

**Aplicação de Técnicas de *Machine Learning* no
Âmbito do Diagnóstico da Ocorrência de
Curtos-Circuitos Estatóricos em Motores de
Indução Trifásicos**

Carlos Fernando Ferreira Mosquera Fernandez

Dissertação para obtenção do Grau de Mestre em

Engenharia e Gestão Industrial

(2º ciclo de estudos)

Orientador: Professor Doutor António João Marques Cardoso

Outubro de 2021

Agradecimentos

Certa vez, ouvi que a dissertação de mestrado é um caminhar no deserto. Trata-se, realmente, de um trabalho solitário e muito árduo que acaba por gerar muitas dúvidas sobre se seremos capazes de levá-lo avante.

Não estive, no entanto, a caminhar sozinho. Consegui alcançar o meu objetivo graças ao apoio de várias pessoas, às quais gostaria de fazer um agradecimento em especial.

Ao Professor Marques Cardoso, agradeço por aceitar o tema desta dissertação e por me orientar dentro dos critérios científicos, que sempre pautaram a sua carreira profissional. Pois, sem dúvida, essas orientações ajudaram-me a esforçar-me cada vez mais, para alcançar os objetivos deste trabalho.

Agradeço ao Mestre Fernando Bento, que me orientou e me ajudou com os ensaios laboratoriais. E, sobretudo, agradeço os seus incentivos constantes: “Não desanimes...”, “Força...”.

Ao Hugo Antunes, agradeço por me ter disponibilizado o seu tempo, modelos, explicações e até mesmo dados, que serviram de base para o presente trabalho.

Ao Doutor Khaled Laadjal, que esteve sempre presente nos momentos de dificuldade tanto a nível técnico como até mesmo pessoal, e que sempre me incentivou a continuar.

Ao André Barcelos, que sempre fez críticas muito construtivas com respeito a este trabalho e que nunca se cansou de perguntar: “Porquê?”. Esta pergunta que aparentemente é simples leva-nos realmente a pensar sobre o propósito do que estamos a fazer, onde pretendemos chegar e, principalmente, o “porquê” de escolhermos este caminho.

A todos os membros do CISE e, em especial à Dra. Sylvie e à Raquel, por me apoiarem sempre e estarem sempre disponíveis para me ajudar durante esta jornada.

A minha esposa, Fabiana Oliveira e família, por me apoiarem nos meus sonhos.

Obrigado a todos!

Resumo

A Indústria 4.0 traz associado o conceito de fábricas inteligentes, que surgem para atender a uma necessidade crescente de elevada flexibilidade e eficiência na fabricação dos produtos. Os motores de indução trifásicos já são utilizados em larga escala pela indústria e terão um papel primordial nas *Smart Factories*. A fim de mantê-los em funcionamento, com elevado grau de fiabilidade, e de reduzir os custos associados a paragens e intervenções, torna-se necessária a identificação de avarias, ainda em estágio precoce, de forma a programar a manutenção antes da paragem total daqueles equipamentos.

Os métodos de diagnóstico de avarias *online* vêm sendo alvo de estudo há muitos anos. A avaliação dos resultados da aplicação desses métodos de diagnóstico depende de um especialista para uma interpretação e um diagnóstico precisos. Este trabalho aborda o uso de Inteligência Artificial para a deteção de avarias em motores elétricos, de forma automatizada, ainda em estágio precoce. Foram utilizados algoritmos de *Machine Learning*, nomeadamente *Support Vector Machines* e *Decision Trees* para a deteção de curtos-circuitos entre espiras, de forma preditiva, baseando-se em dados reais adquiridos no Laboratório de Sistemas Electromecatrónicos do CISE - Centro de Investigação em Sistemas Electromecatrónicos.

Como principal diferença em relação a outros trabalhos, será apresentada uma abordagem baseada no *Extended Park's Vector Approach*, que já é atualmente utilizada para esta finalidade. Trata-se, portanto, de uma abordagem híbrida e tem como principal objetivo obter o máximo de eficiência com o mínimo de *features*, a partir de uma ferramenta já utilizada para este fim.

Como resultado, é apresentado um sistema de diagnóstico automatizado para a deteção de curtos-circuitos estatóricos em motores de indução trifásicos que utiliza *Support Vector Machine* e *Decision Tree*, contemplando cenários de funcionamento distintos, nomeadamente condição normal e condição de desequilíbrio na rede de alimentação, relativo ao deslocamento de fase ou à amplitude das tensões. Após a otimização dos modelos, será realizada a validação em dados novos e os seus resultados serão discutidos.

Palavras-chave

Diagnóstico de avarias estatóricas, Curto-circuito, *Motor Current Signature Analysis*, *Machine learning*, *Power Current Signature*, *Park's Vector Approach*, *Extended Park's Vector Approach*.

Abstract

Industry 4.0 brings the concept of intelligent factories, that arise to meet a growing need for high flexibility and efficiency in the manufacture of products. Three-phase induction motors are already used on a large scale by the industry and will play a key role in Smart Factories. In order to keep them in operation with a high degree of reliability, and to reduce the costs associated with stoppages and interventions, it is necessary to identify faults at an early stage, in order to schedule maintenance before the electric machine stops completely.

Online fault diagnostics methods have been studied for many years. However, the assessment of their results depends on an expert for accurate interpretation and diagnosis. This study addresses the use of Artificial Intelligence to detect malfunctions in electric motors, in an automated way, at an early stage. Machine Learning algorithms, namely Support Vector Machines and Decision Trees, were used for the detection of stator inter-turn short circuits, in a predictive way, based on real data acquired at the Laboratory of Electromechatronic Systems of CISE - Electromechatronic Systems Research Centre.

As a main difference in relation to other works, an approach based on the Extended Park's Vector Approach, which is currently used for this purpose, will be presented. Therefore, it is a hybrid approach, and its main objective is to obtain maximum efficiency with minimum features, from a tool already used for this purpose.

As a result, an automated diagnostic system, that uses Support Vector Machine and Decision Tree, is presented for the detection of stator short circuits in three-phase induction motors, which includes scenarios of normal and unbalance in the phase or amplitude of the supply voltages. After optimizing the models, validation will be performed on raw data and its results will be discussed.

Keywords

Stator fault diagnostics, Short-circuit, Motor Current Signature Analysis, Machine learning, Power Current Signature, Park's Vector Approach, Extended Park's Vector Approach.

Sumário

CAPÍTULO 1	21
1 INTRODUÇÃO.....	21
1.1 ENQUADRAMENTO.....	21
1.2 MOTIVAÇÃO.....	23
1.3 OBJETIVO	24
1.4 METODOLOGIA	24
1.4.1 <i>Pesquisa Exploratória</i>	25
1.4.2 <i>Projeto de Investigação</i>	25
1.4.3 <i>Execução do Projeto</i>	26
1.5 ESTRUTURA DA DISSERTAÇÃO	27
CAPÍTULO 2.....	29
2 MÁQUINAS ELÉTRICAS	29
2.1 MOTORES DE INDUÇÃO TRIFÁSICOS DE ROTOR EM GAIOLA DE ESQUILO.....	29
2.2 ESTATOR	30
2.3 ROTOR	30
2.4 VELOCIDADE SÍNCRONA (<i>ns</i>)	31
2.5 POTÊNCIA E RENDIMENTO.....	31
CAPÍTULO 3.....	35
3 MÉTODOS DE DIAGNÓSTICO DE AVARIAS EM MOTORES DE INDUÇÃO TRIFÁSICOS.....	35
3.1 INTRODUÇÃO.....	35
3.2 MÉTODOS DE DIAGNÓSTICO	35
3.3 FUNCIONAMENTO DA MCSA	36
3.4 DIAGNÓSTICO BASEADO EM MCSA.....	37
3.4.1 <i>Curtos-Circuitos entre Espiras no Estator</i>	37
3.5 DIAGNÓSTICO BASEADO EM PVA E EPVA	39
3.6 DIAGNÓSTICO BASEADO EM COMPONENTES SIMÉTRICAS.....	43
CAPÍTULO 4.....	45
4 INTELIGÊNCIA ARTIFICIAL.....	45
4.1 <i>MACHINE LEARNING</i>	46
4.1.1 <i>Decision Tree (DT)</i>	46
4.1.2 <i>Support Vector Machine (SVM)</i>	48
4.2 FEATURE EXTRACTION	49

4.2.1	<i>Remoção da Média (Componente DC)</i>	51
4.2.2	<i>Dados Estatísticos – Domínio do Tempo</i>	52
4.2.3	<i>Dados Estatísticos – Domínio da Frequência</i>	54
4.3	FEATURE SELECTION	55
4.3.1	<i>Correlação</i>	55
4.3.2	<i>Valor p</i>	56
4.4	MÉTRICAS DE AVALIAÇÃO E VALIDAÇÃO	57
4.4.1	<i>Regressão</i>	58
4.4.2	<i>Classificação</i>	59
4.4.3	<i>Generalização e Validação Cruzada</i>	60
CAPÍTULO 5		63
5	AQUISIÇÃO DE DADOS EXPERIMENTAIS	63
5.1	MOTOR DE INDUÇÃO TRIFÁSICO	63
5.2	FONTE AC PROGRAMÁVEL	64
5.3	SIMULAÇÃO DA CARGA.....	64
5.4	SISTEMA DE AQUISIÇÃO DOS DADOS	65
5.5	DESCRIÇÃO DOS DADOS ADQUIRIDOS	66
CAPÍTULO 6		69
6	APLICAÇÃO DO MACHINE LEARNING	69
6.1	PRIMEIRA FASE.....	70
6.1.1	<i>Dados do módulo Vetor de Park das correntes estatóricas</i>	71
6.1.2	<i>Feature Extraction e Windowing do módulo do Vetor de Park</i>	73
6.1.3	<i>Resultados dos modelos: fase 1</i>	86
6.2	SEGUNDA FASE.....	87
6.2.1	<i>Modelo com distúrbios na alimentação: Fase</i>	87
6.2.2	<i>Modelo com distúrbios na alimentação: extração de novas features</i>	90
6.2.3	<i>Resultado modelo: fase 2</i>	92
6.3	TERCEIRA FASE.....	93
6.3.1	<i>Modelo com distúrbios na alimentação: Amplitude</i>	93
6.3.2	<i>Resultado modelo: fase 3</i>	95
6.4	QUARTA FASE	96
6.4.1	<i>Modelo com distúrbios na alimentação: Amplitude e Fase</i>	96
6.5	QUINTA FASE.....	100
6.5.1	<i>Testes de validação em dados conhecidos</i>	100
6.5.2	<i>Testes de validação em dados desconhecidos</i>	101
6.6	DISCUSSÃO	103
CAPÍTULO 7		107

7	CONCLUSÃO	107
	CAPÍTULO 8	109
8	TRABALHOS FUTUROS	109
9	APÊNDICES	115
9.1	APÊNDICE A	115
9.2	APÊNDICE B	117
9.3	APÊNDICE C	119

Lista de Figuras

FIGURA 1 - ESQUEMA DE FUNCIONAMENTO DO SMART MOTOR [6].	23
FIGURA 2 - FLUXO DO PROCESSO DE INVESTIGAÇÃO DESTA DISSERTAÇÃO.	25
FIGURA 3 - VISTA EM CORTE DE UM MOTOR DE INDUÇÃO TRIFÁSICO [7].	29
FIGURA 4 - LIGAÇÃO EM TRIÂNGULO [7].	32
FIGURA 5 - LIGAÇÃO EM ESTRELA [7].	32
FIGURA 6 - TRIÂNGULO DAS POTÊNCIAS (CARGA ATIVA E REATIVA) [7].	33
FIGURA 7 - PROCESSO DE MONITORIZAÇÃO DA CONDIÇÃO DO MOTOR, <i>ONLINE</i> (ADAPTADO DE [9]).	36
FIGURA 8 - SISTEMA DE MONITORIZAÇÃO DE CORRENTE ESTATÓRICA (ADAPTADO DE [5]).	37
FIGURA 9 - MODOS DE FALHA NOS ESTADORES DE MOTORES ELÉTRICOS DE INDUÇÃO TRIFÁSICOS (ADAPTADO DE [10]).	38
FIGURA 10 - CURTO-CIRCUITO ESTATÓRICO ENTRE ESPIRAS NO ENROLAMENTO [10].	38
FIGURA 11 - COMPONENTES DA FREQUÊNCIA CORRESPONDENTES A CURTOS-CIRCUITOS ESTATÓRICOS [5].	39
FIGURA 12 - COMPARAÇÃO DO VETOR DE PARK NUM MOTOR SEM AVARIAS, E COM DIFERENTES SEVERIDADES DE CURTOS-CIRCUITOS ESTATÓRICOS.	41
FIGURA 13 - (A) CORRENTES EQUILIBRADAS; (B) TRANSFORMAÇÃO PARA DQO; (C), A CORRENTE A APRESENTA UMA AMPLITUDE MAIS ELEVADA QUE AS OUTRAS, NA ORDEM DE 50%; (D) NOTA-SE QUE ESTE EFEITO É OBSERVADO TAMBÉM NO MÓDULO DO VETOR DE PARK NA FIGURA D. DEVE-SE NOTAR QUE AS COMPONENTES D E Q NÃO ESTÃO NO MESMO REFERENCIAL DO ROTOR.	42
FIGURA 14 - (A) OSCILAÇÕES NO MÓDULO DO VETOR DE PARK DEVIDO A DESEQUILÍBRIOS NAS CORRENTES; (B) ANÁLISE ESPECTRAL DO MÓDULO DO VETOR DE PARK. É POSSÍVEL OBSERVAR A COMPONENTE DC NA FREQUÊNCIA 0 HZ.	42
FIGURA 15 - IMAGEM ILUSTRATIVA DAS COMPONENTES DE SEQUÊNCIA POSITIVA, NEGATIVA E ZERO [16].	44
FIGURA 16 - INTELIGÊNCIA ARTIFICIAL VS <i>MACHINE LEARNING</i> E <i>DEEP LEARNING</i> [19].	45
FIGURA 17 - EXEMPLO GRÁFICO DE UMA <i>DECISION TREE</i> [22].	47
FIGURA 18 - DIFERENTES CRITÉRIOS DE CLASSIFICAÇÃO [21].	48
FIGURA 19 - <i>SUPPORT VECTOR MACHINE</i> HIPERPLANE [21].	48
FIGURA 20 - A MARGEM É A MAIOR REGIÃO QUE PODEMOS COLOCAR, SEPARANDO AS CLASSES SEM HAVER PONTOS DENTRO, ONDE A CAIXA É FEITA DE DUAS LINHAS PARALELAS AO LIMITE DE DECISÃO [24].	49
FIGURA 21 - ESQUEMA DE <i>FEATURE EXTRACTION</i> DO PVA E DO EPVA (ADAPTADO DE [25]).	51

FIGURA 22 - SINAL A COM COMPONENTE DC: (A) REPRESENTAÇÃO NO DOMÍNIO DO TEMPO; (B) REPRESENTAÇÃO NO DOMÍNIO DA FREQUÊNCIA.	51
FIGURA 23 - SINAL A APÓS REMOÇÃO DA COMPONENTE DC: (A) NO DOMÍNIO DO TEMPO; (B) NO DOMÍNIO DA FREQUÊNCIA.	52
FIGURA 24 - COMPARAÇÃO ENTRE DIFERENTES CORRELAÇÕES ENTRE X E Y: (A) FORTE CORRELAÇÃO NEGATIVA; (B) CORRELAÇÃO FRACA; (C) FORTE CORRELAÇÃO POSITIVA (ADAPTADO DE [27]).	56
FIGURA 25 - CURVA DE DISTRIBUIÇÃO T. IDENTIFICAÇÃO DA REGIÃO $T = 2$, CORRESPONDENTE A UMA DENSIDADE PROBABILÍSTICA DE $0,02963 \times 2$, OU APROXIMADAMENTE 6% (ADAPTADO DE [29]).	57
FIGURA 26 - (A) REGRESSÃO LINEAR POLINOMIAL, COM 1 GRAU DE LIBERDADE NA CONDIÇÃO DE <i>UNDERFIT</i> ; (B) REGRESSÃO LINEAR POLINOMIAL ADEQUADA, COM 4 GRAUS DE LIBERDADE; (C) REGRESSÃO LINEAR POLINOMIAL, COM 20 GRAUS DE LIBERDADE EM CONDIÇÃO DE <i>OVERFIT</i> (ADAPTADO DE [30]).	58
FIGURA 27 - (A) RESULTADO DE UM AGRUPAMENTO (<i>CLUSTERING</i>) CLÁSSICO COM $K = 2$; (B) RESULTADO DE UM AGRUPAMENTO BASEADO EM DENSIDADE; (C) CLASSIFICAÇÃO BASEADA EM BAYES (ADAPTADO DE:[30]).	60
FIGURA 28 - DESENHO ESQUEMÁTICO DO K-FOLD CROSS-VALIDATION COM $K = 10$ (ADAPTADO DE [32]).	61
FIGURA 29 - CURVA HIPOTÉTICA DE APRENDIZAGEM DE UM CLASSIFICADOR PRECISÃO X. TAMANHO DO CONJUNTO DE DADOS DE TREINO. VALIDAÇÃO CRUZADA COM $K = 5$	61
FIGURA 30 - MOTOR DE INDUÇÃO TRIFÁSICO UTILIZADO NOS ENSAIOS EXPERIMENTAIS: (A) FOTO ILUSTRATIVA DO MOTOR; (B) MONTAGEM REAL NA BANCADA DE ENSAIOS.	63
FIGURA 31 - FONTE AC PROGRAMÁVEL.	64
FIGURA 32 - (A) DINAMÔMETRO;(B) SISTEMA DE CONTROLO.	65
FIGURA 33 - (A) SISTEMA DE AQUISIÇÃO DE DADOS DE CORRENTE; (B) SOFTWARE DE PROCESSAMENTO DE DADOS E DIAGNÓSTICO DE AVARIAS.	66
FIGURA 34 - BANCADA DE SIMULAÇÃO DAS AVARIAS DE CURTOS-CIRCUITOS ENTRE ESPIRAS NO MOTOR E DESEQUILÍBRIOS DE FASE E AMPLITUDE NA REDE DE ALIMENTAÇÃO.	67
FIGURA 35 -PERTURBAÇÕES NO MÓDULO DO VETOR DE PARK, CAUSADAS POR UM CURTO-CIRCUITO ENTRE 3 ESPIRAS.	71
FIGURA 36 - AMPLIAÇÃO DA FIGURA 35 ONDE É POSSÍVEL VERIFICAR COM MAIOR CLAREZA AS PERTURBAÇÕES CAUSADAS NO MÓDULO DO VETOR DE PARK DEVIDO AO CURTO-CIRCUITO ENTRE 3 ESPIRAS.	72

FIGURA 37 - <i>DATASET</i> QUE CONTÉM O VETOR DE PARK EM DUAS CONDIÇÕES DE CARGA, 7 NM E 13,6 NM, EM DIFERENTES CONDIÇÕES DE SEVERIDADE DE CURTO-CIRCUITO 00, 03, 06, 09, 12 ESPIRAS.	73
FIGURA 38 - CORRELAÇÃO ENTRE AS 13 <i>FEATURES</i> EXTRAÍDAS E A CLASSE DE AVARIAS, DISPOSTA EM ORDEM DECRESCENTE.....	75
FIGURA 39 - P-VAL ENTRE AS <i>FEATURES</i> E AS CLASSES DE AVARIA (00, 03, 06, 09, 12).....	76
FIGURA 40 - MAPA DE CALOR DE CORRELAÇÃO <i>FEATURES</i> X <i>FEATURES</i> X CLASSE.	77
FIGURA 41 - ITERAÇÕES DURANTE A OTIMIZAÇÃO DE PARÂMETROS DO MODELO SVM ECO LINEAR, REALIZADO NO MATLAB.	78
FIGURA 42 - APLICAÇÃO DO DESVIO PADRÃO: (A) DESVIO PADRÃO NORMALIZADO DO MÓDULO DO VETOR DE PARK, EM DADOS REAIS; (B) DESVIO PADRÃO DO MÓDULO DO VETOR DE PARK, GERADO ATRAVÉS DE MODELO DE SIMULAÇÃO NO SIMULINK MATLAB. JANELA DE 40 PONTOS (0,02 S).....	79
FIGURA 43 - JUSTAPOSIÇÃO DA MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ) DO MÓDULO DO VETOR DE PARK PARA UMA JANELA DE 0,005 S (10 PONTOS) APLICADAS AOS ARQUIVOS DE DADOS.	81
FIGURA 44 - JUSTAPOSIÇÃO DA MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ) DO MÓDULO DO VETOR DE PARK PARA UMA JANELA DE 0,02 S (40 PONTOS) APLICADAS AOS ARQUIVOS DE DADOS.	81
FIGURA 45 - JUSTAPOSIÇÃO DA MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ) DO MÓDULO DO VETOR DE PARK PARA UMA JANELA DE 0,04 S (80 PONTOS) APLICADAS AOS ARQUIVOS DE DADOS.	82
FIGURA 46 - JUSTAPOSIÇÃO DA MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ) DO MÓDULO DO VETOR DE PARK PARA UMA JANELA DE 0,1 S (200 PONTOS) APLICADAS AOS ARQUIVOS DE DADOS.	82
FIGURA 47 - COMPARAÇÃO DA CORRELAÇÃO ENTRE O FEAT19 E AS FEAT6, FEAT3 E FEAT1 X CLASSES.	83
FIGURA 48 - VALOR P DAS <i>FEATURES</i> X CLASSE.....	84
FIGURA 49 - COMPARAÇÃO DE DIFERENTES ALGORITMOS DE <i>MACHINE LEARNING</i> BASEADO EM EPVA PARA UMA JANELA DE 0,1 S, EM CONDIÇÕES NORMAIS DE ALIMENTAÇÃO (<i>TOOLBOX CLASSIFICATION LEARNER</i> DO MATLAB).	85
FIGURA 50 - JUSTAPOSIÇÃO DA MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ) DO MÓDULO DO VETOR DE PARK PARA UMA JANELA DE 0,1 S (200 PONTOS) APLICADAS AOS ARQUIVOS DE DADOS, COM DIFERENTES DESFASAMENTOS NAS TENSÕES DE ALIMENTAÇÃO, 0, 1 2 E 3 GRAUS.	88

FIGURA 51 - RESULTADO DA APLICAÇÃO DO LINEAR SVM, KERNEL: LINEAR, KERNEL SCALE: AUTOMATIC, BOX CONSTRAINT LEVEL: 10, MULTICLASS METHOD: ONE-VS-ONE, STANDARIZE DATA: TRUE.	89
FIGURA 52 - APLICAÇÃO DA CORRELAÇÃO E P-VAL DAS <i>FEATURES</i> X CLASSE: (A) CORRELAÇÃO ENTRE AS <i>FEATURES</i> X CLASSE; (B) P-VAL ENTRE AS <i>FEATURES</i> X CLASSE.....	90
FIGURA 53 - MAPA DE CALOR <i>FEATURES</i> X <i>FEATURES</i> X CLASSE.	91
FIGURA 54 - TOOLBOX DO MATLAB PARA A CRIAÇÃO DO MODELO DE DIAGNÓSTICO DE AVARIAS COM DISTÚRBO DE FASE NA ALIMENTAÇÃO. COMPARAÇÃO ENTRE ALGORITMOS DE <i>MACHINE LEARNING</i>	92
FIGURA 55 - (A) CORRELAÇÃO;(B) VALOR P.	94
FIGURA 56 - MAPA DE CALOR <i>FEATURES</i> X <i>FEATURES</i> X CLASSE.	94
FIGURA 57 - COMPARAÇÃO ENTRE A IMPLEMENTAÇÃO DO ALGORITMO DE <i>DECISION TREE</i> X SVM LINEAR (BOX CONSTRAINT: 10).	95
FIGURA 58 - APLICAÇÃO DA CORRELAÇÃO E P-VAL DAS <i>FEATURES</i> X CLASSE: (A) CORRELAÇÃO ENTRE AS <i>FEATURES</i> X CLASSE; (B) P-VAL ENTRE AS <i>FEATURES</i> X CLASSE.....	97
FIGURA 59 - MAPA DE CALOR <i>FEATURES</i> X <i>FEATURES</i> X CLASSE.	98
FIGURA 60 - RESULTADO DA APLICAÇÃO DO LINEAR SVM, KERNEL: LINEAR, KERNEL SCALE: AUTOMATIC, BOX CONSTRAINT LEVEL: 10, MULTICLASS METHOD: ONE-VS-ONE, STANDARIZE DATA: TRUE E DECISION TREE. FEATURES: vPARK E FEAT13.....	98
FIGURA 61 - MATRIZ CONFUSA DO MODELO DA APLICAÇÃO DO LINEAR SVM, KERNEL: LINEAR, KERNEL SCALE: AUTOMATIC, BOX CONSTRAINT LEVEL: 10, MULTICLASS METHOD: ONE-VS-ONE, STANDARIZE DATA: TRUE. FEATURES: vPARK E FEAT13.....	99
FIGURA 62 - FLUXO DE DECISÕES DOS DIFERENTES MODELOS PARA OBTENÇÃO DO DIAGNÓSTICO DE CURTOS-CIRCUITOS ENTRE ESPIRAS.	100
FIGURA 63 - FUNCIONAMENTO DO PROGRAMA DE DETEÇÃO AUTOMÁTICA CURTOS-CIRCUITOS ESTATÓRICOS EM MOTORES DE INDUÇÃO TRIFÁSICOS.	101
FIGURA 64 - RESULTADOS DA CLASSIFICAÇÃO REAL: SH03_I4.60_230_PH120_ISH4.5.....	102
FIGURA 65 - RESULTADOS DA CLASSIFICAÇÃO REAL: SH09_I2.58_230_PH119_ISH4.5.....	102
FIGURA 66 - RESULTADOS DA CLASSIFICAÇÃO REAL: SH09_I2.58_230_PH119_ISH4.5.	103
FIGURA 67 - RESULTADOS DA CLASSIFICAÇÃO REAL: SH09_I4.60_229_PH120_ISH4.5.....	103
FIGURA 68 - AUMENTO NA MAGNITUDE DO EPVÁ DA CORRENTE DEVIDO AO DESALINHAMENTO DO ACOPLAMENTO. AMOSTRA EXTRAÍDA DO MOTOR EM CONDIÇÃO SAUDÁVEL, EM CONDIÇÕES NORMAIS DE ALIMENTAÇÃO, EM DIFERENTES CONDIÇÕES DE CARGA, 7 Nm E 13,6 Nm, DURANTE O SEGUNDO ENSAIO.....	105

Lista de Tabelas

TABELA 1 - RESUMO DA EXECUÇÃO DO PROJETO DE DISSERTAÇÃO.	26
TABELA 2 - DESCRIÇÃO DE UMA <i>DECISION TREE</i> , EM FORMA DE TABELA.	47
TABELA 3 - EXPLICAÇÃO RESUMIDA DE DIFERENTES TIPOS DE TREINO, COM <i>UNDERFIT</i> , ADEQUADO E <i>OVERFIT</i>	59
TABELA 4 - CARACTERÍSTICAS DO SINAL AMOSTRADO.	67
TABELA 5 - CARACTERÍSTICAS DO SINAL AMOSTRADO.	67
TABELA 6 - CARACTERÍSTICAS DO SINAL AMOSTRADO.	68
TABELA 7 - ARQUIVOS COM OS DADOS ADQUIRIDOS DE TENSÃO E CORRENTES DO MOTOR. FREQUÊNCIA DE AMOSTRAGEM: 2 KHZ. TAMANHO: 2000 PONTOS. TOTAL DE ARQUIVOS: 10.	70
TABELA 8 - DATASET GERADO APÓS O PROCESSO DE <i>FEATURE EXTRACTION</i> , CONSIDERANDO 40 PONTOS POR LINHA.....	74
TABELA 9 - CORRELAÇÃO ENTRE AS <i>FEATURES</i> E AS <i>CLASSES</i>	75
TABELA 10 - CORRELAÇÃO ENTRE AS <i>FEATURES</i> E AS <i>CLASSES</i>	76
TABELA 11 - <i>FEATURES</i> SELECIONADAS COM MENOR REDUNDÂNCIA E MAIOR RELEVÂNCIA.	77
TABELA 12 - COMPARAÇÃO ENTRE AS DIFERENTES <i>FEATURES</i> , PARA A REDUÇÃO DIMENSIONAL.	78
TABELA 13 - DETALHES DAS AMOSTRAS DA FIGURA 42.	79
TABELA 14 - <i>FEATURES</i> DO MÓDULO DO VETOR DE PARK.....	83
TABELA 15 - <i>DATASET</i> COM A MAGNITUDE ESPECTRAL AO DOBRO DA FREQUÊNCIA FUNDAMENTAL (100 HZ).....	84
TABELA 16 - RELAÇÃO DO TEMPO DE AMOSTRAGEM (JANELA) COM AS TAXAS DE ACERTO E O NÚMERO DE TREINO POR CLASSE DE AVARIA.....	85
TABELA 17 - <i>DATASET</i> COM A ANÁLISE ESPECTRAL DO VETOR DE PARK E OUTRAS <i>FEATURES</i> (JANELA DE 0,1 S).....	90
TABELA 18 - <i>DATASET</i> COM A ANÁLISE ESPECTRAL DO VETOR DE PARK E OUTRAS <i>FEATURES</i> (JANELA DE 0,1 S).....	91
TABELA 19 - <i>DATASET</i> COM A ANÁLISE ESPECTRAL DO VETOR DE PARK E OUTRAS <i>FEATURES</i> (JANELA DE 0,1 SEGUNDO).....	95
TABELA 20 - <i>DATASET</i> COM A ANÁLISE ESPECTRAL DO VETOR DE PARK E OUTRAS <i>FEATURES</i> (JANELA DE 0,1 SEGUNDO).....	97
TABELA 21 - SELEÇÃO DAS <i>FEATURES</i> COM BASE NO MAPA DE CALOR DA CORRELAÇÃO.....	98
TABELA 22 - CLASSIFICAÇÃO DO MODELO GERADO EM DADOS DESCONHECIDOS.....	101

TABELA 23 - TABELA CONTENDO OS ARQUIVOS COM OS DADOS ADQUIRIDOS DE TENSÃO E CORRENTES DO MOTOR. FREQUÊNCIA DE AMOSTRAGEM: 2 KHZ. TAMANHO: 2.000 PONTOS. INCLUÍDO DESFASAMENTO NA ALIMENTAÇÃO, TOTALIZA 34 ARQUIVOS.....	115
TABELA 24 - TABELA CONTENDO OS ARQUIVOS COM OS DADOS ADQUIRIDOS DE TENSÃO E CORRENTES DO MOTOR. FREQUÊNCIA DE AMOSTRAGEM: 2 KHZ. TAMANHO: 2.000 PONTOS. INCLUÍDO DESEQUILÍBRIOS DE AMPLITUDE (SUBTENSÃO) NA ALIMENTAÇÃO, TOTALIZA 45 ARQUIVOS.	117
TABELA 25 - TABELA CONTENDO OS ARQUIVOS COM OS DADOS ADQUIRIDOS DE TENSÃO E CORRENTES DO MOTOR. FREQUÊNCIA DE AMOSTRAGEM: 2 KHZ. TAMANHO: 2.000 PONTOS. INCLUÍDO DESEQUILÍBRIOS DE AMPLITUDE (SUBTENSÃO) E DE DESFASAMENTO NA ALIMENTAÇÃO, TOTALIZA 14 ARQUIVOS.....	119

Lista de Acrónimos e Abreviaturas

CISE – Centro de Investigação em Sistemas Electromecatrónicos

FFT – *Fast Fourier Transform*

HCA – *Hierarchical Cluster Analysis*

LSE – Laboratório de Sistemas Electromecatrónicos

MCSA – *Motor Current Signature Analysis*

SVM – *Support Vector Machines*

UBI – Universidade da Beira Interior

EPVA – *Extended Park's Vector Approach*

PVA – *Park's Vector Approach*

IoT – *Internet of Things*

CSA – *Current Signature Analysis*

VSA – *Voltage Signature Analysis*

IPSA – *Instantaneous Power Signature Analysis*

IPRS – *Instantaneous Reactive Power Signature Analysis*

IPAS – *Instantaneous Active Power Signature Analysis*

SVM – *Support Vector Machine*

DT – *Decision Tree*

Capítulo 1

1 Introdução

1.1 Enquadramento

Durante os séculos XVIII e XIX, ocorreram as pesquisas e desenvolvimentos que levaram à criação dos motores elétricos como os conhecemos hoje. Podemos destacar as pesquisas na área de geração de campo magnético a partir de uma corrente elétrica de Hans Christian Oersted, em 1820, bem como as pesquisas sobre o eletromagnetismo, de William Sturgeon, em 1825. Como não havia distribuição de energia elétrica para a população, nessa época, a descoberta da bateria por Alessandro Volta em 1800, deu um impulso extra para que as pesquisas, ainda incipientes, na área dos motores elétricos, progredissem.

Muitos inventores e pesquisadores trabalhavam em paralelo ao redor do mundo. Muitas descobertas, ainda que pequenas, eram realizadas todos os dias. Até que, finalmente, se apresentou o primeiro motor elétrico funcional, em maio de 1834, por Moritz Jacobi. A partir desse momento, vários outros protótipos de motores foram apresentados, mas sempre alimentados por baterias, o que tornava a sua operação muito cara.

Entre os anos de 1856 e 1893, houve progressos nos dínamos, geradores de eletricidade DC e dos motores, que convertem eletricidade em trabalho e que utilizavam o mesmo princípio construtivo. Entre os anos de 1885 e 1893, houve avanços significativos em sistemas de geração de energia trifásicos e motores de indução trifásicos [1].

Atualmente, os motores elétricos de indução trifásicos são amplamente utilizados na indústria, devido à sua simplicidade de construção, robustez e baixo custo. São utilizados em vários setores, como geração de energia, aplicações diversas na indústria, guindastes, movimentação de esteiras transportadoras, movimentação de cargas, entre outros. Alcançando um montante significativo, em termos financeiros, a utilização dos motores representa um papel muito importante na economia mundial [2]. Aproximadamente 69% da energia elétrica é consumida mundialmente no setor industrial e 46% da energia consumida no mundo é usada para alimentar motores elétricos [3].

Portanto, avarias em motores podem representar prejuízos enormes, diretamente associados aos custos de manutenção do equipamento ou indiretamente, devido aos custos originados devido à paragem total da produção. De entre as estratégias de manutenção existentes, a curativa, a periódica e a condicionada, esta última é a que melhor se coaduna com as características da ocorrência de avarias [4].

A manutenção condicionada é uma manutenção de carácter preventivo, entretanto, condicionada a algum evento predeterminado. Esta informação deve ser atualizada e precisa, dando a possibilidade de intervir antes do seu agravamento ou da paragem total do equipamento [4].

Os métodos de diagnóstico precoce adequam-se melhor às avarias de carácter evolutivo. Contudo, para se proceder a um diagnóstico preciso e inequívoco, é necessária não apenas a obtenção dos dados que serão analisados para o diagnóstico, como também, são necessários critérios previamente estabelecidos que permitam avaliar a necessidade ou não de intervenção de manutenção [4].

Muita investigação tem sido realizada nesta área, e muitas técnicas de aquisição de parâmetros de motores foram desenvolvidas. Considerando-se as técnicas não invasivas, podemos destacar as que são baseadas em *Motor Current Signature Analysis* (MCSA), que pode ser realizada *online*, sem interrupção da produção e com o motor dentro das condições normais de operação, podendo detetar as principais causas de falhas conhecidas para motores elétricos, como, excentricidade dinâmica ou estática, barras fraturadas ou fissuras em anéis do rotor, falha no estator (circuito aberto, em curto-circuito em uma ou mais bobinas) e conexão do estator anormal [5].

Para todos os casos acima mencionados, apesar de serem técnicas já conhecidas, são de difícil interpretação, sendo necessário um elevado grau de conhecimento e experiência prévia. Neste contexto, são introduzidos algoritmos de inteligência artificial, com o objetivo de tornar mais automatizado e eficiente o diagnóstico e a análise, através da MCSA [5].

Paralelamente, a Quarta Revolução Industrial, ou simplesmente, Indústria 4.0, traz consigo um conceito há muito perseguido pela humanidade, que é o de fábricas inteligentes. As fábricas inteligentes ou *Smart Factories* surgem para atender uma necessidade crescente de elevada flexibilidade e eficiência na fabricação dos produtos. A fim de manter a fábrica do futuro em operação contínua, torna-se ainda mais importante a deteção de avarias de forma precoce.

A *Internet of Things* (IoT), é a base para toda a Indústria 4.0, pois permite a interconexão de diferentes dispositivos. Logo, uma fábrica inteligente, utilizando um algoritmo de otimização de processos, pode levar em conta a condição de fiabilidade de motores de vários equipamentos, a produção da empresa, as ordens de produção em fila, entre outras

informações armazenadas no *Big Data*, e, assim, tomar decisões automatizadas para redirecionar a produção para outro equipamento e gerar um alerta para a sua manutenção, e atuar no momento mais adequado, em termos financeiros, permitindo assim a máxima eficiência da fábrica como um todo [6]. A Figura 1 descreve o funcionamento do *Smart Motor*.



Figura 1 - Esquema de funcionamento do SMART MOTOR [6].

1.2 Motivação

Os motores de indução trifásicos são usados em larga escala pela indústria, e a Quarta Revolução Industrial traz consigo uma demanda ainda maior. A fim de mantê-los em funcionamento com elevado grau de fiabilidade, e por forma a reduzir os custos associados a paragens e intervenções, torna-se necessário que seja possível diagnosticar a ocorrência de avarias ainda em estágio precoce de desenvolvimento, de modo a programar a manutenção antes da paragem total do equipamento.

Um método amplamente utilizado para o diagnóstico de avarias em motores elétricos é baseado em *Motor Current Signature Analysis* (MCSA). Destacam-se também, *Current Signature Analysis* (CSA), *Voltage Signature Analysis* (VSA), *Extended Park's Vector Approach* (EPVA), *Instantaneous Power Signature Analysis* (IPSA) and *Instantaneous Power Active and Reactive Signature* (IPRS) e (IPAS). As falhas que podem ser detetadas com a MCSA são excentricidade, barras rotóricas fraturadas, rolamentos danificados, curtos-circuitos estatóricos e até mesmo desgaste em equipamento, evidenciando uma elevada versatilidade na deteção dos modos de falha. Porém, para o diagnóstico, torna-se

necessário um especialista no assunto, com elevada experiência para a correta interpretação dos dados obtidos.

Neste trabalho, serão utilizadas técnicas de *Machine Learning* baseadas em EPVA para a criação de modelos que sejam capazes de efetuar um diagnóstico da ocorrência de curtos-circuitos entre espiras do enrolamento estatórico de motores de indução trifásicos, de forma automatizada.

O estudo conta com uma base de dados reais, adquiridos no Laboratório de Sistemas Electromecatrónicos (LSE) do Centro de Investigação em Sistemas Electromecatrónicos (CISE).

1.3 Objetivo

Pretende-se aplicar técnicas de *Machine Learning* baseadas na EPVA das correntes e das tensões, para geração de um modelo eficiente de diagnóstico da ocorrência de curtos-circuitos em motores de indução trifásicos, utilizando SVM e *Decision Tree* (DT), que seja capaz de funcionar em condições normais da rede de alimentação e em condições de desequilíbrio de amplitude ou fase.

1.4 Metodologia

Este trabalho destina-se à obtenção de grau académico de Mestre em Engenharia e Gestão Industrial, sendo abordado o método de investigação científica que visa a implementar a tecnologia de inteligência artificial, para otimização do processo de diagnóstico de avarias em motores de indução trifásicos, sendo a sua aplicação diretamente associada ao campo das Ciências Aplicadas. A investigação seguirá um processo indutivo e contará com a pesquisa em trabalhos pré-existentes sobre o assunto.

A sua contribuição é fundamentada em uma abordagem híbrida, com a utilização de um método de diagnóstico tradicionalmente utilizado, associado à inteligência artificial. Para além disso, este documento segue um critério de metodologia científica, que será abordado em maiores detalhes neste tópico.

A metodologia utilizada é ilustrada na Figura 2.



Figura 2 - Fluxo do processo de investigação desta dissertação.

1.4.1 Pesquisa Exploratória

A questão fundamental da pesquisa é: “É possível a aplicação de inteligência artificial para o diagnóstico de avarias em motores de indução trifásicos?”. Para responder a esta questão foi realizada uma pesquisa exploratória na base de dados do IEEE Explore, MDPI, Science Direct, bem como em livros e websites das áreas de inteligência artificial, motores elétricos e técnicas de diagnóstico de avarias, de modo a se obter o atual estado da arte nesta área, bem como as principais informações para a fundamentação dos conceitos prévios à sua implementação. Contudo, a pesquisa revelou que a abordagem de inteligência artificial aplicada ao diagnóstico de avarias, na sua maioria, se focava em técnicas de processamento de sinais e estatística.

A hipótese é fundamentada nesta falta de informação ou “gap” pertinente à aplicação de inteligência artificial, nomeadamente o *Machine Learning*, em conjunto com uma técnica já conhecida de diagnóstico, nomeadamente o *Extended Park’s Vector Approach* (EPVA), o que motivou e fundamentou este projeto de investigação e sugeriu uma nova pergunta de pesquisa: “É possível a aplicação de inteligência artificial para o diagnóstico de avarias em motores elétricos de indução trifásicos a partir do EPVA?”.

1.4.2 Projeto de Investigação

Para a comprovação da hipótese levantada, uma parte fundamental do projeto de investigação é a determinação dos meios necessários para a sua realização, em especial, em casos experimentais. Os ensaios serão realizados no CISE | LSE, onde se dispõe de motores de indução trifásicos dentro das características esperadas, equipamentos para a simulação

de carga, controlo de velocidade do motor e sistemas de aquisição de dados com elevada precisão, de forma a gerar as amostras que, compiladas, irão gerar o banco de dados necessários para a conclusão do projeto.

1.4.3 Execução do Projeto

A execução do projeto baseou-se na aplicação de algoritmos de *Machine Learning* nos dados reais adquiridos em laboratório, dentro de condições controladas e pré-estabelecidas. Para esta finalidade, foi utilizado o software de nome comercial Matlab, bem como a *ToolBox Classification Learner*. Esta aplicação foi dividida em 5 fases, as quais são descritas em maiores detalhes na Tabela 1.

Tabela 1 - Resumo da execução do projeto de dissertação.

Fases	Descrição
Fase 1	Abordagem inicial, onde se implementa o <i>Machine Learning</i> com o EPVA, num motor sem desequilíbrios nas tensões de alimentação, de modo a comprovar a eficiência de ambos, conforme descrito na literatura.
Fase 2, 3 e 4	Após a conclusão da fase inicial, a proposta é de realizar o diagnóstico de avaria em motores, em condições de desequilíbrio nas tensões de alimentação. Entretanto, esta é justamente uma deficiência do EPVA que se espera ser mitigada com o <i>Machine Learning</i> .
Fase 5	Fase na qual os dados serão validados, utilizando-se como principal critério, amostras de validação desconhecidas e aleatórias que serão processadas, de forma automatizada pelo modelo de <i>Machine Learning</i> .

Após a aplicação das fases de 1 a 5, os seus resultados serão analisados e comparados quantitativamente, com base no critério de precisão e generalização. Todo este processo será descrito na presente dissertação.

1.5 Estrutura da Dissertação

Esta dissertação encontra-se dividida em 8 capítulos, os quais serão brevemente comentados abaixo:

Capítulo 1: Este capítulo traz informações pertinentes quanto ao enquadramento, à motivação, objetivo e estrutura da dissertação. Deste modo, é possível encontrar os principais conceitos, explorados de uma forma resumida, de forma a explicar o conteúdo deste documento, bem como seus objetivos.

Capítulos 2, 3 e 4: Estes capítulos visam organizar a base de conhecimento necessária para o entendimento deste trabalho. Logo, o capítulo 2 revela os conceitos básicos de funcionamento de máquinas elétricas. O capítulo 3, traz os principais conceitos relativos aos métodos de diagnóstico atualmente utilizados, com maior ênfase em MCSA, PVA e EPVA. O capítulo 4, identifica as principais técnicas e métricas associadas ao *Machine Learning*.

O Capítulo 5 descreve os equipamentos e detalha o método de aquisição das amostras utilizadas neste trabalho.

O Capítulo 6 acaba por reunir todo o conceito, anteriormente exposto, a fim de se obter uma aplicação prática dos algoritmos de *Machine Learning* para o diagnóstico de avarias relativas a curtos-circuitos estatóricos. Devido à complexidade desta implementação, este capítulo foi dividido em 5 secções, bem como as discussões dos resultados.

Os Capítulos 7 e 8, apresentam as conclusões e os trabalhos futuros.

Capítulo 2

2 Máquinas Elétricas

2.1 Motores de Indução Trifásicos de Rotor em Gaiola de Esquilo

De um modo geral, os motores elétricos são dispositivos capazes de transformar energia elétrica em energia mecânica. Dentre os seus diversos tipos construtivos, pode-se destacar o motor de indução, que é amplamente utilizado na indústria, devido ao seu baixo custo, facilidade de transporte, limpeza, simplicidade de comando, construção simples e versátil; o qual pode ser dimensionado de acordo com as diferentes necessidades de carga e com excelentes rendimentos, na conversão da energia elétrica em energia mecânica.

De acordo com o seu princípio construtivo, a sua alimentação pode ser monofásica ou trifásica, a partir de tensões alternadas. O estator é a sua parte fixa, enquanto que o rotor é a sua parte móvel. Um exemplo de motor de indução trifásico pode ser visto em detalhe na Figura 3.

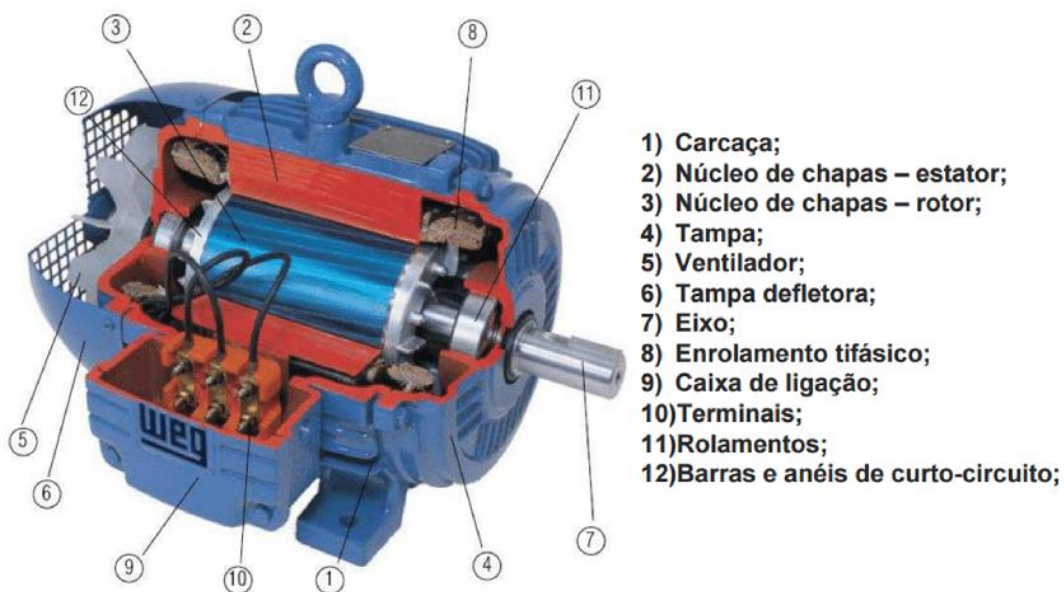


Figura 3 - Vista em corte de um motor de indução trifásico [7].

2.2 Estator

Nos motores de indução trifásicos, o estator é responsável pela geração do campo magnético girante. O estator é alimentado por uma corrente alternada, que percorre os seus enrolamentos, dispostos a 120° (ou $2\pi/3$) entre si, e que formam o seu circuito elétrico. Com o objetivo de aumentar a sua eficiência magnética, os enrolamentos são inseridos dentro de um material ferroso, umas chapas de ferro com poucas perdas magnéticas que, assim, formam o circuito magnético estatórico. Estas chapas têm um formato circular e são revestidas de um verniz eletricamente isolante. As chapas podem possuir diferentes quantidades de ranhuras, (36, 48, 54, etc.), e possuem um formato semifechado. Também é comum a adoção de canais de ventilação que facilitam a passagem de ar, de forma a arrefecer o motor durante seu regime de funcionamento. As suas bobinas são compostas por condutores enrolados que formam as espiras e que são isolados com esmalte. Para reforçar o isolamento, é inserido no interior das ranhuras um material isolante plástico, que forra a sua parte inferior e ficam imobilizados na ranhura por material isolante plástico. Em motores de potência elevada, alternativamente, os enrolamentos estatóricos são formados por bobinas rígidas, compostas por barras de cobre isolado [8].

2.3 Rotor

Nos motores de indução trifásicos, o rotor é responsável por receber o campo magnético girante do estator e, devido ao seu princípio construtivo, é gerada uma fem induzida que, por sua vez, é transformada num binário. Na sua construção, geralmente em formato cilíndrico, o núcleo magnético rotórico, de maneira similar ao estator, é composto por chapas com material similar às utilizadas no núcleo estatórico. Estas, por sua vez, possuem ranhuras fechadas próximas da superfície externa e, a quantidade de ranhuras rotóricas está relacionada com o número de ranhuras estatóricas, com o objetivo de evitar o aparecimento de ruídos ou harmónicos indesejados durante o regime de funcionamento do motor. Existem dois tipos de construção do rotor, nomeadamente de “gaiola de esquilo” ou bobinado. O rotor em gaiola de esquilo é composto por uma estrutura que lembra uma gaiola e o material constituinte é o alumínio, ou ligas de alumínio. No rotor em gaiola de esquilo, os condutores encontram-se, permanentemente, em curto-circuito. O rotor do tipo bobinado, como o seu próprio nome indica, possui um conjunto de enrolamentos, geralmente ligados em Y, e os seus terminais podem ser acedidos, diferentemente do rotor em gaiola de esquilo. Devido ao seu princípio construtivo, o rotor bobinado pode ter a

resistência das bobinas alterada através de resistências externas, e como consequência, podem alterar-se as características de velocidade e binário [8].

2.4 Velocidade Síncrona (n_s)

A velocidade síncrona é determinada pela velocidade de rotação do campo girante, e pode ser calculada de acordo com o número de polos ($2p$) e a frequência (f_1) da rede, em Hertz [7].

$$n_s = \frac{60 \cdot f_1}{p} = \frac{120 \cdot f_1}{2p} \text{ (rpm)} \quad (1)$$

onde:

n_s : Velocidade de sincronismo;

f_1 : Frequência da rede (50 Hz para países europeus);

p : Número de pares de polos do motor;

O deslizamento (s) é a relação entre as velocidades de rotação síncrona (n_s) e a velocidade do rotor (n).

$$s = \frac{n_s - n}{n_s} \times 100 \text{ [%]} \quad (2)$$

A frequência da fem induzida no rotor, f_2 é obtida pelo deslizamento e pela frequência da rede:

$$f_2 = s \cdot f_1 \quad (3)$$

2.5 Potência e Rendimento

A fonte de alimentação fornece ao motor elétrico uma potência elétrica P_{ele} que, no caso dos motores de indução trifásicos, é a soma das potências das três fases, onde cada fase, P_f é obtida através da tensão por fase V_f e corrente por fase I_f .

$$P_{ele} = 3 \times P_f = 3 \times V_f \times I_f \times \cos(\varphi) \text{ [W]} \quad (4)$$

Ligando-se os três sistemas monofásicos entre si, obtém-se um sistema equilibrado com as três tensões de igual valor desfasadas de igual ângulo (120°). A ligação entre fases poderá ser feita em triângulo (Δ) ou em estrela (γ).

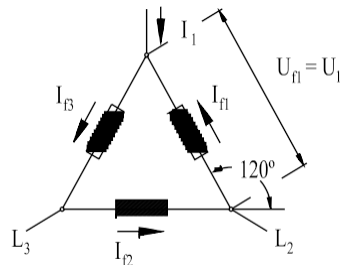


Figura 4 - Ligação em triângulo [7].

Observando a Figura 4, ligação em triângulo, temos:

$$U = U_f \text{ e } I = \sqrt{3} \cdot I_f \quad (5)$$

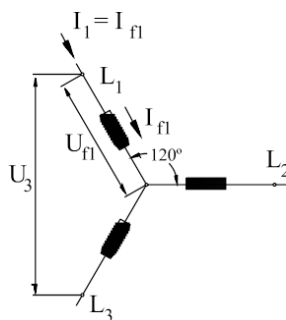


Figura 5 - Ligação em estrela [7].

Observando a Figura 5, ligação em estrela, temos:

$$U = \sqrt{3} \cdot U_f \text{ e } I = I_f \quad (6)$$

Numa rede elétrica, um motor pode ser aproximado a um circuito resistivo, capacitivo e indutivo. Contudo, é possível desprezar as componentes resistivas e capacitivas, de modo a aproximar o motor a um circuito indutivo equivalente. Deste modo, a sua corrente apresenta-se atrasada em relação à tensão de alimentação e, portanto, a sua potência pode ser dividida em potência ativa (P) e potência reativa (Q). A potência aparente (S) é dada pela soma vetorial da potência ativa e reativa. A potência reativa acaba por não contribuir na realização de trabalho.

A Figura 6 representa a relação existente entre as potências, ativa, reativa e aparente.

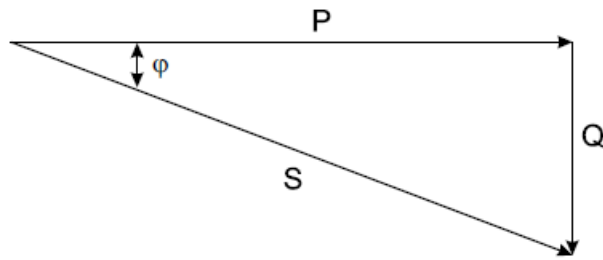


Figura 6 - Triângulo das potências (carga ativa e reativa) [7].

Temos:

$$P = \sqrt{3} \times V \times I \times \cos(\varphi) [W] \quad (7)$$

$$S = \sqrt{3} \times V \times I [V \cdot A] \quad (8)$$

$$Q = \sqrt{3} \times V \times I \times \sin(\varphi) [V \cdot Ar] \quad (9)$$

$$\cos(\varphi) = \frac{P}{S} \quad (10)$$

O fator de potência $\cos(\varphi)$, conforme a Figura 6, é obtido pela razão direta entre a potência ativa e a aparente e é um indicador importante no aproveitamento de energia elétrica extraída da rede. Como o motor de indução é um circuito altamente indutivo, importa que este fator seja tendencialmente 1, pois idealmente pretende-se uma potência reativa a menor possível dado que não realiza trabalho [7].

O rendimento define a eficiência na transformação de energia realizada pelo motor, ou seja, a transformação de energia elétrica em energia mecânica. Portanto, temos:

$$\eta = \frac{P_u}{P_a} \cdot 100 = \frac{736 \cdot P (CV)}{\sqrt{3} \cdot V \cdot I \cdot \cos(\varphi)} \cdot 100 [\%] \quad (12)$$

onde:

P_u : Potência mecânica útil disponível no veio de saída do motor

P_a : Potência elétrica absorvida na rede

Capítulo 3

3 Métodos de Diagnóstico de Avarias em Motores de Indução Trifásicos

3.1 Introdução

As máquinas elétricas exercem um papel fundamental na execução de atividades em geração de energia, estações de bombeamento de água, aplicações industriais, guindastes, transportes e movimentações em geral. Além disso, estão relacionadas com elevados investimentos financeiros, bem como com segurança e fiabilidade [2].

A Indústria 4.0 é baseada em automação, que por sua vez, é altamente dependente de motores elétricos, necessários para qualquer movimentação, seja de veículos, esteiras, robôs, entre outros. A consideração destes aspetos traz consigo uma necessidade ainda maior de monitorização da condição e diagnóstico de avarias em tempo real. Para além disso, as informações recolhidas poderão ser, posteriormente, processadas de forma a gerar um prognóstico, e assim, programar a intervenção no momento mais oportuno, do ponto de vista económico e técnico [2], [6].

3.2 Métodos de Diagnóstico

De entre as mais diversas técnicas de deteção de avarias, a *Motor Current Signature Analysis* ou Análise de Assinatura da Corrente do Motor (MCSA) é um método de análise *online* de corrente para detetar problemas num motor de indução trifásico enquanto ele ainda está operacional e em serviço. A MCSA é considerada o método de deteção de avarias mais amplamente utilizado, pois permite detetar as falhas comuns da máquina elétrica, como o curto-circuito entre espiras, as barras de rotor quebradas ou fraturadas, deterioração do rolamento, entre outras [9].

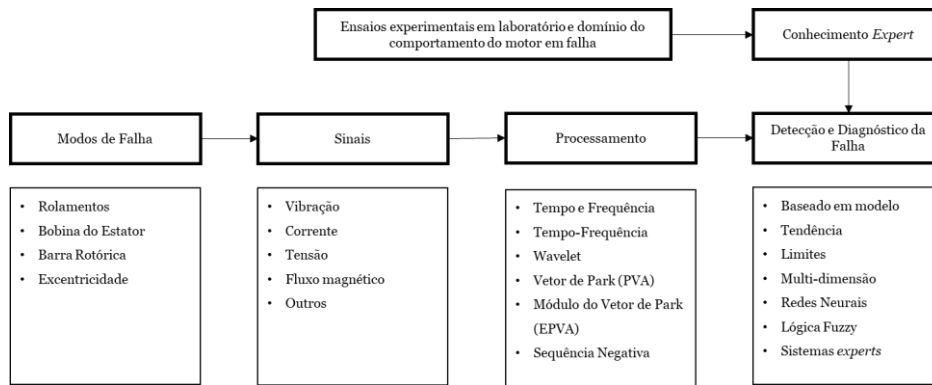


Figura 7 - Processo de monitorização da condição do motor, *online* (Adaptado de [9]).

A Figura 7 descreve o processo de diagnóstico. Nos diversos modos de falha que pode ser de rolamentos, bobinas do estator, barras rotóricas ou de excentricidade, o motor exibe um comportamento característico. Para perceber se o comportamento do motor é normal ou anormal, é preciso a aquisição de sinais do motor, que podem ser vibração, corrente, tensão, fluxo magnético, entre outros. Entretanto, estes sinais precisam ser processados para que exista uma compreensão sobre os mesmos. Neste caso, podem ser processados no domínio do tempo, tal como na PVA, da frequência, onde se pode obter informações do EPVA, do tempo-frequência, para obtenção de *wavelets*, entre outras informações. Com base neste conjunto de informações, somado com a experiência anterior de ensaios e modelação da avaria, é possível a sua deteção e diagnóstico, que podem ser auxiliados, por modelos, tendências, limites, análise multi-dimensional (com vários parâmetros), redes neuronais, lógica *fuzzy*, e sistemas periciais.

3.3 Funcionamento da MCSA

A MCSA tem como princípio a monitorização da corrente fornecida ao motor. Esta, por sua vez, alimenta os enrolamentos do estator que são usados como transdutores, adquirindo sinais (através de fem induzida) do rotor, mas que também revelam informações importantes a respeito do próprio estator.

Os sinais são adquiridos através de sondas e gravados no domínio do tempo, que podem ser então levados a um analisador de espectro ou um sistema especializado, conforme ilustra a Figura 8.

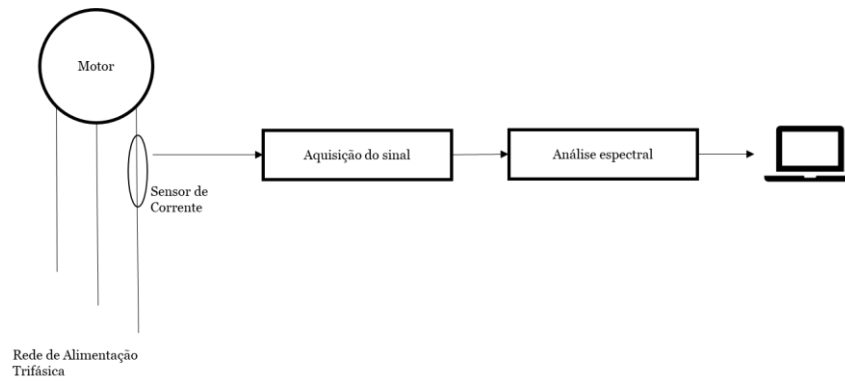


Figura 8 - Sistema de monitorização de corrente estatórica (Adaptado de [5]).

As correntes de um motor ideal deveriam corresponder a ondas puramente sinusoidais. Entretanto, na prática, vários harmónicos são detetados. Algumas avarias nos componentes dos motores produzem anomalias correspondentes no campo magnético e alteram a indutância própria do motor, sendo refletidas no espectro de corrente de alimentação do motor como bandas laterais. Com base nestas assinaturas, é possível identificar avarias no motor, bem como estimar a sua severidade.

3.4 Diagnóstico baseado em MCSA

Os modelos matemáticos para a deteção de avarias para cada modo de falha serão apresentados nas próximas secções, bem como a explicação sobre a avaria.

3.4.1 Curtos-Circuitos entre Espiras no Estator

O curto-circuito entre espiras caracteriza-se pelo curto-circuito entre condutores do mesmo enrolamento e da mesma fase devido à degradação do material isolante. É uma avaria considerada em estágio inicial que levará a máquina à paragem total. Porém, deve-se salientar que, dependendo da severidade do curto-circuito, poder levar segundos para a sua degradação total ou para que o acionamento de outros meios de proteção seja ativado. A falha pode aparecer no final da bobina ou ainda dentro da cavidade. As características deste tipo de falha são ideais neste caso de estudo, pois podem ser detetadas ainda em estágio inicial e têm carácter evolutivo [2].

A Figura 9 exibe um desenho esquemático dos possíveis modos de falha associados a avarias estatóricas em motores de indução trifásicos. A Figura 10 ilustra uma bobina real com espiras em curtos-circuitos.

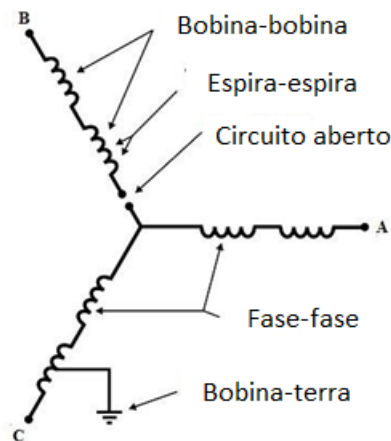


Figura 9 - Modos de falha nos estatores de motores elétricos de indução trifásicos (Adaptado de [10]).

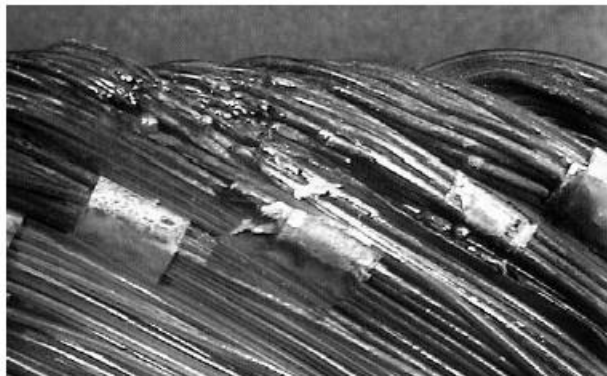


Figura 10 - Curto-circuito estático entre espiras no enrolamento [10].

Quando existe um curto-circuito estático, a resistência da fase afetada é diminuída proporcionalmente com a severidade do curto-circuito. Logo, um motor de indução trifásico com este tipo de falha, alimentado com três fases simétricas, possui uma fase com maior consumo de corrente que as outras duas, o que gera um campo magnético girante assimétrico. Isto gera uma alteração na frequência do terceiro harmônico, que pode ser detetada com um aumento na sua amplitude [2].

Contudo, o aumento da amplitude do terceiro harmônico está associado também a outros desequilíbrios tais como um fornecimento de tensão desequilibrado, assimetria entre as bobinas do motor, conexões com elevada resistência, entre outros fatores. O aumento da amplitude não implica diretamente a severidade da falha. Todos estes fatores se somam para tornarem a monitorização através do terceiro harmônico não confiável, especialmente em casos de baixa severidade, onde a sua análise pode levar a falsos alarmes [2].

A frequência que revela este tipo de falha é dada pela seguinte expressão:

$$f_{st} = f_g \left(\frac{n}{p} (1 - s) \pm k \right) \quad (13)$$

onde:

f_{st} : componente da frequência relacionada com o curto-circuito

f_g : frequência da alimentação

n : 1, 2, 3...

p : número de pares de polos

k : 1, 3, 5...

Nota-se que, neste caso, o modelo indica uma região de frequência, portanto, os sinais obtidos no domínio do tempo devem ser processados e passados ao domínio da frequência, para que seja possível a sua análise. A Figura 11 ilustra a identificação de harmônicos relativos a curtos-circuitos estatóricos:

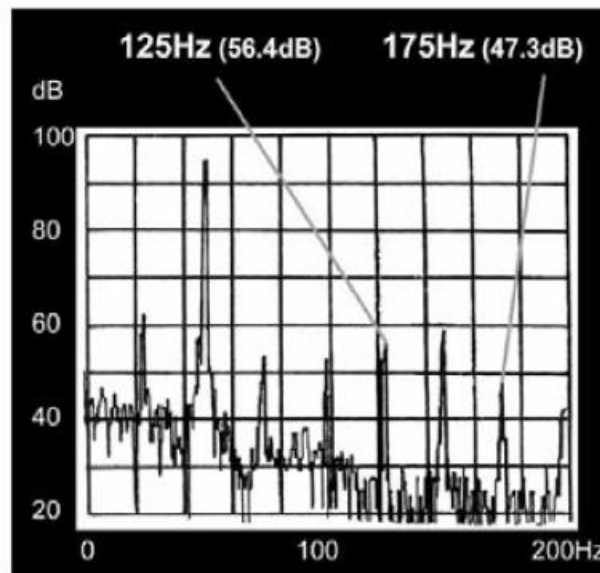


Figura 11 - Componentes da frequência correspondentes a curtos-circuitos estatóricos [5].

3.5 Diagnóstico Baseado em PVA e EPVA

A transformação direta-quadratura-zero (DQZ, DQ0 ou dq0) é um tensor que rotaciona o referencial de três elementos vetoriais com o objetivo de simplificar a análise. A transformada dq0 é o produto da transformada de Clarke e da transformada de Park, que foi proposta, inicialmente, no ano de 1929 por Robert H. Park.

Esta transformação é frequentemente utilizada no contexto da engenharia eletrotécnica em circuitos trifásicos. A transformação pode ser utilizada para rotacionar o referencial de uma corrente alternada, transformando-o em uma corrente contínua, de modo a simplificar os cálculos em inversores para o controlo de máquinas elétricas trifásicas e depois transformar o sinal novamente em corrente alternada. Na análise de máquinas elétricas trifásicas, a transformação pode ser aplicada na análise do estator e do rotor num único referencial, eliminando assim o efeito de indutâncias variantes no tempo, transformando-o num sistema invariante no tempo [11], [12].

No caso de motores elétricos trifásicos alimentados por apenas três condutores, com neutro flutuante, em caso de funcionamento normal, apresentará a componente homopolar nula. O cálculo das suas componentes d e q, relativas à corrente de alimentação (I_a , I_b e I_c), são as seguintes [4]:

$$I_d = \left(\frac{\sqrt{2}}{\sqrt{3}}\right)I_a - \left(\frac{1}{\sqrt{6}}\right)I_b - \left(\frac{1}{\sqrt{6}}\right)I_c \quad (14)$$

$$I_q = \left(\frac{1}{\sqrt{2}}\right)I_b - \left(\frac{1}{\sqrt{2}}\right)I_c \quad (15)$$

$$\text{Módulo} = \sqrt{I_d^2 + I_q^2} \quad (16)$$

Para condições normais de funcionamento, onde não há avarias, o sistema pode ser descrito da seguinte maneira:

$$I_d = \left(\frac{\sqrt{6}}{2}\right)I_{max} \text{sen}(wt) \quad (17)$$

$$I_q = \left(\frac{\sqrt{6}}{2}\right)I_{max} \text{sen}(wt - 2\pi) \quad (18)$$

onde:

I_{max} : valor máximo da corrente elétrica do sistema de alimentação trifásico (A);

w : frequência angular relativa ao sistema de alimentação (rad/seg);

t : tempo (seg).

Utilizando-se esta abordagem (Park's Vector Approach ou PVA), é possível a deteção de avarias através da análise das formas geométricas formadas pelas projeções das componentes d (no eixo x) e q (no eixo y), que formam o Vetor de Park [13]. Em condições normais de funcionamento, a forma é circular. Em condições de avaria estatórica, as equações (23) e (24) já não são válidas, a forma torna-se cada vez mais elíptica, de acordo com o incremento na severidade [2]. A partir do modelo do motor, foram simuladas

diferentes severidades de curtos-circuitos, e seu efeito pode ser analisado através da Figura 12.

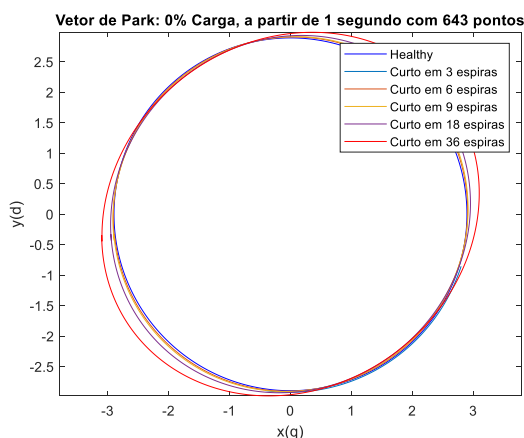


Figura 12 - Comparação do Vetor de Park num motor sem avarias, e com diferentes severidades de curtos-circuitos estáticos.

Entretanto, o PVA possui limitações para a detecção correta da severidade, pois é afetado também por outros parâmetros [2].

Logo, o EPVA mostra vantagens latentes em relação ao PVA. A análise espectral do módulo do Vetor de Park revela uma elevada precisão na correta classificação de diferentes severidades de curtos-circuitos estáticos [2]. Em condições normais de funcionamento, o módulo do Vetor de Park não apresenta oscilações e o seu espectro possui apenas uma componente DC. Quando ocorrem desequilíbrios, tanto em fase quanto em amplitudes, nas fases a, b ou c, estes são refletidos na forma de uma componente DC dominante e uma componente AC que aumenta de acordo com a severidade do desequilíbrio [14]. Esta componente AC tem a sua ocorrência no dobro da frequência fundamental, para o caso particular desta avaria, funciona como um indicador de desequilíbrios que pode ser aplicado tanto nas correntes, como nas tensões de alimentação. A Figura 13, ilustra o comportamento do módulo do Vetor de Park nas correntes a, b e c, em condições equilibradas e desequilibradas.

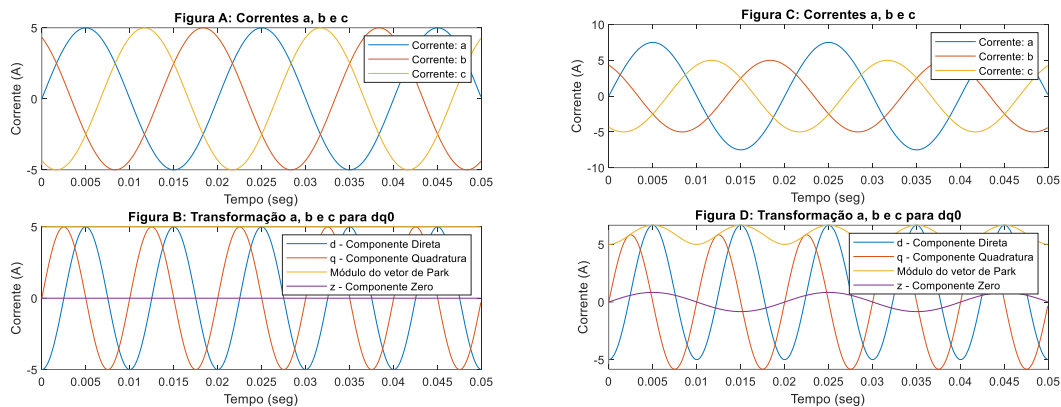


Figura 13 - (A) correntes equilibradas; (B) transformação para dq0; (C), a corrente a apresenta uma amplitude mais elevada que as outras, na ordem de 50%; (D) Nota-se que este efeito é observado também no módulo do Vetor de Park na Figura D. Deve-se notar que as componentes d e q não estão no mesmo referencial do rotor.

A Figura 13 revela também o comportamento das componentes d, q e zero, que neste caso, não estão no mesmo referencial do rotor, servindo apenas para ilustrar a relação entre as componentes d e q e seu respetivo módulo, de acordo com a equação (22). É possível notar que o desequilíbrio nas correntes a, b e c, resulta em desequilíbrio nas componentes d e q e que, por fim, afeta o módulo do Vetor de Park, que passa a ter uma oscilação. O módulo do Vetor de Park no domínio do tempo e seu espectrograma é exibido isoladamente na Figura 14.

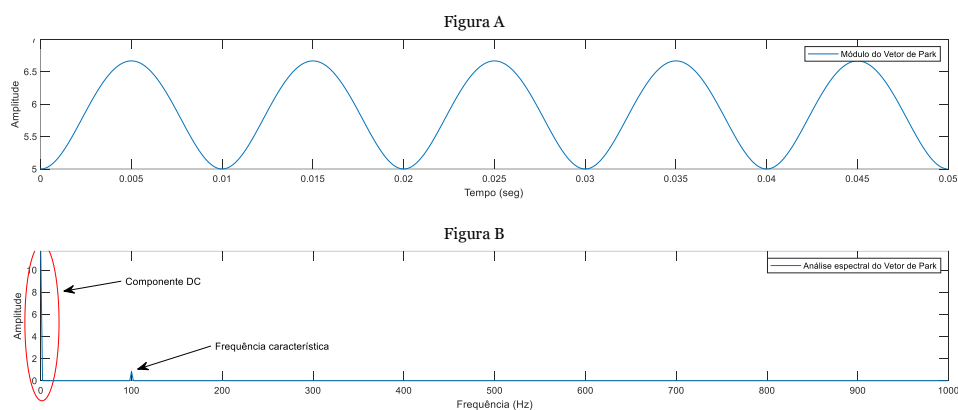


Figura 14 - (A) oscilações no módulo do Vetor de Park devido a desequilíbrios nas correntes; (B) análise espectral do módulo do Vetor de Park. É possível observar a componente DC na frequência 0 Hz.

A análise espectral do módulo do Vetor de Park apresenta um aumento na amplitude na região do dobro da frequência fundamental (2fs). No caso exposto nas Figura 13 e Figura 14, a frequência fundamental é de 50 Hz, e a sua resposta na frequência está em 100 Hz. As magnitudes no dobro da frequência fundamental (2fs) podem ser comparadas de forma a identificar a severidade dos curtos-circuitos com elevada precisão [14].

3.6 Diagnóstico Baseado em Componentes Simétricas

O método de componentes simétricas é usado para simplificar a análise de falhas, de modo a converter um sistema trifásico desequilibrado em dois conjuntos de fasores equilibrados e um conjunto de fasor monofásico, ou componentes simétricos.

Esses conjuntos de fasores são chamados de componentes de sequência positiva, negativa e zero. Estas componentes permitem uma análise simples e objetiva de sistemas de energia com falhas ou em condições de desequilíbrio. Pode ser aplicado em motores de indução trifásicos para a detecção de curtos-circuitos estatóricos, onde a componente de sequência negativa é analisada. Entretanto, esta componente apresenta limitações, pois reflete as avarias não só relativas aos curtos-circuitos, mas também aquelas pertinentes à construção do próprio motor e outros desequilíbrios de alimentação [15].

A conversão de um sistema abc para componentes simétricas é dado por:

$$\begin{bmatrix} I_0 \\ I_1 \\ I_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & \alpha & \alpha^2 \\ 1 & \alpha^2 & \alpha \end{bmatrix} \begin{bmatrix} I_A \\ I_B \\ I_C \end{bmatrix} \quad (19)$$

ou,

$$I_0 = \frac{1}{3}(I_A + I_B + I_C) \quad (20)$$

$$I_1 = \frac{1}{3}(I_A + \alpha I_B + \alpha^2 I_C) \quad (21)$$

$$I_2 = \frac{1}{3}(I_A + \alpha^2 I_B + \alpha I_C) \quad (22)$$

onde:

$$\alpha = 1 \angle 120^\circ$$

$$\alpha^2 = 1 \angle 240^\circ$$

I_0 = sequência zero

I_1 = sequência positiva

I_2 = sequência negativa

Nota-se que as componentes simétricas foram dadas em termos de correntes, I. Entretanto, o mesmo se aplica para tensões V. Na Figura 15, é possível observar o desenho esquemático das componentes simétricas.

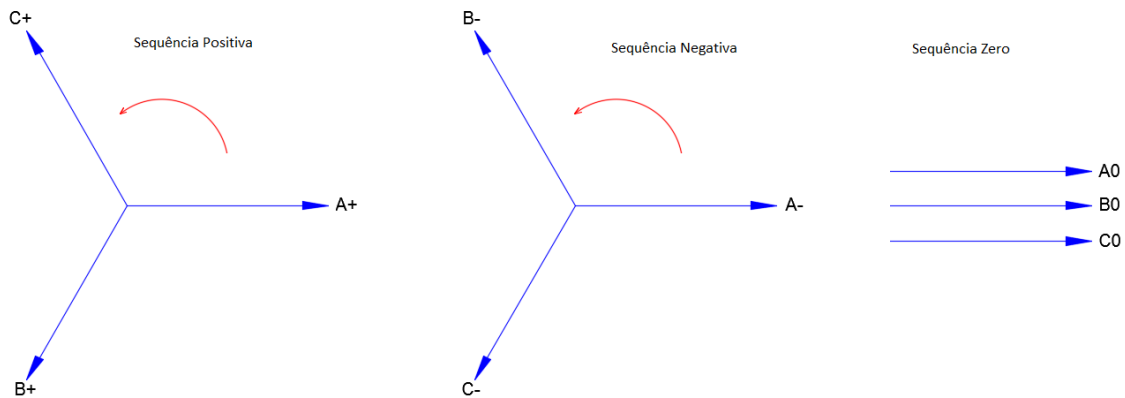


Figura 15 - Imagem ilustrativa das componentes de sequência positiva, negativa e zero [16].

De forma resumida, as componentes podem ser interpretadas da seguinte maneira:

- Em condições ideais, todos os geradores e sistemas de potência devem gerar apenas componentes de sequência positiva.
- Em condições equilibradas num sistema trifásico, as componentes de sequência negativa e zero são nulas.
- As sequências negativas nas tensões resultam em sequência negativa nas correntes dos motores, o que gera um fluxo contrário ao movimento do motor, causando um aumento na temperatura do motor.
- As componentes zero de correntes de um motor são usadas para a detecção do fluxo de corrente para o neutro. Este princípio é utilizado para a detecção de circulação de correntes para a terra, em sistemas com o neutro aterrado.
- Em sistemas de conexão em estrela (Y), com ponto neutro flutuante (neutro não conectado à terra), a componente zero é nula.
- As componentes de sequência negativas são usadas nas tensões e nas correntes para detecção de desequilíbrios.

Capítulo 4

4 Inteligência Artificial

A discussão no campo da inteligência artificial iniciou-se na década de 50, em carácter mais científico, com a apresentação do artigo de Alan Turing, intitulado “*Computing Machinery and Intelligence*”. Desde então, houve muita investigação e avanços, tanto em termos de algoritmos como em termos computacionais [17].

Hoje em dia, já existem *softwares* que usam a inteligência artificial, para aumentar a sua eficiência, num universo gigante de dados, chamado de *Big Data*, tirando assim a grande vantagem de efetuar cálculos através de computadores. Utilizando algoritmos de *Machine Learning*, podem criar-se correlações de forma automática ou controlada, como no caso dos filtros “*anti-spam*” dos *e-mails*, recomendadores de filmes na Netflix, detetores de fraudes em cartões de crédito, entre outros [17].

A inteligência artificial pode ser definida como o conjunto de teorias e técnicas usadas para criar máquinas capazes de simular a inteligência. Tanto as máquinas como os programas são associados à inteligência humana. Contudo, pode haver confusão quanto à utilização dos termos. Assim, o *Machine Learning* está dentro do campo de estudos da inteligência artificial. O *Deep Learning* está inserido no *Machine Learning* [18], como ilustrado na Figura 16.

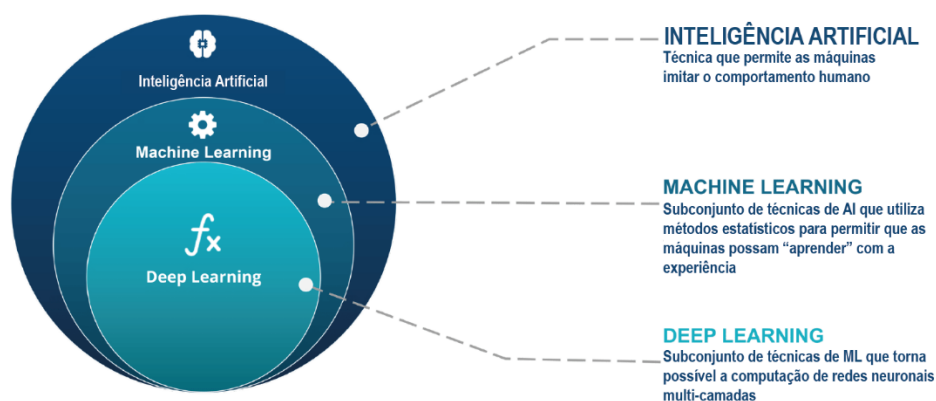


Figura 16 - Inteligência artificial vs *Machine Learning* e *Deep Learning* [19].

4.1 Machine Learning

O *Machine Learning* é um campo da inteligência artificial muito discutido no mundo da computação. É ainda conhecido por “aprendizagem estatística”. O seu objetivo é tornar a máquina ou o computador, capaz de fornecer soluções para problemas complicados através do processamento de uma quantidade astronômica de informações. Isso oferece uma oportunidade de analisar e destacar as correlações que existem entre duas ou mais dadas situações e de prever as suas diversas implicações [18].

Segundo Whele [17], podemos classificar os sistemas de aprendizagem de máquina de acordo com o tipo de supervisão recebida, que pode ser supervisionado, não supervisionado ou ambos. Os algoritmos baseados em aprendizagem supervisionada utilizam uma base de dados previamente conhecida. Os mais conhecidos são: *random forests*, *decision trees*, e *support vector machines (SVM)*.

Os sistemas de aprendizagem de máquina não supervisionados são baseados num conjunto de dados desconhecidos, onde o algoritmo é responsável por descobrir padrões no conjunto de dados, o que auxilia no apoio a decisões. Os algoritmos comumente utilizados para o agrupamento automático, de acordo com os padrões identificados, são: *K-means*, *hierarchical cluster analysis (HCA)* [17].

Para além disso, também podem ser utilizados sistemas chamados semi-supervisionados, que utilizam ambas as abordagens acima mencionadas [17].

4.1.1 Decision Tree (DT)

No seu artigo “Induction of Decision Trees” [20], Quinlan descreve a utilização de uma abordagem do uso de *Machine Learning* em *Decision Tree*. Uma árvore de decisão ou “Decision Tree” (DT) é uma estrutura semelhante a uma árvore, em que um nó interno representa o teste de um atributo, cada ramificação representa o resultado do teste, cada nó folha representa o rótulo da classe e uma decisão é tomada após o cálculo de todos os atributos. Um caminho da raiz para uma folha representa as regras de classificação. Assim, uma árvore de decisão consiste em três tipos de nós [21]:

- Raiz (*Root*)
- Ramo (*Branch*)
- Folha (rótulo da classificação) (*Leaf*)

A “Decision Tree” é um algoritmo que se destaca pela facilidade de interpretação dos resultados apresentados e que prevê regras que levam a um evento ou decisão. Para além disso, apresenta também um menor custo computacional, comparativamente com outros

algoritmos, o que fez sua implementação ser muito elevada nos últimos anos [22]. A Figura 17 ilustra o esquema de funcionamento de uma *Decision Tree*.

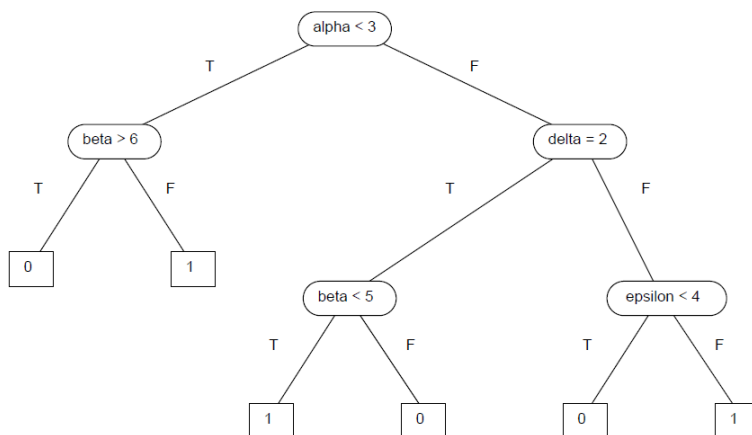


Figura 17 - Exemplo gráfico de uma *Decision Tree* [22].

A Figura 17 apresenta uma *Decision Tree*, onde alfa é o nó principal (raiz), beta, delta e épsilon são os ramos ou ramificações, e as decisões são 0 ou 1. Deste modo, é possível seguir um caminho lógico. Por exemplo, se 1 corresponder a efetuar uma ação e “0” corresponder a não efetuar, pode dizer-se que a “Decision Tree” é baseada na estrutura modelada da árvore, supondo os valores de acordo com a Tabela 2.

Tabela 2 - Descrição de uma *Decision Tree*, em forma de tabela.

Passo	Ação	Resultado	Caminho
1	Calcular Alpha	Alpha = 6	F
2	Calcular Delta	Delta = 3	F
3	Calcular Epsilon	Epsilon = 2	T
4		Decisão Final =	0

Os atributos de cada nó são escolhidos de forma heurística ou estatística da medida da impureza, por exemplo, com a medida da impureza de Gini ou ganho de informação entropia. Gini é definido, por:

$$Gini = 1 - \sum_i (p_i)^2, \text{ onde } p_i \text{ é a probabilidade de cada classe} \quad (23)$$

$$Entropy = -p \log_2(p) - q \log_2(q) \quad (24)$$

onde:

p = probabilidade de sucesso de um nó.

q = probabilidade de falha de um nó.

4.1.2 Support Vector Machine (SVM)

Proposto por Vladimir N. Vapnik e Alexey Ya. Chervonenkis, em 1963, o conceito do SVM é criar um hiperplano que separe duas classes de maneira ótima, tendo assim uma margem máxima entre o hiperplano e as observações [23].

A Figura 18 ilustra três possibilidades de classificação. Observa-se que existem duas classes e uma linha de separação entre elas. Assim, a função da linha de separação, ou seja, a classificação é obtida [21].

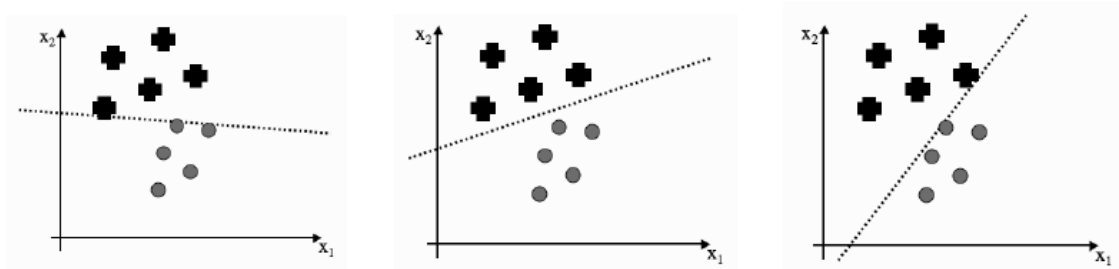


Figura 18 - Diferentes critérios de classificação [21].

Entretanto, para que a função seja otimizada, pode analisar-se mais cuidadosamente a linha de separação, conforme ilustrado na Figura 19.

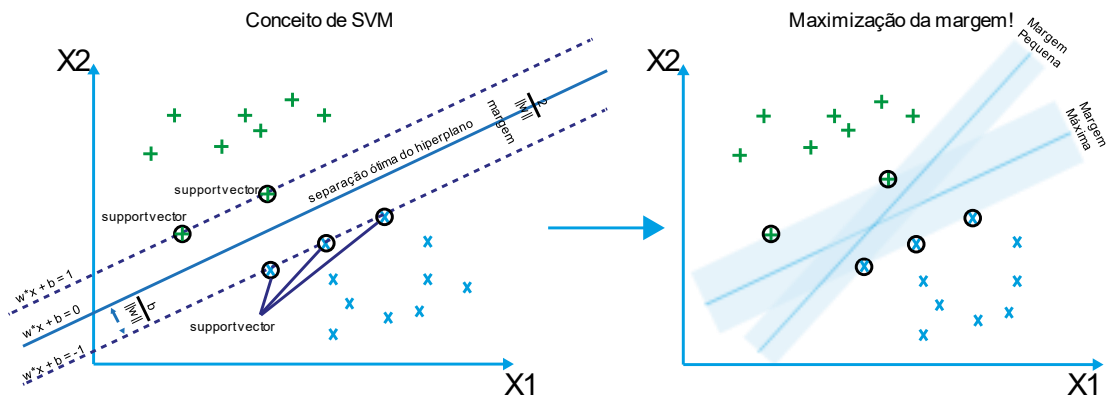


Figura 19 - Support Vector Machine Hiperplane [21].

A Figura 19 ilustra os *Support Vector Machines*, bem como a margem ótima de separação. O conceito de SVM baseia-se no conceito de *margin*. Deste modo, a separação entre as classes não é feita apenas por uma linha divisória, mas sim por uma margem. O critério da classificação é definido como sendo a margem máxima que separe as duas classes conforme ilustrado na Figura 20.

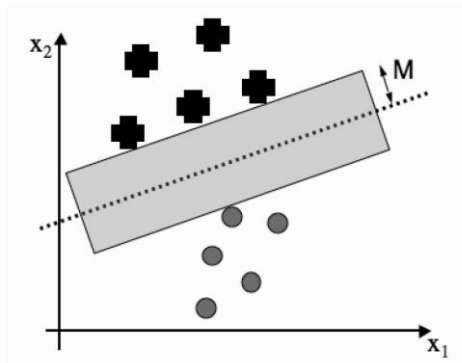


Figura 20 - A margem é a maior região que podemos colocar, separando as classes sem haver pontos dentro, onde a caixa é feita de duas linhas paralelas ao limite de decisão [24].

Para uma dada margem de valor M , pode dizer-se que qualquer ponto x , onde $w^T x + b \geq M$ é um “mais”, enquanto $w^T x + b \leq -M$ é um “círculo”. O hiperplano que separa as duas classes é dado por: $w^T x + b = 0$. Uma vez que o hiperplano ideal maximiza a margem, converte-se automaticamente num problema de otimização quadrática convexa:

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 \\ & \text{sujeito a } y_i(w^T x_i + b) \geq 1 \end{aligned} \quad (25)$$

A equação final do SVM pode ser escrita matematicamente, como:

$$L = \sum_i d_i - \frac{1}{2} \sum_i \alpha_i \alpha_j y_i y_j (X_i X_j) \quad (26)$$

4.2 Feature Extraction

Para fins de descrição e análise, um sinal pode ser definido simplesmente como uma função matemática associada ao tempo e à frequência. De acordo com as propriedades estatísticas do próprio sinal, o sinal pode ser dividido em diferentes categorias. É comum que os sinais sejam divididos em estáticos e não estáticos, onde o estático denota que as propriedades estatísticas de um sinal são invariáveis com o tempo. Para sinais determinísticos, isso basicamente implica que eles são compostos inteiramente de algumas sinusóides com frequências discretas e, portanto, os seus espectros de frequência consistem em linhas discretas nas frequências dessas sinusóides. Uma vez que as frequências, amplitudes e fases iniciais (ou seja, no tempo zero) dessas sinusóides são conhecidas, o valor do sinal pode ser determinado a qualquer momento. Assim, eles são chamados de sinais determinísticos.

Os métodos tradicionais de processamento de sinais podem ser divididos em três categorias:

- domínio do tempo;
- domínio da frequência;
- domínio do tempo-frequência.

O pré-processamento do sinal tenta cancelar o ruído no sinal e melhorar a relação entre o ruído e o sinal, preparando-o para o pós-processamento, que é adequado para qualquer categoria de sinal. O processamento de sinais no domínio do tempo inclui principalmente a análise de correlação e extração de características estatísticas, no domínio do tempo. A análise de autocorrelação pode capturar as características periódicas de um sinal, enquanto que a análise de correlação cruzada pode medir o grau de relevância entre dois sinais diferentes. Além disso, os recursos estatísticos no domínio do tempo podem ser extraídos para refletir as propriedades estatísticas de um sinal de sua forma de onda, no domínio do tempo. O processamento de sinal no domínio da frequência geralmente denota a Transformada de Fourier (FT) baseada na série de Fourier e, posteriormente, a transformada rápida de Fourier (FFT). As suas funções básicas são um conjunto de funções trigonométricas. Assim, a técnica da FT limita-se ao processamento de sinais estáticos. Supondo que um sinal num curto período de tempo é estático, algumas características do domínio da frequência podem ser extraídas para refletir as propriedades estatísticas do domínio da frequência dos espectros de frequência ou potência do sinal. Na prática, um sinal não estático não pode ser completamente descrito no domínio do tempo ou no domínio da frequência, sozinho. Para refletir a distribuição de um sinal não estático, tanto no domínio do tempo, quanto no domínio da frequência, as técnicas de processamento de sinal no domínio do tempo-frequência foram desenvolvidas rapidamente. Existem muitas técnicas representativas, como transformada de Fourier de curto prazo (STFT), distribuição de Wigner-Ville (WVD), transformada de *wavelet* (WT), transformada de pacote *wavelet* (WPT) e até mesmo transformada de Hilbert-Huang (HHT). Essas técnicas são amplamente aplicadas ao diagnóstico de avarias mecânicas. Visto que diferentes técnicas de processamento de sinais caracterizam desempenhos distintos, é importante para diferentes sinais selecionar uma técnica de processamento de sinal adequada para adquirir informações mais úteis.

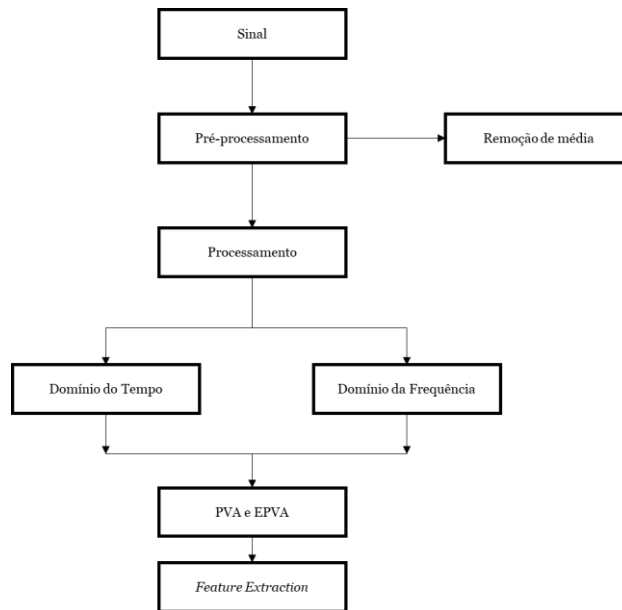


Figura 21 - Esquema de *feature extraction* do PVA e do EPVA (Adaptado de [25]).

A Figura 21 descreve, de um modo geral, o processo de *feature extraction*, baseado em PVA e em EPVA do domínio do tempo e no domínio da frequência, respectivamente.

4.2.1 Remoção da Média (Componente DC)

O processamento de sinais no domínio do tempo para a extração de *features*, usa principalmente a análise de correlação e dados estatísticos temporais. Algumas vezes, também se torna necessário um pré-processamento, para a remoção de componentes de corrente direta (DC) ou média, tendência ou o uso de filtro para sinais com elevado ruído.

Na Figura 22, temos um sinal que possui a influência da componente DC.

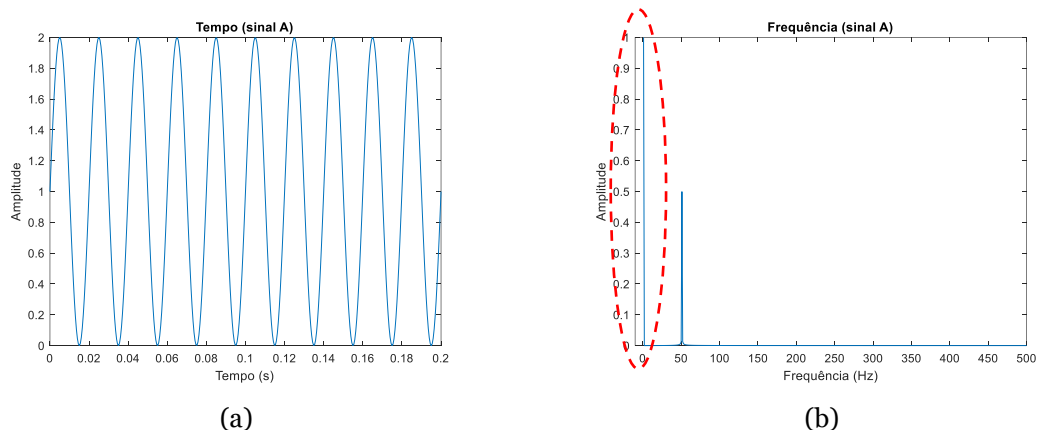


Figura 22 - Sinal A com componente DC: (a) representação no domínio do tempo; (b) representação no domínio da frequência.

A Figura 22 (a) possui um sinal sinusoidal de média 1 do domínio do tempo, ao passo que a Figura 22 (b) corresponde à sua análise espectral, onde a frequência 0 Hz reflete a média.

Na Figura 23, temos o mesmo sinal, entretanto, com a remoção da componente DC.

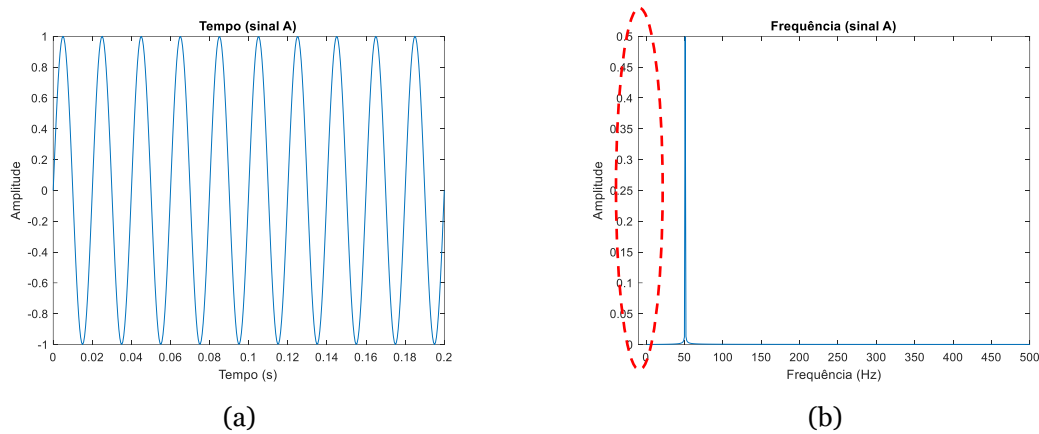


Figura 23 - Sinal A após remoção da componente DC: (a) no domínio do tempo; (b) no domínio da frequência.

A componente DC pode ser removida, através da remoção da sua média no domínio do tempo, e reflete-se no domínio da frequência, como é ilustrado pela Figura 23 (a) e (b), respectivamente. Esta técnica é especialmente útil na comparação de sinais com diferentes componentes DC [26].

4.2.2 Dados Estatísticos – Domínio do Tempo

Quando uma falha estatórica ocorre, ela gera perturbações que podem ser observadas nas correntes medidas num motor. As amplitudes e distribuições podem ser alteradas no domínio do tempo. Portanto, os recursos estatísticos no domínio do tempo podem ser extraídos de suas formas de para refletir tal condição [26].

Os principais indicadores são: valor médio (X_m), desvio padrão (X_{sd}), mediana (X_{med}), amplitude da raiz (X_{root}), raiz quadrada média (X_{rms}), pico máximo absoluto (X_{peak}), valor de pico a pico ($X_{peak2peak}$), a variância (X_{var}) e a raiz da soma dos quadrados (X_{rssq}) [24]. Entretanto, em avarias com menor severidade é recomendado o uso de indicadores, como: assimetria ($X_{skewness}$), curtose ($X_{kurtosis}$), fator de crista (X_{crest}) [26]. Seguem-se as suas respectivas fórmulas.

A média, é dada por:

$$Xm = \frac{\sum_{n=1}^N x(n)}{N} \quad (27)$$

O desvio padrão é definido como:

$$Xsd = \sqrt{\frac{\sum_{n=1}^N (x(n) - Xm)^2}{N - 1}} \quad (28)$$

Amplitude da raiz, dada por:

$$Xroot = \left(\frac{\sum_{n=1}^N \sqrt{|x(n)|}}{N} \right)^2 \quad (29)$$

Raiz quadrada média, é dada por:

$$Xrms = \left(\frac{\sum_{n=1}^N x(n)}{N} \right)^2 \quad (30)$$

O pico máximo absoluto, é dado por:

$$Xpeak = \max |x(n)| \quad (31)$$

A assimetria (*skewness*), é dada por:

$$Xskewness = \frac{\sum_{n=1}^N (x(n) - Xm)^3}{(N - 1)X^3sd} \quad (32)$$

A curtose (*kurtosis*), é dada por:

$$Xkurtosis = \frac{\sum_{n=1}^N (x(n) - Xm)^4}{(N - 1)X^4sd} \quad (33)$$

A fator de crista (*crest*), é dada por:

$$Xcrest = \frac{Xpeak}{Xrms} \quad (34)$$

A mediana, é dada por:

$$X_{med} = \frac{n + 1}{2} \quad (35)$$

A variância, é dada por:

$$X_{var} = \frac{\sum_{n=1}^N (x(n) - Xm)^2}{N - 1} \quad (36)$$

A raiz da soma dos quadrados, é dada por:

$$X_{rssq} = \sqrt{\sum_{n=1}^N |x(n)|^2} \quad (37)$$

O valor pico a pico, é dado por:

$$X_{peak2peak} = X_{rms} 2\sqrt{2} \quad (38)$$

4.2.3 Dados Estatísticos – Domínio da Frequência

O processamento do sinal no domínio do tempo possui algumas deficiências, que podem ser superadas através da análise espectral. Pequenas variações no formato, amplitude ou na frequência podem ser facilmente identificadas através do seu espectro, cuja obtenção se baseia na teoria da transformada de Fourier (FT), onde um sinal periódico é dividido numa soma de vários componentes sinusoidais, em virtude da série de Fourier. O objetivo desta transformada é passar um sinal periódico contínuo para o domínio da frequência. Entretanto, em termos computacionais, os sinais são discretos, o que leva à utilização da transformada discreta de Fourier (DFT), que possui a mesma finalidade anteriormente descrita para um sinal contínuo, porém aplica-se a sinais discretos [25]. Pode ser calculada através da seguinte equação:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{i2\pi}{N}kn} = \sum_{n=0}^{N-1} x_n \left(\cos\left(\frac{2\pi}{N}kn\right) - i \operatorname{sen}\left(\frac{2\pi}{N}kn\right) \right) \quad (39)$$

Deve ter-se em conta a frequência de amostragem, que, segundo o teorema de Nyquist, as frequências que são possíveis de serem obtidas estão associadas à metade do comprimento do sinal.

4.3 Feature Selection

A seleção das melhores *features* (*feature selection*) e consequentemente a redução dimensional, acaba por trazer inúmeros benefícios, visto que a complexidade de um algoritmo classificador ou regressor está diretamente associada ao número de entradas ou *features*. Este fator determina o tempo, complexidade e o número de treino necessário para a geração de um modelo adequado. Para além disso, ainda que o algoritmo de *Machine Learning* seja capaz de dar mais ênfase para aquelas *features* que são mais importantes, ainda restam outras *features* que não possuem grande relevância e podem reduzir a sua performance, tanto em termos de treino, quanto na sua posterior precisão em relação a novos dados. Como vantagem adicional, para modelos que possuem poucas *features*, é possível ter uma ideia mais precisa sobre o que acontece dentro dos algoritmos, o que torna a sua otimização mais fácil e eficiente [25].

4.3.1 Correlação

Podemos citar a correlação e o valor p como duas ferramentas estatísticas largamente utilizadas para identificar corretamente a correlação ou associação entre um dado conjunto de variáveis X e Y. Deve-se notar, contudo, que a correlação não exprime uma relação de causalidade, e que este cálculo identifica apenas a correlação linear.

O coeficiente de correlação de Pearson de X e Y, denotado por $Cor(X, Y)$ ou ρ_{xy} , é definido como:

$$\rho_{xy} = Cor(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}} \quad (40)$$

onde:

\bar{X} e \bar{Y} são as médias, respetivamente de X e Y.

A Figura 24 mostra a comparação entre diferentes correlações entre X e Y. A Figura 24 (a), à esquerda, exhibe uma forte correlação negativa. A Figura 24 (b), ao centro, indica uma correlação fraca. A Figura 24 (c), exhibe uma forte correlação positiva.

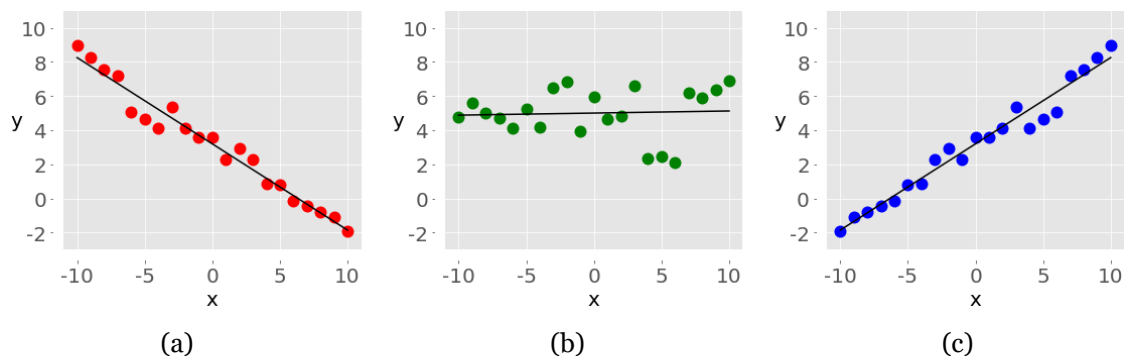


Figura 24 - Comparação entre diferentes correlações entre X e Y: (a) Forte correlação negativa; (b) Correlação fraca; (c) Forte correlação positiva (Adaptado de [27]).

A correlação de duas variáveis, X e Y é expressa num intervalo de -1 a +1. Caso o coeficiente de correlação seja ρ_{xy} ou $Cor(X, Y) = 0$, existe uma correlação linear muito fraca ou nula entre estas duas variáveis. Caso a resposta seja igual a +1, a correlação é proporcional (positiva) e caso a resposta seja -1 a correlação é inversamente (negativa) proporcional [28]. Contudo, quando os coeficientes de correlação assumem valores mais reduzidos, como por exemplo, + 0,2 ou - 0,12, pode ser mais difícil a interpretação da existência ou não da correlação. Neste caso, uma alternativa é a aplicação de um teste estatístico, para a determinação do valor p.

4.3.2 Valor p

O valor p varia de 0 a 1 e tem como principal função a identificação da hipótese nula ou da simples significância da correlação entre duas variáveis, X e Y, por exemplo, e que utilizem um critério de confiança, por exemplo, 0,05 (ou 95%). Neste caso, o teste da hipótese nula, passa a ser o contrário da correlação. Caso o resultado seja menor que 0,05, existe uma probabilidade maior que 95% de confiança de que exista correlação entre as variáveis X e Y. Caso o valor seja de maior que 0,05, a confiança de que existe a correlação é cada vez menor. Logo, quanto menor o valor de p, maior a correlação, quanto maior o valor de p, menor é a correlação.

Na prática, quando utilizadas para o propósito do *Machine Learning*, a correlação resulta em valores que podem ser comparados de acordo com a sua importância. O valor p, por sua vez, indica a confiança na correlação e acaba por funcionar como outro critério de seleção, confirmando ou rejeitando o critério anterior. Portanto, em conjunto, as duas informações criam um critério de avaliação de *features*, de modo a descartar as redundantes, e selecionar aquelas que têm maior relevância. O valor p pode ser calculado a partir da correlação e é dado como [28]:

$$t = \rho \sqrt{\frac{n-2}{1-\rho^2}} \quad (41)$$

onde:

n : grau de liberdade (número de amostras)

ρ : Correlação entre X e Y

Nota-se que o valor encontrado em t está de acordo com tabela t Student de distribuição e que também depende do grau de liberdade. Esta tabela deve ser consultada para encontrar a densidade probabilística da hipótese nula.

A Figura 25 mostra um exemplo de uma curva de distribuição T, com 20 graus de liberdade (GL), onde $t = -2$ e $+2$, quando consideradas as duas caudas. Portanto, o valor t convertido resulta em $0,02963 + 0,02963 = 0,05926$, ou aproximadamente, 6%, que corresponde ao valor p .

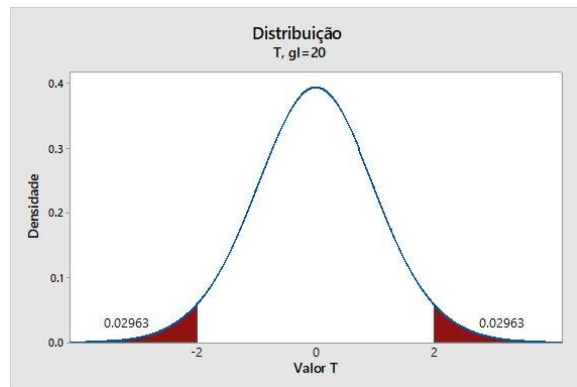


Figura 25 - Curva de distribuição T. Identificação da região $t = 2$, correspondente a uma densidade probabilística de $0,02963 \times 2$, ou aproximadamente 6% (Adaptado de [29]).

Deve notar-se que no caso deste exemplo, a probabilidade da hipótese nula é muito baixa, na ordem de 6%. Contudo, fica fora do valor normalmente utilizado de significância de 95%.

4.4 Métricas de Avaliação e Validação

Um ponto muito importante para a avaliação e comparação são as métricas utilizadas para esta finalidade. Entretanto, deve ter-se em consideração o tipo de aplicação que se utiliza, onde as variáveis que se querem determinar podem ser caracterizadas tanto como quantitativas como qualitativas. Alguns exemplos de variáveis quantitativas incluem a idade, a altura, o peso, o salário de uma pessoa, entre outros. No caso das variáveis

qualitativas, podemos citar o caso do género de uma pessoa, masculino ou feminino, a marca de um produto, A, B ou C, o diagnóstico de avaria em motor elétrico, em boas condições, 3 espiras em curto-circuito, 12 espiras em curto-circuito, com desequilíbrio de fase na tensão de alimentação, entre outros. Apesar de não ser uma distinção absoluta, em geral, os problemas quantitativos são tratados com o nome de regressão, ao passo que os problemas qualitativos são tratados como problemas de classificação [30].

4.4.1 Regressão

Um dos principais objetivos da utilização de métricas de avaliação de modelos de previsão é obter uma boa ideia da sua taxa de erro. Entretanto, dentro dos dados conhecidos, ou seja, aqueles utilizados no treino, esta tarefa acaba por ser relativamente mais simples. Entretanto, em dados desconhecidos, ou seja, aqueles que não foram utilizados no treino e que servem de validação, esta tarefa acaba por ser mais difícil.

Para a regressão, uma das principais métricas de avaliação é o *mean square error* (MSE), e é dado por [30]:

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2 \quad (42)$$

onde:

N = Número de observações

y_i = Resposta real

$\hat{f}(x_i)$ = Resposta prevista pelo modelo

A Figura 26 ilustra o MSE relativo ao treino e, posteriormente, a generalização do modelo.

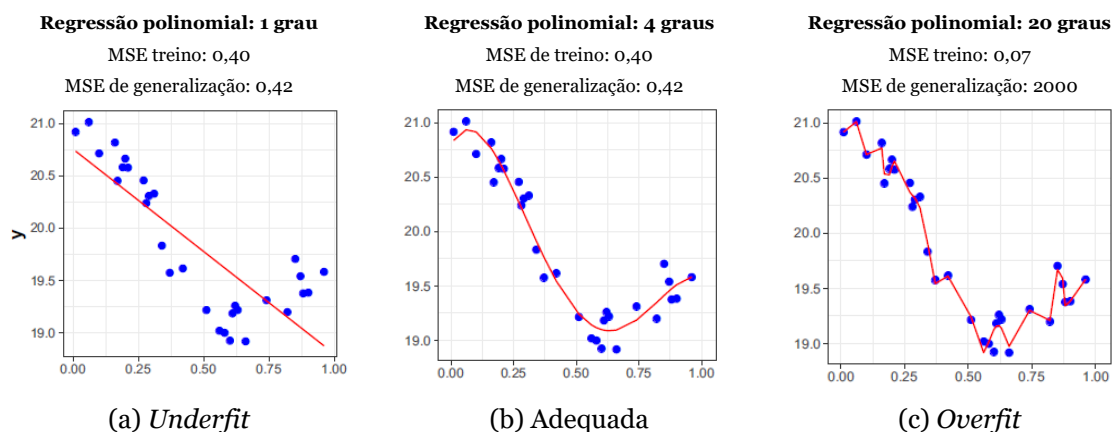


Figura 26 - (a) Regressão linear polinomial, com 1 grau de liberdade na condição de *underfit*; (b) Regressão linear polinomial adequada, com 4 graus de liberdade; (c) Regressão linear polinomial, com 20 graus de liberdade em condição de *overfit* (Adaptado de [30]).

Deste modo, a Figura 26 ilustra três diferentes condições, que podem ser resumidas na Tabela 3.

Tabela 3 - Explicação resumida de diferentes tipos de treino, com *Underfit*, Adequado e *Overfit*.

Figura	Descrição
(a)	Exibe uma regressão linear polinomial, com 1 grau de liberdade na condição de <i>underfit</i> . A sua principal característica é o seu elevado erro, tanto nos dados de treino quanto nos dados de validação (generalização).
(b)	Exibe uma regressão linear polinomial adequada, com 4 graus de liberdade. A sua principal característica é a sua baixa taxa de erro, tanto nos dados de treino quanto nos dados de validação (generalização).
(c)	Exibe uma regressão linear polinomial, com 20 graus de liberdade em condição de <i>overfit</i> . A sua principal característica é a sua taxa de erro muito próxima de zero, nos dados de treino. Entretanto, nos dados de validação (generalização) exibe um erro muito elevado.

4.4.2 Classificação

Apesar do método para classificação ser diferente da regressão, é necessária a utilização de métricas de avaliação do desempenho dos modelos, de maneira análoga ao item de regressão, descrito anteriormente. Assim, é possível avaliar o seu desempenho, compará-lo a outros modelos e estimar a sua generalização.

Para a classificação, uma das principais métricas de avaliação é a taxa de erro, e é dada por [30]:

$$\overline{\text{erro}} = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i) \quad (43)$$

onde:

n : Número total de observações

y_i : Resposta real

\hat{y}_i : Resposta prevista pelo modelo

\neq : Nota-se que neste caso é utilizado o operador de comparação de diferença, ou seja, a resposta é 1 quando a condição é verdadeira, $I(y_i \neq \hat{y}_i)$ e zero, quando $I(y_i = \hat{y}_i)$. Desta maneira, é possível encontrar o erro absoluto, e ao dividi-lo pelo número total de observações obtém-se o erro percentual.

A precisão é dada pelo inverso da taxa de erro:

$$\text{Precisão} = 1 - \overline{\text{erro}} \quad (44)$$

A Figura 27 exhibe 3 condições de classificação:

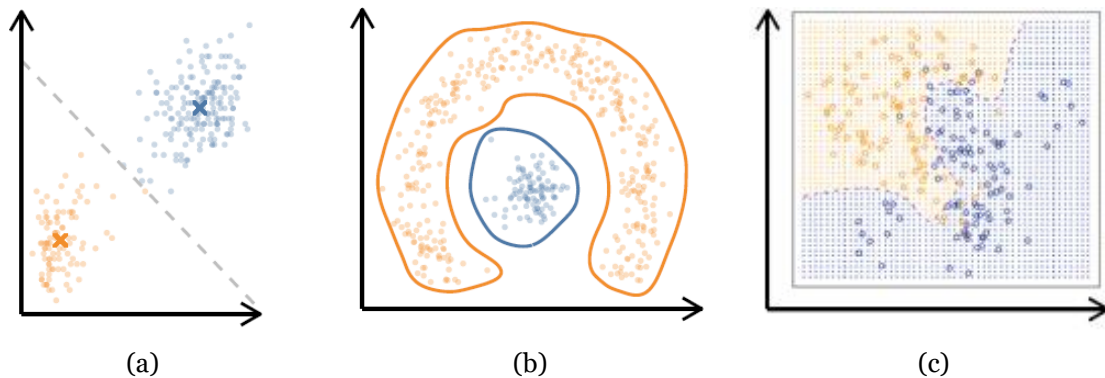


Figura 27 - (a) Resultado de um agrupamento (*clustering*) clássico com $k = 2$; (b) Resultado de um agrupamento baseado em densidade; (c) Classificação baseada em Bayes (Adaptado de:[30]).

A Figura 27(a), apresenta uma classificação binária, linear, onde é possível identificar que não há um erro elevado. A Figura 27(b), apresenta uma classificação polinomial, onde não apresenta erros de classificação. A Figura 27(c), apresenta uma classificação binária polinomial, entretanto, os seus dados não podem ser separados como os anteriores. A princípio, a Figura 27(c) apresenta maior probabilidade de ter problemas de *overfit*, pois no sentido em que se diminui o erro nos dados de treino, o oposto pode ocorrer nos dados de validação. Desta forma, deve haver um compromisso entre a taxa de erro e a generalização [30]

4.4.3 Generalização e Validação Cruzada

Diferentes estratégias são utilizadas para tentar prever a generalização. Os seus objetivos são o de estimar a precisão (ou o erro) de um determinado modelo num conjunto de dados diferente daquele utilizado no treino. A validação cruzada ou *cross-validation* (CV), consiste na divisão dos dados em duas partes (uma ou mais vezes), onde uma parte é utilizada para o treino do modelo e outra para a validação do mesmo, e desta maneira, estimar o erro e generalizá-lo a partir da média dos diferentes erros. Esta estratégia é conhecida como *K-Fold Cross-Validation* [31]. A Figura 28, resume a sua estratégia de funcionamento.

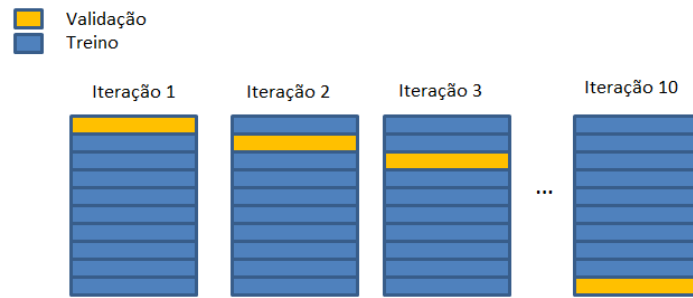


Figura 28 - Desenho esquemático do K-Fold Cross-Validation com $k = 10$ (Adaptado de [32]).

Na Figura 28, é possível observar um *dataset*, que utiliza a técnica de validação cruzada. Na iteração 1 *dataset*, é subdividido em 10 partes ($k = 10$), onde 9 são utilizadas para treino, e o primeiro subconjunto é utilizado para validação, e tem o seu erro calculado. A iteração 2, segue o mesmo princípio, entretanto, é o segundo subconjunto de dados que é utilizado para a validação, e tem o seu erro calculado. Desta forma, são realizadas 10 iterações ($k = 10$), a partir da qual são gerados 10 erros, de onde é calculada a sua média [33].

A validação cruzada, apesar de ser uma boa estratégia, e de ser comumente utilizada para medir o desempenho de modelos de previsão, pode sofrer uma influência negativa do tamanho do *dataset* utilizado [33]. A Figura 29 ilustra o tamanho do conjunto de treino x precisão (1- erro):

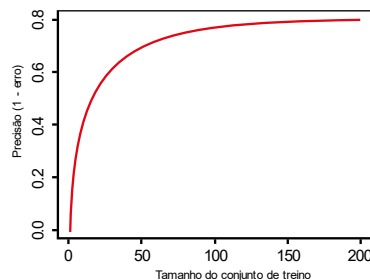


Figura 29 - Curva hipotética de aprendizagem de um classificador Precisão x. Tamanho do conjunto de dados de treino. Validação cruzada com $k = 5$.

Na Figura 29, é possível observar que o tamanho do conjunto de dados de treino influencia na precisão. Caso o conjunto de dados de treino seja muito pequeno, pode ser induzida uma baixa precisão. No caso do exemplo acima, é possível notar que para um tamanho de 25 amostras de treino, para uma validação cruzada com $k = 5$, nota-se que a precisão fica em torno dos 40%. Para 100 amostras, nas mesmas condições, a precisão passa a ser de 70%. Entretanto, entre 100 e 200, o ganho com um conjunto de dados de treino maior passa a diminuir, até alcançar o seu máximo de 80%. Este comportamento, mostra um compromisso entre o tamanho da amostra e possíveis erros de generalização devido a um conjunto de dados treino inadequados.

Capítulo 5

5 Aquisição de Dados Experimentais

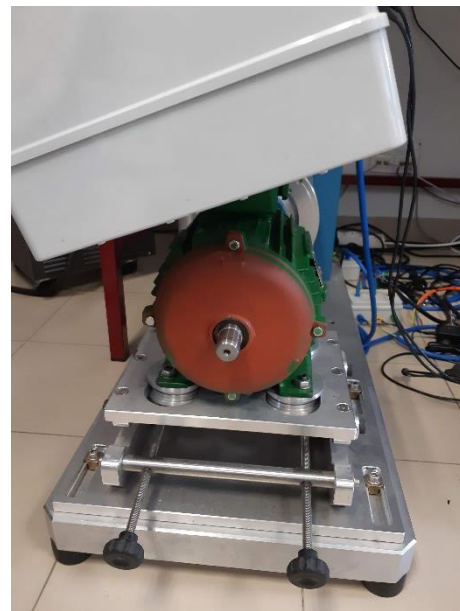
Para a aquisição de dados reais do motor de indução trifásico, foi necessário montar uma bancada de ensaio. Na Figura 30 estão descritos os elementos utilizados para o ensaio e aquisição dos sinais de corrente e tensão.

5.1 Motor de Indução Trifásico

Foi utilizado um motor elétrico de indução trifásico, com ligação em estrela (Y) e ponto neutro flutuante.



(a)



(b)

Figura 30 - Motor de indução trifásico utilizado nos ensaios experimentais: (a) Foto ilustrativa do motor; (b) Montagem real na bancada de ensaios.

Fabricante / Modelo: WEG / W22

Descrição: Motor ligado em estrela (Y), 400 VAC, 50 Hz, 1435 rpm, 4,56 A, PF 0,80, IE Code IE3, Rendimento 87%.

Observação: Motor montado com simulador de curtos-circuitos estáticos.

5.2 Fonte AC Programável

Para a alimentação do motor, foi utilizada uma fonte AC programável. A Figura 31 apresenta a sua descrição em detalhe.



Figura 31 - Fonte AC Programável.

Fabricante / Modelo: Chroma / Model 61705 (12000 VA)

Descrição: Output 12 kVA, 3 ϕ (61705); Tensão: 0-150 V/0-300 V; Frequência: 15~1.2 kHz; Ângulo de Fase: 0~360° Programável; Capacidade de medir: V, Irms, Ipk, Iinrush, P, PF, CF, etc.

5.3 Simulação da Carga

Para a simulação da carga, foi utilizado um dinamómetro (Figura 32.a) e o seu respetivo sistema de controlo (Figura 32.b).



(a)



(b)

Figura 32 - (a) Dinamômetro;(b) Sistema de controlo.

Dinamómetro:

Fabricante / Modelo: Magtroll / Model HD-815-8NA (Hysteresis Dynamometer)

Descrição: Binário desde 18 Nm a 56,5 Nm; Conta com 16 valores pré-programados; Testes desde sem carga até teste de motor parado.

Sistema de Controlo:

Fabricante / Modelo: DSP7000 High Speed Programmable Controller

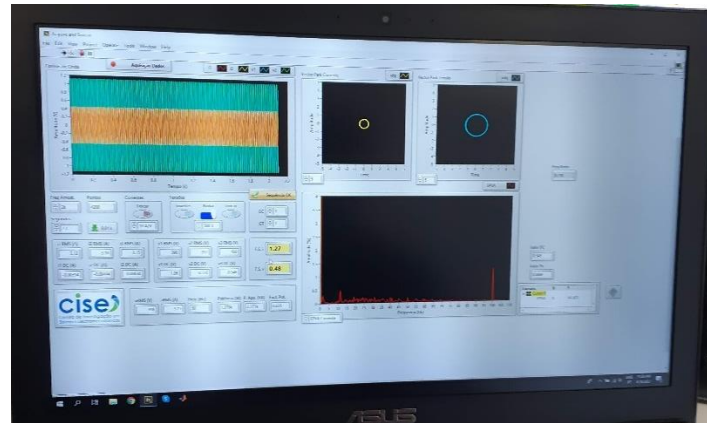
Descrição: DSP7001 Single Channel: Pode testar um único equipamento de malha de controlo fechado. Possui alarmes para potência, velocidade, binário, temperatura, fluxo de ar, fluxo de água, sobrecarga elétrica e entradas externas.

5.4 Sistema de Aquisição dos Dados

Para a aquisição dos dados, foi utilizado um sistema de aquisição e de processamento dos dados (Figura 33).



(a)



(b)

Figura 33 - (a) Sistema de aquisição de dados de corrente; (b) Software de processamento de dados e diagnóstico de avarias.

Sistema de Aquisição de Dados:

Fabricante / Modelo: Tektronix TCPA300

Descrição: Sensor amplificador de corrente, DC to 100 MHz; Escala e unidades automáticas.

Sistema de Processamento de Dados e Diagnóstico de Avarias:

Fabricante / Modelo: CISE

Descrição: O CISE utiliza um sistema proprietário de diagnóstico de avarias com base em PVA e EPVA.

5.5 Descrição dos Dados Adquiridos

A bancada de testes foi configurada com todos os equipamentos supracitados, de forma a assegurar a correta simulação das avarias no motor. A Figura 34 ilustra a bancada de testes utilizada.



Figura 34 - Bancada de simulação das avarias de curtos-circuitos entre espiras no motor e desequilíbrios de fase e amplitude na rede de alimentação.

Com os equipamentos anteriormente descritos, os dados foram então amostrados seguindo as características resumidas na Tabela 4.

Tabela 4 - Características do sinal amostrado.

Fase a e b		Tempo	Comprimento
Corrente	@ 2 kHz	1,024 s	2.048 pontos
Tensão	@ 2 kHz	1,024 s	2.048 pontos

As características de simulação das avarias estão resumidas na Tabela 5.

Tabela 5 - Características do sinal amostrado.

Desfasamento	0, 1, 2 e 3 graus
Amplitude	230, 228, 226 e 224 V*
Curtos-circuitos	00, 03, 06, 09 e 12 espiras

*As tensões referentes ao desequilíbrio de amplitudes foram consideradas na bobina do motor.

As condições de carga, estão resumidas na Tabela 6.

Tabela 6 - Características do sinal amostrado.

Carga	Percentual Nominal
7 Nm	51 %
13,6 Nm	100 %

A análise combinatória das simulações de desfasamento, amplitude e curtos-circuitos deveria gerar um banco de dados com $4 \times 4 \times 5 = 80$ possibilidades. Se forem considerados os dois níveis de carga, o *dataset* final deveria contar com 160 arquivos.

Entretanto, o *dataset* possui algumas características que, resumidamente, são:

- Não há desequilíbrio de amplitude nas tensões de alimentação de um motor em condições normais, ou seja, sem curtos-circuitos.
- Não há desequilíbrio de fase nas tensões de alimentação de um motor em condições normais, ou seja, sem curtos-circuitos.
- Os desequilíbrios de fase e de amplitude que ocorrem de maneira conjunta, não foram considerados nos testes.
- As classes possíveis são: 00 (saudável), 03, 06, 09 e 12 espiras. Para além disso, foram também recriadas condições de avarias nas tensões de alimentação, tanto em amplitudes, 224 V, 226 V, 228 V e 230 V (saudável), quanto em fase, 0 (saudável), 1, 2 e 3 graus.

Capítulo 6

6 Aplicação do Machine Learning

Este capítulo visa demonstrar a aplicação de algoritmos de inteligência artificial associados ao módulo do Vetor de Park para a classificação de avarias em curtos-circuitos.

A primeira fase tem como objetivo criar um modelo baseado em SVM para a detecção e classificação automática de curtos-circuitos estatóricos em motores de indução trifásicos, bem como determinar a sua severidade, que varia de 00 (bom estado), 03, 06, 09 a 12 espiras em curto-circuito, porém, em condições de alimentação equilibrada (normais). Para além disso, há também a determinação do compromisso entre o tempo mínimo de amostragem (*windowing*) e a precisão do diagnóstico, bem como a determinação das melhores *features* tanto no domínio do tempo, quanto no domínio da frequência.

A segunda fase tem como objetivo criar um modelo baseado em SVM, para a detecção e classificação automática de curtos-circuitos estatóricos em motores de indução trifásicos, bem como determinar a sua severidade, que varia de 00 (bom estado), 03, 06, 09 a 12 espiras em curto-circuito. Entretanto, desta vez, conta também com desequilíbrios de fase nas tensões de alimentação do motor, e que variam de 0º, 1º, 2º e 3º. Utiliza-se o EPVA nas correntes e nas tensões de alimentação.

A terceira fase tem como objetivo criar um modelo baseado em SVM, para a detecção e classificação automática de curtos-circuitos estatóricos em motores de indução trifásicos, bem como determinar a sua severidade, que varia de 00 (bom estado), 03, 06, 09 a 12 espiras em curto-circuito. Entretanto, desta vez, conta também com desequilíbrios relativos às tensões de alimentação do motor, e que variam em amplitudes (subtensão) de 224 V, 226 V, 228 V e 230 V. Utiliza-se o EPVA nas correntes e nas tensões de alimentação.

A quarta fase tem como objetivo criar um modelo baseado em SVM, para a detecção e classificação automática de desequilíbrios nas tensões de alimentação do motor de indução trifásico. Logo, os quatro modelos criados podem funcionar de forma automatizada, redirecionado o diagnóstico de avaria para seu respetivo modelo otimizado naquele cenário. No caso de um motor em condições normais de alimentação, o modelo treinado em condições normais gera o diagnóstico. No caso de um motor com desequilíbrios relativos à amplitude das tensões de alimentação, é diagnosticado pelo modelo com treino em

condições de desequilíbrios de amplitudes. No caso de um motor cuja alimentação apresenta desequilíbrios de fase, tal condição é diagnosticada através do modelo com treino em condições de desequilíbrios de fase. Utiliza-se o EPVA e a média das tensões de alimentação.

Quinta fase: testes com dados novos, tem como objetivo validar os modelos, bem como identificar as ineficiências nos modelos gerados.

6.1 Primeira Fase

O *dataset* utilizado nesta fase conta com 10 arquivos, que contêm as diferentes condições de avaria do motor, que vão desde o motor em boas condições, 00 (bom estado), 03, 06, 09 a 12 espiras em curto-circuito. A Tabela 7 detalha estes arquivos:

Tabela 7 - Arquivos com os dados adquiridos de tensão e correntes do motor. Frequência de amostragem: 2 kHz. Tamanho: 2000 pontos. Total de arquivos: 10.

Nome do arquivo	Espiras em curto-circuito	Carga	Classificação
Healthy_I2.60_f2000	0	7 Nm	0
Healthy_I4.56_f2000_13.6Nm	0	13,6 Nm	0
sh03_I2.58_230_Ish_4.5_f2000	03	7 Nm	3
sh03_I2.58_230_Ish_4_f2000	03	7 Nm	3
sh06_I2.58_230_ph120_Ish_4.6_f2000	06	7 Nm	6
sh06_I4.58_230_ph120_Ish_4.6_f2000	06	13,6 Nm	6
sh09_I2.60_230_ph120_Ish_4.7_f2000	09	7 Nm	9
sh09_I4.60_230_ph120_Ish_4.7_f2000	09	13,6 Nm	9
sh12_I2.60_230_ph120_Ish_4.7_f2000	12	7 Nm	12
sh12_I4.62_230_ph120_Ish_4.7_f2000	12	13,6 Nm	12

A classificação está relacionada com a identificação das classes que se pretendem identificar com o algoritmo de *Machine Learning*. Neste caso, existem cinco classes apresentadas abaixo:

- Classe 0:** sem curto-circuito;
- Classe 3:** curto-circuito em 3 espiras;
- Classe 6:** curto-circuito em 6 espiras;
- Classe 9:** curto-circuito em 9 espiras;
- Classe 12:** curto-circuito em 12 espiras.

Deste modo, é necessário processar todos os arquivos, gerando um único *dataset* e identificando-o de acordo com a sua classificação antes de ser processado pelo algoritmo SVM.

6.1.1 Dados do módulo Vetor de Park das correntes estatóricas

Uma avaliação eficiente do equilíbrio das correntes pode ser feito através da aplicação do Vetor de Park que terá como resultado as componentes diretas (d), em quadratura (q) e zero, que neste caso, representam as três fases da corrente do motor. As investigações apontam que o módulo Vetor de Park sofre alterações devido a desequilíbrios no motor. É, portanto, uma excelente ferramenta para abordar juntamente com o algoritmo de *Machine Learning*.

Em ensaios teóricos, utilizando um modelo, tornam-se evidentes os distúrbios causados no módulo do Vetor de Park devido aos curtos-circuitos estatóricos. A Figura 35 apresenta a simulação de um curto-circuito entre 3 espiras, onde se pode verificar o comportamento do módulo do Vetor de Park, as componentes d, q e zero.

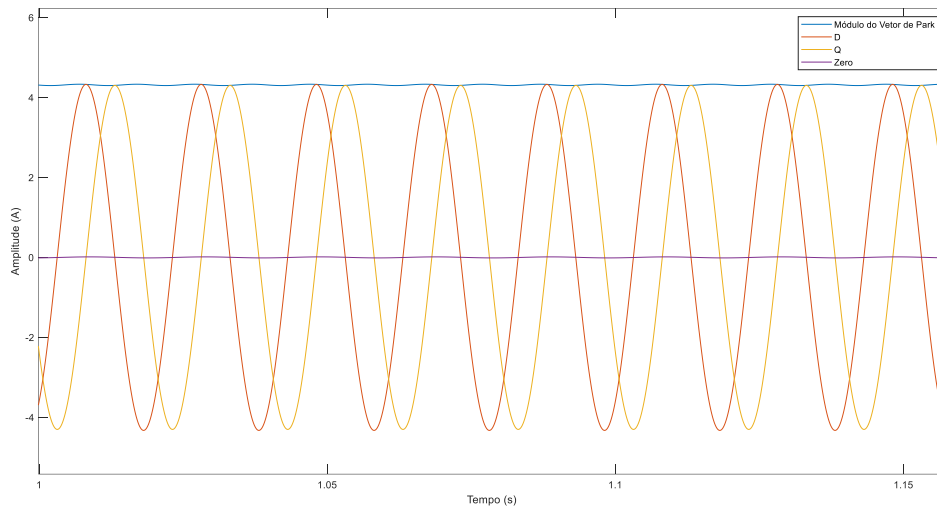


Figura 35 – Perturbações no módulo do Vetor de Park, causadas por um curto-circuito entre 3 espiras.

A Figura 36 apresenta em maior detalhe as perturbações causadas pelo curto-circuito entre 3 espiras.

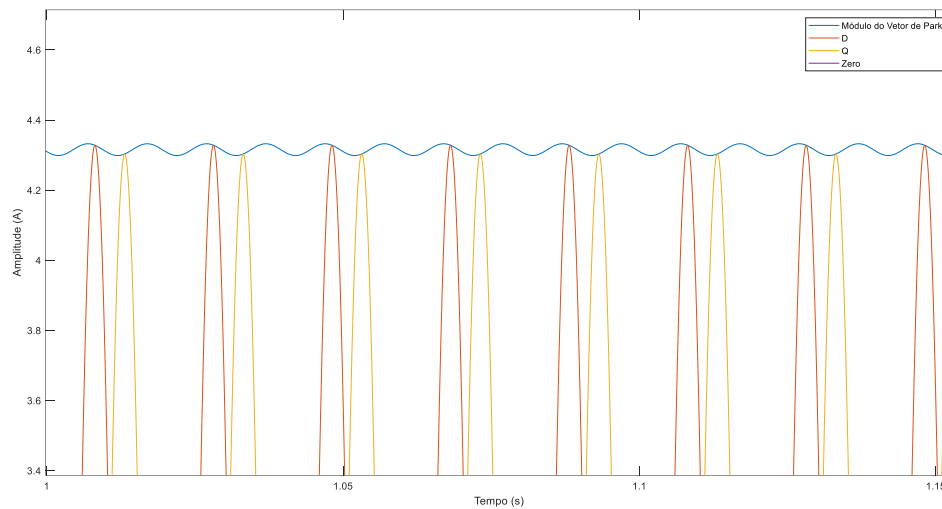


Figura 36 - Ampliação da Figura 35 onde é possível verificar com maior clareza as perturbações causadas no módulo do Vetor de Park devido ao curto-circuito entre 3 espiras.

A Figura 35 e a Figura 36 foram simuladas a partir de um modelo e servem para ilustrar como o módulo do Vetor de Park é afetado por avarias de curtos-circuitos. Note-se que as componentes d e q não estão sincronizadas com a velocidade de rotação e, deste modo, apresentam oscilações equivalentes ao dobro da frequência fundamental. Estas figuras também demonstram como o desequilíbrio nas componentes d e q são refletidos no módulo do Vetor de Park. Entretanto, os dados reais adquiridos a 2 kHz podem apresentar distorções causadas por baixa taxa de aquisição, ou ainda, devido à ruídos.

Para a construção do *dataset*, as correntes a e b foram adquiridas, sendo a corrente c calculada ($i_c = -i_a - i_b$). Em cada período de 0,02 s foram aplicados os cálculos para encontrar d e q, que posteriormente foram utilizados para calcular o módulo do Vetor de Park. Os períodos foram sobrepostos para facilitar o entendimento do gráfico. A última coluna de dados contém a sua classificação, que vai de 00 a 12 espiras em curto-circuito. Os resultados de sua aplicação são apresentados na Figura 37.

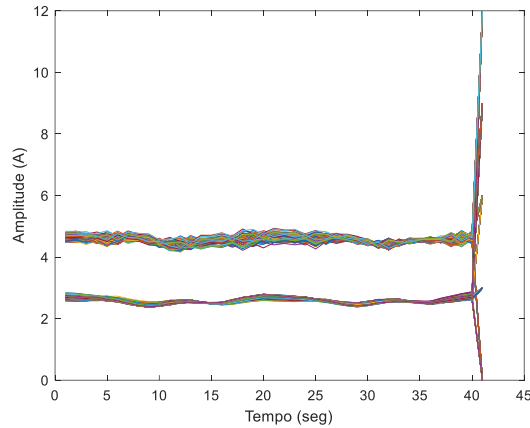


Figura 37 - *Dataset* que contém o Vetor de Park em duas condições de carga, 7 Nm e 13,6 Nm, em diferentes condições de severidade de curto-circuito 00, 03, 06, 09, 12 espiras.

Na próxima seção será discutida a extração das *features* do módulo do Vetor de Park, de modo a classificar uma avaria de curtos-circuitos estatóricas, bem como a sua classificação quanto à severidade. Em seguida, as *features* selecionadas serão aplicadas ao SVM.

6.1.2 *Feature Extraction* e *Windowing* do módulo do Vetor de Park

Este capítulo tem como objetivo extrair as características mais relevantes das correntes do motor de indução trifásico, através do módulo do Vetor de Park para determinar a existência ou não de curtos-circuitos estatóricas, de forma a criar um modelo automático de classificação. Para isso, serão extraídas características no tempo, na frequência, e será realizada uma análise do sinal para determinar o melhor *windowing* onde serão aplicados algoritmos de *Machine Learning* para testar sua eficiência.

6.1.2.1 *Features* no Tempo

Quando uma avaria estatórica ocorre, gera perturbações que podem ser observadas nas correntes medidas num motor. As amplitudes e distribuições podem ser alteradas no domínio do tempo. Portanto, os recursos estatísticos no domínio do tempo podem ser extraídos de suas formas para refletir tal condição [26]. Os principais indicadores serão implementados para a criação de um *dataset* no módulo do Vetor de Park das correntes do motor. A Tabela 8 ilustra as *features* e as classes.

Tabela 8 - Dataset gerado após o processo de *feature extraction*, considerando 40 pontos por linha.

Features do módulo do Vetor de Park	sh00_07	sh12_07
1 - Assimetria (<i>skwness</i>)	-0,3910	0,2747
2 - Curtose	2,0107	1,9762
3 - Mediana	0,0153	-0,0382
4 - Média	0,0000	0,0000
5 - Variância	0,0039	0,0218
6 - Desvio Padrão	0,0628	0,1478
7 - Raiz da soma dos quadrados (rssq)	0,3925	0,9228
8 - Raiz quadrada média (rms)	0,0621	0,1459
9 - Pico máximo absoluto	0,1241	0,2923
10 - rms / Pico máximo absoluto	0,5001	0,4992
11 - Pico máximo absoluto / rssq	0,3161	0,3167
12 -rms / Mediana	4,0540	-3,8199
13 - (rssq) / Mediana	8,1056	-7,6517
Classe (00, 03, 06, 09 e 12 espiras em curto-circuito)	00	12

É importante que o *dataset* contenha as informações pertinentes ao caso de estudo. Além disso, deve estar organizado de forma que seja compreensível a sua interpretação tanto pelos algoritmos, quanto para a sua interpretação. Note-se que antes da aplicação dos dados estatísticos, a componente DC foi removida do sinal.

Todos os índices acima descritos visam descrever com precisão o comportamento da avaria, revelada pelas perturbações no módulo do Vector de Park nas correntes. Entretanto, nem todos são significativos e foram, portanto, removidos aqueles que contribuem negativamente ou muito pouco para a melhoria na taxa de precisão.

Para realizar esta operação, todas as 13 *features* foram classificadas de acordo com a sua importância em relação à sua classificação. Para este critério, foi primeiramente utilizada a correlação, conforme demonstrado na Figura 38.

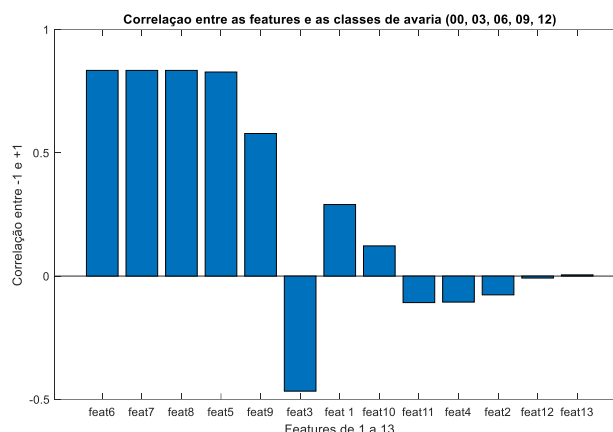


Figura 38 - Correlação entre as 13 *features* extraídas e a classe de avarias, disposta em ordem decrescente.

Na Figura 38, foi realizada a correlação entre as *features* e a classe, os seus resultados foram apresentados de forma decrescente de forma que os valores mais elevados sejam aqueles que têm a maior associação ou correlação com as classes. Deve notar-se que os valores de correlação variam de -1 a +1, onde os valores positivos estão correlacionados de maneira proporcional e os valores negativos são inversamente proporcionais. Assim, as *features* 6, 7, 8, 18 e 5, possuem a maior correlação, acima de 0,8. São seguidas das *features*, 14, 17, 15, 9, 3, 16 e 1, que ficam no intervalo intermediário, entre 0,2 e 0,8. As *features* com menor correlação estão abaixo de 0,2, sendo elas 10, 11, 4, 2, 12 e 13. Através da análise de correlação, é possível concluir que não faz sentido utilizar *features* que não estão associadas às avarias.

A Tabela 9 resume os três grupos de correlação entre as *features* e as classes.

Tabela 9 - Correlação entre as *features* e as classes.

Intervalo de Correlação	<i>Features</i>	Análise qualitativa
Correlação > 0,80	6, 7, 8 e 5	Elevada correlação
0,20 > Correlação > 0,80	9, 3 e 1	Média correlação
Correlação < 0,20	10, 11, 4, 2, 12 e 13	Reduzida ou nenhuma correlação

Após a análise de correlação nas *features* no módulo do Vetor de Park das correntes, é possível notar que existem três classes de correlação, a elevada, a média e a reduzida. Para definir o critério para esta escolha, foi utilizado o p-val, de forma a testar a probabilidade de não correlação, conforme a Figura 39.

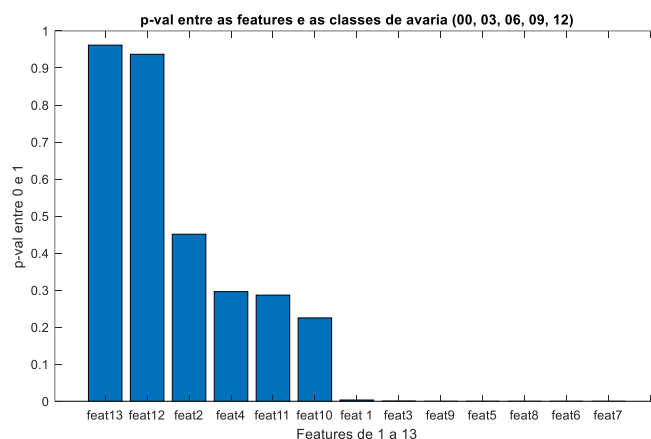


Figura 39 - p-val entre as *features* e as classes de avaria (00, 03, 06, 09, 12).

O valor p, por norma, utiliza o critério de corte de 0,05, ou seja, 95% de certeza de que as *features* não estão correlacionadas com as classes. Portanto, as *features* 13, 12, 2, 4, 11 e 10 são descartadas a partir deste critério, conforme ilustrado na Figura 39. Logo, as *features* que possuem correlação e que podem ser aplicadas em algoritmos de *Machine Learning* são 6, 7, 8, 5, 9, 3 e 1 (em ordem decrescente de correlação), e estão de acordo com a correlação elevada e média da Tabela 10, que resume os resultados do valor p.

Tabela 10 - Correlação entre as *features* e as classes.

Critério pval	<i>Features</i>	Análise qualitativa
Maior que 0,05	13, 12, 2, 4, 11 e 10	Descartam-se
Menor que 0,05	6, 7, 8, 5, 9, 3 e 1	Podem ser utilizadas

Como resultados de ambos os testes, nota-se que existe uma correlação entre as *features* 1, 3, 9, 5, 8, 6 e 7 e a classe. Entretanto, algumas *features* podem carregar os modelos de *Machine Learning* com informações redundantes, desta maneira, será realizado mais uma seleção, desta vez, levando-se em conta o critério de redundância, onde as *features* serão confrontadas entre elas mesmas, com respeito à sua correlação, conforme o mapa de calor exibido na Figura 40.

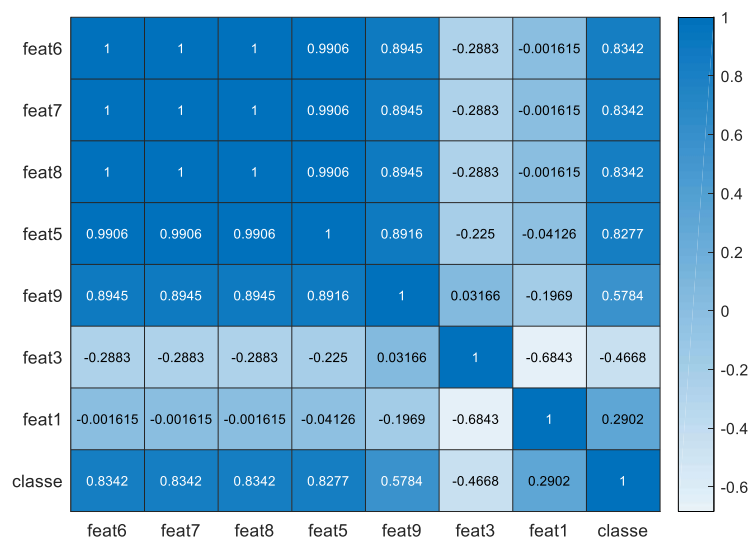


Figura 40 - Mapa de calor de correlação *features x features x classe*.

A análise da Figura 40 revela uma correlação entre as *features x features x classe*. Deste modo, revela as informações que são redundantes. Por exemplo, no caso da *feature 7 e 8*, ambas possuem elevada correlação entre elas e a classe, com valor de 0,8342. Entretanto, as *feature 6 x feature 7* possuem um valor de correlação de 1, sendo este o maior valor possível de ser alcançado, o que revela que as duas *features* carregam basicamente a mesma informação e que uma pode substituir a outra. Caso sejam utilizadas as duas *features*, corre-se o risco de levar o algoritmo de *Machine Learning* a uma sobrecarga de dados (*overfitting*) e, em consequência, trazer resultados errados na classificação. Após a análise do mapa de calor, as *features* selecionadas foram listadas na Tabela 11.

Tabela 11 - *Features* selecionadas com menor redundância e maior relevância.

Classes de Redundância	<i>Features</i>	<i>Features</i> Selecionadas
Critério de Corte > 0,85	6, 7, 8, 5, 9	6
Menor que 0,85	3 e 1	3 e 1
	Total Selecionado	6, 3, 1

Deste modo, as *features* que possuem a máxima relevância e a mínima redundância, a partir dos critérios anteriormente expostos, foram selecionadas. São, nomeadamente, as *features* 6, 3, 16 e 1.

Com as *features* selecionadas, é possível o seu treino, e tendo em conta as suas características, foi escolhido o algoritmo SVM ECOC do MATLAB, pois pode fazer uma

classificação com mais de duas classes. Outro ponto importante do treino, é a escolha dos hiper parâmetros, que será realizado de forma a encontrar o mínimo local em função do erro gerado pelo modelo, conforme ilustrado na Figura 41.

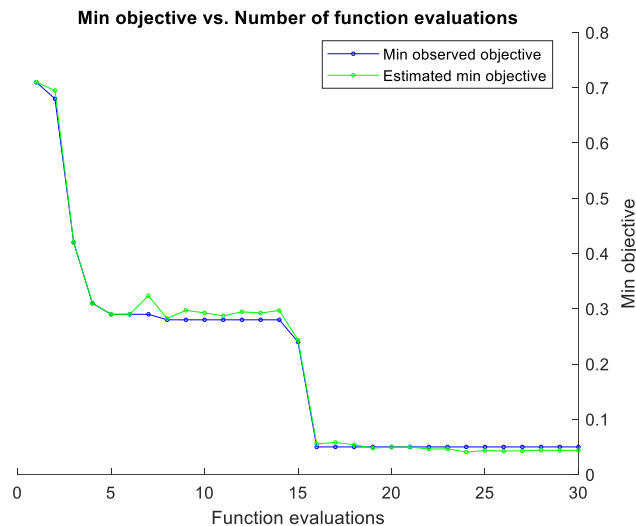


Figura 41 - Iterações durante a otimização de parâmetros do modelo SVM ECOC Linear, realizado no Matlab.

Após a realização da análise das *features* extraídas no módulo do Vetor de Park, foi realizada uma comparação da sua eficiência quando aplicadas ao SVM. O seu indicador de precisão foi baseado na validação cruzada de 50 vezes. Os seus resultados são apresentados na Tabela 12.

Tabela 12 - Comparação entre as diferentes *features*, para a redução dimensional.

<i>Features</i>	Precisão (PCA OFF)	Precisão (PCA ON: 4/18)
Todos os 18 <i>features</i> supracitadas	82,0%	86,0%
<i>Features</i> : 6, 3, 16 e 1	89,0%	83,0%

Os dados da Tabela 12 demonstram uma precisão de 89,0%, com apenas 3 *features* num intervalo de 0,02 s, o que indica que o módulo Vetor de Park é um excelente indicador.

Os ruídos são refletidos nas características estatísticas que são a base do treino do algoritmo de *Machine Learning*. Na Figura 42, pode notar-se a média do desvio padrão, que tem como função principal servir de indicador da amplitude do módulo do Vetor de Park.

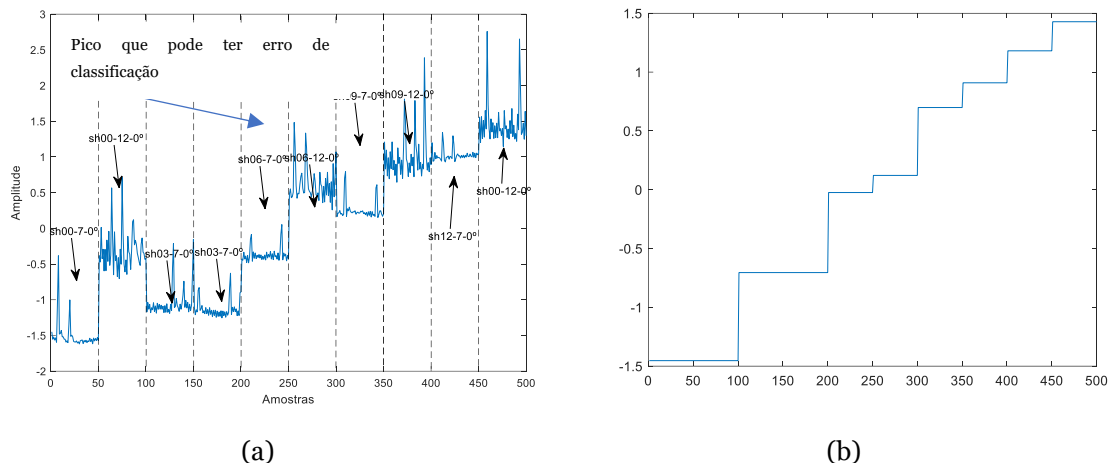


Figura 42 - Aplicação do desvio padrão: (a) Desvio padrão normalizado do módulo do Vetor de Park, em dados reais; (b) Desvio padrão do módulo do Vetor de Park, gerado através de modelo de simulação no Simulink Matlab. Janela de 40 pontos (0,02 s).

Cada ponto no gráfico reflete um período de 0,02 s (40 pontos). Note-se que na Figura 42(a) existem picos que podem ser classificados erradamente, o que diminui a precisão e, para além disso, há influência da carga. Entretanto, a Figura 42(b), obtida através de simulação revela pouca ou nenhuma influência da carga sobre a amplitude do módulo Vetor de Park. A Tabela 13 descreve em maiores detalhes o conteúdo da Figura 42.

Tabela 13 - Detalhes das amostras da Figura 42.

Amostras (eixo x de 1 a 500)	Condição do Motor
0 a 50	Sh00_7
51 a 100	Sh00_12
101 a 150	Sh03_7
151 a 200	Sh03_7
201 a 250	Sh06_7
251 a 300	Sh06_12
301 a 350	Sh09_7
351 a 400	Sh09_12
401 a 450	Sh12_7
451 a 500	Sh12_12

Para mitigar os problemas provenientes de distorções no sinal do módulo do Vetor de Park e melhorar a precisão, será realizada, na próxima secção, uma análise espectral com diferentes *windowings*. Deve lembrar-se que os dados testados não incluem problemas da rede de alimentação, refletidos na corrente.

6.1.2.2 Features no Domínio da Frequência

A extração de dados no domínio da frequência acaba por revelar ainda mais informações pertinentes para o problema proposto. Muitas abordagens diferentes foram realizadas neste contexto. Podem destacar-se os trabalhos realizados na Universidade de Paderborn, na Alemanha, que disponibilizou os seus dados, tendo sido publicado um artigo contendo informação acerca dos métodos utilizados para que fosse referência na área de deteção de avarias em rolamentos de motores [34] e, onde foram utilizadas frequências de amostragem superiores a 40 kHz. A revisão realizada por MILJKOVIĆ [5], discute também a utilização de inteligência artificial no âmbito da deteção de diversos tipos de avarias relacionadas com o motor de indução trifásico. Uma informação muito útil neste documento diz respeito às frequências de interesse ao nível do MCSA, que vão desde 0 a 5 kHz, e ainda à frequência mínima de aquisição de sinais que é de 10 kHz.

Estes artigos revelam a importância de uma elevada frequência para um diagnóstico preciso da avaria. Outros pontos que devem ser considerados são o compromisso entre o tamanho da amostra, a taxa de amostragem, o poder de processamento e o tempo necessário para o diagnóstico.

O desafio para a operação no domínio da frequência é revelar os distúrbios causados pela avaria, no menor intervalo de tempo possível, com menor processamento e com elevada precisão. O trabalho no domínio do tempo, no tópico anterior, revela que existem picos da amplitude média do módulo do Vetor de Park para um período (0,02 s). Uma estratégia é aumentar a janela, em vez de utilizar um período, utilizar mais pontos, desta maneira os picos serão atenuados. Quanto maior a janela, maior será a atenuação de ruídos. O ensaio inicial reduziu a janela, ou seja, o tempo de observação do módulo do Vetor de Park para 0,005 s ou 10 pontos. A esta fração do sinal, foi aplicada a FFT, o seu resultado foi normalizado, e a sua magnitude espectral na componente de 100 Hz foi extraída. Sendo realizado este processo para cada um dos arquivos de dados, os mesmos foram justapostos para gerar a Figura 43. Com uma amostra tão pequena, a magnitude espectral ao dobro da frequência fundamental (100 Hz) apresenta elevada imprecisão, o que prejudica o diagnóstico correto e a classificação das diferentes severidades. Para além disso, as diferenças entre os picos máximos e mínimos são tão grandes, que é impossível a sua correta classificação. Esta interpretação pode ser feita através da Figura 43.

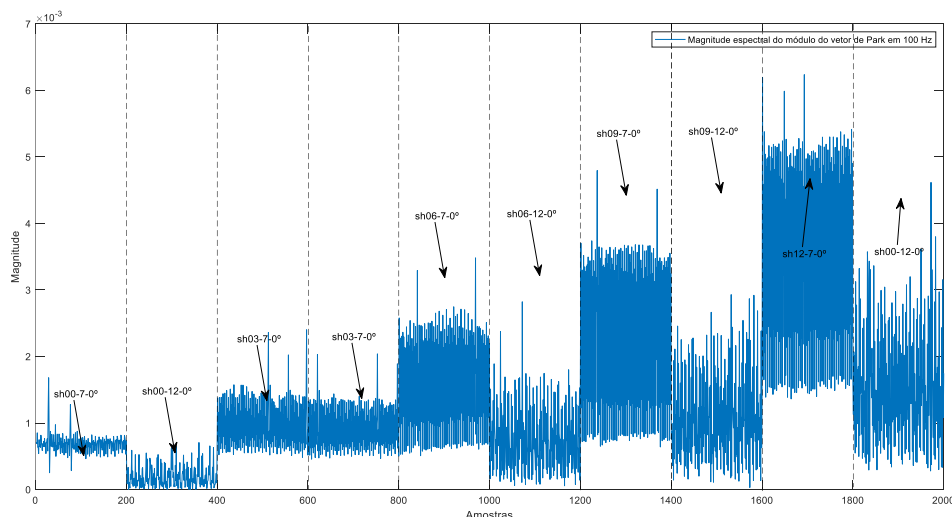


Figura 43 - Justaposição da magnitude espectral ao dobro da frequência fundamental (100 Hz) do módulo do Vetor de Park para uma janela de 0,005 s (10 pontos) aplicadas aos arquivos de dados.

Uma segunda análise foi realizada com uma janela correspondente a um período, ou seja, 0,02 s (40 pontos). Os resultados estão representados na Figura 44, onde é possível notar que a influência da carga foi reduzida. Entretanto, existe ainda uma grande diferença entre os picos máximos e mínimos, especialmente em condições com cargas mais elevadas, o que leva a uma possível condição de classificação errônea.

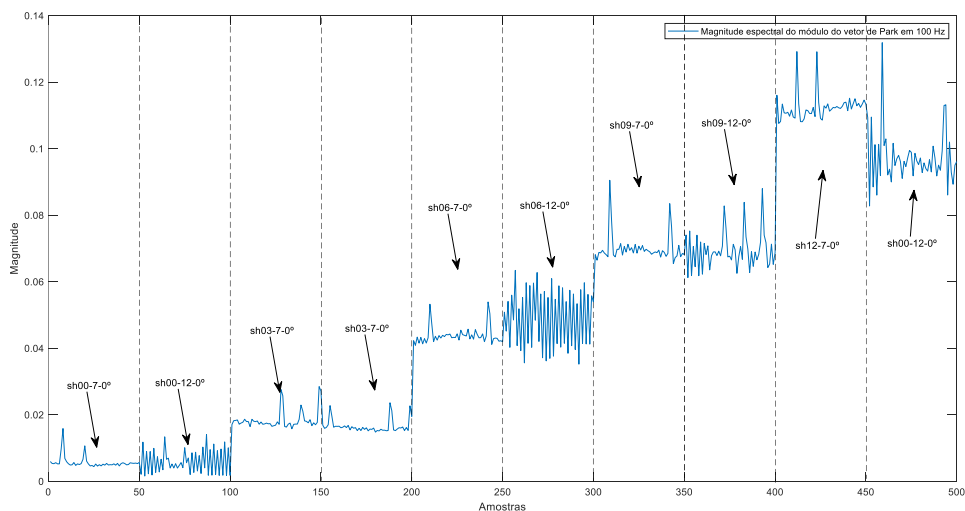


Figura 44 - Justaposição da magnitude espectral ao dobro da frequência fundamental (100 Hz) do módulo do Vetor de Park para uma janela de 0,02 s (40 pontos) aplicadas aos arquivos de dados.

Para melhorar ainda mais estes dados, a próxima janela escolhida foi de duas vezes o período, desta forma, uma janela de 0,04 s (80 pontos), conforme pode ser observado na Figura 45.

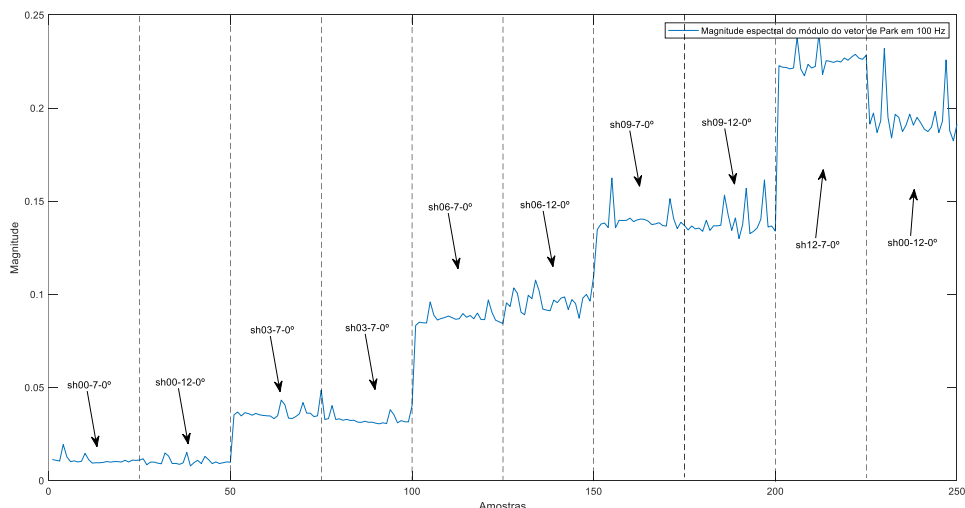


Figura 45 - Justaposição da magnitude espectral ao dobro da frequência fundamental (100 Hz) do módulo do Vetor de Park para uma janela de 0,04 s (80 pontos) aplicadas aos arquivos de dados.

A partir do teorema de Nyquist e para obtermos dados concretos sobre os 100 Hz, é necessária uma amostragem de pelo menos 200 pontos. Desta forma, foi inserido no Matlab um único *feature* de dados, que conta com 200 pontos de amostragem, ou seja, 0,1 segundo. Os seus resultados podem ser observados na Figura 46.

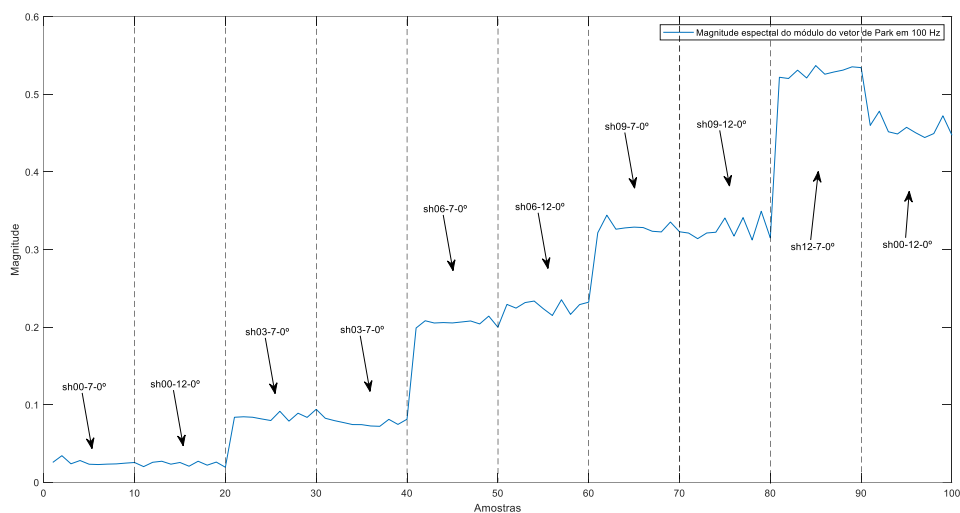


Figura 46 - Justaposição da magnitude espectral ao dobro da frequência fundamental (100 Hz) do módulo do Vetor de Park para uma janela de 0,1 s (200 pontos) aplicadas aos arquivos de dados.

A análise da Figura 46 revela uma separação inequívoca dos diversos graus de severidade para a avaria de curto-circuito estatórico, não havendo picos que se possam confundir com

outras classes e não sofrendo influência da carga para o caso de 00, 03, 06 e 09 espiras em curto-circuito. A única exceção é para o caso de 12 espiras que não se apresenta na mesma amplitude média em diferentes cargas.

Contudo, não há uma dificuldade de classificação, em nenhum dos casos, devido a picos ou ruídos. Deste modo e de maneira antecipada, esta *feature* carrega informações pertinentes e muito precisas relativas ao curto-circuito estatórico.

Para comprovar esta hipótese, foi realizado uma comparação entre a correlação da magnitude espectral do módulo do Vetor de Park ao dobro da frequência fundamental e outras *features* extraídas do módulo do Vetor de Park anteriormente, no domínio do tempo.

Tabela 14 - Features do módulo do Vetor de Park.

Features
1 - Assimetria (<i>skwness</i>)
3 - Mediana
6 - Desvio Padrão
19 - Magnitude espectral do módulo do Vetor de Park em 2fs
Classe (00, 03, 06, 09 e 12 espiras em curto-circuito)

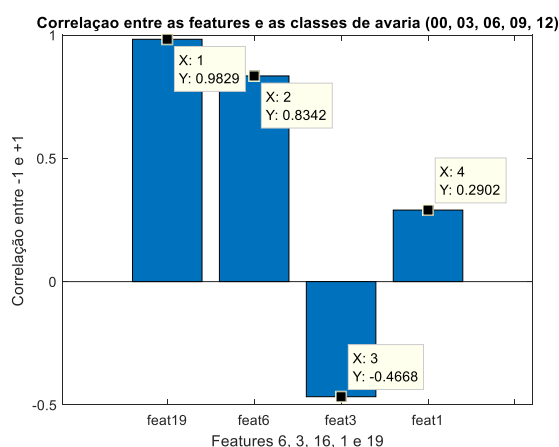


Figura 47 - Comparação da correlação entre o feat19 e as feat6, feat3 e feat1 x classes.

A Figura 47 revela a forte correlação existente entre as *features* (Tabela 14) e as classes de avaria. A magnitude espectral do módulo do Vetor de Park ao dobro da frequência fundamental (feat19) destaca-se, pois alcança o valor de 0,9829, seguido pelo desvio padrão do módulo do Vetor de Park, com 0,8342 (feat6), pela mediana (feat3), com -0,4668, e, em último, a assimetria, com 0,2902 (feat1). Note-se que a correlação varia de -1 a +1 e que a feat19 alcança um valor de correlação muito próximo de 1.

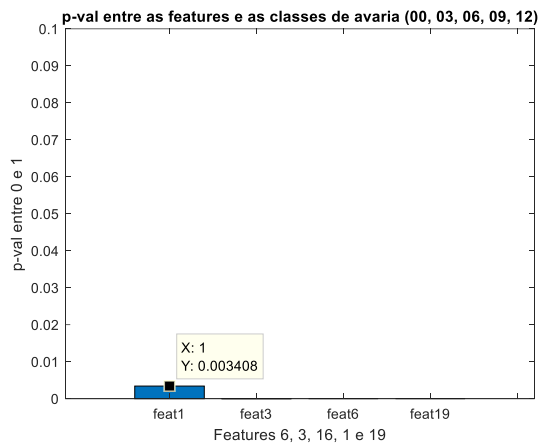


Figura 48 - Valor p das *features* x classe.

Na Figura 48, referente à análise do valor p, nota-se que todas as *features*, 19, 6, 3 e 1, passam no teste da hipótese nula, com 95% de segurança. Entretanto, como a *feat19* apresenta uma elevada correlação, foi decidido criar um novo modelo, baseado apenas nesta *feature*.

Assim, foi criado um novo *dataset*, que substitui o anterior e que conta com apenas uma *feature*, nomeadamente o EPVA aplicado às correntes (*feat19*), segundo [14]. A Tabela 15, descreve em detalhe este novo *dataset*:

Tabela 15 - *Dataset* com a magnitude espectral ao dobro da frequência fundamental (100 Hz).

Nomes	Descrição
Feat19	Magnitude espectral do módulo do Vetor de Park em 100 Hz, com janela de 0,1 segundo (200 pontos), baseada em EPVA.
Classe	(00, 03, 06, 09 e 12 espiras em curto-circuito)

Deve ser observado que ao aumentar a janela para 200, o número de amostras que serão inseridas no algoritmo de *Machine Learning* é reduzido proporcionalmente. Neste caso, temos apenas 10 pontos de amostra por arquivo de dados. Também, se deve notar que este valor de 10 amostras está no limite, pois a sua redução induz os algoritmos a erros por falta de amostras para o treino, ainda que os ruídos, característicos devido a amostras muito curtas, sejam atenuados.

Os dados foram inseridos no Matlab, e apresentaram os resultados revelados pela Figura 49.

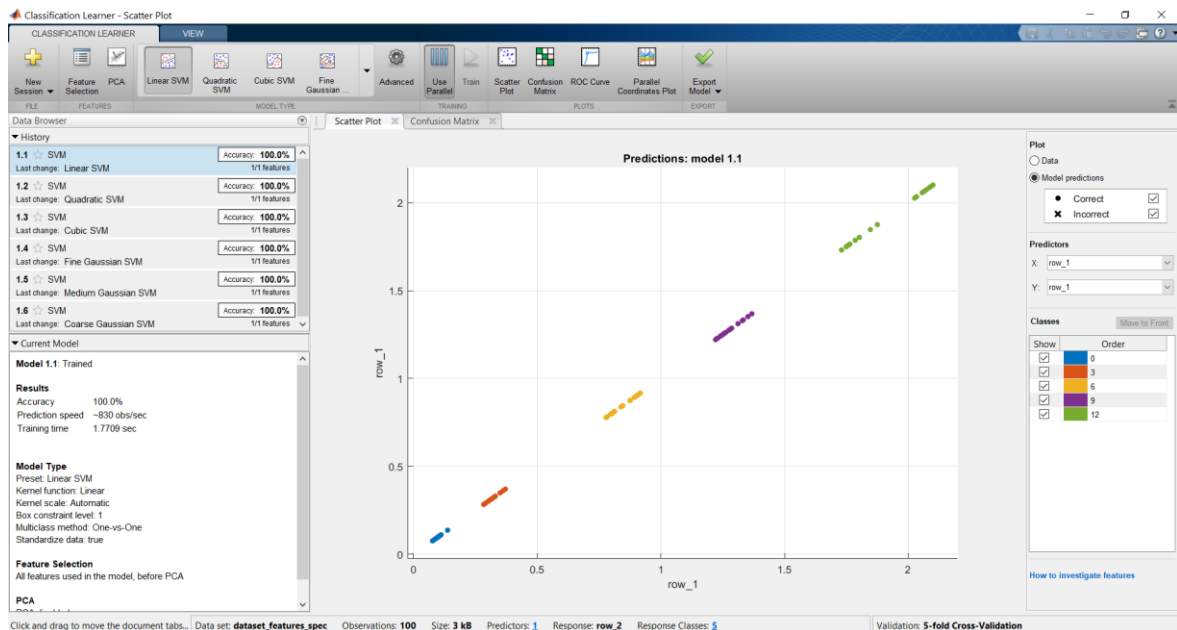


Figura 49 - Comparação de diferentes algoritmos de *Machine Learning* baseado em EPVA para uma janela de 0,1 s, em condições normais de alimentação (*ToolBox Classification Learner* do Matlab).

Na Figura 49, pode notar-se que do lado esquerdo se apresenta a precisão de todos os algoritmos de SVM disponíveis na biblioteca do Matlab, na sua configuração padrão, e que todos apresentam 100% de precisão. É possível notar também, ao centro, que as diferentes classes se apresentam de maneira concentrada entre elas e com distâncias bem definidas umas das outras. A exceção é a classe de 12 espiras em curto-circuito, representada pela cor verde, que apresenta uma maior dispersão. Esse comportamento já era previsível, devido à análise prévia deste *feature*.

Deve-se ainda salientar que, com esta abordagem, é possível o treino de algoritmos de *Machine Learning*, nomeadamente o SVM, comumente abordado na literatura, de forma a alcançar uma classificação ou um diagnóstico de curto-circuito entre espiras, com 100% de probabilidade de acerto, dentro dos dados que foram cedidos pelo CISE, tendo em vista o motor em questão. É preciso destacar também, que este diagnóstico pode ser obtido já a partir de 0,04 s, e apresenta uma boa segurança a partir de 0,1 s, com uma taxa de aquisição de 2 kHz, conforme pode ser destacado na Tabela 16.

Tabela 16 - Relação do tempo de amostragem (janela) com as taxas de acerto e o número de treino por classe de avaria.

Tempo de amostragem (janela)	Taxa de Acerto PCA OFF	Taxa de acerto PCA ON	Treino / Classe
0,005 s (10 pontos)	* 41,0 %	41,1 %	200 x 2 = 400 / classe
0,01 s (20 pontos)	* 83,1 %	* 83,3 %	100 x 2 = 200 / classe

0,02 s (40 pontos)	* 98,2 %	97,8 %	50 x 2 = 100 / classe
0,03 s (60 pontos)	* 98,5 %	* 98,5 %	33 x 2 = 66 / classe
0,04 s (80 pontos)	** 100 %	** 100 %	25 x 2 = 50 / classe
0,05 s (100 pontos)	* 97,0 %	97,0 %	20 x 2 = 40 / classe
0,06 s (120 pontos)	** 100 %	** 100 %	16 x 2 = 32 / classe
0,075 s (150 pontos)	** 100 %	** 100 %	13 x 2 = 26 / classe
0,08 s (160 pontos)	** 100%	* 100 %	12 x 2 = 24 / classe
0,0835 s (167 pontos)	* 100 %	* 100 %	11 x 2 = 22 / classe
0,0865 s (173 pontos)	* 100 %	* 100 %	11 x 2 = 22 / classe
0,1 s (200 pontos)	** 100%	* 100 %	10 x 2 = 20 / classe

* Cross-validation fold: 5, Preset: Linear SVM, Kernel: Gaussian, Kernel scale: Automatic, Box constraint level: 2, Multiclass method: One-vs-All, Standarize data: true.

** Todos os algoritmos de SVM, com condição automática do Matlab.

Nota-se também, de acordo com a Tabela 16, que o *windowing* ideal está diretamente relacionado com o período da frequência fundamental. Esse efeito pode ser notado entre a janela de 0,04 s (80 pontos) e de 0,05 s (100 pontos). Esta última, ainda que conte com mais pontos de amostragem, carrega distúrbios relacionados ao *windowing* fora dos múltiplos da frequência fundamental. Este efeito acaba por ser reduzido com o aumento da janela, e que, já a partir de 0,075 s (150 pontos), apresenta resultados com exatidão de * 100%. Para obter maior segurança e menor influência do efeito negativo de uma janela não sincronizada com a frequência fundamental, os dados devem ser considerados a partir de 0,1 s (200 pontos), obtendo-se, assim, uma precisão de ** 100% em todos os algoritmos de SVM, na condição automática da *Toolbox Machine Learning* do Matlab.

6.1.3 Resultados dos modelos: fase 1

Após o processo de extração de *features* do módulo do Vetor de Park aplicado às correntes estatóricas do motor de indução trifásico, são comprovados os diferentes estudos relativos ao módulo do Vetor de Park e à aplicação de SVM para a classificação de avarias em motores elétricos trifásicos, os quais apontam para uma maior eficiência da análise espectral na região do dobro da frequência fundamental em detrimento da análise no tempo, que fez o modelo de *Machine Learning* ir de 89% para 100% de acerto. Existe, também, uma janela mínima, ou seja, um tempo mínimo de duas vezes o período da frequência fundamental, para uma confiança de 100% de precisão na classificação das diferentes severidades de curtos-circuitos estatóricos. Um tempo maior reduz a influência da frequência de operação do motor. Deve ser observado que não é possível a distinção entre avarias na rede de

alimentação. Este aspeto será abordado no próximo tópico. Como resumo dos resultados, temos:

- Desempenho do modelo no domínio do tempo (*features* 1, 3 e 6): 89%;
- Desempenho do modelo no domínio da frequência (feat 19): 100%;
- Total de arquivos selecionados: 10;
- Deficiências: Faz a deteção apenas em condições normais de alimentação;
- Não é capaz de detetar curtos-circuitos equilibrados nas três fases;
- Janela mínima de 0,1 s (200 pontos) @ 2 kHz;
- A partir de 0,125 s (250 pontos) @ 2 kHz o desempenho dos algoritmos de *Machine Learning* começam a diminuir por falta de dados;
- Identificação em 5 classes: 00, 03, 06, 09, 12.

6.2 Segunda Fase

6.2.1 Modelo com distúrbios na alimentação: Fase

A abordagem no domínio da frequência tem vantagens latentes em relação ao domínio do tempo. Com apenas uma *feature* (feat19) que revela a magnitude espectral do módulo do Vetor de Park ao dobro da frequência fundamental [14], apresentam-se resultados de classificação muito satisfatórios, com uma janela adequada (0,1 segundo), com uma taxa de precisão de 100%, para diferentes algoritmos de *Machine Learning*. Portanto, nesta secção, o algoritmo criado será submetido não só às condições de avaria estatóricas, mas também será submetido a condições de avaria na rede de alimentação (desfasamento) para efetuar a sua classificação. Assim, o grau de dificuldade no diagnóstico é aumentado. Serão, portanto, incluídas novas classes (00, 03, 06, 09, 12, 31, 32, 33, 61, 62, 63, 91, 92, 93, 121, 122, 123), bem como novos dados adquiridos no CISE. Os novos dados foram precedidos de “...” para facilitar o seu reconhecimento. A sua descrição está detalhada no Apêndice A, Tabela 23.

Esta nova tabela reúne, portanto, arquivos de dados que contêm os dados anteriores relativos às diferentes severidades dos curtos-circuitos estatóricos, que variam de 00 (no caso de motor saudável), 03, 06, 09 a 12 espiras em curto-circuito. Para além disso, foram adicionadas informações pertinentes aos diferentes desequilíbrios de fase das tensões de alimentação, que variam de 0 (no caso de alimentação saudável), 1, 2 a 3 graus de desfasamento.

Desta forma, cria-se uma forma de categorização específica para os diferentes casos, onde a primeira casa decimal da classificação corresponde à quantidade de espiras em curto-circuito, e o segundo dígito diz respeito ao desfasamento em graus. Por exemplo, uma classe 033, diz respeito à classificação de um motor com 3 espiras em curto-circuito e com 3 graus de desfasamento na alimentação, 122 diz respeito a 12 espiras em curtos-circuitos e 2 graus de desfasamento, uma classe 09 diz respeito a 9 espiras em curto-circuito e 0 graus de desfasamento. Deste modo, a classificação proposta visa classificar os diferentes tipos de avaria de um modo sistemático e de fácil entendimento. Salienta-se que neste conjunto de dados, não há um motor sem avaria de curtos-circuitos e com avaria de desequilíbrio na amplitude das tensões de alimentação.

Com a adição destes novos dados, foi necessária a criação de um novo *dataset*, com novas classes. Deste modo, a componente espectral do módulo do Vetor de Park em 2fs foi extraída das correntes, para cada amostra como realizado anteriormente, gerando assim o gráfico ilustrado na Figura 50.

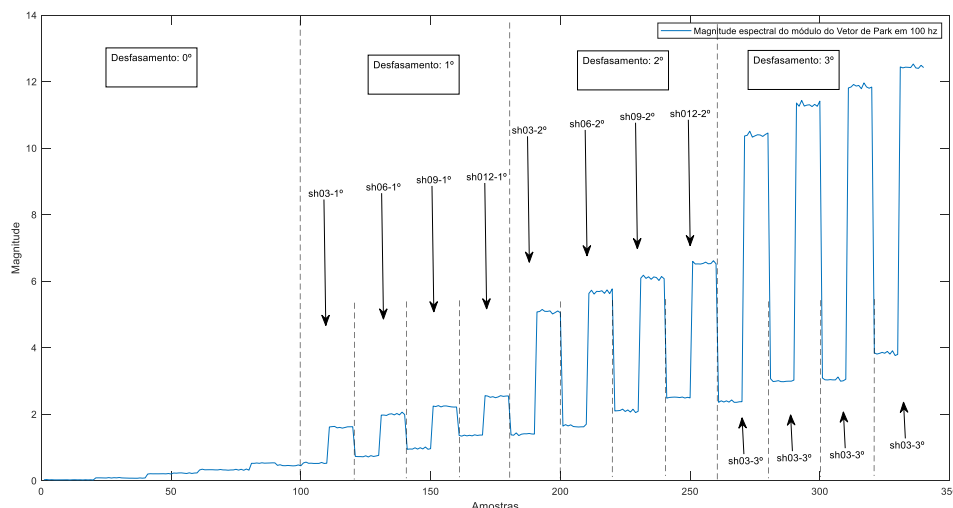


Figura 50 – Justaposição da magnitude espectral ao dobro da frequência fundamental (100 Hz) do módulo do Vetor de Park para uma janela de 0,1 s (200 pontos) aplicadas aos arquivos de dados, com diferentes desfasamentos nas tensões de alimentação, 0, 1 2 e 3 graus.

A Figura 50 demonstra que, diferentemente do cenário anterior, em condições de desfasamento nas tensões de alimentação do motor, a carga afeta drasticamente a magnitude espectral ao dobro da frequência fundamental. Note-se que a partir da amostra 100, temos condições de carga que se alternam de 10 em 10 amostras, entre 7 Nm e 13,6 Nm até à amostra 340.

Esse efeito também se reflete no desempenho do algoritmo de *Machine Learning*, que tem uma precisão reduzida para 35,3%, quando aplicado este novo *dataset* no Matlab, com Cross-validation fold: 50, Preset: Linear SVM, Kernel: Linear, Kernel scale: Automatic, Box constraint level: 10, Multiclass method: One-vs-one, Standarize data: true, como pode ser observado na Figura 51.

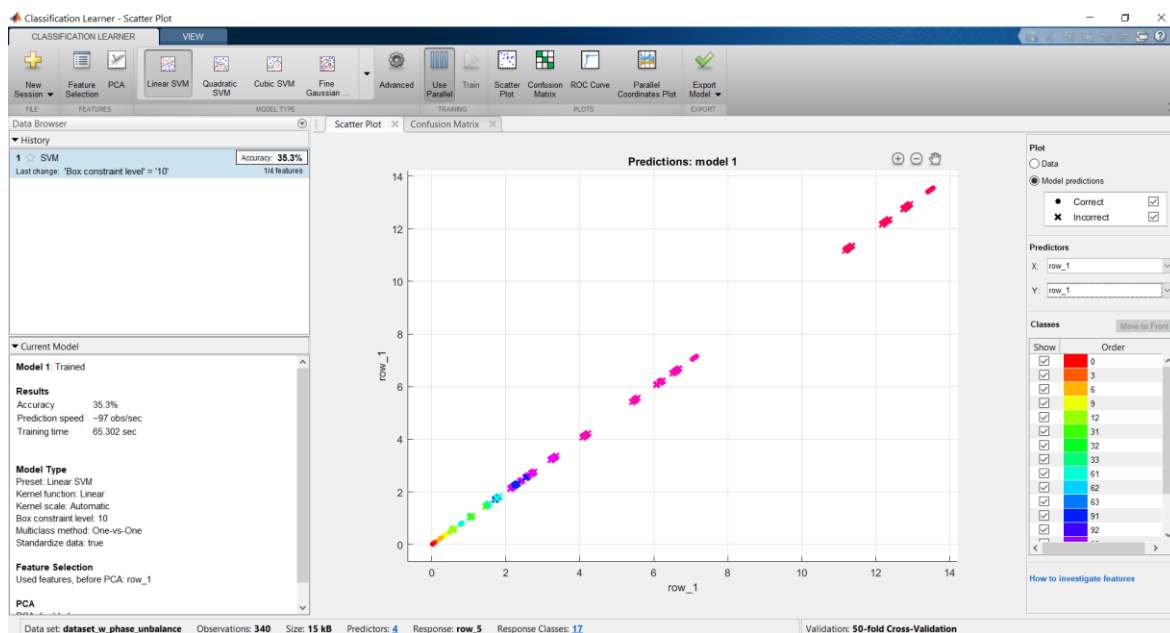


Figura 51 - Resultado da aplicação do Linear SVM, Kernel: Linear, Kernel scale: Automatic, Box constraint level: 10, Multiclass method: One-vs-one, Standarize data: true.

Deve ter-se em conta que a magnitude espectral relacionou apenas a amplitude do módulo do Vetor de Park aplicado às correntes com as avarias referentes a curtos-circuitos estatóricos. Como é ilustrado na Figura 51, há uma componente extra que é o desfasamento nas tensões de alimentação, que se reflete na amplitude do módulo do Vetor de Park e, consequentemente, na magnitude espectral em 2fs.

Desta forma, é possível afirmar que, com base no módulo do Vetor de Park, com apenas uma *feature*, nomeadamente a magnitude espectral em 100 Hz, não é possível a classificação com uma boa taxa de precisão, para avarias de curtos-circuitos entre espiras e desfasamento nas tensões de alimentação do motor de indução trifásico.

Portanto, novas *features* devem ser extraídas de forma a melhorar a sua precisão para quaisquer algoritmos de *Machine Learning*. Na próxima secção serão descritas novas *features*.

6.2.2 Modelo com distúrbios na alimentação: extração de novas *features*

Como foi visto na secção anterior, não foi possível o diagnóstico preciso utilizando-se apenas a *feat19* aplicado às correntes de alimentação. As taxas de precisão são extremamente reduzidas, na ordem de 35,3%. Optou-se, portanto, por extrair novas *features* que pudessem estar associadas às avarias, de forma a aumentar a precisão do modelo. As novas *features* foram extraídas de forma a maximizar a eficiência do modelo. A Tabela 17, resume as *features* utilizadas:

Tabela 17 - Dataset com a análise espectral do Vetor de Park e outras *features* (janela de 0,1 s).

Nome	Descrição
Feat11	EPVA - Magnitude espectral do módulo do Vetor de Park das correntes em 2fs (100 Hz)
Feat1v	EPVA - Magnitude espectral do módulo do Vetor de Park das tensões em 2fs (100 Hz)
iPos	Sequência positiva das correntes
iNeg	Sequência negativa das correntes
Feat10	Desvio padrão dos ângulos de fase das tensões
Feat13	Média das tensões
Feat14	Desvio padrão das tensões
Classe	(00, 03, 06, 09, 12, 31, 32, 33, 61, 62, 63, 91, 92, 93, 121, 122, 123)

O critério de escolha foi o mesmo que anteriormente proposto. Logo, foi realizada a correlação e o p-val das *features* x classe, conforme a Figura 52.

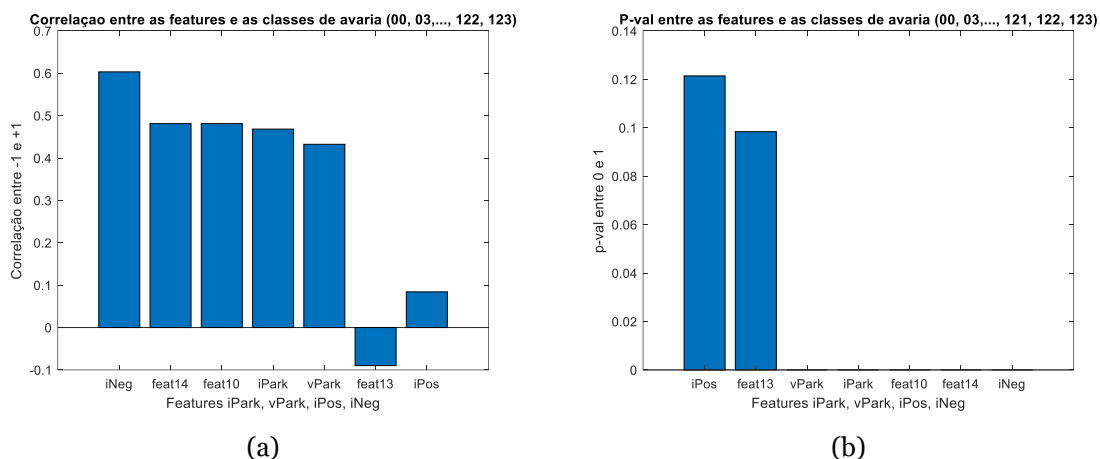


Figura 52 - Aplicação da correlação e p-val das *features* x classe: (a) correlação entre as *features* x classe; (b) p-val entre as *features* x classe.

Na Figura 52, pode observar-se que, de acordo com os critérios de correlação, temos de forma decrescente de importância a componente de sequência negativa (iNeg), o desvio

padrão da amplitude das tensões (feat14), o desvio padrão dos ângulos de fase das tensões (feat10) e a magnitude espectral do módulo do Vetor de Park para as correntes. As *features* feat13 e iPos, média das tensões e sequência positiva, respetivamente, apresentam uma correlação muito reduzida, e, pelo critério do p-val, não são significativos, podendo, portanto, ser excluídos.

Deste modo, o próximo critério é de exclusão de *features* redundantes, através da correlação entre cada uma das *features* e a classe (*features x features x classe*), como pode ser visto no mapa de calor, ilustrado na Figura 53.

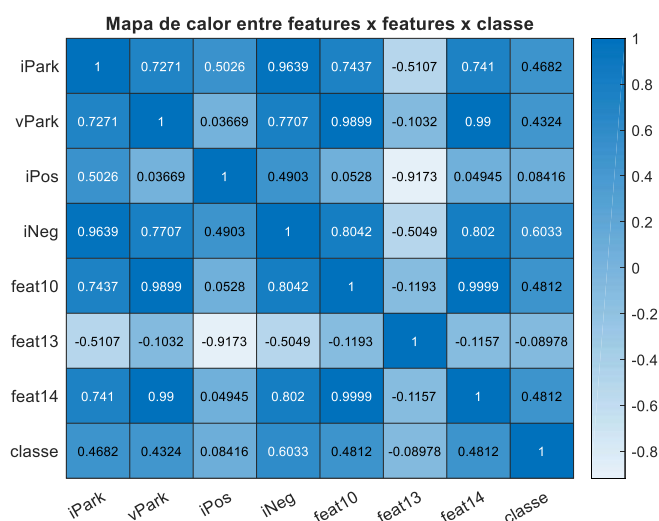


Figura 53 - Mapa de calor *features x features x classe*.

As *features* que possuem correlação maior que 0,85 serão removidas, de modo a remover a redundância. A Tabela 18 indica o critério e as *features* selecionadas.

Tabela 18 - Dataset com a análise espectral do Vetor de Park e outras *features* (janela de 0,1 s).

Critério	Feature	Inclusão
p-val > 0,5	iPos	-
p-val > 0,5	Feat13	-
Correlação entre <i>features</i> > 0,85	iPark x iNeg	iPark
Correlação entre <i>features</i> > 0,85	vPark x Feat10	vPark
Correlação entre <i>features</i> > 0,85	Feat14 x vPark	-
	Features Selecionadas	iPark e vPark

Como resultado, as *features* iPark e vPark, que representam o EPVA das correntes e das tensões, respetivamente, possuem uma elevada correlação com as avarias contidas nas

classes, de forma que são selecionadas para a utilização no modelo, enquanto a iPos, iNeg, feat10, feat13 e feat14 podem ser desprezadas.

6.2.3 Resultado modelo: fase 2

Após a pré-seleção das *features* ser realizada, estes dados foram inseridos na *Toolbox* do Matlab, e os seus resultados são apresentados na Figura 54.

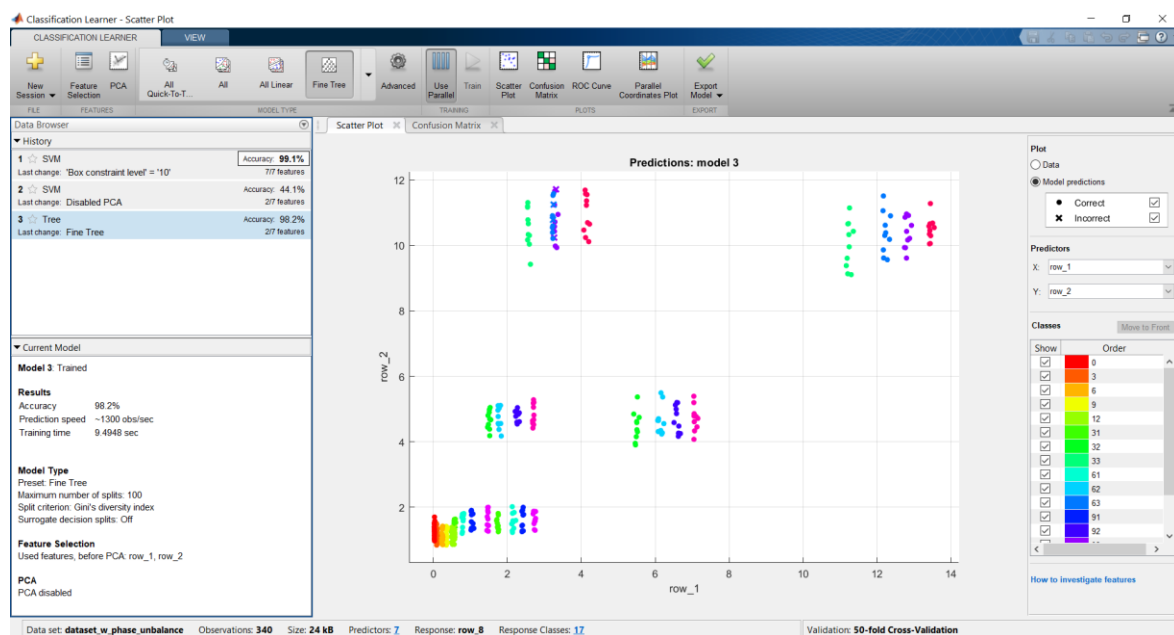


Figura 54 - Toolbox do Matlab para a criação do modelo de diagnóstico de avarias com distúrbio de fase na alimentação. Comparação entre algoritmos de *Machine Learning*.

A Figura 54, exibe 3 possíveis modelos de classificação utilizando algoritmos de *Machine Learning*. No primeiro, utiliza-se o SVM Linear com Box-constraint: 10. Neste, foram inseridas todas as 7 *features* e ele apresenta uma precisão de 98,1%. No seguinte, foi realizado um novo treino com as mesmas condições, utilizando-se apenas a iPark e vPark, que se selecionou com o critério da correlação. Neste último, o resultado foi de uma precisão de 44,1%. Essa diferença no resultado deve-se às *features* redundantes, que acabam por levar a um *overfitting*, o que leva a valores errados de precisão.

Com uma precisão real do SVM Linear realmente reduzida, optou-se por utilizar a “Decision tree” que, em condições semelhantes, após a seleção das *features* sem redundância, alcançou valores de 98,2%, dentro do conjunto de dados testados. Deve-se salientar que estes dados não estão misturados com dados de desequilíbrios de amplitude, e que este modelo tem esta taxa de acerto associada à suas classes, ou seja, só é capaz de detetar as seguintes classes de avarias: 00, 03, 06, 09, 12, 31, 32, 33, 61, 62, 63, 91, 92, 93, 121, 122,

123. Também não é capaz de detetar desequilíbrios apenas nas tensões de alimentação, pois estes dados não existiam no conjunto de dados disponibilizado. Em resumo:

- A precisão do modelo está dependente da qualidade da seleção das *features*.
- *Features* redundantes geram elevada precisão nos testes com dados conhecidos, entretanto, levam a erros em dados desconhecidos (*overfitting*).
- *Decision Tree* possui maior precisão do que o SVM, com as *features* escolhidas com o critério de correlação e valor p.
- *Decision Tree* com precisão de 98,2%, dentro do *dataset* conhecido.

6.3 Terceira Fase

6.3.1 Modelo com distúrbios na alimentação: Amplitude

Aqui, o objetivo é criar um modelo baseado em SVM, para a deteção e classificação automática de curtos-circuitos estatóricos em motores de indução trifásicos, bem como determinar a sua severidade, que varia de 00 (bom estado), 03, 06, 09 a 12 espiras em curtos-circuitos. Entretanto, desta vez, também se conta com desequilíbrios relativos às tensões de alimentação do motor, e que variam em amplitude (subtensão) de 224 V, 226 V, 228 V e 230 V. Tal como no caso anterior, usar-se-á o iPark e o vPark como *features* para a geração do modelo de *Machine Learning*.

Serão, portanto, incluídos novos arquivos. As classes serão mantidas como classes (00, 03, 06, 09, 12). Os novos dados foram precedidos de “...” para facilitar o seu reconhecimento, e estão detalhados no Apêndice B, Tabela 24.

Após esta nova organização dos dados, criou-se um novo *dataset*. O critério para a escolha das *features* foi o mesmo que anteriormente proposto. Logo, foi realizada a correlação e o p-val das *features* x classe, conforme a Figura 55.

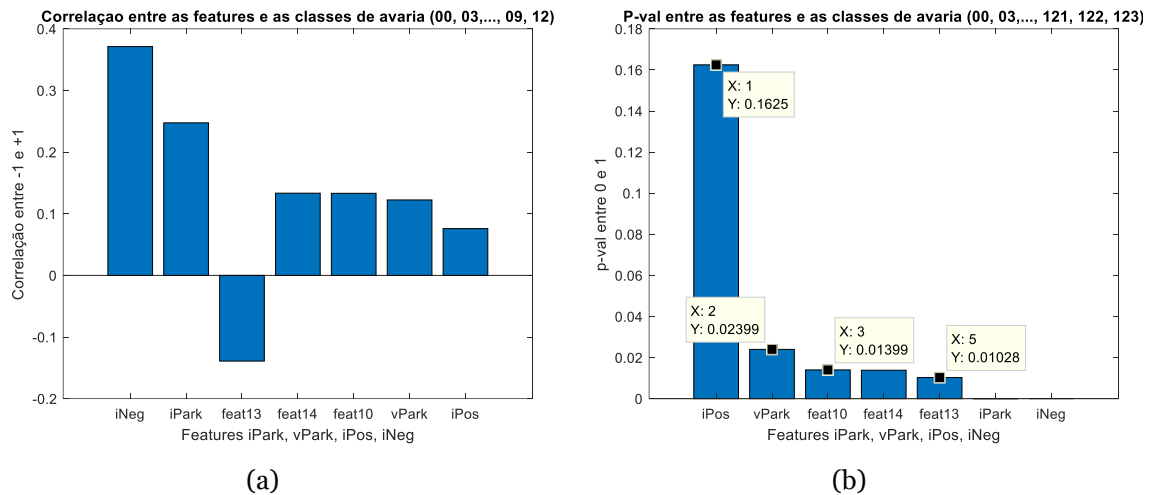


Figura 55 - (a) correlação;(b) valor p.

Na Figura 55, nota-se que a correlação entre as *features* e as classes apresentam-se em ordem decrescente, iNeg, iPark, feat13, feat14, feat10, vPark, iPos. Contudo, de acordo com o valor p, o iPos é descartado. Na Figura 56, é mostrado o mapa de calor entre as *features* x *features* x classe.

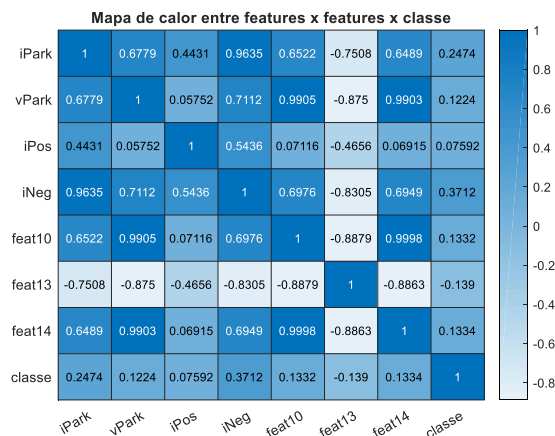


Figura 56 - Mapa de calor *features* x *features* x classe.

A partir do mapa de calor exibido na Figura 56, é possível notar uma elevada correlação, maior que 0,85, entre iPark x iNeg, vPark x feat10 x feat14 x feat13, o que resulta na decisão de manter o iPark e o vPark, conforme ilustrado na Tabela 19.

Tabela 19 - Dataset com a análise espectral do Vetor de Park e outras *features* (janela de 0,1 segundo).

Critério	Feature	Inclusão
p-val > 0,5	iPos	-
Correlação entre <i>features</i> > 0,85	iPark x iNeg	iPark
Correlação entre <i>features</i> > 0,85	vPark x Feat10	vPark
Correlação entre <i>features</i> > 0,85	Feat14 x vPark	-
Correlação entre <i>features</i> > 0,85	Feat13 x vPark	-
Correlação entre <i>features</i> > 0,85	Feat13 x vPark	-
	Features Seleccionadas	iPark e vPark

Nota-se que o valor de correlação do iNeg é superior ao iPark. Entretanto, para manter a coerência entre os diferentes modelos, optou-se por utilizar o iPark.

6.3.2 Resultado modelo: fase 3

A implementação foi realizada com as mesmas *features* anteriormente extraídas, para o caso de desequilíbrios de fase. A implementação na *Toolbox* do Matlab é apresentada na Figura 57.

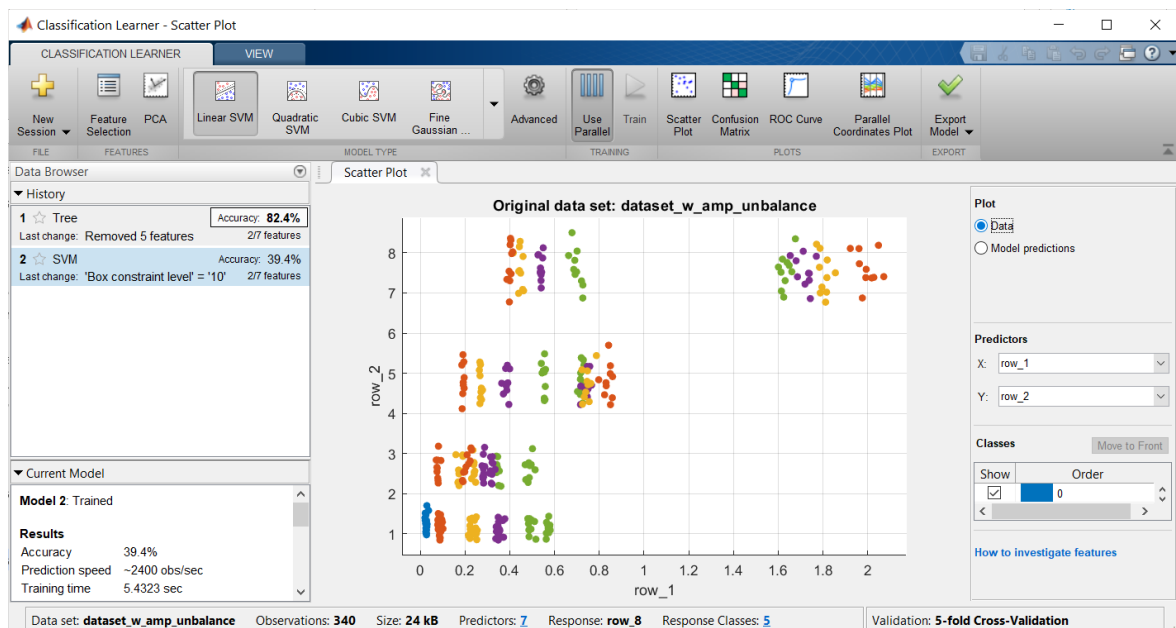


Figura 57 - Comparação entre a implementação do algoritmo de *Decision Tree* x SVM Linear (Box constraint: 10).

Em termos comparativos, a *Decision Tree* teve um desempenho superior, com 82,4% de precisão. O SVM Linear teve um desempenho realmente reduzido, de 39,4%, e foi, portanto, descartado. Em resumo:

- O *Decision Tree* obteve uma precisão de 82,4%;
- O SVM Linear, Box constraint: 10, obteve uma precisão de 39,4%;
- O iPark e o vPark, foram escolhidos a partir do processo de *feature selection*, tanto para desequilíbrios de amplitude, quanto para desequilíbrios de fase.

6.4 Quarta Fase

6.4.1 Modelo com distúrbios na alimentação: Amplitude e Fase

Nesta quarta fase, o objetivo é criar um modelo baseado em SVM, para a detecção e classificação automática de desequilíbrios nas tensões de alimentação do motor de indução trifásico. Logo, os quatro modelos criados podem funcionar de forma automatizada, redirecionando o diagnóstico de avaria para o seu respectivo modelo otimizado naquele cenário. No caso de um motor em condições normais de alimentação, o modelo treinado em condições normais gera o diagnóstico. No caso de um motor com desequilíbrios relativos à amplitude das tensões de alimentação, é diagnosticado pelo modelo com treino em condições de desequilíbrios de amplitudes. No caso de um motor com desequilíbrios relativos ao desfasamento das tensões de alimentação, é diagnosticado pelo modelo com treino em condições de desequilíbrios de fase.

Como não existiam casos isolados de desequilíbrios nas tensões de alimentação, foram selecionados arquivos com as avarias de curtos-circuitos mais reduzidos, ou seja, 3 espiras. Um novo conjunto de dados foi organizado, para a geração de um novo *dataset*. Os seus detalhes estão descritos no Apêndice C, Tabela 25.

Neste caso, por se tratar de um modelo para a avaliação das tensões de alimentação, não serão utilizadas *features* de análise das correntes. Assim, as novas *features* estão discriminadas na Tabela 20.

Tabela 20 - Dataset com a análise espectral do Vetor de Park e outras *features* (janela de 0,1 segundo).

Nome	Descrição das <i>Features</i>
Feat1v	Magnitude espectral do módulo do Vetor de Park das tensões em 2fs (100 Hz)
Feat10	Desvio padrão dos ângulos de fase das tensões
Feat13	Média das tensões
Feat14	Desvio padrão das tensões
Classe	(224, 226, 228, 230, 1230, 2230, 3230)

Desta forma, uma nova análise das *features* torna-se necessária. A Figura 58 ilustra o resultado da correlação e o valor p.

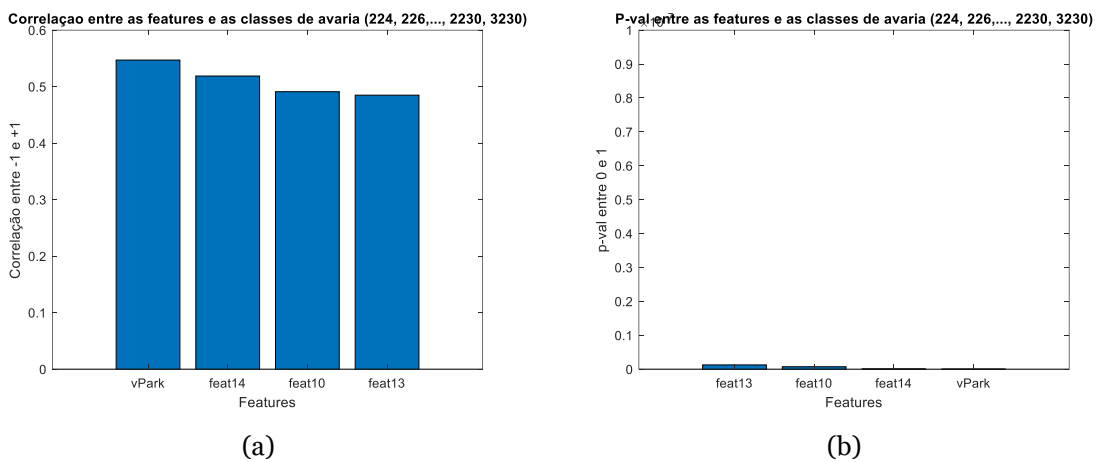


Figura 58 - Aplicação da correlação e p-val das *features* x classe: (a) correlação entre as *features* x classe; (b) p-val entre as *features* x classe.

Após a análise de correlação e do valor p, torna-se necessária a análise de redundância através do mapa de calor, ilustrado na Figura 59.

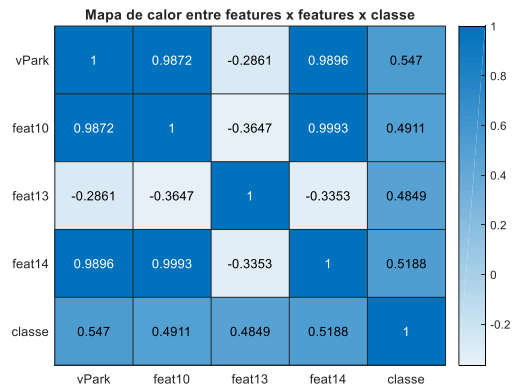


Figura 59 - Mapa de calor *features x features x classe*.

A Tabela 21 resume os critérios de inclusão e as *features* que foram selecionadas.

Tabela 21 - Seleção das *features* com base no mapa de calor da correlação.

Critério	Feature	Inclusão
Correlação entre <i>features</i> > 0,85	vPark x Feat10 x Feat14	vPark
	Features Selecionadas	vPark e Feat13

Logo, as *features* selecionadas foram introduzidas na *Toolbox* do Matlab e os seus resultados são apresentados na Figura 60.

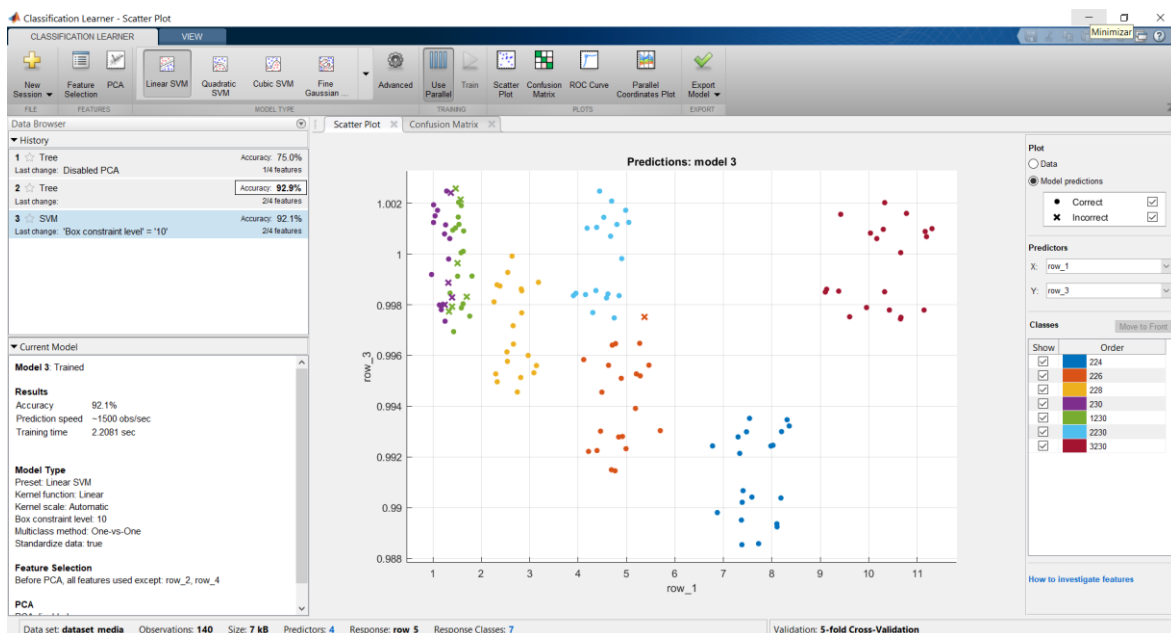


Figura 60 - Resultado da aplicação do Linear SVM, Kernel: Linear, Kernel scale: Automatic, Box constraint level: 10, Multiclass method: One-vs-One, Standarize data: true e Decision Tree. Features: vPark e Feat13.

A Figura 60, revela que o EPVA aplicado às tensões como uma única *feature* tem um reduzido desempenho para a classificação correta dos desequilíbrios relativos à amplitude e à fase, com um desempenho médio de 75%, utilizando-se o algoritmo de *Decision Tree*. Entretanto, quando associado com a média das tensões (*feat13*) é capaz de ter o seu desempenho melhorado, como é demonstrado na Figura 60. Nota-se ainda que o SVM Linear apresenta uma precisão de 92,1%, e que o *Decision Tree* teve um desempenho de 92,9%.

A matriz confusa apresentada na Figura 61 mostra que os erros se concentraram nas regiões de 230, condição normal e em 1230, condição de desfasamento de 1 grau na alimentação, mas com 230 v de amplitude.

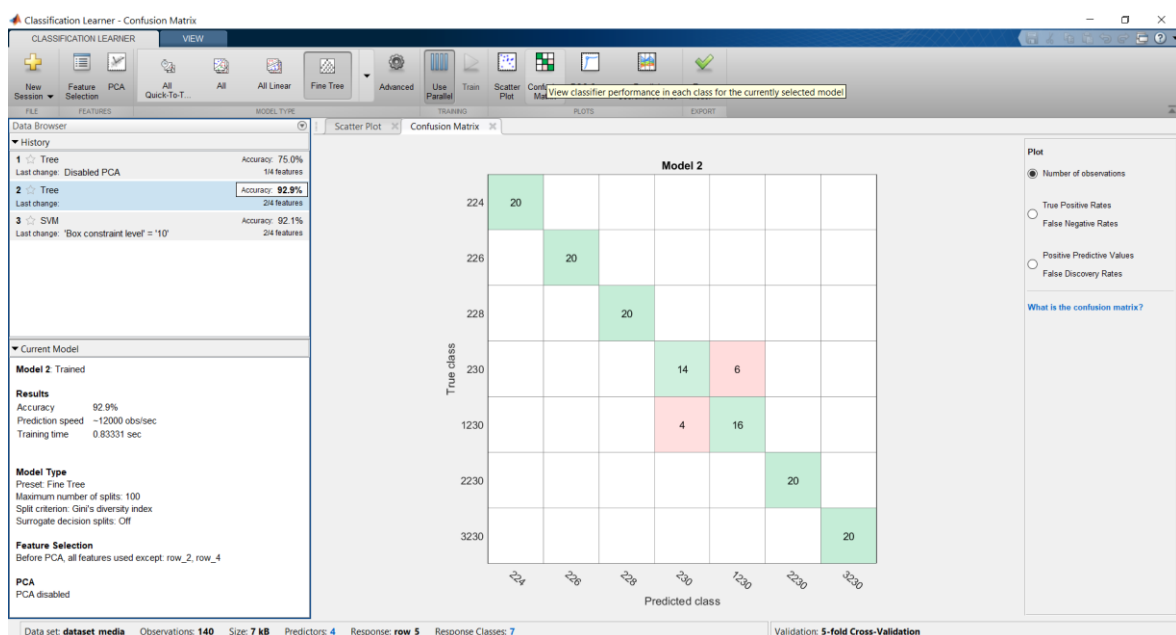


Figura 61 - Matriz confusa do modelo da aplicação do Linear SVM, Kernel: Linear, Kernel scale: Automatic, Box constraint level: 10, Multiclass method: One-vs-One, Standarize data: true. Features: vPark e Feat13.

Os resultados da implementação da quarta fase estão resumidos abaixo:

- Para a correta classificação entre desequilíbrios de tensão relativos às amplitudes ou desfasamento, o EPVA, isoladamente, associado ao SVM Linear apresenta resultados de 75% de precisão.
- Após o processo de *feature selection*, com a adição da média das tensões (*feat13*), sua precisão é de 92,1% e de 92,6%, para o SVM Linear e o *Decision Tree*, respetivamente.
- Apresenta uma certa dificuldade na classificação de avarias de desfasamento de 1 grau e a alimentação em condições normais (ver Figura 61).

6.5 Quinta Fase

Esta secção tem como objetivo comprovar os resultados obtidos através dos modelos gerados, após os ensaios anteriores. Neste caso, foram gerados 4 modelos que serão submetidos a amostras para quais não foram previamente treinados, de forma a demonstrar sua precisão.

O modelo de Rede faz a análise da Rede de alimentação e, em caso normal, redireciona o diagnóstico de avaria no motor para o Modelo Normal. Em caso de avaria na alimentação, o diagnóstico é redirecionado para o Modelo de Amplitude ou para o Modelo de Fase, de acordo com a avaria diagnosticada na Rede de Alimentação. O desenho esquemático de funcionamento dos 4 modelos encadeados é apresentado na Figura 62.

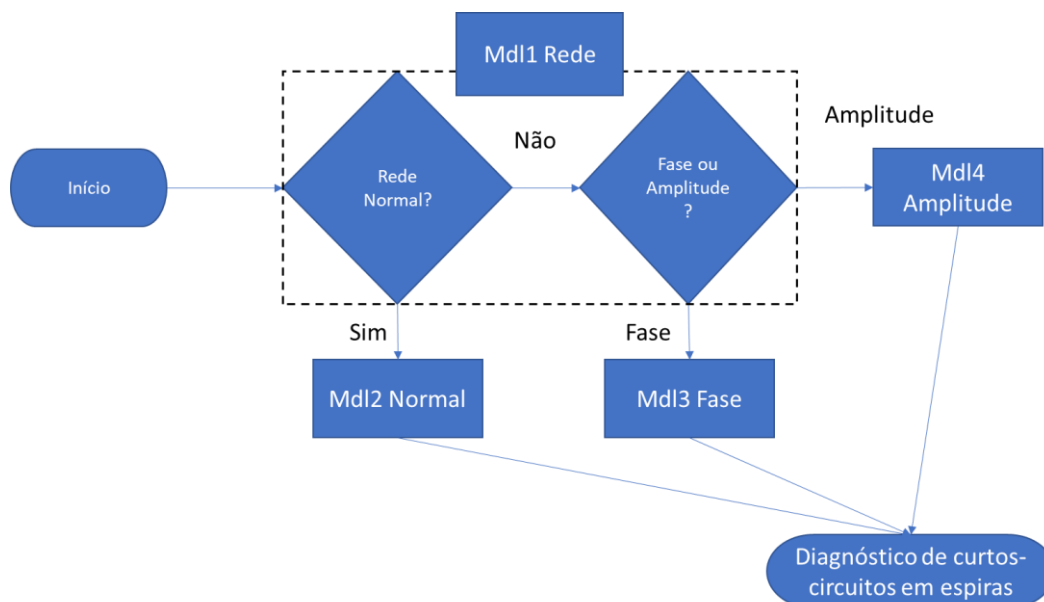


Figura 62 - Fluxo de decisões dos diferentes modelos para obtenção do diagnóstico de curtos-circuitos entre espiras.

6.5.1 Testes de validação em dados conhecidos

Como primeiro teste de auto-validação, foi inserido um arquivo no programa de diagnóstico. O arquivo faz parte dos dados utilizados para treino, nomeadamente “Healthy_I4.56_f2000_13.6Nm” e traz dados sobre um motor em condições normais de funcionamento. Deve-se notar, que os 4 modelos criados estão alinhados de maneira a apresentar os resultados relativos a tensões de alimentação desequilibradas, bem como a sua classificação, presença ou não de curtos-circuitos estatóricos, bem como a sua classificação quanto à severidade. Deve-se também salientar que o modelo gera 10

respostas, pois o algoritmo divide a amostra de 1 s em janelas de 0,1 s, e cada uma delas é submetida ao processo de diagnóstico da avaria. A explicação pertinente à janela, ou tempo ideal de amostragem pode ser visto em maiores detalhes no item 6.1.2.

Os resultados são apresentados na Figura 63.

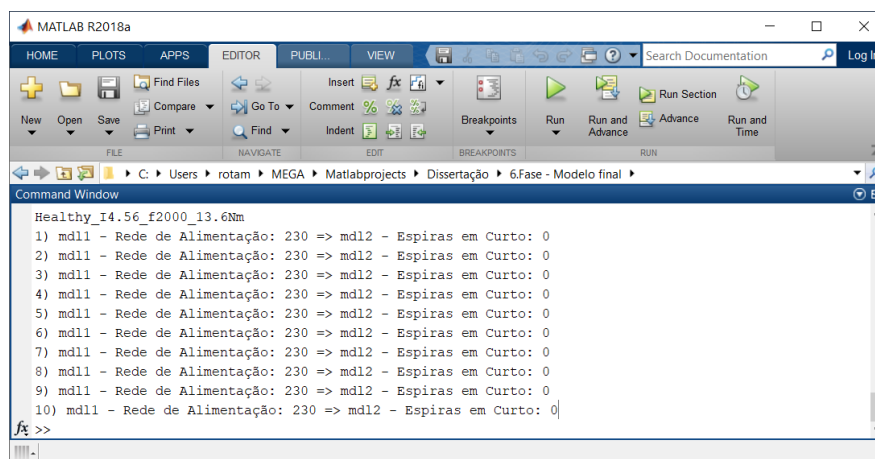


Figura 63 - Funcionamento do programa de detecção automática curtos-circuitos estáticos em motores de indução trifásicos.

6.5.2 Testes de validação em dados desconhecidos

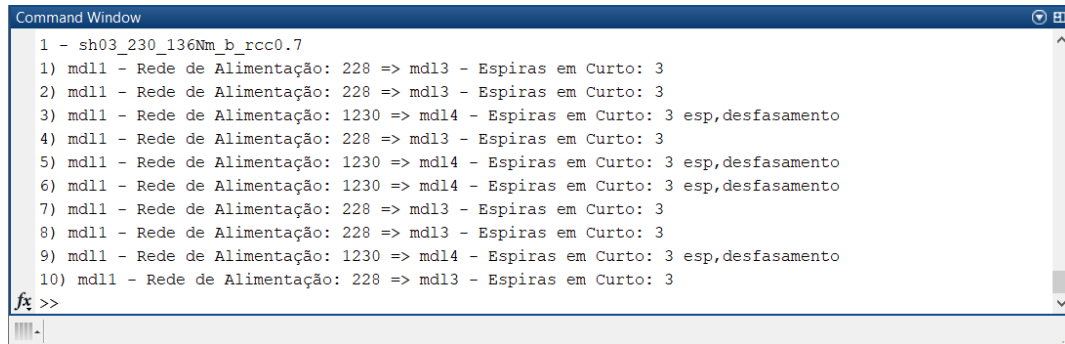
Após testes de validação preliminar no próprio *dataset* que foi utilizado para o treino e otimização do algoritmo, foi realizado um novo teste para validação, desta vez em novos dados, adquiridos no mesmo motor. Os seus resultados são apresentados na Tabela 22.

Tabela 22 - Classificação do modelo gerado em dados desconhecidos.

Classificação Real	Diagnóstico da Rede	Diagnóstico do Motor	% Acerto
1) Sh03_I4.60_230_ph12 o_ish4.5	228	3 espiras	0 % / 100 %
2) Sh09_I2.58_230_ph11 9_ish4.5	1230	3 espiras	100 % / 0 %
3) Sh09_I4.60_229_ph12 o_ish4.5	228	6 espiras	0 % / 0 %
4) Sh09_I2.58_228_ph12 o_ish4.5	228 / 1230	6 espiras	50 % / 0 %
5) Healthy	230	0	100 %

Análise dos erros de diagnóstico:

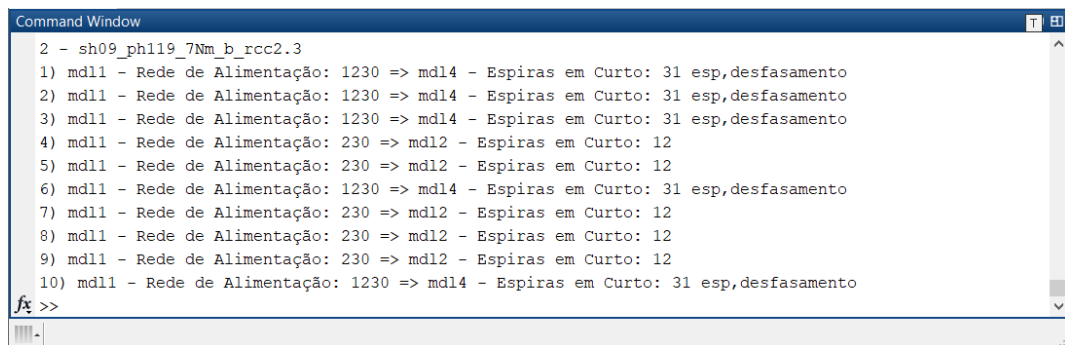
- 1) No primeiro teste notou-se uma instabilidade do algoritmo para o diagnóstico correto das tensões de alimentação embora tenha conseguido acertar na quantidade de espiras em curto-circuito. Assim, o modelo 1, de desequilíbrios nas tensões de alimentação, precisa ser melhorado. Entretanto, os modelos 3 e 4 foram capazes de acertar a severidade do curto-circuito.



```
Command Window
1 - sh03_230_136Nm_b_rcc0.7
1) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
2) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
3) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 3 esp,desfasamento
4) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
5) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 3 esp,desfasamento
6) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 3 esp,desfasamento
7) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
8) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
9) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 3 esp,desfasamento
10) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 3
fx >>
```

Figura 64 - Resultados da classificação real: Sh03_I4.60_230_ph120_ish4.5.

- 2) Nota-se, no segundo teste, uma certa instabilidade e este apresenta resultados variados. O modelo 1, avalia a rede com desfasamento de 1 grau (1230) e por vezes avaria a rede como normal (230). Entretanto, os modelos 2 e 4 erram na quantidade de espiras avariadas. O modelo 4 indica 3 espiras com 1 grau de desfasamento, enquanto que o modelo 2 indica 12 espiras, sem desequilíbrios na alimentação.



```
Command Window
2 - sh09_ph119_7Nm_b_rcc2.3
1) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 31 esp,desfasamento
2) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 31 esp,desfasamento
3) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 31 esp,desfasamento
4) mdl1 - Rede de Alimentação: 230 => mdl2 - Espiras em Curto: 12
5) mdl1 - Rede de Alimentação: 230 => mdl2 - Espiras em Curto: 12
6) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 31 esp,desfasamento
7) mdl1 - Rede de Alimentação: 230 => mdl2 - Espiras em Curto: 12
8) mdl1 - Rede de Alimentação: 230 => mdl2 - Espiras em Curto: 12
9) mdl1 - Rede de Alimentação: 230 => mdl2 - Espiras em Curto: 12
10) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 31 esp,desfasamento
fx >>
```

Figura 65 - Resultados da classificação real: Sh09_I2.58_230_ph119_ish4.5.

- 3) O modelo 1 classifica a rede de alimentação com 228 V, o que reflete um desequilíbrio na alimentação. Contudo, o modelo 3 apresenta-se bastante consistente na apresentação dos resultados em 6 espiras, e apresenta 9 espiras, uma única vez. Erra, portanto, a severidade do curto-circuito que é de 9 espiras.

```
Command Window
3 - sh09_229_136Nm_b_rcc2.3
1) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
2) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
3) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
4) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
5) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 9
6) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
7) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
8) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
9) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
10) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
fx >>
```

Figura 66 - Resultados da classificação real: Sho9_I2.58_230_ph119_ish4.5.

- 4) A rede de alimentação apresenta-se um pouco inconsistente, e aponta para uma tensão de alimentação desequilibrada com 228 V. Contudo, também apresenta 2 resultados com 1 grau de defasamento e amplitudes equilibradas (1230). O modelo 3 e o modelo 4 foram capazes de manter a consistência e apontaram para 6 espiras em curto-circuito, quando na verdade se tratava de 9 espiras em curto-circuito.

```
Command Window
4 - sh09_228_7Nm_b_rcc2.3
1) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
2) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
3) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 6 esp, defasamento
4) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
5) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
6) mdl1 - Rede de Alimentação: 1230 => mdl4 - Espiras em Curto: 6 esp, defasamento
7) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
8) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
9) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
10) mdl1 - Rede de Alimentação: 228 => mdl3 - Espiras em Curto: 6
fx >>
```

Figura 67 - Resultados da classificação real: Sho9_I4.60_229_ph120_ish4.5.

- 5) Este resultado teve uma ótima taxa de acerto, de 100%. Contudo, este foi o resultado de um motor em condições saudáveis e sem desbalanços nas tensões de alimentação.

6.6 Discussão

O capítulo 6, de aplicação do *Machine Learning* foi subdividido em 5 fases, que vão desde a construção de um método para a criação de modelos até à sua validação com novos dados. Alguns pontos importantes são destacados em seguida, para o caso de alimentação equilibrada:

- Foi possível comprovar uma maior eficiência entre as *features* no domínio da frequência, com base no EPVA apenas nas correntes, em detrimento das *features* do domínio do tempo.
- Para o caso do motor com tensões de alimentação equilibradas, foi possível obter elevadas taxas de precisão (100%) para o diagnóstico e classificação da severidade dos curtos-circuitos estatóricos, de forma automatizada, com base nos diferentes modelos de *Machine Learning* da *Toolbox Classification Learner* do Matlab. Deve destacar-se que o *windowing* adequado está diretamente ligado à elevada precisão, devido à análise espectral realizada no EPVA.
- Foi comprovada a eficiência tanto em termos da aplicabilidade do *Machine Learning*, quanto do EPVA, o que confirma a hipótese levantada no projeto de pesquisa.

No caso do motor com desequilíbrios nas tensões de alimentação, os pontos mais importantes são resumidos abaixo:

- Comprovou-se que estes desequilíbrios interferem no EPVA e, conseqüentemente, dificultam um correto diagnóstico.
- Outras *features* foram introduzidas para mitigar este problema, verificando-se que o EPVA das correntes e das tensões se destaca dentre as demais. Deve notar-se, que as componentes de sequência negativa em alguns casos tiveram uma correlação com as classes maior que o EPVA. Entretanto, optou-se por descartá-las para manter a coerência ao longo do trabalho. Contudo, esta poderá ser uma proposta para trabalhos futuros.
- O *Decision Tree* teve uma melhor generalização em relação ao SVM Linear.

Validação dos modelos gerados e resultados:

- Para a validação, 4 modelos gerados foram agrupados de forma a trabalharem de maneira interdependente. Como vantagem, oferece a possibilidade de facilitar a aplicação e sua otimização.
- A precisão foi muito menor que a prevista, apresentando erros de classificação.
- Uma proposta de melhoria seria a utilização da componente de sequência negativa e a criação de um modelo único, ao invés dos 4 modelos separados.
- Outro fator que pode levar ao aumento de precisão, é uma janela maior, que filtre pequenos transitórios, prejudiciais para o correto diagnóstico.

Observações:

Um ponto em que se deve ter uma atenção especial neste estudo, é para o caso da diferença ténue de apenas 3 espiras. No caso de um motor saudável, a deteção de 3 espiras pode ser causada por um transitório ou fatores externos ao motor, e ser facilmente confundida com um curto-circuito. O mesmo se pode dizer para o caso da classificação correta do número de espiras em curto-circuito, ou seja, um motor com 3 espiras pode ter transitórios que podem ser confundidos com 6 espiras, por exemplo. Durante a fase de validação, em alguns casos o algoritmo deu respostas consistentes, e noutras situações, pareceu estar no limiar entre duas classes, apresentando uma resposta duvidosa.

Deve-se ressaltar que este tipo de deteção é de elevada dificuldade e extremamente sensível. A Figura 68 ilustra esta sensibilidade.

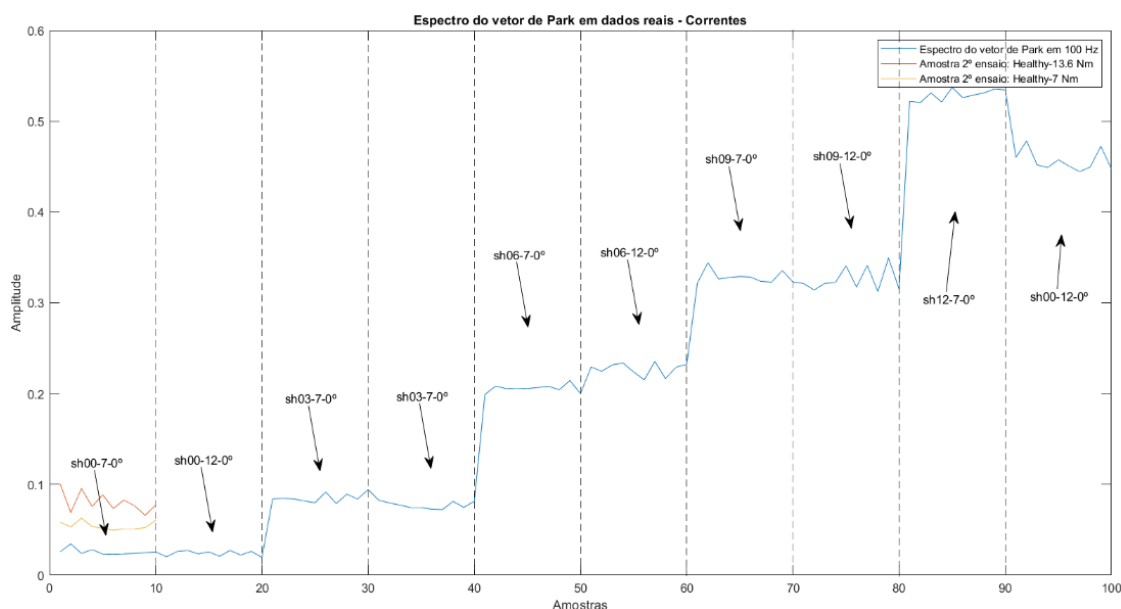


Figura 68 - Aumento na magnitude do EPVA da corrente devido ao desalinhamento do acoplamento. Amostra extraída do motor em condição saudável, em condições normais de alimentação, em diferentes condições de carga, 7 Nm e 13,6 Nm, durante o segundo ensaio.

A Figura 68 exibe um pequeno incremento na magnitude do EPVA aplicado à corrente do motor durante o segundo ensaio, em condições normais de alimentação, em diferentes condições de carga, 7 Nm e 13,6 Nm. Nota-se que, neste caso, um pequeno desalinhamento no acoplamento do motor e da carga acaba por ser refletido no EPVA. Neste caso particular, a amostra com 7 Nm apresentou um resultado que variava entre 00 (motor saudável) e 03 espiras. No caso da amostra com 13,6 Nm, o resultado do modelo foi constante de 03 espiras em curto-circuito, quando na realidade o motor estava em condição saudável. Portanto, os fatores externos contribuem negativamente para erros no correto diagnóstico de avarias.

Capítulo 7

7 Conclusão

Com os avanços na área de *Machine Learning*, o uso da Inteligência Artificial para aplicações de detecção automática de avarias em motores elétricos tornou-se um foco de estudo, com inúmeras publicações nos últimos anos. Para a sua implementação, este trabalho contou com dados reais adquiridos no Laboratório de Sistemas Electromecatrónicos do CISE (CISE | LSE), com as diversas condições de avaria relacionadas com os curtos-circuitos estatóricos 00 (saudável), 03, 06, 09 e 12 espiras. Para além disso, foram também recriadas condições de avaria nas tensões de alimentação, tanto em amplitudes, 224 V, 226 V, 228 V e 230 V (saudável), quanto em desfasamento, 0 (saudável), 1, 2 e 3 graus. Desta maneira, foi gerado um *dataset* ou conjunto de dados, que foi processado por algoritmos de *Machine Learning*, com o objetivo de se obter um classificador automático para o diagnóstico de avarias no motor.

Foi demonstrada, ao longo deste trabalho, a possibilidade de uma implementação híbrida de *Machine Learning* com o EPVA. Foram comparados os algoritmos de SVM Linear e o *Decision Tree*, onde ambos apresentaram boa precisão nos dados conhecidos ou dados de treino, em especial para os casos em que não havia desequilíbrios nas tensões de alimentação. Apesar de ambos terem a sua precisão mais reduzida do que esperado nas amostras de validação, a generalização no algoritmo de *Decision Tree* demonstrou ser mais eficiente que no SVM Linear.

Como principal limitação estão os elevados erros de classificação relacionados com os desequilíbrios nas tensões de alimentação, um fato muito comum na indústria. Os erros de classificação nestas condições devem, portanto, ser corrigidos antes que um algoritmo possa fazer o diagnóstico automático de avarias em motores de indução trifásicos.

Capítulo 8

8 Trabalhos Futuros

O presente trabalho contou com algoritmos como SVM Linear e *Decision Trees* para o diagnóstico automático de avarias. Entretanto, devido à sua complexidade, esta abordagem mostrou-se pouco eficiente face às condições impostas a um motor em condições reais de operação, simulado a partir de desequilíbrios na alimentação.

Portanto, uma abordagem mais consistente, em termos de *feature extraction* torna-se necessária. Destacam-se as *features* de tempo-frequência, como *Short-Time Fourier Transform* (STFT), distribuição de Wigner–Ville (WVD), transformada de Hilbert-Huang (HHT), análise de *wavelets* (WT), entre outras. Também existe a possibilidade de se incluir outros indicadores como as componentes simétricas e potência ativa e reativa para a obtenção de mais *features*. Há ainda a possibilidade de aplicações de algoritmos de *Deep Learning* para aumentar a precisão na classificação e junção de todos os diferentes modelos em apenas um modelo, capaz de identificar as diversas avarias, tanto na rede de alimentação quanto as relativas aos curtos-circuitos do motor.

Bibliografia

- [1] Karlsruhe Institute of Technology, “The invention of the electric motor,” *Web*, p. www.eti.kit.edu/english/1390.php, 2017, [Online]. Available: www.eti.kit.edu/english.
- [2] A. J. M. Cardoso, *Diagnosis and Fault Tolerance of Electrical Machines, Power Electronics and Drives*. IET, London, UK, 2019.
- [3] P. Waide and C. U. Brunner, “Energy-Efficiency Policy Opportunities for Electric Motor-Driven Systems,” *Int. energy agency*, p. 132, 2011, doi: 10.1787/5kkg52gb9gjd-en.
- [4] A. J. M. Cardoso, *Diagnóstico de Avarias em Motores de Indução Trifásicos*. Coimbra Editora, Coimbra, Portugal, 1991.
- [5] D. Miljkovic, “Brief review of motor current signature analysis,” *IEEE Ind. Appl. Mag.*, vol. 5, no. 1, pp. 14–26, 2015, [Online]. Available: <https://hrcak.srce.hr/148715>.
- [6] S. Dol and R. Bhinge, “SMART motor for industry 4.0,” *2018 IEEMA Eng. Infin. Conf. eTechNxt 2018*, pp. 1–6, 2018, doi: 10.1109/ETECHNXT.2018.8385291.
- [7] WEG, “Motores Elétricos - Guia de Especificação,” p. 68, 2014, [Online]. Available: www.weg.net.
- [8] M. V. Guedes, “O Motor de Indução Trifásico selecção e aplicação Faculda de de Engenharia da Universidade do Porto. O Motor de Indução Trifásico selecção e aplicação.” 1994.
- [9] V. Jose, G. Jose, P. G. Scholar, and V. Jose, “Induction Motor Fault Diagnosis Methods: A Comparative Study Induction Motor Fault Diagnosis Methods: A Comparative Study,” no. August, 2015.
- [10] A. H. Bonnett, “Root cause failure analysis for AC induction motors in the petroleum and chemical industry,” *Rec. Conf. Pap. - Annu. Pet. Chem. Ind. Conf.*, pp. 1–13, 2010, doi: 10.1109/PCIC.2010.5666831.
- [11] R. H. PARK, “Two-Reaction Theory of Synchronous Machines: Generalized Method of Analysis-Part I,” *Trans. Am. Inst. Electr. Eng.*, vol. 48, no. 3, pp. 716–727, 1929, doi: 10.1109/T-AIEE.1929.5055275.
- [12] R. H. Park, “Two-Reaction Theory of Synchronous Machines—ii,” *Trans. Am. Inst. Electr. Eng.*, vol. 52, no. 2, pp. 352–354, 1933, doi: 10.1109/T-AIEE.1933.5056309.
- [13] A. J. Marques Cardoso, S. M. A. Cruz, and D. S. B. Fonseca, “Inter-turn stator winding fault diagnosis in three-phase induction motors, by Park’s vector approach,” *IEEE*

- Trans. Energy Convers.*, vol. 14, no. 3, pp. 595–598, 1999, doi: 10.1109/60.790920.
- [14] S. M. A. Cruz and A. J. Marques Cardoso, “Stator winding fault diagnosis in three-phase synchronous and asynchronous motors, by the extended park’s vector approach,” *IEEE Trans. Ind. Appl.*, vol. 37, no. 5, pp. 1227–1233, 2001, doi: 10.1109/28.952496.
- [15] M. Ben Khader, G. Champenois, and S. Tnani, “Electrical Power and Energy Systems Reliable stator fault detection based on the induction motor negative sequence current compensation,” *Int. J. Electr. Power Energy Syst.*, vol. 95, pp. 490–498, 2018, doi: 10.1016/j.ijepes.2017.09.008.
- [16] Power Engineering Study Resource, “Sequence Components – Voltage Disturbance,” 2020. <http://voltage-disturbance.com/power-engineering/sequence-components/> (accessed Jul. 06, 2021).
- [17] H. Wehle, “ML – AI- COGNITIVE,” no. July, 2017.
- [18] A. Chahal and P. Gulia, “Machine learning and deep learning,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 4910–4914, 2019, doi: 10.35940/ijitee.L3550.1081219.
- [19] Top 6 Sites, “AI vs Machine Learning vs Deep Learning – Top 6 Sites,” 2020. <https://top6sites.com/2020/05/15/ai-vs-machine-learning-vs-deep-learning/> (accessed Aug. 20, 2020).
- [20] J. . QUINLAN, “Simplifying decision trees,” *Int. J. Hum. Comput. Stud.*, vol. 51, no. 2, pp. 497–510, Aug. 1999, doi: 10.1006/ijhc.1987.0321.
- [21] M. Swamynathan, *Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics using Python. Second Edition*. 2019.
- [22] P. D. Turney, “Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm,” *J. Artif. Intell. Res.*, vol. 2, no. February 1995, pp. 369–409, 1995, doi: 10.1613/jair.120.
- [23] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.
- [24] M. Stephen, *Machine Learning An Algorithmic Perspective Second Edition*. 2014.
- [25] Y. Lei, *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. 2016.
- [26] E. Alpaydin, *Introduction to Machine Learning MIT*, 3thd Editi. 2014.
- [27] Real Python, “NumPy, SciPy, and Pandas: Correlation With Python – Real Python,” 2020. <https://realpython.com/numpy-scipy-pandas-correlation-python/> (accessed Jul. 02, 2021).

- [28] D. Selvamuthu and D. Das, “Single-Factor Experimental Design,” in *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*, Singapore: Springer Singapore, 2018, pp. 223–264.
- [29] Minitab, “Como entender os testes t: valores-t e distribuições-t,” 2019. <https://blog.minitab.com/pt/como-entender-os-testes-t-valores-t-e-distribuicoes-t> (accessed Jul. 03, 2021).
- [30] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.
- [31] B. Gu and C. X. Ling, “Generalized Error Path Algorithm,” *Pattern Recognit.*, vol. 120, p. 108112, 2021, doi: 10.1016/j.patcog.2021.108112.
- [32] Towards Data Science, “Cross Validation Explained: Evaluating estimator performance. | by Rahil Shaikh | Towards Data Science,” 2018. <https://towardsdatascience.com/cross-validation-explained-evaluating-estimator-performance-e51e5430ff85> (accessed Jul. 10, 2021).
- [33] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.
- [34] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, “Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: a benchmark data set for data-driven classification,” *Third Eur. Conf. Progn. Heal. Manag. Soc. 2016*, no. Cm, pp. 152–156, 2016.

9 Apêndices

9.1 Apêndice A

Tabela 23 - Tabela contendo os arquivos com os dados adquiridos de tensão e correntes do motor. Frequência de amostragem: 2 kHz. Tamanho: 2.000 pontos. Incluído desfasamento na alimentação, totaliza 34 arquivos.

Nome do arquivo	Espiras em curtos-circuitos	Carga Nm	Classificação	Desfasamento na rede de alimentação
Healthy_I2.60_f2000	00	7	0	0
Healthy_I4.56_f2000_13.6Nm	00	13,6	0	0
sh03_I2.58_230_Ish_4.5_f2000	03	7	3	0
sh03_I2.58_230_Ish_4_f2000	03	7	3	0
sh06_I2.58_230_ph120_Ish_4.6_f2000	06	7	6	0
sh06_I4.58_230_ph120_Ish_4.6_f2000	06	13,6	6	0
sh09_I2.60_230_ph120_Ish_4.7_f2000	09	7	9	0
sh09_I4.60_230_ph120_Ish_4.7_f2000	09	13,6	9	0
sh12_I2.60_230_ph120_Ish_4.7_f2000	12	7	12	0
sh12_I4.62_230_ph120_Ish_4.7_f2000	12	13,6	12	0
sh03_I2.58_230_ph117_Ish_4.5_f2000	03	7	33	3
sh03_I2.58_230_ph118_Ish_4.5_f2000	03	7	32	2
sh03_I2.58_230_ph119_Ish_4.5_f2000	03	7	31	1
sh03_I4.56_230_ph117_Ish_4.7_f2000	03	13,6	33	3
sh03_I4.56_230_ph118_Ish_4.7_f2000	03	13,6	32	2
sh03_I4.56_230_ph119_Ish_4.7_f2000	03	13,6	31	1
sh06_I2.58_230_ph117_Ish_4.7_f2000	06	7	63	3
sh06_I2.58_230_ph118_Ish_4.7_f2000	06	7	62	2
sh06_I2.58_230_ph119_Ish_4.7_f2000	06	7	61	1
sh06_I4.58_230_ph117_Ish_4.7_f2000	06	13,6	63	3
sh06_I4.58_230_ph118_Ish_4.7_f2000	06	13,6	62	2
sh06_I4.58_230_ph119_Ish_4.7_f2000	06	13,6	61	1
sh09_I2.60_230_ph117_Ish_4.7_f2000	09	7	93	3
sh09_I2.60_230_ph118_Ish_4.7_f2000	09	7	92	2

sh09_I2.60_230_ph119_Ish_4.7_f2000	09	7	91	1
sh09_I4.60_230_ph117_Ish_4.7_f2000	09	13,6	93	3
sh09_I4.60_230_ph118_Ish_4.7_f2000	09	13,6	92	2
sh09_I4.60_230_ph119_Ish_4.7_f2000	09	13,6	91	1
sh12_I2.60_230_ph117_Ish_4.7_f2000	12	7	123	3
sh12_I2.60_230_ph118_Ish_4.7_f2000	12	7	122	2
sh12_I2.60_230_ph119_Ish_4.7_f2000	12	7	121	1
sh12_I4.62_230_ph117_Ish_4.7_f2000	12	13,6	123	3
sh12_I4.62_230_ph118_Ish_4.7_f2000	12	13,6	122	2
sh12_I4.62_230_ph119_Ish_4.7_f2000	12	13,6	121	1

9.2 Apêndice B

Tabela 24 - Tabela contendo os arquivos com os dados adquiridos de tensão e correntes do motor. Frequência de amostragem: 2 kHz. Tamanho: 2.000 pontos. Incluído desequilíbrios de amplitude (subtensão) na alimentação, totaliza 45 arquivos.

Nome do arquivo	Espiras em curtos-circuitos	Carga Nm	Classificação	Tensão equilibrada / desequilibrada
Healthy_I2.60_f2000	00	7	0	230
Healthy_I4.56_f2000_13.6Nm	00	13,6	0	230
sh03_I2.58_230_Ish_4.5_f2000	03	7	3	230
sh03_I2.58_230_Ish_4_f2000	03	7	3	230
sh06_I2.58_230_ph120_Ish_4.6_f2000	06	7	6	230
sh06_I4.58_230_ph120_Ish_4.6_f2000	06	13,6	6	230
sh09_I2.60_230_ph120_Ish_4.7_f2000	09	7	9	230
sh09_I4.60_230_ph120_Ish_4.7_f2000	09	13,6	9	230
sh12_I2.60_230_ph120_Ish_4.7_f2000	12	7	12	230
sh12_I4.62_230_ph120_Ish_4.7_f2000	12	13,6	12	230
...sh03_I2.58_224_Ish_4.5_f2000	03	7	3	224
...sh03_I2.58_226_Ish_4.5_f2000	03	7	3	226
...sh03_I2.58_228_Ish_4.5_f2000	03	7	3	228
...sh03_I2.58_230_Ish_4.5_f2000	03	7	3	230
...sh03_I2.58_230_Ish_4_f2000	03	7	3	230
...sh03_I4.56_224_ph120_Ish_4.7_f2000	03	13,6	3	224
...sh03_I4.56_226_ph120_Ish_4.7_f2000	03	13,6	3	226
...sh03_I4.56_228_ph120_Ish_4.7_f2000	03	13,6	3	228
...sh03_I4.56_228_ph120_Ish_4.7_f20000	03	13,6	3	228
...sh06_I2.58_224_ph120_Ish_4.6_f2000	06	7	6	224
...sh06_I2.58_226_ph120_Ish_4.6_f2000	06	7	6	226
...sh06_I2.58_228_ph120_Ish_4.6_f2000	06	7	6	228
...sh06_I2.58_230_ph120_Ish_4.6_f2000	06	7	6	230
...sh06_I4.58_224_ph120_Ish_4.6_f2000	06	13,6	6	224
...sh06_I4.58_226_ph120_Ish_4.6_f2000	06	13,6	6	226
...sh06_I4.58_228_ph120_Ish_4.6_f2000	06	13,6	6	228
...sh06_I4.58_230_ph120_Ish_4.6_f2000	06	13,6	6	230
...sh09_I2.60_224_ph117_Ish_4.7_f2000	09	7	9	224
...sh09_I2.60_224_ph120_Ish_4.7_f2000	09	7	9	224

...sh09_I2.60_226_ph120_Ish_4.7_f2000	09	7	9	226
...sh09_I2.60_228_ph120_Ish_4.7_f2000	09	7	9	228
...sh09_I2.60_230_ph120_Ish_4.7_f2000	09	7	9	230
...sh09_I4.60_224_ph120_Ish_4.7_f2000	09	13,6	9	224
...sh09_I4.60_226_ph120_Ish_4.7_f2000	09	13,6	9	226
...sh09_I4.60_228_ph119_Ish_4.7_f2000	09	13,6	9	228
...sh09_I4.60_228_ph120_Ish_4.7_f2000	09	13,6	9	228
...sh09_I4.60_230_ph120_Ish_4.7_f2000	09	13,6	9	230
...sh12_I2.60_224_ph120_Ish_4.7_f2000	12	7	9	224
...sh12_I2.60_226_ph120_Ish_4.7_f2000	12	7	12	226
...sh12_I2.60_228_ph120_Ish_4.7_f2000	12	7	12	228
...sh12_I2.60_230_ph120_Ish_4.7_f2000	12	7	12	230
...sh12_I4.62_224_ph120_Ish_4.7_f2000	12	13,6	12	224
...sh12_I4.62_226_ph120_Ish_4.7_f2000	12	13,6	12	226
...sh12_I4.62_228_ph120_Ish_4.7_f2000	12	13,6	12	228
...sh12_I4.62_230_ph120_Ish_4.7_f2000	12	13,6	12	230

9.3 Apêndice C

Tabela 25 - Tabela contendo os arquivos com os dados adquiridos de tensão e correntes do motor. Frequência de amostragem: 2 kHz. Tamanho: 2.000 pontos. Incluído desequilíbrios de amplitude (subtensão) e de defasamento na alimentação, totaliza 14 arquivos.

Nome do arquivo	Espiras em curtos-circuitos	Carga Nm	Classificação
Healthy_I2.60_f2000	00	7	0230
Healthy_I4.56_f2000_13.6Nm	00	13,6	0230
...sh03_I2.58_230_ph117_Ish_4.5_f2000	03	7	3230
...sh03_I4.56_230_ph117_Ish_4.7_f2000	03	13,6	3230
...sh03_I2.58_230_ph118_Ish_4.5_f2000	03	7	2230
...sh03_I4.56_230_ph118_Ish_4.7_f2000	03	13,6	2230
...sh03_I2.58_230_ph119_Ish_4.5_f2000	03	7	1230
...sh03_I4.56_230_ph119_Ish_4.7_f2000	03	13,6	1230
...sh03_I2.58_224_Ish_4.5_f2000	03	7	224
...sh03_I2.58_226_Ish_4.5_f2000	03	7	226
...sh03_I2.58_228_Ish_4.5_f2000	03	7	228
...sh03_I4.56_224_ph120_Ish_4.7_f2000	03	13,6	224
...sh03_I4.56_226_ph120_Ish_4.7_f2000	03	13,6	226
...sh03_I4.56_228_ph120_Ish_4.7_f2000	03	13,6	228