                         | 12        |
| 2.2.3. <i>Long-Term</i> .....                           | 12        |
| 2.3. IDENTIFICAÇÃO DE FATORES COM IMPACTO NA CARGA..... | 13        |
| 2.4. TÉCNICAS PARA PREVISÃO DE CONSUMO.....             | 14        |
| 2.5. PREVISÃO BASEADA EM <i>MACHINE LEARNING</i> .....  | 16        |
| 2.5.1. <i>K-Means</i> .....                             | 17        |
| 2.5.2. <i>Redes Neurais</i> .....                       | 17        |
| 2.6. METODOLOGIA DE DESENVOLVIMENTO .....               | 21        |
| <b>3. CASO DE ESTUDO</b> .....                          | <b>24</b> |
| 3.1. CONTEXTO.....                                      | 24        |
| 3.2. MODELAÇÃO CONCEPTUAL.....                          | 25        |
| 3.3. CARACTERIZAÇÃO DA POPULAÇÃO.....                   | 26        |
| 3.4. ANÁLISE DE CONSUMO DE ENERGIA.....                 | 27        |
| <b>4. DESENVOLVIMENTO</b> .....                         | <b>34</b> |
| 4.1. PREVISÃO DE CONSUMOS DE ENERGIA .....              | 34        |
| 4.2. <i>CLUSTERING</i> .....                            | 35        |
| 4.3. MODELAÇÃO DAS REDES NEURONAIS .....                | 37        |
| 4.3.1. <i>Variáveis de Entrada e Saída</i> .....        | 37        |
| 4.3.2. <i>Conjunto de Treino e Validação</i> .....      | 39        |
| 4.3.3. <i>Erros de previsão</i> .....                   | 40        |
| <b>5. ANÁLISE DE RESULTADOS</b> .....                   | <b>41</b> |
| 5.1. <i>CLUSTERING</i> .....                            | 41        |

|  |           |
|--|-----------|
| 5.1.1. Centros e associações de clientes a clusters .....            | 42        |
| 5.1.2. Análise de Clusters .....                                     | 44        |
| 5.1.3. Clientes Residenciais .....                                   | 44        |
| 5.1.4. Clientes Empresariais .....                                   | 46        |
| 5.2. TREINO DA REDE NEURONAL.....                                    | 47        |
| 5.3. RESULTADOS DA PREVISÃO .....                                    | 50        |
| 5.3.1. Clientes Residenciais .....                                   | 50        |
| 5.3.1. Clientes Empresariais .....                                   | 54        |
| 5.3.2. Análise de Grandes Desvios .....                              | 57        |
| <b>6. CONCLUSÕES E TRABALHO FUTURO .....</b>                         | <b>62</b> |
| 6.1. CONCLUSÕES .....  | 62        |
| 6.2. TRABALHO FUTURO.....  | 64        |
| <b>7. REFERÊNCIAS .....</b>  | <b>65</b> |
| <b>8. ANEXOS .....</b>   | <b>71</b> |
| 8.1. ACRÓNIMOS .....   | 71        |
| 8.2. TECNOLOGIA RELEVANTE.....                                       | 71        |
| 8.3. VARIÁVEIS ASSOCIADAS A INQUÉRITOS DE CLIENTES RESIDENCIAIS..... | 72        |
| 8.4. VARIÁVEIS DE INQUÉRITOS ASSOCIADAS A CLIENTES EMPRESARIAIS..... | 73        |
| 8.5. DISTRIBUIÇÃO DA POPULAÇÃO PELAS VARIÁVEIS DOS INQUÉRITOS.....   | 74        |



# Índice

## Lista de Figuras

|   |    |
|---|----|
| FIGURA 1 - INFRAESTRUTURA DA <i>SMART GRID</i> . ADAPTADO DE [4].....   | 7  |
| FIGURA 2: REDE ELÉTRICA ATUAL [1].....  | 8  |
| FIGURA 3 - <i>MULTILAYER FEED-FORWARD ANN</i> .....   | 18 |
| FIGURA 4 – FUNCIONAMENTO DE UM NEURÓNIO ARTIFICIAL.....   | 19 |
| FIGURA 5 – FUNÇÕES DE ATIVAÇÃO DOS NEURÓNIOS .....  | 19 |
| FIGURA 6 – FRAMEWORK CRISP-DM.....  | 22 |
| FIGURA 7 – MODELO CONCEPTUAL DE CARACTERIZAÇÃO DE CASO DE ESTUDO.....   | 25 |
| FIGURA 8 – DISTRIBUIÇÃO DE CLIENTES POR TIPO.....   | 26 |
| FIGURA 9 – RELAÇÃO ENTRE O CONSUMO TOTAL AGREGADO POR MÊS E TEMPERATURA MÉDIA .....                               | 27 |
| FIGURA 10 – RELAÇÃO ENTRE PERFIL DA TEMPERATURA MÉDIA E PERFIL DE CONSUMO POR HORA.....                           | 28 |
| FIGURA 11 – CONSUMO MÉDIO POR HORA AO LONGO DE CADA MÊS PARA O ANO 2010.....                                      | 29 |
| FIGURA 12 – CONSUMO MÉDIO POR PARA CLIENTES RESIDENCIAIS PARA DIAS DA SEMANA.....                                 | 30 |
| FIGURA 13 – CONSUMO MÉDIO POR HORA PARA CLIENTES RESIDENCIAIS PARA OS FINS-DE-SEMANA.....                         | 30 |
| FIGURA 14 – CONSUMO MÉDIO PARA CLIENTES RESIDENCIAIS PARA CADA ESTAÇÃO DO ANO .....                               | 31 |
| FIGURA 15 – CONSUMO MÉDIO POR HORA PARA CLIENTES EMPRESARIAIS PARA DIAS DA SEMANA.....                            | 32 |
| FIGURA 16 - CONSUMO MÉDIO POR HORA PARA CLIENTES EMPRESARIAIS PARA OS FINS-DE-SEMANA .....                        | 32 |
| FIGURA 17 – CONSUMO MÉDIO PARA CLIENTES EMPRESARIAIS PARA CADA ESTAÇÃO DO ANO.....                                | 33 |
| FIGURA 18 – METODOLOGIA DO TRABALHO DESENVOLVIDO .....  | 35 |
| FIGURA 19 – NÚMERO DE <i>CLUSTERS</i> CLIENTES RESIDENCIAIS .....   | 41 |
| FIGURA 20 – NÚMERO DE <i>CLUSTERS</i> CLIENTES EMPRESARIAIS.....  | 41 |
| FIGURA 21 – CENTROS DOS <i>CLUSTERS</i> PARA CLIENTES RESIDENCIAIS .....  | 42 |
| FIGURA 22 - CENTROS DOS <i>CLUSTERS</i> PARA CLIENTES EMPRESARIAIS.....   | 43 |
| FIGURA 23 – ANÁLISE DO NÚMERO DE CLIENTE POR TIPO DE HABITAÇÃO E <i>CLUSTER</i> .....                             | 44 |
| FIGURA 24 – ANÁLISE DO NÚMERO DE CLIENTES POR RENDIMENTO E <i>CLUSTER</i> .....                                   | 45 |
| FIGURA 25 - ANÁLISE DO NÚMERO DE CLIENTES POR ESTADO PROFISSIONAL E <i>CLUSTER</i> .....                          | 45 |
| FIGURA 26 - ANÁLISE DO NÚMERO DE CLIENTES DISTRIBUÍDOS POR CLASSES DE NÚMERO DE EMPREGADOS E <i>CLUSTER</i> ..... | 46 |
| FIGURA 27 - ANÁLISE DO NÚMERO DE CLIENTES POR VOLUME DE NEGÓCIOS E <i>CLUSTER</i> .....                           | 46 |
| FIGURA 28 – ANÁLISE DE VARIÁVEIS DE TREINO COM BASE NO MAPE PARA CLIENTES RESIDENCIAIS <i>CLUSTER #1</i> .....    | 47 |
| FIGURA 29 - ANÁLISE DE VARIÁVEIS DE TREINO COM BASE NO MAPE PARA CLIENTES RESIDENCIAIS <i>CLUSTER #2</i> .....    | 48 |
| FIGURA 30 - ANÁLISE DE VARIÁVEIS DE TREINO COM BASE NO MAPE PARA CLIENTES EMPRESARIAIS <i>CLUSTER #1</i> .....    | 48 |
| FIGURA 31 - ANÁLISE DE VARIÁVEIS DE TREINO COM BASE NO MAPE PARA CLIENTES EMPRESARIAIS <i>CLUSTER #2</i> .....    | 49 |
| FIGURA 32 - ERRO MÉDIO DE PREVISÃO PARA <i>CLUSTER #1</i> DE CLIENTES RESIDENCIAIS.....                           | 50 |

|   |    |
|---|----|
| FIGURA 33 – CLIENTES RESIDENCIAIS CLUSTER # 1 DIA 15-07-2010.....   | 51 |
| FIGURA 34 – CLIENTES RESIDENCIAIS CLUSTER #1 DIA 21-12-2010 .....   | 51 |
| FIGURA 35 - ERRO MÉDIO DE PREVISÃO PARA CLUSTER #2 DE CLIENTES RESIDENCIAIS.....                              | 52 |
| FIGURA 36 – CLIENTES RESIDENCIAIS CLUSTER #2 DIA 16/10/2010 .....   | 52 |
| FIGURA 37 – CLIENTES RESIDENCIAIS CLUSTER #2 DIA 16/12/2010 .....   | 53 |
| FIGURA 38 - ERRO MÉDIO DE PREVISÃO PARA CLUSTER #1 DE CLIENTES EMPRESARIAIS .....                             | 54 |
| FIGURA 39 – CLIENTES EMPRESARIAIS CLUSTER #1 DIA 15/07/2010 .....   | 55 |
| FIGURA 40 – CLIENTES EMPRESARIAIS CLUSTER #1 DIA 13/11/2010 .....   | 55 |
| FIGURA 41 - ERRO MÉDIO DE PREVISÃO PARA CLUSTER #2 DE CLIENTES EMPRESARIAIS .....                             | 56 |
| FIGURA 42 – CLIENTES EMPRESARIAIS CLUSTER #2 DIA 27/08/2010.....  | 56 |
| FIGURA 43 – CLIENTES EMPRESARIAIS CLUSTER #2 DIA 14/12/2010.....  | 57 |
| FIGURA 44 – CONSUMO REAL VS PREVISÃO PARA CLIENTES RESIDENCIAIS, CLUSTER #1 DIA 25-12-2010 .....              | 58 |
| FIGURA 45 – RELAÇÃO ENTRE DIA COM ERRO DE PREVISÃO E DIAS ANTERIORES – CLIENTES RESIDENCIAIS, CLUSTER #1..... | 59 |
| FIGURA 46 - CONSUMO REAL VS PREVISÃO PARA CLIENTES EMPRESARIAIS, CLUSTER #1 DIA 25-10-2010 .....              | 60 |
| FIGURA 47 - RELAÇÃO ENTRE DIA COM ERRO DE PREVISÃO E DIAS ANTERIORES – CLIENTES EMPRESARIAIS, CLUSTER #1..... | 60 |
| FIGURA 48 – NÚMERO DE CLIENTES POR ESTADO PROFISSIONAL.....   | 74 |
| FIGURA 49 – NÚMERO DE CLIENTES POR TARIFÁRIO .....  | 74 |
| FIGURA 50 – NÚMERO DE CLIENTES POR TIPO DE HABITAÇÃO .....  | 74 |
| FIGURA 51 – NÚMERO DE CLIENTES POR ESTADO DE AQUISIÇÃO DE HABITAÇÃO .....                                     | 75 |
| FIGURA 52 – NÚMERO DE CLIENTES POR NÍVEL DE FORMAÇÃO.....   | 75 |
| FIGURA 53 – NÚMERO DE CLIENTES POR CLASSE DE RENDIMENTO.....  | 75 |
| FIGURA 54 – NÚMERO DE CLIENTES POR TIPO DE EMPRESA .....  | 76 |
| FIGURA 55 – NÚMERO DE CLIENTES POR CLASSE DE NÚMERO DE FUNCIONÁRIOS .....                                     | 76 |
| FIGURA 56 – NÚMERO DE CLIENTES POR CLASSE DE RECEITA.....   | 77 |

# Índice

---

## Lista de Tabelas

|   |    |
|---|----|
| TABELA 1 - COMPARAÇÃO DAS CARACTERÍSTICAS DA REDE EXISTENTE COM A <i>SMART GRID</i> [1] ..... | 9  |
| TABELA 2 BENEFÍCIOS DOS <i>SMART METERS</i> . ADAPTADO DE [4].....                            | 10 |
| TABELA 3 - EXEMPLOS DE ALGORITMOS DE PREVISÃO .....   | 15 |
| TABELA 4 – VARIÁVEIS QUE COMPÕEM A REDE NEURONAL .....  | 37 |
| TABELA 5 – PARÂMETROS OBTIDOS NO PROCESSO DE TREINO .....                                     | 49 |
| TABELA 6 – ANÁLISE DE GRANDES DESVIOS (CONSUMO REAL VS CONSUMO PREVISTO).....                 | 57 |
| TABELA 7 - VARIÁVEIS ASSOCIADAS A INQUÉRITOS DE CLIENTES RESIDENCIAIS.....                    | 72 |
| TABELA 8 - 8.4. VARIÁVEIS DE INQUÉRITOS ASSOCIADAS A CLIENTES EMPRESARIAIS .....              | 73 |

# Capítulo 1

---

## 1. Introdução

A energia elétrica é um bem indispensável à existência da atual sociedade. Nas nossas casas utilizamos energia para iluminação, aquecimento, refrigeração, conservação de alimentos e equipamentos de lazer. No entanto, a maior parte dos consumidores não se preocupa muito com a eletricidade até esta falhar ou até a fatura ter um valor elevado.

O mercado da energia a nível global tem sofrido alterações significativas nas últimas décadas devido à liberalização, ao crescimento da competição entre as várias *Utilities* (empresas que operam no negócio da energia), aos avanços tecnológicos e às regulamentações governamentais no que diz respeito à eficiência. Num contexto de modernização e acelerados avanços tecnológicos neste sector, o conceito das *Smart Grids* [1] surge associado às novas capacidades de gestão de redes energéticas inteligentes, através da integração de mecanismos automatizados de comunicação que possibilitam alcançar uma visão global do estado da rede energética e da totalidade dos seus componentes. As políticas de sustentabilidade e os avanços tecnológicos vêm potenciar o aparecimento deste novo conceito de rede de energia. Com a adoção das *Smart Grids* todo o modelo de negócio e a relação com todos os *stakeholders* é alterado e a rede fica capacitada de inteligência quando utilizados os dados provenientes das comunicações com equipamentos. Estas alterações são possíveis através da incorporação de filosofias, conceitos e tecnologias que permitem ligar a rede de energia às redes de informação permitindo um fluxo bidirecional de informação originando assim grandes quantidades de dados que serão passíveis de ser tratados [2].

As *Utilities* passam assim a ter disponíveis volumes massivos de informação, capazes de suportar a estratégia de otimização de todo o ciclo de produção, distribuição e comercialização de energia. Ao contrário de falarmos de tecnologia, de um produto ou de um serviço, o conceito de *Smart Grids* é, na sua essência, uma visão do futuro da infraestrutura da distribuição de energia. Através da introdução destes conceitos nas redes de energia pretende-se dotar a rede de inteligência utilizando capacidades sensoriais ao longo da rede. Estes sensores permitem às *Utilities* conhecer e gerir de forma mais adequada a rede. Através desse conhecimento é possível o redireccionamento de energia

em caso de sobrecarga, diminuindo potenciais falhas e ainda um controlo eficiente do aumento de necessidades de energia sem alterar infraestrutura. Para além disso as *Smart Grids* visam capacitar a rede de mecanismos de comunicação em tempo real *end-to-end* desde o consumidor até à *Utility* criando com isso novas oportunidades de desenvolvimento de aplicações, melhorando a qualidade de energia evitando sobre e subtensões e também saber com precisão quais os consumos de cada cliente em curtos intervalos de tempo, que permitem a criação de serviços mais direcionados e personalizados de acordo com o perfil de cada cliente.

Quando se fala em *Smart Grids*, surge associado um dos equipamentos potenciadores de gerar informação, os *Smart Meters* ou contadores inteligentes. Estes equipamentos são os responsáveis pela comunicação entre os clientes e as empresas. Definem-se como aparelhos elétricos que medem o consumo de energia em pequenos intervalos de tempo, armazenando informação e comunicando de forma direta com a empresa. Com a adoção destes equipamentos uma quantidade significativa de novos dados passa a estar disponível. Os *Smart Meters* permitem recolher e fornecer para a rede grandes volumes de dados relacionados com as suas operações e consumos. A introdução destes equipamentos em casa dos consumidores irá permitir às *Utilities* tirar partido destes dados e fazer uma gestão mais eficiente da rede. A capacidade de recolher leituras de consumo de energia de cada cliente, por exemplo, permite às empresas perceber os hábitos de consumo dos seus clientes contribuindo para uma melhor gestão das necessidades de energia. Para além de consumos, os *Smart Meters* permitem obter mais informações acerca da qualidade da energia e do funcionamento da rede que permite diagnosticar e antecipar eventuais problemas.

Um dos objetivos principais das *Utilities* é gerir de forma eficiente a oferta e procura de energia. Através dos contadores tradicionais apenas é possível medir o consumo total de energia desconhecendo por completo em que período a energia foi consumida, ou apenas em períodos definidos como horas de ponta e horas de vazio, para o caso dos bi-horários. A introdução de *Smart Meters* fornece uma forma de medir especificamente a quantidade da energia consumida num dado período de tempo [3] obtendo assim um conhecimento mais preciso do consumo global do sistema ao longo do tempo. É de grande importância para as empresas que atuam no setor de geração, transmissão e distribuição de energia elétrica conhecer as necessidades futuras de energia. Esse conhecimento é útil em várias das atividades diárias efetuadas por estas empresas, especialmente se permitirem otimizar

o escalonamento da produção de energia elétrica e a distribuição de energia. Além disso, permite também agendar tarefas de manutenção que não perturbem o fornecimento normal da rede.

### **1.1. Motivação**

A tomada de decisões é um facto quotidiano que desempenha um papel relevante dentro das empresas. Atualmente, o alto grau de competitividade no meio empresarial exige a capacidade de tomar decisões rápidas e precisas. A qualidade da tomada de decisão tem relação direta com os dados disponíveis para o agente responsável pelas decisões e com a sua habilidade em extrair destes dados informações relevantes. A temática do *Predictive Analytics* potencia a tomada de decisão antecipando os acontecimentos, através, por exemplo, da análise de padrões em dados históricos que ajudam a antecipar problemas ganhando assim vantagens competitivas. O facto de ser possível prever acontecimentos futuros é diferenciador na medida que permitirá atuar em antecipação relativamente a determinados acontecimentos inesperados, tais como, falhas na rede em função de eventos ocorridos, prever picos de consumos de energia para determinadas zonas agilizando os mecanismos necessários para dar resposta às necessidades, ou à própria gestão sustentável e eficiente da geração e transmissão de energia em função das necessidades previstas. Através das técnicas de previsão é possível extrair informações dos dados recolhidos no passado e que permitirão a modelação analítica do comportamento da rede e dos seus agentes. Os consumos dos clientes são um exemplo de acontecimentos que podem ser modelados. Tendo em conta a suposição de uma continuidade nesse comportamento, e utilizando dados com uma granularidade mais fina do que até agora é feito, torna-se possível a realização de previsões cuja qualidade e precisão serão superiores àquelas que são feitas até agora baseando-se apenas em consumos mais agregados.

## 1.2. Objetivos

Os Modelos Preditivos assumem uma importância vital para a indústria das *Utilities* podendo ser utilizados para muitas aplicações. De entre as várias aplicações o equilíbrio entre a procura e a produção de energia assume cada vez mais um papel importante alinhado com o objetivo global da sustentabilidade económica e ambiental. A realização do presente trabalho tem como principal objetivo o desenvolvimento e teste de uma metodologia que permite criar um modelo de previsão de consumo de energia para um horizonte de curto prazo de forma a tornar possível uma melhor gestão das operações no contexto das redes de energia elétrica. Os consumos de energia, tipicamente, apresentam comportamentos que podem ser explicados por diversas variáveis, como por exemplo dimensões temporais, como a hora do dia em que se consome energia ou a própria estação do ano, e características meteorológicas. O modelo de previsão basear-se-á na utilização de dados históricos de consumos associados a variáveis que possam influenciar o consumo de forma a criar modelos mais precisos e adequados.

Em suma os objetivos do presente trabalho são os seguintes:

- Perceber as variáveis que influenciam o consumo;
- Utilizar algoritmos de segmentação de forma a organizar o consumo de energia em função de vários padrões típicos de consumo;
- Prever consumos de energia para diferentes grupos de clientes para um horizonte temporal de curto prazo (neste caso próximas 24h);
- Avaliar o output da metodologia utilizada de forma a perceber a capacidade desta responder à previsão do consumo de energia.

### 1.3. Contribuições

A principal contribuição a nível académico desta dissertação, está relacionada com o estudo sistemático, utilização e aplicação de métodos e algoritmos de análise preditiva que têm vindo a ganhar cada vez mais o seu espaço tanto no meio académico como no meio profissional. Em termos de inovação, este trabalho não aborda temas totalmente inovadores pois a previsão de consumos de energia já é feita há algum tempo, mas por outro lado, utiliza estes métodos de previsão em conjunto com dados provenientes de equipamentos que até agora não tinham a capacidade de gerar a quantidade de informação agora disponível, indo de encontro a toda a inovação que está a ocorrer nas redes de energia elétrica.

Outra das contribuições passa pelo contributo que este trabalho pretende fornecer ao meio profissional. Alinhado com os desafios da Novabase, que surge aqui como a entidade onde me encontro a desenvolver o meu caminho profissional, mais especificamente no negócio específico da energia, este trabalho pretende contribuir para o desenvolvimento da oferta na nessa mesma área procurando captar conhecimento na temática do *Predictive Analytics* de forma a conseguir perceber e responder às necessidades do mercado. Através deste trabalho é possível enriquecer o conhecimento dentro do meio profissional procurando criar valor útil para atuar na área do *Predictive Analytics* em vários contextos e geografias potenciando novas estratégias e soluções de negócio.



## 1.4. Estrutura do Documento

O presente trabalho apresenta uma estrutura de capítulos definida da seguinte forma:

O Capítulo 2 começa por apresentar o estado da arte relativamente aos novos conceitos de gestão da rede elétrica, a *Smart Grid*. Em seguida são apresentadas as diferentes técnicas utilizadas para a precisão de consumos de energia e as variáveis consideradas relevantes em diferentes estudos.

O Capítulo 3 apresenta as características principais do conjunto de dados de *Smart Meters* utilizados. Neste são apresentados os resultados do processo de exploração de dados efetuado com vista a perceber as dinâmicas de consumo dentro do universo de clientes e quais as variáveis que têm maior influência e podem explicar o consumo de energia.

No Capítulo 4 são apresentados os desenvolvimentos dos modelos e dos algoritmos utilizados para previsão de consumos de energia justificando a sua utilização.

O Capítulo 5 apresenta uma análise dos resultados obtidos utilizando métricas que permitam avaliar a capacidade e a precisão da previsão dos consumos.

No Capítulo 6 são apresentadas as conclusões dando ênfase aos objetivos que foram alcançados e apresentadas algumas possibilidades para trabalho futuro.

# Capítulo 2

## 2. Estado da Arte

### 2.1. *Smart Grid*

A rede elétrica está organizada hierarquicamente, onde as centrais elétricas, estas no topo da hierarquia, produzem energia que irá ser transportada e distribuída até aos clientes finais (Figura 2: Rede Elétrica Atual ). O sistema tradicional é estritamente unidirecional, em termos de comunicações, onde as informações não circulam, de ponto a ponto da rede, nem é possível captar conhecimento em tempo real e de forma direta acerca do que se passa nos pontos de destino da energia. Esta rede está sobredimensionada para resistir a eventuais picos de necessidade de energia mas se estes picos são pouco frequentes o sistema torna-se ineficiente [1].

O conceito de *Smart Grid* [1] define-se como um conjunto de tecnologias, conceitos e abordagens que permitem substituir os silos hierárquicos da geração, transmissão e distribuição de energia por um sistema integrado e inteligente onde os processos de negócio, objetivos e necessidades de todos os *stakeholders* são suportadas pela partilha de dados.

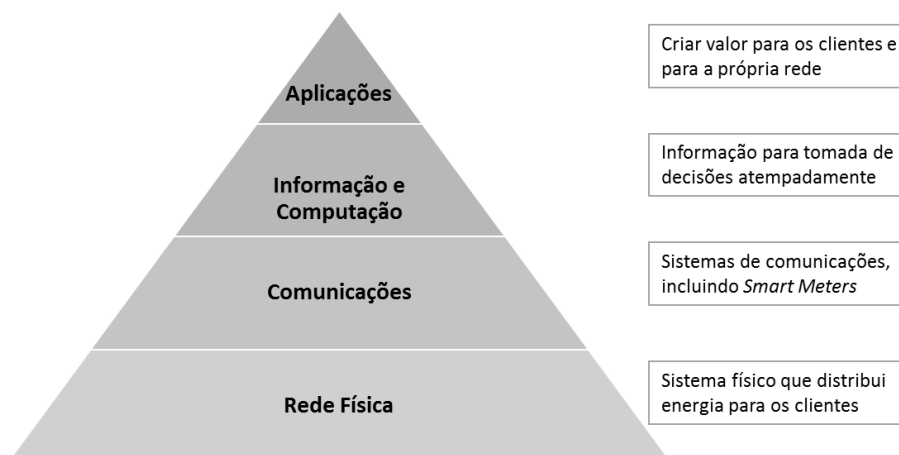
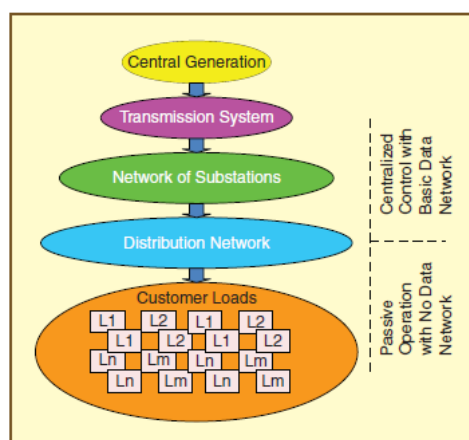


Figura 1 - Infraestrutura da *Smart Grid*. Adaptado de [4]

Esta nova visão das redes de energia permite fornecer às *Utilities* uma visão global e controlo acerca de todos os ativos de rede e serviços. Estas empresas atravessam grandes desafios, como por exemplo, a diversificação da geração de energia, a capacidade de dar resposta às necessidades crescentes de consumo e a necessidade de redução da pegada de carbono [1]. Endereçar estes desafios torna-se difícil com a atual infraestrutura da rede elétrica dado a sua natureza unidirecional que dificulta a fiel perceção do que acontece na rede. Para permitir um melhor controlo e monitorização, as *Smart Grids* introduzem na rede a capacidade de comunicação entre equipamentos, gerando um conjunto de novos dados que podem ser utilizados pelos sistemas de informação. Com o aparecimento das *Smart Grids* aparece também uma nova arquitetura denominada por *Advanced metering infrastructure* (AMI) [1]. Esta arquitetura define-se como uma infraestrutura automatizada baseada em comunicações. Estas comunicações são bidirecionais entre os *Smart Meters*, que estão ligados à rede através de um endereço IP, e a *Utility*. A arquitetura AMI não se resume a tecnologia mas sim a uma integração de várias tecnologias (e.g. redes de comunicações, aplicações de gestão de dados) em que o objetivo passa por fornecer às *Utilities* informações sobre a rede, como por exemplo, consumos de energia em intervalos de tempo de alta resolução, estado de funcionamento dos equipamentos, eventos ocorridos na rede, entre outros. Desta forma, as *Utilities* ganham a capacidade de obter instantaneamente informações acerca da rede e dos seus clientes [5]. Este conhecimento permite às *Utilities* controlarem o estado de funcionamento dos mais variados sensores e equipamentos na rede, bem como as necessidades de consumo de energia para o futuro.



**Figura 2: Rede Elétrica Atual [1]**

Tal como podemos observar na Tabela 1 - Comparação das características da rede existente com a *Smart Grid*, as *Smart Grids* apresentam um conjunto de novas capacidades fazendo com que as comunicações e a gestão da informação assumam um papel vital na gestão da rede elétrica. Para além da introdução de comunicações, outra característica importante a destacar passa pela introdução de sensores ao longo da rede que permitem a geração de dados em tempo real, de forma a melhorar a gestão e monitorização da rede de forma distribuída. Uma função crítica para o negócio, como a gestão das necessidades de consumo, poderá ser realizada de forma mais precisa através da introdução destes conceitos, funcionalidades e equipamentos nomeadamente os *Smart Meters*.

**Tabela 1 - Comparação das características da rede existente com a *Smart Grid* [1]**

|                                      | Grid existente                    | Smart Grid                        |
|--------------------------------------|-----------------------------------|-----------------------------------|
| <b>Tipologia</b>                     | Electromecânica                   | Digital                           |
| <b>Comunicações</b>                  | Comunicação Unidirecional         | Comunicação Bidirecional          |
| <b>Capacidade de Geração</b>         | Geração Centralizada              | Geração Distribuída               |
| <b>Organização</b>                   | Hierárquica                       | Rede                              |
| <b>Sensores</b>                      | Poucos Sensores                   | Sensores Pervasivos               |
| <b>Gestão da Rede</b>                | Cega                              | Monitorizada                      |
| <b>Restauro</b>                      | Restauro Manual                   | Restauro Automático               |
| <b>Resiliência</b>                   | Falhas e Cortes                   | Adaptativa                        |
| <b>Testes</b>                        | Testes e Verificações Manuais     | Testes e Verificações Automáticos |
| <b>Controlo</b>                      | Controlo Limitado                 | Controlo Pervasivo                |
| <b>Facilidades para Consumidores</b> | Poucas Escolhas para Consumidores | Muitas escolhas para consumidores |

Os *Smart Meters* são, à imagem dos contadores tradicionais de energia, dispositivos eletrónicos usados pelas *Utilities* para medir os consumos de energia. A diferença fundamental entre os *Smart Meters* e os contadores tradicionais é a sua capacidade de comunicação de forma bidirecional, que permite para além de tratar de valores de consumo, obter mais informações acerca do estado do equipamento e da rede quase em tempo real e de forma mais granular. Este conjunto de novos dados apresenta vários benefícios não só para as *Utilities* mas também para os próprios clientes (Tabela 2 Benefícios dos *Smart Meters*. Adaptado de ).

**Tabela 2 Benefícios dos *Smart Meters*. Adaptado de [4]**

| <i>Stakeholders</i>   | Benefícios  |
|---|---|
| <b>Clientes</b>   | <ul style="list-style-type: none"> <li>• Fácil acesso a dados para gestão da energia</li> <li>• Maior precisão na faturação</li> <li>• Ajustes e melhorias ao nível das tarifas</li> <li>• Melhoramento na restauração em caso de falha</li> <li>• Qualidade de dados da energia</li> </ul>                               |
| <b>Atendimento ao Cliente e Operações no Terreno</b>                                | <ul style="list-style-type: none"> <li>• Redução de custo nas leituras</li> <li>• Redução das viagens fora do ciclo de leitura</li> <li>• Eliminação dos equipamentos de leitura manual</li> <li>• Redução das operações via <i>call center</i></li> <li>• Redução na operação do corte e religação da energia</li> </ul> |
| <b>Serviços de Ciclo de Receita - Faturação, Contabilidade, Proteção da Receita</b> | <ul style="list-style-type: none"> <li>• Redução dos ajustes face à previsão de faturação</li> <li>• A deteção antecipadamente da adulteração do medidor ou roubo</li> <li>• Redução dos erros de estimação na faturação</li> </ul>   |
| <b>Transmissão e distribuição</b>   | <ul style="list-style-type: none"> <li>• Gestão mais eficiente nos transformadores de energia</li> <li>• Melhorar a eficiência, fiabilidade do serviço e perdas</li> <li>• Dados que permitam melhorar o desenho da rede de energia</li> <li>• Dados de qualidade de energia para as áreas de serviço</li> </ul>          |
| <b>Marketing &amp; Previsão de Carga</b>  | <ul style="list-style-type: none"> <li>• Redução de custos de recolha de dados de consumos</li> </ul>   |
| <b>Utility em geral</b>   | <ul style="list-style-type: none"> <li>• Redução de queixas</li> <li>• Melhoria da segurança nas instalações do cliente e perfil de risco</li> <li>• Redução de incidentes de segurança com funcionários</li> </ul>   |
| <b>Stakeholders externos</b>  | <ul style="list-style-type: none"> <li>• Benefícios ambientais</li> <li>• Apoio às iniciativas de <i>Smart Grid</i></li> </ul>  |

## 2.2. Previsão de Consumos de Energia

No contexto das redes de energia, funcionalidades como o *load forecasting* (a previsão do consumo de energia, previsão de cargas) assumem uma importância vital para a gestão das redes de transporte e distribuição de eletricidade [6]. O termo *forecasting* surge associado à capacidade de prever um determinado acontecimento ou variável no futuro, normalmente através do estudo do comportamento temporal da variável observando, acontecimentos passados que possam influenciar essa variável. O seu interesse está associado à gestão e planeamento da rede mas também à previsão de necessidades de energia que permitam gerir melhor a própria produção. A previsão de cargas advém de diversas necessidades que o elevado e rápido crescimento das redes de transporte e distribuição têm vindo a seguir [7] [8].

A definição do termo *load forecasting* pode ser vista como a forma de estimar a quantidade de energia que será necessária para um determinado horizonte temporal definido. Esta previsão não é simples e direta dado que não é possível medir e prever a energia elétrica através de um único parâmetro. Para que seja possível prever o consumo de energia será necessário definir o horizonte temporal para o qual irá ser previsto o consumo de energia. Os horizontes temporais de previsão podem ser classificados em três classes [9]: *short-term load forecasting* (STLF), *medium-term load forecasting* (MTLF) e *long-term load forecasting* (LTLF). O STLF é definido como a capacidade de previsão até as próximas 24h, o MTLF remete para previsões que vão desde o próximo dia até ao próximo ano e o LTLF desde o próximo ano até aos 10 anos seguintes. Estes horizontes temporais podem ter ligeiras alterações, segundo [6] o STLF mantém o horizonte desde 1 hora até às próximas 24 horas mas os horizontes MTLF e LTLF são definidos respetivamente por um dia até algumas semanas e alguns meses até um ano. Cada um destes horizontes pode ser utilizado para diferentes fins [10] que vão desde a gestão do dia-a-dia até à capacidade de planear investimentos a longo prazo. Algumas das finalidades dos três tipos de previsão são abordadas nas subsecções seguintes.

### **2.2.1. Short-Term**

O STLF é essencial para gerir de melhor forma a geração e transmissão de energia. As previsões de curto prazo são utilizadas como *inputs* de algoritmos de geração e transmissão tentando minimizar custos bem como gerir a entrega de energia de forma segura mesmo ocorrendo falhas ou comportamentos inesperados por parte dos consumidores [11]. Este horizonte temporal é também importante quando estamos perante mercados de energia em que previsões de curto prazo nos permitem negociar as compras e vendas de energia em função das necessidades ou excessos de produção. Uma previsão adequada permite prevenir sobrecarga na rede reduzindo a ocorrência de falhas, uma vez que podemos tomar ações para que esta subcarga não seja severa e provoque uma falha no fornecimento de energia.

### **2.2.2. Medium-Term**

MTLF poderá ser útil para planear a expansão das centrais de geração, planear operações de manutenção, realizar melhorias nos sistemas e planear estratégias de aquisição de combustível mais eficientes [11], [12]. Por exemplo, sabendo que o consumo de um determinado dia da próxima semana será mais reduzido que o habitual, podemos gerar menos energia para esse dia, bem como planear intervenções na rede.

### **2.2.3. Long-Term**

LTLF é essencial também para planear a expansão do sistema elétrico num futuro. É estratégico perceber a evolução do consumo de energia, para ter um plano eficiente de construção, ou aquisição de novas centrais e unidades de geração ou inativação de unidades. Ao nível da distribuição, conseguimos perceber se a atual infraestrutura suportará a carga necessária a longo prazo, e quais os investimentos que serão necessários face aos futuros consumos [13]. A previsão *Long-Term* é fulcral na estratégia de geração de energia, ou seja, de entre as várias formas de geração de energia, umas mais amigas do ambiente do que outras, para que esta seja sustentável, eficiente e rentável para a *Utility*.

De entre as várias metodologias e modelos para previsão de cargas referenciadas na literatura, a maioria destas focam-se na previsão de consumos com base na agregação do consumo de energia ao nível de uma sub-rede (ao nível de uma central ou transformador).

Neste nível de agregação, não conseguimos identificar perfis de consumo dos consumidores, nem ter a previsão precisa do consumo de cada cliente de modo a aumentar a eficiência na faturação, em campanhas, na segurança da distribuição, ou seja, em todos os serviços ligados diretamente ao cliente [14]. Com esta abordagem não está a ser retirada toda a potencialidade dos *Smart Meters*, pois não são analisados dados ao nível de cada equipamento, cuja granularidade é drasticamente superior, permitindo, entre outras coisas, prever consumos nos vários períodos temporais para toda a hierarquia de regionalização – região; posto de transformação; cliente individual.

### **2.3. Identificação de fatores com impacto na carga**

Existem vários fatores que poderão ter impacto direto para o nível de consumo [15]. De entre os fatores identificados destacam-se os fatores temporais, nomeadamente em termos de dia da semana, hora do dia, mês ou estação do ano [16]. Estes fatores são cíclicos, como é o caso dos fins-de-semana, e tendem a ser sazonais (associados a diferentes períodos do ano). Outro fator de relevo está relacionado com as condições meteorológicas, especificamente a temperatura. Acontecimentos extraordinários, como jogos de futebol ou outros eventos sociais têm impacto na alteração do perfil típico de consumo. Isto pode levar a que seja verificado um aumento de consumo em certas zonas específicas, levando a uma sobrecarga e ou uma diminuição em outras durante um período de tempo. Identificam-se também fatores económicos e demográficos, como a área geográfica (meio urbano ou meio rural) ou o crescimento económico de uma determinada área ou região. Outro fator que tem impacto no consumo é a própria oferta tarifária da empresa de comercialização de energia pois poderá incentivar o consumo de pico em certas horas do dia [17].



## 2.4. Técnicas para previsão de consumo

A gestão da procura de energia é fundamental para uma alocação de recursos mais adequados às necessidades. Nos últimos anos muitos foram os investigadores que focaram os seus trabalhos no campo da previsão. Tal como se viu anteriormente, o desafio de previsão de consumos é complexo pois a procura de energia não é linear e pode ser afetada por diversos fatores [7].

O consumo de energia assume um comportamento que pode ser analisado sob o ponto de vista de uma série temporal. As séries temporais podem ser definidas como uma sequência de observações de uma variável ao longo do tempo. Dito de outra forma, é uma sequência de pontos em ordem sucessiva durante um período de tempo.

Muitas técnicas e métodos foram estudados e utilizados para fazer essa gestão da procura de energia com o objetivo de prever de forma mais precisa quais as necessidades futuras. Tradicionalmente, métodos como séries temporais [18], [19], regressões [20], *fuzzy logic*, algoritmos genéticos ou redes neuronais [20] são utilizados com esse fim. Em [21] são explicadas várias técnicas e como cada uma poderá contribuir para prever o consumo de energia bem como as suas vantagens e desvantagens. Entre os primeiros métodos estatísticos usados para este fim, encontram-se as regressões lineares, processos estocásticos, *Auto-Regressive and Moving Average* (ARMA) e métodos de Box-Jenkins [22]. O uso de técnicas de previsão que têm por base o comportamento de uma série temporal é muito comum em várias áreas e a previsão de consumos de energia não é exceção. Este tipo de técnica assume que há correlação e uma tendência nos dados usados. Dentro dos algoritmos de séries temporais, os métodos mais usados são: ARMA (autorregressivo de média móvel), ARIMA (autorregressivo integrado de média móvel), ARMAX (autorregressivo de média móvel com variáveis exógenas) e o modelo ARIMAX (autorregressivo integrado de média móvel com variáveis exógenas) [23].

De forma a ter em conta variáveis que podem influenciar o consumo, as redes neuronais artificiais (ANN- *Artificial neural networks*) ganharam uma importância na previsão de consumos demonstrando capacidade de apresentar bons resultados [24]. As redes neuronais para além de mostrarem a capacidade de aprenderem conseguem também modelar e ter em conta características não lineares entre os consumos e outras variáveis externas.

Para além de aplicações que utilizam estas técnicas de previsão direta e individualmente, podemos encontrar também relações entre várias técnicas de *data mining* de forma a obter melhores resultados nas previsões [25].

Os consumos de energia dos *Smart Meters* poderão ser agregados de diferentes formas: por área geográfica, de acordo com a central de transformação a que está associado a cada contador, e virtualmente poderão ser criados grupos tendo em conta consumidores com características semelhantes [25]. Aumentar a precisão da previsão de consumo com base nos valores agregados por grupo não é um tópico novo. Em [17] é estudado o impacto positivo da previsão com base em grupos e a influência que previsões mais precisas têm nos mercados de energia. As técnicas de segmentação têm um papel importante na previsão de consumo na medida em que serão gerados modelos mais afinados aos diferentes padrões de consumo, ao invés de generalizar para todo o universo. Em [25] os autores utilizaram o algoritmo de *K-Means* para segmentar o universo de *Smart Meters* em grupos com comportamentos semelhantes e seguidamente utilizaram o algoritmo de redes neuronais de forma a prever o consumo por cada grupo.

Alguns exemplos de algoritmos de previsão podem ser encontrados em:

**Tabela 3 - Exemplos de algoritmos de previsão**

| <b>Técnica</b>          | <b>Referências</b>                |
|-------------------------|-----------------------------------|
| <b>Clustering + ANN</b> | [25]                              |
| <b>ANN</b>              | [9] , [16], [24] [26], [27], [28] |
| <b>ARIMA</b>            | [18], [19]                        |
| <b>Regression + ANN</b> | [20]                              |
| <b>ARIMA + ANN</b>      | [23]                              |

## 2.5. Previsão baseada em *Machine Learning*

As técnicas de previsão baseadas em aprendizagem automática (*machine learning*) ganharam uma grande importância ao longo dos tempos, existindo vários autores que utilizam estas técnicas para a previsão de consumos de energia. Entre os algoritmos que ganham maior destaque, as Redes Neurais Artificiais ganharam particular destaque [29]. A grande vantagem deste tipo de algoritmo está relacionada com a sua capacidade de representar relações lineares e não lineares entre as variáveis e aprender com essas relações [26].

Existem vários tipos de ANNs caracterizadas pela sua topologia e método de aprendizagem. No que diz respeito à previsão de consumo de curto prazo, que é o principal objeto de estudo deste trabalho, o tipo que ganha maior destaque é o algoritmo de *Backpropagation* [29].

A conjugação de vários algoritmos tem vindo a ganhar destaque, e nesse sentido este trabalho pretende seguir essa abordagem. Serão utilizados os algoritmos de *K-Means*, para formar grupos de clientes cujo seu perfil de consumo se assemelhe, e posteriormente o consumo agregado por grupo será utilizado para treinar uma rede neuronal.

### 2.5.1. K-Means

O algoritmo de *K-Means* [30] é um dos mais simples e mais utilizados métodos para segmentar informação e o seu objetivo passa por classificar cada entrada de dados num determinado *cluster* dentro do número de *clusters* definido à priori. O seu funcionamento [31] passa por seleccionar aleatoriamente um centro para cada *cluster*, com base nos dados de entrada, e seguidamente associar os outros pontos de entrada ao centro mais semelhante, isto é, cuja distância para o centro é mais próxima. O passo seguinte passa por recalcular o centro do *cluster* com base nos dados que foram associados e voltar a associar todos os pontos do conjunto de dados ao centro mais próximo. Este processo de recalcular os centros de cada *cluster* é executado até não existirem alterações de associações de dados a *clusters*. De acordo com o funcionamento acima descrito, é necessário definir previamente um número fixo de *clusters* para que cada ponto dos dados de entrada seja associado a um determinado *cluster*.

### 2.5.2. Redes Neurais

Uma Rede Neuronal Artificial pretende sintetizar o funcionamento das redes neuronais biológicas, onde cada neurónio está em constante comunicação com os outros neurónios através de ligações sinápticas. De uma forma muito resumida, as redes neuronais caracterizam-se por unidades consideradas neurónios, que possuem funções de ativação individuais, e que estão conectadas entre si de uma forma hierárquica, através de ligações com um determinado peso também individual [31].

As ANNs possuem inúmeras vantagens, tais como: tolerância a ruído nos dados, capacidade de classificar ou prever padrões cujo seu resultado não tenha sido treinado bem como perceber relações entre variáveis cujo conhecimento acerca da sua relação é muito reduzido [31]. Outras características importantes deste tipo de abordagem são o facto de terem a capacidade de processar um volume massivo de dados numa estrutura multicamada existindo a possibilidade de paralelizar o processamento, a existência de um processo de treino, ao contrário de outros processos analíticos ou matemáticos [32], e ainda a grande capacidade que este tipo de algoritmos têm de processar e prever variáveis contínuas.

### 2.5.2.1. Multilayer feed-forward Neural Network

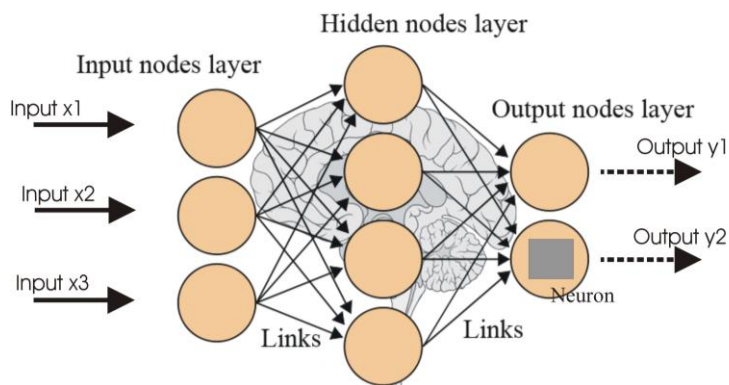


Figura 3 - Multilayer feed-forward ANN

Este tipo de rede neuronal caracteriza-se por ter uma camada de entrada, uma ou mais camadas escondidas (*hidden layers*) e uma camada de saída [31]. Cada uma destas camadas é composta por neurónios onde:

- A camada de entrada tem o número de neurónios correspondentes ao número de variáveis de entrada;
- A(s) camada(s) escondidas(s) tem um número de neurónios que deverá ser definido e que pode variar, influenciando o desempenho da rede de modo a evitar o subdimensionamento (pondo em causa o desempenho da rede)
- O número de neurónios de saída corresponde ao número de variáveis de saída.

Neste tipo de ANN, os neurónios não interagem com os neurónios da mesma camada, e estão ligados a todos os neurónios da camada seguinte tendo como ação transmitir um sinal de saída para os neurónios da próxima camada (*feed-forward*).

Relativamente ao funcionamento específico de cada neurónio (Figura 3), estes têm uma lógica de funcionamento semelhante entre si. Os pesos (representados por  $w_{nj}$  na Figura 4) e a *Bias* (representado por  $\theta$ ) são iniciados aleatoriamente com valores pequenos (tipicamente entre -1.0 e 1.0 ou 0 e 1). Cada valor do tuplo de entrada que chega ao neurónio é multiplicado pelo peso da respetiva ligação ao neurónio. Em seguida, todos estes valores são somados obtendo-se o *net input* (Figura 4 – Funcionamento de um neurónio artificial). Cada neurónio possui ainda uma função de ativação [33] (*activation*

function na Figura 5 – Funções de ativação dos neurónios) de forma a calcular o valor à saída do neurónio.

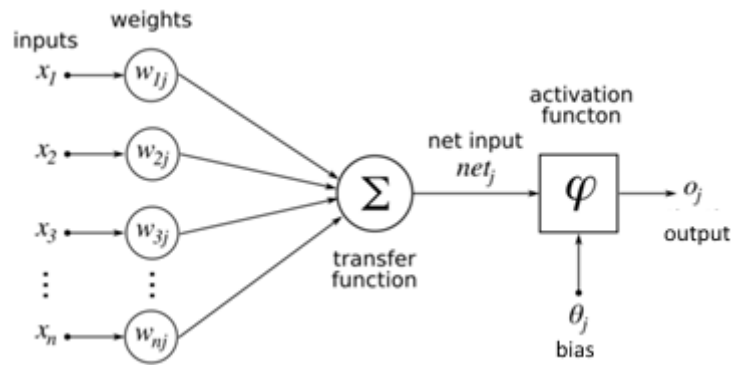


Figura 4 – Funcionamento de um neurónio artificial

Existem diversos tipos de função de ativação tais como logísticas, funções de degrau (*threshold activation*), funções lineares ou tangente hiperbólica (Figura 5 – Funções de ativação dos neurónios).

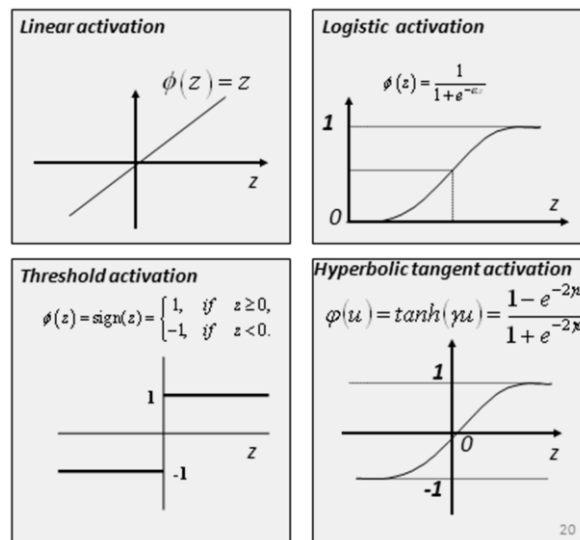


Figura 5 – Funções de ativação dos neurónios

Finalmente, podemos representar a expressão que reflete o funcionamento de cada neurónio da rede neuronal. O funcionamento de cada neurónio depende dos outputs dos neurónios das camadas anteriores, dos pesos associados a cada ligação entre os neurónios da camada anterior e os neurónios da próxima camada, do valor da *bias* que cada neurónio

possui e da sua função de ativação. Este comportamento pode ser expresso matematicamente pela seguinte expressão:

$$O_j = \vartheta(\sum_1^n W_{nj} \cdot X_n + \theta_j) \quad (1)$$

Onde:

$\vartheta$  é a função de activação;

$W_{nj}$  é o peso da ligação entre o neurónio  $n$  da camada anterior e o neurónio  $j$  ;

$X_n$  é o valor à saída do neurónio  $n$  da camada anterior;

$\theta$  é a *bias* do neurónio. Este parâmetro funciona como um *threshold* de forma a variar os valores do neurónio;

$O_j$  é o valor de saída do neurónio  $j$ .

Antes de podermos treinar a rede neuronal é necessário escolher a topologia da própria rede, isto é, é necessário definir o valor dos parâmetros que podemos controlar, tais como: o número de neurónios da camada de entrada, o número de neurónios de cada uma das camadas escondidas, as funções de activação dos neurónios da camada escondida, o número de neurónios da camada de saída e as suas funções de activação. Não existem regras bem definidas acerca de quais os parâmetros a utilizar para cada caso, tipicamente a escolha destes valores recorre a tentativa e erro.

### 2.5.2.2. Treino da Rede Neuronal

Uma das características mais importantes durante o processo de treino é garantir que a rede neuronal adquira a capacidade de generalização. Esta característica é importante para que a rede consiga ser eficiente na previsão de valores que não constem do conjunto de treino. De uma forma simples, aumentando o número de neurónios da rede, podemos dotar a rede neuronal de uma boa capacidade de previsão para o conjunto de treino. A olho nu pode parecer bom mas poderá ter um efeito negativo a que chamamos *overfitting* [28]. O conceito de *overfitting* define-se como a incapacidade da rede fazer generalizações. Isto é, a rede adapta-se demasiado aos valores de treino e perde a capacidade de fazer previsões satisfatórias para valores que não pertençam ao conjunto de treino.



### 2.5.2.2.1. Backpropagation

O objetivo de uma rede neuronal é produzir sinais à saída da rede, através do processamento dos sinais de entrada, de forma a aproximar os resultados da rede dos resultados desejados. Sendo as redes neuronais metodologias de aprendizagem supervisionada [31], é possível durante o processo de treino verificar qual o erro associado à previsão. Durante esta fase, a rede conhece os valores de entrada e os respectivos valores saída. O erro é dependente da topologia escolhida para a rede, dos pesos das ligações e das *bias* dos neurónios. Este método de treino “aprende” através de um processo iterativo de processamento dos vários tuplos de entrada comparando o valor estimado pela rede para as variáveis de saída com os valores conhecidos durante o treino. Para cada um dos tuplos de entrada, os pesos de cada ligação são afinados de forma a minimizar o erro quadrático médio entre as previsões da rede neuronal e o valor real para o valor alvo. Este processo de afinar os pesos de cada ligação e *bias* é feito de forma retroativa, ou seja, desde a camada de saída até à camada de entrada, passando por cada camada escondida.

## 2.6. Metodologia de desenvolvimento

De forma a seguir uma metodologia estável e fidedigna o processo de previsão de consumos de energia nesta dissertação, irá seguir uma metodologia de *data mining* baseada no *state-of-the-art framework* CRISP-DM [34] (*Cross Industry Standard Process for Data Mining*). Esta *framework* divide o processo de *Data Mining* em seis fases, *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation e Deployment* (Figura 6 – Framework CRISP-DM). Este é um processo iterativo em que a sua sequência de fases não é estritamente definida, podendo sempre avançar ou voltar para trás para uma fase anterior e repetir o processo. As setas existentes na Figura 6 – Framework CRISP-DM representam as mais importantes e frequentes dependências em termos de cada fase. O círculo exterior reforça a ideia que o processo de *data mining* não é um ciclo fechado. Após o *deployment* é necessário revalidar os modelos e processos através da introdução de novos dados significativos. Cada fase do processo é definida da seguinte forma:



**Figura 6 – Framework CRISP-DM**

***Business understanding:*** esta é a fase inicial em que o foco é perceber o problema e os objetivos do ponto de vista do negócio convertendo os conhecimentos num problema de *data mining*.

***Data understanding:*** baseia-se na exploração dos dados criando conhecimento acerca dos mesmos. É nesta fase que se identificam problemas de qualidade de dados, descobrir e captar *insights* sobre os dados e detetar *subsets* interessantes de forma a formular hipóteses.

***Data preparation:*** a fase de preparação de dados abrange todas as tarefas necessárias de forma a construir, partindo dos dados iniciais, o *data set* que servirá de entrada no(s) modelo(s). Estas tarefas incluem manipulação e transformações de dados.

***Modeling:*** nesta fase, várias técnicas de modelação serão seleccionadas e aplicadas e os seus parâmetros serão calibrados e ajustados de forma a obter valores ótimos. Tipicamente, existem várias técnicas para o mesmo problema, ainda assim, algumas técnicas exigem alguns requisitos ao nível da preparação de dados e poderá ser necessário voltar à fase anterior.

***Evaluation:*** ao alcançar esta fase, o modelo já estará construído e é necessário avaliar minuciosamente os resultados de forma a perceber se o modelo responde aos objetivos

identificados. No final desta fase deverá ser possível medir a qualidade dos resultados e decidir quando à sua utilização ou não.

**Deployment:** A fase de *Deployment* de projetos de *data mining* pretende organizar e apresentar os conhecimentos obtidos através dos modelos de forma útil e simples para o negócio. Dependendo dos requisitos esta fase poderá ser simples, como a criação de um relatório ou complexa, como estender o processo de *data mining* a toda a empresa.

# Capítulo 3

## 3. Caso de Estudo

### 3.1. Contexto

O trabalho realizado utiliza como caso de estudo um conjunto de dados de consumos de energia reais de uma *Utility* da Irlanda publicados pela Comissão para a Regulação de Energia [35]. Estes dados foram obtidos através da *Irish Social Science Data Archive* (ISSDA) [36], que publicou dados anonimizados, de *Smart Meters* refletindo o consumo de clientes com uma periodicidade de 30 em 30 minutos. Os dados obtidos abrangem um período temporal que vai desde 14 de Julho de 2009 até 31 de Dezembro de 2010 refletindo o comportamento de mais de 4000 clientes residenciais e mais de 400 clientes empresariais. Para além dos dados relativos a consumos de energia, foram realizados inquéritos a cada um dos clientes, residenciais (*Residential*) e empresariais (SME – *Small and Medium Enterprise*), com o objetivo de recolher informação complementar e perceber hábitos e preocupações de cada consumidor. Através da análise dos dados dos inquéritos foi possível extrair variáveis interessantes, que pudessem fornecer algum detalhe acerca de cada cliente. Estas variáveis são por exemplo, o sector de atividade de cada cliente empresarial, o número de colaboradores da organização ou o número de habitantes de uma residência, no caso dos clientes residenciais. As variáveis retiradas dos inquéritos, que caracterizem a população, tornam-se interessantes na medida que nos fornecessem mais detalhe acerca da população a estudar e que permitem, em última instância, perceber o impacto que cada variável tem no perfil ou amplitude de consumo de energia. Informações relacionadas com condições meteorológicas, que têm impacto significativo no consumo, não constam originalmente no conjunto de dados disponíveis. Apesar de identificado o país, as cidades ou locais específicos de consumo são desconhecidos pelo que dificulta o enriquecimento de dados utilizando variáveis demográficas. De forma a acrescentar informação relevante aos dados disponíveis, foi estudada a hipótese de introduzir a média de temperaturas de algumas cidades da Irlanda (através de uma amostra que pretendia captar várias cidades) para o período de dados disponível. Esta análise poderá ser encontrada na secção 3.4 onde é possível observar que existe uma grande relação entre os dados de consumo e as temperaturas recolhidas.

### 3.2. Modelação Conceptual

Do ponto de vista conceptual, e tendo em conta o caso de estudo apresentado, o consumo de energia pode ser analisado sob vários prismas. Onde foi consumida a energia? Quando foi consumida? Quem foi o consumidor? Por que razão é consumida mais ou menos energia num determinado período? Cada uma destas perguntas/prismas pode ser respondida cruzando dimensões de análise com o valor de energia consumida. Desta forma, é possível perceber a distribuição do consumo em função de cada dimensão. Uma destas dimensões é a dimensão temporal que tipicamente é caracterizada com uma hierarquia (*When?*) que nos permite olhar para o valor do consumo agregado a vários níveis (Ano, estação do ano, mês, Semana, dia, ou horariamente).

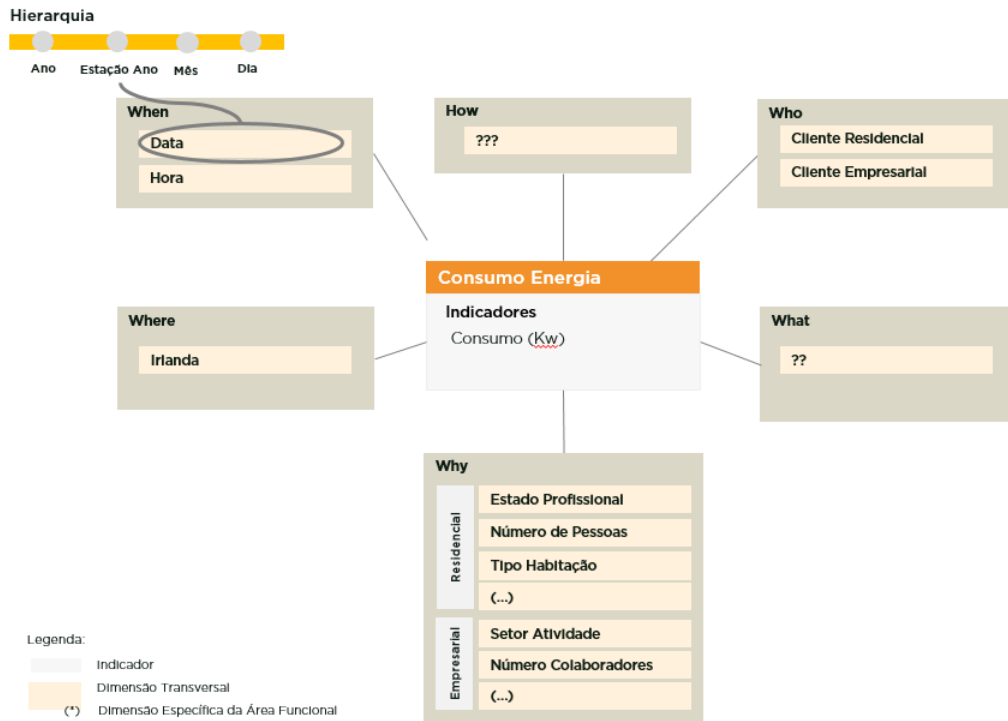
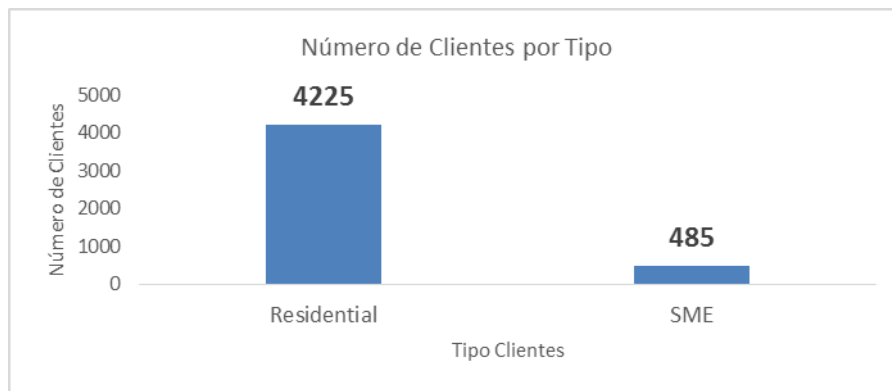


Figura 7 – Modelo conceptual de caracterização de caso de estudo

### 3.3. Caracterização da População

A população está dividida em dois tipos de clientes, Residenciais e SME (*Small and Medium Enterprise*). Esta distribuição da população encontra-se distribuída da seguinte forma:



**Figura 8 – Distribuição de clientes por tipo**

A Figura 8 – Distribuição de clientes por tipo mostra claramente um enviesamento da amostra para uma das classes. Neste caso, a classe de clientes residenciais corresponde a cerca de 90% da população estudada. Dada esta distribuição optou-se por estudar os consumos dos dois tipos de clientes separadamente e utilizar uma metodologia de previsão de consumos orientada a cada um dos tipos.

No anexo Distribuição da população pelas variáveis dos inquéritos 8.5 são apresentadas outras estatísticas sobre a população e como esta se divide em função das restantes variáveis de análise. No entanto, estas variáveis categóricas não foram incluídas nos modelos de *clustering* nem na modelação das redes neuronais. Esta abordagem foi adotada com o objetivo da segmentação da população ser feita unicamente através do seu perfil de consumo de energia, uma vez que nem sempre os *datasets* que as *utilities* possuem para fazer previsão de consumos contem informação sobre a população em estudo. Após a segmentação de clientes por perfil de consumo, é possível perceber como cada uma das variáveis categóricas está distribuída em cada um dos *clusters* e perceber o impacto ou como cada uma se relaciona com cada um dos perfis de consumo identificados. Existem estudos [37] que utilizam este tipo de caracterização da população para perceber a sua relação com perfis típicos de consumo. Esta associação é útil na medida em que para novos clientes, cujo seu perfil de consumo não é conhecido, permite numa fase inicial, determinar qual o perfil de consumo de um determinado cliente.

### 3.4. Análise de Consumo de Energia

O consumo de energia elétrica caracteriza-se tipicamente por ser uma série temporal. Esta assenta em intervalos de tempo regulares, uniformes e ocorrendo de forma sucessiva ao longo do tempo. Tal como referido anteriormente, a temperatura é um dos fatores cujo impacto no consumo de energia é significativo, estando diretamente relacionada com o aumento ou diminuição do consumo de energia elétrica. A Figura 9 – Relação entre o consumo total agregado por mês e temperatura média representa o consumo de energia e a temperatura média, de algumas cidades da Irlanda, ao longo do tempo, em função do mês, entre Julho de 2009 e Dezembro de 2010. Através desta análise é possível observar que a temperatura segue uma tendência inversamente proporcional à amplitude do consumo de energia. Isto é, temperaturas mais altas, observadas nos meses tipicamente mais quentes, apresentam consumos de energia mais baixos e vice-versa.

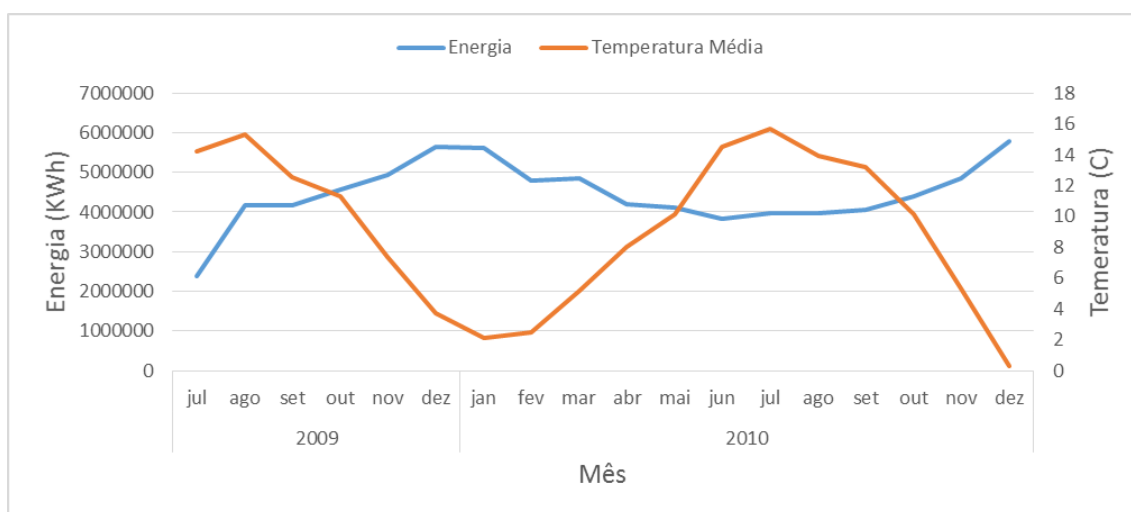
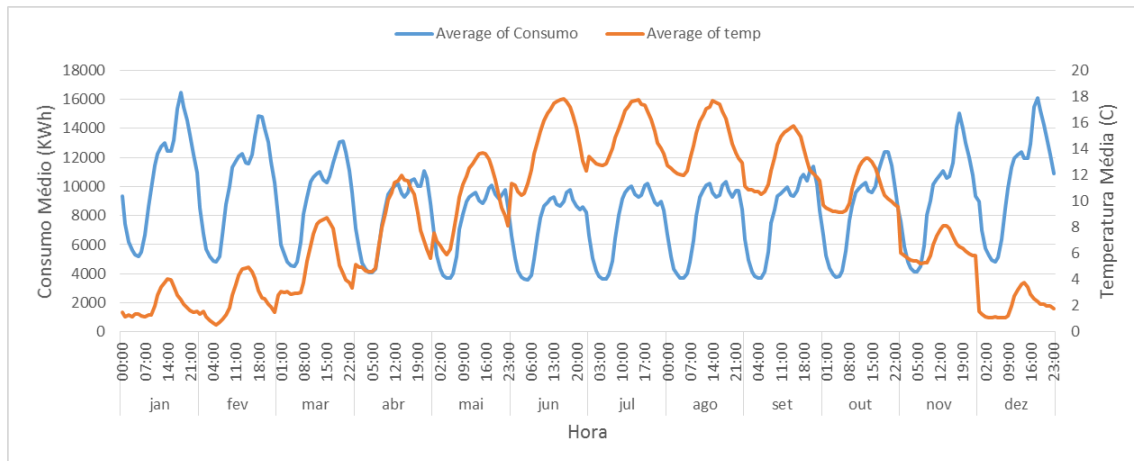


Figura 9 – Relação entre o consumo total agregado por mês e temperatura média

Dado esta correlação negativa forte entre a temperatura e o consumo, podemos concluir que a temperatura têm impacto no valor de energia consumida e pode ser determinante para uma previsão mais adequada. No entanto também existe uma correlação forte entre a temperatura e a dimensão temporal, ou seja, a variação da temperatura está relacionada com a hora do dia, mês ou estação do ano. Embora exista relação entre a temperatura e a dimensão temporal não existem evidências de alteração do perfil diário tendo em conta a

temperatura e/ou o mês/estação do ano, ou seja, há um aumento ou diminuição do consumo de energia distribuído de igual forma ao longo do dia, independentemente da temperatura ou do mês. Através da Figura 10 – Relação entre perfil da temperatura média e perfil de consumo por hora podemos observar que onde se evidencia maior impacto é ao nível da amplitude de consumo de energia que se apresenta quase uniformemente distribuída pelo perfil diário.

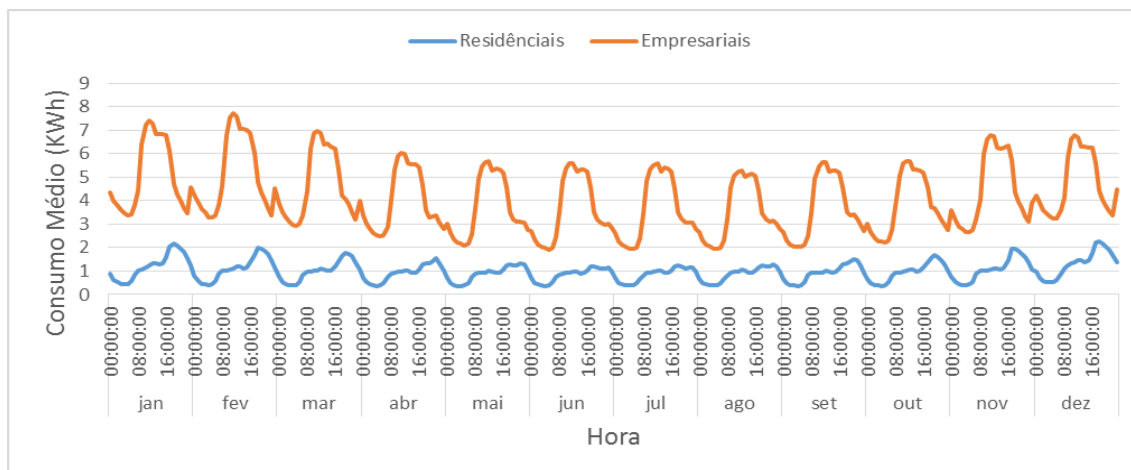


**Figura 10 – Relação entre perfil da temperatura média e perfil de consumo por hora**

Uma vez que a previsão de consumo será efetuada com base em cada tipo de cliente (Residencial e Empresarial) é necessário perceber como varia o perfil de consumo ao longo do tempo para cada tipo de cliente e de que forma os fatores como a dimensão temporal influenciam esse perfil e a própria amplitude do consumo.

A Figura 11 – Consumo médio por hora ao longo de cada mês para o ano 2010, representa para cada tipo de cliente o consumo médio mensal por hora. Tal como se pode perceber cada tipo de cliente tem o perfil de consumo e amplitude diferentes, no entanto verifica-se a existência um padrão para cada tipo ao longo das horas do dia.

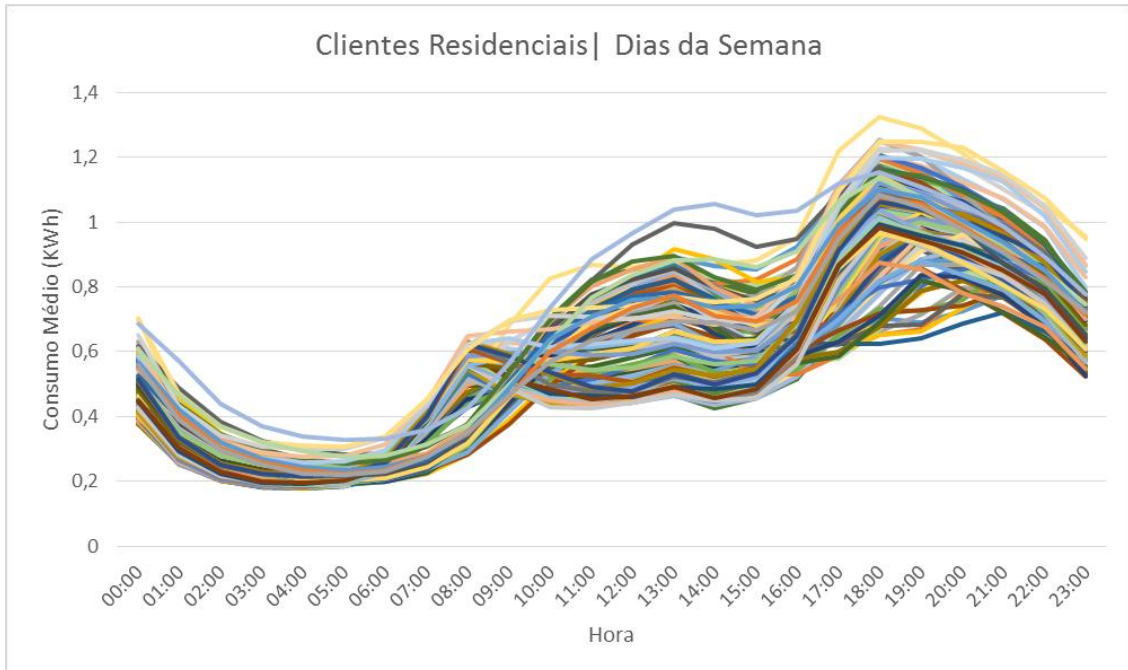




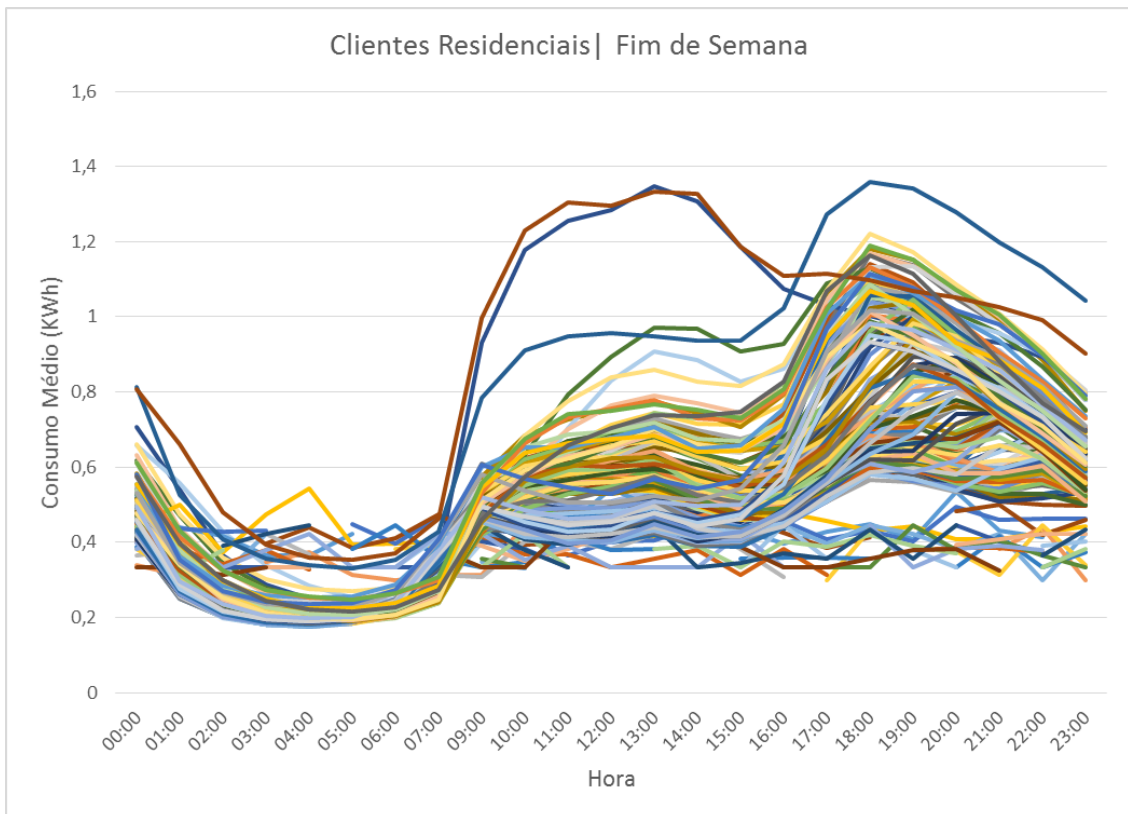
**Figura 11 – Consumo médio por hora ao longo de cada mês para o ano 2010.**

O tipo de dia tem também influência no padrão de consumo de energia. Os hábitos de consumo variam conforme o dia da semana, nomeadamente quando observamos a diferença existente entre dias tipicamente pertencentes ao período laboral e o fim-de-semana. Observando a Figura 12 – Consumo médio por hora para clientes residenciais para que representa o consumo médio ao longo de todo o período para dias da semana (segunda a sexta) para os clientes residenciais, conseguimos perceber que o comportamento é bastante semelhante em termos de perfil, variando em amplitude, aspeto que já foi justificado anteriormente de acordo com as diferenças de consumo ao longo dos vários meses. É possível observar através do gráfico que existem dois momentos onde a alteração do perfil é mais considerável. O período que ocorre entre 07:00 e 09:00, onde se verifica que o consumo de energia tem um aumento substancial e sofre novamente um pico no período compreendido entre as 17:00 e 20:00. Estes períodos estão intimamente ligados com as dinâmicas típicas dos clientes residenciais, nomeadamente associados ao período em que as pessoas acordam e o período em que tipicamente regressam a casa. No que diz respeito aos fins-de-semana o comportamento apresenta algumas diferenças, essencialmente no segmento de clientes residenciais cujo comportamento é mais dinâmico. É possível notar que em termos de perfil global apresenta bastantes semelhanças quando comparado com a análise dos dias de semana, mas existem dias que apresentam desvios mais acentuados face ao perfil habitual. Observando o consumo médio por cada estação do ano (Figura 14 – Consumo médio para clientes residenciais para cada estação do ano), que neste caso é a agregação dos vários dias pertencentes a

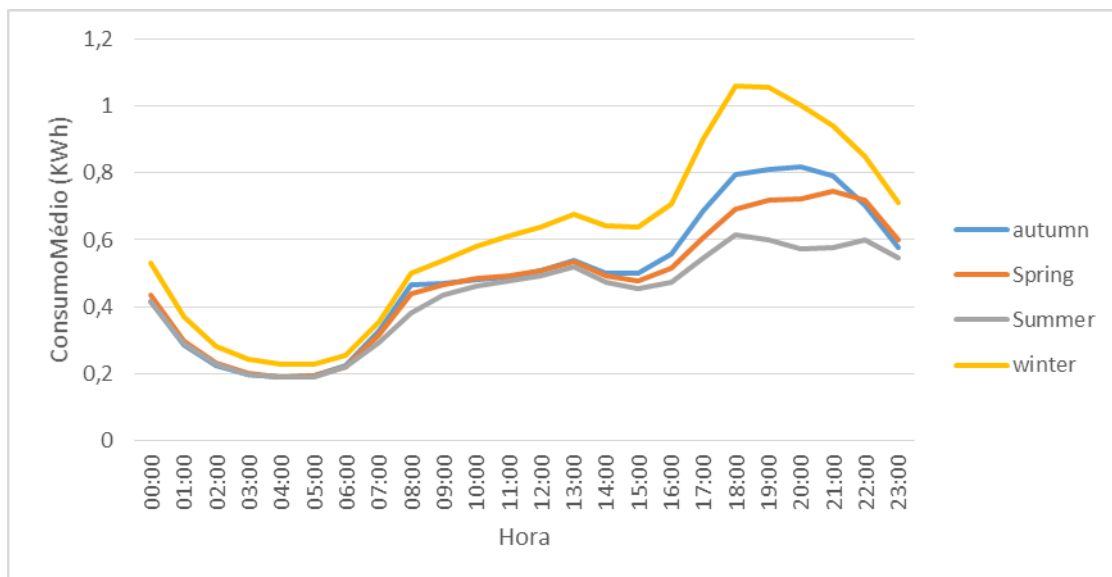
cada uma das estações, é possível observar que ao nível do perfil este é mantido ao longo das horas do dia variando a amplitude em função da temperatura.



**Figura 12 – Consumo médio por para clientes residenciais para dias da semana**

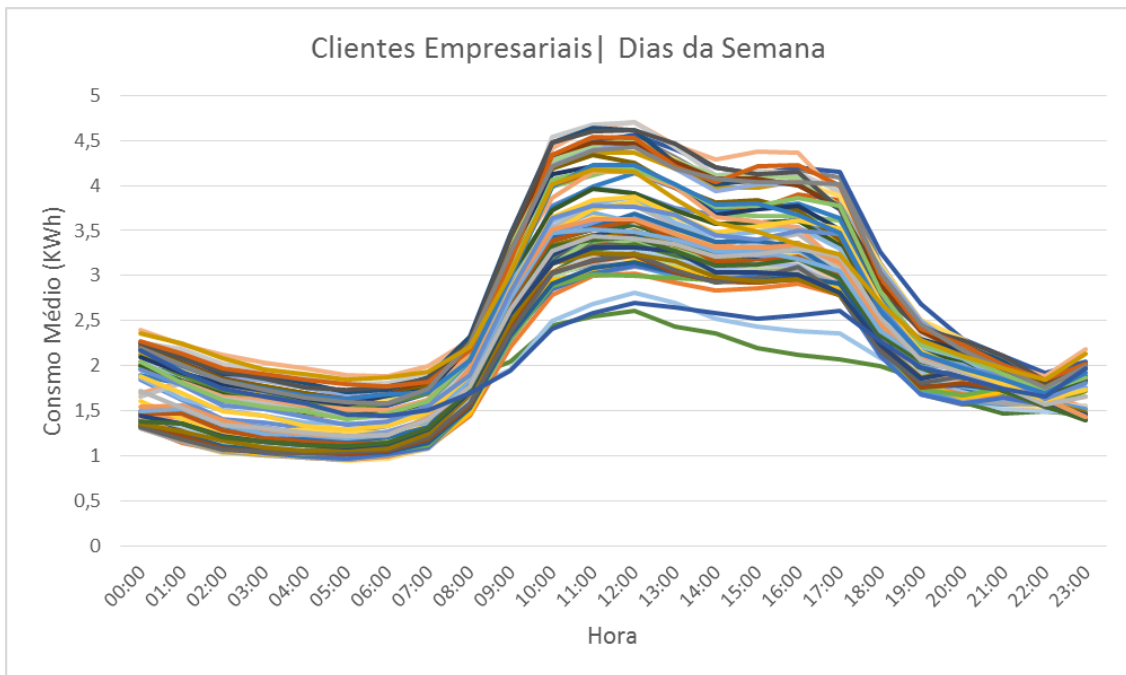


**Figura 13 – Consumo médio por hora para clientes residenciais para os fins-de-semana**

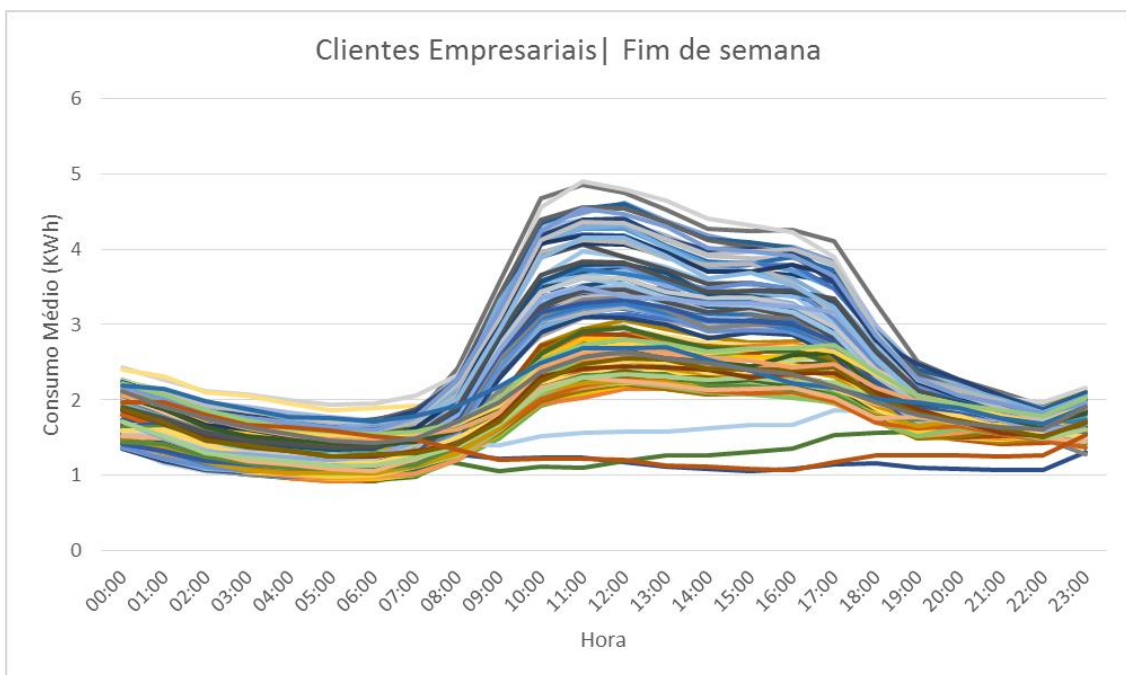


**Figura 14 – Consumo médio para clientes residenciais para cada estação do ano**

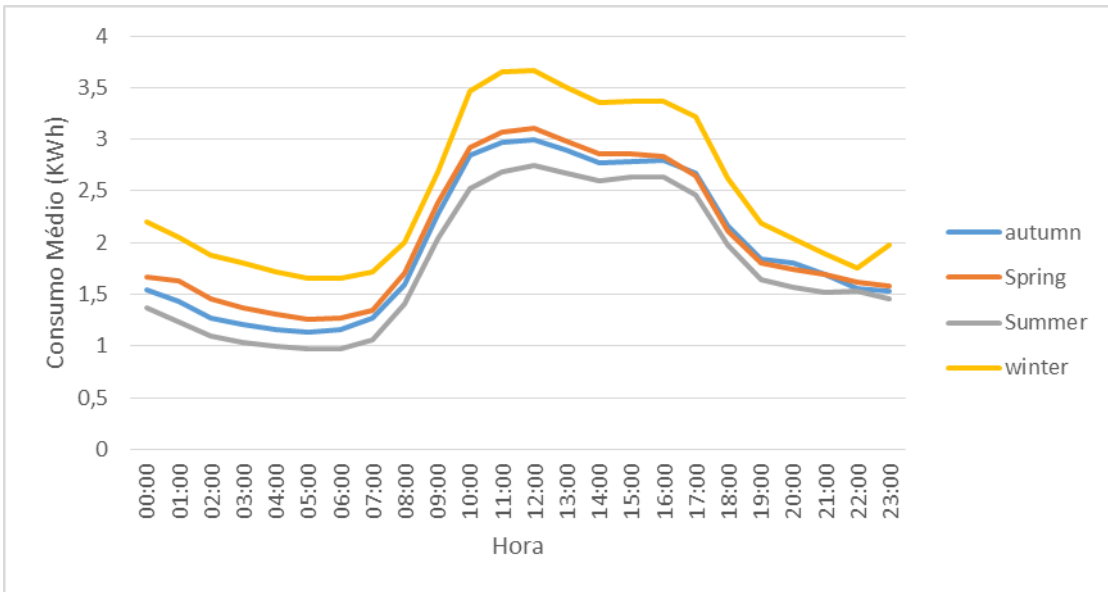
Relativamente aos clientes empresariais, cujo comportamento pode ser analisado na Figura 15 – Consumo médio por hora para clientes empresariais para dias da semana e Figura 16 - Consumo médio por hora para clientes empresariais para os fins-de-semana, e observando os dados de acordo com as mesmas condições, as conclusões são globalmente as mesmas. É possível observar que ao longo de todo o período existe uma semelhança significativa em termos de perfil de consumo com algumas diferenças relativamente a amplitudes. O consumo de energia para os clientes empresariais ao longo do dia é bastante mais uniforme existindo uma clara distinção entre horas que pertencem ao horário laboral típico, normalmente entre o período das 08:00 e 18:00, e horas de pós laboral. Tal como nos clientes residenciais a altura do ano tem influência na amplitude de energia consumida. É possível observar através da Figura 17 – Consumo médio para clientes empresariais para cada estação do ano que em alturas do ano que pertencem a estações mais frias o consumo é elevado ao passo que em estações mais amenas o consumo é mais reduzido.



**Figura 15 – Consumo médio por hora para clientes empresariais para dias da semana**



**Figura 16 - Consumo médio por hora para clientes empresariais para os fins-de-semana**



**Figura 17 – Consumo médio para clientes empresariais para cada estação do ano**

# Capítulo 4

## 4. Desenvolvimento

### 4.1. Previsão de Consumos de Energia

A metodologia adotada neste trabalho passa por um processo de quatro fases distintas. A primeira fase passa pela divisão do conjunto de dados em clientes residenciais e empresarias. A divisão é argumentada tanto pela disparidade do cardinal de clientes residenciais face ao cardinal de clientes empresariais, já referida anteriormente, como pelas diferenças bastantes significativas encontradas ao nível do perfil de consumo. Através desta divisão esperam-se construir modelos de previsão mais específicos e direcionados ao tipo de perfil, tendo em conta as suas características específicas.

O segundo passo surge na sequência do anterior e prende-se essencialmente com o mesmo objetivo, isto é, conseguir criar modelos mais específicos de acordo com os perfis de consumo dos clientes existentes na população. Este passo pretende segmentar os dados, ou seja, criar grupos cujo perfil de consumo é semelhante ao longo do tempo. A criação de grupos ou *clustering* é feito através dos dados de consumo médio hora a hora para cada estação do ano para cada cliente. O algoritmo de *clustering* terá como entradas 24 x 4, ou seja, 96 atributos representando o valor médio de consumo para cada uma das 24h a cada estação do ano. Desta forma, conseguimos aproximar os clientes cujo seu perfil é semelhante ao longo do tempo e dessa forma criar modelos de previsão mais específicos.

Após a identificação dos *clusters*, procedeu-se ao estudo dos mesmos relacionando-os com as variáveis dos inquéritos, ou seja, perceber se os clientes para cada um dos clusters obtidos estão uniformemente distribuídos. Desta forma, podemos afirmar que a previsão de consumo de cliente, se focará unicamente no perfil de consumo, e não nas variáveis categóricas dos clientes.

O passo seguinte, passa pelo treino de uma rede neuronal específica para cada grupo identificado. O consumo dos clientes é agregado por cada um dos *clusters* e será esse consumo agregado, em conjunto com outras variáveis (temporais e a temperatura), que servirá como *inputs* para o treino das várias redes neuronais. Desta forma o modelo de previsão será mais focado e direcionado ao perfil típico dos vários clientes.

O último passo, passa pela validação do modelo de previsão, utilizando dados de teste, que não tenham sido introduzidos no processo de treino e verificar qual a precisão do modelo.

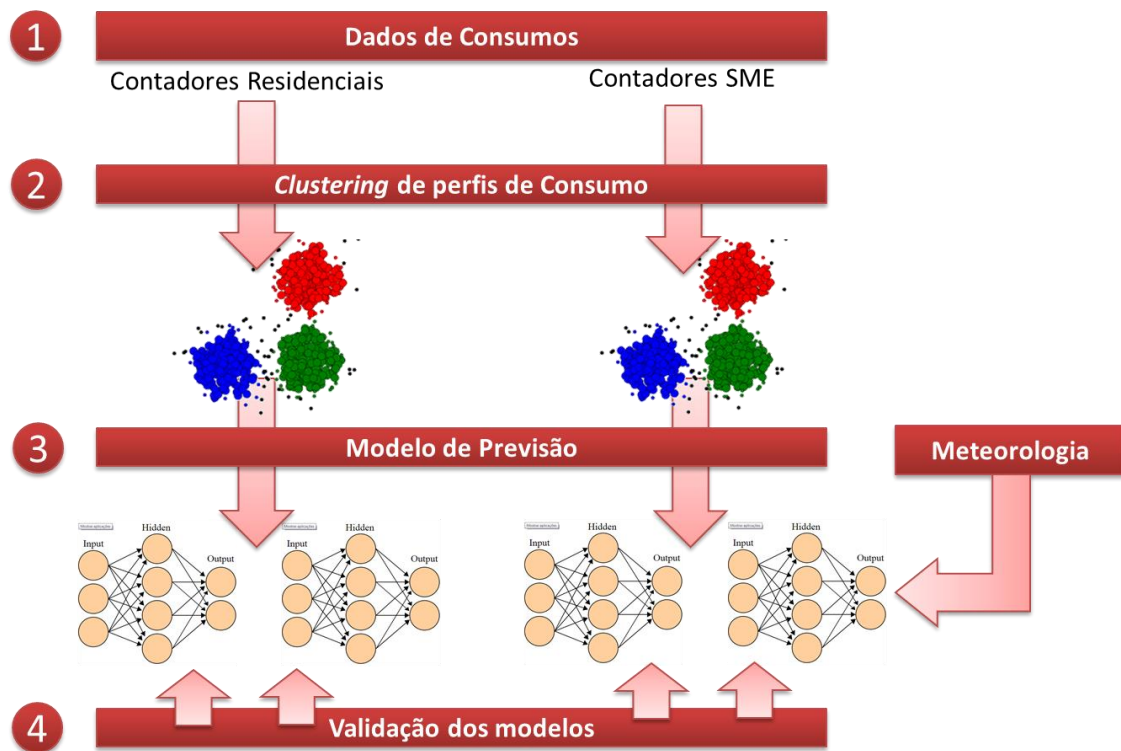


Figura 18 – Metodologia do trabalho desenvolvido

## 4.2. Clustering

No contexto de *machine learning* o processo de *clustering* ou segmentação de dados enquadra-se nos métodos de aprendizagem não supervisionada e o seu objetivo passa por fazer agrupamentos automáticos de dados de acordo com o seu grau de semelhança. Cada agrupamento ou *cluster*, é composto por um conjunto de dados que apresentam semelhanças entre si e disparidades entre os dados de outro qualquer *cluster* identificado [25]. Este tipo de método é utilizado em várias áreas e para vários fins, como por exemplo, detetar *outliers* (dados que apresentam comportamentos bastante diferentes face aos demais), agrupar clientes semelhantes com base nas suas características, entre outros. Este tipo de método pode ser utilizado individualmente como “ferramenta” de geração de novos *insights* sobre os dados ou pode servir de pré-processamento para outros métodos e técnicas que podem ganhar com a utilização e segmentação de dados [31].

No contexto deste trabalho, esta técnica é utilizada como um método de pré-processamento de dados, com o objetivo de agrupar clientes cujo consumo apresenta semelhanças ao longo do tempo. O objetivo passa por agregar os consumos de todos os clientes pertencentes a cada segmento identificado, de forma a “dividir” o consumo de energia com base em perfis típicos e dessa forma construir modelos de previsão mais direcionados e focados. Ao invés de um modelo global de previsão de energia, serão gerados vários modelos de previsão tendo em conta características particulares de cada perfil típico de consumo identificado.

Dado que o algoritmo utilizado, neste caso o *K-Means*, necessita de conhecer qual o número de grupos que irá gerar, foi necessário utilizar um método para a interpretação e validação do processo de *clustering*. O método utilizado foi o *Silhouette Index* [38] que pretende medir a coesão dos *clusters* criados de forma a tornar cada grupo o mais compacto possível mas devidamente separados entre si. Esta métrica é definida por:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Onde,

$a(i)$  Representa a média da desigualdade, em termos de distância, entre o objeto  $i$  e todos os outros dentro do mesmo *cluster*. Quanto menor este valor, melhor é a associação do objeto ao seu *cluster*;

$b(i)$  Representa a média da desigualdade, em termos de distância, entre o objeto  $i$  e todos os outros elementos de outros *clusters*;

E que pode ser simplificado em:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{se } a(i) < b(i) \\ 0, & \text{se } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{se } a(i) > b(i) \end{cases} \quad (3)$$

Observando a equação, e dado que o objetivo será maximizar a aproximação entre objetos do mesmo *cluster* e a separação de objetos de diferentes *clusters*, estamos em condições de referir que quando mais próximo de 1 este índice se encontrar melhor será a divisão do *dataset*.



### 4.3. Modelação das Redes Neurais

#### 4.3.1. Variáveis de Entrada e Saída

De forma a treinar a rede neuronal foi necessário definir um conjunto de variáveis de entrada que de alguma forma estão relacionadas e podem explicar o consumo de energia (variável de saída). O conjunto de variáveis selecionadas surgiu da análise efetuada aos dados no Capítulo 3. A tabela abaixo é um excerto dos dados modelados de forma a servir de *input* para a rede neuronal.

| hora   | Mês     | dia_semana | estacao   | temp   | flag_fds | estacao_7 |
|--------|---------|------------|-----------|--------|----------|-----------|
| 1      | 7       | 3          | 3         | 7,8    | 0        | 3         |
| 2      | 7       | 3          | 3         | 10,6   | 0        | 3         |
| 3      | 7       | 3          | 3         | 12,0   | 0        | 3         |
| 4      | 7       | 3          | 3         | 12,4   | 0        | 3         |
| 5      | 7       | 3          | 3         | 12,8   | 0        | 3         |
| temp_7 | cons_7  | dia_sem_1  | estacao_1 | temp_1 | cons_1   | Consumo   |
| 11,8   | 1612,45 | 2          | 3         | 11,4   | 1661,73  | 1604,60   |
| 11,8   | 846,74  | 2          | 3         | 11,1   | 870,90   | 833,80    |
| 11,4   | 729,34  | 2          | 3         | 10,4   | 734,08   | 714,27    |
| 10,7   | 682,78  | 2          | 3         | 10,3   | 681,17   | 672,08    |
| 10,3   | 680,25  | 2          | 3         | 10     | 673,48   | 672,69    |

Tabela 4 – Variáveis que compõem a rede neuronal

Onde:

Hora – hora do dia [1-24]

Mês – Mês do ano [1-12]

dia\_semana – Dia da semana no qual o consumo foi efectuado [1-7]

estacao – Estação do ano [1-4]

temp – Temperatura Média à hora do consumo

flag\_fds – Indicador de dia da semana ou fim de semana

estacao\_7 – Estação do ano da semana anterior no mesmo dia

temp\_7 – Temperatura Média à mesma hora do consumo na semana anterior

cons\_7 – Consumo de energia à mesma hora na semana anterior

dia\_sem\_1 – Dia da semana anterior

estacao\_1 – Estação do ano do dia anterior

temp\_1 – Temperatura média à mesma hora do consumo no dia anterior

cons\_1 – Consumo de energia à mesma hora no dia anterior

Consumo – Consumo de energia (Variável de Saída)

De modo a homogeneizar os valores possíveis para cada uma das variáveis é necessário recorrer a um processo de normalização das mesmas. Este processo é importante de forma a distribuir de igual forma o peso de cada variável de *input*. O objetivo passa por garantir que variáveis com grande dispersão de valores não causem entropia ao modelo, isto é, minimizar a importância que valores grandes, em termos absolutos, influenciem excessivamente o modelo.

Com o objetivo de normalizar as variáveis numa escala entre [0;1] em [39] é apresentado o seguinte método:

$$V_i = \frac{(x_i - \min(x))}{(\max(x) - \min(x))} \quad (4)$$

Onde,

$V_i$  é o valor normalizado da variável

$x_i$  é o valor real da variável

$\min(x)$  é o valor mínimo para todas as observações da variável  $x$

$\max(x)$  é o valor máximo para todas as observações da variável  $x$

Os valores normalizados são então submetidos ao modelo para o treino do mesmo e o *output* do modelo deverá ser desnormalizado para que se torne em valores com significado real.

### **4.3.2. Conjunto de Treino e Validação**

O treino da rede neuronal, sendo este um algoritmo de aprendizagem supervisionada [31], necessita de um conjunto de dados significativo e que espelhe de forma correta o comportamento do objeto em estudo. No caso das previsões de consumo é necessário contar com dados de consumo referentes ao passado para que seja possível prever o consumo de energia no futuro. De acordo com a metodologia acima descrita, foi treinada uma rede neuronal para cada *cluster* identificado. Dessa forma, o conjunto de dados disponível inicialmente foi dividido por cada um dos tipos de clientes (Residenciais e Empresariais) e *clusters* de forma a tornar possível treinar, validar/testar cada uma das redes neuronais.

Cada um dos conjuntos é composto por um vetor de dados de entrada, que correspondem às variáveis de *input* e uma de *output*, que corresponde ao valor da energia.

#### **4.3.2.1. Conjunto de Treino**

O conjunto de treino é utilizado para treinar a rede neuronal, de forma que esta aprenda os comportamentos necessários e fazer previsões para dados futuros [31]. Este conjunto de dados é representado por um *subset* dos dados originais e é formado pelos dados de consumo, agregado por *cluster* e por hora para cada tipo de cliente e estão compreendidos entre 2009-07-14 e 2010-07-13. Este período contempla um ano de dados, pelo que tem informação considerada relevante para o consumo de energia, como as várias estações do ano, meses e os vários dias da semana. Este é um conjunto vital para o processo de aprendizagem, pois recebe as variáveis de entrada e o valor de consumo e é através dele que a rede neuronal ajusta os pesos e *bias* tentando minimizar o erro da previsão.

#### **4.3.2.2. Conjunto de Validação/Teste**

O conjunto de validação/teste tem que ser necessariamente diferente do conjunto de treino, só desta forma é garantida a independência na validação do modelo [31]. Este conjunto de dados é representado pelo restante período disponível que está compreendido entre 2010-07-14 e 2010-12-31 e pretende testar o modelo, de forma a estimar a quantidade de energia consumida e posteriormente confrontar esta estimação com o valor efetivo de consumo e avaliar a *performance* do modelo.

### 4.3.3. Erros de previsão

Para se escolher um método de previsão ou outro é importante utilizar uma medida de erro com a finalidade de se encontrar aquele método que melhor atenda os resultados esperados. Existem na literatura diversas medidas de erro disponíveis de entre as quais podemos citar comparação gráfica, diagramas de dispersão, coeficiente de correlação, erro quadrático acumulado, raiz do erro médio quadrático (RMSE), erro percentual médio absoluto (MAPE), entre outros. Neste trabalho em específico foi usado o MAPE para avaliar as previsões. Este é calculado da seguinte forma:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - Ft}{A_t} \right| \quad (5)$$

Onde,  $A_t$  é o valor real e  $Ft$  é o valor resultante da previsão do processo.

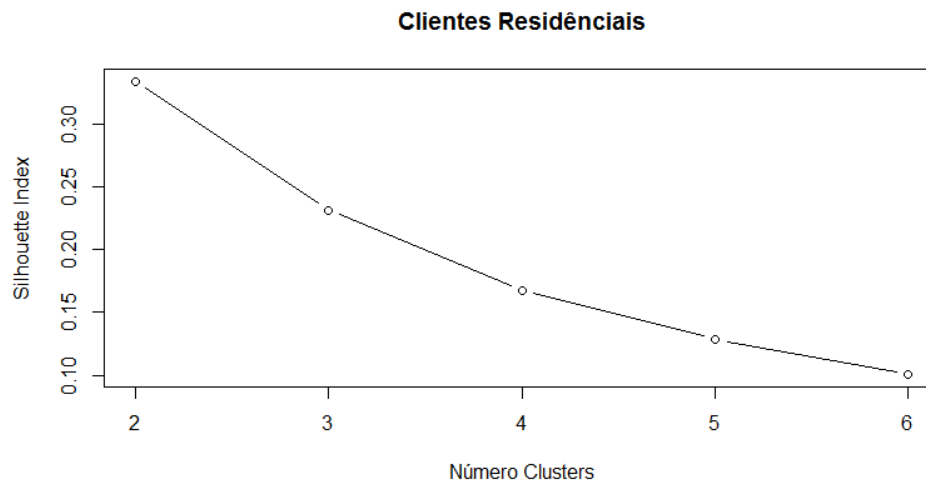
Esta métrica indica um valor percentual do erro e é de grande utilidade pois revela informação que varia entre 0 e 1 sendo possível uma análise qualitativa e comparativa entre vários sistemas ou configurações da rede de previsão.

# Capítulo 5

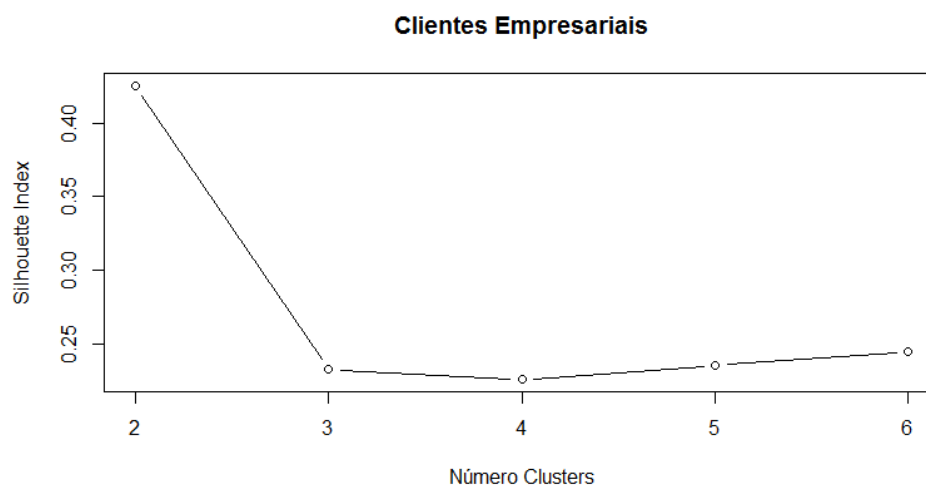
## 5. Análise de Resultados

### 5.1. Clustering

Através do método de cálculo do *Silhouette Index* [38] é possível identificar o número de *clusters* ideal para cada tipo de clientes. Através de uma análise gráfica torna-se simples perceber qual o número de *clusters* onde a métrica assume o valor mais elevado, representado o K ideal de acordo com os dados presentes no *dataset*.



**Figura 19 – Número de *Clusters* clientes Residenciais**



**Figura 20 – Número de *clusters* Clientes Empresariais**

Através da análise da Figura 19 – Número de *Clusters* clientes Residenciais e Figura 20 – Número de *clusters* Clientes Empresariais é possível selecionar o número de *clusters* que apresente o melhor *Silhouette Index*, representando a melhor divisão dos dados. Em ambos os casos foi possível perceber que a divisão em dois grupos apresenta o melhor resultado. Dessa forma, o conjunto de dados inicial passará a estar dividido em quatro partes: divisão de cada tipo de cliente (Residenciais e Empresariais) em dois perfis de consumo distintos.

### 5.1.1. Centros e associações de clientes a *clusters*

As imagens abaixo mostram o resultado do processo de segmentação de clientes em diferentes *clusters*. A Figura 21 – Centros dos *clusters* para clientes residenciais e Figura 22 - Centros dos *clusters* para clientes empresariais apresentam o centro de cada um dos *clusters* identificados para cada um dos tipos de clientes. Os centros podem ser vistos como o perfil típico de consumo médio para todos os clientes pertencentes a um determinado *cluster*. Também é possível observar que ao nível do perfil, isto é, comportamento das curvas, existem apenas algumas diferenças residuais mas é ao nível da amplitude que podem ser encontradas maiores diferenças justificando esta divisão.

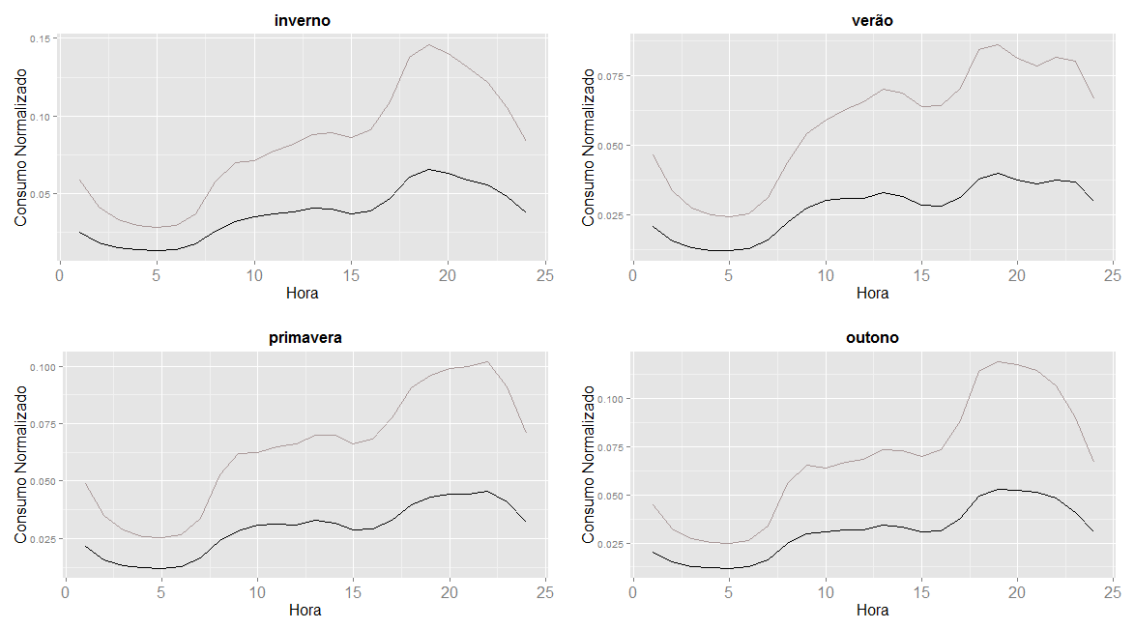
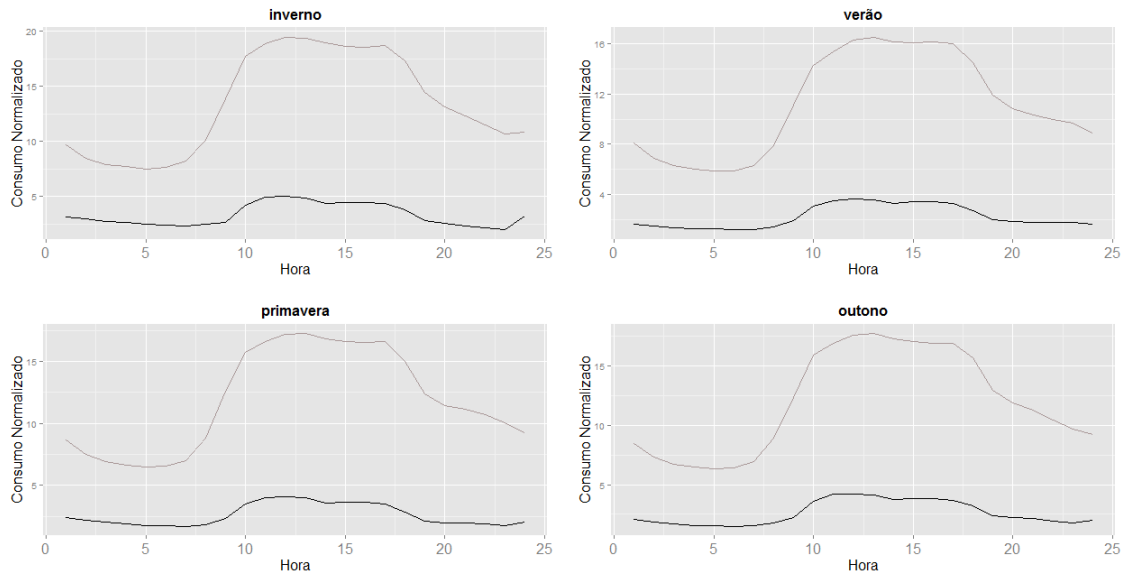


Figura 21 – Centros dos *clusters* para clientes residenciais



**Figura 22 - Centros dos *clusters* para clientes empresariais**

### 5.1.2. Análise de Clusters

De forma a perceber a distribuição de cliente por cada um dos *clusters* a população será estudada em função de algumas características recolhidas através dos inquéritos (ver anexos: 8.3, 8.4, 8.5) que podem ser encontradas na secção de anexos. Desta forma, podemos perceber se o processo de *clustering* utilizado, com base no perfil de consumo, está relacionado com outras características da população. Foram escolhidos alguns exemplos de características para cada um dos tipos de clientes, neste caso, para os clientes residenciais a análise recai sobre o seu tipo de habitação, rendimento e estado profissional, e para clientes empresariais a análise é feita através do número de empregados e volume de negócios apresentado.

### 5.1.3. Clientes Residenciais

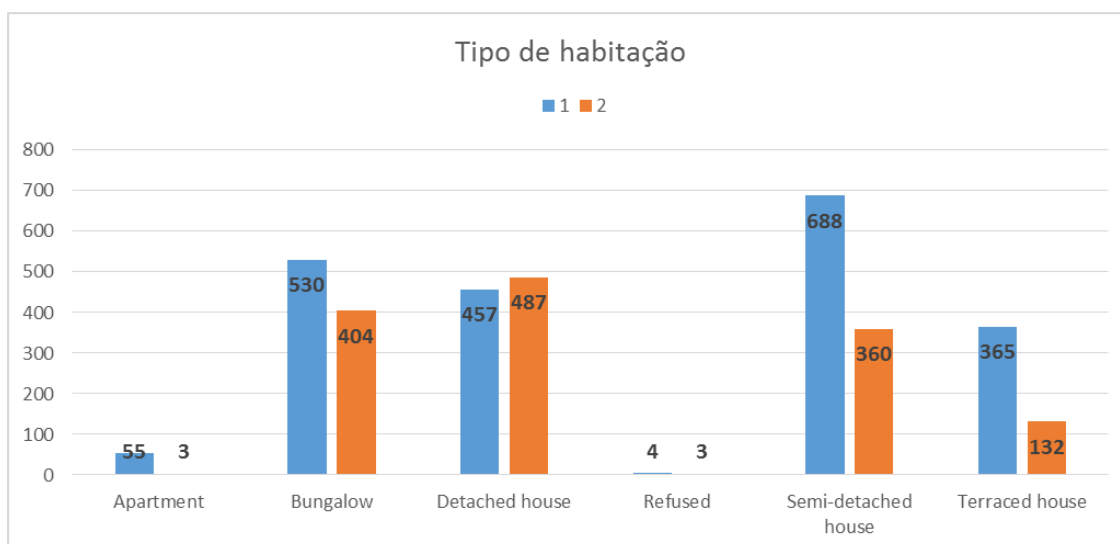
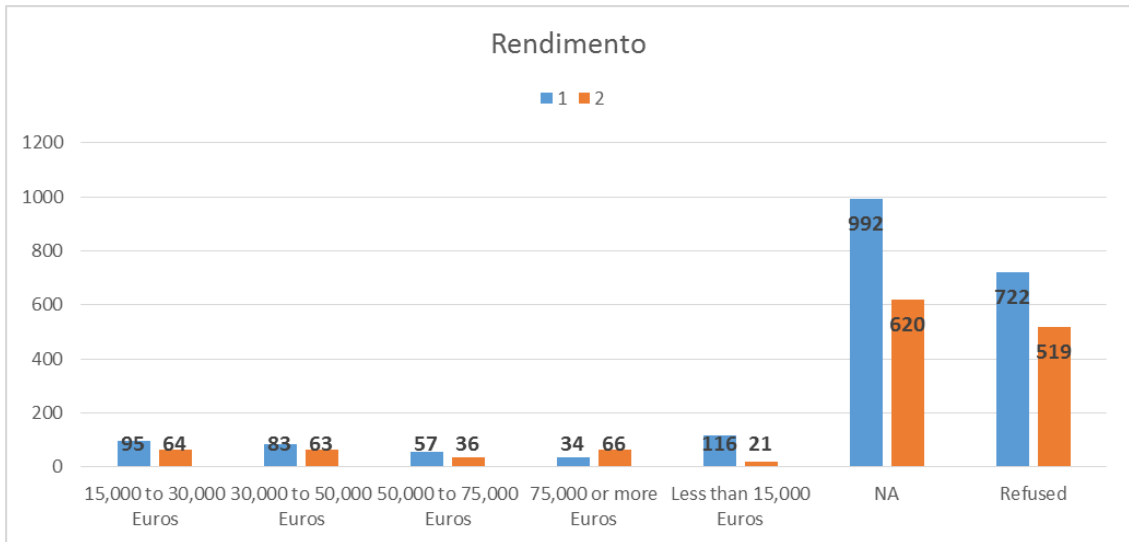


Figura 23 – Análise do número de cliente por tipo de habitação e *cluster*



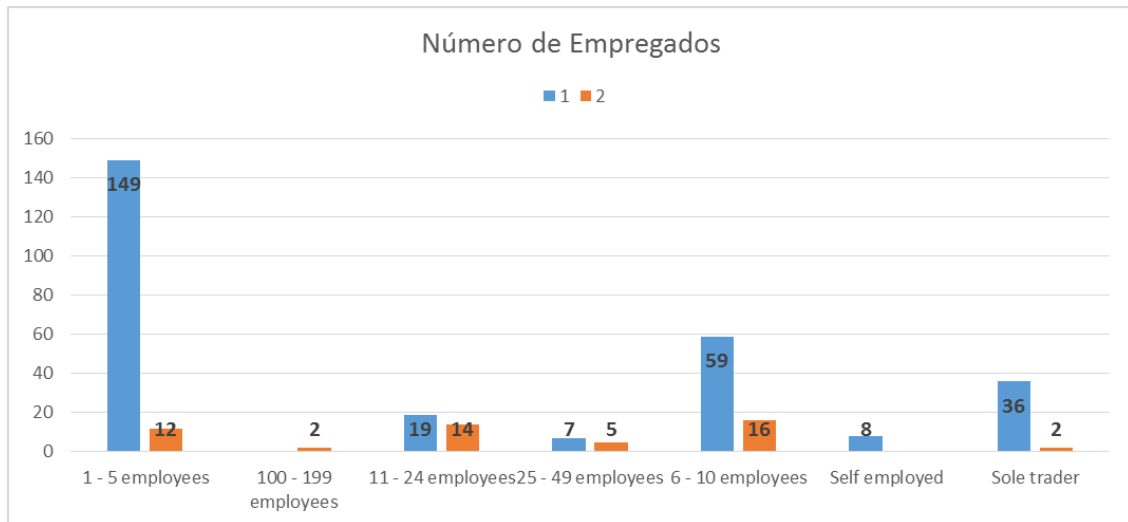


**Figura 24 – Análise do número de clientes por rendimento e cluster**

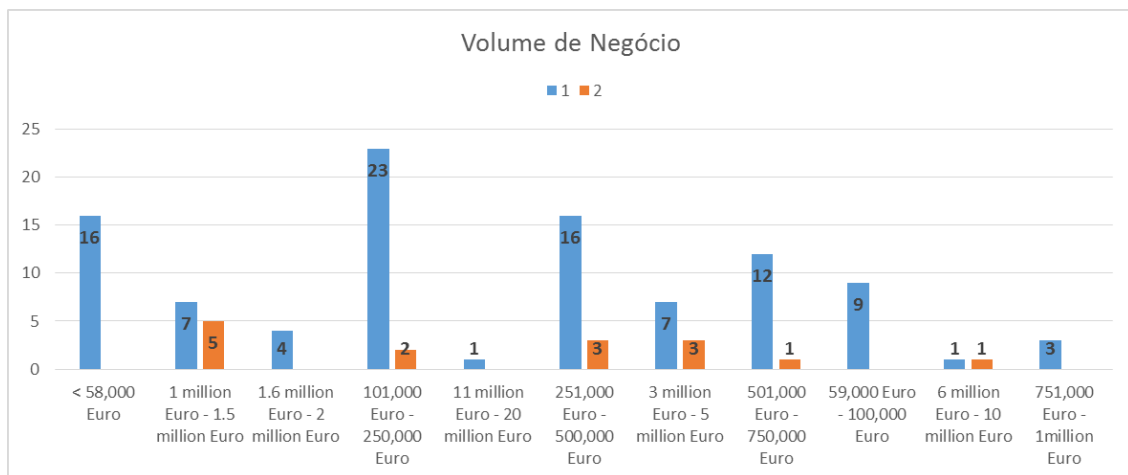


**Figura 25 - Análise do número de clientes por estado profissional e cluster**

### 5.1.4. Clientes Empresariais



**Figura 26 - Análise do número de clientes distribuídos por classes de número de empregados e *cluster***



**Figura 27 - Análise do número de clientes por volume de negócios e *cluster***

Através do conjunto de características demográficas estudadas não existe evidência que exista uma característica da população que esteja significativamente representada num só *cluster*. É possível observar que apesar de existirem características diferentes para cada cliente, estas encontram-se bem distribuídas e representadas em cada um dos *clusters* identificados. Esta análise permite justificar a utilização do processo de *cluster* em função do perfil de consumo típico de cada cliente excluindo as suas características inquiridas.

## 5.2. Treino da Rede Neuronal

Como referido no capítulo 4, a metodologia adotada pressupõe o treino de várias redes neuronais, de acordo com o tipo de cliente e com o perfil identificado no processo de segmentação. Especificamente foram treinadas quatro redes neuronais, duas por cada tipo de cliente, sendo o processo de treino semelhante entre elas.

Para cada uma das redes neuronais foram testadas várias configurações, nomeadamente em termos de número de neurónios da camada escondida e o parâmetro de *decay* (parâmetro utilizado para evitar o *overfitting* da rede, assumindo os valores de 0,01, 0,02 e 0,03). Para que fosse seleccionada a rede que apresentasse melhores resultados foi calculada para cada configuração de treino uma métrica que analisa o erro percentual médio da previsão (MAPE  $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - Ft}{A_t} \right|$  ( 5 )).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - Ft}{A_t} \right| \quad ( 5 )$$

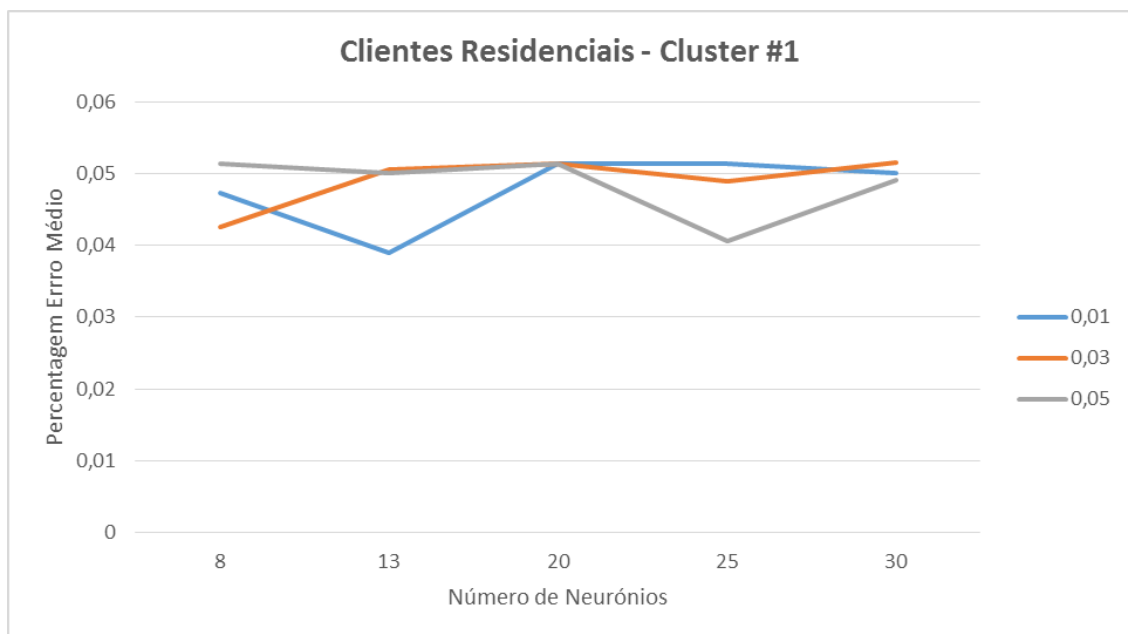
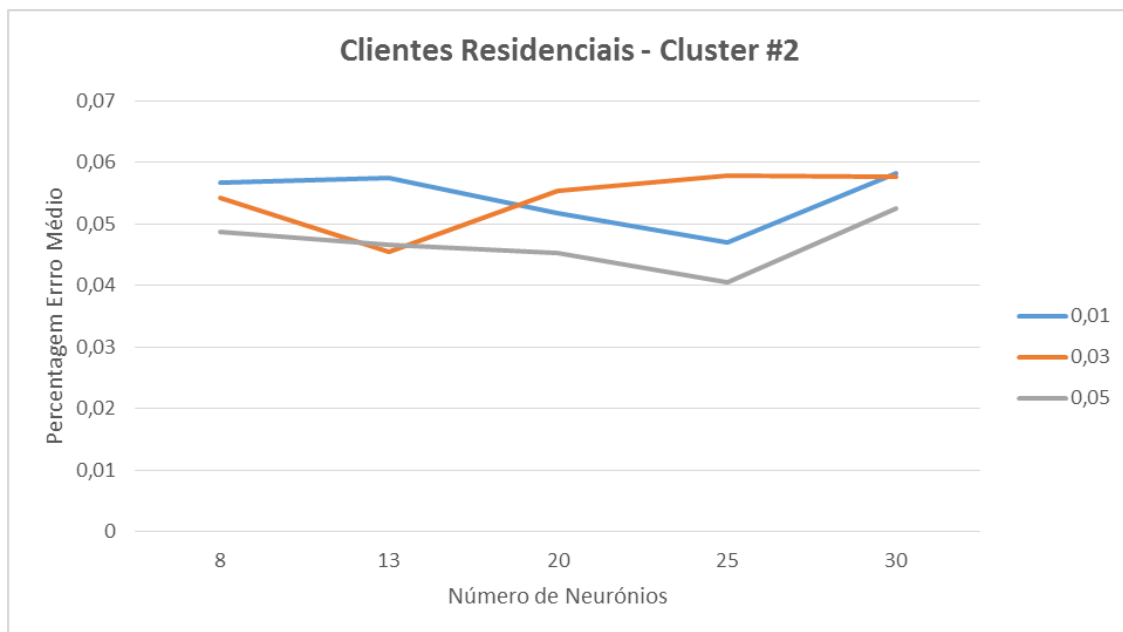
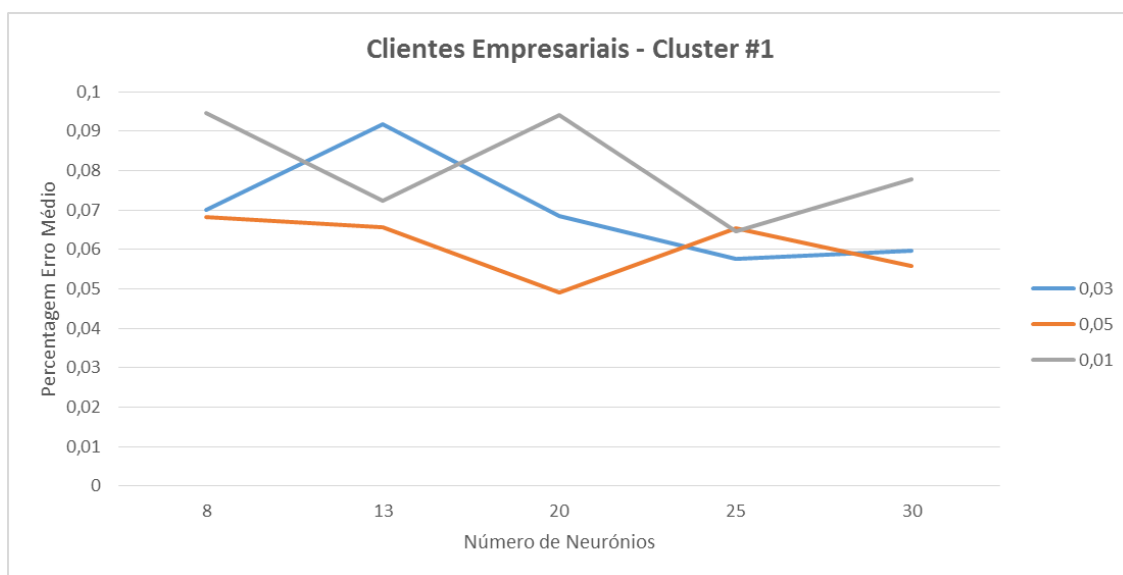


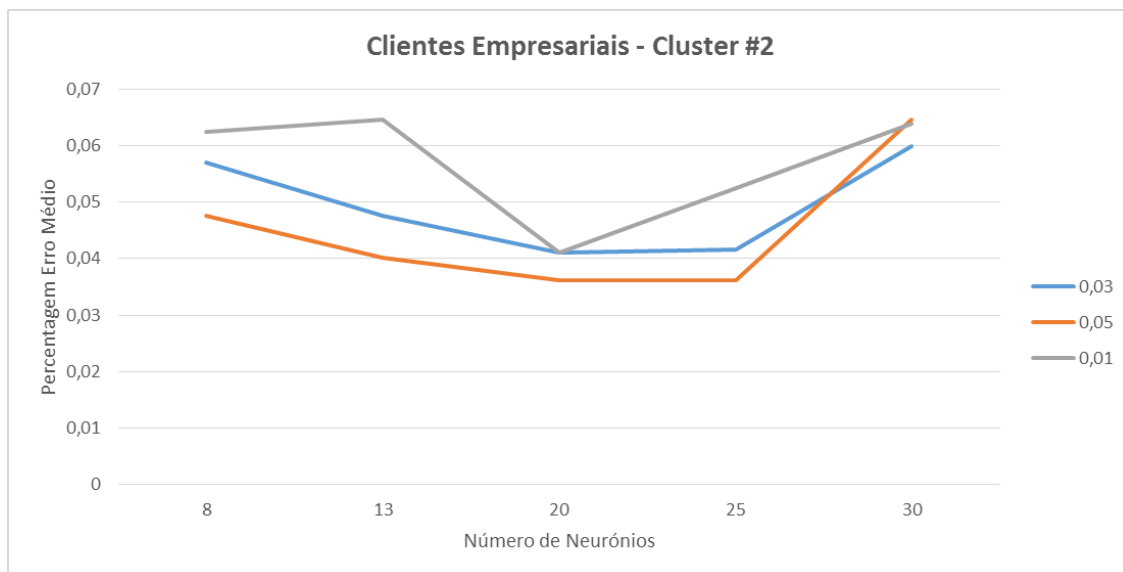
Figura 28 – Análise de variáveis de treino com base no MAPE para Clientes Residenciais *cluster #1*



**Figura 29 - Análise de variáveis de treino com base no MAPE para Clientes Residenciais cluster #2**



**Figura 30 - Análise de variáveis de treino com base no MAPE para Clientes Empresariais cluster #1**



**Figura 31 - Análise de variáveis de treino com base no MAPE para Clientes Empresariais *cluster #2***

Após analisados os resultados, e através do valor mínimo de percentagem de erro médio concluiu-se a seguinte configuração para cada uma das redes neuronais:

| Rede Neuronal              | <i>Decay</i> | Número de Neurónios | MAPE    |
|----------------------------|--------------|---------------------|---------|
| Clientes Residenciais – C1 | 0.01         | 13                  | ~ 3.9 % |
| Clientes Residenciais – C2 | 0.05         | 25                  | ~ 4.0 % |
| Clientes Empresariais – C1 | 0.05         | 20                  | ~ 4.9 % |
| Clientes Empresariais – C2 | 0.05         | 20                  | ~ 3.6 % |

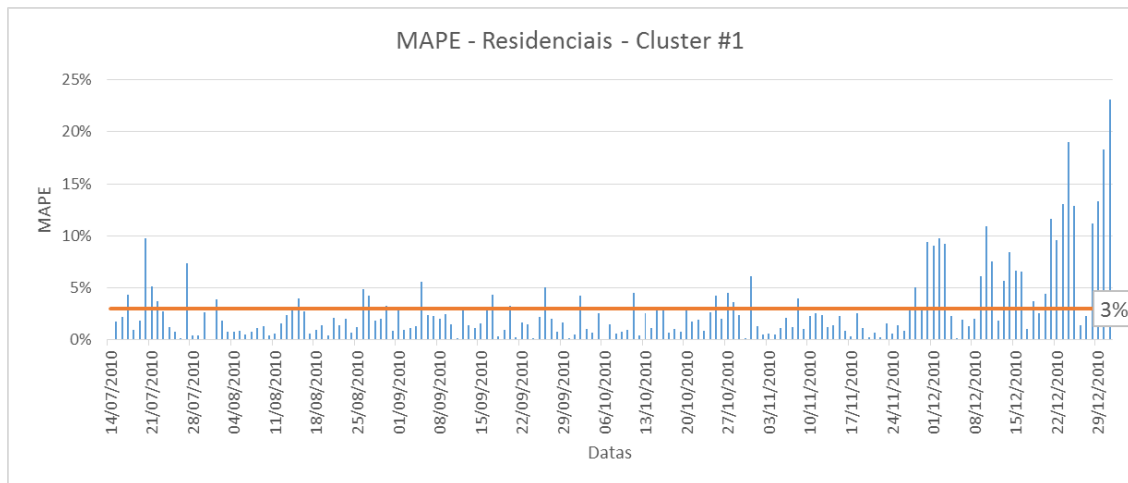
**Tabela 5 – Parâmetros obtidos no processo de treino**

Os parâmetros identificados como sendo os mais favoráveis foram utilizados para o treino da rede neuronal que será utilizada como modelo final para o sistema de previsão. Cada um dos modelos resultantes será então posto à prova, neste caso verificando a sua capacidade de generalização utilizando conjunto de dados de teste.

### 5.3. Resultados da Previsão

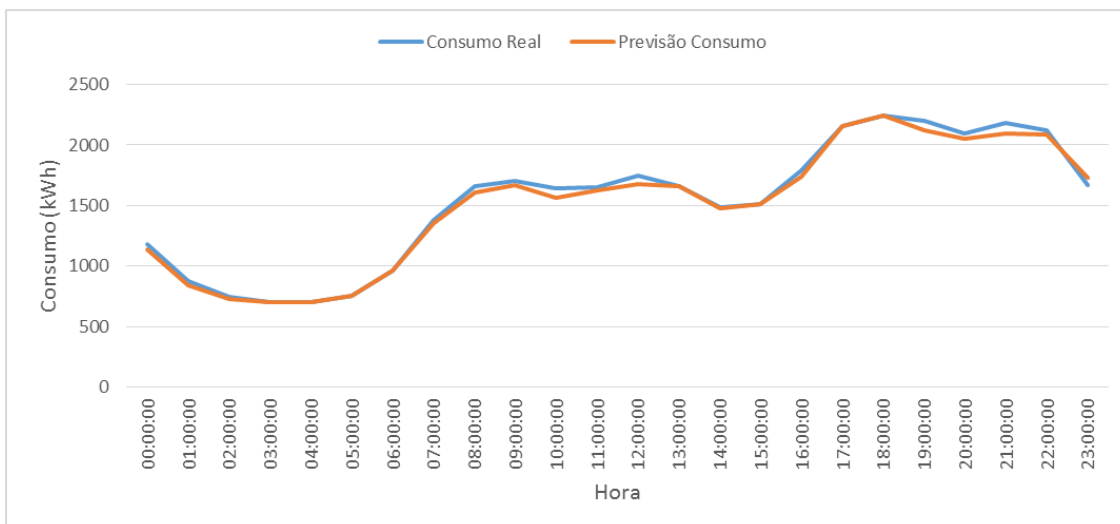
De forma a conseguir demonstrar o resultado de previsão, e dado que a previsão é bastante granular, isto é, valores de consumo de energia para cada uma das 24h de cada um dos dias, inicialmente será mostrado o erro de previsão médio (MAPE - ( 6 )) dia-a-dia para todo o período de testes, e depois serão demonstrados em detalhe alguns dias, escolhidos de forma arbitrária, com o objetivo de perceber o desvio entre o valor real de consumo e o valor previsto através do processo de previsão.

#### 5.3.1. Clientes Residenciais



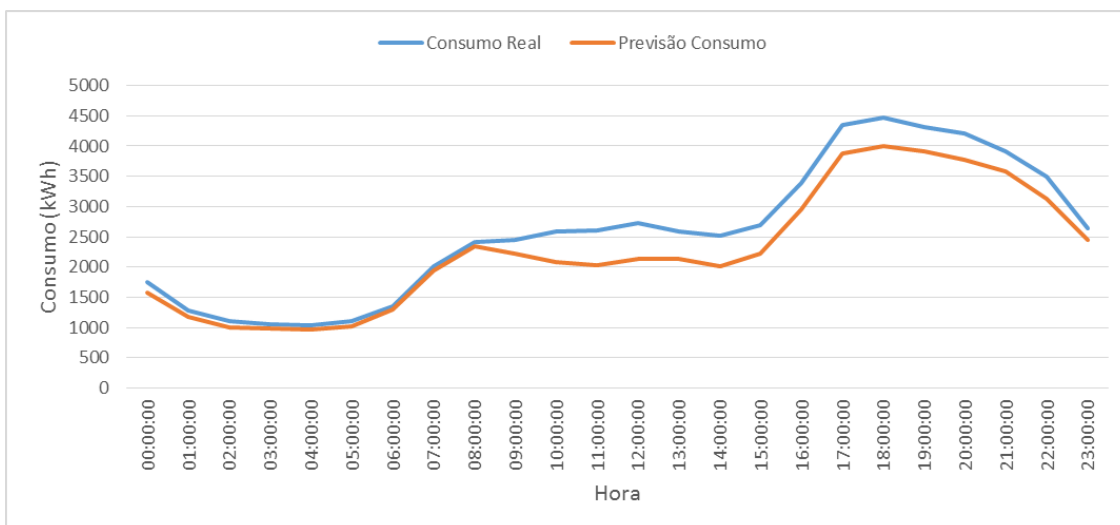
**Figura 32 - Erro Médio de Previsão para Cluster #1 de Clientes Residenciais**

Analisando a Figura 32 - Erro Médio de Previsão para Cluster #1 de Clientes Residenciais, que mostra uma visão geral dos resultados de previsão, é possível observar que globalmente os resultados são satisfatórios, resultando em um valor médio de erro para o período considerado de aproximadamente 3%. Ainda assim, este é um valor global e é necessário avaliar a previsão em mais detalhe e alinhada com os objetivos iniciais (previsão de consumo hora a hora para as próximas 24h). Essa análise poderá ser vista nas imagens que se seguem mostrando a comparação entre os valores reais de consumo e os valores previstos pelo processo de previsão. Os dias foram selecionados arbitrariamente, sem qualquer critério.



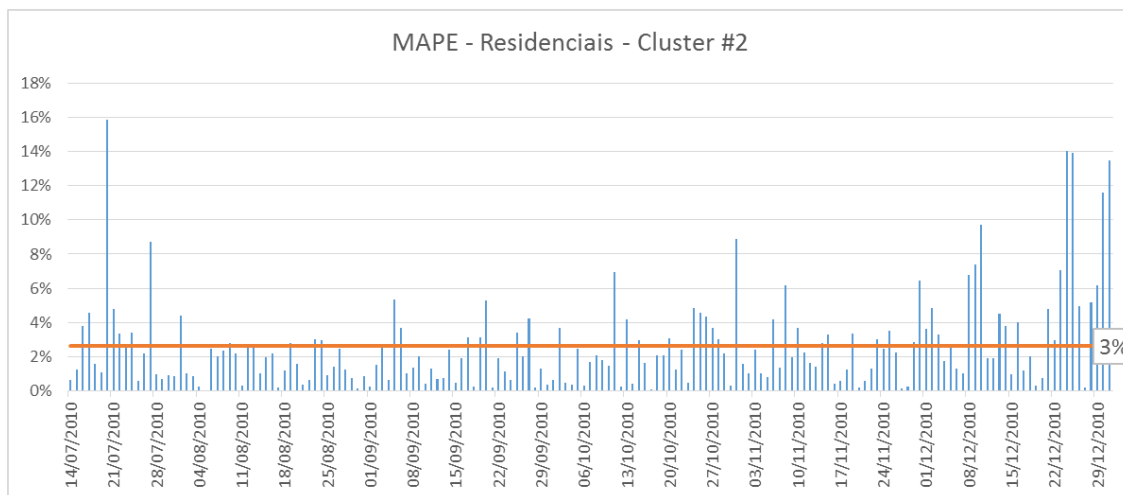
**Figura 33 – Clientes Residenciais Cluster # 1 dia 15-07-2010**

Na Figura acima, podemos observar o resultado de previsão para o dia 15/07/2010 para o *cluster #1* dos clientes Residenciais. Ao observar as duas curvas, real e prevista, concluímos que os valores são bastante próximos o que revela uma precisa previsão do consumo para este dia.



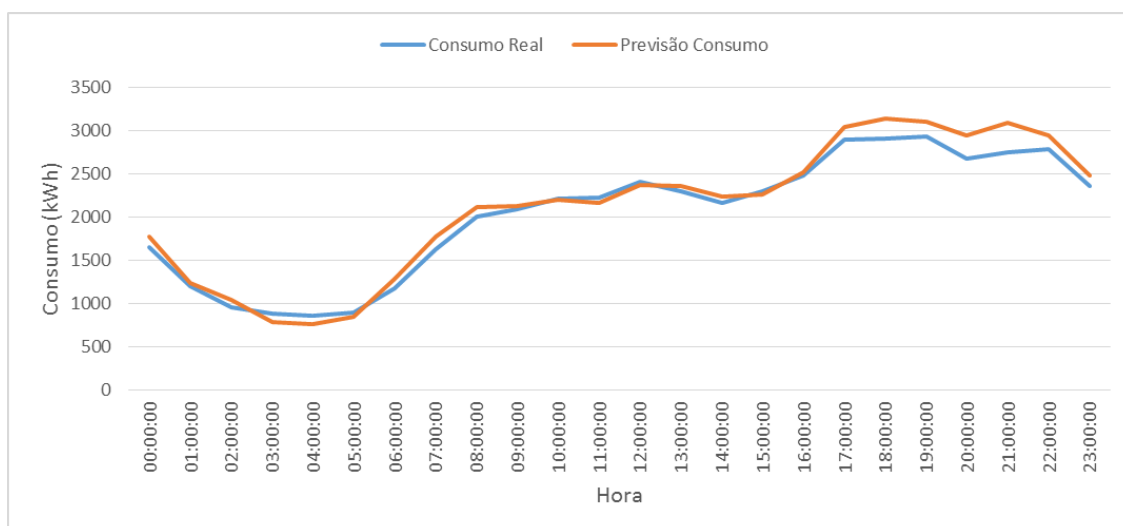
**Figura 34 – Clientes Residenciais Cluster #1 dia 21-12-2010**

Na Figura 34 – Clientes Residenciais Cluster #1 dia 21-12-2010, podemos observar o resultado de previsão para o dia 21/12/2010 para o *cluster #1* dos clientes Residenciais. Ao observar as duas curvas, real e prevista, verifica-se uma semelhança bastante vincada em termos do seu perfil, mas existem algumas diferenças em termos de amplitude. Ainda assim, a previsão parece bastante ajustada face ao valor real de consumo.



**Figura 35 - Erro Médio de Previsão para Cluster #2 de Clientes Residenciais**

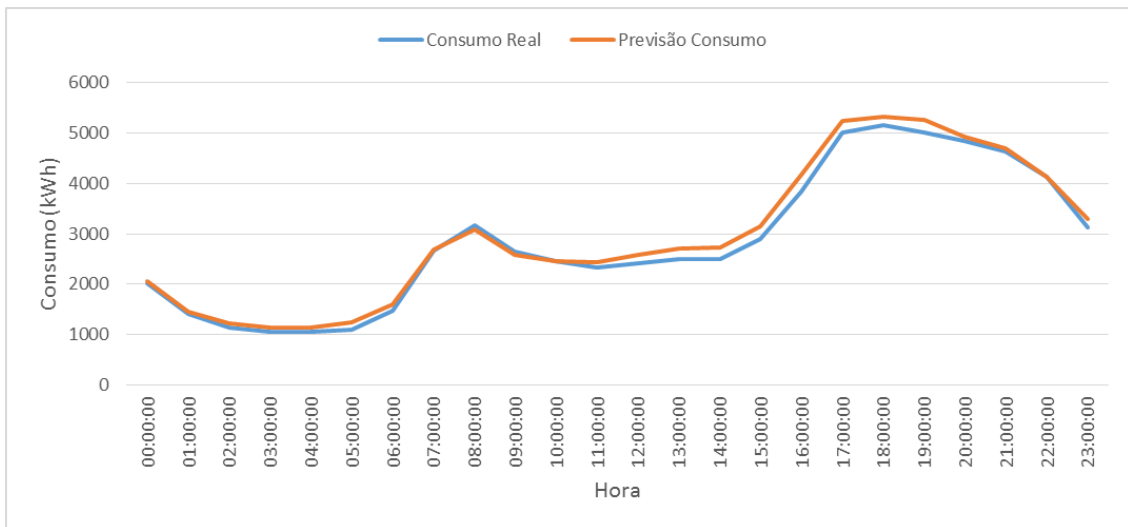
Da mesma forma que anteriormente demonstrado a Figura 35 - Erro Médio de Previsão para Cluster #2 de Clientes Residenciais, representa a visão geral dos resultados de previsão para o *cluster #2*, é possível observar que neste caso os resultados também são satisfatórios, ainda que existam maiores oscilações em termos de MAPE ao longo do tempo.



**Figura 36 – Clientes Residenciais cluster #2 dia 16/10/2010**

Na Figura 36 – Clientes Residenciais cluster #2 dia 16/10/2010, podemos observar o resultado de previsão para o dia 16/10/2010 para o *cluster #2* dos clientes Residenciais. Ao observar as duas curvas, real e prevista, para este dia podemos concluir que a previsão também apresenta resultados satisfatórios.

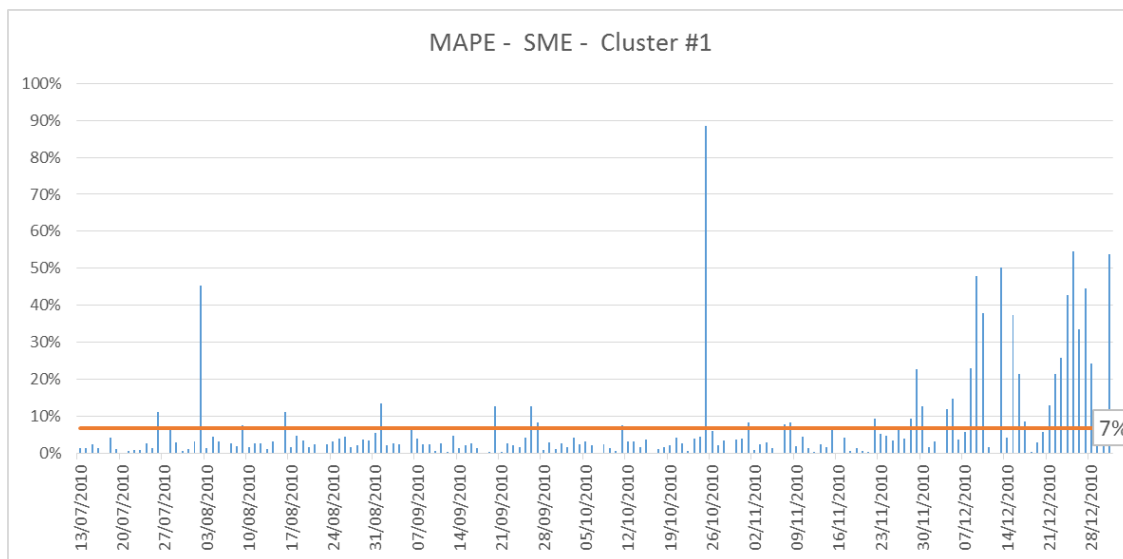




**Figura 37 – Clientes Residenciais Cluster #2 dia 16/12/2010**

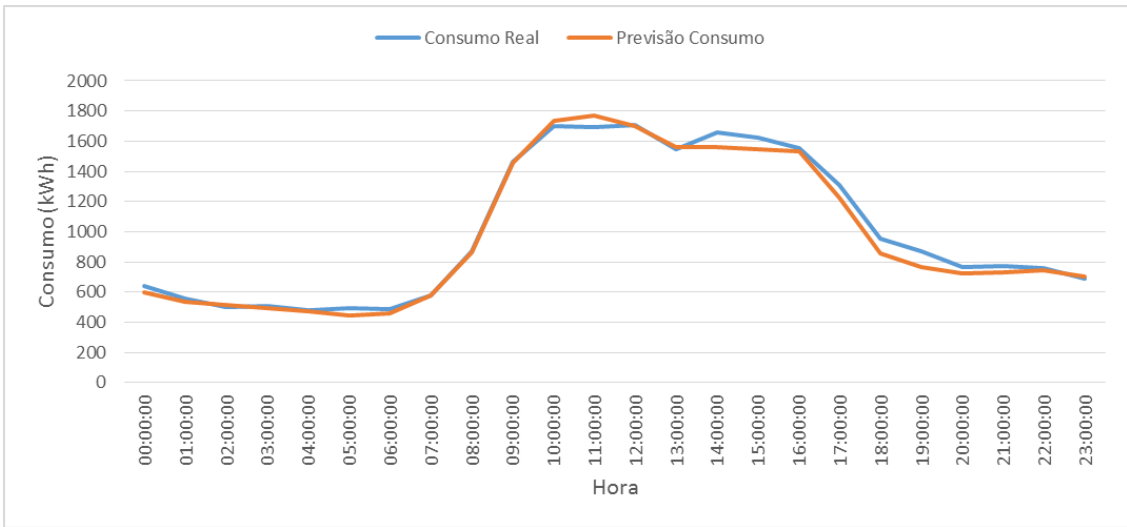
Na Figura 37, podemos observar o resultado de previsão para o dia 16/12/2010 para o *cluster #2* dos clientes Residenciais. Ao observar as duas curvas, real e prevista, concluímos que os valores são bastante próximos o que revela também uma precisa previsão do consumo para este dia.

### 5.3.1. Clientes Empresariais



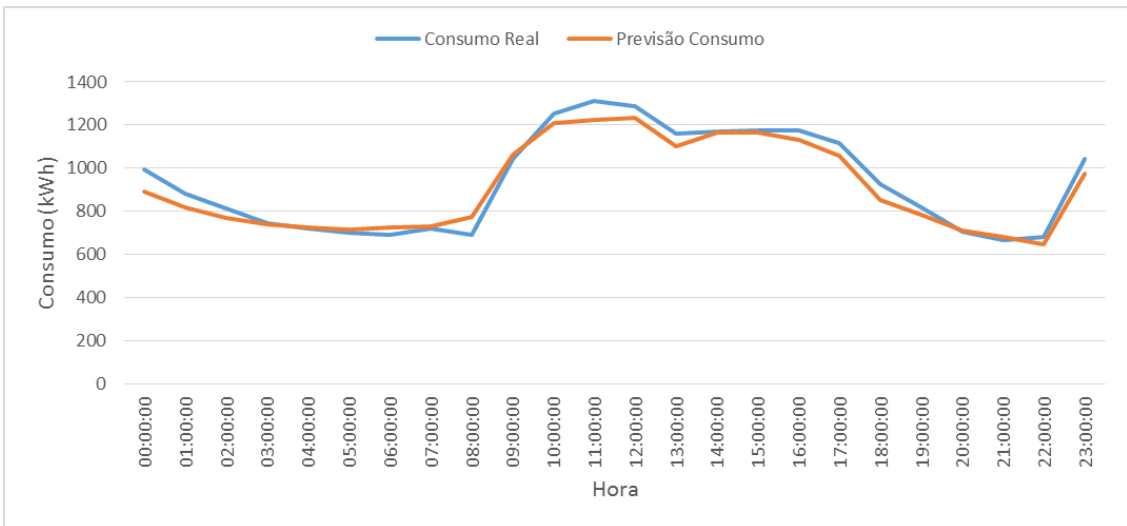
**Figura 38 - Erro Médio de Previsão para *Cluster #1* de Clientes Empresariais**

Utilizando a mesma abordagem de visualização de resultados, o resultado para o *cluster #1* dos Clientes Empresariais pode ser analisado na Figura 38 - Erro Médio de Previsão para *Cluster #1* de Clientes Empresariais. Neste caso, o valor de erro global aumentou substancialmente, passando para aproximadamente 7%. Este valor é justificado por grandes oscilações, existindo dias específicos cuja previsão não foi de todo satisfatória. Mais adiante serão analisados exemplos de dias cujo erro de previsão é maior de forma a tentar perceber a justificação para estes resultados.



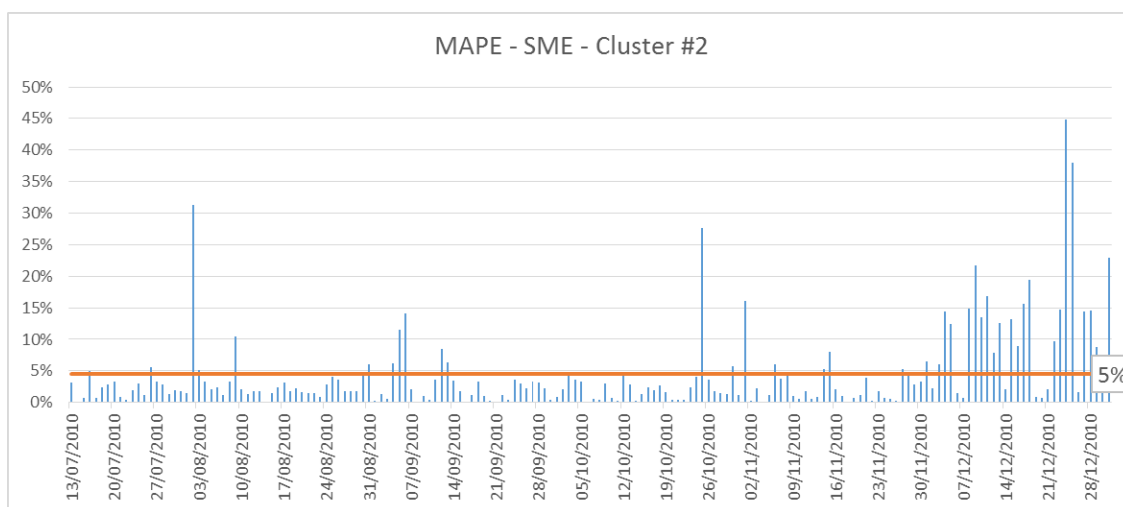
**Figura 39 – Clientes Empresariais Cluster #1 dia 15/07/2010**

Na Figura 39, podemos observar o resultado de previsão para o dia 15/07/2010 para o *cluster #1* dos clientes Empresariais. Ao observar as duas curvas, real e prevista, concluímos mais uma vez que os valores são bastante próximos, estando a previsão bastante afinada face ao consumo real.



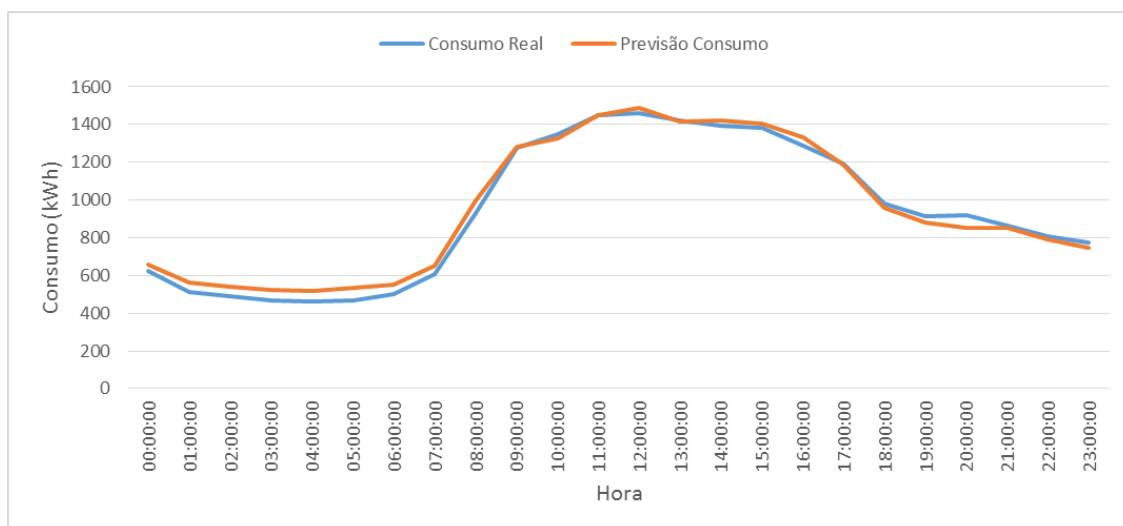
**Figura 40 – Clientes Empresariais Cluster #1 dia 13/11/2010**

Na Figura 40, podemos observar o resultado de previsão para o dia 13/11/2010 para o *cluster #1* dos clientes Empresariais. Ao observar as duas curvas, real e prevista, concluímos mais uma vez que os valores são bastante próximos, estando a previsão bastante afinada face ao consumo real.



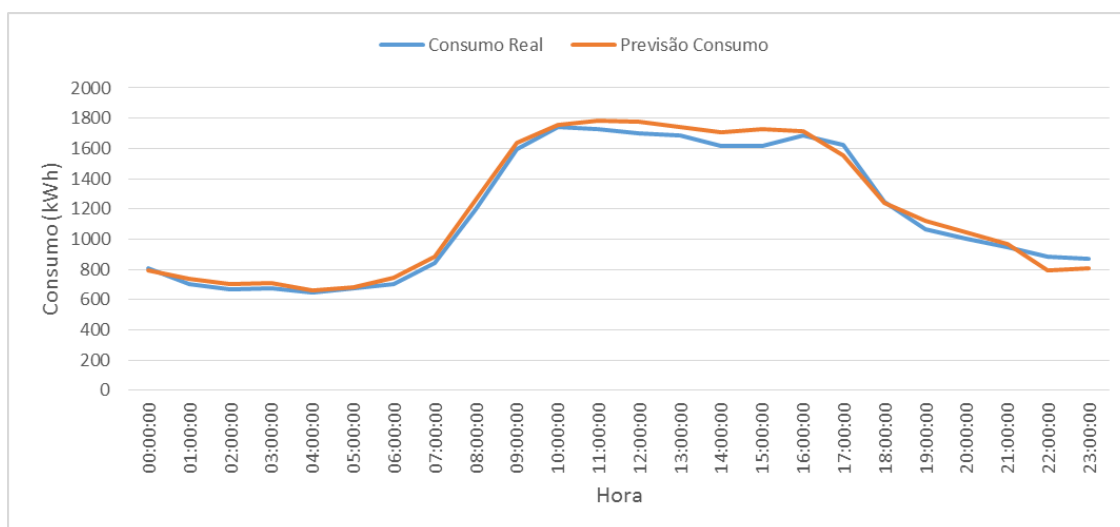
**Figura 41 - Erro Médio de Previsão para Cluster #2 de Clientes Empresariais**

Por fim, o resultado acima mostra para a precisão de previsão para o *cluster #2* dos Clientes Empresariais. Neste caso, o valor de erro global médio aproxima-se de 5% para todo o universo de dados. Mais uma vez, para o caso dos clientes Empresariais as oscilações são mais acentuadas. Não obstante, a nível geral a metodologia utilizada conseguiu resultados satisfatórios.



**Figura 42 – Clientes Empresariais cluster #2 dia 27/08/2010**

Na Figura 42 – Clientes Empresariais *cluster #2* dia 27/08/2010, podemos observar o resultado de previsão para o dia 27/08/2010 para o *cluster #2* dos clientes Empresariais. O resultado observado mostra-se bastante preciso ao longo de todas as 24h do dia.



**Figura 43 – Clientes Empresariais *cluster* #2 dia 14/12/2010**

Na Figura 43 – Clientes Empresariais *cluster* #2 dia 14/12/2010 podemos observar o resultado de previsão para o dia 14/12/2010 para o *cluster* #2 dos clientes Empresariais. Neste caso, o resultado também é bastante positivo observando-se uma grande proximidade entre os valores de consumo real e a previsão.

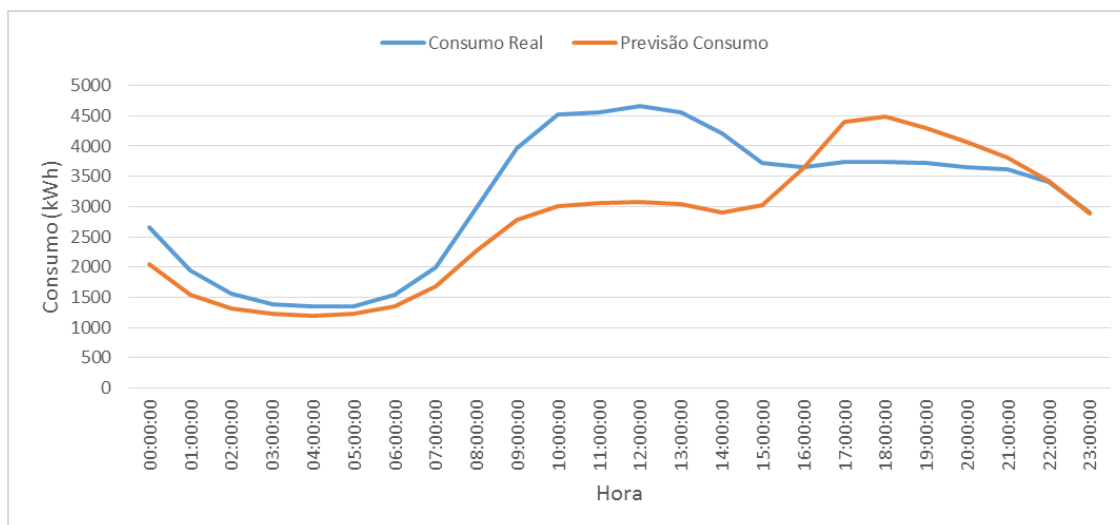
### 5.3.2. Análise de Grandes Desvios

De acordo com a análise da Figura 32 - Erro Médio de Previsão para Cluster #1 de Clientes Residenciais, Figura 35 - Erro Médio de Previsão para Cluster #2 de Clientes Residenciais, Figura 38 - Erro Médio de Previsão para *Cluster* #1 de Clientes Empresariais e Figura 41 - Erro Médio de Previsão para *Cluster* #2 de Clientes Empresariais, onde estão representados os resultados globais de previsão para cada um dos *clusters* para cada tipo de cliente, podemos observar alguns dias cuja previsão não se mostrou acertada. De forma a perceber esse comportamento vai ser escolhido um dia, cujo resultado não seja satisfatório, para cada tipo de cliente (neste caso será escolhido o *cluster* #1 de cada tipo de cliente a título de exemplo). Serão utilizadas algumas comparações com dados históricos de forma a justificar e perceber tais erros.

| Análise | Data       | Tipo de Cliente | Cluster | MAPE (24h) |
|---------|------------|-----------------|---------|------------|
| 1       | 25/12/2010 | Residencial     | #1      | 17,86%     |
| 2       | 25/10/2010 | Empresarial     | #1      | 93,12%     |

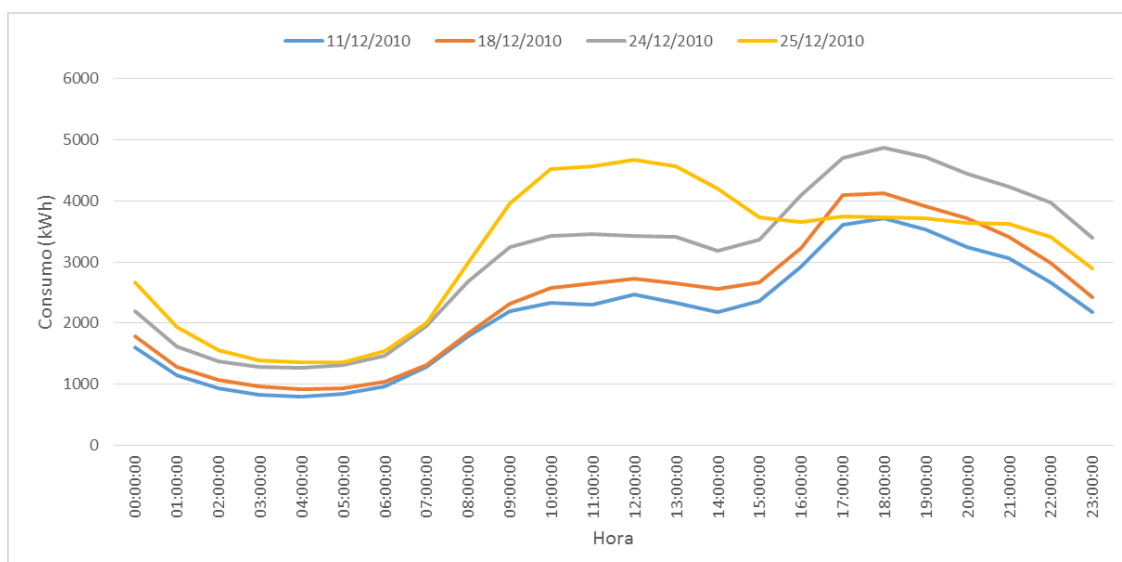
**Tabela 6 – Análise de Grandes Desvios (Consumo Real VS Consumo Previsto)**

### 5.3.2.1. Análise 1



**Figura 44 – Consumo Real VS Previsão para Clientes Residenciais, Cluster #1 dia 25-12-2010**

Tal como se pode observar na Figura 44 – Consumo Real VS Previsão para Clientes Residenciais, Cluster #1 dia 25-12-2010 existem grandes diferenças entre o consumo real e o consumo que foi previsto pela rede neuronal. De forma a perceber este comportamento, a Figura 45 mostra a comparação entre o dia 25/12/2010 e os dias anteriores que têm impacto no consumo. Isto é, foram selecionados o dia anterior e o mesmo dia da semana da semana anterior (variáveis que são utilizadas para o treino da rede neuronal e que têm impacto no consumo).



**Figura 45 – Relação entre dia com erro de previsão e dias anteriores – Clientes Residenciais, Cluster #1**

Ao observar a Figura 45 – Relação entre dia com erro de previsão e dias anteriores – Clientes Residenciais, *Cluster #1* é possível observar que a dinâmica do consumo do dia 25/12/2010 tem algumas diferenças face ao dia anterior e dias comuns das duas semanas anteriores (apresentando estas dinâmicas semelhantes ao nível do perfil).

Esta alteração poderá justificar-se, pelo facto de 25/12/2010 ser um dia especial. Este dia é feriado e a dinâmica destes tipos de dias é necessariamente diferente. Neste caso, erradamente a rede neuronal não aprendeu as dinâmicas dos feriados dado que estas datas não foram classificadas como tal.



### 5.3.2.1. Análise 2

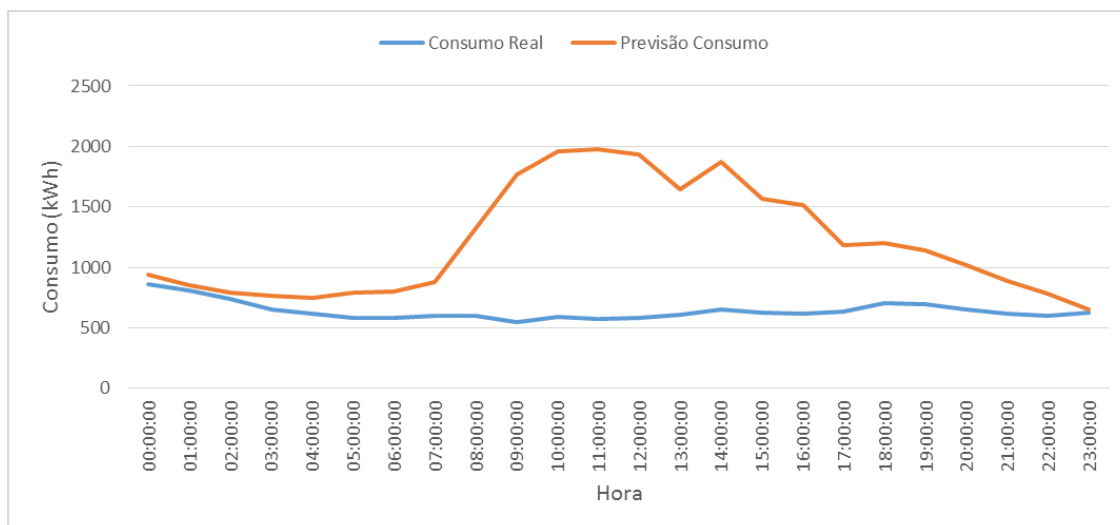


Figura 46 - Consumo Real VS Previsão para Clientes Empresariais, Cluster #1 dia 25-10-2010

Da mesma forma que no caso anterior, o consumo previsto para dia 25/10/2010 para o cluster #1 dos Clientes Empresariais apresenta grandes diferenças face ao valor real. Utilizando a mesma abordagem que no caso anterior, a imagem Figura 47 - Relação entre dia com erro de previsão e dias anteriores – Clientes Empresariais, Cluster #1 apresenta uma análise dos dias anteriores e dias homólogos das semanas anteriores.

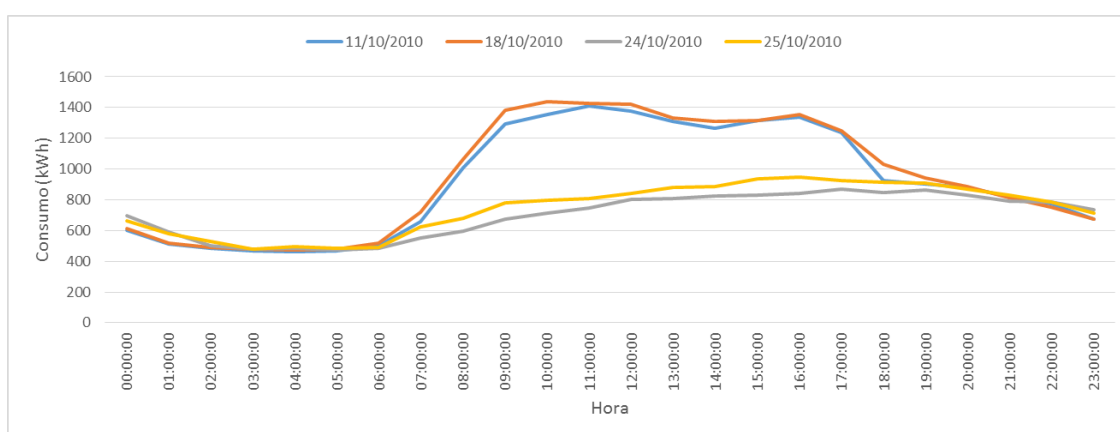


Figura 47 - Relação entre dia com erro de previsão e dias anteriores – Clientes Empresariais, Cluster #1

Mais uma vez esta diferença poderá ser justificada por alguma anomalia no consumo registado, já que neste dia as diferenças são bastante acentuadas, ou pela mesma justificação verificada no caso anterior. Foi verificado que o dia 25/10/2010 é também feriado na Irlanda [40]. Mais uma vez, e dado que os dados não foram enriquecidos com este tipo de características, este tipo de dinâmicas não foi aprendida através do processo de treino da rede neural e dessa forma não está capacitada a generalizar este tipo de características.

# Capítulo 6

## 6. Conclusões e Trabalho Futuro

### 6.1. Conclusões

A principal motivação do presente trabalho foi a construção de uma metodologia capaz de prever consumos de energia no curto prazo utilizando dados reais de consumo de clientes de vários tipos.

Foi realizado um estudo teórico acerca de métodos e metodologias utilizadas em sistemas de previsão, tendo sido adotada a técnica de aprendizagem automática: Rede Neuronal. Esta escolha justifica-se por ser uma técnica recorrentemente utilizada em estudos semelhantes. Outro fator com relevo na escolha deste tipo de metodologia, recai na sua capacidade de relacionar variáveis que à partida poderão, à primeira vista, não estar relacionadas e pela sua capacidade de adaptação às suas dinâmicas conseguindo ganhar a capacidade de generalização.

No que diz respeito à parametrização da rede e com base no pacote de software utilizado (Anexo 8.2), foi utilizada uma rede neuronal com uma camada escondida e testadas várias configurações de neurónios procurando perceber qual a configuração que desempenhava uma previsão mais adequada.

Após alguma análise de literatura e através de análises gráficas foi possível identificar as variáveis relevantes utilizadas para o treino da rede. A temperatura prevista referente ao dia de previsão; dia da semana referente ao dia de previsão; hora referente à hora do dia da previsão; o consumo verificado na semana anterior à mesma hora e o consumo verificado no dia anterior à mesma hora, foram as variáveis identificadas como sendo mais relevantes.

A metodologia utilizada como caso de estudo apresentou-se interessante na medida em que conseguiu responder aos objetivos propostos e permitiu testar e avaliar novas potencialidades. Utilizar consumos de clientes divididos por tipo de cliente e por perfil de consumo fez sentido na medida em que estamos a captar as características fundamentais e específicas de cada tipo de cliente excluindo a potencial entropia existente quando misturados vários tipos de clientes.

Desta forma, e recorrendo à metodologia de *Clustering + ANN* foi também possível ganhar mais conhecimento específico nesta área de *data mining* e análises preditivas potenciando a criação de novas oportunidades a nível empresarial.

Avaliando os resultados finais referentes à previsão de consumo de energia, estes demonstraram-se bastante interessantes visto a percentagem de erro médio (MAPE) rondar valores significativamente baixos. Isto permite-nos ter um grau de confiança elevado na capacidade de generalização do modelo. Os resultados deste trabalho, quando divididos por tipo de cliente e perfil de consumo (cada *Cluster*) apresentam resultados próximos dos identificados acima rondando os 3% de erro percentual médio para o caso dos clientes residenciais e 5% e 7 % para os clientes empresariais (cluster #1 e cluster #2, respetivamente).

## 6.2. Trabalho Futuro

Tal como foi identificado na metodologia CRISP-DM, o processo de *data mining*, e neste caso específico associação aos algoritmos preditivos não é um processo estanque e exige que seja revisitado ao longo do tempo. No contexto dos algoritmos de previsão é extremamente difícil alcançar a previsão ideal, e o objetivo passa por tentar diminuir as diferenças existentes entre o valor real e o valor previsto.

Como trabalho futuro, e de forma a tentar aproximar o valor de previsão face ao real, um dos próximos passos, passaria por estudar em maior detalhe os dias em que os desvios do valor consumido e da previsão foi maior, com o objetivo perceber o que poderia estar a causar tal impacto. Neste caso, deveriam ser utilizadas informações do quotidiano para enriquecer os dados. A introdução dos tipos de dias, nomeadamente dos feriados globais e específicos de cada país, apresentou-se como um fator importante a ter em conta que não foi considerado neste trabalho. Outro ponto importante passa pelo processo de “re-treinar” as redes neuronais ao longo do tempo. Estas deviam ser alimentadas com novos dados de forma a criar novo conhecimento existente e continuar a responder com precisão às previsões de consumo.

Outro fator relevante como trabalho futuro passaria por utilizar o conhecimento obtido através dos questionários efetuados aos clientes para realizar outro tipo de análises. Como referi anteriormente no capítulo de análise exploratória, existem estudos que pretendem segmentar clientes em perfis de consumo semelhantes com base nas suas características socioeconómicas. De acordo com a metodologia discutida no presente trabalho, a capacidade de associar indivíduos a perfis de consumo sem conhecer os seus hábitos de consumo, poderia ser útil para classificar novos clientes e atribuí-los a um dos segmentos já identificados.

Ainda como estratégia futura, este tipo de previsão de curto prazo, poderia ser utilizada como *inputs* de outros algoritmos e mecanismos que pretendem melhorar a qualidade da rede elétrica ou a qualidade do serviço prestado. Um conceito interessante e que está a surgir com o advento das *smart grids* passa pela capacidade de construir planos tarifários dinâmicos e ajustáveis ao longo do tempo utilizando os perfis de consumo e previsão de energia consumida na rede tentando gerir, com base nas tarifas, a carga em certos períodos horários. Do ponto de vista comercial seria importante perceber onde existirão horas de pico no dia seguinte e tentar motivar clientes a consumirem em outros horários do dia, ajustando o preço da tarifa dinamicamente.



# Bibliografia

---

## 7. Referências

- [1] H. Farhangi, “The path of the smart grid,” *IEEE Power and Energy Magazine*, vol. 8, pp. 18-28, 2010.
- [2] Office of Electricity Delivery & Energy Reliability, “the SMART GRID: an introduction,” [Online]. Available: <http://energy.gov/oe/downloads/smart-grid-introduction-0>. [Acedido em 08 12 2014].
- [3] J. S. (ed), *Smart Building Systems for Architects, Owners, and Builders* Elsevier, 2010.
- [4] EEI-AEIC-UTC, “Smart Meters and Smart Meter Systems: A Metering Industry Perspective,” Edison Electric Institute (EEI), 2011.
- [5] H. Sui, Y. Sun e W.-J. Lee, “A demand side management model based on advanced metering infrastructure,” em *Electric Utility Deregulation and Restructuring and Power Technologies ({DRPT}), 2011 4th International Conference on*, 2011.
- [6] K. a. K. Metaxiotis, A. a. Askounis e J. D. and Psarras, “Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher,” *Energy Conversion and Management*, vol. 44, pp. 1525-1534, 2003.
- [7] K. Y. Lee, Y. T. d Cha e J. H. Park, “Short-term load forecasting using an artificial neural network,” *Power Systems, {IEEE} Transactions on*, vol. 7, pp. 124-132, 1992.
- [8] H. K. Alfares e M. Nazeeruddin, “Electric load forecasting: Literature survey and classification of methods,” *International Journal of Systems Science*, vol. 33, pp. 23-34, 2002.

- [9] C. Xia, J. Wang e K. {McMenemy}, “Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks,” *International Journal of Electrical Power & Energy Systems*, vol. 32, pp. 743-750, 2010.
- [10] D. J. Pedregal e J. R. Trapero, “Mid-term hourly electricity forecasting based on a multi-rate approach,” *Energy Conversion and Management*, vol. 51, pp. 105-111, 2010.
- [11] T. Yalcinoz e U. Eminoglu, “Short term and medium term power distribution load forecasting by neural networks,” *Energy Conversion and Management*, vol. 46, pp. 1393 - 1405, 2005.
- [12] M. Ghiassi, D. K. Zimbra e H. Saidane, “Medium term system load forecasting with a dynamic artificial neural network model,” *Electric Power Systems Research*, vol. 5, pp. 302 - 316, 2006.
- [13] N. Amjady e F. Keynia, “Mid-term load forecasting of power systems by a new prediction method,” *Energy Conversion and Management*, vol. 48, pp. 2678 - 2687, 2008.
- [14] M. E. a. B. S. D. Dyson, M. D. Tabone e D. S. Callaway, “Using smart meter data to estimate demand response potential, with application to solar energy integration,” *Energy Policy*, vol. 73, pp. 607-619, 2014.
- [15] K. Y. Lee, Y. T. Cha e J. H. Park, “Short-term load forecasting using an artificial neural network,” *Power Systems, {IEEE} Transactions on*, vol. 7, pp. 124-132, 1992.
- [16] R. D. S. M. Rodrigues, *Previsão dos consumos na {RNT} considerando a produção distribuída—Desenvolvimento de um procedimento de apoio à decisão do Gestor do sistema*, 2014.
- [17] T. A. Heberlein e G. Warriner, “The influence of price and attitude on shifting residential electricity consumption from on- to off-peak periods,” *Journal of Economic Psychology*, vol. 4, pp. 107 - 130, 1983.



- [18] J.-C. Lu, D.-x. niu e z.-y. jia, "A study of short term load forecasting based on arima-ann," *Machine Learning and Cybernetics*, vol. 5, pp. 3183 - 3187, 2004.
- [19] P. Bunnoon, K. Chalermyanont e C. Limsakul, "Mid Term Load Forecasting of the Country Using Statistical Methodology: Case Study in Thailand," *International Conference on Signal Processing Systems*, pp. 924 - 928, 2009.
- [20] U. b. filik e M. kurban, "A new approach for the short term load forecasting with autoregressive and artificial neural network models," *International journal of computational intelligence research*, vol. 3, pp. 66-71.
- [21] L. Suganthi e A. A. Samuel, "Energy models for demand forecasting - A review," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 1223 - 1240, 2012.
- [22] G. E. P. Box e G. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-Day, Incorporated, 1990.
- [23] O. Valenzuela, I. a. R. F. Rojas, H. Pomares, L. Herrera, A. Guillen e L. a. P. M. Marquez, "Hybridization of intelligent techniques and {ARIMA} models for time series prediction," *Fuzzy Sets and Systems*, vol. 159, pp. 821-845, 2008.
- [24] H. Hippert, C. Pedreira e R. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *Power Systems, IEEE Transactions on*, vol. 16, pp. 44-55, 2001.
- [25] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang e R. L. Szabados, "Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities," *IEEE Transactions on Smart Grid*, 2014.
- [26] M. Hayati e Y. Shirvany, "Artificial neural network approach for short term load forecasting for Illam region," *World Academy of Science, Engineering and Technology*, vol. 28, pp. 280 - 284, 2007.
- [27] E. González-Romera, M. Jaramillo-Morán e D. Carmona-Fernández, "Monthly electric energy demand forecasting with neural networks and Fourier series," *Energy Conversion and Management*, vol. 49, pp. 3135-3142, 2008.

- [28] Z. S. Chan, H. Ngan, A. Rad, A. David e N. Kasabov, “Short-term ANN load forecasting from limited data using generalization learning strategies,” *Neurocomputing*, vol. 70, pp. 409-419, 2006.
- [29] Y. Rui e A. A. El-Keib, “A review of {ANN}-based short-term load forecasting models,” em *System Theory, 1995., Proceedings of the Twenty-Seventh Southeastern Symposium on*, 1995.
- [30] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. Volume 31, pp. 651-666, 2010.
- [31] J. Han e M. Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 2006.
- [32] P. C. E, S. R. C e H. S. Hippert, “Neural Networks For Short-Term Load Forecasting:,” em *Power Systems, IEEE Transactions*, 2001.
- [33] R. Bessa, J. G. d. Sousa e others, “Treino on line de redes neuronais com critérios de informação aplicado à previsão eólica,” Faculdade de Engenharia da Universidade do Porto, 2009.
- [34] P. Chapman, J. Clinton e R. Kerber, *CRISP-DM 1*, 2000.
- [35] “CER,” Commission for Energy Regulation, [Online]. Available: <http://www.cer.ie/>. [Acedido em 19 Janeiro 2015].
- [36] “ISSDA,” [Online]. Available: <https://www.ucd.ie/issda/>. [Acedido em 19 Fevereiro 2015].
- [37] S. M. V. J. M. C. S. M. M. F. M. Joaquim L. Viegas, “Electricity demand profile prediction based on household characteristics,” 2015.
- [38] P. J. Rousseeuw, “Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis,” *Computational and Applied Mathematics*, p. 53–65, 1986.
- [39] A. F. R. Lapedes, “How neural nets work,” em *Neural Information Processing Systems, American Institute of Physics*, New York, 1988.
- [40] Bell Laboratories, “R Language,” [Online]. Available: <https://www.r-project.org/about.html>. [Acedido em 18 10 2015].

- [41] P. Bunnoon, K. Chalermyanont e C. Limsakul, “A computing model of artificial intelligent approaches to mid-term load forecasting: a state-of-the-art survey for the researcher,” *International Journal of Engineering Technology*, vol. 2, pp. 94-100, 2010.
- [42] M. A. Peeples, “R Script for K-Means Cluster Analysis,” 2011. [Online]. Available: <http://www.mattpeeples.net/kmeans.html>. [Acedido em 17 05 2015].



# Anexos

---

## 8. Anexos

### 8.1. Acrónimos

| Acrónimo | Descrição   |
|----------|---|
| AMI      | <i>Advanced Metering Infrastructure</i>                     |
| STLF     | <i>Short-Term Load Forecast</i>                             |
| MTLF     | <i>Mid-Term Load Forecast</i>                               |
| LTLF     | <i>Long Term Load Forecast</i>                              |
| ANN      | <i>Artificial Neural Network – Rede Neuronal Artificial</i> |

### 8.2. Tecnologia Relevante

Os processos de *clustering* e a previsão de consumo de energia foram criados através da linguagem R [40]. Esta linguagem é *open source* que apresenta um vasto conjunto de pacotes de software que permitem realizar as mais variadas análises estatísticas.

No contexto deste trabalho os principais pacotes de software utilizados foram:

- K Means (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html>)
  - Para segmentação da população em diferentes grupos de consumidores com consumos de energia semelhantes;
- Nnet (<https://cran.r-project.org/web/packages/nnet/nnet.pdf>)
  - Treino e Teste da rede neuronal;

### 8.3. Variáveis Associadas a inquéritos de clientes residenciais

Tabela 7 - Variáveis Associadas a inquéritos de clientes residenciais

| Estado Profissional |  |
|---------------------|--|
| 1                   | An employee                            |
| 2                   | Self-employed (with no employees)      |
| 3                   | Self-employed (with no employees)      |
| 4                   | Unemployed (actively seeking work)     |
| 5                   | Unemployed (not actively seeking work) |
| 6                   | Retired                                |
| 7                   | Cared: Looking after relative family   |

| Classe Social |  |
|---------------|--|
| 1             | High or Intermediate managerial, administrative or professional  |
| 2             | Supervisory, clerical and junior managerial, administrative or professional  |
| 3             | Skilled manual workers   |
| 4             | Semi and unskilled manual workers or State pensioners, casual or lowest grade workers, unemployed with state benefits only |
| 5             | Farmers  |
| 6             | Refused  |

| Número de Pessoas na habitação |     |
|--------------------------------|-----|
| 0                              | 0   |
| 1                              | 1   |
| 2                              | 2   |
| 3                              | 3   |
| 4                              | 4   |
| 5                              | 5   |
| 6                              | 6   |
| 7                              | + 7 |

| Tarifário |                |
|-----------|----------------|
| E         | Control        |
| A         | Tariff A       |
| B         | Tariff B       |
| C         | Tariff C       |
| D         | Tariff D       |
| W         | Weekend tariff |

| Estado Aquisição Habitação |                                |
|----------------------------|--------------------------------|
| 1                          | Rent (from a private landlord) |
| 2                          | Rent (from a local authority)  |
| 3                          | Own Outright (not mortgaged)   |
| 4                          | Own with mortgage etc          |
| 5                          | Other                          |

| Tipo Habitação |                     |
|----------------|---------------------|
| 1              | Apartment           |
| 2              | Semi-detached house |
| 3              | Detached house      |
| 4              | Terraced house      |
| 5              | Bungalow            |
| 6              | Refused             |

| Rendimento |                        |
|------------|------------------------|
| 1          | Less than 15,000 Euros |
| 2          | 15,000 to 30,000 Euros |
| 3          | 30,000 to 50,000 Euros |
| 4          | 50,000 to 75,000 Euros |
| 5          | 75,000 or more Euros   |
| 6          | Refused                |

| Nível de Formação |  |
|-------------------|--|
| 1                 | No formal education                              |
| 2                 | Primary  |
| 3                 | Secondary to Intermediate Cert Junior Cert level |
| 4                 | Secondary to Leaving Cert level                  |
| 5                 | Third level                                      |
| 6                 | Refused  |

## 8.4. Variáveis de inquéritos associadas a clientes empresariais

**Tabela 8 - 8.4. Variáveis de inquéritos associadas a clientes empresariais**

| Sector |  | Faturação |                                   |
|--------|--|-----------|-----------------------------------|
| 1      | Agriculture, forestry and fishing  | 1         | < 58,000 Euro                     |
| 2      | Industry Mining and quarrying Manufacturing Electricity, gas, steam and air conditioning supply Water supply, sewerage, waste management and remediation activities  | 2         | 59,000 Euro - 100,000 Euro        |
| 3      | Construction   | 3         | 101,000 Euro - 250,000 Euro       |
| 4      | Wholesale and retail trade; repair of motor vehicles and motorcycles   | 4         | 251,000 Euro - 500,000 Euro       |
| 5      | Business and Professional Services Information and communication Financial and insurance activities Real estate activities Professional, scientific and technical activities Administrative and support service activities Public administration and defence; compulsory social security Education Human health and social work activities | 5         | 501,000 Euro - 750,000 Euro       |
| 6      | Other Transportation and storage Accommodation and food service activities Leisure hotels Arts, entertainment and recreation Other service activities Activities of households as employers; undifferentiated goods- and services producing activities of households for own use Activities of extra-territorial organisations and bodies  | 6         | 751,000 Euro - 1million Euro      |
|        |  | 7         | 1 million Euro - 1.5 million Euro |
|        |  | 8         | 1.6 million Euro - 2 million Euro |
|        |  | 9         | 3 million Euro - 5 million Euro   |
|        |  | 10        | 6 million Euro - 10 million Euro  |
|        |  | 11        | 11 million Euro - 20 million Euro |
|        |  | 12        | over 20million euro               |
|        |  | 13        | Refused                           |

| Número de Colaboradores |                     |
|-------------------------|---------------------|
| 1                       | Sole trader         |
| 2                       | Self employed       |
| 3                       | 1 - 5 employees     |
| 4                       | 6 - 10 employees    |
| 5                       | 11 - 24 employees   |
| 6                       | 25 - 49 employees   |
| 7                       | 50 - 99 employees   |
| 8                       | 100 - 199 employees |
| 9                       | 200+ employees      |

## 8.5. Distribuição da população pelas variáveis dos inquiridos

Nos gráficos seguintes, são apresentadas as distribuições das variáveis com maior relevância subdivididas por tipo de cliente. No caso dos clientes Residenciais temos o estado profissional; o número de habitantes; o tipo de habitação; a classe social; o rendimento; nível de formação; estado de aquisição da habitação ou tarifário. No caso dos clientes empresariais temos disponíveis informações acerca do seu sector de atividade; o número de colaboradores; e o nível de faturação da empresa.

### Clientes Residenciais



Figura 48 – Número de clientes por estado profissional

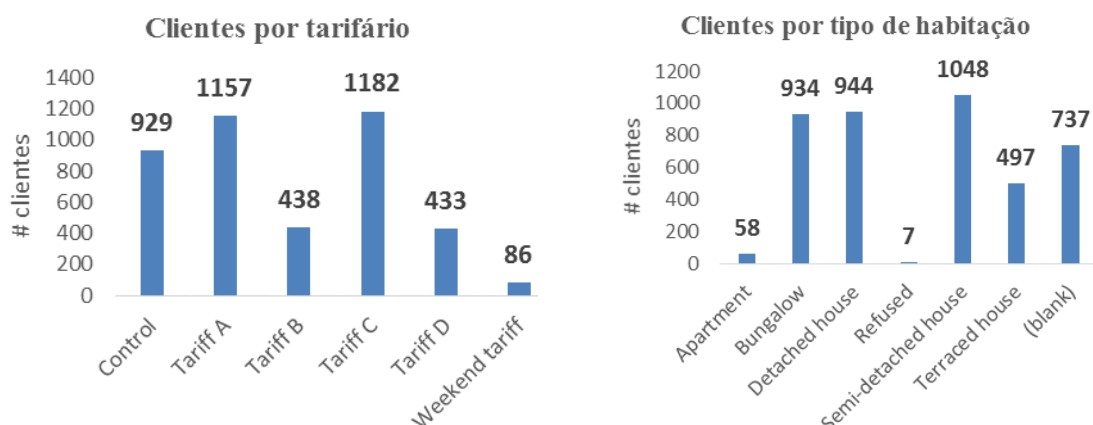
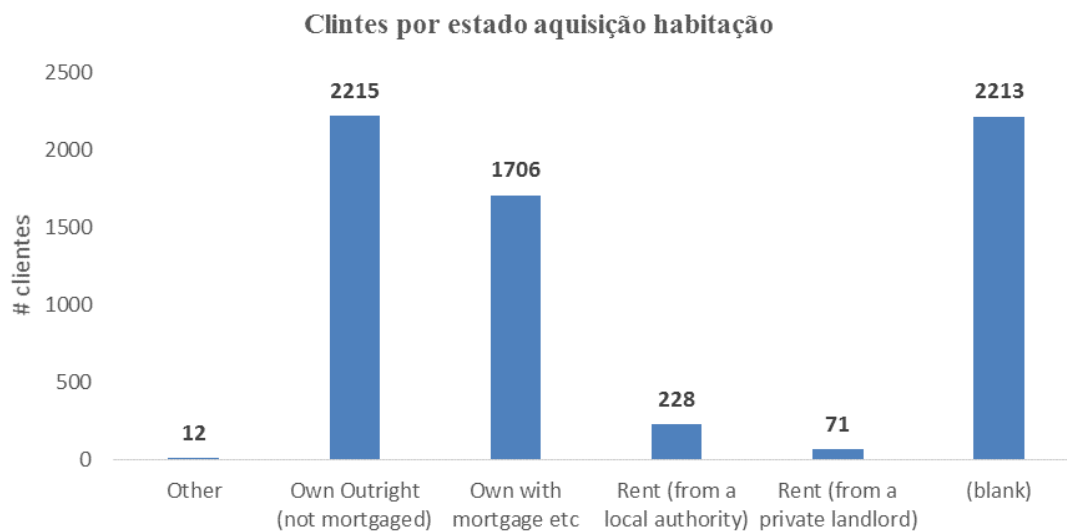


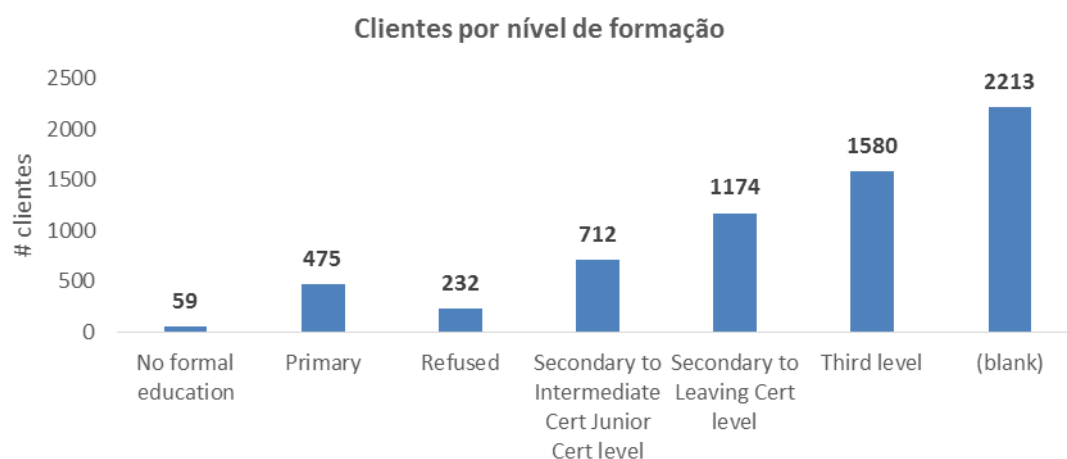
Figura 49 – Número de clientes por tarifário

Figura 50 – Número de clientes por tipo de habitação

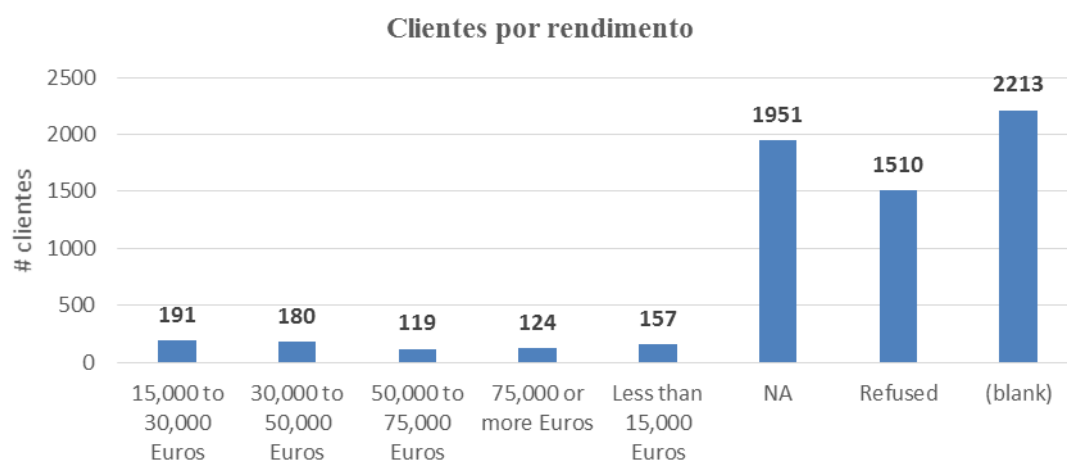




**Figura 51 – Número de clientes por estado de aquisição de habitação**

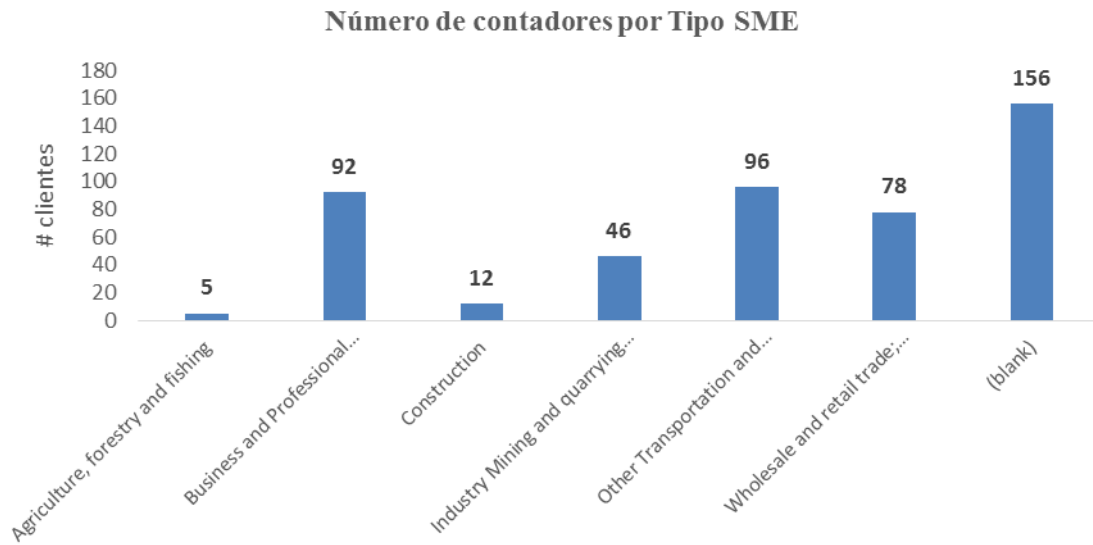


**Figura 52 – Número de clientes por nível de formação**

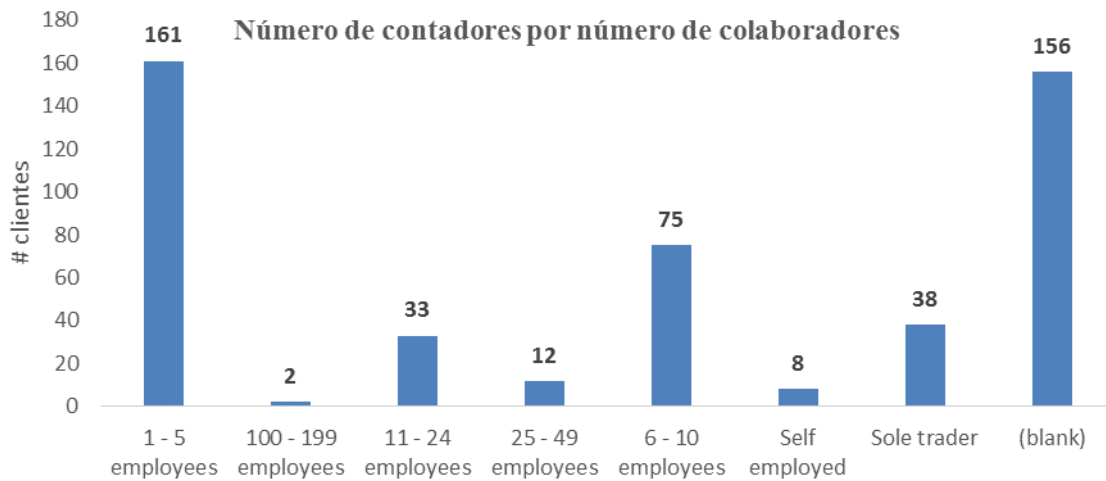


**Figura 53 – Número de clientes por classe de rendimento**

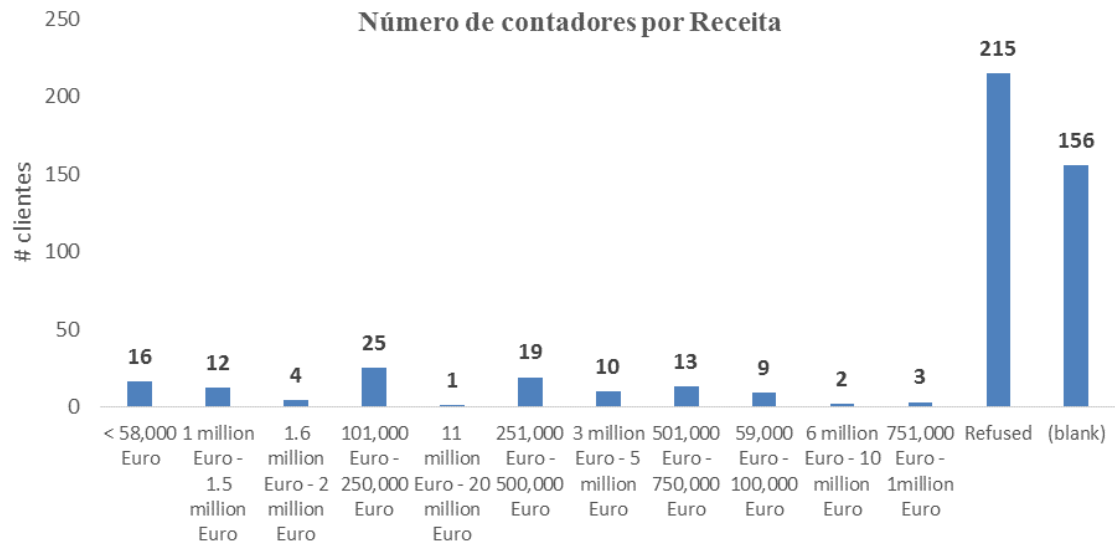
## Cientes Empresariais



**Figura 54 – Número de clientes por tipo de empresa**



**Figura 55 – Número de clientes por classe de número de funcionários**



**Figura 56 – Número de clientes por classe de receita**