

## **MESTRADO**

# **MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA E EMPRESARIAL**

## **TRABALHO FINAL DE MESTRADO**

### **RELATÓRIO DE ESTÁGIO**

**ANÁLISE DE DADOS DE MANUTENÇÃO – ESTIMAÇÃO DE  
PROBABILIDADE DE FALHAS**

**INÊS PATRÍCIA CANELAS VITORINO**

**OUTUBRO - 2017**

**MESTRADO EM**  
**MÉTODOS QUANTITATIVOS PARA A DECISÃO ECONÓMICA**  
**E EMPRESARIAL**

**TRABALHO FINAL DE MESTRADO**  
**RELATÓRIO DE ESTÁGIO**

**ANÁLISE DE DADOS DE MANUTENÇÃO – ESTIMAÇÃO DE**  
**PROBABILIDADE DE FALHAS**

**INÊS PATRÍCIA CANELAS VITORINO**

**ORIENTAÇÃO:**

**DOUTOR NUNO SANTOS**

**PROF. DOUTORA MARGARIDA MARIA GONÇALVES VAZ PATO**

**OUTUBRO - 2017**



## AGRADECIMENTOS

Após estes longos meses de trabalho e dedicação, deixo aqui os mais sinceros agradecimentos a todos aqueles que me apoiaram e que, de alguma forma, contribuíram para que todo este processo fosse concluído.

Em especial, quero deixar um grande agradecimento aos meus pais por todo o apoio que me deram e por me terem proporcionado a oportunidade de realizar o mestrado, sem eles nada disto teria sido possível. Às minhas irmãs agradeço por terem sempre acreditado nas minhas capacidades e por todos os conselhos e força incondicional que me transmitiram.

Agradeço também às coordenadoras do Mestrado de MQDEE, pelo apoio direto e indireto que me prestaram ao longo deste processo. Em especial à professora Margarida Vaz Pato por ter sido minha orientadora e por toda a disponibilidade e conselhos que me transmitiu.

À equipa da PSE gostaria também de deixar um agradecimento, especialmente ao Dr. Nuno Santos por tudo aquilo que me ensinou e por toda a disponibilidade que demonstrou ao longo deste estágio. Assim como à minha colega Sara que não só estagiou comigo como me aturou e apoiou durante estes 6 meses.

Agradeço também a todos aqueles que conviveram comigo de forma mais regular. Ao meu namorado por todo o apoio e força de vontade que me transmitiu quando mais precisei. À Andreia, à Inês, às Joanas e à Micaela, por terem ouvido as minhas lamúrias e queixúmes e terem estado sempre lá a dar-me força. Assim como a todos os outros colegas e amigos que sempre me apoiaram ao longo deste processo. Um grande obrigada a todos!

## RESUMO

O presente trabalho final de mestrado resulta de uma parceria entre o ISEG e a empresa PSE – Produtos e Serviços de Estatística, Lda., tendo por base a integração num projeto sob a forma de estágio com a duração de 6 meses (março a agosto de 2017).

Sendo a PSE uma empresa de prestação de serviços focada na implementação de novas soluções, este estágio baseia-se no desenvolvimento de modelos analíticos para a gestão da manutenção de um cliente da PSE, isto é, na análise e identificação de padrões e comportamentos de um conjunto de ativos de modo a conseguir determinar, de forma antecipada, a necessidade de serviços de manutenção, ou seja, prever a ocorrência de novas falhas nesses ativos. O estudo e a previsão de ocorrências de manutenção tem uma importância central para a redução de custos, a disponibilidade dos ativos e, conseqüentemente, a produção. Desta forma, o projeto tem como título, “Análise de Dados de Manutenção – Estimação de Probabilidade de Falhas”.

Mais especificamente, o projeto prende-se com a análise de dados de manutenção na área hospitalar. Por motivos de confidencialidade de dados, toda a informação do cliente e outras que possam identificar localizações, técnicos, etc., serão anonimizadas e/ou omitidas.

Para desenvolvimento do projeto, foram disponibilizados dados de manutenção relativos ao ano de 2016, nomeadamente dados do inventário dos ativos, custos de manutenção, manutenções corretivas e preventivas que foram realizadas.

O projeto foi dividido em duas fases fundamentais:

Preparação e exploração dos dados – com o objetivo de descrever e caracterizar estatisticamente os principais indicadores e potenciais associações na manutenção;

Modelização – com o objetivo de criar modelo(s) que permita(m) conjugar tanto as condições intrínsecas aos equipamentos, como o seu histórico de manutenção e intervenções e as suas condições atuais, por forma a identificar indicadores avançados de possibilidade de falha.

Posteriormente haverá a implementação dos resultados, que corresponderá a implementação técnica do(s) modelo(s) preditivo(s) nos sistemas do cliente por forma a antecipar as necessidades de manutenção preventiva.

O projeto foi desenvolvido recorrendo ao *software Microsoft Excel, IBM SPSS Statistics e IBM SPSS Modeler*.

**Palavras-Chave:** Ativos; Ordens de serviço; Intervenções de manutenção; Manutenção corretiva; Manutenção preventiva; Previsão; Falhas; Árvores de decisão.

## **ABSTRACT**

The present master's thesis is the result of a partnership between ISEG and the company PSE – Produtos e Serviços de Estatística, Lda., and it was developed based on the integration of a six-month internship project (between March and August 2017).

PSE is a company focused on the implementation of new solutions, that is why this internship subject is to develop analytical models for the management of the maintenance of one of PSE's customers by analysing and identifying patterns and behaviours of a set of assets in order to determine, in advance, the need for maintenance services by predicting the occurrence of new faults in these assets. The study and prediction of maintenance needs is crucial to achieve costs reduction, assets availability and, consequently, production. Therefore, the project is entitled "Maintenance Data Analysis – Estimate Failure Probability".

More specifically, the project deals with the analysis of maintenance data in the hospital field. For confidentiality of data reasons, all customer information, locations, technicians, etc, will be omitted.

For the development of this project, maintenance data for the year 2016 were made available, namely data on assets inventory, maintenance costs and corrective and preventive maintenance measures that were performed.

The project was divided into two fundamental parts:

Setting and analysis of data – with the aim to describe and determine the main indicators and potential associations in the maintenance;

Modeling – with the aim to create a model(s) that allows the association of the primary condition of the equipment, its maintenance history, past interventions and its current conditions, in order to identify advanced indicators of the chance of failure.

Subsequently, the results will be implemented, which will correspond to the technical implementation of the predictive model(s) in the customer systems in order to anticipate the preventive maintenance needs.

The project was developed using *Microsoft Excel*, *IBM SPSS Statistics* and *IBM SPSS Modeler software*.

**Keywords:** Assets; Service orders; Maintenance intervention; Corrective maintenance; Preventive maintenance; Prediction; Faults; Decision trees.

## ÍNDICE

<b>AGRADECIMENTOS</b> .....	<b>II</b>
<b>RESUMO</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>V</b>
<b>ÍNDICE</b> .....	<b>VI</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>VII</b>
<b>ÍNDICE DE TABELAS</b> .....	<b>VII</b>
<b>LISTA DE ABREVIATURAS</b> .....	<b>VIII</b>
<b>CAPÍTULO 1: INTRODUÇÃO</b> .....	<b>1</b>
1.1. ENQUADRAMENTO DO TEMA .....	1
1.2. MANUTENÇÃO DE ATIVOS .....	1
1.3. OBJETIVO DO ESTÁGIO .....	2
<b>CAPÍTULO 2: APRESENTAÇÃO DA EMPRESA</b> .....	<b>3</b>
<b>CAPÍTULO 3: ENQUADRAMENTO TEÓRICO</b> .....	<b>5</b>
3.1. INTRODUÇÃO À MANUTENÇÃO DE ATIVOS .....	6
3.2. TIPOS DE MANUTENÇÕES .....	6
3.3. TIPOS DE FALHAS .....	8
3.4. CAUSAS DE FALHAS .....	8
3.5. CURVA DE MORTALIDADE (OU <i>BATHTUB CURVE</i> ) .....	9
3.6. PREVISÃO DE FALHAS .....	10
3.7. OUTRAS METODOLOGIAS APLICÁVEIS .....	11
<b>CAPÍTULO 4: ANÁLISE EMPÍRICA</b> .....	<b>12</b>
4.1. BASES DE DADOS .....	13
4.2. ANÁLISE DE VARIÁVEIS .....	14
4.3. ESCOLHA DE VARIÁVEIS PARA O MODELO .....	17
<b>CAPÍTULO 5: RESOLUÇÃO DO CASO PRÁTICO</b> .....	<b>18</b>
5.1. INTRODUÇÃO À METODOLOGIA .....	19
5.2. DESCRIÇÃO DO MODELO .....	20
5.3. REQUISITOS DE IMPLEMENTAÇÃO .....	21
5.5. FUNCIONAMENTO E REGRAS DO ALGORITMO <i>CHAID</i> .....	22

5.6. ANÁLISE DA ÁRVORE DE DECISÃO .....	29
5.7. RESULTADOS OBTIDOS .....	32
<b>CAPÍTULO 6: CONCLUSÃO .....</b>	<b>35</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>36</b>
<b>ANEXOS .....</b>	<b>39</b>
ANEXO A – ÁRVORE DE DECISÃO .....	39
ANEXO B – MODELOS TESTADOS .....	40

### ÍNDICE DE FIGURAS

Figura 1 – Metodologia da PSE .....	4
Figura 2 – Diagrama de políticas de manutenção .....	7
Figura 3 – Curva de mortalidade (ou de sobrevivência) .....	10
Figura 4 – Ciclo de vida do projeto .....	19
Figura 5 – <i>Stream</i> do modelo .....	20
Figura 6 – Tabela de contingência target x preventiva .....	20
Figura 7 – Proporção de corretivas e de preventivas.....	21
Figura 8 – Importância de preditores.....	30
Figura 9 – Primeiro troço da árvore de decisão .....	30
Figura 10 – Regras para o critério 0 [mode 0] .....	31
Figura 11 – Regras para o critério 1 [mode 1] .....	32
Figura 12 – % <i>Gain</i> do modelo \$R-Target.....	33
Figura 13 – <i>Lift</i> do modelo \$R-Target.....	33
Figura 14 – Resultados obtidos .....	34

### ÍNDICE DE TABELAS

Tabela 1 – Esperança média de vida dos ativos .....	15
Tabela 2 – Distribuição das horas trabalhadas.....	17
Tabela 3 – Distribuição dos custos de manutenção .....	17

## LISTA DE ABREVIATURAS

CHAID – *Chi-squared Automatic Interaction Detection*

EAM – *Enterprise Asset Management*

PMML – *Predictive Model Markup Language*

PSE – Produtos e Serviços de Estatística

SQL – *Structured Query Language*

## CAPÍTULO 1: INTRODUÇÃO

Neste primeiro capítulo será feita uma introdução sobre o tema em estudo, uma breve explicação relacionada com a manutenção de ativos e os objetivos do estágio enquanto TFM (trabalho final de mestrado).

### 1.1. Enquadramento do Tema

Atualmente, existe um grande desenvolvimento a nível de *software*, nomeadamente de análise de dados, pelo facto de ser cada vez maior o nível de conhecimento das organizações e haver cada vez mais informação armazenada com necessidade de ser trabalhada. Ao ritmo dos dias que correm, a velocidade de armazenamento de nova informação é muito superior à velocidade a que a mesma poderia ser analisada por meios manuais.

Esta carência de análise fez com que a procura de soluções que automatizem processos aumentasse com o objetivo de melhorar sobretudo a produtividade nas organizações, assim como fornecer um melhor apoio na tomada de decisão. [12]

Isto levou a que as organizações percebessem que, grande quantidade de informação não é sinónimo de boa informação, ou seja, boa informação é aquela que nos fornece resultados concretos em tempo determinado. A informação que já foi trabalhada é a que vai contribuir para melhorar a estratégia organizacional. [4]

Na manutenção, volume e diversidade de dados criam também a oportunidade para a adoção de ferramentas de análise de dados e aplicação de técnicas de *data mining*, seja para fins mais exploratórios, seja para fins mais confirmatórios, tanto para a descrição e caracterização dos fenómenos, como para análises de previsão de tendências e comportamentos futuros. [4]

Em termos práticos e, tendo em conta o enquadramento feito, o objetivo deste projeto é o de conseguir identificar novos indicadores e prever ocorrências através da análise de dados da informação disponibilizada.

### 1.2. Manutenção de Ativos

Antes de mais, porque se faz e para que serve a manutenção de ativos? A manutenção é feita porque não há ativos que sejam perfeitos, todos têm um prazo de bom funcionamento que depende de uma grande variedade de fatores e serve para

permitir que esse prazo de funcionamento seja mais extenso, de forma a usufruir dos mesmos durante o máximo tempo possível com a melhor qualidade e segurança.

“A manutenibilidade é um parâmetro do *design* do sistema e a manutenção é o resultado do *design*.” [2] A manutenção de um ativo requer diagnóstico e reparação, é um conjunto de ações que, interligadas, pretendem restituir o ativo às suas condições iniciais de funcionamento.

Quando se fala em manutenção, os conceitos de fiabilidade e manutenibilidade são muito importantes. A fiabilidade está intrinsecamente relacionada com a duração do equipamento, nomeadamente se falha com frequência ou não e a manutenibilidade não é mais que a preocupação em criar equipamentos que sejam facilmente reparáveis, ou seja, a partir do momento em que ocorre a falha, quão rápida e fácil é a sua manutenção. Quando estes dois conceitos se cruzam, fiabilidade e manutenibilidade, conseguimos saber a disponibilidade do equipamento uma vez que não é mais que a junção dos dois. [2]

Conseguir prevenir ou antecipar serviços de manutenção beneficia o negócio em todos os aspetos, entre outros, na redução de custos e aumento da produção, no aumento do tempo de disponibilidade dos ativos, na redução do trabalho de emergência não planeado (avarias de última hora), assim como na redução de acidentes de trabalho.

Mediante os conceitos apresentados e para o intuito deste trabalho, é necessário fazer a distinção entre dois tipos de manutenção, nomeadamente manutenção corretiva e manutenção preventiva.

Tal como os próprios nomes indicam, manutenção corretiva é aquela que corrige depois de ter ocorrido a falha. É uma intervenção reativa que tipicamente se traduz na indisponibilidade desse ativo ou equipamento. Manutenção preventiva é precisamente o oposto. É a que procura, por antecipação, evitar que ocorram as falhas. Tradicionalmente, este tipo de manutenção está normalmente associado e indexado ao tempo de operação, isto é, tal como a manutenção de um automóvel, um equipamento tem manutenções preventivas programadas em função do tempo de operação. [17]

### 1.3. Objetivo do Estágio

A realização deste estágio tem como objetivo aplicar conhecimentos que tenham sido adquiridos ao longo do mestrado de MQDEE no âmbito de um projeto real

empresarial, no caso, na integração de modelos preditivos na gestão da manutenção hospitalar. Mais do que modelos preditivos de forma lata, a finalidade é identificar os fatores e respetivos níveis que podem indiciar uma falha, e obter diariamente essa possibilidade de ocorrência para os ativos do hospital em causa.

O âmbito do estágio passa exclusivamente pela fase de descrição e caracterização dos principais indicadores de manutenção e manutenibilidade, e pela modelização dos dados. Em suma, a implementação tecnológica encontra-se fora de âmbito do projeto de estágio.

A preparação e exploração dos dados requer um estudo intenso dos mesmos, uma análise pormenorizada de cada variável explicativa de forma a conseguir perceber quais as que têm maior impacto e aquelas que melhor podem vir a explicar o aparecimento das falhas. Assim como houve necessidade de calcular novas variáveis explicativas a partir das já existentes.

Este trabalho encontra-se dividido em seis capítulos. O capítulo da introdução onde é feito um breve enquadramento sobre o tema e é explicado o objetivo do projeto. O segundo capítulo refere-se à apresentação da empresa onde foi realizado o estágio assim como do *software* utilizado. Posteriormente, no capítulo três é feito todo o enquadramento teórico (revisão bibliográfica) sobre manutenção de ativos e metodologias. Os capítulos quarto e quinto, dizem respeito à vertente mais prática do trabalho, nomeadamente às bases de dados e respetivas variáveis analisadas e estudadas e, seguidamente, à explicação do modelo desenvolvido, regras do *software*, algoritmos utilizados e resultados obtidos. No capítulo seis são tiradas as conclusões referentes a todo este projeto. Por fim, encontram-se as referências bibliográficas seguidas pelos anexos.

## **CAPÍTULO 2: APRESENTAÇÃO DA EMPRESA**

Com base numa estrutura de capital unicamente portuguesa, a PSE – Produtos e Serviços de Estatística, Lda., especializada em análise de dados, deu início à sua

atividade em 1994. Tendo como principais objetivos “implementar soluções e prestar serviços de consultoria de natureza analítica, contribuindo dessa forma para uma maior eficácia, certeza e valor na tomada de decisão e nos processos de negócio dos seus clientes e parceiros.” [19]

Atualmente, uma necessidade cada vez mais presente nas empresas e nos seus clientes é a questão analítica e a exploração dos dados para obter novos resultados que sejam, de alguma forma, mais apelativos que os atuais. Tendo isso em consideração, com mais de 1.500 clientes registados, a PSE tem experiência em inúmeros setores de atividade e nas mais diversas áreas operacionais. [19]

A abordagem metodológica da PSE em qualquer projeto onde se insere passa por “PSE – Perceber, Solucionar e Executar”, mais concretamente, perceber o problema, desenhar uma solução e colocá-la em prática de acordo com o serviço a prestar. [20]



É uma empresa que tem experiência em vários setores de atividade, nomeadamente “(telecomunicações, banca, seguros, retalho, distribuição, energia, grande consumo, governo, saúde, etc..) e nas mais diferentes áreas funcionais (marketing, risco, produção e operações, distribuição, logística, qualidade, financeira, etc...)”. [19] A PSE apresenta uma equipa qualificada na incorporação de inteligência preditiva nas organizações.

Como tal, a empresa trabalha com diverso *software* na área da análise de dados, nomeadamente o *IBM SPSS Statistics* e o *IBM SPSS Modeler* que, são utilizados neste trabalho e passam, seguidamente, a ser explicados.

### **Software IBM SPSS Statistics**

Este *software* permite ao analista validar a sua informação e escolher as técnicas que pretende utilizar, de forma mais conveniente e rápida, para responder às questões

necessárias. Considerado como “uma ferramenta modular e integrada que cobre todo o processo analítico”. [21]

Consegue ser dos mais utilizados pelo facto de ter à disposição uma série de técnicas e métodos estatísticos que facilitam o acesso aos dados. É extremamente fácil e intuitivo fazer transformações nos dados, assim como selecionar e aplicar procedimentos estatísticos para a obtenção de resultados. [21]

### **Software IBM SPSS Modeler**

Este *software* pode facilitar a tomada de decisão por parte dos seus utilizadores através de vários métodos que utilizam algoritmos avançados que englobam análise de entidade, análise de texto e otimização. É um programa extremamente completo em termos de análise preditiva e foi configurado de forma a “agregar inteligência preditiva às decisões tomadas por indivíduos, grupos, sistemas e pela empresa”. [8]

## **CAPÍTULO 3: ENQUADRAMENTO TEÓRICO**

Neste capítulo é feita uma abordagem mais teórica sobre manutenção de ativos. São aprofundados temas como os diferentes tipos de manutenção que existem, as

falhas que podem ocorrer ao longo da vida útil dos ativos, os motivos pelos quais essas falhas podem ocorrer (causas), sobre a durabilidade dos mesmos (tempo de vida dos ativos), sobre previsão de falhas, a metodologia que foi utilizada assim como outras metodologias passíveis de aplicar a estes estudos.

### 3.1. Introdução à Manutenção de Ativos

De acordo com Moubray [16], a manutenção é uma combinação de características técnicas, administrativas e de direção que pretendem manter os ativos fiáveis durante o seu ciclo de vida e, quando necessário, restaurá-los. A gestão da manutenção servirá para caracterizar o processo de liderança de direção de uma organização.

A manutenção contribui mais do que nunca para que uma empresa consiga alcançar os seus objetivos. Cada vez mais, a grande maioria dos processos empresariais é executada através de meios computacionais e maquinarias que, necessitam de ter a garantia de bom funcionamento. Por isso, ter uma manutenção adequada, mantém o custo do ciclo de vida dos ativos baixo, assim como contribui para um desempenho geral positivo nas empresas, consoante os autores Waeyenbergh e Pintellon [26].

### 3.2. Tipos de Manutenções

Segundo Assis [2], existem intervenções de manutenção de duas naturezas distintas: manutenções corretivas (para as falhas casuais e imprevisíveis) e manutenções preventivas (para falhas provenientes de degradação progressiva) que se subdividem noutros tipos de manutenções de acordo com as diversas variantes da falha, como representa a figura 2.

Por outro lado, existem autores como Wireman [27] que, têm outro ponto de vista no que refere a este tema e fazem referência a quatro naturezas distintas de intervenções de manutenção: reativa, corretiva, preventiva e preditiva.

Para este estudo em concreto, será aprofundada a teoria de Assis uma vez que o projeto apresenta apenas manutenções corretivas e preventivas para análise.

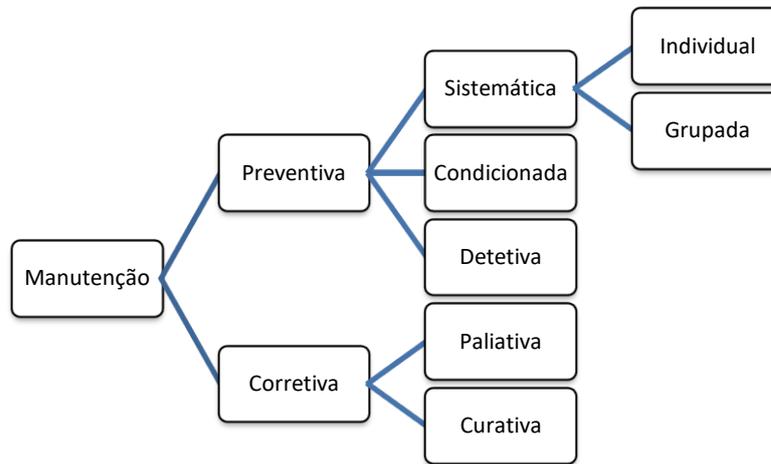


Figura 2 – Diagrama de políticas de manutenção

[Fonte: Baseado no livro de Assis [1], cap.5, p. 205]

A manutenção preventiva, também designada por pró-ativa ou planeada, pretende evitar (prevenir) que um ativo falhe. A manutenção corretiva ou reativa, pretende restituir as condições normais de funcionamento dos ativos. [2]

Posto isto, a manutenção preventiva subdivide-se em sistemática, condicionada e detetiva, como mostra a figura 2.

Quando se fala em intervenções sistemáticas significa que ocorrem de forma regular (ou a partir de determinada idade do ativo) enquanto que as condicionadas, por norma, só ocorrem no fim da vida útil dos mesmos, altura em que, ao observar as tendências de degradação, é possível prever determinadas ocorrências. As intervenções detetivas, como o próprio nome indica, tentam encontrar falhas que estejam disfarçadas, ou seja, quando determinado componente de um ativo já não funciona nas devidas condições mas não se encontram sinais visíveis nem de impedimento para que o ativo em si deixe de funcionar. Este tipo de falha costuma ser revelada a partir de testes de funcionamento no local ou só mesmo depois de se desmontar o equipamento. É o caso, por exemplo, de aparelhos de medida, válvulas de segurança ou monitorizadas, etc. [2]

Por outro lado, temos a manutenção corretiva que se subdivide em paliativa e curativa. “A manutenção corretiva é, muitas vezes, um “paliativo” e a manutenção curativa a “cura” (por exemplo, um disjuntor dispara frequentemente, uma intervenção paliativa limita-se a rearmar o disjuntor, enquanto uma intervenção curativa procura a(s) causa(s) dos disparos e elimina-a(s)).” [2]

Posteriormente, a manutenção preventiva sistemática pode ser de dois tipos, individual ou grupada. A individual é a manutenção feita componente a componente e a grupada é realizada quando deixa de ser apenas um componente e passam a ser vários em simultâneo. [2]

### 3.3. Tipos de Falhas

De acordo com Assis [2], a ocorrência de uma falha significa que houve uma interrupção no desempenho do equipamento, usualmente designada por desgaste ou degradação de determinado elemento, de forma a considerar o seu funcionamento insuficiente.

Quando ocorre uma falha de um equipamento, esta pode ser de maneira súbita (falha catastrófica), ou lenta e progressiva (falha por degradação ou deriva). [2]

Se a falha for proveniente da variação repentina (ocorrência imprevisível) de uma ou mais características de um equipamento, por norma com origem externa, é designada por falha catastrófica e, conseqüentemente, costuma avariar o equipamento. Por outro lado, se a falha for proveniente da variação contínua de uma ou mais características de um equipamento, por norma com origem interna, é designada como falha por degradação. Este tipo de falhas também pode ocorrer através de influências externas, por exemplo outro qualquer equipamento com o qual este interaja.

### 3.4. Causas de Falhas

Almeida *et al.* [1], referem que a causa da falha é uma parte essencial da informação necessária para se conseguir evitar que a mesma reincida. Uma falha que possa provocar a interrupção do funcionamento do equipamento pode dar origem a prejuízos extremamente elevados.

Desta forma, Assis [2] enumera várias causas que podem levar um equipamento a falhar, nomeadamente: erros de projeto (deficiente conceção), má seleção de material, defeitos de fabrico, manutenção inadequada (ou omissa), deficiente operação e condições de ambiente imprevistas.

Relativamente a erros de projeto (deficiente conceção), estes ocorrem na fase de projeto dos equipamentos e podem ser devido a:

- Omissão de algumas características operacionais;

- Informação incompleta das cargas e das condições ambientais em que irá funcionar;
- Cálculos errados e ausência de procedimentos e especificações de engenharia por norma reconhecidos como minimizadores da probabilidade de falha.

A má seleção de material caracteriza-se por escolhas inadequadas face às condições de serviço (carga e meio) e costuma ser dos erros mais comuns.

Os defeitos de fabrico ocorrem durante o processo de construção dos equipamentos e podem ser: porosidades e inclusões nas fundições, tensões residuais de maquinaria, descarbonização, porosidades, fissuras e outros defeitos de soldadura, desalinhamentos, falhas de aperto e outros defeitos de montagem.

Na manutenção inadequada (ou omissa) as falhas de manutenção mais frequentes, são: falta de lubrificação ou a sua prática ser feita em intervalos de tempo despropositados, remontagens inadequadas (desalinhamentos, *e.g.*) e desleixo ou ignorância na monitorização dos ativos (temperatura, pressões, vibrações, etc).

A deficiente operação acontece quando, por acidente, desleixo ou negligência, os ativos são sujeitos a períodos excessivos de carga, pressão ou temperatura, *e.g.*, o que contribui para uma degradação mais acentuada dos mesmos, logo à ocorrência de falhas.

As condições de ambiente imprevistas também podem fazer com que os ativos tenham uma degradação mais acentuada pelo facto de os equipamentos ficarem sujeitos a armazenagens prolongadas ou a várias operações de manuseamento, embalagem ou transporte.

Um equipamento encontra-se sujeito a qualquer um destes tipos de situações, propícias à aceleração do desgaste e degradação, a partir do momento em que deixa as instalações do fabricante para chegar ao(s) respetivo(s) cliente(s). [2]

### 3.5. Curva de Mortalidade (ou *bathtub curve*)

A curva de mortalidade é também conhecida por curva da “banheira” (ou *bathtub curve*).

De acordo com Assis [2], esta curva é representada através de três períodos específicos na vida dos ativos: o período de infância, o período de maturidade e o período de degradação, que se encontram respetivamente numerados na figura 3.

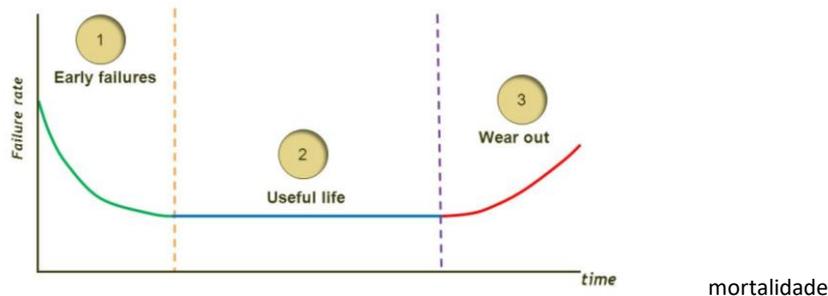


Figura 3 – Curva de

mortalidade

[Fonte: Imagem do Google referente a *bathtub curve*]

Quando ocorre uma falha no período de infância (1 – *Early failures*) por norma será numa fase inicial da vida do ativo, tal como o próprio nome indica. É precisamente nesta altura, quando os equipamentos entram em funcionamento pela primeira vez, que apresentam uma taxa repentina de falhas, motivo pelo qual  $h(t)$  é decrescente. Esta fase tende a passar rapidamente até entrar num período mais estável, o de maturidade ou período de vida útil (2 – *Useful life*).

No período de maturidade ocorrem falhas casuais, quando ocorrem ( $h(t)$  é constante e igual a  $\lambda$  que representa a taxa de falhas). Nesta fase as falhas advêm de fatores extrínsecos, por exemplo do facto de se forçar o equipamento a realizar um esforço mais acentuado que o habitual, o que pode originar falhas mais graves. É este período que define a vida útil dos ativos por ser o mais constante e prolongado.

No período de degradação ou de envelhecimento (3 – *Wear out*) é quando ocorrem mais falhas ( $h(t)$  é crescente), principalmente com intervalos de tempo mais curtos uma vez que é nesta fase que a taxa de falhas aumenta de forma acentuada devido à degradação natural dos ativos (fadiga, corrosão, desgaste, etc). Estes começam a deixar de funcionar e são necessárias substituições de equipamentos.

### 3.6. Previsão de Falhas

Uma previsão apresenta duas características: a estabilidade e a resposta. Apesar de serem ambos atributos desejáveis numa previsão, podem entrar em conflito entre si, segundo a opinião de Assis [2].

Estabilidade é a propriedade de não haver irregularidades na previsão para que haja uma evolução consistente da mesma. Quando a previsão tem estabilidade significa que é sólida e equilibrada e é tanto maior quanto maior for o número de dados históricos a considerar em análise.

Por outro lado, a resposta é a capacidade de adaptação da previsão às alterações do nível base do indicador (no caso de surgimento de tendência). Quanto maior for o peso atribuído aos dados históricos mais atuais, maior será o nível da resposta.

A qualidade de uma previsão é considerada mediante indicadores que consigam indicar o erro (ou desvio) de uma previsão quando relacionada com a realidade à *posteriori*. No caso de previsão temporal, interessam, para este estudo, dois métodos, o método de alisamento exponencial e o método de regressão linear.

De acordo com Taylor [25], o método de alisamento exponencial é fortemente utilizado quando o principal objetivo do estudo em causa requer análises a grandes conjuntos de informação, como, por exemplo, o controlo de inventário.

O método de regressão linear, segundo Henriques [6], é um método que serve para modelar a relação entre duas ou mais variáveis, através de uma equação linear. Uma das variáveis é a independente (ou explicativa) e a(s) outra(s) a(s) dependente(s). Este método é maioritariamente aplicado para previsões sob o pressuposto da continuidade (correlação duradoura entre as variáveis).

Este último método será o utilizado conforme explicação no capítulo da análise de resultados (quinto capítulo), onde é construída uma árvore de regressão.

### 3.7. Outras Metodologias Aplicáveis

Existem outras metodologias aplicáveis quando se trata do estudo de previsões e estimações, para além das que foram referidas no ponto anterior, que se baseiam em métodos determinísticos e em métodos probabilísticos (ou estocásticos).

Segundo Santos [22], os métodos determinísticos consistem no estudo de determinados fatores que afetam os dados em análise e na compreensão dos seus mecanismos de atuação. Posteriormente, estes fatores são transformados em fórmulas que demonstram a sua ação ao longo do tempo, até se atingir um valor mínimo de desempenho que seja admissível.

Campozano *et al.* [3] referem que os métodos determinísticos são baseados em modelos matemáticos que permitem atingir o mesmo resultado com determinada condição inicial, sem considerar a possibilidade de existência de aleatoriedade.

De acordo com Hovde [7], para o estudo dos métodos determinísticos não se considera a avaliação de incerteza ou a variação estatística dos fatores, além disso os valores utilizados devem ser exatos.

Dentro da classe dos métodos determinísticos, o método fatorial é dos que tem melhor receptividade na comunidade científica e isso deve-se, essencialmente, à sua elevada operacionalidade. Este método permite fazer estimações sob determinadas condições específicas, sempre com base nos dados disponíveis. [22]

Por outro lado, os métodos probabilísticos (ou estocásticos), segundo Santos [22] têm, como base, o cálculo probabilístico que estabelece a probabilidade de ocorrência de mudanças de estado dos elementos de forma a tentar superar incertezas. Campozano *et al.* [3] referem que estes métodos, em contraste com os métodos determinísticos, fornecem estimativas de probabilidade para os resultados.

Apesar de serem promissores, estes métodos ainda não têm grande margem de aplicação tendo em conta o seu grau de complexidade por necessitarem de grandes quantidades de informação para serem fiáveis e terem demasiada dependência do trabalho de campo, tendo em conta o ponto de vista de Sousa [24].

Dentro da classe dos métodos probabilísticos, o baseado no modelo de Markov (*Markov Chain*) é bastante utilizado e reconhecido. Este modelo é um processo estocástico que se caracteriza pelo simples facto de os acontecimentos passados não influenciarem os acontecimentos do futuro, a previsão depende unicamente dos acontecimentos atuais. [29]

O nome deste modelo foi dado em homenagem ao matemático russo Andrey Markov. [26]

Existem vários estudos realizados, tanto por autores de livros como de teses, que têm como base estes métodos de análise probabilística e/ou estocástica, como por exemplo Silva [23], Zhao [28], Casimiro [5], Peña [18] e Montgomery *et al.* [15].

#### **CAPÍTULO 4: ANÁLISE EMPÍRICA**

Neste capítulo é feita uma análise mais aprofundada primeiramente às bases de dados fornecidas e depois às variáveis mais relevantes para o estudo em concreto. No ponto 4.1. é explicada a informação que foi fornecida nos vários ficheiros e no ponto

4.2. são revelados os grandes números, onde se percebe de forma mais concreta a dimensão do projeto.

#### 4.1. Bases de Dados

Para a realização do estudo dos ativos, foram fornecidas bases de dados com informação referente a todos os ativos utilizados e que necessitaram de manutenção, ao longo do ano 2016, nomeadamente, ficheiros referentes ao inventário de ativos, custos de manutenção, manutenções corretivas e preventivas.

- **Inventários**, este ficheiro é composto por todos os ativos que pertencem à organização em estudo. Cada um deles é diferenciado através de um código, ou seja, por muito que sejam ativos semelhantes, apenas existe um ativo por código, por exemplo, para o código “1001C9995” existe o ativo “Coletores de inox - Rede de abastecimento de água” e assim sucessivamente. Este ficheiro é composto por 17 variáveis nomeadamente, “Ativo”, “Código de localização”, “Descrição de localização”, “Localização”, “Descrição”, “Organização”, “Status”, “Designação de projeto”, “Fabricante”, “Modelo”, “Número de série”, “Departamento”, “Data de início de operação”, “Data de retirada”, “Preparação”, “Família” e “Tempo de vida (em anos)”. Estas variáveis descrevem e caracterizam de forma minuciosa cada um dos ativos em estudo.

- **Custos**, este ficheiro é constituído por todos os custos provenientes de manutenções corretivas e preventivas ocorridas no ano 2016. Esses custos encontram-se associados a “ordens de serviço”, nome pelo qual é designado cada “pedido de manutenção”. Uma ordem de serviço pode ser constituída por um ou mais pedidos de manutenção que, por norma são referentes ao mesmo ativo ou semelhantes. Este ficheiro é composto por 12 variáveis nomeadamente, “Ordem de serviço”, “Tipo”, “Organização”, “Moeda”, “Custo de materiais”, “Custo de mão-de-obra”, “Custo da ferramenta”, “Custo total”, “Horas trabalhadas”, “Status”, “Data de término programada” e “Descrição”. Estas variáveis permitem determinar quais as ordens de serviço que têm um maior peso a nível de custos e de horas trabalhadas em serviços de manutenção. Estas variáveis são de extrema importância uma vez que distinguem as

ordens de serviço corretivas das preventivas, transmitindo uma nova percepção à empresa sobre as manutenções que mais afetam os custos.

- **Manutenções corretivas**, este ficheiro é composto por todas as ordens de serviço de manutenções corretivas que foram efetuadas entre janeiro e dezembro do ano 2016. Este ficheiro tem mais variáveis explicativas do que o das manutenções preventivas, nomeadamente, “Ordem de serviço”, “Descrição de avaria”, “Nome do chamador”, “Descrição grau”, “Compartimento reportado”, “Compartimento reportado no pedido”, “Descrição do equipamento localização”, “Código do equipamento”, “Descrição do equipamento”, “Status”, “Mão-de-obra (horas)”, “Custo (peças)”, “Empregados” e “Mês”, o que perfaz um total de 14 variáveis. Estas permitem que haja alguma segmentação relativamente a manutenções semelhantes, facilitando análises às mesmas.

- **Manutenções preventivas**, este ficheiro é constituído por todas as ordens de serviço de manutenções preventivas que ocorreram entre janeiro e setembro do ano 2016. Para estas manutenções têm-se as seguintes 14 variáveis, “Ordem de serviço”, “Código de MP”, “Descrição”, “Equipamento compartimento”, “Localização”, “Localização arquitetura”, “Descrição localização”, “Rota”, “Descrição do equipamento”, “Tarefas”, “Data de término da atividade”, “Status”, “Nº relatório” e “Mês”. Assim como acontece no ficheiro anterior, estas variáveis permitem que haja alguma segmentação relativamente a manutenções semelhantes, facilitando análises às mesmas.

#### 4.2. Análise de Variáveis

Mediante as bases de dados apresentadas, há que fazer o enquadramento das variáveis mais relevantes no que diz respeito à manutenção de ativos. Para um total de 65.535 ativos disponíveis, sabe-se que, em relação ao ano 2016:

- O hospital é constituído por 42 departamentos diferentes;
- Existem cerca de 715 famílias de ativos e 864 tipos de fabricantes distintos;
- Cerca de 28.818 ativos têm uma esperança média de vida de 40 anos;
- Houve um total de 2.714 ordens de serviço corretivas e 19.572 preventivas;

- Encontram-se 536 localizações de equipamentos para as ordens de serviço corretivas e 1.981 para as ordens de serviço preventivas;
- Existem 23 funcionários para a manutenção das ordens de serviço corretivas;
- Foram gastas aproximadamente 5.441 horas em manutenções corretivas e 36.043 horas em manutenções preventivas;
- Para os custos de manutenção foram gastos cerca de 87.738€ com as manutenções corretivas e cerca de 489.633€ com as preventivas.

### **Ativos e Famílias de Ativos**

A representar 37% do total de ativos, existem 13 famílias de equipamentos (do total das 715 famílias) com mais de 1.000 elementos por família, nomeadamente, “Tomada de energia dupla de calha”, “Grelha AVAC”, “Detetor de incêndio iónico”, “Armadura de emergência em LED”, “Alçapão”, “Interruptor simples de iluminação”, “Teto em gesso cartonado liso”, “Sinalizador de alarme de incêndio”, “Parede de gesso cartonado”, “Porta interior”, “Grelha AVAC”, “Rodapé em meia cana na continuação pavimento vinílico até cota 100mm” e “Armadura de iluminação dupla tipo - T3”.

Mais de 80% das famílias de ativos são constituídas por menos de 100 ativos, ou seja, por norma uma família de ativos é constituída por poucos equipamentos.

Relativamente à esperança média de vida dos ativos, a representar 44% do total dos 65.535 ativos, temos aqueles que se espera que durem cerca de 40 anos.

Esperança de vida	Nº de ativos	Percentagens (%)
Entre 0 e 10 anos	15.616	23,828
Entre 11 e 30 anos	21.098	32,194
40 anos	28.818	43,973
≥ 50 anos	3	0,005
TOTAL	65.535	100

Tabela 1 – Esperança média de vida dos ativos

Como exemplo de ativos com esperança média de vida de 40 anos temos, “Transformador de potência”, “Pára-raios”, “Parede de grelha em lâminas de alumínio”, “Pinos plástico”, “Rede de rega - piso B2”, etc.

### **Ordens de Serviço Corretivas**

Para um total de 2.714 ordens de serviço existem 107 trabalhadores designados por "chamadores" que são os funcionários que fazem o registo dos pedidos de manutenção às ordens de serviço.

De forma a atribuir diferentes prioridades às manutenções que vão ocorrendo, a representar 2.472 ordens de serviço, equivalentes a 91% do total de corretivas, encontra-se a designação "baixa prioridade, há disponibilidade do serviço", ou seja, a grande maioria das manutenções não provoca grande transtorno ao normal funcionamento do hospital.

Existem ainda 23 funcionários que são os técnicos de manutenção, entre engenheiros, canalizadores, técnicos, eletricitas, engenheiros elétricos (tal como na empresa são designados estes profissionais), técnicos de elevadores, técnicos de escadas rolantes, etc. Aqueles que têm maior número de horas de trabalho acumuladas são os eletricitas, com cerca de 880 horas gastas em manutenções.

### **Ordens de Serviço Preventivas**

Como a quantidade de ordens de serviço preventivas é muito superior às corretivas, pelo facto de haver ativos que requerem inúmeras manutenções preventivas ao longo da sua vida útil, existe distinção entre o total de ordens de serviço e o total de intervenções de manutenção, 19.572 e 406.533, respetivamente. Pode dizer-se que praticamente não existem ordens de serviço preventivas com apenas uma intervenção de manutenção.

A representar 16% do total de intervenções de manutenção, existem 2 localizações, nomeadamente, "circulação" (53.598 manutenções) e "estacionamento" (11.724 manutenções).

### **Horas Trabalhadas**

No que diz respeito às horas anuais trabalhadas, as manutenções preventivas corresponderam a um total de 36.043,69 horas o que equivale a uma média aproximada de 3.000 horas/mês e as manutenções corretivas corresponderam a um total de

5.441,67 horas, o que equivale a 453 horas/mês, de acordo com a tabela 2 representada abaixo.

Corretiva		Preventiva	
Horas de trabalho manutenção		Horas de trabalho manutenção	
Média mensal	453,47	Média mensal	3.003,64
Média semanal	104,64	Média semanal	693,15
Média diária	20,92	Média diária	138,63
TOTAL	5.441,67	TOTAL	36.043,69

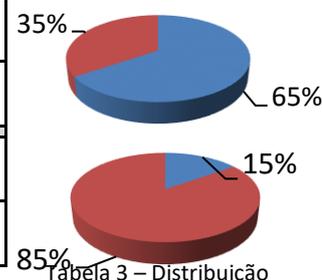
Tabela 2 – Distribuição das horas trabalhadas

### Custos de Manutenção

Quanto aos custos de manutenção, em termos de mão-de-obra destacam-se as manutenções preventivas com 483.261,30€ e relativamente aos materiais destacam-se as manutenções corretivas com 12.097,65€.

Como se pode ver pela tabela 3, os custos totais das manutenções preventivas são bastante superiores aos custos totais das manutenções corretivas, correspondendo a 85% contra 15%, respetivamente.

Custos	Valores (€)	Tipos de manutenção
Mão-de-obra	75.640,88	■ Corretiva
	483.261,30	■ Preventiva
Materiais	12.097,65	■ Corretiva
	6.372,41	■ Preventiva
Custos totais	87.738,53	■ Corretiva
	489.633,71	■ Preventiva



dos custos de manutenção

### 4.3. Escolha de Variáveis para o Modelo

Analisadas as 57 variáveis iniciais pertencentes às bases de dados apresentadas, foi necessário criar novas variáveis nomeadamente para aprofundar o estudo das já existentes.

Aquilo que se pretende é, através das variáveis existentes, perceber quais delas podem ajudar a melhor conseguir antecipar a necessidade de manutenção de falhas através do modelo, ou seja, quais as variáveis relevantes para o estudo pretendido.

Posto isto, foram analisadas várias variáveis explicativas, entre elas:

- Tempo médio de reparo da falha;
- Nº de ativos da mesma família;
- Nº de ações preventivas antes da falha;
- Desvio padrão dos tempos de reparação;
- Nº de técnicos em ações preventivas;
- Nº de técnicos em ações corretivas;
- Tempo desde a última falha;
- Tempo médio entre falhas preventivas;
- Se todos os ativos são do mesmo modelo e/ou fabricante.

Como para a grande maioria das variáveis que se tentou criar não se conseguiu dispor de toda a informação necessária para a criação das mesmas, apenas foram escolhidas cinco variáveis explicativas para o modelo, quatro das quais são novas variáveis, mais concretamente, médias de variáveis anteriormente estudadas. São elas, “N\_preventivas\_antes\_falha\_mean”, “Dias\_desde\_ultima\_preventiva\_mean”, “Dias\_entre\_preventivas\_mean”, “N\_falhas\_mesma\_localizacao\_mean” e “N\_ordens\_servico”.

É com base nestas cinco variáveis que, no capítulo seguinte será explicado o modelo construído e respetiva árvore de decisão.

## **CAPÍTULO 5: RESOLUÇÃO DO CASO PRÁTICO**

Este capítulo começa com uma breve introdução ao problema, para lembrar o que irá ser feito, seguida pela descrição do modelo aplicado. No ponto 5.3. é feita uma

breve referência aos requisitos de implementação e no ponto 5.4. encontra-se a explicação do algoritmo utilizado e de toda a sua metodologia de funcionamento. No ponto 5.5. é analisada a árvore de decisão que foi gerada através do modelo e, por fim, encontram-se os resultados obtidos, no ponto 5.6.

### 5.1. Introdução à Metodologia

O que se pretende é criar um modelo que permita, através do histórico de manutenções e intervenções, assim como das suas condições atuais, identificar indicadores de ocorrência de falha, por forma a antecipar as necessidades de manutenção preventiva.

A principal finalidade é identificar os fatores e respetivos níveis que podem indiciar uma falha e obter diariamente uma possibilidade de ocorrência para os ativos do hospital em causa.

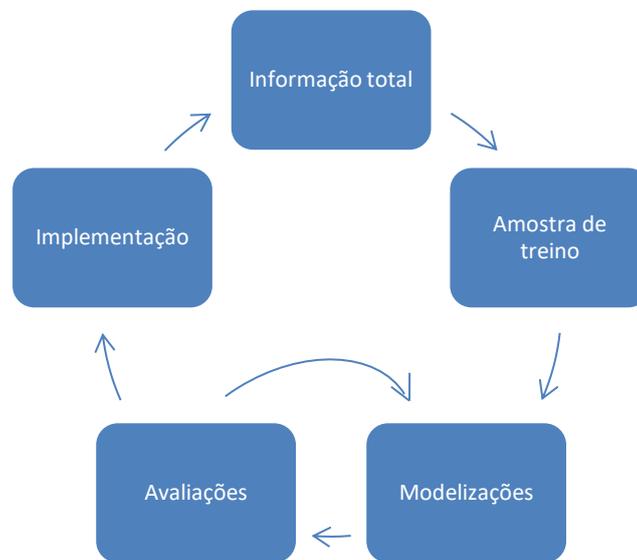


Figura 4 – Ciclo de vida do projeto

Como mostra a figura 4, através da informação total disponível sobre os ativos, foi selecionada uma amostra de treino que foi submetida a várias modelizações e avaliações por forma a encontrar a melhor para posterior implementação.

### 5.2. Descrição do Modelo

Com o auxílio do *software IBM SPSS Modeler*, criou-se o modelo que se encontra representado abaixo, na figura 5.

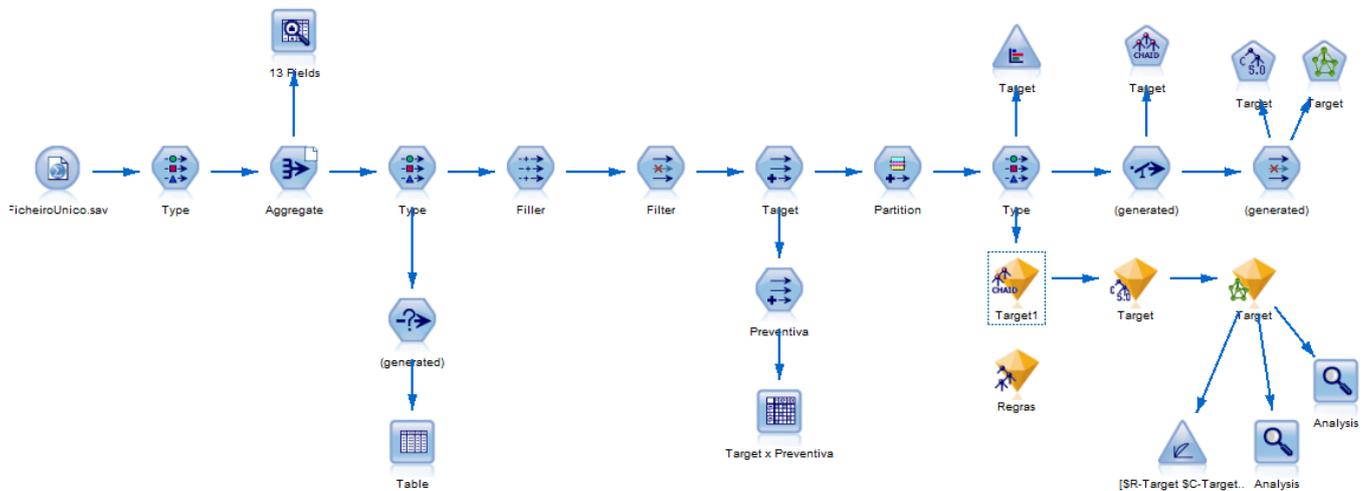


Figura 5 – Stream do modelo

Nesta *stream* foram criadas duas variáveis binárias, denominadas por “*target*” (alvo) e “*preventiva*”. Destas duas variáveis, a variável objetivo é a *target* que assume o valor 1 quando o equipamento está perante manutenções corretivas e o valor 0 em caso contrário (manutenções preventivas).

Estas duas variáveis binárias deram origem a uma tabela de contingência, onde se encontra representada a variável *target* em linha e a variável *preventiva* em coluna. Desta tabela resulta um  $\chi^2$  (qui-quadrado) igual a 104,316, com 1 grau de liberdade (df = 1). Quanto maior for o valor do qui-quadrado, menor será a probabilidade de as variáveis serem independentes.

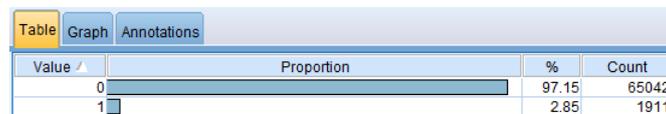
Esta tabela permite concluir que, de um total de 66.953 equipamentos, apenas 1.901 tiveram simultaneamente manutenções corretivas e preventivas, tal como mostra a figura 6.

Matrix Appearance Annotations			
Preventiva			
Target	0	1	Total
0	3998	61044	65042
1	10	1901	1911
Total	4008	62945	66953

Figura 6 – Tabela de contingência *target* x *preventiva*

Tendo em conta os resultados anteriores, a amostra em análise foi repartida através do nó de análise “*partition*”, que serve para facilitar a avaliação dos dados. Este nó serviu para dividir a amostra total em duas de 50%, sendo uma a amostra de treino e a outra amostra de teste. A amostra de treino vai servir para originar o modelo e depois será com a amostra de teste que se vai testar o mesmo. Este método de análise funciona tanto melhor quanto maior for a semelhança entre os dados das amostras. [13]

Esta repartição veio validar os resultados obtidos com a tabela de contingência da página anterior, onde cerca de 97% dos equipamentos corresponderam a manutenções preventivas (65.042 equipamentos) e os restantes 3% às manutenções corretivas (1.911 equipamentos), como mostra a figura seguinte.



Value	Proportion	%	Count
0	97.15	97.15	65042
1	2.85	2.85	1911

Figura 7 – Proporção de corretivas e de preventivas

Como estes resultados são extremamente desequilibrados para uma análise posterior, foi utilizado o nó “*balance*”. Este nó de análise serve precisamente para equilibrar amostras com resultados demasiado distintos pelo facto de haver técnicas de análise que tendem a ignorar parte da amostra quando esta é demasiado baixa. [14] Para o estudo em concreto, caso não fosse utilizado este nó, os 2,85% da amostra total, seriam desconsiderados para análises posteriores porque o sistema não os iria ter em consideração por considerar um valor irrelevante.

Posteriormente, com o auxílio do nó *Chi-squared Automatic Interaction Detection (CHAID)*, foi construída uma árvore de decisão.

### 5.3. Requisitos de Implementação

O objetivo da unidade hospitalar é criar um modelo preditivo que diariamente seja executado sobre informação atualizada proveniente do ambiente e sistema de informação *Enterprise Asset Management (EAM)*, com a finalidade de gerar a probabilidade de ocorrência de falha para cada equipamento.

Isto significa que em função das características técnicas do EAM, o modelo preditivo a implementar tem obrigatoriamente de ser passível de integração em

formato *standard* (PMML - *Predictive Model Markup Language* ou SQL - *Structured Query Language*).

Adicionalmente, o modelo tem o requisito fundamental de ser legível e gerível pelos utilizadores de negócio.

Originalmente, o algoritmo *CHAID* do *modeler* permite automaticamente transformar o modelo estimado em formatos *standard* PMML 4.0 ou SQL sendo desta forma o instrumento de estimação eleito.

### 5.5. Funcionamento e Regras do Algoritmo *CHAID*

As fórmulas e explicações referidas neste ponto do trabalho encontram-se presentes no “*IBM SPSS Modeler 16 Algorithms Guide*”, nas referências bibliográficas em [10].

Como é natural, existem algoritmos deste *software* que explicam a forma como os nós trabalham. Para perceber melhor todo este processo, será explicado o funcionamento do nó até à construção da árvore de decisão.

É necessário referir que existem vários processos de classificação que podem ser utilizados através do *software modeler* para fazer a análise pretendida às variáveis. Foi escolhido o algoritmo *CHAID* em detrimento de qualquer outro apenas pelo simples facto de, numa análise inicial, este ter sido aquele com o qual foram obtidos resultados mais favoráveis. Tendo em conta a complexidade de cada processo e a escassez de tempo, não foi possível realizar análises mais complexas para, eventualmente, tomar uma melhor decisão sobre os diferentes algoritmos do *software*.

#### Algoritmo *CHAID*

*CHAID* foi criado, em meados de 1980, por Kass [11]. Este algoritmo estatístico é uma técnica altamente eficiente em matéria de segmentação e construção de árvores de decisão.

Este algoritmo testa a significância estatística com o auxílio de um teste de independência qui-quadrado ( $\chi^2$ ) ou de um teste *F* de forma a identificar divisões óptimas. O que diferencia a escolha entre estes testes é a unidade da variável *target*. Neste caso como esta é categórica (quando o número de valores distintos dos dados é desconhecido) é utilizado um teste qui-quadrado, se fosse contínua (quando são

intervalos conhecidos de valores numéricos, que tanto podem ser números inteiros, números reais ou datas) seria utilizado um teste  $F$ .

Ao utilizar o teste de significância estatística como critério, o algoritmo *CHAID* avalia todas as variáveis em estudo como potenciais preditores (alvos). Aquelas que forem consideradas estatisticamente homogêneas (variáveis similares) em relação à variável alvo (*target*) são agrupadas e todas as outras variáveis (heterogêneas) são mantidas independentes.

Em seguida, é selecionado o melhor preditor relativamente à variável alvo para formar o primeiro ramo da árvore de decisão de modo a que, cada uma das ramificações forme um grupo de valores homogêneos do campo selecionado. Este processo continua de forma recursiva até que a árvore esteja totalmente formada.

Este processo é explicado de forma mais pormenorizada em “Nós de Divisão”.

### **Nós de Divisão**

Quando todas as variáveis se encontram devidamente agrupadas, significa que foram definidos os campos preditores para a construção da árvore de decisão. Cada um dos campos definidos é avaliado consoante o grau de associação que os mesmos têm com a variável alvo, *target*, uma vez que esta é a variável principal. Essa associação é feita com base no valor  $p$  ajustado, do teste estatístico de associação.

O preditor que tenha a associação mais forte (aquele que tiver valor  $p$  inferior) será o comparado com o limite de divisão  $\alpha_{split}$ . Caso o valor  $p$  seja inferior ou igual a  $\alpha_{split}$ , esse campo é o selecionado como campo de divisão para o nó atual da árvore. As categorias que forem agregadas são as que originam os “nós filhos” da divisão.

Feita a divisão ao nó atual, o processo é igual para os nós filhos até que uma ou mais regras de paragem sejam acionadas e nenhuma outra divisão possa ser feita.

Os cálculos diretamente envolvidos na construção do modelo vão ser descritos seguidamente.

### **Campo Alvo Contínuo (Teste F)**

Para modelos com um campo alvo conhecido, o valor  $p$  é calculado com base num padrão estandardizado ANOVA Teste F que compara o campo alvo com as várias categorias do campo preditor.

O teste F é baseado através da seguinte fórmula:

$$F = \frac{\sum_{i=1}^I \sum_{n \in D} w_n f_n I(x_n = i) (\bar{y}_i - \bar{y})^2 / (I - 1)}{\sum_{i=1}^I \sum_{n \in D} w_n f_n I(x_n = i) (y_n - \bar{y}_i)^2 / (N_f - I)}$$

e o valor  $p$  é dado por  $p = \Pr(F(I - 1, N_f - I) > F)$ , onde,

$$\bar{y}_i = \frac{\sum_{n \in D} w_n f_n y_n I(x_n = i)}{\sum_{n \in D} w_n f_n I(x_n = i)}, \bar{y} = \frac{\sum_{n \in D} w_n f_n y_n}{\sum_{n \in D} w_n f_n}, N_f = \sum_{n \in D} f_n$$

e  $F(I - 1, N_f - I)$  é uma variável aleatória que segue uma distribuição F com  $(I - 1)$  e  $(N_f - I)$  graus de liberdade.

### **Campo Alvo Nominal (Teste Qui-quadrado)**

Durante a fase da separação da amostra, todas as variáveis são consideradas no cálculo do valor  $p$ , para que os dados sejam utilizados na sua totalidade.

Para um exemplo, em que o campo *target*,  $Y$ , é definido como categórico, será testada a hipótese nula de independência de  $X$  e  $Y$ . Para o efeito, é utilizada uma tabela de contingência (contagem) com as classes de  $Y$  em coluna e as classes preditivas de  $X$  em linha.

As frequências observadas e esperadas são utilizadas para aplicação do qui-quadrado e o valor  $p$  é definido com base na estatística calculada.

### **Teste de Pearson (Qui-quadrado)**

Este teste é baseado na seguinte fórmula:

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}$$

onde  $n_{ij} = \sum_n f_n I(x_n = i) \wedge (y_n = j)$  são as frequências observadas e  $\hat{m}_{ij}$  são as frequências esperadas de cada linha ( $x_n = i, y_n = j$ ) para o modelo de teste de independência descrito abaixo. O valor  $p$  correspondente é calculado por  $p =$

$\Pr(\chi_d^2 > \chi^2)$ , onde  $\chi_d^2$  segue uma distribuição qui-quadrado com  $d = (J - 1)(I - 1)$  graus de liberdade.

### Teste da Máxima Verosimilhança (Likelihood-ratio)

Este teste é aplicado tendo por base as frequências esperadas e as observadas:

$$G^2 = 2 \sum_{j=1}^J \sum_{i=1}^I n_{ij} \ln(n_{ij}/\hat{m}_{ij})$$

onde o valor  $p$  é calculado como  $p = \Pr(\chi_d^2 > G^2)$ .

### Teste Qui-quadrado para Frequências Esperadas

Quando o modelo não tem o peso de cada observação, calcula-se:  $\hat{m}_{ij} = \frac{n_i n_j}{n..}$

onde,

$$n_i = \sum_{j=1}^J n_{ij}, n_j = \sum_{i=1}^I n_{ij}, n.. = \sum_{j=1}^J \sum_{i=1}^I n_{ij}.$$

Quando o peso de cada observação é especificado, a frequência esperada para a hipótese nula de independência apresenta-se sob a forma:

$$m_{ij} = \bar{w}_{ij}^{-1} \alpha_i \beta_j$$

onde,  $\alpha_i$  e  $\beta_j$  são parâmetros a estimar, e

$$\bar{w}_{ij} = \frac{w_{ij}}{n_{ij}}, w_{ij} = \sum_{n \in D} w_n f_n I(x = i \wedge y_n = j).$$

As estimativas dos parâmetros,  $\hat{\alpha}_i$  e  $\hat{\beta}_j$ , assim como  $\hat{m}_{ij}$  são calculadas tendo como base o seguinte método:

- 1) Inicialização,  $k = 0$ ,  $\alpha_i^{(0)} = \beta_j^{(0)} = 1, m_{ij}^{(0)} = \bar{w}_{ij}^{-1}$
- 2)  $\alpha_i^{(k+1)} = \frac{n_i}{\sum_j w_{ij}^{-1} \beta_j^{(k)}} = \alpha_i^{(k)} \frac{n_i}{\sum_j m_{ij}^{(k)}}$
- 3)  $\beta_j^{(k+1)} = \frac{n_j}{\sum_i w_{ij}^{-1} \alpha_i^{(k+1)}}$
- 4)  $m_{ij}^{(k+1)} = \bar{w}_{ij}^{-1} \alpha_i^{(k+1)} \beta_j^{(k+1)}$
- 5) Se,  $\max_{i,j} |m_{ij}^{(k+1)} - m_{ij}^{(k)}| < \epsilon$ , verifica-se a condição de paragem e o *output* é  $\alpha_i^{(k+1)}, \beta_j^{(k+1)}$  e  $m_{ij}^{(k+1)}$  como estimativa final de  $\hat{\alpha}_i, \hat{\beta}_j$  e  $\hat{m}_{ij}$ . Caso contrário, incrementa-se  $K$  e repete-se o processo desde o ponto 2.

### **Campo *Target* (Alvo) – Modelo de Efeitos em Linha**

Se o campo *target*  $Y$  for ordinal, a hipótese nula de independência de  $X$  e  $Y$  é testada pelo modelo de efeitos em linha (*row effects model*), sendo as linhas as categorias de  $X$  e as colunas as de  $Y$ . Assim, são estimados os dois conjuntos de frequências esperadas,  $\widehat{m}_{ij}$  (sob a hipótese de independência  $\widehat{m}_{ij}$ ).

A estatística do rácio da verosimilhança é calculada como,

$$H^2 = 2 \sum_{i=1}^I \sum_{j=1}^J \widehat{m}_{ij} \ln \left( \widehat{m}_{ij} / \widehat{m}_{ij} \right)$$

e o valor  $p$  é calculado da seguinte forma,  $p = \Pr(\chi_{I-1}^2 > H^2)$ .

### **Ajuste de Bonferroni**

O valor  $p$  ajustado é calculado como o valor  $p$  a multiplicar por um multiplicador de *Bonferroni*. Este multiplicador é um fator de correção que controla o valor  $p$  em diversos testes estatísticos e é utilizado, precisamente, devido aos sucessivos testes de comparações que são efetuados pelo algoritmo *CHAID*.

Supondo que um campo preditivo tem  $I$  categorias e é reduzido a  $r$  categorias depois de ser feita a junção dos campos. O multiplicador de *Bonferroni*,  $B$ , é o número de modos possíveis para as  $I$  categorias serem unidas em  $r$  categorias.

Para  $r = I, B = 1$ . Para  $2 \leq r < I$ , tem-se:

$$B = \begin{cases} \binom{I-1}{r-1} & \text{Preditores ordinal} \\ \sum_{v=0}^{r-1} (-1)^v \frac{(r-v)^I}{v!(r-v)!} & \text{Preditores nominal} \\ \binom{I-2}{r-2} + r \binom{I-2}{r-1} & \text{Ordinal com valores em falta} \end{cases}$$

### **Manipulação em Branco (*Blank Handling*)**

Se existirem campos de destino em branco, se todos os campos preditores estiverem em branco, se os campos forem especificados mas ainda assim estiverem em branco, iguais a zero ou com valores negativos, esses registos são ignorados.

Preditor ordinal – o algoritmo gera o melhor conjunto de categorias com toda a informação existente. Posteriormente, identifica a categoria mais semelhante à categoria em branco. Em seguida, são calculados dois valores  $p$ , um para o conjunto de categorias que nasceu da junção entre os campos em branco e aqueles que são mais similares e outro para o conjunto de categorias que se formou pela adição da categoria em branco com uma categoria em separado. Dos dois, será válido o menor valor  $p$ .

Preditor nominal – as categorias em branco são tratadas do mesmo modo que aquelas que estão em análise.

### **Regras de Paragem (Stopping Rules)**

Estas regras controlam a maneira como o algoritmo *CHAID* decide quando parar de dividir os nós da árvore. O crescimento da árvore prossegue até que cada folha do nó da árvore desencadeie pelo menos uma regra de paragem. Qualquer uma das seguintes condições impedem que um nó seja dividido:

- O nó é puro (todos os registos têm o mesmo valor para o campo de destino);
- Todos os registos do nó têm o mesmo valor para os campos preditores utilizados pelo modelo;
- A profundidade da árvore para o nó atual é a profundidade máxima da árvore (por defeito ou especificação do utilizador);
- O número de registos no nó é menor que o tamanho mínimo do nó principal (por defeito ou especificação do utilizador);
- O número de registos em qualquer um dos nós filhos, resultante da melhor separação do nó é menor do que o tamanho mínimo do nó filho (por defeito ou especificação do utilizador);
- A melhor divisão para o nó produz um valor  $p$  que é maior do que o valor  $\alpha_{\text{dividido}}$  (por defeito ou especificação do utilizador).

### **Riscos de Estimação**

Estas estimativas descrevem o risco de erro nos valores previstos para os nós da árvore e para a árvore como um todo.

Para árvores de classificação (com campo alvo simbólico), a estimativa de risco  $r(t)$  de um nó  $t$  é calculada como,

$$r(t) = \frac{1}{N_f} \sum_j N_{f,j}(t) C(j \times (t)|j)$$

onde,  $C(j \times (t)|j)$  é o custo de classificar de forma errada um registo com valor alvo  $j$ , quando  $j \times (t)$ ,  $N_{f,j}(t)$  é a soma dos pesos de frequência para os registos no nó  $t$  na categoria  $j$  (ou o número de registos se nenhum peso de frequência for definido), e  $N_f$  é a soma do peso de frequência para todos os registos nos dados de treino.

É de realçar que os pesos não são considerados no cálculo das estimativas de risco.

Para árvores de regressão (com campo alvo numérico), a estimativa de risco  $r(t)$  de um nó  $t$  é calculada como,

$$r(t) = \frac{1}{N_f(t)} \sum_{i \in t} f_i (y_i - \bar{y}(t))^2$$

onde,  $f_i$  é o peso de frequência para registo  $i$  (registo atribuído ao nó  $t$ ),  $y_i$  é o valor do campo destino para o registo  $i$  e  $\bar{y}(t)$  é a média ponderada do campo alvo para todos os registos do nó  $t$ .

Em ambos os casos, árvores de classificação ou de regressão, a estimativa de risco  $r(t)$  para a árvore ( $T$ ) é calculada através da soma das estimativas de risco dos nós terminais  $r(t)$ :

$$R(T) = \sum_{t \in T'} r(t)$$

onde,  $T'$  é o conjunto de nós terminais da árvore.

Para o caso em concreto, é o risco de estimação para árvores de regressão que é relevante.

### Modelo Gerado - Árvores de Regressão

Para estes casos, os valores apresentados nos nós terminais dizem respeito a médias ponderadas do valor da categoria *target* (campo numérico).

Média ponderada que se calcula através de:  $\bar{y}(t) = \frac{1}{N_w(t)} \sum_{i \in t} w_i f_i y_i$ , onde  $N_w(t)$  é definido como  $N_w(t) = \sum_{i \in t} w_i f_i$ .

## 5.6. Análise da Árvore de Decisão

Após análise ao algoritmo *CHAID*, resta perceber o funcionamento da árvore de decisão apresentada pelo modelo.

Uma árvore de decisão é um importante meio de suporte para ajudar as empresas a tomar decisões, principalmente a nível financeiro e de vendas. Estes diagramas conseguem especificar todas as sequências possíveis de decisão lógica assim como as possibilidades mais instáveis ao longo de um projeto. A utilização deste método tem várias vantagens, entre elas o facto de ser relativamente simples de compreender todo o seu processo a nível de raciocínio. [9]

Tal como foi referido no ponto anterior do trabalho, tendo em consideração a variável *target*, este algoritmo avalia as variáveis preditivas e une as que são homogéneas (semelhantes) e mantém as heterogéneas (diferentes). Depois, de entre as variáveis preditivas, escolhe a que melhor representa a amostra de dados, a mais homogénea, e cria o primeiro ramo da árvore. Os ramos “filhos” desse ramo que foi criado, serão amostras dessa variável preditiva com valores ainda mais homogéneos. Este processo ocorre de forma repetitiva até a árvore estar totalmente formada. [10]

Assim, para a construção da árvore foram selecionadas a variável alvo, *target*, e cinco variáveis de entrada para análise probabilística, que são: “N\_preventivas\_antes\_falha\_mean”, “Dias\_desde\_ultima\_preventiva\_mean”, “Dias\_entre\_preventivas\_mean”, “N\_falhas\_mesma\_localizacao\_mean” e “N\_ordens\_servico”.

Desta forma, o primeiro troço da árvore de decisão corresponde à variável “N\_ordens\_servico”, pelo facto de ter sido considerada a mais preditiva e a que melhor explica a variável *target*, tal como se pode ver na figura 8, representada na página seguinte.

A ordem pela qual as variáveis se encontram dispostas na árvore vai de encontro a este grau de importância.

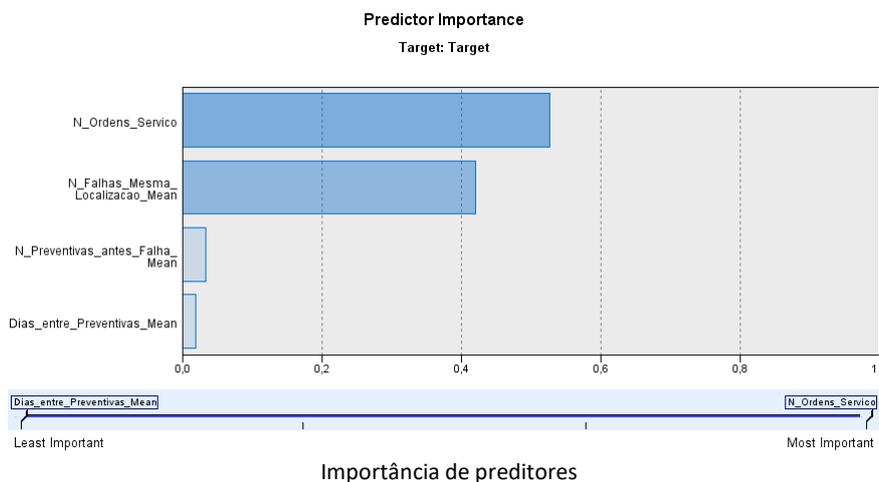


Figura 8

Posto isto, a figura 9, representa a variável *target*, onde se pode ver que o valor p ajustado é zero (quanto menor for o valor p, maior é a significância do mesmo) o que significa que a hipótese é significativa, com um  $\chi^2$  igual a 11.831,987 e 1 grau de liberdade. Quanto maior for o valor do qui-quadrado, menor será a probabilidade de as variáveis serem independentes.

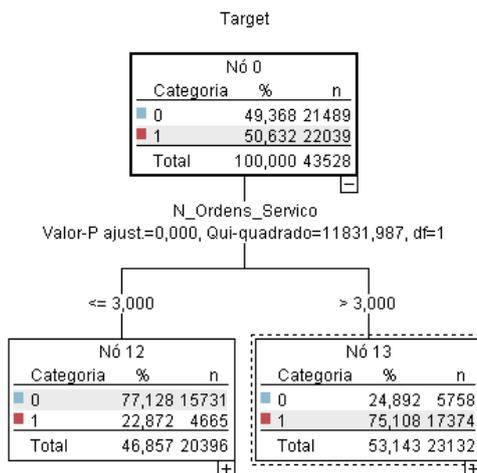


Figura 9 – Primeiro troço da árvore de decisão

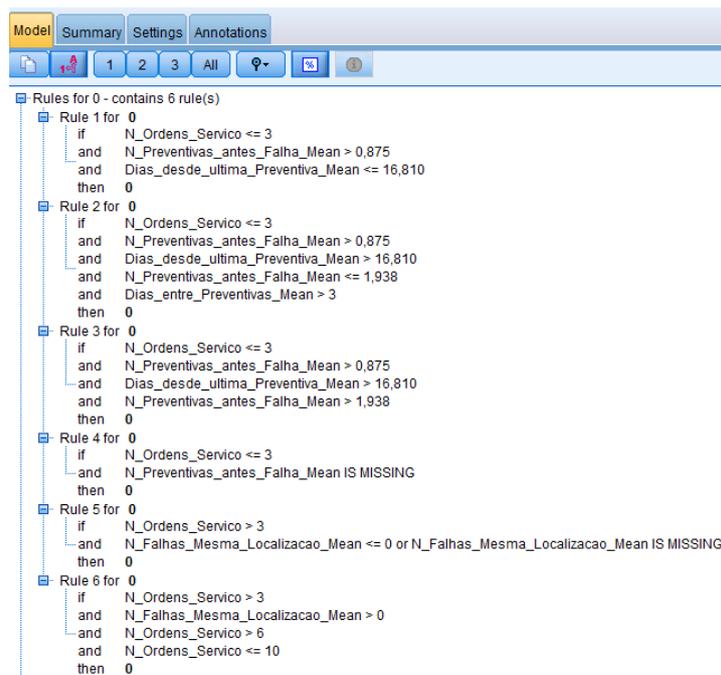
Cerca de 51% da amostra total ponderada de ativos, equivalente a 22.039 equipamentos (representado a vermelho na legenda), referem-se a manutenções corretivas e os 49% que representam 21.489 equipamentos (representado a azul na

legenda), dizem respeito a manutenções preventivas. Nos troços seguintes da árvore, a leitura a fazer será equivalente à do troço inicial.

No “Anexo A – Árvore de Decisão” encontra-se a figura da árvore de decisão totalmente construída, com um total de cinco ramificações.

### Regras de Análise

Para a construção da árvore de decisão foi necessário definir à *posteriori* determinadas regras para a análise de probabilidades, nomeadamente, as apresentadas na figura 10.



```

Model Summary Settings Annotations
1 2 3 All
Rules for 0 - contains 6 rule(s)
Rule 1 for 0
  if N_Ordens_Servico <= 3
  and N_Preventivas_antes_Falha_Mean > 0,875
  and Dias_desde_ultima_Preventiva_Mean <= 16,810
  then 0
Rule 2 for 0
  if N_Ordens_Servico <= 3
  and N_Preventivas_antes_Falha_Mean > 0,875
  and Dias_desde_ultima_Preventiva_Mean > 16,810
  and N_Preventivas_antes_Falha_Mean <= 1,938
  and Dias_entre_Preventivas_Mean > 3
  then 0
Rule 3 for 0
  if N_Ordens_Servico <= 3
  and N_Preventivas_antes_Falha_Mean > 0,875
  and Dias_desde_ultima_Preventiva_Mean > 16,810
  and N_Preventivas_antes_Falha_Mean > 1,938
  then 0
Rule 4 for 0
  if N_Ordens_Servico <= 3
  and N_Preventivas_antes_Falha_Mean IS MISSING
  then 0
Rule 5 for 0
  if N_Ordens_Servico > 3
  and N_Falhas_Mesma_Localizacao_Mean <= 0 or N_Falhas_Mesma_Localizacao_Mean IS MISSING
  then 0
Rule 6 for 0
  if N_Ordens_Servico > 3
  and N_Falhas_Mesma_Localizacao_Mean > 0
  and N_Ordens_Servico > 6
  and N_Ordens_Servico <= 10
  then 0
  
```

Figura 10 – Regras

para o critério 0

[mode 0]

Estas seis regras dizem respeito ao critério 0, ou seja, às ramificações que resultam em probabilidade a favor de manutenções preventivas.

A primeira regra diz que se a variável “N\_ordens\_servico” for inferior ou igual a 3, o “N\_preventivas\_antes\_falha\_mean” for superior a 0,875 e “Dias\_desde\_ultima\_preventiva\_mean” for inferior ou igual a 16,810, então a probabilidade será a favor do critério 0, intervenção de manutenção preventiva. E assim sucessivamente para a leitura das restantes regras.

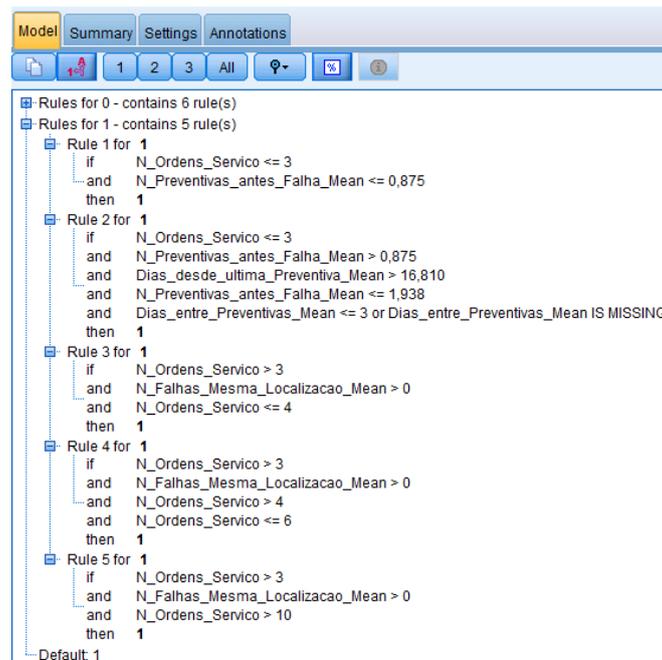


Figura 11 – Regras

para o critério 1

[mode 1]

Por outro lado, temos agora a figura 11 com as cinco regras que dizem respeito ao critério 1, ou seja, às ramificações que resultam em probabilidade a favor de manutenções corretivas.

A primeira regra diz que se a variável “N\_ordens\_servico” for inferior ou igual a 3 e o “N\_preventivas\_antes\_falha\_mean” for inferior ou igual a 0,875, então a probabilidade será a favor do critério 1, intervenção de manutenção corretiva. E assim sucessivamente para a leitura das restantes regras.

### 5.7. Resultados Obtidos

Para a análise de resultados há que relembrar que a amostra total das manutenções de equipamentos foi dividida em duas amostras equivalentes, a amostra de treino (50%) e a amostra de teste (50%). Com a amostra de treino gerou-se o modelo analisado e com a amostra de teste foram testados resultados.

Estes resultados obtidos dizem respeito ao melhor modelo encontrado (\$R-Target). No “Anexo B – Modelos Testados”, estão representados, para além deste, os outros modelos que foram gerados e testados.

Os gráficos apresentados abaixo, na figura 12, comparam o modelo preditivo,  $\$R$ -Target, com a variável objetivo, Target, que se encontra representada pela reta vermelha.

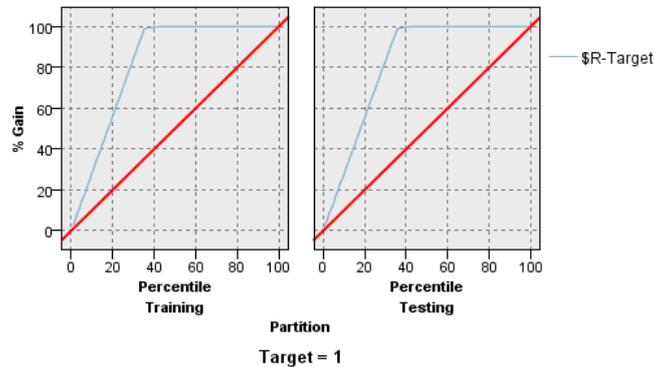


Figura 12 – % *Gain* do modelo  $\$R$ -Target

A tendência da amostra de treino é bastante semelhante à da amostra de teste, o que significa não só que os dados das duas amostras são semelhantes, como que os modelos preditivos têm comportamentos muito idênticos. Estes gráficos sugerem que os resultados são fiáveis.

Ao analisar os gráficos pode concluir-se que os 20% com melhor classificação (percentil 20) classificam corretamente cerca de 60% dos casos em estudo. Assim como, os 40% (percentil 40) melhor classificados correspondem a uma classificação correta de cerca 100% dos casos em estudo, em ambas as amostras. Logo, a proporção de casos de sucesso no modelo  $\$R$ -Target é total a partir do percentil 40. Quanto mais íngreme for esta curva, maior será o ganho.

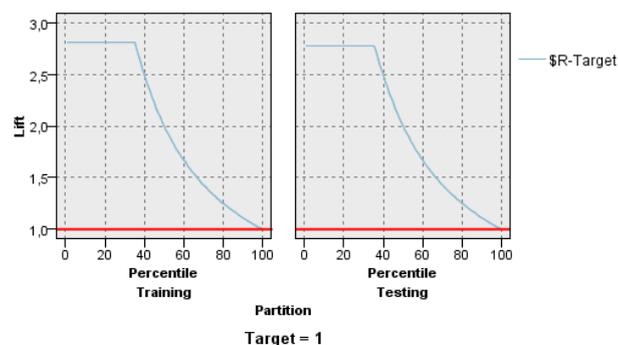


Figura 13 – *Lift* do modelo  $\$R$ -Target

Quanto aos gráficos referentes a *lift*, pode concluir-se que, cerca de 37% (percentil 37) dos casos mais bem classificados têm um *lift* de 2,75. A partir dos 40% até aos 100%, a tendência é de aproximação do valor 1.

Results for output field Target

Comparing \$R-Target with Target

'Partition'	1_Training		2_Testing	
Correct	21.901	65,74%	21.996	65,39%
Wrong	11.413	34,26%	11.643	34,61%
Total	33.314		33.639	

Figura 14 – Resultados obtidos

Em termos gerais, ao fazer a comparação entre a amostra de treino e a amostra de teste, a percentagem de concordâncias é similar e explica cerca de 65% dos casos em estudo, o que significa que se consegue prever a ocorrência de falha em 65% dos equipamentos, através do modelo criado, \$R-Target.

## **CAPÍTULO 6: CONCLUSÃO**

Tendo em conta o modelo e resultados que foram apresentados e analisados, pode dizer-se que o objetivo deste estágio foi concretizado.

Mediante a informação disponibilizada pelo cliente da PSE, conseguiu-se criar um modelo que conjuga condições intrínsecas dos equipamentos com aquelas que pertencem ao seu histórico de manutenção, assim como as suas condições atuais. Este modelo consegue antecipar, de forma diária, a ocorrência de falha de determinados equipamentos do hospital. Tal como se pode concluir nos resultados obtidos, cerca de 65% dos equipamentos que foram estudados são passíveis de antecipação de ocorrência de falhas.

Desta forma, a unidade hospitalar irá conseguir certamente reduzir os custos mensais relativos a manutenções, diminuir as horas de trabalho dos técnicos e aumentar o tempo de disponibilidade e bom funcionamento de grande parte dos seus equipamentos.

Em termos de gestão interna, qualquer empresa, principalmente no ramo industrial (com grande número de equipamentos/maquinarias), beneficiaria em ter um modelo preditivo com estas características.

Em suma, modelos que apresentem soluções benéficas tendem sobretudo a aumentar a produtividade nas organizações, assim como a conseguir fornecer um melhor apoio na tomada de decisão.

## REFERÊNCIAS BIBLIOGRÁFICAS

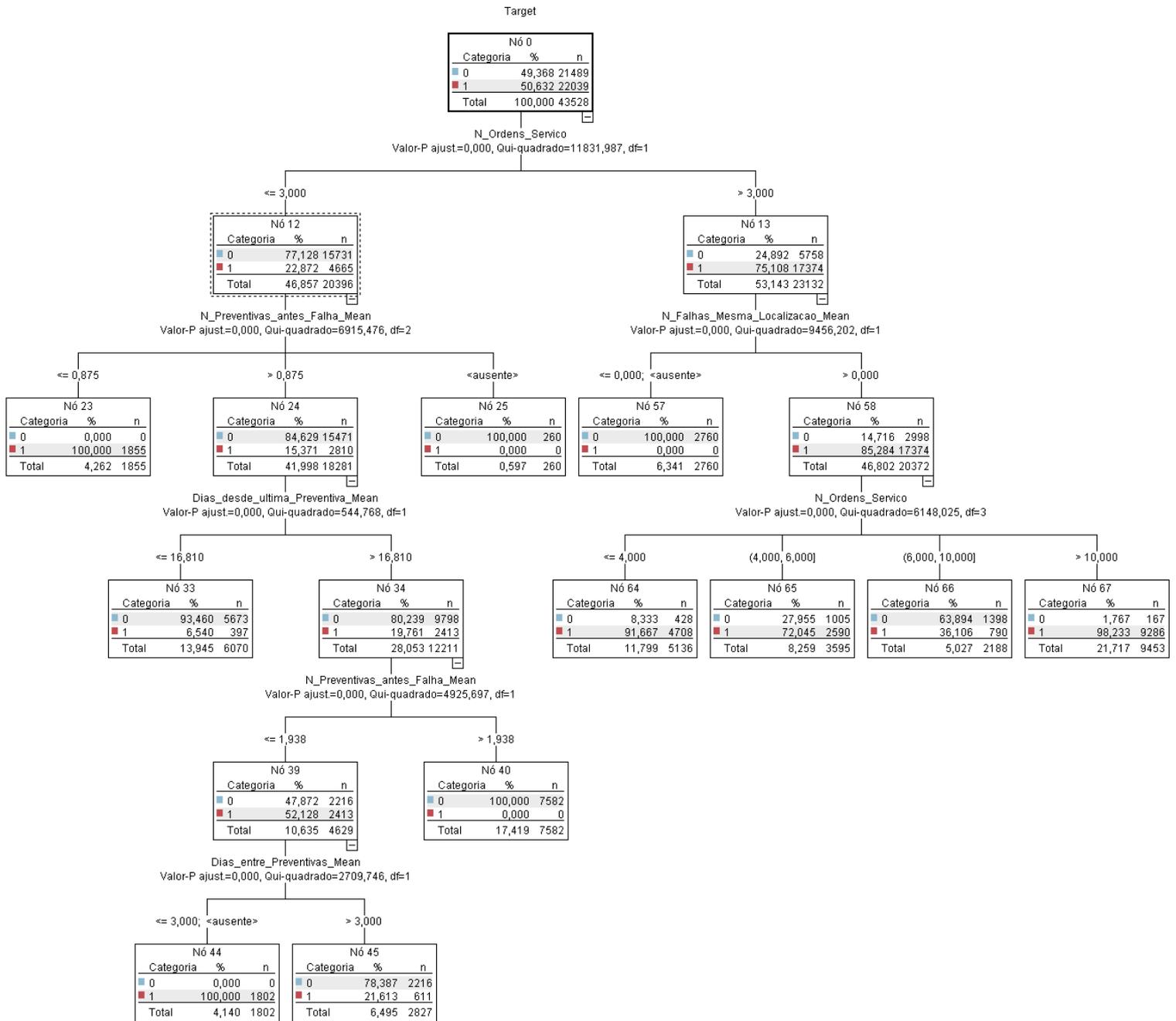
- [1] Almeida, D.A. de, Leal, F., Pinho, A.F. e Fagundes, L.D., “Gestão do Conhecimento na Análise de Falhas: Mapeamento de Falhas através de Sistema de Informação”, Production, Brasil, vol. 16.1, pp 171-188, 2006.
- [2] Assis, R., “Apoio à Decisão em Manutenção na Gestão de Ativos Físicos”, 2ª Edição Atualizada, LIDEL – Edições Técnicas, Lda., Lisboa, Cap. 1, 2, 5 e 8, 2014.
- [3] Campozano, L., Sánchez, E., Aviles, A. e Samaniego, E., “Evaluation of Infilling Methods for Time Series of Daily Precipitation and Temperature: The Case of the Ecuadorian Andes”, Maskana, vol. 5.1, pp 99-115, 2015.
- [4] Cardoso, O.N.P. e Machado, R.T.M., “Gestão do Conhecimento usando Data Mining: Estudo de Caso na Universidade Federal de Lavras”, Revista de Administração Pública, Rio de Janeiro, vol. 42.3, pp 495-528, 2008.
- [5] Casimiro, M.P.G., "Análisis de Series Temporales: Modelos ARIMA", Departamento de Economía Aplicada III (Econometría y Estadística), Facultad de Ciencias Económicas y Empresariales, Universidad del País Vasco, 2009.
- [6] Henriques, C., “Análise de Regressão Linear Simples e Múltipla”, Documento do Departamento de Matemática, Escola Superior de Tecnologia de Viseu, 2010/2011.
- [7] Hovde, P.J., “Evaluation of the Factor Method to Estimate the Service Life of Building Components”, CIB World Building Congress, Gaevle, Junho, 1998.
- [8] IBM – International Business Machines (2017). IBM. Disponível em: <https://www.ibm.com/br-pt/marketplace/spss-modeler> [acedido em 24/07/2017]
- [9] “IBM SPSS Modeler 17 Modeling Nodes”. IBM. Chapter 6 – “Decision Trees”. Disponível em:  
<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/17.0/en/ModelerModelingNodes.pdf> [acedido em 31/08/2017]
- [10] “IBM SPSS Modeler 16 Algorithms Guide”. IBM. Chapter 10 – “CHAID Algorithms”. Disponível em:  
<ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/16.0/en/AlgorithmsGuide.pdf> [acedido em 24/08/2017]
- [11] Kass, G., “An Exploratory Technique for Investigating Large Quantities of Categorical Data”, Applied Statistics, South Africa, vol. 29.2, pp 119–127, 1980.

- [12] Madureira, E.M.G.V. de, “Análise de Mercado: Clustering”, Trabalho Final de Mestrado em Decisão Económica e Empresarial, Instituto Superior de Economia e Gestão, de Lisboa, 2016.
- [13] Menú Help do Software Modeler – Nó “Partition”. Disponível em:  
[http://127.0.0.1:52066/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/partition\\_overview.htm](http://127.0.0.1:52066/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/partition_overview.htm) [acedido em 20/09/2017]
- [14] Menú Help do Software Modeler – Nó “Balance”. Disponível em:  
[http://127.0.0.1:52066/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/balance\\_overview.htm](http://127.0.0.1:52066/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/balance_overview.htm) [acedido em 19/09/2017]
- [15] Montgomery, D.C., Jennings, C.L. e Kulahci, M., “Introduction to Time Series Analysis and Forecasting”, 2<sup>nd</sup> Edition, John Wiley & Sons, Inc., New Jersey, 2008.
- [16] Moubray, J., “Reliability - Centered Maintenance”, 2<sup>nd</sup> Edition, Industrial Press Inc., New York, pp 1-18, 1997.
- [17] Otani, M. e Waltair V.M., "A Proposta de Desenvolvimento de Gestão da Manutenção Industrial na busca da Excelência ou Classe Mundial", Revista Gestão Industrial, Brasil, vol. 4.2, pp 1-16, 2008.
- [18] Peña, D., Tiao, G. C. e Tsay, R. S., “A Course in Time Series Analysis”, John Wiley & Sons, Inc., New York, 2001.
- [19] PSE – Produtos e Serviços de Estatística (2017). PSE. Disponível em:  
<http://www.pse.pt/pse/> [acedido em 19/07/2017]
- [20] PSE – Produtos e Serviços de Estatística (2017). PSE. Disponível em:  
<http://www.pse.pt/metodologia/> [acedido em 20/07/2017]
- [21] PSE – Produtos e Serviços de Estatística (2017). PSE. Disponível em:  
<http://www.pse.pt/ibm-spss-statistics-2/> [acedido em 20/07/2017]
- [22] Santos, M.R.P., "Metodologias de Previsão da Vida Útil de Materiais, Sistemas ou Componentes da Construção: Revisão Bibliográfica", Dissertação em Engenharia Civil – Especialização em Construções, Faculdade de Engenharia Universidade do Porto, 2010.
- [23] Silva, D.F.F. da, “Detecção e Previsão de Falhas de Equipamentos de Produção Industrial”, Associação de Politécnicos do Norte (APNOR), Instituto Politécnico do Porto, 2013.

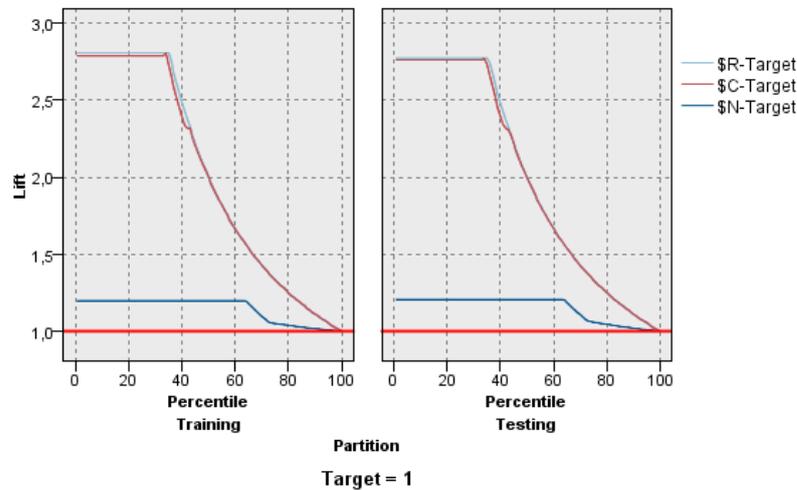
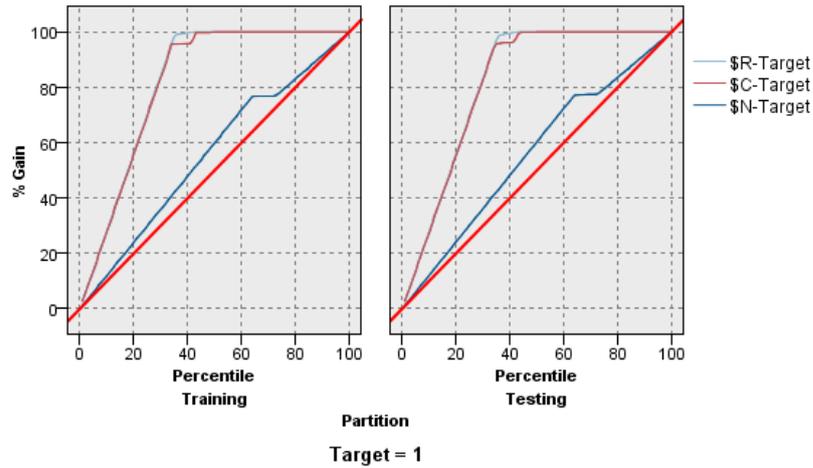
- [24] Sousa, R.D.B., “Previsão da Vida Útil dos Revestimentos Cerâmicos Aderentes em Fachada”, Dissertação de Mestrado em Engenharia Civil, Instituto Superior Técnico, Lisboa, 2008.
- [25] Taylor, J.W., "Exponential Smoothing with a Damped Multiplicative Trend", International Journal of Forecasting, Oxford, vol. 19, pp 715-725, 2003.
- [26] Waeyenbergh, G. e Pintelon, L., “A Framework for Maintenance concept Development”, International Journal of Production Economics, Belgium, vol. 77.3, pp 299-313, 2002.
- [27] Wireman, T., “Developing Performance Indicators for Managing Maintenance”, Industrial Press Inc., New York, Chapter 1, 2005.
- [28] Zhao, J., Xu, L. e Liu, L., “Equipment Fault Forecasting Based on ARMA Model”, Proceedings of the IEEE Conference on International Conference on Mechatronics and Automation, Harbin, China, 2007.
- [29] “5 – Cadeia de Markov”, Site Portal Action. Disponível em: <http://www.portalaction.com.br/processo-estocastico/cadeia-de-markov> [acedido em 09/10/2017].

ANEXOS

Anexo A – Árvore de Decisão



Anexo B – Modelos Testados



Results for output field Target

Individual Models

Comparing \$R-Target with Target

'Partition'	1_Training		2_Testing	
Correct	21.901	65,74%	21.996	65,39%
Wrong	11.413	34,26%	11.643	34,61%
Total	33.314		33.639	

Comparing \$C-Target with Target

'Partition'	1_Training		2_Testing	
Correct	19.972	59,95%	20.024	59,53%
Wrong	13.342	40,05%	13.615	40,47%
Total	33.314		33.639	

Comparing \$N-Target with Target

'Partition'	1_Training		2_Testing	
Correct	892	2,68%	882	2,62%
Wrong	32.422	97,32%	32.757	97,38%
Total	33.314		33.639	