

Análise de Dados para Previsão de Micro Produção de Energia Solar e Eólica

Artur Jorge Ferreira da Costa Dias

Dissertação apresentada à Escola Superior de Tecnologia e de Gestão

Instituto Politécnico de Bragança

Para obtenção do grau de Mestre em

Energias Renováveis e Eficiência Energética

Trabalho realizado sob a orientação de:

Professor Doutor Paulo Jorge Pinto Leitão

Bragança, Novembro de 2015

Mestrado em Energias Renováveis e Eficiência Energética

2014/2015

Agradecimentos

Gostaria de aproveitar estes espaço para agradecer a todas as pessoas que contribuíram diretamente ou indiretamente para o sucesso deste trabalho. Especialmente pelo apoio e colaboração de algumas pessoas, evidenciando as seguintes:

Em primeiro ao meu orientador, Professor Doutor Paulo Leitão, por todo o tempo dispensado no desenvolvimento desta dissertação, como por todo o conhecimento que me tentou inculcar, e pelas críticas e motivação que me ajudaram na conclusão desta dissertação.

A todas as pessoas do Laboratório de Controlo, Automação e Robótica (LCAR) pela amizade, palavras de incentivo e pelos momentos de aprendizagem. Principalmente ao Jonas Queiroz por toda a disponibilidade e apoio, pois foi sempre uma pessoa incansável e sempre pronto a ajudar, em que lhe desejo as maiores felicidades no seu caminho rumo ao Doutoramento.

Aos meus amigos de longa data que sempre me apoiaram e me incentivaram ao longo desta dissertação, com uma palavra amiga e de apoio, e sobretudo pela amizade de sempre.

Aos amigos de Bragança pela amizade, por todos os momentos vividos ao longo da minha vida de estudante, e sobretudo por me fazerem acreditar em mim e não me deixarem desistir, nos momentos que estava mais em baixo e com vontade de desistir.

“A maior recompensa pelo nosso trabalho não é o que nos
pagam por ele, mas aquilo em que ele nos transforma”

John Ruskin

Resumo

A liberalização do mercado energético, combinado com a utilização de energias renováveis, tem provocado uma grande transformação no sistema elétrico de energia, tornando-se numa rede inteligente, distribuída e sustentável, designada por “smart grid”. Neste novo paradigma, a rede de energia elétrica deixa de estar assente numa estrutura centralizada e passa a ser assente numa estrutura descentralizada, compreendendo micro produção distribuída, assente em fontes renováveis de energia, como são a produção solar fotovoltaica e eólica.

Neste novo paradigma energético, as tecnologias de informação e comunicação assumem particular relevância na gestão inteligente e eficiente dos recursos e infraestruturas, e ainda integrar os consumidores como participantes ativos na gestão do sistema elétrico de energia (incluindo também os veículos elétricos e a produção usando energias renováveis). O conhecimento do comportamento da produção solar fotovoltaica e eólica para horizontes temporais relativamente curtos tem grande importância nesta gestão inteligente, conduzindo a ganhos económicos.

Neste trabalho foi realizada a análise dos dados de produção da micro rede do IPB, proveniente de sistemas solar fotovoltaicos, correlacionando esses dados com dados meteorológicos de forma a criar modelos de previsão da produção mais fiáveis. Estes modelos de previsão da energia produzida permitem a monitorização em tempo real dos equipamentos de produção, permitindo a deteção de avarias ou degradação na sua operação, assim como uma previsão a produção de energia elétrica a curto e médio prazo, permitindo balancear mais eficazmente o fluxo de energia da micro rede de energia elétrica em análise. Para o efeito foi utilizada a plataforma WEKA e vários algoritmos de análise de dados, nomeadamente, redes neurais artificiais, árvores de decisão, regras e regressão linear. No qual foram obtidos bons modelos de previsão quer para previsão a curto e a médio prazo.

Palavras-Chave: Smart grids, análise de dados, análise preditiva, previsão de produção de energia renovável, energias renováveis.

Abstract

The liberalization of the energy market, combined with the use of renewable energy, has caused a major transformation in the electric power system, making it a smart grid, distributed and sustainable. In this new paradigm, the power grid is no longer based on a centralized structure and is now based on a decentralized structure, comprising micro distributed generation, based on renewable energy sources, as are solar photovoltaic and wind production.

In this new energy paradigm, the information and communication technologies are particularly relevant in the intelligent and efficient management of resources and infrastructure, and also integrate consumers as active participants in the electrical system power management (also including electric vehicles and production using energy renewable). Knowledge of the behaviour of the solar photovoltaic and wind production for relatively short time frames is of great importance in this intelligent management, leading to economic gains.

This work was carried out micro analysis of production data network IPB from solar photovoltaic systems, by correlating these data with weather data so as to create predictive models of the most reliable output. These models forecast the energy produced allow real-time monitoring of production equipment, enabling the detection of faults or degradation in its operation, as well as a forecast production of electricity in the short and medium term, allowing balance more effectively the flow energy micro power grid under review. For this purpose it used WEKA platform and various data analysis algorithms, namely, artificial neural networks, decision trees, rules, and linear regression. In which it was obtained good predictive models both for the short and medium-term forecast.

Keywords: Smart grids, data analysis, predictive analysis, renewable energy production forecast, renewable energy.

Índice

1	INTRODUÇÃO.....	1
1.1	CONTEXTUALIZAÇÃO.....	1
1.2	PROBLEMA EXISTENTE.....	2
1.3	OBJETIVOS DO TRABALHO.....	2
1.4	ORGANIZAÇÃO DO DOCUMENTO.....	3
2	CONCEITOS ASSOCIADOS A ANÁLISE DE DADOS.....	5
2.1	PROSPECÇÃO DE DADOS.....	5
2.2	ANÁLISE DE DADOS APLICADA ÀS SMART GRIDS.....	7
3	FERRAMENTAS DE ANÁLISE E ALGORITMOS.....	8
3.1	FERRAMENTAS DE SUPORTE À ANÁLISE DE DADOS.....	8
3.2	ALGORITMO DE PREVISÃO.....	9
3.2.1	<i>Redes neurais artificiais</i>	9
3.2.2	<i>Regressão linear</i>	11
3.2.3	<i>Árvores de decisão</i>	12
3.2.4	<i>Regras</i>	14
3.3	MEDIDAS DE DESEMPENHO.....	15
4	DESCRIÇÃO DO CASO DE ESTUDO.....	17
4.1	<i>SISTEMA EÓLICO SILOS</i>	18
4.2	<i>SISTEMA SOLAR FOTOVOLTAICO ESTIG</i>	19
4.3	<i>SISTEMA HÍBRIDO ESTIG</i>	20
4.4	SISTEMA SOLAR FOTOVOLTAICO ESA.....	22
4.5	SISTEMA SOLAR FOTOVOLTAICO ESE.....	23
5	PRÉ-PROCESSAMENTO NA ANÁLISE DE DADOS.....	25
5.1	LIMPEZA DOS DADOS.....	25
5.2	PERCEÇÃO DOS DADOS DO SISTEMA SOLAR FOTOVOLTAICO ESTÁTICO.....	26
5.3	PERCEÇÃO DOS DADOS DO SISTEMA FOTOVOLTAICO SEGUIDOR.....	30

5.4	PERCEÇÃO DOS DADOS DO SISTEMA EÓLICO.....	31
5.5	UTILIZAÇÃO DA FERRAMENTA <i>WEKA</i>	32
6	ANÁLISE DE RESULTADOS.....	35
6.1	SISTEMA SOLAR FOTOVOLTAICO ESTÁTICO.....	35
6.1.1	<i>Monitorização da operação do sistema solar fotovoltaico estático.....</i>	<i>35</i>
6.1.2	<i>Previsão da produção de energia a curto prazo.....</i>	<i>38</i>
6.1.3	<i>Previsão da produção de energia a médio prazo.....</i>	<i>41</i>
6.2	SISTEMA FOTOVOLTAICO SEGUIDOR.....	43
6.2.1	<i>Monitorização da operação do sistema fotovoltaico seguidor.....</i>	<i>43</i>
6.2.2	<i>Previsão da produção de energia a curto prazo.....</i>	<i>44</i>
6.2.3	<i>Previsão da produção de energia a médio prazo.....</i>	<i>46</i>
6.3	SISTEMA EÓLICO.....	48
6.3.1	<i>Monitorização da operação do sistema eólico.....</i>	<i>48</i>
6.3.2	<i>Previsão da produção de energia a curto prazo.....</i>	<i>51</i>
6.3.3	<i>Análise para caso de avaria.....</i>	<i>54</i>
6.4	ANÁLISE COMPARATIVA.....	55
6.4.1	<i>Comparação das previsões a curto prazo.....</i>	<i>56</i>
6.4.2	<i>Comparação das previsões a médio prazo.....</i>	<i>57</i>
7	CONCLUSÃO.....	59
7.1	CONCLUSÕES.....	59
7.2	DESENVOLVIMENTO FUTURO.....	60
8	REFERÊNCIAS.....	62
9	ANEXO.....	66
9.1	ANEXO A.....	66
9.2	ANEXO B.....	69
9.3	ANEXO C.....	70
9.4	ANEXO D.....	71
9.5	ANEXO E.....	72

Índice de Figuras

FIGURA 1.1-EVOLUÇÃO DAS REDES DE ENERGIA	1
FIGURA 2.1-METODOLOGIA	6
FIGURA 3.1- UM MODELO SIMPLIFICADO DE UM NEURÓNIO BIOLÓGICA	10
FIGURA 3.2- MODELO NÃO-LINEAR DE UM NEURÔNIO	10
FIGURA 3.3-REDE NEURAL SIMPLES.....	11
FIGURA 3.4- LINHA DE MELHOR AJUSTE PARA REGRESSÃO LINEAR.....	12
FIGURA 3.5-ÁRVORE DE CLASSIFICAÇÃO SIMPLES	14
FIGURA 3.6- ÁRVORE MODELO E AS REGRAS PARA UM CONJUNTO DE DADOS	15
FIGURA 4.1-MICRO REDE DO IPB.....	17
FIGURA 4.2-COMPONENTES DA UNIDADE DE PRODUÇÃO EÓLICA SILOS.....	19
FIGURA 4.3- COMPONENTES DA UNIDADE DE PRODUÇÃO SOLAR FOTOVOLTAICA ESTIG.	20
FIGURA 4.4- COMPONENTES DA UNIDADE DE PRODUÇÃO HÍBRIDA ESTIG.	21
FIGURA 4.5- COMPONENTES DA UNIDADE DE PRODUÇÃO SOLAR FOTOVOLTAICA ESA.	22
FIGURA 4.6- COMPONENTES DA UNIDADE DE PRODUÇÃO SOLAR FOTOVOLTAICA ESE.	23
FIGURA 5.1- PROCESSO DE EXECUÇÃO PARA OBTENÇÃO DOS MODELOS DE PREVISÃO.	25
FIGURA 5.2- EXEMPLO DE FUNCIONAMENTO DA ESTIMAÇÃO DA TEMPERATURA AMBIENTE PARA UM PERÍODO DE 5 MINUTOS E 15 MINUTOS.....	28
FIGURA 5.3-PREPARAÇÃO DO CONJUNTO DE DADOS DE TREINO NO SOFTWARE WEKA	33
FIGURA 5.4-AMBIENTE GRÁFICO PARA APLICAÇÃO DOS ALGORITMOS.....	34
FIGURA 6.1-COMPARAÇÃO DO ERRO ABSOLUTO RELATIVO PARA PERÍODOS DE SOL E NUBLADOS	37
FIGURA 6.2- MONITORIZAÇÃO DO SISTEMA FOTOVOLTAICO ESTÁTICO USANDO O M5P PARA UM PERÍODO DE UMA ANO.....	37
FIGURA 6.3-COMPARAÇÃO DO ERRO ABSOLUTO RELATIVO EM RELAÇÃO A POTÊNCIA CA PARA PREVISÃO DE UM PERÍODO DE TEMPO DE 5 MINUTOS E 15 MINUTOS.....	39
FIGURA 6.4-COMPARAÇÃO DAS PREVISÕES USANDO O M5 RULES COM POTÊNCIA CA REAL, PARA UM PERÍODO DE TEMPO DE 5 E 15 MINUTOS (DOIS DIAS).	40

FIGURA 6.5- ERRO PERCENTUAL PARA PERÍODOS DE 5 E 15 MINUTOS ENTRE A POTÊNCIA EM CA PREVISTA E A REAL (DOIS DIAS).....	40
FIGURA 6.6-PREVISÃO A MÉDIO PRAZO DA POTÊNCIA EM CA PREVISTA USANDO M5RULES (SETE DIAS).	42
FIGURA 6.7-ERRO ABSOLUTO ENTRE A POTÊNCIA EM CA PREVISTA E A REAL (SETE DIAS).....	42
FIGURA 6.8- MONITORIZAÇÃO DO SISTEMA FOTOVOLTAICO SEGUIDOR USANDO O ALGORITMO M5P PARA O PERÍODO DE UM ANO.....	43
FIGURA 6.9- COMPARAÇÃO DO ERRO MÉDIO ABSOLUTO DA POTÊNCIA EM CA PREVISTA PARA UM PERÍODO DE 5 E 15 MINUTOS	46
FIGURA 6.10- COMPARAÇÃO DO ALGORITMO MP PARA UM PERÍODO DE 5 E 15 MINUTOS PARA PREVISÃO DA POTÊNCIA EM CA (TRÊS DIAS).....	46
FIGURA 6.11- PREVISÃO A MÉDIO PRAZO DA POTÊNCIA EM CA PREVISTA USANDO MULTILAYER PERCEPTON (6 DIAS).	48
FIGURA 6.12- COMPARAÇÃO DO ERRO ABSOLUTO RELATIVO PARA PERÍODOS COM MUITO E POUCO VENTO.	50
FIGURA 6.13- MONITORIZAÇÃO DA POTÊNCIA EM CA DO SISTEMA EÓLICO USANDO M5 RULES POR ESTAÇÃO DO ANO.....	50
FIGURA 6.14- COMPARAÇÃO DO ALGORITMOS M5 RULES PARA 5 MINUTOS.....	52
FIGURA 6.15- COMPARAÇÃO DO ALGORITMO M5 RULES PARA UM PERÍODO DE TEMPO DE 15 MINUTOS EM RELAÇÃO A POTÊNCIA EM CA REAL.....	53
FIGURA 6.16-PREVISÃO DA POTÊNCIA EM CA PARA O CASO DE NÃO PRODUÇÃO ENERGIA.....	54
FIGURA 6.17-PREVISÃO PARA UM CASO DE AVARIA.....	55
FIGURA 9.1-ANÁLISE SEMANAL DA IRRADIAÇÃO E POTÊNCIA EM CA.....	66
FIGURA 9.2- PROBLEMA NA POTÊNCIA EM CA E NA IRRADIAÇÃO SOLAR DEVIDO AO SOMBREAMENTO DE UMA ÁRVORE.	67
FIGURA 9.3- RESULTADO DA POTÊNCIA EM CA PREVISTA USANDO O ALGORITMO M5P SEM PROBLEMA INICIAL DE SOMBREAMENTO REFERENTE A TABELA 9.1.....	68

Índice de Tabelas

TABELA 4.1- TABELA COMO OS PARÂMETROS DISPONÍVEIS NAS WEBBOXES.....	18
TABELA 5.1- PARÂMETROS USADOS PARA A CRIAÇÃO DO MODELO PREVISÃO DO SISTEMA FOTOVOLTAICO ESTÁTICO	27
TABELA 5.2- PARÂMETROS USADOS PARA A CRIAÇÃO DOS MODELOS PREVISÃO DO SISTEMA FOTOVOLTAICO SEGUIDOR	30
TABELA 5.3- PARÂMETROS USADOS PARA A CRIAÇÃO DOS MODELOS PREVISÃO DO SISTEMA EÓLICO	31
TABELA 6.1- MODELOS DE PREVISÃO PARA O SISTEMA SOLAR FOTOVOLTAICO, CASO DE MONITORIZAÇÃO DE SEMANA SOL.....	36
TABELA 6.2- MODELOS DE PREVISÃO PARA O SISTEMA SOLAR FOTOVOLTAICO, CASO DE MONITORIZAÇÃO DE SEMANA NUBLADA	36
TABELA 6.3- COMPARAÇÃO DOS ALGORITMOS PARA UM PERÍODO DE TEMPO DE 5 MINUTOS....	38
TABELA 6.4- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO DE UM PERÍODO DE TEMPO DE 15 MINUTOS.	38
TABELA 6.5- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO A MÉDIO PRAZO.	41
TABELA 6.6- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO 5 MINUTOS PARA SEGUIDOR SOLAR.	44
TABELA 6.7- COMPARAÇÃO DOS ALGORITMOS PARA UMA PREVISÃO DE UM PERÍODO DE TEMPO DE 15 MINUTOS PARA O SEGUIDOR SOLAR	45
TABELA 6.8- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO A MÉDIO PRAZO PARA O SISTEMA FOTOVOLTAICO SEGUIDOR.	47
TABELA 6.9- RESULTADOS PARA O PERÍODO DE UMA SEMANA COM MUITO VENTO.....	49
TABELA 6.10- RESULTADOS PARA O PERÍODO DE UMA SEMANA COM POUCO VENTO.....	49
TABELA 6.11- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO DE PERÍODOS DE 5 MINUTOS PARA O SISTEMA EÓLICO.	51
TABELA 6.12- COMPARAÇÃO DOS ALGORITMOS PARA PREVISÃO DE PERÍODOS DE 15 MINUTOS PARA O SISTEMA EÓLICO	52

TABELA 6.13-COMPARAÇÃO DOS MELHORES MODELOS DE PREVISÃO DE PRODUÇÃO ENERGIA PARA MONITORIZAÇÃO	56
TABELA 6.14- COMPARAÇÃO DOS MELHORES MODELOS DE PREVISÃO DE PRODUÇÃO PARA UM PERÍODO DE 5 MINUTOS.....	56
TABELA 6.15- COMPARAÇÃO DOS MELHORES MODELOS DE PREVISÃO DE PRODUÇÃO PARA UM PERÍODO DE 15 MINUTOS	57
TABELA 6.16- COMPARAÇÃO DOS MELHORES MODELOS DE PREVISÃO A MÉDIO PRAZO	58
TABELA 9.1-TESTE PARA VERACIDADE DA TÉCNICA APLICADA PARA REMOÇÃO DO SOMBREAMENTO	67

Abreviaturas

CRISP-DM	<i>Cross industry standard process for data mining</i>
ESA	Escola Superior Agrária
ESE	Escola Superior de Educação
ESTIG	Escola Superior de Tecnologia e Gestão
GE	<i>General Electric</i>
GNU	<i>General Public License</i>
GPS	<i>Global Positioning System</i>
HyRef	Hybrid Renewable Energy Forecasting
IBM	<i>International Business Machines</i>
IPB	Instituto Politécnico de Bragança
P_{ac}	Potência em corrente alternada
PLC	<i>Programmable Logic Controller</i>
PVGIS	<i>Photovoltaic Geographical Information System</i>
RNAs	Redes Neurais Artificiais
V_{ca}	Tensão em corrente alternada

1 Introdução

A elevada dependência energética de recursos fósseis, e as preocupações com questões ambientais, que vieram a aumentar devido ao uso excessivo de recursos fósseis, representam os principais motivos que levaram ao crescimento da produção de energia através de fontes renováveis. No entanto, a mudança de fontes de energia de recursos fósseis para fontes renováveis, não é suficiente para mudar as questões ambientais. É importante uma utilização de forma racionalizada da energia que é produzida.

1.1 Contextualização

A alteração do paradigma energético surgiu com o aparecimento de pequenos produtores energéticos, usando micro produção baseada em fontes de energia renováveis, como sejam sistemas solares fotovoltaicos ou microturbinas eólicas. Este facto obrigou à alteração da arquitetura tradicional das redes de energia elétrica, assente numa estrutura centralizada, para uma arquitetura descentralizada, conduzindo ao aparecimento das redes inteligentes, designadas por “*smart grid*”. Estas redes oferecem uma maior integração, e um melhor controlo sustentável sobre a rede, com serviços de maior qualidade [1]. A Figura 1.1 ilustra esta evolução [2].

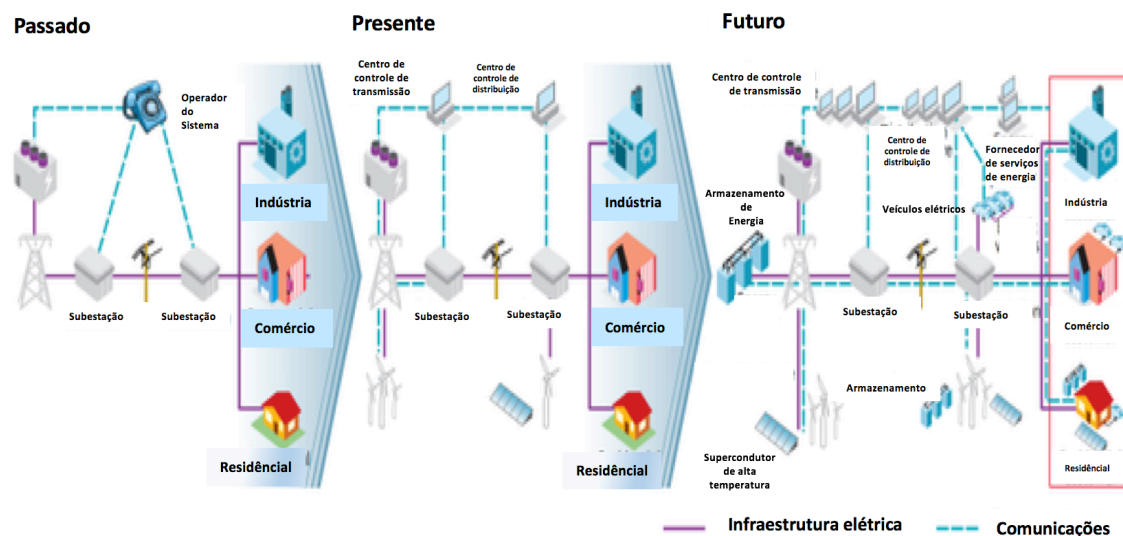


Figura 1.1-Evolução das redes de energia (igual de [2]).

A *smart grid* é uma rede elétrica inteligente e distribuída que usa as tecnologias de informação e comunicação mais avançada na gestão inteligente e eficiente dos recursos e infraestruturas, e ainda na integração dos consumidores como participantes ativos na

gestão do sistema elétrico de energia. As redes inteligentes são capazes de coordenar as necessidades e as capacidades de todos os produtores e operadores da rede, de forma tão eficiente quanto possível, minimizando os custos e os impactos ambientais e ao mesmo tempo maximizando a confiabilidade do sistema.

1.2 Problema existente

Um problema existente nas redes inteligentes, é em relação aos agentes produtores, terem a capacidade de saber com base nos consumos dos anos anteriores as necessidades energéticas que precisam de satisfazer para os clientes finais. Mas como sabemos, na produção de energia, as energias renováveis são as primeiras a entrar no plano para satisfazer o consumo, e muitas das vezes quem faz a gestão das redes de energia não tem capacidade de saber o que vai ser produzido por parte dos pequenos e grandes produtores de energias renováveis. Então a compensação é feita por parte de energias não renováveis. Assim, seria de todo conveniente conseguir prever a quantidade de energia que será produzida para os sistemas de produção fotovoltaica e eólica. No entanto, esta previsão não é simples e linear e requer a utilização de métodos de análise de dados e aplicação de modelos de previsão para os diferentes sistemas estudados. O conhecimento do comportamento da produção fotovoltaica e eólica para horizontes temporais curtos tem grande importância para os produtores bem como para os operadores de rede do sistema elétrico, contribuindo para utilização da energia de forma eficiente.

1.3 Objetivos do trabalho

O objetivo deste trabalho centra-se na análise dos dados de forma a ser possível obter os modelos de previsão de produção de energia elétrica assente em fontes renováveis, nomeadamente sistemas solares fotovoltaicos e turbinas eólicas. Esta previsão da energia produzida permitirá a monitorização em tempo real dos sistemas de produção, permitindo a deteção de avarias ou a degradação dos sistemas de produção, assim como a previsão da produção de energia elétrica a curto e médio prazo, permitindo balancear mais eficazmente o fluxo de energia da micro rede de energia elétrica em análise.

A micro rede em análise é a do Instituto Politécnico de Bragança (IPB), que compreende sistemas solares fotovoltaicos e eólicos, em que os dados de produção

serão correlacionados com os dados meteorológicos (por exemplo, irradiação solar, temperatura e velocidade do vento). Para o efeito, foi utilizada a plataforma WEKA e vários algoritmos de análise de dados, nomeadamente, redes neurais artificiais, árvores de decisão, regras e regressão linear, que utilizaram dados históricos e atuais relativos aos dispositivos de produção de energia solar fotovoltaica, dados de previsão meteorológica e dados meteorológicos colecionados por estações meteorológicas existentes no campus do IPB.

1.4 Organização do documento

Este documento encontra-se organizado em seis capítulos. No presente capítulo é apresentada uma introdução do tema da dissertação, o problema existente, os objetivos delineados para a sua solução e a organização do documento.

O segundo capítulo apresenta os conceitos associados à análise de dados, onde é abordado o conceito e as vantagens de aplicação da análise de dados. É ainda detalhado o processo de análise de dados, desde a recolha dos dados, ao tratamento dos dados, à aplicação dos algoritmos de análise de dados e por fim à análise dos resultados.

No terceiro capítulo são descritas, de uma forma breve, algumas ferramentas de análise de dados e alguns algoritmos de previsão, sendo dada especial atenção à ferramenta WEKA e aos algoritmos redes neurais artificiais, árvores de decisão, regras e regressão linear utilizados neste trabalho.

O quarto capítulo descreve o caso de estudo deste trabalho, nomeadamente a micro rede do IPB que é composta por vários sistemas solar fotovoltaicos estáticos, um sistema solar fotovoltaico seguidor e dois sistemas eólicos. É também detalhado o conjunto de instrumentação de medida associado a cada sistema, nomeadamente referente à medição de irradiação solar e velocidade do vento.

O quinto e sexto capítulo descrevem a aplicação dos algoritmos utilizados neste trabalho para a criação dos modelos de previsão, sendo discutidas as melhores abordagens para os diferentes cenários analisados, tendo por base os resultados experimentais obtidos. Neste capítulo é ainda realizada uma análise para alguns casos de avaria que surgiram.

Por fim, no sétimo capítulo são apresentadas as principais conclusões referentes ao trabalho desenvolvido e apresentadas algumas direções para o desenvolvimento de um trabalho futuro.

Em anexo encontram-se listado o *dataset* utilizado no desenvolvimento deste trabalho e que resultou da aquisição de dados na micro rede do IPB no período de 01 de Janeiro de 2012 a 31 Setembro de 2015.

2 Conceitos Associados a Análise de Dados

A cada segundo que passa são guardadas e armazenadas grandes quantidades de dados, que primeiramente não sabemos o seu significado nem como os interpretar. Uma primeira abordagem é a sua representação gráfica, de forma a permitir visualizar como são distribuídos e como variam as relações entre eles. Ao conseguir observar os dados somos capazes de os analisar e saber o que esses dados significam e até detetar padrões [3].

Historicamente, o conceito de analisar dados e encontrar padrões em dados está associado a uma variedade de diferentes nomes, incluindo prospeção de dados, origem de conhecimentos, descoberta de informação, recolha da informação, arqueologia de dados e processamento de dados padrão [4].

A análise de dados é cada vez mais importante na sociedade para resolução de problemas complexos que possuem muita informação de difícil interpretação, nomeadamente, medicina, alimentação, transportes, sistemas de produção, clima e energia. Esta importância é reconhecida pela Comissão Europeia no seu programa de investigação e desenvolvimento Horizon 2020, sendo o *Big Data* uma das áreas de maior relevo. As maioria das empresas estão alinhadas com este desafio e importância da análise de dados na melhoria da sua produtividade e competitividade, e em especial empresas relacionadas com a produção de energia baseada em fontes renováveis [5].

2.1 Prospeção de dados

A prospeção de dados (em inglês *Data Mining*), um subcampo interdisciplinar da ciência da computação, é umas das técnicas para encontrar e descrever padrões estruturais nos dados, constituindo uma ferramenta para ajudar a explicar os dados e fazer previsões a partir deles [6]. A prospeção de dados é um processo computacional que pode usar um grande conjuntos de dados, e envolver métodos no cruzamento da inteligência artificial, aprendizado de máquina, estatística e sistemas de base de dados, de forma a inspecionar, limpar, transformar e modelar dados com o objetivo de extrair informação útil e ajudar na tomada de decisões. A análise de dados pode ser abordada de diversas maneiras e engloba diversas técnicas. O método Inteligência de negócios (em inglês *Business intelligence*) abrange a análise de dados mas depende fortemente da agregação e com foco em informações de negócios. Em aplicações estatísticas, a análise

de dados divide-se por estatística descritiva, análise exploratória de dados e análise de dados de confirmação [4].

A metodologia de prospeção de dados mais relevante é a *Cross Industry Standard Process for Data Mining* (CRISP-DM) que é constituída por seis fases para resolução do problema onde vai ser aplicada a análise de dados, tal como ilustra a Figura 2.1.

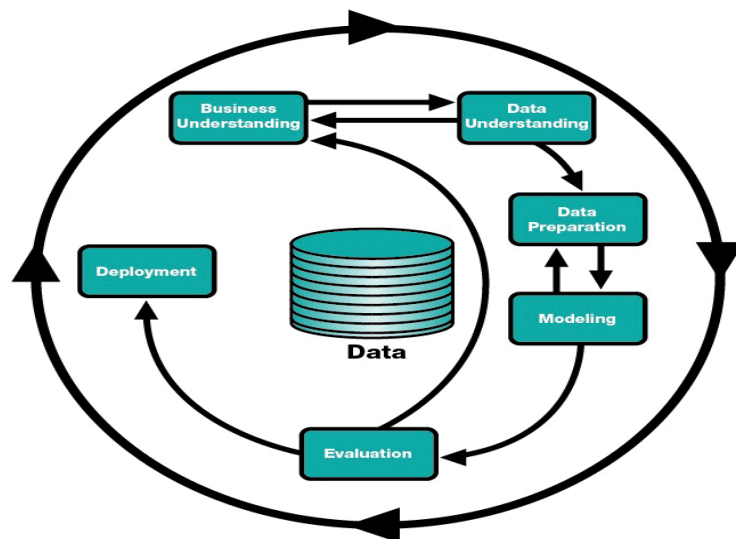


Figura 2.1-Metodologia CRISP-DM [7].

A metodologia CRISP-DM [8] tem com primeira fase a compreensão do negócio ou (do Inglês *Business understanding*), que tem como função entender o que o cliente pretende realizar numa perspetiva de negócio. Na segunda fase está presente a compreensão dos dados (do Inglês *Data understanding*), onde é realizada a aquisição e familiarização dos dados . Numa terceira etapa, é executada a preparação dos dados (do Inglês *Data preparation*), que é referente à decisão sobre o volume e o tipo de dados a usar no processo de análise.

Na quarta fase temos a modelação (do Inglês *modeling*) que é a fase onde é selecionado o algoritmo ou a modelação a utilizar na análise de dados, por exemplo se uma rede neural ou uma técnica baseada em regras. A quinta etapa consiste na avaliação (do Inglês *Evaluation*), que consiste na avaliação do modelo de forma a verificar se o modelo alcança adequadamente os objetivos pretendidos. Por fim, numa última fase temos a implantação (*Deployment*), que pode variar de um simples relatório ou um complexo processo de prospeção de dados repetível em toda a organização [7][9].

2.2 Análise de Dados aplicada às Smart grids

A análise de dados pode ser aplicada em diversas situações no contexto das smart grids, como por exemplo na previsão de produção de energia por parte de um sistema solar fotovoltaico [10], na previsão de carga elétrica para saber a disponibilidade de energia, isto de forma o sistema operar de forma econômica e com maior fiabilidade [11]. Pode ser também utilizada para prever o consumo de aparelhos domésticos usando um modelo probabilístico, uma rede de sensores sem fios, e medindo o consumo de energia por parte dos eletrodomésticos 24 horas por dia. Esta análise permite prever quais os dispositivos mais utilizados, em que momento do dia e durante quanto tempo, para uma melhor gestão da energia consumida [12]. A análise de dados pode ser também aplicada para prever os preços da energia e a procura por parte dos consumidores [13].

A empresa VESTAS [14] coopera com IBM no projeto designado por “Hybrid Renewable Energy Forecasting” (HyRef) [15]. Neste projeto são utilizadas tecnologias avançadas de visão artificial, onde a aquisição de imagens de câmaras direcionadas para o céu permite rastrear o movimentos de nuvens, enquanto sensores em turbinas eólicas monitorizam a temperatura, e a velocidade e direção do vento. A aplicação de técnicas de análise de dados aos dados colecionados permite obter previsões climáticas exatas no local, para um mês em avanço ou de 15 em 15 minutos.

A utilização da análise de dados permite a obtenção de previsões meteorológicas locais, suportando a perceção da energia que pode ser redirecionada para a rede de energia ou armazenada [15] e assim melhorar a gestão da rede elétrica.

3 Ferramentas de Análise e Algoritmos

Neste capítulo serão descritas algumas ferramentas de análise de dados, sendo dada especial atenção à ferramenta WEKA e aos algoritmos redes neurais artificiais, árvores de decisão, regras e regressão linear utilizados neste trabalho.

3.1 Ferramentas de suporte à Análise de Dados

A análise de dados pode ser realizada através da utilização de diversas ferramentas de *software*, livres ou comerciais, que fornecem um ambiente de suporte à implementação de modelos de prospeção de dados. Estas ferramentas fornecem uma maior interatividade entre os utilizadores e os dados, através de um ambiente gráfico amigável e a disponibilização de algumas funcionalidades já integradas para um maior apoio e rapidez de resposta.

Na área do *software* livre existe disponível a ferramenta *Orange* com base em prospeção de dados e aprendizagem máquina, em que é utilizado como base a linguagem de programação *python*, a ferramenta *Tanagra* com orientação para a visualização, e a ferramenta *OpenNN* mais direcionada para o ensino e constituindo uma biblioteca aberta com base em redes neurais. A ferramenta R [16] usa linguagem de programação e ambiente de desenvolvimento de *software* para computação estatística e gráfica. A ferramenta KNIME [17] permite fazer a montagem do pré-processamento dos dados num ambiente gráfico, a ferramenta RapidMiner [18] é muito usada para aplicações comerciais e industriais, e a ferramenta Weka [19] consiste numa *framework* desenvolvida e escrita na linguagem *Java*[®], o qual disponibiliza diversos algoritmos de aprendizagem de máquina.

Em relação a *softwares* comerciais, é possível encontrar o IBM SPSS, o *Microsoft Analysis*[®] e o Oracle Data.

Nesta dissertação foi usado o *Weka*[®], versão 3.7.12, desenvolvido na Universidade de *Waikato*, na Nova Zelândia, que consiste num *software* livre (distribuído sob os termos da GNU *General Public License*), de código aberto e amplamente utilizado e suportado por uma grande comunidade, a qual dispõe de diversos conteúdos e documentos que auxiliam na aprendizagem e utilização do mesmo. O *Weka*[®] tem a vantagem de poder ser usado em diferentes plataforma e sistemas operativos *Linux*[®], *Windows*[®], *Macintosh*[®].

O *Weka*[®] contém um ambiente de trabalho apelativo que fornece um amplo suporte para todo o processo de *data mining* experimental, incluindo a preparação dos dados de entrada, a avaliação estatística dos regimes de aprendizagem, e a visualização dos dados de entrada e dos resultados da aprendizagem. A ferramenta fornece também uma variedade de algoritmos de aprendizagem que podem ser imediatamente aplicados para uma obtenção rápida de resultados, assim como métodos para pré-processamento e avaliação dos resultados [6].

Este conjunto de ferramentas diversificadas, abrangentes e disponível através de uma interface comum, permite aos utilizadores a comparação de diferentes métodos e algoritmos, e identificar aqueles que são mais adequados para o problema em questão.

3.2 Algoritmo de Previsão

Neste subcapítulo são descritos alguns algoritmos usados para realizar a previsão a curto prazo (5 e 15 minutos) e a médio prazo (hora a hora), da produção de energia de fontes renováveis, nomeadamente redes neurais artificiais, árvores de decisão, regras e regressão linear.

3.2.1 Redes neurais artificiais

O estudo das redes neurais artificiais (RNAs) é um dos principais ramos da inteligência artificial. Durante os últimos anos, tem havido um aumento relativo no interesse nas redes neurais artificiais. As RNAs são adequadas para a realização de algumas tarefas, especificamente, tarefas que envolvem conjuntos de dados incompletos, informação distorcida e para problemas bastantes complexos e mal definidos, em que normalmente os seres humanos decidem numa base intuitiva. As RNAs podem aprender a partir de exemplos, e são capazes de lidar com problemas não lineares e apresentam alguma vigor e alguma tolerância a falhas [20]. Com isso, podem ser aplicadas em inúmeras situações, como sejam a previsão de meteorologia [21] e a previsão das tendências do mercado [22].

As RNAs baseiam o seu princípio de funcionamento na analogia com o cérebro humano, sendo este composto por 100 bilhões de neurónios interligados [23]. A Figura 3.1 ilustra um modelo simplificado de um neurónio biológico. No cérebro existe um fluxo de informação codificada (usando meios eletroquímicos, os chamados neurotransmissores) das sinapses no sentido do axónio. O axónio de cada neurónio

transmite informação para um certo número de outros neurónios. O neurônio recebe informações nas sinapses de um grande número de outros neurónios. Estima-se que cada neurônio pode receber até 10.000 estímulos de outros neurónios. Grupos de neurónios são organizados em subsistemas e a integração desses subsistemas forma o cérebro.

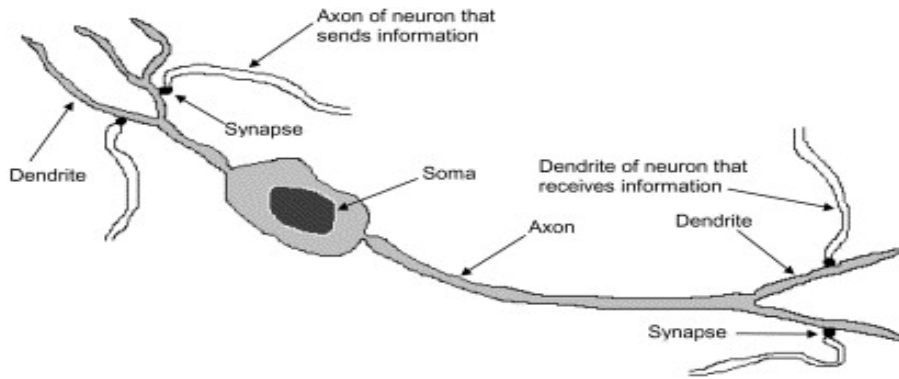


Figura 3.1- Um modelo simplificado de um neurônio biológica [24].

A Figura 3.2 apresenta um modelo altamente simplificado de um neurônio artificial, que pode ser utilizado para ilustrar alguns aspetos importantes do neurônio biológico real. Uma RNA é um grupo de neurónios artificiais interligadas, interagindo uns com os outros de uma forma concertada. Neste sistema, a excitação é aplicada à entrada da rede. A realização de uma operação adequada resulta numa saída desejada. Nas sinapses, existe uma acumulação de algum potencial, o que no caso dos neurónios artificiais é modelada como uma ligação pesada. Os pesos associados às ligações são continuamente modificados com base em regras de aprendizagem adequadas [24].

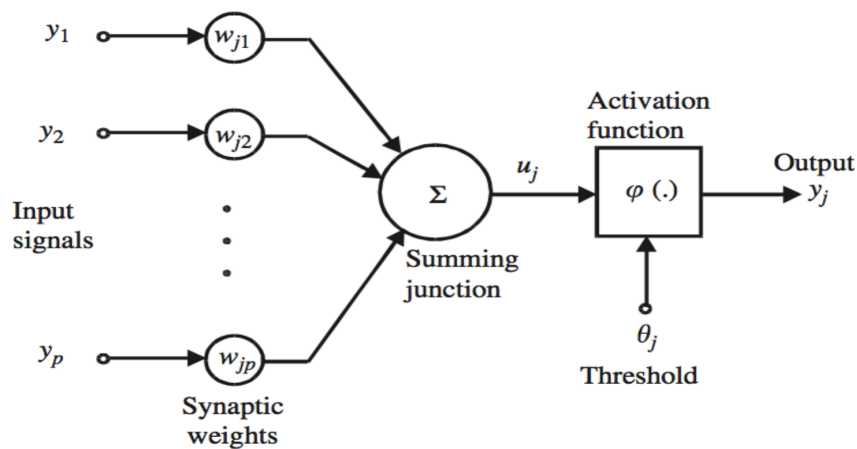


Figura 3.2- Modelo não-linear de um neurônio (adaptado de [20]).

As RNAs são compostas por uma ou mais entradas, uma saída e uma ou mais camadas escondidas, como podemos visualizar na Figura 3.3. Cada neurónio é ligado a partir da camada de entrada para um neurónio a partir da camada oculta, e da camada oculta de cada nó é ligado a um neurónio da camada de saída.

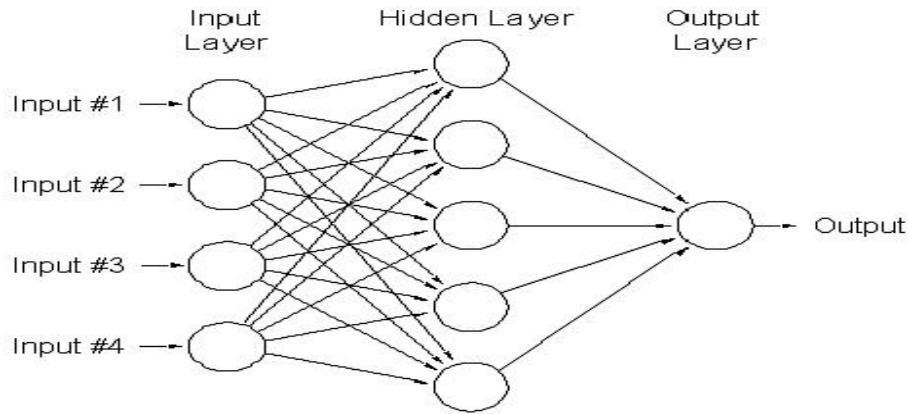


Figura 3.3-Rede Neural Simples [25].

Uma camada de entrada representa a informação em bruto que é alimentada para dentro da rede. Cada valor único de entrada para a rede é duplicado e enviado para baixo, para os nódulos na camada escondida. A camada oculta aceita os dados a partir da camada de entrada.

A RNA usa o valor de entrada e modifica esse valor, que é enviado para a camada de saída, mas também será modificado por algum peso de ligação entre a camada oculta e de saída. A informação do processo da camada de saída é recebido da camada oculta e produz uma saída. Esta saída é processada pela função de ativação [25].

3.2.2 Regressão linear

Em estatística, regressão linear é uma abordagem para modelar a relação entre um escalar variável dependente e uma ou mais variáveis independentes. Na regressão linear, as relações são modeladas utilizando funções de previsão lineares cujo modelo desconhecido são parâmetros estimados a partir dos dados.

Neste modelo, o valor a prever é expresso como uma combinação linear dos atributos, com pesos pré-determinados:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k \quad (3.1)$$

no qual x é o valor a prever e w_i são os pesos, determinados numa fase de treino, e a_k são os atributos.

A regressão linear é um método excelente, para previsão numérica simples, e tem sido utilizada em aplicações estatísticas há muitas décadas. É claro, os modelos lineares tem uma desvantagem de linearidade. Isto se os dados exibirem uma dependência não-linear, a linha reta de melhor ajuste será encontrada. Onde melhor é interpretado como a diferença mínima da média quadrática. Esta linha pode não se encaixar muito bem. No entanto, se os modelos lineares servir bem como blocos de construção para métodos de aprendizagem mais complexos [6][26].

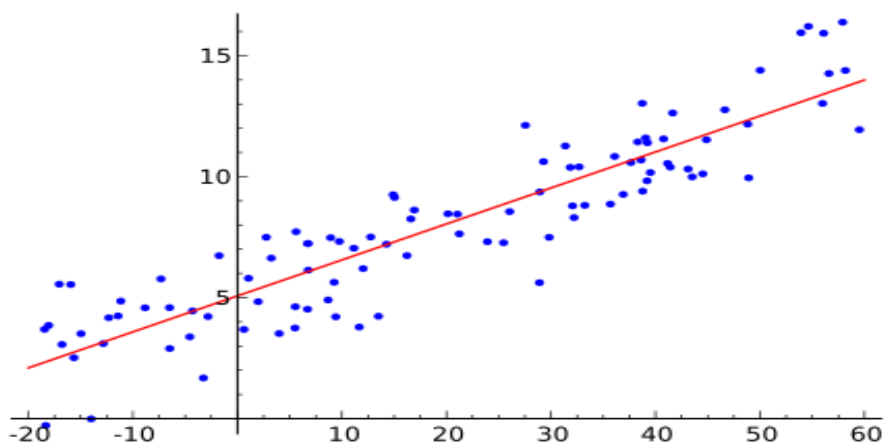


Figura 3.4- Linha de melhor ajuste para regressão linear (igual de [27])

Outra forma simples é a representação de um modelo linear, a saída do qual é simplesmente a soma dos valores de atributos, exceto que os pesos são aplicados a cada atributo antes de adicioná-los em conjunto. O truque é chegar a bons valores para os pesos que fazem a saída do modelo coincidir com a saída desejada. Aqui, a saída e as entradas de atributos são valores numérico. Os modelos lineares são mais fáceis de visualizar em duas dimensões, onde eles são o mesmo que desenhar uma reta através de um conjunto de pontos de dados. Como mostra a linha ajustada representada na Figura 3.4 do desempenho do “melhor ajuste” para a equação de previsão.

3.2.3 Árvores de decisão

A árvore de decisão usa um modelo de árvore em que a variável alvo pode usar um conjunto finito de valores no qual são chamados de árvore de classificação. Nesta estruturas de árvore, as folhas representam combinações de características que levam a esses graus de classe. As árvores de decisão, onde a variável alvo pode usar valores

contínuos (tipicamente números reais) são chamadas de árvores de regressão. As árvores de classificação são frequentemente usadas em diversas áreas, por exemplo finanças, marketing, engenharia e medicina [28].

Nós em uma árvore de decisão envolvem testar um atributo específico. Normalmente, o ensaio compara um valor de atributo com uma constante. Nós da folha da uma classificação que se aplica a todas as instâncias que atingem a folha, ou um conjunto de classificações, ou uma distribuição de probabilidades sobre todas as classificações possíveis. Para classificar uma instância desconhecido, ele será encaminhado para baixo da árvore de acordo com os valores dos atributos testados em nós sucessivos, e quando uma folha é alcançada a instância é classificada de acordo com a classe atribuída à folha.

Por vezes os valores de atributo são divididos em subconjuntos, e os ramos de árvore apenas em duas formas, dependendo do valor que se encontra no subconjunto. Nesse caso, o atributo pode ser testado mais do que uma vez num caminho. Se o atributo for numérico, o teste num nó geralmente determina se o seu valor é maior ou menor do que uma constante predeterminada, dando uma divisão de duas vias. Alternativamente, uma divisão de três vias pode ser utilizada, caso em que existem várias possibilidades diferentes.

Uma alternativa para um atributo de valores inteiros seria uma separação tripartida para menor que, igual a, e maior do que. Uma alternativa para um atributo de valor real, para o que é igual a uma opção não é tão significativa, seria testar contra um intervalo, em vez de uma única constante, novamente dando uma separação tripartida: a seguir, dentro de, e acima. Um atributo numérico é frequentemente testado várias vezes em qualquer caminho dado para baixo da árvore da raiz às folhas, cada teste envolvendo uma constante diferente.

A árvore de classificação ou árvores de decisão é um algoritmo que cria um guia passo-a-passo de como se determina a saída de uma nova instância de dados. A árvore que ele cria é exatamente isto, uma árvore em que em cada nó da árvore representa um ponto onde a decisão deve ser tomada com base na entrada, e move-se para o próximo nó e no próximo até chegar a uma folha que diz que o previsto saída. Vamos analisar um exemplo na Figura 3.5 [29].

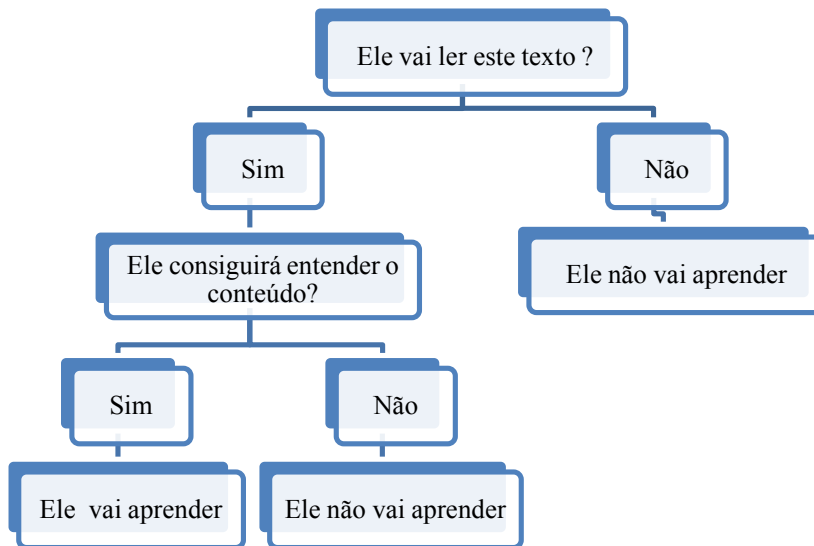


Figura 3.5-Árvore de classificação simples (adaptado de [29]).

Esta árvore de classificação simples procura responder à questão "Ele entenderá árvores de classificação?" Em cada nó, responde à pergunta e continua na ramificação até encontrar uma folha que responda sim ou não.

Esse modelo pode ser usado para qualquer conjunto de dados desconhecidos, isto se for capaz de prever esse conjunto de dados desconhecidos e aprender sobre árvores de classificação perguntando-lhe somente questões simples. Aparentemente, essa é a grande vantagem de uma árvore de classificação e não necessita de muitas informações sobre os dados para criar uma árvore em que pode ser muito exata e muito informativa.

3.2.4 Regras

O método de gerar regras com base em árvores, que é conhecido também como M5'Rules, é simples e compreende a aplicação sucessiva de um modelo em árvore a um conjunto de dados de forma a serem construídas regras de aprendizagem. Após a aplicação do modelo ao conjunto de dados treinados para uma árvore e aprendida, uma regra é construída e a árvore é rejeitada. Todos os casos abrangidos pela regra são removidos do conjunto de dados. O processo é aplicado de forma repetida aos restante casos e termina quando todos os casos são cobertos por uma ou mais regras.

Em contraste com a árvore parcial, que emprega a mesma estratégia para a previsão precisa, o método M5'Rules constrói árvores completas em vez de árvores parcialmente exploradas. A construção de árvores parciais conduz a uma maior

eficiência computacional, e não afeta o tamanho e a precisão das regras resultantes. Este objetivo concentra-se na geração de regras usando modelos lineares.

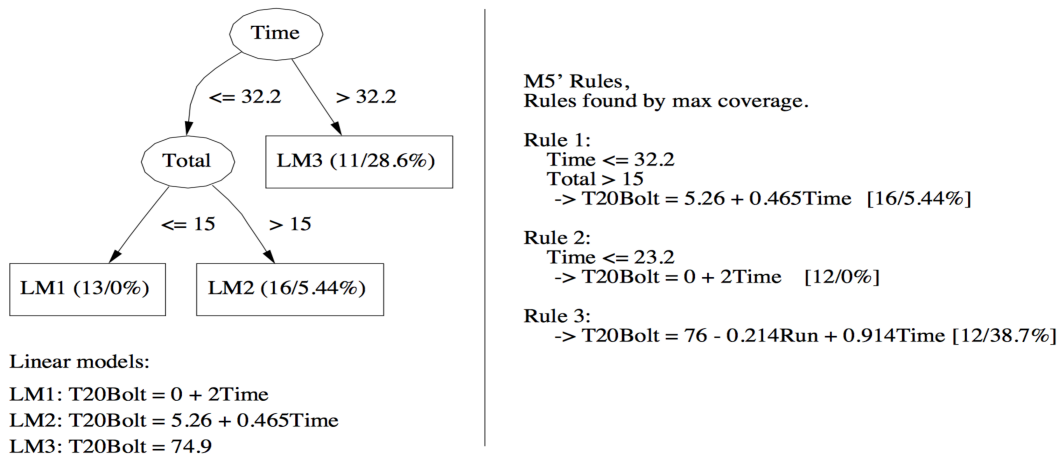


Figura 3.6- Árvore modelo e as regras para um conjunto de dados [30].

Até agora, temos descrito uma abordagem geral para extrair regras com base árvores, aplicáveis a qualquer classificação ou regressão. Resta determinar, em cada fase, que a folha na árvore é o melhor candidato para além do conjunto de regras. A abordagem mais óbvia, deve escolher a folha que cobre a maioria dos exemplos. A Figura 3.6 mostra uma árvore produzida pelo algoritmo M5'Rules e as regras geradas por M5'Rules usando a experiência de cobertura para o conjunto de dados.

3.3 Medidas de desempenho

Na avaliação da aplicação dos algoritmos de análise de dados para criação dos modelos de previsão, são usualmente utilizados os seguintes critérios (sendo P os valores previstos, a os valores atuais, n o número total de valores e i refere-se ao valor numérico da previsão para a instância de teste):

- **Coefficiente de correlação:** mede a correlação estatística entre os valores previstos e os valores atuais. O coeficiente de correlação varia entre 0 e 1 sendo 0 o pior valor a obter e 1 o melhor valor para resultados correlacionados [31].

$$\frac{S_{PA}}{\sqrt{S_P S_A}}, \text{ onde } S_{PA} = \frac{\sum_i (P_i - \bar{P})(a_i - \bar{a})^2}{n - 1}, S_P = \frac{\sum_i (P_i - \bar{P})^2}{n - 1}, \quad (3.1)$$

$$e S_A = \frac{\sum_i (a_i - \bar{a})^2}{n - 1}$$

- **Erro médio absoluto:** média dos erros absolutos, sendo que quanto menor for este erro, menor será a diferença entre os valores previstos e os valores reais. [32].

$$\frac{|P_1 - a_1| + \dots + |P_n - a_n|}{n} \quad (3.2)$$

- **Erro médio quadrático:** indica o quão distante, em média, o conjunto estimado está do parâmetro a ser estimado [31].

$$\frac{(P_1 - a_1)^2 + \dots + (P_n - a_n)^2}{n} \quad (3.3)$$

- **Erro absoluto relativo:** erro total absoluto, com o mesmo tipo de normalização dos restantes erros, em que \bar{a} é o valor médio ao longo dos dados usados para treino [33].

$$\frac{|P_1 - a_1| + \dots + |P_n - a_n|}{|a_1 - \bar{a}| + \dots + |a_n - \bar{a}|} \quad (3.4)$$

- **Raiz quadrada do erro médio quadrático:** indica a diferença entre os valores (amostra e valores da população)[34].

$$\sqrt{\frac{(P_1 - a_1)^2 + \dots + (P_n - a_n)^2}{n}} \quad (3.5)$$

Estes critérios são todos importantes na análise de dados a realizar, cada um dos quais fornecendo uma indicação particular. Por exemplo, o erro médio absoluto pode ser uma medida importante, visto fornecer uma perceção em média, da diferença entre o valor previsto e o valor real; no entanto, o erro absoluto relativo é também importante pois fornece uma ideia absoluta de forma percentual de qual o erro de previsão.

4 Descrição do caso de estudo

A micro rede usada como caso de estudo neste trabalho é designada por VERCampus - Campus Vivo de Energias Renováveis, com uma potência de pico total de 68.3 kWp, que fica situada no Instituto Politécnico de Bragança [35]. Esta micro rede é composta por vários sistemas de produção conforme se encontra ilustrado na Figura 4.1, nomeadamente várias unidades solar fotovoltaica estática, uma unidade solar fotovoltaica seguidor e duas unidades de microprodução com turbinas eólicas. No Anexo B encontra-se uma descrição mais detalhada da micro rede.

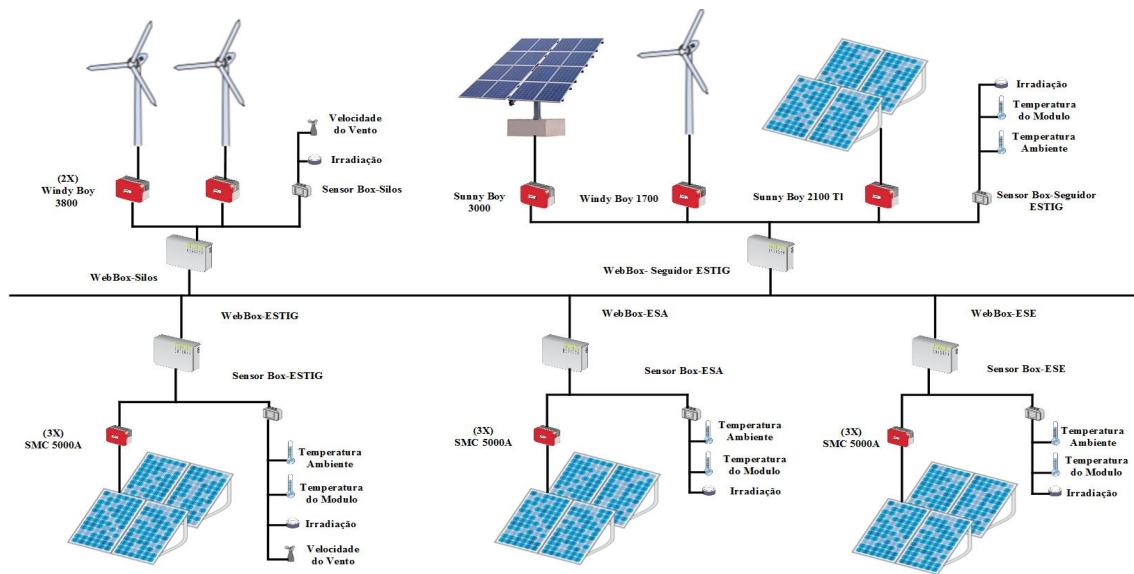


Figura 4.1-Micro rede do IPB.

A monitorização da produção de energia na micro rede VERCampus é efetuada através de um sistema integrado que permite ter acesso aos dados colecionados por cada uma das unidades, quer local quer remotamente via *web*. A informação disponibilizada inclui, entre outros, a energia produzida diariamente, o total de energia produzida, o número de horas de funcionamento, a potência injetada na rede, a frequência da rede, a temperatura ambiente, a temperatura dos módulos, o nível de irradiação, a velocidade do vento, e o estado de operação do equipamento. A informação da temperatura ambiente, temperatura do módulo, nível de irradiação e velocidade do vento são obtidas pelos sensores que estão ligados a um equipamento designado por *sensorbox*. A *sensorbox* e o inversor de cada sistema de produção de energia estão ligados a uma unidade designada de *webbox*, -SMA Solar Technology que coleciona e disponibiliza os dados remotamente.

A Tabela 4.1 ilustra os parâmetros mais relevantes colecionados pelas *webboxes*. Os dados são adquiridos de 5 em 5 segundos e colecionados pela *webbox* de cinco em cinco minutos (fazendo uma média dos valores para esse período) [36].

Tabela 4.1- Tabela como os parâmetros disponíveis nas webboxes.

Parâmetros	Descrição	Webbox-ESTIG	Webbox-Seguidor ESTIG	Webbox-Silos	Webbox-ESA	Webbox-ESE
TimeStamp	Horas do sistema	①✓	①✓	①✓	①✓	①✓
IntSolIrr	Irradiação Interna	①✓	①✓	①✓	①✓	①✓
TmpAmb	Temperatura Ambiente	①✓	①✓	⊗✗	①✓	①✓
TmpMdul	Temperatura do Modulo	①✓	①✓	⊗✗	①✓	①✓
WindVel	Velocidade do Vento	①✓	⊗✗	①✓	①✗	⊗✗
Pac	Potência em corrente alternada	①✓	①✓	①✓	①✓	①✓

✓-Contem dados; ①-Funciona corretamente; ✗-Não contem dados; ⊗-Não funciona corretamente

De notar que em algumas *webboxes*, alguns parâmetros não se encontram disponíveis mesmo existindo os sensores na *sensorbox*, provavelmente devido a avaria desses sensores.

Neste caso de estudo, vamos considerar cinco sistemas de produção de energia renovável, cada um tendo associado uma *webbox*, e que serão detalhados de seguida.

4.1 Sistema Eólico Silos

O sistema de produção eólico encontra-se localizado nos silos do campus do IPB e compreende duas turbinas FORTIS (modelo Montana) com potência de 5kW (para uma velocidade do vento de 17m/s) [37], tal como ilustrado na Figura 4.2.

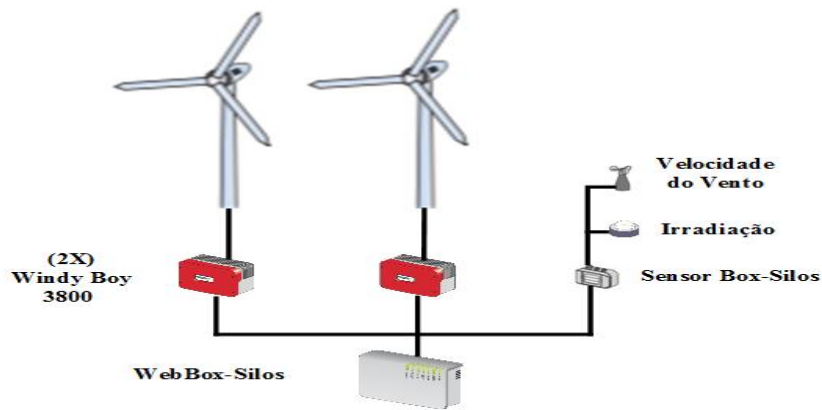


Figura 4.2-Componentes da unidade de produção eólica Silos.

Cada turbina é ligada a um inversor *Windy Boy 3800*, com potência em CC máxima de 4.040W, tensão CC máxima de 500V, corrente CC máxima de 20A. Na saída do inversor (tensão de 220-24Vca e frequência de 50Hz) é apresentada uma potência em CA máxima de 3800 W. A *webbox* responsável pela aquisição e colecionar os dados na unidade de produção eólica Silos, além de colecionar informação relativa à produção de energia, possui também sensores que permitem a medição da velocidade do vento e da irradiação solar.

4.2 Sistema Solar Fotovoltaico ESTIG

A unidade de produção solar fotovoltaica instalada na Escola Superior de Tecnologia e Gestão (ESTIG) [36] compreende 84 módulos fotovoltaicos da marca GE Energy (modelo GEPVp-205-M), com um tipo de célula policristalina e com uma potência de pico de 205Wp, que perfazem uma potência de pico de 17,22 kWp formando três matrizes, tal como é ilustrado na Figura 4.3. Cada matriz é constituída por duas fileiras em paralelo, cada uma com 14 módulos ligados em série.

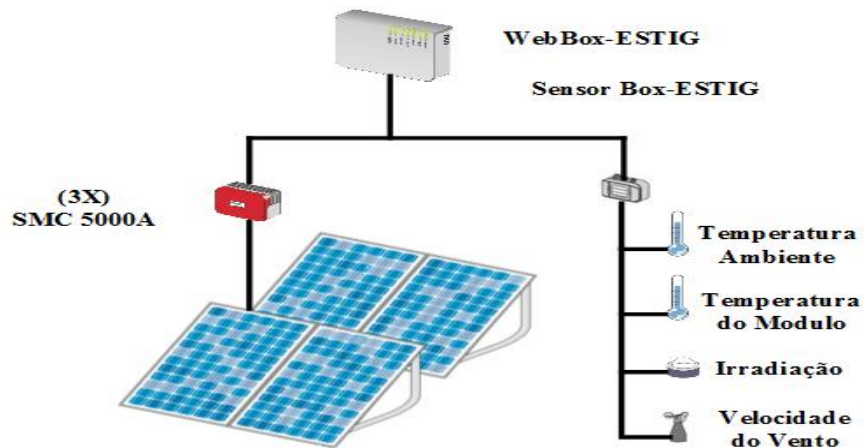


Figura 4.3- Componentes da unidade de produção solar fotovoltaica ESTIG.

Cada matriz está ligada a um inversor que permite a injeção da energia produzida na rede elétrica. O inversor fotovoltaico utilizado é da marca SMA, modelo SMC 5000, com potência em CC máxima de 5750W, tensão em CC máxima de 600V e corrente máxima de 24A. Na saída do inversor para a rede elétrica existe uma tensão de 220-240Vca, frequência de 50Hz e potência em CA máxima de 5500W.

A unidade de produção solar fotovoltaica ESTIG disponibiliza ainda um *sensorbox* na qual se pode obter as medições da temperatura ambiente, da temperatura dos módulos, da irradiação e da velocidade do vento.

4.3 Sistema Híbrido ESTIG

A unidade de produção híbrida instalada na ESTIG [38] é constituída por três elementos de produção de energia elétrica, tal como ilustrado na Figura 4.4: um painel solar fotovoltaico seguidor, uma turbina eólica e um conjunto de nove painéis solares fotovoltaicos fixos.

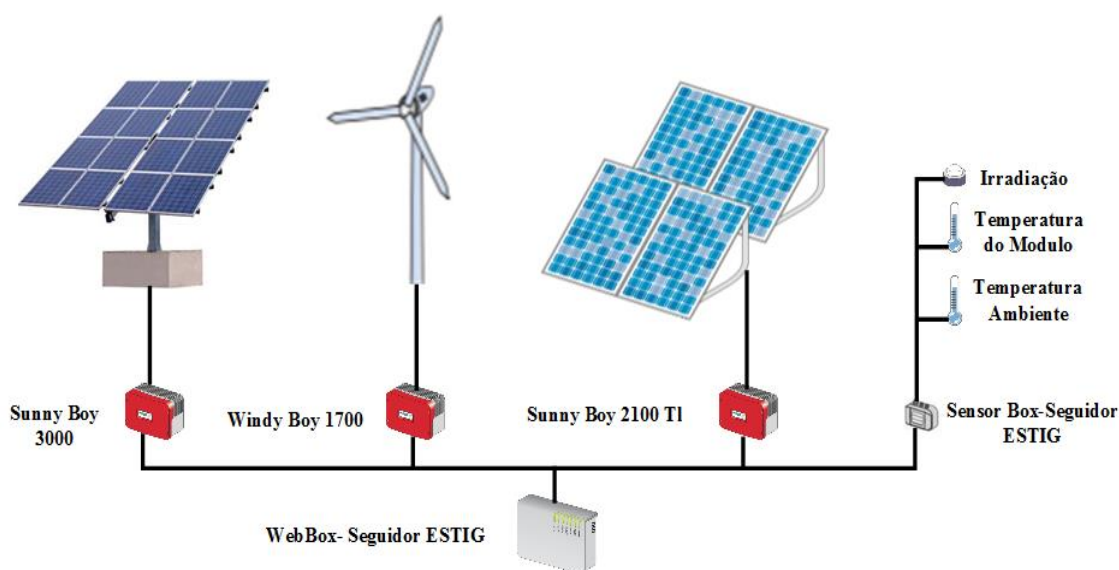


Figura 4.4- Componentes da unidade de produção híbrida ESTIG.

O painel solar fotovoltaico seguidor é constituído por 16 módulos fotovoltaicos ligados em série, de modelo Kyocera 190GHT-2, com um tipo de célula policristalina e uma potência de pico de 190Wp. O painel está ligado a um inversor fotovoltaico da marca SMA, modelo Sunny Boy 300, de potência em CC máxima de 2700W, tensão CC máxima de 600V e corrente máxima de 12A. Na saída do inversor para a rede elétrica existe uma tensão de 220-240Vca, frequência de 50Hz e uma potência em CA máxima de 2.500W. O seguidor solar movimenta-se em dois eixos através do controlo por um PLC (Programmable Logic Controller) e é orientado por GPS (Global Positioning System) [38].

A turbina eólica é uma turbina da marca Fortis, modelo Passaat, utilizando uma potência de pico de 1.400Wp (a uma velocidade de vento de 16m/s), estando ligada a um retificador específico, equipado com uma resistência para dissipar picos de potência. A turbina possui um inversor da marca SMA, modelo Windy Boy 1700, com uma potência em CC máxima de 1.850W, tensão CC máxima 400V e corrente CC máxima 12,6A. Na saída do inversor para a rede elétrica temos a tensão de 220-240Vca, frequência de 50Hz e potência CA máxima de 1550W [39].

O conjunto de módulos solar fotovoltaicos fixos, são formados por nove módulos da marca Fluitecnick, modelo FTS-220P, com um tipo de célula policristalina com uma potência de 220Wp e que estão ligados em série com um inversor da marca SMA,

modelo Sunny Boy 2100 TI, com uma potência em CC máxima de 2200W, uma tensão CC máxima 600V e uma corrente CC máxima 12A. Na saída do inversor para a rede elétrica temos uma tensão de 220-240Vca, frequência de 50Hz e potência CA máxima de 1950W.

A unidade de produção híbrida da ESTIG inclui um *sensorbox* que adquire informação sobre a temperatura ambiente, a temperatura dos módulos e a irradiação incidente nos painéis, sendo todos estes valores guardados na *webbox*.

4.4 Sistema Solar Fotovoltaico ESA

A unidade de produção solar fotovoltaica instalada na Escola Superior Agrária (ESA) [36], representada na Figura 4.5 é formada por 84 módulos fotovoltaicos, da marca GE Energy (modelo GEPVp-205-M) com um tipo de células policristalinas e com uma potência de pico de 205Wp, que perfazem uma potência de pico de 17,22kWp isto formando três matrizes. Cada matriz é constituída por duas fileiras em paralelo, cada uma com 14 módulos ligados em série. Cada matriz está ligada a um inversor fotovoltaico, para a injeção da energia na rede elétrica.

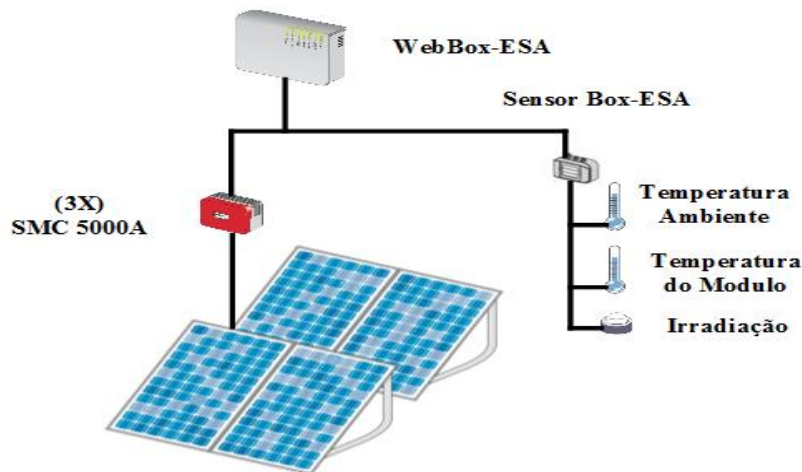


Figura 4.5- Componentes da unidade de produção solar fotovoltaica ESA.

O inversor da marca SMA, modelo SMC 5000A com potência em CC máxima de 5.750W, tensão CC máxima de 600V e corrente máxima de 24A. Na saída do inversor para a rede elétrica temos a tensão de 220-240Vca, a frequência de 50Hz e com potência em CA máxima de 5500W. A unidade possui um também conjunto de sensores que fornecem informação sobre a temperatura ambiente, a temperatura dos módulos e a irradiação.

4.5 Sistema Solar Fotovoltaico ESE

Por fim, a unidade de produção solar fotovoltaica da Escola Superior de Educação (esse) [36], representada na Figura 4.6, é formada por 84 módulos fotovoltaicos da marca GE Energy (modelo GEPVp-205-M) com um tipo de células policristalina e com uma potência de pico de 205Wp, que perfazem uma potência de pico de 17,22kWp, isto formando três matrizes. Cada matriz é constituída por duas fileiras em paralelo, cada uma com 14 módulos ligados em série. Cada matriz ligada a um inversor, para a injeção da energia na rede elétrica.

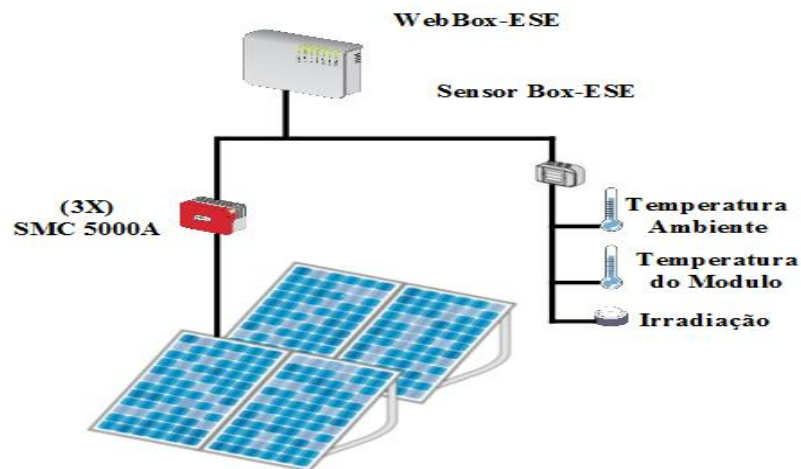


Figura 4.6- Componentes da unidade de produção solar fotovoltaica ESE.

O inversor da marca SMA, modelo SMC 5000A, apresenta uma potência em CC máxima de 5750W, tensão CC máxima de 600V e corrente máxima de 24A. Na saída do inversor para a rede elétrica temos uma tensão de 220-240Vca, frequência de 50Hz e potência em CA máxima de 5.500W [36].

A unidade possui um também conjunto de sensores que fornecem informação sobre a temperatura ambiente, a temperatura dos módulos e a irradiação.

Neste trabalho pretende-se realizar a análise dos dados de produção dos vários sistemas solar fotovoltaicos e eólicos da micro rede do IPB, correlacionando esses dados adquiridos com os dados meteorológicos de forma a criar modelos de previsão da produção mais fiáveis. Para o efeito foram definidas três situações:

1. Monitorização em tempo real dos equipamentos de produção de energia, permitindo a deteção de avarias ou degradação na sua operação

2. Previsão da produção de energia elétrica a curto prazo, isto é para os próximos 5 e 15 minutos, permitindo reagir rapidamente a eventuais desvios na produção.
3. Previsão da produção de energia elétrica a médio prazo, isto é para a próxima hora, permitindo balancear mais eficazmente o fluxo de energia na micro rede.

Nos capítulos que se seguem serão descritos os procedimentos realizados para realizar a análise de dados e efetuada a análise dos resultados obtidos.

5 Pré-Processamento na Análise de Dados

Nesta seção serão abordadas algumas atividades realizadas com o objetivo de preparação dos dados para posterior análise e avaliação, nomeadamente a limpeza dos dados e a perceção.

A Figura 5.1 ilustra o processo de execução para criação dos modelos de previsão. Numa primeira fase, é realizada a aquisição e armazenamento dos dados provenientes dos sistemas de produção de energia elétrica, os quais, combinados com os dados de previsão meteorológica, por exemplo provenientes do *Open Weather* [40] e *Weather underground* [41], são introduzidos na plataforma *Weka*. Aqui é realizada uma limpeza dos dados e aplicados os algoritmos de previsão com o objetivo de obter o melhor modelo de previsão para o sistema em questão.

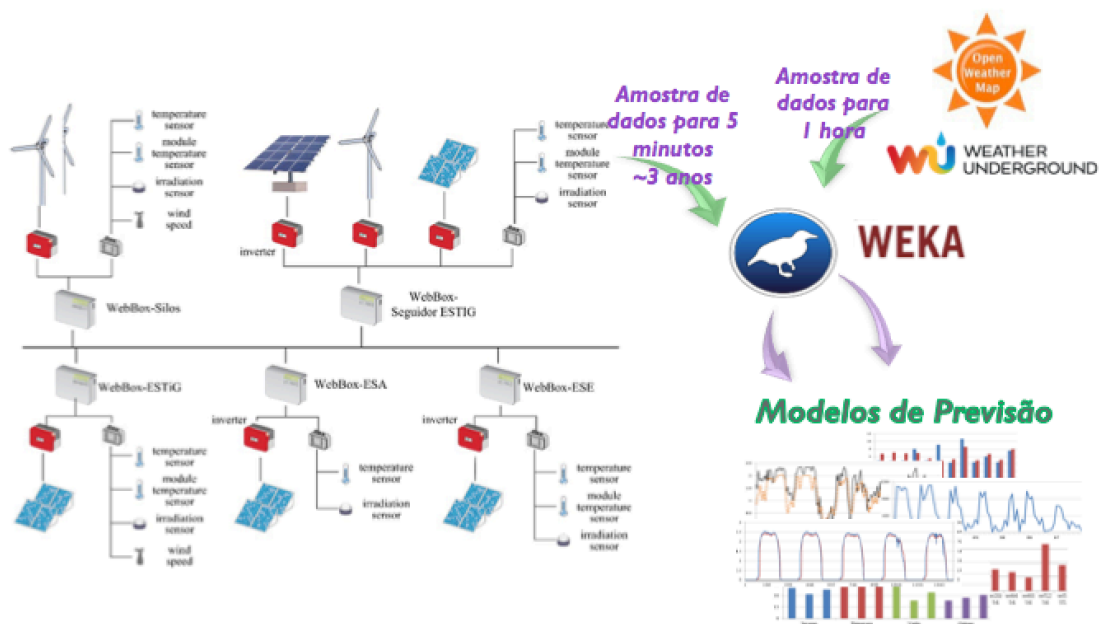


Figura 5.1- Processo de execução para obtenção dos modelos de previsão.

5.1 Limpeza dos dados

A preparação dos dados para aplicação dos algoritmos com o objetivo de criação dos modelo de previsão, requer um tratamento inicial dos dados para todos os sistemas de produção analisados. Esse tratamento é aplicado aos conjuntos de dados de treino e de validação.

Neste tratamento inicial são removidos as instâncias com valores vazios, e com valor zero, que possam existir em ambos os conjuntos de dados analisados. É também importante remover os dados onde a potência em corrente alternada (Pac) é inferior a 1W, pois para a criação dos modelos de previsão de energia elétrica só são usados os dados quando o sistema se encontra em produção, e isso acontece quando existe irradiação solar ou vento. A razão de excluir a potência em CA inferior a 1W é que podemos ter algum erro de medição por parte do sensor ou por produção irrelevante e isso pode afetar a criação dos modelo de previsão.

5.2 Percepção dos dados do Sistema Solar Fotovoltaico Estático

A Tabela 5.1 sintetiza os vários parâmetros, provenientes de diferentes fontes, que serão utilizados na análise de dados para o sistema solar fotovoltaico estático. A fonte de dados mais relevante consiste nos parâmetros colecionados por sensores incorporados no sistema fotovoltaico, nomeadamente a potência, temperatura ambiente e temperatura dos módulos PV. Estes parâmetros foram medidos e registados para cada 5 minutos desde os últimos 3 anos, constituindo os dados históricos da operação de painéis fotovoltaicos. Outra fonte de dados inclui um conjunto de dados meteorológicos obtidos a partir de sites de meteorologia, tais como o *Open Weather* [40] e o *Weather underground* [41] [42]. A partir destes sites, é possível recuperar os dados meteorológicos históricos, tais como a temperatura ambiente, velocidade do vento e cobertura de nuvens, bem como as previsões meteorológicas. As previsões geralmente não são muito precisas (quando o período de previsão é largo), o que afeta o modelo de previsão [43].

Tabela 5.1- Parâmetros usados para a criação do modelo previsão do sistema fotovoltaico estático

Fonte da Informação	Parâmetro	Descrição
Sistema de Produção	Irradiação Solar	Irradiação solar (W/m^2) medida por sensor instalado no sistema
	Potência em corrente alternada	Potência de corrente alternada (W) produzida pelo sistema
	Temperatura ambiente	Temperatura ($^{\circ}C$) medida por sensor instalado no sistema
	Temperatura dos módulos	Temperatura ($^{\circ}C$) dos módulos fotovoltaicos
	Velocidade do vento	Velocidade do vento (m/s) medida por sensor instalado no sistema
OpenWeathermap.org +Wunderground.com	Temperatura ambiente	Uma aproximação da temperatura local ($^{\circ}C$) com base em estações de perto do local, geralmente para períodos de tempo de 1 hora
	Humidade	Concentração de vapor de água no ar em %
	Cobertura das nuvens	Cobertura percentual (%) de céu com nuvens
	Velocidade do vento	Velocidade do vento local (m/s)
	Irradiação calculada	Irradiação calculada com base na dedução de algumas formulas

Após a realização de diversas experiências, no caso de monitorização, para a situação de sol (verão) foram considerados todos os dados disponíveis para a implementação do treino dos modelos de previsão e escolhida a semana de 25 a 31 de Março de 2015 para a realização dos testes. Relativamente à situação de nublada (inverno), foram considerados igualmente todos os dados disponíveis para o treino e escolhida a semana de 24 a 30 de Novembro de 2014 para os testes.

Na previsão da produção de energia a curto e a médio prazo foram considerados todos os dados disponíveis para o treino e usada a semana de 4 a 13 de Março de 2015 para os testes. Nesta situação, alguns parâmetros tiveram que ser estimados. De facto, nos *websites* de meteorologia, geralmente, são fornecidas previsões para diferentes períodos de tempo de horas e dias, com a finalidade de saber o clima para os próximos períodos de tempo. Na criação dos modelos de previsão para intervalos de 5-15 minutos à frente, foi necessário estimar os parâmetros para o conjunto de dados a utilizar, antes de se realizar a análise de dados no *software Weka*. Estes parâmetros foram estimados com base nos valores anteriores medidos pelos sensores do sistema solar fotovoltaico estático. Estes valores tiveram de ser estimado devido a falta de informação nos sites

meteorológicos para curtos prazos. A equação 5.1 formaliza a função de previsão \hat{x}_t que é referente ao valor estimado para qualquer um dos parâmetros em consideração (e.g., irradiação solar e temperatura ambiente).

$$\hat{x}_t = x_t + \frac{x_t - x_{t-1}}{2} \quad (5.1)$$

onde, x_t é o valor medido no instante atual e x_{t-1} é o valor medido no instante anterior ao atual. A Figura 5.2 representa como a equação 5.1 foi usada para fazer a estimativa dos valores do parâmetro temperatura ambiente, para os períodos de 5 e 15 minutos.

Temperatura em horas (HH:MM)	Temperatura Ambiente (°C)	Previsão para 1 passo à frente	Previsão para 3 passos à frente
09:15	13.603		
09:20	13.305		
09:25	13.337		
09:30	12.724	13.353	
09:35	11.768	12.4175	
09:40	11.923	11.29	13.353

Figura 5.2- Exemplo de funcionamento da estimação da temperatura ambiente para um período de 5 minutos e 15 minutos.

Neste exemplo, o período das 09:25 corresponde ao período corrente. Logo, utilizando a equação 5.1, o valor estimado da temperatura ambiente para o período das 9:30 (5 minutos à frente) e 9:40 (15 minutos à frente) corresponde ao valor atual acrescido de uma taxa de aumento ou diminuição, relativa à tendência apresentada pelos dois últimos valores. Esta tendência é definida em função de metade da diferença entre o valor atual e o valor imediatamente anterior.

Na previsão a médio prazo, foi usado para treino todos os dados disponíveis e para teste os dados de da semana de 4 a 10 de Março de 2015. Relativamente aos dados de previsão meteorológica usados, estes dados são fornecidos por vários *sites* de medição meteorológica para períodos de hora a hora Estes *websites* costumam fazer previsões para a temperatura ambiente, para a humidade, para a velocidade do vento, para a cobertura das nuvens, entre outros, que não influenciam a produção de energia solar fotovoltaica. No entanto, a irradiação solar na superfície, que é o principal fator que influencia a produção de energia solar fotovoltaica, não se encontra disponível.

Assim, a fim de obter uma aproximação para o valor deste parâmetro, foi utilizada a irradiação solar extraterrestre horária, calculada de acordo com a equação 5.2 [43][44].

$$I_0 = I_{sc} \cdot \left(1 + 0.033 \cos \left(\frac{360^\circ d_n}{365} \right) \right) \cdot (\sin \delta \cdot \sin \phi + \cos \delta \cdot \cos \phi \cdot \cos \omega) \quad (5.2)$$

em que, I_{sc} é o valor da constante solar (1367W/m^2), d_n é o dia do ano contado a partir do início do ano, δ é a declinação solar em graus decimais calculado de acordo com a equação 5.3, ϕ é a latitude geográfica em graus decimais, correspondente à localização do sistema solar fotovoltaico, ω é o ângulo hora em graus decimais e calculado de acordo com a equação 5.4.

$$\delta = 23.45^\circ \cdot \sin \left(360^\circ \frac{284 + d_n}{365} \right) \quad (5.3)$$

$$\omega = 15 \cdot (h_s - 12) \quad (5.4)$$

onde, o parâmetro h_s é o tempo de sol em horas, calculado por:

$$h_s = H_{UTC} + ET - \frac{\textit{longitude}}{15} \quad (5.5)$$

H_{UTC} são as coordenadas universais do tempo, em horas, calculas através da equação 5.6.

$$H_{UTC} = OT - SA \quad (5.6)$$

onde, OT é o tempo padrão local em horas (o tempo mostrado por um relógio) e SA é o tempo em horas, pelo qual os relógios são adiantados do fuso horário local. Na União Europeia, SA consiste geralmente uma hora durante o inverno e outono e duas horas durante a primavera e o verão.

O parâmetro ET corresponde à “Equação do Tempo” em horas, obtido a partir da equação 5.7.

$$ET = \frac{9.87 \cdot \sin 2\beta - 7.53 \cdot \cos \beta - 1.5 \cdot \sin \beta}{60} \quad (5.7)$$

onde, β , em graus decimais, é dado por :

$$\beta = \frac{360^\circ \cdot (d_n - 81)}{364}$$

A *longitude*, em graus decimais, corresponde à localização do sistema solar fotovoltaico e é calculada através da equação 5.8.

$$longitude = LL - LS \quad (5.8)$$

onde, LL consiste na longitude local em graus decimais, e LS a referência (standard) da longitude, em graus decimais, do fuso horário local (positivo em direção ao oeste e negativa em direção ao leste do meridiano de Greenwich).

5.3 Perceção dos dados do Sistema Fotovoltaico seguidor

A Tabela 5.2 sintetiza os parâmetros utilizados para a criação dos modelos de previsão para o sistema solar fotovoltaico seguidor.

Tabela 5.2- Parâmetros usados para a criação dos modelos previsão do sistema fotovoltaico seguidor

Fonte da Informação	Parâmetro	Descrição
Sistema de Produção	Irradiação Solar	Irradiação solar (W/m^2) medida por sensor instalado no sistema
	Potência em corrente alternada	Potência de corrente alternada (W) produzida pelo sistema
	Temperatura ambiente	Temperatura ($^\circ C$) medida por sensor instalado no sistema
	Temperatura dos módulos	Temperatura ($^\circ C$) dos módulos fotovoltaicos
	Velocidade do vento	Velocidade do vento (m/s) medido por sensor instalado no sistema
OpenWeathermap.org +Wunderground.com	Temperatura ambiente	Temperatura local ($^\circ C$) com base na estações de perto do local, geralmente para um período de tempo de 1 hora
	Humidade	Concentração de vapor de água no ar em (%)
	Cobertura das nuvens	Cobertura percentual (%) de céu com nuvens
	Velocidade do vento	Velocidade do vento local (m/s)
	Irradiação (PVGIS)	Irradiação fornecida pela plataforma do Sistema de Informação Geográfica fotovoltaica

No caso da previsão da produção de energia para a monitorização, foram utilizados todos os dados disponíveis para o treino e os dados da 3ª semana de cada mês para o teste. Inicialmente durante esta fase foi realizado um estudo devido a um

sombreamento detetado durante o funcionamento do sistema seguidor que acontecia no início de cada dia (ver Anexo A para mais detalhes).

Na previsão da produção de energia a curto prazo foram usados todos os dados para a fase de treino e a semana de 4 a 13 de Março de 2015 para a fase de teste.

Para se fazer o estudo de produção de energia a médio prazo para o sistema solar fotovoltaico seguidor foi utilizada uma abordagem diferente para calcular a irradiação, pois neste caso não se pode usar a irradiação solar utilizada na abordagem a médio prazo para o sistema solar fotovoltaico estático, uma vez que a irradiação extraterrestre calculada para o sistema fixo é determinada de forma a que o sistema esteja em posição plana e para o caso do sistema solar fotovoltaico seguidor esses valores de irradiação induzem os algoritmos em erro. Desta maneira, optou-se por utilizar os dados da irradiação fornecidos pela plataforma do Sistema de Informação Geográfica fotovoltaica (em inglês *Photovoltaic Geographical Information System - PVGIS*). Nesta plataforma é possível obter a informação da irradiação solar mensal ou diária, tanto para sistemas solar fotovoltaicos estáticos como para sistemas solar fotovoltaicos seguidores.

5.4 Perceção dos dados do Sistema Eólico

A Tabela 5.3 sintetiza os parâmetros utilizados para a criação dos modelos de previsão para o sistema eólico.

Tabela 5.3- Parâmetros usados para a criação dos modelos previsão do sistema eólico

Fonte da Informação	Parâmetro	Descrição
Sistema de Produção	Potência em corrente alternada	Potência de corrente alternada (W) produzida pelo sistema
	Temperatura ambiente	Temperatura (°C) medida por o sensor instalado no sistema
	Velocidade do vento	Velocidade do vento (m/s) medido por sensor instalado no sistema

Na previsão da produção de energia para realizar a monitorização do sistema eólico foram considerados todos os dados disponíveis para a fase de treino e os dados da semana de 22 a 28 de Fevereiro de 2015 (identificada como sendo uma semana com muito vento) e da semana de 25 a 31 de Julho de 2014 (identificada como sendo uma

semana de pouco vento) para a fase de teste. Para o estudo do ano completo são usados os dados da última semana de cada mês para a fase de teste.

Na previsão da produção de energia para previsão a curto prazo foram usados todos os dados disponíveis para a fase de treino e os dados da semana de 22 a 28 de Fevereiro de 2015 para a fase de teste. A previsão a médio prazo não foi considerada pois por falta de tempo para um melhor desenvolvimento, e também devido ao valores obtidos a partir de sites de meteorologia não terem boa correlação com a potência em CA.

5.5 Utilização da Ferramenta *Weka*

No contexto do pré-processamento de dados, a Figura 5.3 ilustra o ambiente gráfico usado pelo *software Weka* para realizar as várias atividades incluídas no pré-processamento dos dados, que posteriormente serão processados através da aplicação dos algoritmos de previsão.

O conjunto de dados pode ser carregado para a plataforma através da opção *open file* (os dados devem estar no formato do tipo *.arff*). Após o carregamento, os dados ficarão visíveis na secção *Attributes*. Usando a opção *FILTER* pode-se escolher o método de tratamento dos dados que vai ser realizado, por exemplo remover os valores de produção inferiores a 1 W.

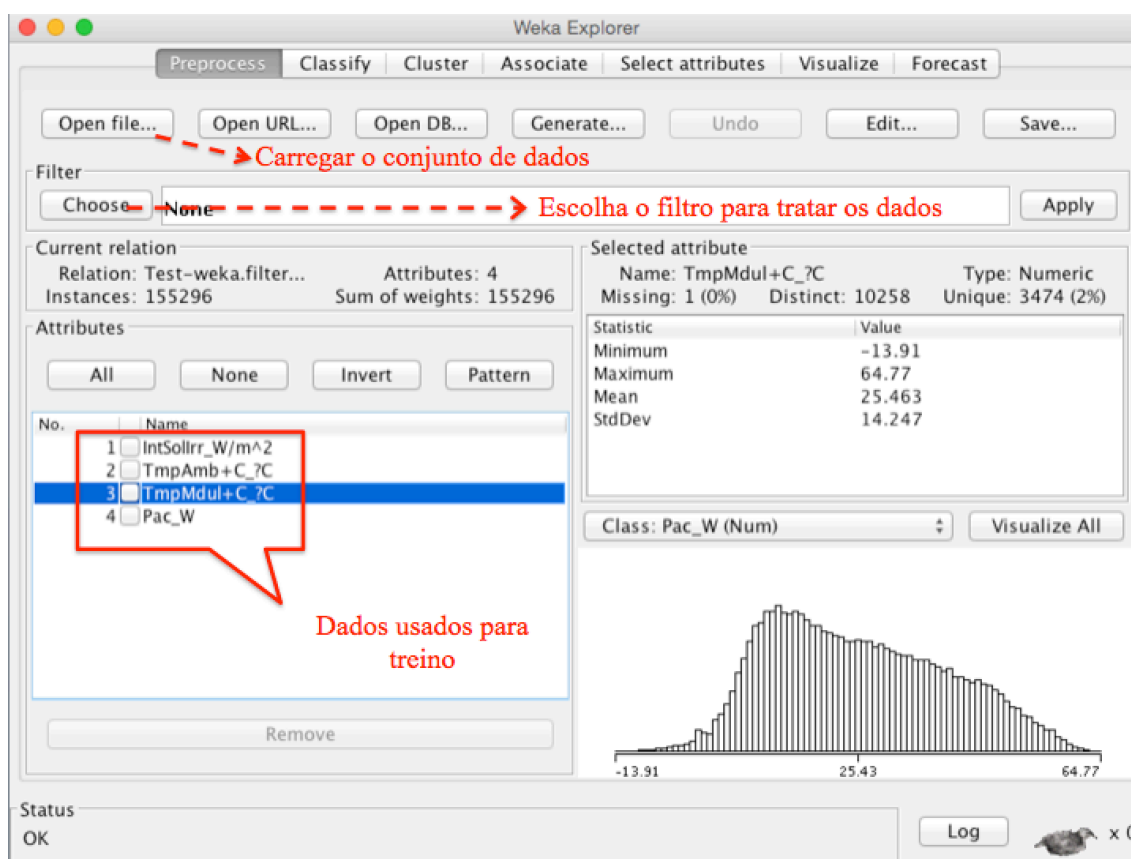


Figura 5.3-Preparação do conjunto de dados de treino no software Weka

A Figura 5.4 ilustra o ambiente gráfico usado pelo *software Weka* para a escolha dos algoritmos de previsão, o qual pode ser realizado através da secção *Classifier*. Neste trabalho, inicialmente foram testados vários algoritmos, mas no final apenas se consideraram os mais relevantes para todos os casos em análise, nomeadamente o M5P (algoritmo baseado em árvores de decisão), o MP - Multilayer Perception (algoritmo baseado em redes neuronais), o M5 Rules (algoritmo baseado em regras) e por fim o LR - Linear Regression (algoritmo baseado em regressão linear).

A opção *supplied test set* permite carregar os dados de treino. A avaliação do modelo aplicando a um determinado algoritmo de previsão ao conjunto de dados de entrada pode ser visualizada na opção *Classifier Output*.

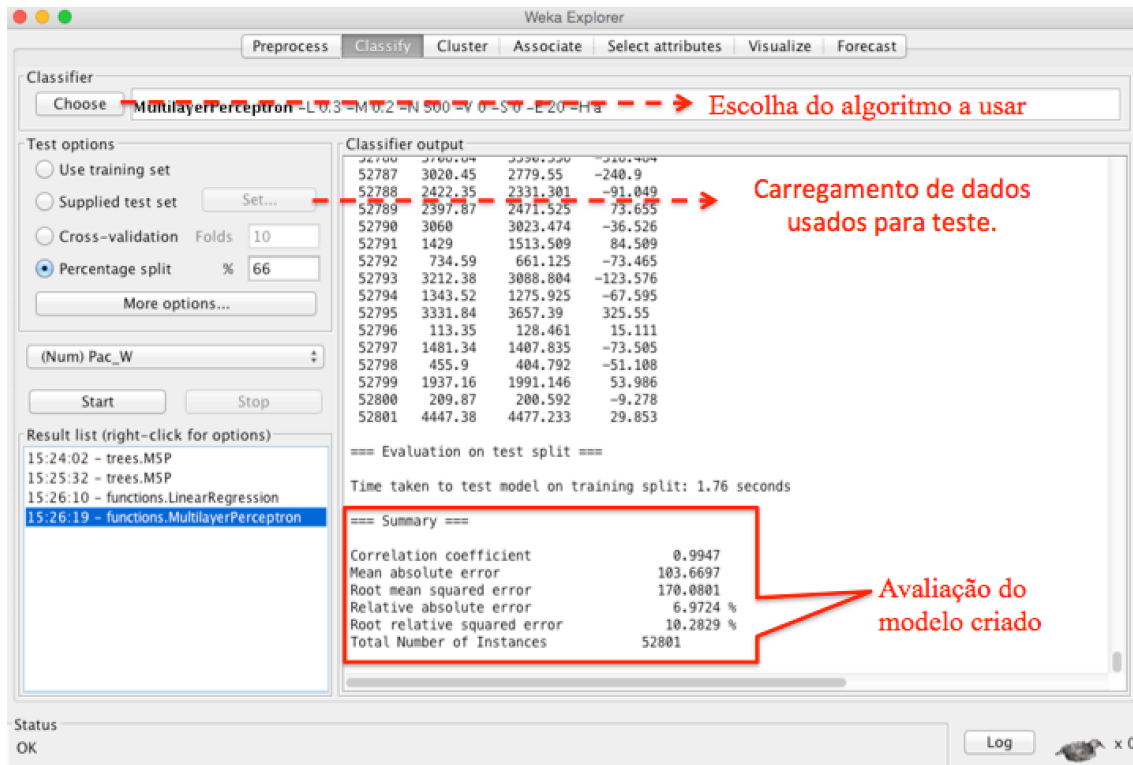


Figura 5.4-Ambiente gráfico para aplicação dos algoritmos

No capítulo que se segue será realizada a análise de resultados após a aplicação dos algoritmos de análise de dados para as diferentes situações definidas na apresentação do caso de estudo.

6 Análise de Resultados

No que diz respeito à análise de resultados relativamente à criação de modelos de previsão para a produção de energia através de fontes renováveis, foram analisadas diferentes abordagens, nomeadamente considerando diferentes algoritmos com diferentes conjuntos de dados, usando a ferramenta *WEKA*. Os resultados para cada uma das abordagens são analisados tendo em consideração o erro entre os dados reais da produção elétrica e os valores previstos.

6.1 Sistema solar fotovoltaico estático

6.1.1 Monitorização da operação do sistema solar fotovoltaico estático

Diferentes experimentações foram realizadas para determinar a melhor configuração para a construção de modelos de previsão: considerando todos os dados disponíveis, e apenas os dados dos últimos meses anteriores. Estas experiências também foram realizadas considerando a combinação de diferentes atributos para identificar qual deles contribuir para a produção da energia do sistema fotovoltaico. Os resultados mostraram que a produção de energia fotovoltaica estática é afetada principalmente pela irradiação solar e temperatura ambiente e dos módulos, enquanto que a velocidade do vento não apresenta uma contribuição significativa.

Neste modelo de previsão, a produção de energia prevista é comparado em tempo real com o valor medido de forma convincente. As diferenças apresentadas por esses valores, ao longo de um determinado período, irá indicar possíveis problemas na operação do sistema fotovoltaico, que podem estar diretamente relacionadas com a unidade de produção de energia, com o sombreamento de algumas células fotovoltaicas ou mesmo com o mau funcionamento do sistema de deteção.

As Tabela 6.1 e Tabela 6.2 resumem os resultados da aplicação dos diferentes algoritmos de previsão para uma semana inteira de períodos de sol e Tabela 6.2 para períodos nublados respetivamente [43].

Tabela 6.1-Modelos de previsão para o sistema solar fotovoltaico, caso de monitorização de semana sol

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	8.97	128.65	0.13	39.65
Coefficiente de correlação	0.998	0.998	0.997	0.998
Erro médio absoluto	71.669	72.154	120.597	87.871
Erro médio quadrático	101.337	102.072	150.396	116.071
Erro absoluto relativo (%)	4.491	4.522	7.558	5.507
Raiz quadrada do erro médio quadrático (%)	5.653	5.694	8.390	6.475

Como seria de esperar, os modelos de previsão para a semana de sol apresentam valores mais elevados para o coeficiente de correlação, o que indica que os modelos atingidos apresentam maior qualidade.

Tabela 6.2- Modelos de previsão para o sistema solar fotovoltaico, caso de monitorização de semana nublada

	<i>M5P</i>	<i>M5 Rules</i>	<i>LR</i>	<i>MP</i>
Tempo de construção (s)	6.25	149.34	0.13	41.76
Coefficiente de correlação	0.991	0.990	0.989	0.990
Erro médio absoluto	69.400	70.610	114.791	92.836
Erro médio quadrático	133.731	136.544	166.841	154.865
Erro absoluto relativo (%)	4.6357	4.7165	7.6676	6.2011
Raiz quadrada do erro médio quadrático (%)	8.462	8.640	10.558	9.800

Mesmo sendo uma semana nublada os resultados foram bons, pois o coeficiente de correlação continua alto, no entanto os erros obtidos tiveram um aumento em comparação com a análise feita para uma semana de sol, tal como é ilustrado na Figura 6.1.

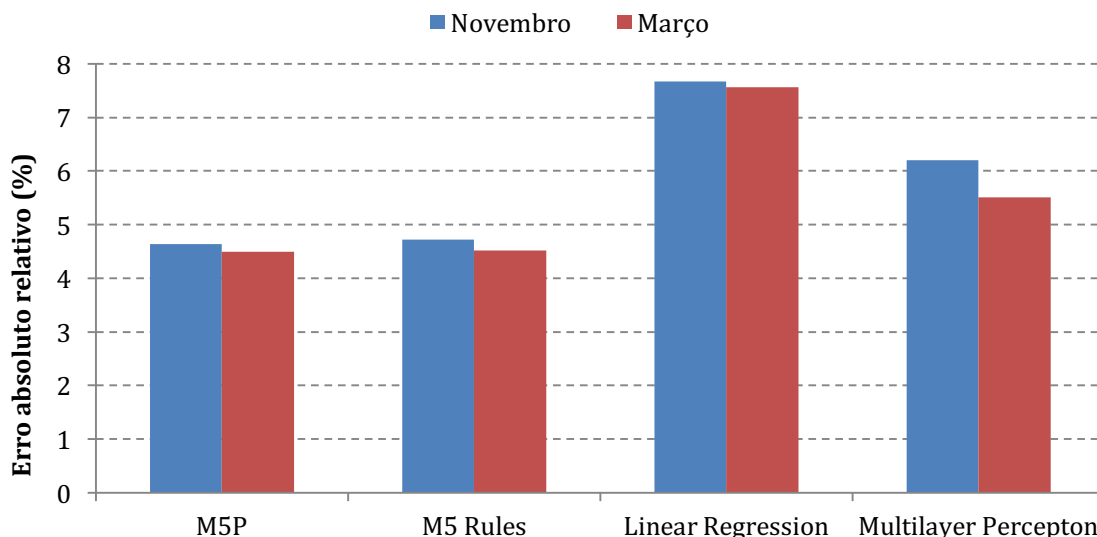


Figura 6.1-Comparação do erro absoluto relativo para períodos de sol e nublados

A análise do coeficiente de correlação e do erro relativo permite verificar que os melhores resultados foram obtidos usando o algoritmo de previsão M5P.

Na Figura 6.2 é possível verificar os resultados obtidos para uma abordagem considerando dados para o período de um ano, usando o algoritmo M5P (no Anexo C encontram-se mais detalhes sobre os resultados obtidos).

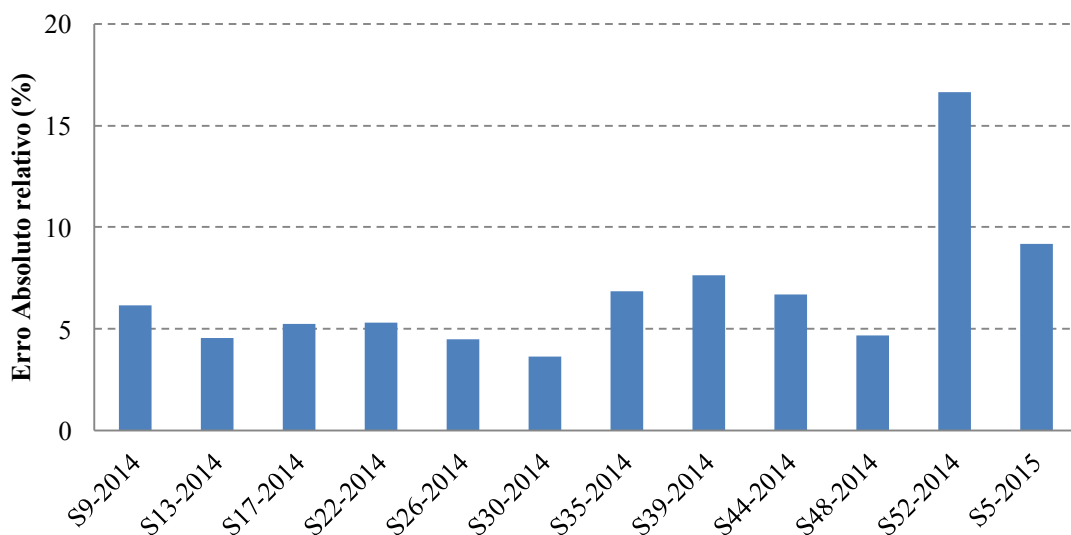


Figura 6.2- Monitorização do sistema fotovoltaico estático usando o M5P para um período de uma ano.

Em termos de erro absoluto, os piores resultados obtidos foram a semana 52 e a semana 5, em que estas semanas são referentes a semanas de inverno, onde os dias são muito nublados e com pouca estabilidade climática (a semana 52 é referente ao mês de

dezembro e a semana 5 é referente ao mês de Janeiro), o que torna mais complicado a previsão por parte do algoritmo

6.1.2 Previsão da produção de energia a curto prazo

A análise para criação dos modelos de previsão a curto prazo considera a previsão para intervalos de tempo de 5 minutos e de 15 minutos à frente. A Tabela 6.3 representa os resultados obtidos para um período de 5 minutos.

Tabela 6.3- Comparação dos algoritmos para um período de tempo de 5 minutos.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	8.9	121.63	0.17	41.39
Coefficiente de correlação	0.991	0.991	0.990	0.990
Erro médio absoluto	171.294	167.035	202.653	245.011
Erro médio quadrático	253.418	249.554	271.193	319.470
Erro absoluto relativo (%)	10.899	10.628	12.895	15.590
Raiz quadrada do erro médio quadrático (%)	14.316	14.098	15.320	18.048

A análise dos resultados permite observar que o coeficiente de correlação é elevado, o que nos indica que os modelos de previsão estão correlacionados. Em relação ao erro médio absoluto, o algoritmo M5 Rules é o que apresenta valor mais baixo, o que significa que o algoritmo prevê mais ou menos 167W do que a produção de energia real. Em relação à abordagem do ano completo, descrita no Anexo C, o valor médio do erro absoluto relativo é de 6.760 %. Para um período de 5 minutos o melhor resultado tem um acréscimo de 4%, o que é um bom resultado visto usar valores previstos. A Tabela 6.4 representa os resultados obtidos para um período de 5 minutos.

Tabela 6.4- Comparação dos algoritmos para previsão de um período de tempo de 15 minutos.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	7.16	143.84	0.13	41.08
Coefficiente de correlação	0.977	0.977	0.975	0.976
Erro médio absoluto	247.106	245.413	265.978	287.846
Erro médio quadrático	364.748	362.191	377.068	417.296
Erro absoluto relativo (%)	15.723	15.616	16.924	18.316
Raiz quadrada do erro médio quadrático (%)	20.605	20.461	21.301	23.574

A análise dos resultados obtidos permite verificar que o coeficiente de correlação baixou em relação ao período de tempo de 5 minutos e em termos de erro médio absoluto um pouco mais elevado. Isto permite concluir que quanto maior for o tempo para previsão maior será o erro entre a potência em CA real e a potência em CA prevista, como seria expectável.

A Figura 6.3 ilustra a comparação gráfica em relação ao erro absoluto relativo para o período de tempo de 5 minutos e para o período de tempo de 15 minutos.

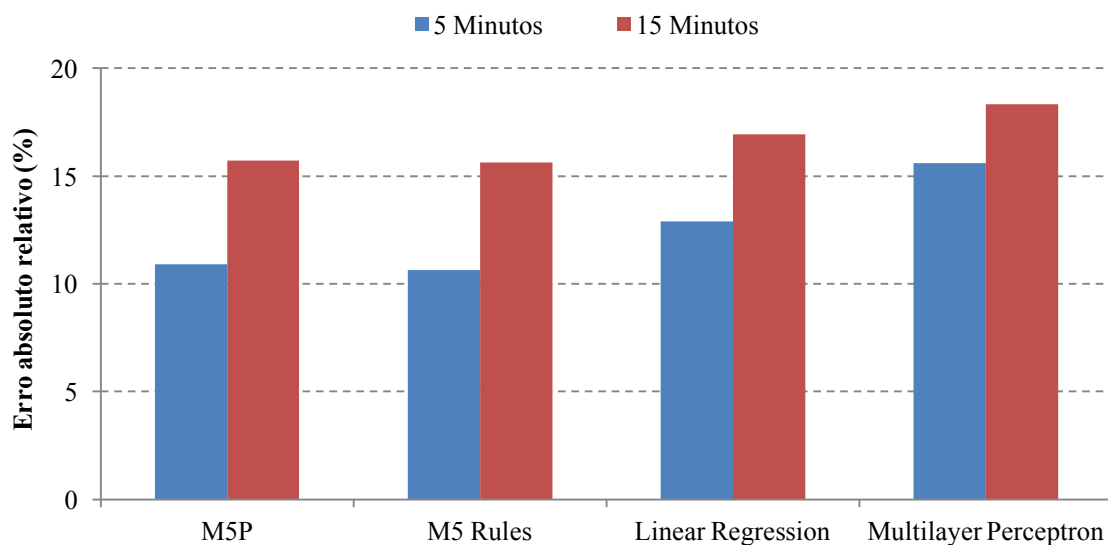


Figura 6.3-Comparação do erro absoluto relativo em relação a potência CA para previsão de um período de tempo de 5 minutos e 15 minutos.

A análise gráfica permite concluir que o melhor algoritmo para fazer uma previsão para ambos os tempos é o M5 Rules, onde são atingidos os melhores resultados, quer em coeficiente de correlação quer o erro médio absoluto. No entanto, este algoritmo apresenta um elevado tempo de processamento para criação do modelo de previsão, quando comparado com os restantes métodos.

A Figura 6.4 ilustra os valores da previsão da potência para os modelos de previsão de 5 minutos e 15 minutos, assim como para a potência real. A observação das curvas permite verificar que os valores previstos ficam próximos dos valores reais o que indica uma boa qualidade dos modelos de previsão [43].

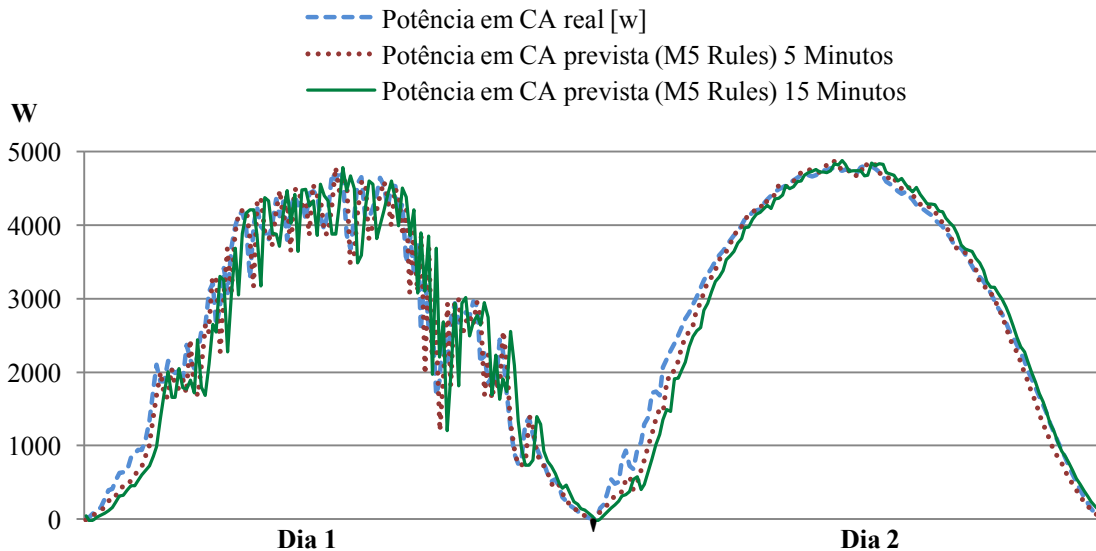


Figura 6.4-Comparação das previsões usando o M5 Rules com potência CA real, para um período de tempo de 5 e 15 minutos (dois dias).

No entanto, o erro associado nestas duas situações é significativo, tal como é ilustrado na Figura 6.5 erro percentual entre as previsões do algoritmo M5 Rules para ambos os períodos de previsão tempo e a potência em CA real. Como é facilmente verificável, em algumas situações do dia o erro é bastante elevado, particularmente ao início e fim do dia. Isto devido o uso de parâmetros estimados, e o algoritmo apresenta alguma dificuldade no reconhecimento da mudanças de dias

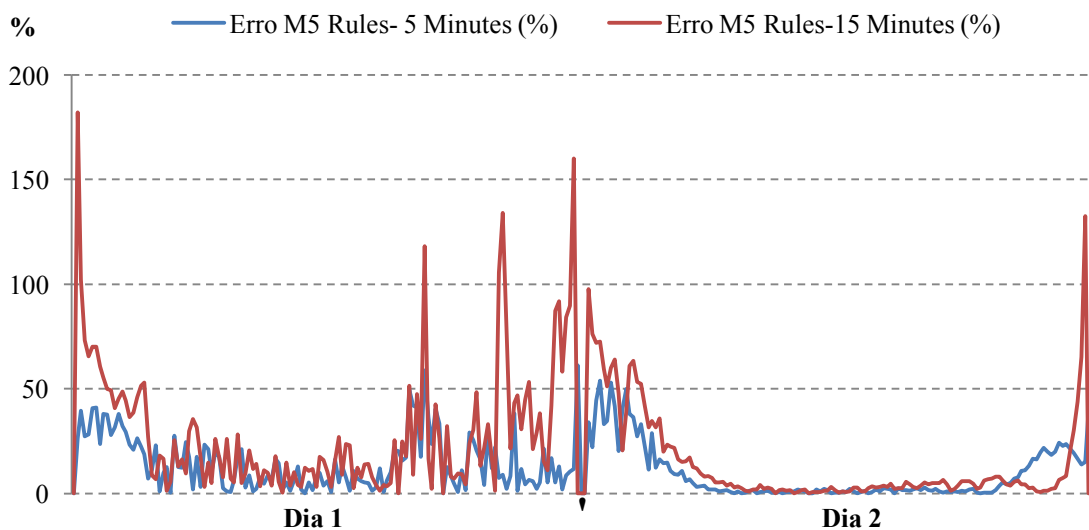


Figura 6.5- Erro percentual para períodos de 5 e 15 minutos entre a potência em CA prevista e a real (dois dias).

6.1.3 Previsão da produção de energia a médio prazo

Para criação do modelo de previsão a médio prazo foram realizados vários testes com vários parâmetros até obter os melhores resultados, os quais estão representados na Tabela 6.5.

Tabela 6.5-Comparação dos algoritmos para previsão a médio prazo.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	0.4	1.4	0.01	1.83
Coefficiente de correlação	0.952	0.952	0.859	0.926
Erro médio absoluto	436.574	429.147	837.181	621.396
Erro médio quadrático	564.695	566.175	961.330	712.293
Erro absoluto relativo (%)	25.997	25.555	49.853	37.003
Raiz quadrada do erro médio quadrático (%)	29.858	29.936	50.830	37.662

A análise destes resultados permite verificar o aumento do erro médio absoluto, o que significa que a potência em CA prevista se distânciava da potência em CA real. Uma causa para este acontecimento foi facto de usar dados de *sites* de previsão meteorológica, os quais já tem um erro associado (pois são previsões). A utilização de um valor calculado para a irradiação extraterrestre calculada também tem influência pois existe um erro associado a este calculo devido à presença de nuvens, sombreamentos, entre outros que podem também influenciar o valor real da irradiação. Contudo, os resultados foram satisfatórios e o melhor algoritmo para aplicar na previsão a médio prazo é o M5 Rules. A Figura 6.6 ilustra as curvas da potência prevista e real [43].

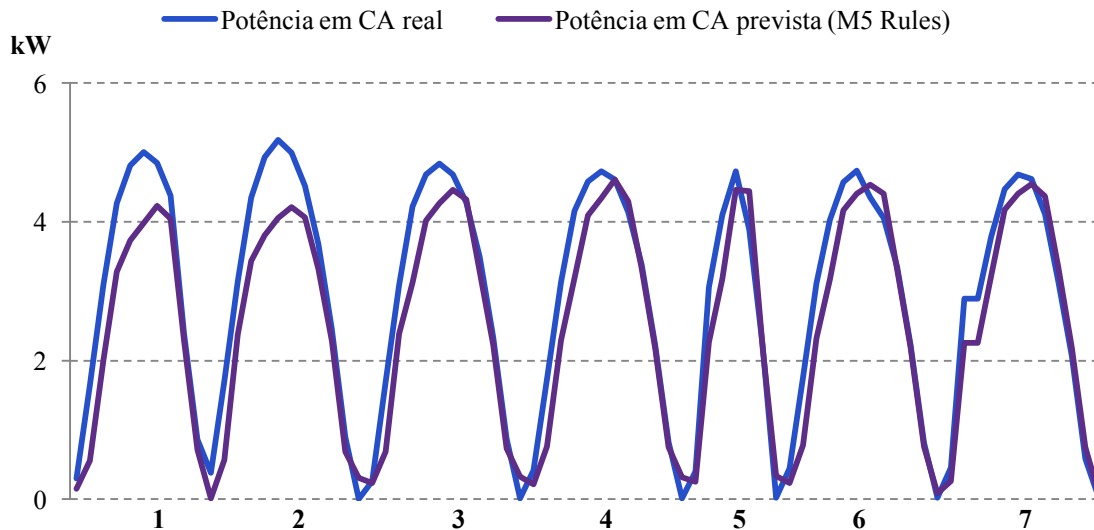


Figura 6.6-Previsão a médio prazo da potência em CA prevista usando M5Rules (sete dias).

A análise da Figura 6.6 permite afirmar que no geral a linha referente aos valores da potência em CA prevista está abaixo da linha correspondente aos valores da potência em CA real. No entanto, é possível verificar que a maior dificuldade do algoritmo é na previsão até ao meio do dia, pois o gráfico é referente a sete dias e cada curva representa um dia, o que leva a um erro elevado na previsão. Ao fim do dia os valores previstos já são idênticos aos valores reais.

A Figura 6.7 ilustra o erro entre a potência em CA prevista e a potência em CA real em watts, para dar uma perceção da diferença em watts entre os dois valores

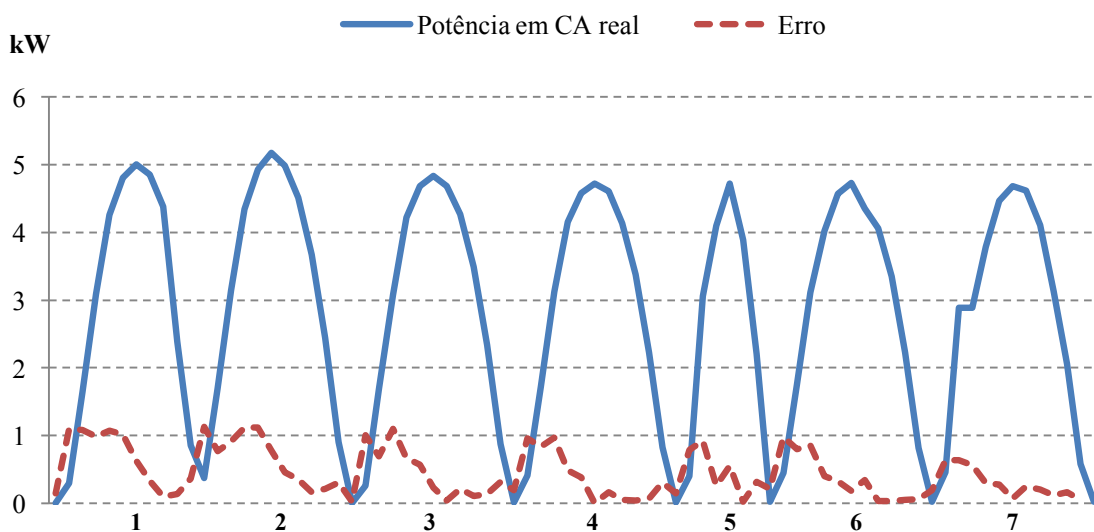


Figura 6.7-Erro absoluto entre a potência em CA prevista e a real (sete dias)

Como podemos observar os valores mais altos do erro no início do dia, até ao ponto de potencia máxima por dia, uma vez que o algoritmo não consegue prever tão claramente para a previsão a médio prazo devido ao facto dos valores serem medidos de hora em hora, o que torna mais difícil a previsão.

6.2 Sistema fotovoltaico seguidor

No que diz respeito à análise de resultados relativamente à criação de modelos de previsão para a produção de energia do sistema fotovoltaico seguidor Os resultados para cada uma das abordagens serão analisados tendo em consideração o erro entre os dados reais da produção elétrica e os valores previstos.

6.2.1 Monitorização da operação do sistema fotovoltaico seguidor

Considerando o conhecimento adquirido no sistema fotovoltaico estático foi aplicado o mesmo modelo de previsão a fim de obter informação para a monitorização do sistema fotovoltaico seguidor (no Anexo D encontram-se mais detalhes os resultados obtidos). Na Figura 6.8 representa o erro absoluto relativo dos resultados obtidos usando o algoritmo MSP para o período de um ano. Os resultados são baixos, com exceção de alguns meses que são difíceis de prever como é o caso da semana 3 de 2015, que é relativo ao mês de janeiro. Uma possível causa para a ocorrência deste erro poderá ser o facto de não haver os dados relativamente ao mês de Janeiro do ano anterior, o que faz com que o algoritmo não reconheça o comportamento para o ano de 2015.

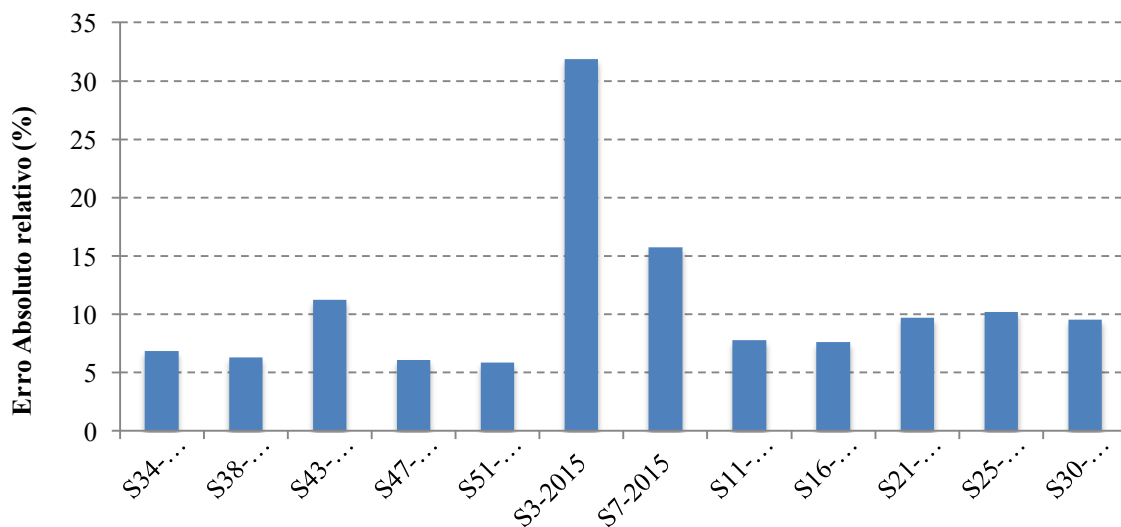


Figura 6.8- Monitorização do sistema fotovoltaico seguidor usando o algoritmo MSP para o período de um ano.

No entanto, relativamente aos resultados obtidos, o coeficiente de correlação continua elevado, o que indica que os valores previstos seguem o mesmo comportamento dos valores reais. O erro médio absoluto é baixo, o que significa que é possível aplicar o algoritmo M5P para previsão da produção de energia para realizar a monitorização do sistema fotovoltaico estático.

6.2.2 Previsão da produção de energia a curto prazo

O objetivo desta previsão é tentar criar uma nova previsão a curto prazo mas com uma elevada precisão, isto devido também aos parâmetros de temperatura ambiente e irradiação solar que não apresentam uma grande variação em intervalos de tempos curtos. A Tabela 6.6 ilustra os resultados para um período de tempo 5 minutos.

Tabela 6.6- Comparação dos algoritmos para previsão 5 minutos para seguidor solar.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	5.42	100.07	21.68	25.24
Coefficiente de correlação	0.947	0.947	0.947	0.959
Erro médio absoluto	173.814	173.514	161.612	138.274
Erro médio quadrático	227.822	226.955	221.614	185.114
Erro absoluto relativo (%)	17.421	17.391	16.198	13.859
Raiz quadrada do erro médio quadrático (%)	21.760	21.678	21.167	17.681

Fazendo uma análise aos resultados para a previsão de um período de tempo de 5 minutos, podemos concluir que os coeficientes de correlação são elevados nos quatro algoritmos, o que significa que os parâmetros de entrada têm uma boa correlação com o parâmetro que o algoritmo pretende prever (isto é a potência). Em termos de erro médio absoluto os valores são baixos, o que é bom e significa que em média os valor previstos andam perto dos valores reais.

Dos quatro algoritmos testados, os algoritmos que se destacam são o LR e o MP . Em relação ao erro médio absoluto para previsão curto prazo para um período de 5 minutos do sistema fotovoltaico estático o algoritmo que apresentou melhores resultados foi o M5 Rules. Em relação ao erro absoluto relativo, o melhor resultado é obtido pelo MP. A Tabela 6.7 ilustra os resultados para um período de tempo de 15 minutos.

Tabela 6.7- Comparação dos algoritmos para uma previsão de um período de tempo de 15 minutos para o seguidor solar

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	5.45	23.28	23.28	22.75
Coefficiente de correlação	0.936	0.937	0.937	0.949
Erro médio absoluto	180.495	181.070	170.996	147.730
Erro médio quadrático	244.018	243.987	238.431	206.204
Erro absoluto relativo (%)	18.067	18.125	17.116	14.787
Raiz quadrada do erro médio quadrático (%)	23.279	23.276	22.746	19.672

Em relação à previsão para um período de tempo de 15 minutos, os resultados obtidos são um pouco mais altos do que nos resultados para o período de tempo de 5 minutos, o que é normal porque quanto mais afastado for da previsão, maior vai ser o erro. Porém, no geral os coeficientes de correlação estão acima dos 0.935, o que quer dizer que os valores previstos seguem o mesmo comportamento do valores reais. Já em relação ao erro médio absoluto e erro absoluto relativo, o algoritmo que obteve melhor resultado foi o MP.

A Figura 6.9 representa a comparação em relação ao erro médio absoluto para um período de tempo de 5 minutos e de 15 minutos, sendo possível afirmar que o modelo de previsão para 5 minutos apresenta menores erros que o modelo de previsão para 15 minutos, e que o melhor algoritmo para ambos os casos é o MP.

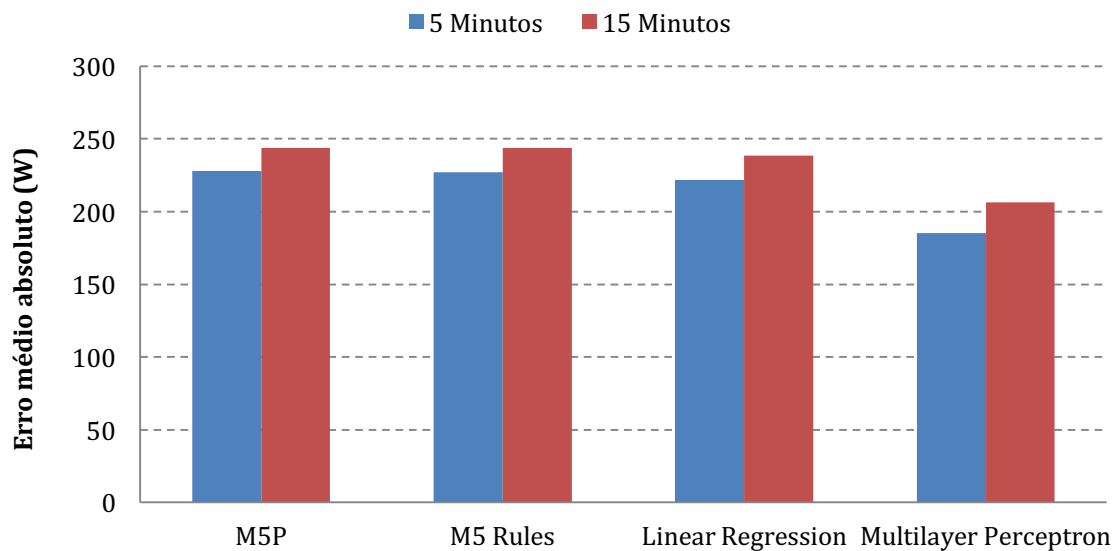


Figura 6.9- Comparação do erro médio absoluto da potência em CA prevista para um período de 5 e 15 minutos

A Figura 6.10 permite comparar o comportamento do algoritmo MP ao criar a previsão para um período de tempo de 5 e 15 minutos e a potência real.

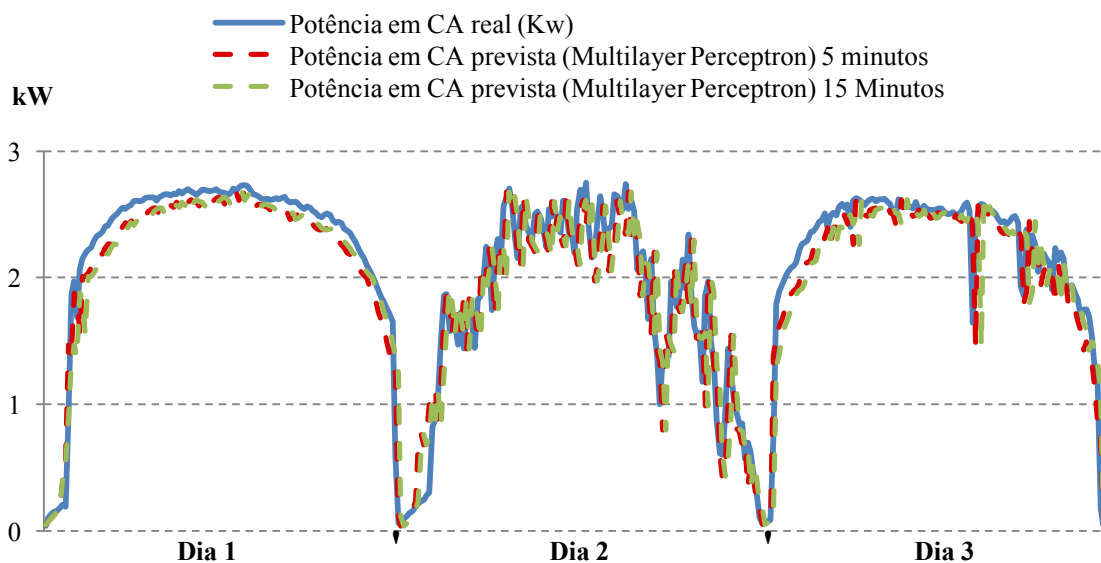


Figura 6.10- Comparação do algoritmo MP para um período de 5 e 15 minutos para previsão da potência em CA (três dias).

Os valores de previsão para ambos os tempos de previsão aproximam-se, em valores absolutos, dos valores da potência em real.

6.2.3 Previsão da produção de energia a médio prazo

A Tabela 6.8 sintetiza os resultados obtidos para a previsão de produção de energia do sistema fotovoltaico seguidor considerando o cenário de médio prazo.

Tabela 6.8- Comparação dos algoritmos para previsão a médio prazo para o sistema fotovoltaico seguidor.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	0.3	0.74	0.01	1.34
Coefficiente de correlação	0.696	0.783	0.830	0.804
Erro médio absoluto	535.059	473.757	526.194	479.515
Erro médio quadrático	703.466	604.735	622.930	702.883
Erro absoluto relativo (%)	49.205	43.567	48.389	44.097
Raiz quadrada do erro médio quadrático (%)	60.193	51.745	53.302	60.143

Como se pode verificar, os resultados do coeficiente de correlação são baixos, ou seja, os valores andam na ordem dos 0.70, o que significa que os modelos de previsão não são muito fiáveis. Em relação ao erro médio absoluto, os valores são mais altos, o que quer dizer que a potência prevista afasta-se um pouco da potência real. Mais uma vez, uma razão para este facto é a utilização de dados de *sites* de previsão meteorológica, uma vez que estes dados contêm um erro associado (valores previstos). Em relação ao erro absoluto os valores também são altos. Uma das possíveis causas para este acontecimento será a utilização de previsões de hora a hora o que não garantem precisão. Neste caso, o melhor algoritmo é o MP.

Como é possível observar na Figura 6.11, a potência prevista usando o algoritmo MP não consegue aproximar-se da potência real, pois a previsão a médio prazo usa dados medidos de hora a hora.

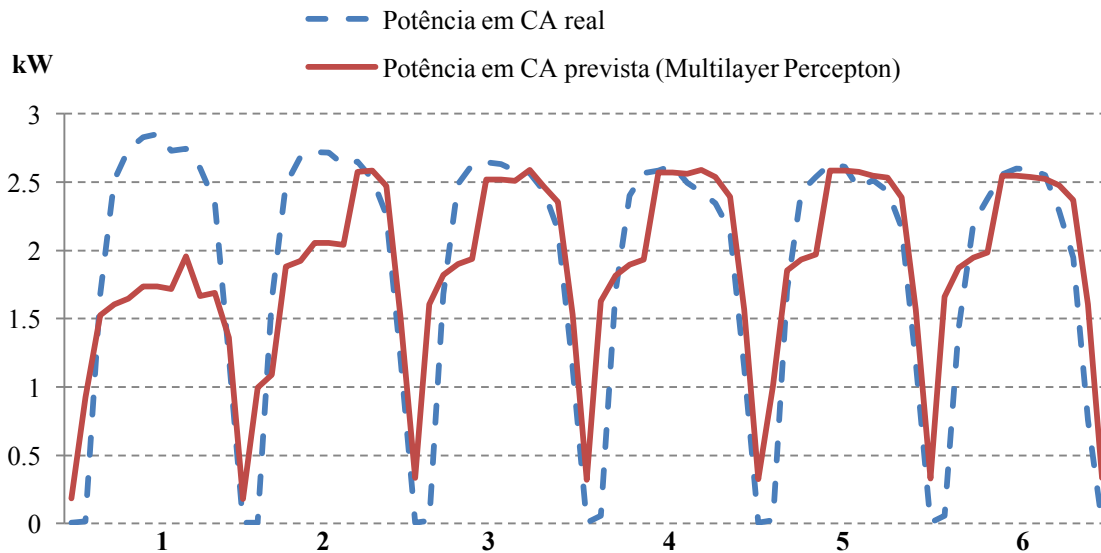


Figura 6.11- Previsão a médio prazo da potência em CA prevista usando Multilayer Perceptron (6 dias).

6.3 Sistema eólico

Em relação ao sistema eólico e já com alguma falta de tempo para um melhor desenvolvimento, foram executados alguns testes para obter os melhores modelos de previsão. Foi realizado um estudo para a monitorização, a previsão a curto prazo e avarias que surgiram durante o estudo do sistema.

6.3.1 Monitorização da operação do sistema eólico.

De maneira a fazer a previsão da produção de energia a fim de obter informação para a monitorização ao sistema eólico, juntamente com conhecimento adquirido nos experimentos realizados para os sistemas fotovoltaicos, foi realizado um estudo para a previsão de uma semana com muito e com pouco vento.

A Tabela 6.9 sumaria os resultados obtidos dos modelos de previsão para uma semana com muito vento e a Tabela 6.10 sumaria os resultados obtidos dos modelos de previsão para uma semana com pouco vento. A análise destes resultados permite verificar que os erros são um pouco mais elevados pois o sistema eólico é um sistema mais complexo de prever isto devido a variação da velocidade do vento.

Tabela 6.9-Resultados para o período de uma semana com muito vento

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	5.01	22.66	0.06	10.49
Coefficiente de correlação	0.892	0.893	0.891	0.893
Erro médio absoluto	237.622	238.567	261.853	243.923
Erro médio quadrático	343.989	344.397	354.377	355.372
Erro absoluto relativo (%)	38.157	38.309	42.048	39.169
Raiz quadrada do erro médio quadrático (%)	44.125	44.177	45.457	45.585

Tabela 6.10- Resultados para o período de uma semana com pouco vento

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	6.8	33.48	0.07	12.17
Coefficiente de correlação	0.848	0.846	0.804	0.856
Erro médio absoluto	95.081	95.668	160.579	91.608
Erro médio quadrático	138.452	139.379	196.480	134.672
Erro absoluto relativo (%)	23.566	23.711	39.799	22.705
Raiz quadrada do erro médio quadrático (%)	32.030	32.244	45.454	31.155

A Figura 6.12 compara os resultados obtidos para as duas semanas, sendo possível verificar que o erro absoluto relativo é menor no mês de julho (que tem pouco vento).

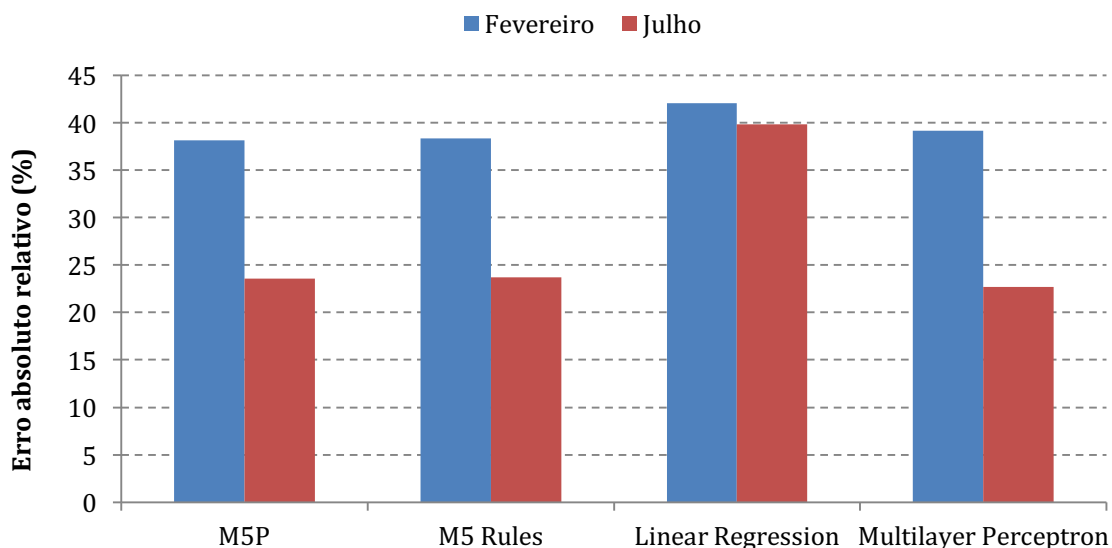


Figura 6.12- Comparação do erro absoluto relativo para períodos com muito e pouco vento.

Usando apenas os algoritmos que apresentaram melhores resultados nos casos anteriores (isto é M5P e M5Rules), foi realizado um estudo para o ano completo. A Figura 6.13 ilustra os resultado do algoritmo M5 Rules (mais detalhe sobre este estudo pode ser encontrado no Anexo E).

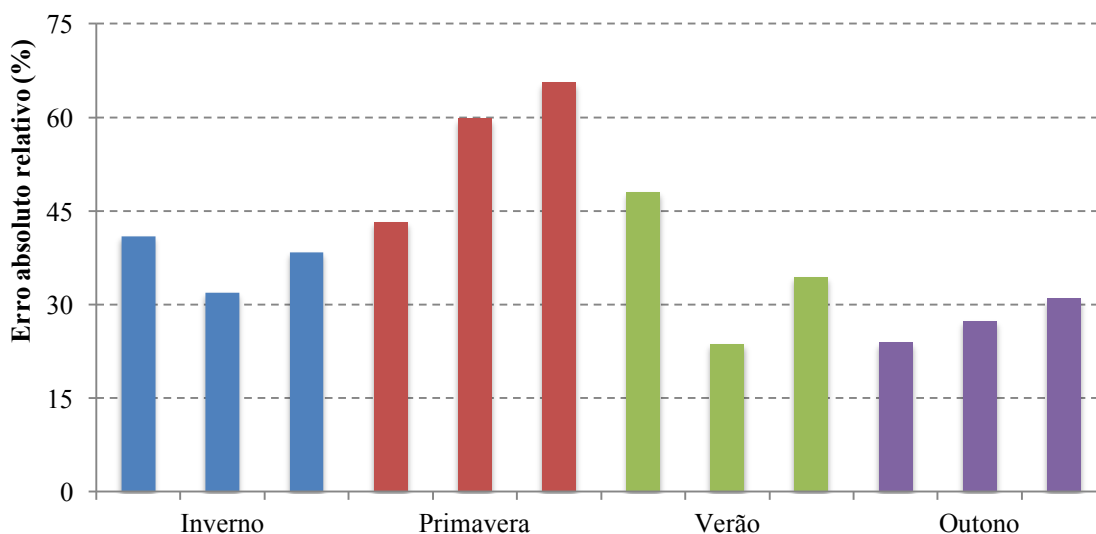


Figura 6.13- Monitorização da potência em CA do sistema eólico usando M5 Rules por estação do ano.

Este estudo permite verificar que, em termos de erro absoluto relativo, a Estação do ano que apresenta melhores resultados é o Outono. Uma possível causa para este acontecimento será estes meses não conterem muita informação sobre o vento. O mês de Julho apresenta o melhor resultado pois é um mês que contém poucos dados sobre vento, tornando fácil de prever.

A estação do ano com pior resultado foi a primavera, com um valor mais alto, isto também devido ao facto de ser uma estação com pouco vento, o que faz com que seja mais difícil para o algoritmo reconhecer algum padrão na produção da energia por parte do sistema eólico. Contudo, no geral para o sistema eólico, os resultados são razoáveis pois é um sistema difícil de prever devido ao facto dos dados não conterem valores tão lineares como no caso do sistema solares fotovoltaicos.

6.3.2 Previsão da produção de energia a curto prazo

De forma similar ao realizado nos outros sistemas, neste foi realizada a mesma abordagem de previsão da produção de energia a curto prazo, em que é feita a previsão para um período de tempo para 5 e 15 minutos. A Tabela 6.11 representa os resultados obtidos para um período de tempo de 5 minutos.

Tabela 6.11- Comparação dos algoritmos para previsão de períodos de 5 minutos para o sistema eólico.

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	7.91	55.78	0.09	15.03
Coefficiente de correlação	0.839	0.841	0.837	0.840
Erro médio absoluto	288.829	287.592	319.326	302.649
Erro médio quadrático	368.877	367.626	392.489	373.357
Erro absoluto relativo (%)	40.252	40.080	44.502	42.178
Raiz quadrada do erro médio quadrático (%)	42.194	42.051	44.895	42.707

Após uma análise à Tabela 6.11, em relação ao erro médio absoluto os valores são altos, o que significa que os valores previstos se aproximam da realidade. Dos quatro algoritmos testados, os que apresentam melhores resultados são M5P e o M5 Rules.. Os valores mais elevados para o erro absoluto relativo E deve-se ao facto do parâmetro com mais importância ser a velocidade do vento e este parâmetro apresentar uma grande instabilidade.

Para uma análise mais precisa do resultado para um período de tempo de 5 minutos é possível comparar graficamente, na Figura 6.14, a potência real e a potência prevista pelo algoritmo M5 Rules.

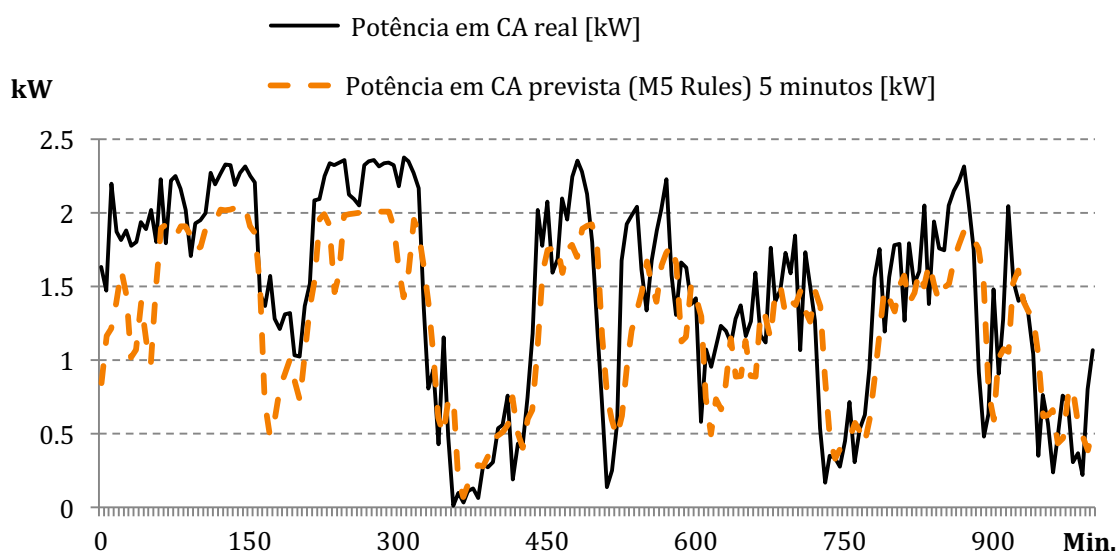


Figura 6.14- Comparação do algoritmos M5 Rules para 5 minutos.

É de notar que existem momentos onde a previsão tem valores abaixo da potência real mas em termos comportamentais, a potência prevista tem a mesma forma que a potência real. Contudo, sendo que uma previsão já apresenta um erro associado, e como os valores da velocidade do vento e da temperatura ambiente também foram estimados, leva a que estes erros conduzam a fracos modelos de previsão.

A Tabela 6.12 representa os resultados obtidos para um período de tempo de 15 minutos.

Tabela 6.12- Comparação dos algoritmos para previsão de períodos de 15 minutos para o sistema eólico

	M5P	M5 Rules	LR	MP
Tempo de construção (s)	8	56.94	0.08	16
Coefficiente de correlação	0.779	0.781	0.780	0.779
Erro médio absoluto	330.957	330.286	351.672	341.371
Erro médio quadrático	423.701	422.569	434.013	425.762
Erro absoluto relativo (%)	46.128	46.034	49.015	47.579
Raiz quadrada do erro médio quadrático (%)	48.463	48.334	49.643	48.699

Como na abordagem anterior foi realizada a previsão para um período de tempo de 15 minutos, em que foi obtido um coeficiente de correlação um pouco mais baixo do que a previsão para o período de tempo de 5 minutos. De facto, os valores desta

previsão andavam na ordem dos 0.80 e na previsão para um período de tempo de 15 minutos os valores andam na ordem dos 0.70. Esta situação é considerada normal uma vez que quanto mais afastado for a previsão, maior será o erro associado aos resultados.

Em relação ao erro médio absoluto os valores têm um comportamento igual ao da previsão para um período de tempo de 5 minutos. Dos algoritmos usados para testar os que obtiveram melhores resultados foram o M5P e o M5 Rules. Os resultados são mais altos do que a previsão para um período de tempo de 5 minutos, em que o melhor resultado para este período é o M5 Rules. Em termos de erro absoluto relativo os melhores resultados obtidos foram o M5P e o M5 Rules.

A Figura 6.15 ilustra a comparação entre a potência real e a potência prevista usando o algoritmo M5Rules para um período de tempo de 15 minutos.

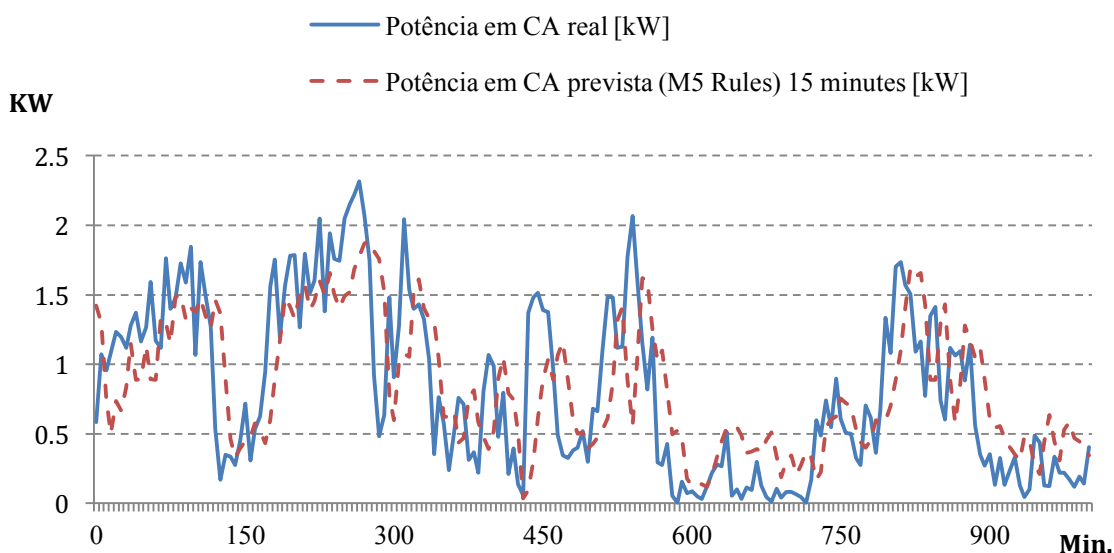


Figura 6.15- Comparação do algoritmo M5 Rules para um período de tempo de 15 minutos em relação a potência em CA real.

Na Figura 6.15 é visível que a potência prevista fica abaixo da potência real mas isso é um erro que pode ser melhorado fazendo um ajuste percentual. Em termos comportamentais, os valores previsto têm a mesma forma que à potência real, o que pode ser considerado com um resultado bastante satisfatório, uma vez que a produção eólica não é tão linear como é o caso dos sistemas solares fotovoltaicos, devido ao grau de incerteza na previsão do vento (velocidade e direção) .

6.3.3 Análise para caso de avaria

Inicialmente, a análise mensal do sistema eólico deparou-se com alguns problemas, em que existiam situações onde o vento não era medido mesmo sabendo-se que havia produção de energia ou o caso de existir vento e não existir produção de energia. Como exemplo, na semana de 27 a 31 de Dezembro de 2014, o sistema eólico não produziu energia, devido à existência de algum problema, ou até devido à uma falha de ligação a rede.

Assim sendo, procedeu-se à previsão dos restantes parâmetros medidos, isto é a velocidade do vento e a temperatura ambiente. Esta análise pode ser bastante interessante ao ponto de prever se o sistema está produzindo corretamente e ajudar a prever algumas avarias na produção. O algoritmo utilizado foi o M5 Rules e os resultados obtidos encontram-se ilustrados Figura 6.16.

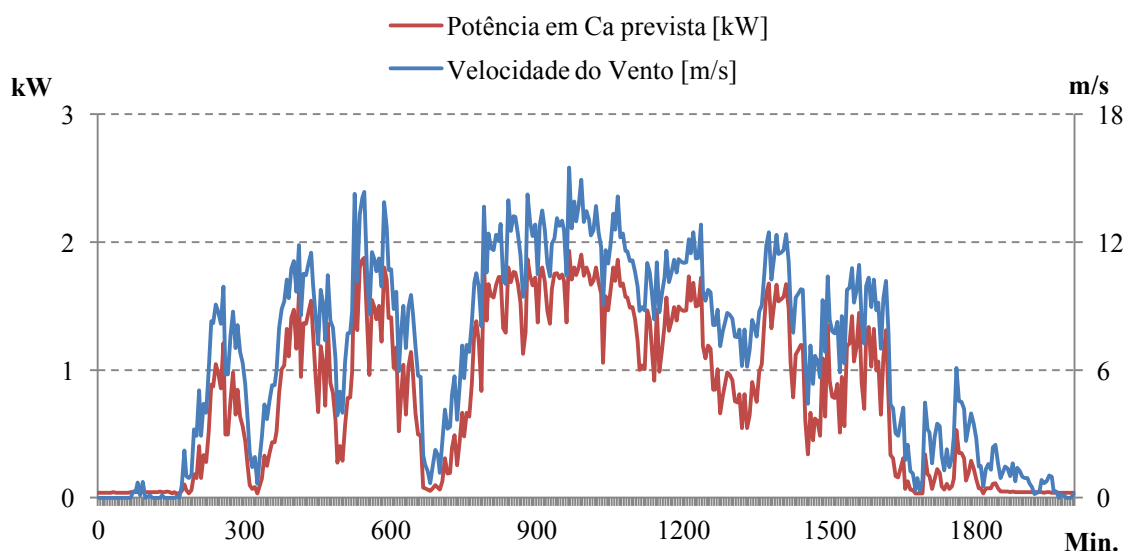


Figura 6.16-Previsão da potência em CA para o caso de não produção energia.

No seguimento das avarias ocorridas a 16 de Setembro de 2015 foi aplicado o modelo de previsão para prever a produção de energia, como se pode visualizar na Figura 6.17.

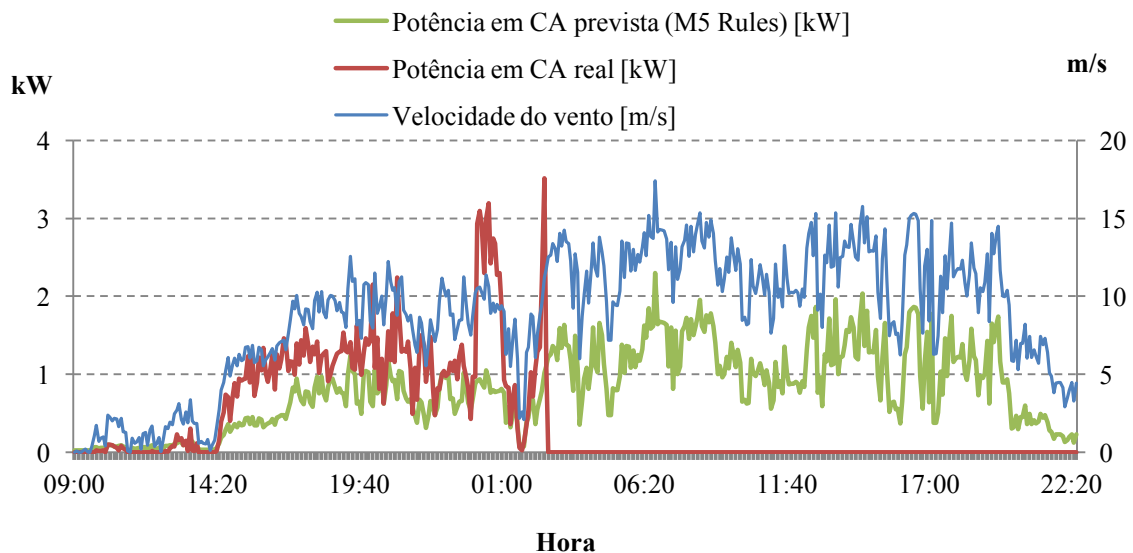


Figura 6.17-Previsão para um caso de avaria.

Neste caso, a potência prevista é inferior à potência real, e a potência real é inferior à velocidade do vento. Contudo, existem momentos antes da avaria em que a potência real fica acima da velocidade do vento e nos instantes a seguir fica abaixo da potência prevista, e de seguida o sistema deixa de funcionar. De forma a prevenir este acontecimento, o sistema computacional responsável pela monitorização do sistema deve ter conhecimento de quando o sistema estiver a produzir um valor acima do valor da velocidade do vento e um valor inferior ao valor que o sistema tenha previsto para a produção para determinadas condições. Desta forma, o sistema pode indicar de uma forma automática a ocorrência de uma avaria, constituindo uma mais valia na deteção em tempo real das avarias.

6.4 Análise Comparativa

Na Tabela 6.13 é realizada uma comparação dos melhores modelos de previsão de energia a fim de obter informação para a monitorização dos sistemas analisados (em que foi escolhida uma semana do mês de Março).

Tabela 6.13-Comparação dos melhores modelos de previsão de produção energia para monitorização

	Fotovoltaico Estático M5P	Fotovoltaico Seguidor M5P	Eólica M5 Rules
Tempo de construção (s)	8.97	5.4	33.48
Coefficiente de correlação	0.998	0.991	0.846
Erro médio absoluto	71.669	72.031	95.668
Erro médio quadrático	101.337	129.499	139.379
Erro absoluto relativo (%)	4.491	7.812	23.711
Raiz quadrada do erro médio quadrático (%)	5.653	12.997	32.244

Como podemos analisar na Tabela 6.13, o modelo que obteve melhor previsão de energia a fim de obter informação para a monitorização, foi o do sistema fotovoltaico estático em que obteve um coeficiente de correlação elevado e valores de erro mais baixos. Este facto é justificado por ser um sistema mais simples e com um comportamento de produção de energia mais regular, o que não acontece em relação ao sistema eólico (que é o que apresenta piores resultados).

6.4.1 Comparação das previsões a curto prazo

Com o objetivo de saber qual o sistema que obteve melhor modelo de previsão para o período de 5 minutos, a Tabela 6.14 compara os melhores resultados de cada sistema analisado.

Tabela 6.14- Comparação dos melhores modelos de previsão de produção para um período de 5 minutos

	Fotovoltaico Estático M5 Rules	Fotovoltaico Seguidor MP	Eólica M5 Rules
Tempo de construção (s)	121.63	25.24	55.78
Coefficiente de correlação	0.991	0.959	0.841
Erro médio absoluto	167.035	138.274	287.592
Erro médio quadrático	249.554	185.114	367.626
Erro absoluto relativo (%)	10.628	13.859	40.080
Raiz quadrada do erro médio quadrático (%)	14.098	17.681	42.051

Como é possível verificar na Tabela 6.14, os sistemas fotovoltaico obtiveram bons resultados, sendo que o sistema fotovoltaico estático obteve o melhor coeficiente de correlação, mas em relação a valores de erro o sistema fotovoltaico seguidor obteve melhores resultado. Em relação ao tempo de construção do modelo, o sistema fotovoltaico seguidor foi o que obteve o valor mais baixo. A Tabela 6.15 compara os melhores resultados de todos os sistemas para o modelo para previsão no período de 15 minutos.

Tabela 6.15- Comparação dos melhores modelos de previsão de produção para um período de 15 minutos

	Fotovoltaico Estático M5 Rules	Fotovoltaico Seguidor MP	Eólica M5 Rules
Tempo de construção (s)	143.84	22.75	56.94
Coefficiente de correlação	0.977	0.949	0.781
Erro médio absoluto	245.413	147.730	330.286
Erro médio quadrático	362.191	206.204	422.569
Erro absoluto relativo (%)	15.616	14.787	46.034
Raiz quadrada do erro médio quadrático (%)	20.461	19.672	48.334

Na previsão para o período de 15 minutos, tal como no caso de 5 minutos, o melhor modelo de previsão foi o do sistema fotovoltaico seguidor, sendo aquele que tem o tempo construção do modelo mais baixo e o que obteve valores mais baixo de erro.

6.4.2 Comparação das previsões a médio prazo

Na comparação a médio prazo só é realizada a comparação para os sistemas fotovoltaicos, visto que não é realizada previsão a médio prazo para o sistema eólico devido à dificuldade de previsão e complexidade.

Como é possível analisar na Tabela 6.16, ambos os sistemas apresentam valores de erro elevados, sendo o sistema fotovoltaico estático o que apresenta melhores resultados em termos de coeficiente de correlação e de erro. O tempo de construção do modelo para ambos os casos é similar e relativamente baixo.

Tabela 6.16- Comparação dos melhores modelos de previsão a médio prazo

	Fotovoltaico Estático M5 Rules	Fotovoltaico Seguidor MP
Tempo de construção (s)	1.4	1.34
Coefficiente de correlação	0.952	0.804
Erro médio absoluto	429.147	479.515
Erro médio quadrático	566.175	702.883
Erro absoluto relativo (%)	25.555	44.097
Raiz quadrada do erro médio quadrático (%)	29.936	60.143

7 Conclusão

7.1 Conclusões

O principal objetivo deste trabalho consistiu na análise de dados relativos à produção de energia elétrica de sistemas solar fotovoltaicos e eólicos provenientes da micro rede do IPB por forma a criar modelos de previsão. Esta análise de dados correlacionou dados históricos e atuais da produção de energia elétrica e dados meteorológicos (por exemplo, irradiação solar, temperatura e velocidade do vento), considerando também dados sobre a previsão meteorológica. Neste âmbito foram identificadas três situações: i) monitorização em tempo real dos equipamentos de produção, permitindo a deteção de avarias ou degradação na sua operação, ii) previsão da produção de energia elétrica a curto prazo, isto é para os próximos 5 e 15 minutos, permitindo reagir rapidamente a eventuais desvios na produção, e iii) previsão da produção de energia elétrica a médio prazo, isto é para a próxima hora, permitindo balancear mais eficazmente o fluxo de energia na micro rede.

Para o efeito foi utilizada a ferramenta Weka e diversos algoritmos de análise de dados, nomeadamente redes neuronais, regras e árvores de decisão, de forma a obter os melhores modelos de previsão.

Na análise de resultados dos sistemas fotovoltaico é importante salientar que durante o período de verão são obtidos melhores resultados o que não acontece no período de inverno. No sistema eólico os resultados são evidentes no período do verão pois é uma estação do ano com pouco vento, com isto podemos concluir que os resultados variam com a estação do ano nos diferentes sistemas.

Os resultados experimentais mostraram que é possível um modelo credível para realizar a monitorização para o sistema solar fotovoltaico estático, assim como para a previsão para curto prazo, quer para um período de tempo de 5 minutos quer para um período de 15 minutos. Em relação à previsão a médio prazo, os resultados não foram tão bons, essencialmente devido à fraca qualidade da previsão meteorológica para períodos mais largos (quanto maior for o tempo de previsão, piores será a qualidade da previsão).

No que diz respeito ao sistema de produção eólico, a construção dos modelos de previsão é mais complexa, essencialmente devido ao elevado grau de incerteza na previsão da velocidade do vento, que assume crucial importância neste modelo. Neste caso, apenas foi aplicado o modelo de previsão a curto prazo, uma vez que as previsões da velocidade do vento apenas foram consideráveis boas para períodos de 5 e 15 minutos. Neste sistema de produção ainda houve a hipótese de analisar alguns casos de avarias que ocorreram durante a análise da produção de energia elétrica, em que foi possível analisar como o sistema se comportava com uma avaria ou até no caso de não existir produção de energia.

7.2 Desenvolvimento futuro

Depois de toda a informação adquirida ao longo do desenvolvimento desta dissertação. E já com todos os testes possíveis executados, a sempre inúmeras situações que poderiam ser feitas. Aqui ficam algumas ideias interessantes, que futuramente podem ser desenvolvidas para melhorar os sistemas analisados ao longo desta dissertação.

Neste trabalho foram apresentadas várias hipóteses para a construção do modelo de previsão da produção no sistema eólico, em que foram obtidos resultados razoáveis, isto devido ao facto de ser um sistema de difícil previsão, e visto também a instabilidade do vento. Futuramente, este tópico deveria ser detalhadamente abordado, com o objetivo de melhorar os resultados de previsão e reduzir os valores dos erros de previsão. No desenvolvimento do estudo para o sistema de produção eólico apareceram algumas avarias, que futuramente poderia ser alvo de estudo, nomeadamente o comportamento do sistema a diferentes situações de avarias. Seria interessante também desenvolver mecanismos que permitam ao sistema de monitorização informar da necessidade de realização de tarefas de manutenção, e ou troca de equipamento, após detetar, por exemplo degradação do desempenho do sistema de produção de energia elétrica.

Seria também interessante integrar no atual processo de análise de dados os dados provenientes das duas estações meteorológicas existentes no IPB, de forma a melhorar os coeficientes de correlação dos modelos de previsão desenvolvidos neste trabalho

Por fim, seria interessante aplicar a análise de dados para criação de modelos de previsão de micro produção de energia elétrica a outras micro redes, a outros sistemas de produção (tipo de potência instalada), possibilitando o desenvolvimento de modelos de previsão mais genéricos.

8 Referências

- [1] A. F. dos Santos, “Energia e a crise. A mudança de paradigma energético”, *Edifícios e Energia*, pp. 29–30, 2014.
- [2] International Energy Agency, “Technology Roadmap Smart Grids”, *Current*, p. 52, 2011.
- [3] P. K. Janert, *Data Analysis with Open Source Tools*. 2011.
- [4] U. Fayyad, G. Piatetsky-Shapiro, e P. Smyth, “From data mining to knowledge discovery in databases”, *AI Mag.*, pp. 37–54, 1996.
- [5] “Energy Big Data Europe.” [Em linha]. Disponível em: <http://www.big-data-europe.eu/energy/>. [Acedido: 28-Out-2015].
- [6] I. H. Witten, E. Frank, e M. a. Hall, *Data Mining: Practical machine learning tools and techniques*. 2011.
- [7] “What is the CRISP-DM methodology.” [Em linha]. Disponível em: <http://www.sv-europe.com/crisp-dm-methodology/>.
- [8] C. Pete, C. Julian, K. Randy, K. Thomas, R. Thomas, S. Colin, e R. Wirth, “Crisp-Dm 1.0”, *Cris. Consort.*, p. 76, 2000.
- [9] C. D. L. Daniel T. Larose, *Data Mining and Predictive Analytics (Wiley Series on Methods and Applications in Data Mining)*, vol. 25, n. 9. 2015.
- [10] V. Lo Brano, G. Ciulla, e M. Di Falco, “Artificial Neural Networks to Predict the Power Output of a PV Panel”, *Int. J. Photoenergy*, vol. 2014, p. 12, 2014.
- [11] A. Badri, Z. Ameli, e A. M. Birjandi, “Application of Artificial Neural Networks and Fuzzy logic Methods for Short Term Load Forecasting”, *Energy Procedia*, vol. 14, n. 2011, pp. 1883–1888, 2012.
- [12] A. Barbato, A. Capone, M. Rodolfi, e D. Tagliaferri, “Forecasting the usage of household appliances through power meter sensors for demand management in the smart grid”, *IEEE Int. Conf. Smart Grid Commun.*, pp. 404–409, 2011.

- [13] A. Motamedi, H. Zareipour, S. Member, e W. D. Rosehart, “Electricity Price and Demand Forecasting in Smart Grids”, *IEEE Trans. Smart Grid*, vol. 3, n. 2, pp. 664–674, 2012.
- [14] “Vestas | Wind it means the world to us.” [Em linha]. Disponível em: <https://www.vestas.com/#!> [Acedido: 03-Set-2015].
- [15] IBM, “IBM Drives the Future of Renewable Energy with New Wind and Solar Forecasting System.” 12-Ago-2013.
- [16] “R: The R Project for Statistical Computing.” [Em linha]. Disponível em: <https://www.r-project.org/>. [Acedido: 02-Nov-2015].
- [17] “KNIME-Plataforma Analytics KNIME.” [Em linha]. Disponível em: <https://www.knime.org/knime>. [Acedido: 02-Nov-2015].
- [18] “RapidMiner - Open Source Predictive Analytics Platform.” [Em linha]. Disponível em: <https://rapidminer.com/>. [Acedido: 02-Nov-2015].
- [19] “Weka - Data Mining with Open Source Machine Learning Software in Java.” [Em linha]. Disponível em: <http://www.cs.waikato.ac.nz/ml/weka/index.html>. [Acedido: 02-Nov-2014].
- [20] M. Bilgili, B. Sahin, e A. Yasar, “Application of artificial neural networks for the wind speed prediction of target station using reference stations data”, *Renew. Energy*, vol. 32, n. 14, pp. 2350–2360, 2007.
- [21] K. Abhishek, M. P. Singh, S. Ghosh, e A. Anand, “Weather Forecasting Model using Artificial Neural Network”, *Procedia Technol.*, vol. 4, pp. 311–318, 2012.
- [22] J. Z. J. Zhang e F. S. F. Shao, “Stock Data Analysis Based on BP Neural Network”, *Intell. Inf. Technol. Appl. Work. 2009 IITAW 09 Third Int. Symp.*, pp. 288–291, 2009.
- [23] M. Kubat, “Neural networks: a comprehensive foundation by Simon Haykin, Macmillan, 1994, ISBN 0-02-352781-7.”, *The Knowledge Engineering Review*, vol. 13, n. 4. pp. 409–412, 1999.

- [24] S. a Kalogirou, “Artificial neural networks in renewable energy systems applications: a review”, *Renew. Sustain. Energy Rev.*, vol. 5, n. 4, pp. 373–401, 2001.
- [25] M. Cilimkovic, “Neural Networks and Back Propagation Algorithm”, *Fett.Tu-Sofia.Bg*, 2010.
- [26] S. Kilic, *Linear regression analysis Theory and computing*. World Scientific Publishing Co. Pte. Ltd., 2013.
- [27] “Linear regression-From Wikipedia, the free encyclopedia.” [Em linha]. Disponível em: https://commons.wikimedia.org/wiki/File:Linear_regression.svg. [Acedido: 26-Out-2015].
- [28] L. Rokach, *Data Mining with Decision Trees, Theory and Application*. 2008.
- [29] M. Abernethy, “Data mining with WEKA, Part 2: Classification and clustering”, *IBM Dev.*, p. None, 2010.
- [30] J. Siekmann, J. Hartmanis, e J. Van Leeuwen, *Lecture Notes in Artificial Intelligence*. Springer Berlin Heidelberg, 1919.
- [31] I. H. Witten, E. Frank, e M. a Hall, *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*. 2011.
- [32] J. Manuel e F. Loureiro, “Desenvolvimento de Modelos de Previsão a Longo Prazo de Preços de Mercado de Electricidade”, 2008.
- [33] R. Jing, Y. Wang, Y. Wu, Y. Hua, X. Dai, e M. Li, “A Research of Predicting the B-factor Base on the Protein Sequence”, *J. Theor. Comput. Sci.*, vol. 2014, Abr. 2014.
- [34] T. Chai e R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE) – Arguments against avoiding RMSE in the literature”, *Geosci. Model Dev.*, pp. 1247–1250, 2014.
- [35] “VERCAMPUS.” [Em linha]. Disponível em: http://vercampus.ipb.pt/portal/page?_pageid=495,292215&_dad=portal&_schema=PORTAL. [Acedido: 09-Set-2014].

- [36] “VERCAMPUS - Unidades Fotovoltaicas.” [Em linha]. Disponível em: http://vercampus.ipb.pt/portal/page?_pageid=495,292181&_dad=portal&_schema=PORTAL. [Acedido: 26-Out-2014].
- [37] “VERCAMPUS - Microprodução com turbinas eólicas.” [Em linha]. Disponível em: http://vercampus.ipb.pt/portal/page?_pageid=495,291958&_dad=portal&_schema=PORTAL. [Acedido: 26-Out-2014].
- [38] “VERCAMPUS-Seguidor Solar.” [Em linha]. Disponível em: http://vercampus.ipb.pt/portal/page?_pageid=495,339773&_dad=portal&_schema=PORTAL. [Acedido: 26-Out-2014].
- [39] “VERCAMPUS-Turbina Eólica.” [Em linha]. Disponível em: http://vercampus.ipb.pt/portal/page?_pageid=495,339782&_dad=portal&_schema=PORTAL. [Acedido: 26-Out-2014].
- [40] “OpenWeatherMap current weather and forecast.” [Em linha]. Disponível em: <http://openweathermap.org/>. [Acedido: 09-Mar-2014].
- [41] “Weather Underground.” [Em linha]. Disponível em: <http://www.wunderground.com/>. [Acedido: 01-Jan-2015].
- [42] R. António, “Gestão Colaborativa em Micro-Redes Inteligentes”, IPB, 2014.
- [43] J. Queiroz, A. Dias, e P. Leitão, “Predictive Data Analytics for Agent-based Management of Electrical Micro Grids”, *IECON'15*, p. 6, 2015.
- [44] C. Monteiro, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, A. Muñoz-Jimenez, e P. M. Lara-Santillan, “Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques”, *Math. Probl. Eng.*, vol. 2013, 2013.

9 ANEXO

9.1 Anexo A

Analisando gráfica de todos meses disponíveis de modo a analisar o comportamento da potência em CA fornecida em relação ao comportamento da irradiação ao longo dos meses. Como pode ser concluído na Figura 9.1 a potência em CA tem o mesmo comportamento da irradiação, mas não podemos deixar de reparar que o comportamento inicial quer da irradiação quer da potência em CA não são lineares, tem um momento inicial que á uma alteração brusca na potência em CA.

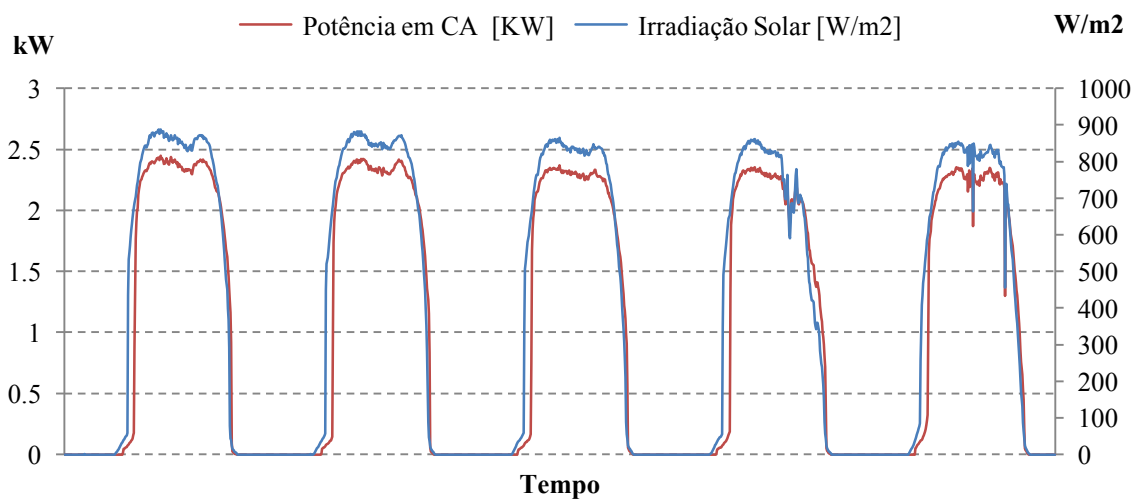


Figura 9.1-Análise semanal da irradiação e potência em CA

Na Figura 9.2 pode ser feita uma análise gráfica para um dia. Em que a situação torna-se ainda mais evidente, onde á uma alteração brusca da potência em CA e a irradiação solar. Conclui-se que era devido a um sombreamento de uma árvore no início do dia, com o sombreamento a potência em CA mantinha-se em baixo e com a movimentação do sol deixava de haver sombreamento, provocando uma elevação da potência em CA, como se pode verificar com uma subida repentina quer por parte da irradiação, quer por parte da potência em CA fornecida. Devido a este problema inicial o estudo do algoritmo foi induzido em erro com a utilização destes dados.

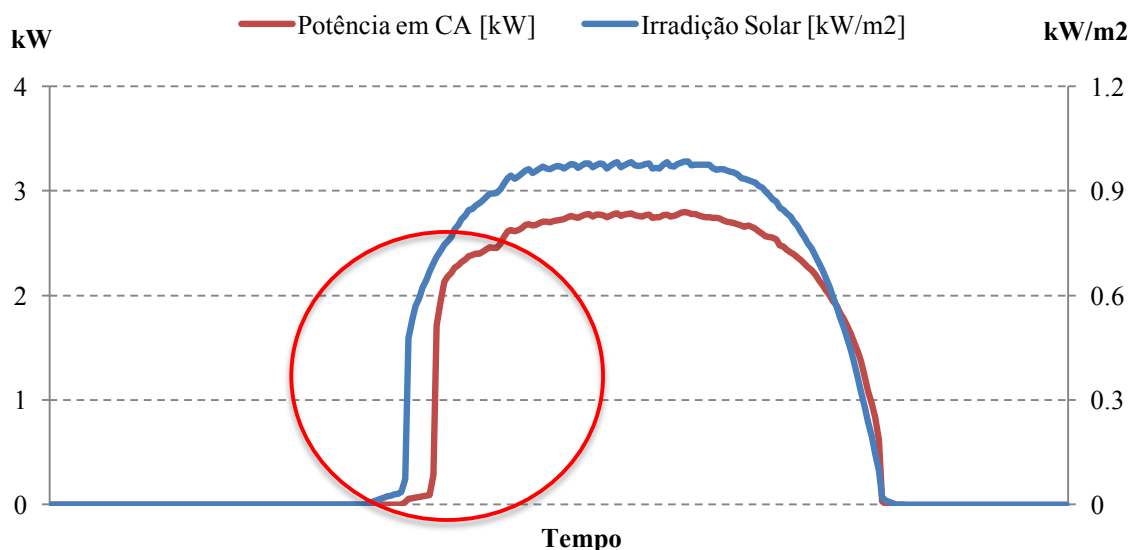


Figura 9.2- Problema na potência em CA e na irradiação solar devido ao sombreamento de uma árvore.

Com o contexto de perceber o problema da primeira abordagem, devido aos valores elevados, optou-se por aplicar uma técnica para o tratamento do sombreamento com o objetivo de melhorar o modelo de previsão. Então optou-se por fazer um mini-teste para verificar a veracidade da técnica. Com a aquisição destes dados construiu-se a Tabela 9.1.

Tabela 9.1-Teste para veracidade da técnica aplicada para remoção do sombreamento

Algoritmo:	<i>M5P</i>
Tempo de construção (s)	8.0
Coefficiente de correlação	0.989
Erro médio absoluto	92.585
Erro médio quadrático	137.721
Erro absoluto relativo (%)	9.243
Raiz quadrada do erro médio quadrático (%)	12.926

Com a realização deste teste foram obtidos resultados favoráveis, ou seja, um alto coeficiente de correlação, um erro médio absoluto baixo e um erro absoluto relativo baixo, isto significa que o sombreamento está a influenciar o estudo por parte do algoritmo e conseqüentemente a influenciar os resultados. Sendo assim, esta técnica

pode ser aplicada para esta situação, de modo a obter-se melhores resultados. Com base nos resultados da Tabela 9.1 realizou-se o gráfico referente a Figura 9.3, de modo a ver o comportamento da potência em CA prevista para cinco dias.

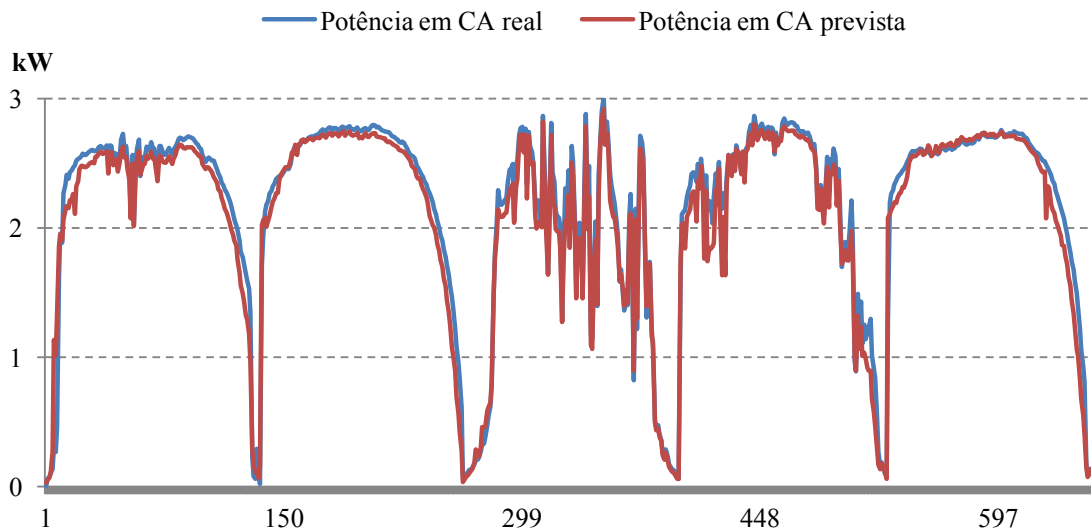
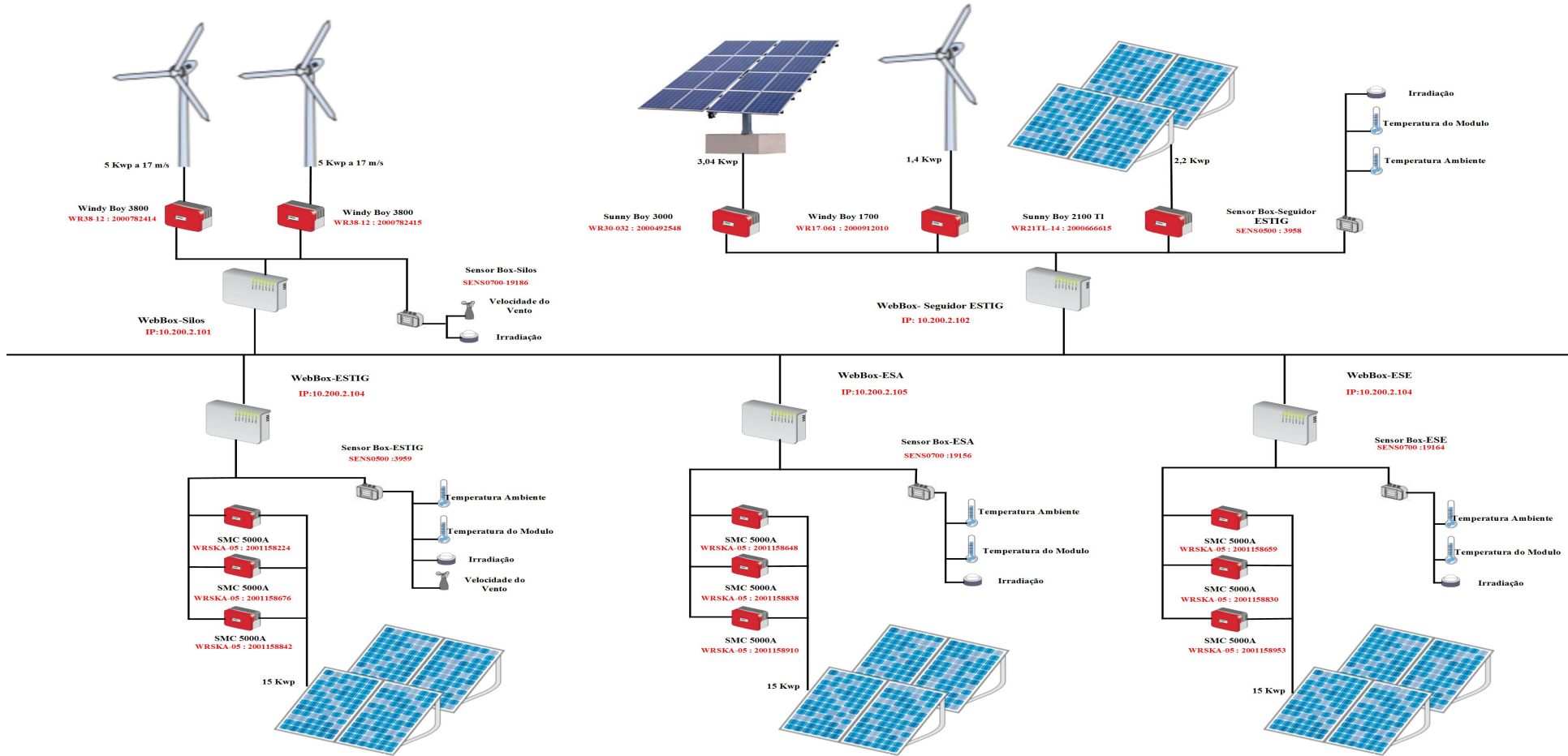


Figura 9.3- Resultado da potência em CA prevista usando o algoritmo MSP sem problema inicial de sombreamento referente a Tabela 9.1.

Como se pode verificar na Figura 9.3 o algoritmo obteve uma boa previsão da potência em CA fornecida, uma vez que conseguiu fazer uma previsão abaixo da potência em CA real.

9.2 Anexo B



9.3 Anexo C

Tabela com resultados aplicando algoritmo MSP para 5 minutos, anualmente, para sistema fotovoltaico estático.

Algoritmo:	MSP-5 minutos											
Semana de teste:	S9-2014	S13-2014	S17-2014	S22-2014	S26-2014	S30-2014	S35-2014	S39-2014	S44-2014	S48-2014	S52-2014	S5-2015
Tempo de construção (s)	4.15	2.44	4.59	5.07	5.05	5.69	6.02	6.44	6.43	6.63	6.7	6.94
Coefficiente de correlação	0.997	0.998	0.998	0.998	0.999	1.000	0.999	0.997	0.997	0.991	0.967	0.991
Erro médio absoluto	82.032	63.239	77.324	76.425	66.606	55.266	97.065	109.651	84.856	69.989	247.935	129.374
Erro médio quadrático	124.598	104.504	118.458	118.003	102.131	70.019	126.067	143.656	112.000	133.734	448.850	209.093
Erro absoluto relativo (%)	6.148	4.564	5.238	5.320	4.485	3.648	6.841	7.653	6.700	4.692	16.651	9.192
Raiz quadrada do erro médio quadrático (%)	8.262	6.687	7.049	7.168	6.167	4.192	7.852	8.970	7.953	8.493	28.319	13.551

9.4 Anexo D

Tabela com resultados aplicando algoritmo M5P para 5 minutos, anualmente, para sistema fotovoltaico seguidor .

Algoritmo:	M5P-5 minutos											
Semana de teste:	S34-2014	S38-2014	S43-2014	S47-2014	S51-2014	S3-2015	S7-2015	S11-2015	S16-2015	S21-2015	S25-2015	S30-2015
Tempo de construção (s)	4.04	4.07	5.31	4.5	5.3	4.92	4.91	5.4	5.48	5.69	5.84	5.98
Coefficiente de correlação	0.989	0.996	0.955	0.966	0.910	0.906	0.977	0.991	0.994	0.981	0.959	0.981
Erro médio absoluto	58.320	57.744	104.811	60.362	65.137	239.891	150.564	72.031	70.017	91.318	90.402	79.478
Erro médio quadrático	118.448	86.869	259.752	155.137	207.325	386.987	216.827	129.499	116.614	148.101	234.702	154.228
Erro absoluto relativo (%)	6.836	6.322	11.226	6.089	5.861	31.861	15.719	7.812	7.614	9.688	10.220	9.519
Raiz quadrada do erro médio quadrático (%)	12.896	8.691	25.342	14.696	17.984	44.614	20.991	12.997	11.685	14.731	25.048	17.164

9.5 Anexo E

Tabela com resultados aplicando algoritmo M5 Rules para 5 minutos, anualmente, para o sistema eólico.

Algoritmo:	M5 Rules-5 minutos											
Semana de teste:	Dezembro 13	Janeiro 14	Fevereiro 14	Março 14	Abril 14	Mai 14	Junho 14	Julho 14	Agosto 14	Setembro 14	Outubro 14	Novembro 14
Tempo de construção (s)	22.28	32.68	22.66	26.56	28.41	22.42	29.64	33.48	41.38	41.61	41.48	65.62
Coefficiente de correlação	0.825	0.911	0.893	0.895	0.878	0.670	0.848	0.846	0.926	0.878	0.785	0.956
Erro médio absoluto	262.999	222.321	238.567	284.575	289.390	211.143	175.846	95.668	154.641	110.178	127.220	180.074
Erro médio quadrático	387.701	314.929	344.397	383.606	411.328	287.038	268.110	139.379	221.729	149.493	150.693	224.682
Erro absoluto relativo (%)	40.925	31.880	38.309	43.156	59.791	65.546	48.027	23.711	34.394	23.903	27.371	30.967
Raiz quadrada do erro médio quadrático (%)	49.379	36.805	44.177	47.864	68.983	75.071	60.240	32.244	39.482	31.451	31.889	30.536