



UNIVERSIDADE CATÓLICA PORTUGUESA

# Modelos de Classificação e Definição de Montante de Crédito de Clientes

Caso do Grupo Nors

Trabalho Final na modalidade de Relatório de Estágio  
apresentado à Universidade Católica Portuguesa  
para obtenção do grau de mestre em Gestão

por

Pedro Carvalho Neto Gabriel Silva

sob orientação de  
Prof. Dra. Rita Moura Bastos de Almeida Ribeiro  
Católica Porto Business School

Março, 2017



Aos meus Pais e irmãs



# Agradecimentos

Aos meus pais, pelos seus conselhos e pelo exemplo que me procuram sempre transmitir em termos de humildade, integridade e constante vontade em aprender.

À minha orientadora, Professora Rita Ribeiro, pela sabedoria, disponibilidade e experiência com que conduziu o desenvolvimento do meu trabalho.

Ao Dr<sup>o</sup> Daniel Silva, da Nors, pela sua disponibilidade, ajuda e conselhos que muito contribuíram para o desenvolvimento deste trabalho.

À Dr<sup>a</sup> Cláudia Araújo, responsável da Nors pelo estágio, que permitiu uma abordagem focando áreas de conhecimento novas na empresa e me deu liberdade para estruturar este trabalho de forma adequada ao estudo realizado.

À Professora Ana Camanho, minha vizinha e docente noutra instituição universitária, pelos seus experientes conselhos, em resultado de trocas de impressões que gentilmente teve comigo.

Ao Professor Luís Pacheco, da Católica Porto Business School, pela sua disponibilidade para diversas opiniões sobre assuntos abordados no trabalho.

Às pessoas que trabalham no Norshare pela sua simpatia e boa disposição, o que permitiu que estes meses fossem muito agradáveis.

Por último, aos meus amigos que me têm acompanhado desde o início do meu percurso académico.



# Resumo

A análise financeira é um fator preponderante na atribuição de crédito, por parte das instituições financeiras. Os modelos de classificação de crédito surgem como forma de facilitar a análise financeira, ao classificar os clientes em categorias de pagador. A aplicação destes modelos tem sido uma prática corrente em instituições financeiras, porque permitem poupar tempo e recursos.

O Grupo Nors concede crédito aos clientes, como forma de fomentar a sua faturação. Assim, a área das Contas a Receber (CaR), dos Serviços Partilhados, está encarregue da análise de crédito dos clientes e tomar as decisões de conceder ou não crédito e que montante disponibilizar ao cliente.

Assim, o presente trabalho tem como objetivo construir e propôr um modelo de classificação de crédito que seja útil à tomada de decisão de crédito da área das CaR e uma ferramenta que indique o montante de crédito a disponibilizar aos clientes que solicitam crédito. Desta forma, foram criados vários modelos de classificação de crédito diferentes, construídos a partir de três métodos distintos, Análise Múltipla Discriminante, Regressão Logística e Redes Neurais Artificiais. Também, foi criado um modelo que indica o montante de crédito a conceder a cada cliente, baseado em algoritmos genéticos.

Constatou-se que as Redes Neurais Artificiais são o método de classificação com melhores resultados para o caso de classificação de crédito do Grupo Nors e que o modelo de definição do montante de crédito de clientes pode constituir uma ferramenta útil para o Grupo, uma vez que os indicadores correntemente utilizados não estão orientados especificamente para esse efeito.

Palavras-chave: Modelos de classificação de crédito; redes neuronais artificiais; regressão logística; análise discriminante; algoritmos genéticos; “*plafond* de crédito” de clientes



# Abstract

Financial analysis is an important feature concerning credit assignment, for financial institutions. Credit Scoring models arise as a way to facilitate the financial analysis needed before conceding credit, as they classify clients in different categories. Financial institutions have been applying these models since they allow them to save time and resources.

Nors Group concedes credit to customers in order to increase sales. So, the department "Contas a Receber"(CaR), a receiving accounts area of the Shared Services of Nors is responsible for customer's credit analysis and for deciding whether or not to concede credit, as well as to define the amount of credit to concede to a customer that asks for credit.

Thus, this work aims to develop and propose a credit scoring model useful to CaR area and a tool to indicate the amount of credit to concede.

Therefore, different credit scoring models were developed, using and comparing three distinguished methods (Multiple Discriminant Analysis, Artificial Neural Networks, Logistic Regression). It was also developed a model using genetic algorithms that indicates the amount of credit to concede to a customer applying for credit.

This work shows that the Artificial Neural Networks are the classification method giving best results for the credit classification at Nors. Also the model that defines the amount of credit to concede to a customer can become a useful tool to CaR, because the usual analytical pointers are not oriented specifically to that goal.

Keywords: Credit Scoring Models; Artificial Neural Networks; Logistic Regression; Discriminant Analysis; genetic algorithms; customer credit plafond



# Índice

Agradecimentos .....	iii
Resumo.....	v
Abstract .....	vii
Índice .....	ix
Índice de Figuras.....	xii
Índice de Tabelas .....	xv
Glossário .....	xviii
1. Introdução.....	20
1.1 Relevância do Tema .....	20
1.2 Objetivo .....	22
1.3 Organização do Trabalho .....	22
2. Revisão de Literatura .....	24
2.1 Modelos de Classificação.....	24
2.1.1 Vantagens e Desvantagens .....	27
2.1.2 Redes Neurais Artificiais .....	28
2.1.3 Análise Multivariada Discriminante .....	35
2.1.4 Regressão Linear Múltipla.....	37
2.1.5 Regressão Logística.....	38
2.1.6 Modelos de Classificação de Crédito .....	38
2.1.7 Exemplos de aplicação de modelos de classificação de crédito .....	42
2.2 Problema de alocação de recursos .....	44
3. Metodologia.....	48
3.1 Procedimento de recolha de dados .....	49
3.2 Base de Dados .....	50
3.3 Outras considerações .....	57
3.4 Limitações do Método.....	58
4. O Grupo Nors.....	60
4.1 História do Grupo .....	60
4.2 Serviços Partilhados da Nors .....	64
4.3 Análise de Crédito das Contas a Receber.....	65
4.4 Desempenho da Análise de Crédito das Contas a Receber .....	72
5. Resultados do desenvolvimento dos modelos de crédito .....	79

5.1 Modelos de Classificação.....	82
5.2 Modelo de Definição de Montante de Crédito.....	99
5.3 Limitações dos resultados .....	109
6. Conclusões .....	113
6.1 Desenvolvimentos futuros .....	115
Bibliografia.....	116
Anexo 1. Gráficos de Dispersão.....	121
Anexo 2. RNA_1 .....	124
Anexo 3. Modelos de Classificação sem variável Histórico Interno .....	128
Anexo 4. Modelos de Classificação (RNA_3; RL_3; AMD_3) .....	131
Anexo 5. RNA_4 .....	138
Anexo 6. Classificação de Crédito do modelo RNA_4.....	142
Anexo 7. Resultados do RiskOptimizer a atribuir crédito aos clientes 1, 2 e 3 ...	144
Anexo 8. Resultados do RiskOptimizer a atribuir crédito aos clientes 1 e 2 .....	149
Anexo 9. Classificações de crédito dos clientes 4, 5, 6, 7 e 8, pelo modelo RNA_1 .....	152



# Índice de Figuras

Figura 1: Perceptrão.....	29
Figura 2: Perceptrão Multicamada .....	31
Figura 3: Organização Funcional Nors .....	62
Figura 4: Organização do Negócio Nors .....	63
Figura 5: Matriz de correlação das variáveis .....	80
Figura 6: Gráfico de Dispersão entre Resultado e Resultado Operacional ..	123
Figura 7: Informações RNA_1 .....	124
Figura 8: Arquitectura RNA_1 .....	126
Figura 9: Resumo do treino da RNA_1 .....	127
Figura 10: Matriz de Classificação RNA_1 .....	127
Figura 11: Informações do modelo RNA_3.....	132
Figura 12: Arquitectura da rede RNA_3.....	133
Figura 13: Resumo do treino do modelo RNA_3 .....	134
Figura 14: Matriz de Classificação do modelo RNA_3.....	134
Figura 15: Informações do modelo RNA_4.....	138
Figura 16: Arquitectura do modelo RNA_4.....	139
Figura 17: Resumo do treino do modelo RNA_4 .....	140
Figura 18: Matriz de classificação do modelo RNA_4.....	140
Figura 19: Probabilidade predita de classificação do modelo RNA_4.....	141
Figura 20: Classificações de crédito do modelo RNA_4.....	143
Figura 21: Distribuição do Valor líquido do Portfólio da CiviParts pós atuação do RsikOptimizer para conceder crédito aos 3 clientes.....	144
Figura 22: Distribuição Pert da tx de risco do Cliente 3 .....	144
Figura 23: Distribuição Pert da tx de risco do Cliente 2 .....	145
Figura 24: Distribuição Pert da tx de risco do Cliente 1 .....	145

Figura 25: Distribuição Cumulativa das Imparidades Previstas do Cliente 3 .....	146
Figura 26: Distribuição Cumulativa das Imparidades Previstas do Cliente 2 .....	146
Figura 27: Distribuição Cumulativa das Imparidades Previstas do Cliente 1 .....	147
Figura 28: Distribuição da Proporção entre FE e <i>Plafond</i> disponibilizado....	147
Figura 29: Modelo RiskOptimizer com restrições definidas.....	148
Figura 30: Distribuição Pert da tx de risco ponderada utilizada para calcular as imparidades previstas dos clientes já existentes no Portfólio da CiviParts .	148
Figura 31: Distribuição do Valor Líquido do Portfólio após a intervenção do RiskOptimizer para conceder crédito aos clientes 1 e 2 .....	149
Figura 32: Distribuição Pert da tx de risco do Cliente 1 .....	149
Figura 33: Distribuição Pert da tx de risco do Cliente 2 .....	150
Figura 34: Distribuição Cumulativa das Imparidades Previstas do Cliente 2 .....	150
Figura 35: Distribuição Cumulativa das Imparidades Previstas do Cliente 1 .....	151
Figura 36: Modelo RiskOptimizer com restrições definidas.....	151
Figura 37: Classificações de crédito do modelo RNA_1.....	152



# Índice de Tabelas

Tabela 1: Responsabilidade de atribuição de crédito .....	69
Tabela 2: Classes de Risco de Crédito .....	70
Tabela 3: Decisão de Crédito .....	73
Tabela 4: Desfecho das decisões de crédito .....	73
Tabela 5: Matriz de classificação das decisões de crédito das CaR .....	74
Tabela 6: Histórico interno dos casos aprovados .....	74
Tabela 7: Histórico interno dos casos rejeitados.....	75
Tabela 8: Desfecho das decisões de crédito de pedidos abertura.....	76
Tabela 9: Desfecho das decisões de crédito de pedidos reabertura.....	76
Tabela 10: Montantes de crédito concedido .....	77
Tabela 11: Matriz de Classificação RNA_1.....	86
Tabela 12: Importância variáveis independentes RNA_1 .....	87
Tabela 13: Função Discriminante do AMD_1 .....	87
Tabela 14: Matriz de classificação do modelo AMD_1.....	88
Tabela 15: Função do modelo RL_1.....	88
Tabela 16: Matriz de classificação do modelo RL_1.....	89
Tabela 17: Matriz de classificação do modelo RNA_3.....	93
Tabela 18: Importância das variáveis independentes do modelo RNA_3.....	94
Tabela 19: Matriz de classificação do modelo RNA_4.....	97
Tabela 20: Importância das variáveis independentes do modelo RNA_4.....	98
Tabela 21: Matriz de classificação RNA_2.....	128
Tabela 22: Importância das variáveis RNA_2.....	128
Tabela 23: Matriz de classificação do modelo RL_2.....	129
Tabela 24: Importância das variáveis do modelo RL_2.....	129
Tabela 25: Matriz de classificação AMD_2.....	130

Tabela 26: Importância das variáveis do modelo AMD_2.....	130
Tabela 27: Matriz de classificação do modelo AMD_3.....	135
Tabela 28: Função discriminante do modelo AMD_3 .....	135
Tabela 29: Matriz de classificação do modelo RL_3.....	136
Tabela 30: Função de regressão do modelo RL_3 .....	136



# Glossário

A - Ativo

AF – Autonomia financeira

AG – Algoritmos genéticos

AMD – Análise multivariada discriminante

CaR – Contas a Receber

CP – Capital próprio

CSP – Centro de serviços partilhados

FM – Fundo de maneo

GRA – Grau de rotação do ativo

MCC – Modelos de classificação de crédito

NFM – Necessidades de Fundo de Maneio

Norshare – Centro de serviços partilhados do Grupo Nors

PMC – Percepção Multicamada

RL – Regressão Logística

RLM – Regressão Linear Múltipla

RNA – Redes Neurais Artificiais

TL – Tesouraria Líquida

VN – Volume de Negócios



# Capítulo 1

## Introdução

### 1.1 Relevância do Tema

O presente estudo foi desenvolvido no âmbito do Mestrado em Gestão, da Católica Porto Business School, tendo-se realizado em contexto de estágio, no Centro de Serviços Partilhados, designado por Norshare, do Grupo Nors. O Grupo Nors é uma empresa líder em soluções de transporte e equipamentos de construção, a nível mundial.

Hoje em dia, o ambiente empresarial é caracterizado por uma forte concorrência. As organizações sentem, cada vez mais, a necessidade de desenvolver vantagens competitivas que lhes permitam destacarem-se das demais. Assim, todas as organizações procuram ferramentas que permitam reduzir custos, otimizar processos e aumentar a competitividade, de forma a reforçar a sua posição.

O Centro de Serviços Partilhados (CSP) da Nors faz parte de um modelo de gestão que tem como finalidade concentrar atividades de suporte, como atividades de recursos humanos e contabilidade, transversais a toda a organização, desenvolvendo-as numa unidade criada com esse intuito, de forma a reduzir custos e a padronizar os processos. A implementação deste tipo de modelos tem vindo a ser uma prática frequente em muitas instituições, já que estes podem constituir uma vantagem comparativa.

Uma das áreas inseridas no CSP do Grupo Nors é a área “Contas a Receber”, na qual se desenvolve a análise de crédito a clientes. Esta análise tem como propósito a tomada da decisão de conceder crédito a clientes e aprovar certo tipo de negócios.

Quanto à decisão de conceder crédito a clientes, esta é tomada por analistas financeiros, de forma subjetiva e com base na sua experiência, incluindo na sua avaliação alguns fatores como: diversos rácios financeiros; o histórico de pagamentos interno do cliente; o montante de processos judiciais em aberto contra a entidade. Este é um problema de classificação, ou seja, o analista financeiro tem de classificar o cliente como “bom pagador” ou “mau pagador”, de forma a tomar a decisão de conceder ou não crédito e que montante conceder.

Um Modelo de Classificação de Crédito (MCC) (*Credit Scoring Model*) é uma ferramenta que permite resolver problemas de classificação. Este tipo de ferramenta estabelece uma classificação (*scoring*) para cada candidato a crédito, que o distinguirá como, bom ou mau, pagador. Inicialmente, estes modelos começaram por ser aplicados em instituições financeiras, na década de 60, com o objetivo de mecanizar a decisão de conceder crédito. De facto, numa altura em que a procura de crédito estava a crescer, tornou-se necessário para estas instituições, recorrer a ferramentas que lhes permitissem uniformizar as suas decisões, reduzir as perdas e aumentar os ganhos. O seu objetivo é a redução do tempo na resolução do problema, pela standardização da informação e decisão, e a redução de custos, ao aumentar os lucros e diminuir as perdas, através de uma previsão mais precisa sobre o comportamento do cliente, do que a atingida por uma avaliação subjetiva, de um analista financeiro. Hoje em dia, quase todos os Bancos utilizam ferramentas do tipo de MCC nas suas operações de crédito, que se tornaram uma ferramenta indispensável a este tipo de atividade. De facto, este é o tipo de ferramenta que as organizações procuram para fazer face à nova realidade competitiva, introduzida pelo mercado global. Nesse sentido, esta ferramenta vai de encontro aos objetivos dos CSP, ao permitir uniformizar processos, poupando tempo e recursos.

Assim, o tema surgiu devido à dificuldade sentida na empresa de tornar a sua decisão de crédito mais mecanizada e uniforme. De facto, a importância deste estudo prende-se com a possibilidade de aplicar um modelo de classificação de

crédito a uma empresa comercial. Desta forma, será possível otimizar a área de análise de crédito, ao mecanizar o processo de decisão de crédito, poupando tempo e custos ao Norshare, e esperando-se também melhores resultados de decisão. Nessas circunstâncias, a empresa estará a aumentar a qualidade do seu serviço, prestando-o aos seus clientes internos a um menor custo, o que gera valor para o Grupo, como um todo.

## 1.2 Objetivo

No início do estágio realizado no Grupo Nors, foi proposto ao aluno a elaboração de um modelo que permitisse distinguir os candidatos a crédito e, em segundo plano, a elaboração de um modelo que identificasse o montante de crédito a conceder a cada candidato de crédito. Nesse sentido, o presente trabalho tem como objetivo desenvolver e aplicar um modelo de classificação de crédito e criar um modelo que indique o montante de crédito a conceder ao candidato de crédito.

No âmbito do problema proposto, este estudo assenta numa investigação quantitativa, usando os métodos de regressão logística (RL), redes neuronais artificiais (RNA) e análise multivariada discriminante (AMD), de forma a comparar os seus resultados e a discutir as vantagens e desvantagens relativas entre os modelos. Nesse sentido, a questão de investigação do presente trabalho passa por escolher um método de classificação, a aplicar ao caso de classificação de crédito do Grupo Nors, e como decidir qual o montante de crédito a atribuir a cada candidato de crédito.

## 1.3 Organização do Trabalho

O trabalho que de seguida se apresentará está dividido em mais 5 capítulos, para além do capítulo de introdução.

O segundo capítulo é composto pela Revisão de Literatura, onde se apresenta, primeiramente, uma análise bibliográfica dos métodos propostos, neste trabalho, para resolver o problema de classificação de crédito e posteriormente, uma análise dos modelos de classificação mais utilizados no contexto de classificação de crédito e de previsão de falência. Também se insere neste capítulo, uma apresentação da literatura de problemas de alocação de recursos e das ferramentas que serão aplicados ao problema de decisão do montante de crédito a conceder.

O terceiro capítulo apresenta a metodologia utilizada no presente trabalho, mais concretamente, as questões de investigação, o tratamento e análise dos dados e outras considerações efetuadas pelo aluno.

O quarto capítulo introduz a situação atual da empresa relativamente ao processo de atribuição de crédito, nomeadamente, os montantes envolvidos, os indicadores utilizados na tomada de decisão, as taxas de incumprimento e atraso por parte dos clientes e por fim, os resultados das decisões tomadas pelo departamento responsável pela análise de crédito.

No quinto capítulo discute-se o processo de aplicação, dos métodos estudados na Revisão de Literatura, à classificação de crédito, no caso do Grupo NORS, bem como, os seus resultados.

Por último, no sexto capítulo são apresentadas as conclusões deste trabalho e indicações para investigação futura.

# Capítulo 2

## Revisão de Literatura

A presente revisão de literatura está dividida em três partes. Na primeira parte, inserem-se os temas que foram estudados para abordar o problema de classificação. Assim, nesta parte insere-se o estudo sobre modelos de classificação, destacando-se as RNA, a RL e a AMD, por serem os três métodos mais frequentemente aplicados. Na segunda parte, inserem-se os temas estudados sobre classificação de crédito, nomeadamente a sua história e exemplos de aplicações. Por último, introduzem-se os temas estudados para resolver o problema do montante de crédito a conceder ao cliente.

### 2.1 Modelos de Classificação

De acordo com Jain et al. (1999), o reconhecimento de padrões é o estudo de como as máquinas conseguem observar o ambiente, aprender a distinguir padrões de interesse com base no seu conhecimento e tomar decisões razoáveis sobre as categorias dos padrões. O mesmo autor defende que este é um problema de reconhecimento, definido como uma tarefa de classificação, sendo que as categorias podem ser definidas pelo criador do sistema (aprendizagem supervisionada) ou podem ser aprendidas com base na semelhança dos padrões (aprendizagem não supervisionada).

As aplicações do estudo de reconhecimento de padrões são muitas, sendo o Data Mining uma delas e consistindo na procura de padrões em informação, através de *inputs* que podem ser representados num espaço multidimensional.

Para desenvolver um sistema de reconhecimento de padrões são desenvolvidos três passos fundamentais, sendo eles a aquisição de dados e processamento dos mesmos, sua representação e a tomada de decisão, conforme

Jain et al. (1999). Estes autores defendem que duas das melhores formas de desenvolver um sistema de reconhecimento de padrões são as redes neuronais e a classificação estatística. Na presente revisão de literatura vão ser estudados métodos de regressão linear, regressão logística, análise discriminante, conforme proposto nos objetivos do trabalho, que constituem formas de classificação estatística utilizados para criar sistemas de reconhecimento de padrões e também RNA.

No fundo estes modelos utilizam características que representam as populações de cada padrão (variáveis independentes), para reconhecer os padrões e classificar novos casos consoante as suas características numa determinada categoria (variável dependente). Watanabe (1985) define um padrão como o oposto de caos, sendo uma entidade, vagamente definida, á qual se pode dar um nome.

O desempenho de um classificador (sistema de reconhecimento de padrões) depende da interrelação entre os tamanhos das amostras representativas de cada padrão, do número de características que definem as populações de cada padrão e a complexidade do classificador, conforme Jain et al. (1999). Os autores defendem que o aumento de características para caracterizar as populações pertencentes a cada padrão pode ter efeitos negativos na capacidade do classificador de estabelecer previsões corretas, se o tamanho da amostra de cada população for relativamente pequeno face ao número de características. Isto porque estes classificadores estimam parâmetros desconhecidos, sendo que o aumento do número de características, *ceteris paribus*, aumenta o número de parâmetros a estimar, o que diminui a sua confiabilidade. No entanto, se o tamanho da amostra representativa de cada padrão aumentar, os parâmetros estimados pelo classificador vão ser mais confiáveis perante uma situação em que o tamanho das amostras seria menor.

De facto, a aplicação de sistemas de reconhecimento de padrões é cada vez maior e podemos observá-la em pormenores das nossas vidas, como por

exemplo, os vídeos sugeridos pelo *Youtube* com base nas nossas preferências pessoais, classificação de crédito, campanhas de marketing ou o modo de piloto automático dos aviões. Tal é possível devido à evolução tecnológica, o que permitiu, tanto recolher mais informação a todos os níveis e armazenamento desta, como trabalhar bases de dados enormes de forma rápida. Segundo um artigo do *The Economist*, em 2013 foram gerados 4.4 zettabytes de dados digitais e espera-se que em 2020 este número aumente para 44 zettabytes.

Isto significa que os sistemas de reconhecimento de padrões criados, nos dias de hoje, utilizam amostras cada vez maiores, o que permite estimar os parâmetros desconhecidos com um nível de confiança maior, traduzindo-se numa capacidade de previsão mais precisa. No entanto, isto leva a que os utilizadores destes modelos, por exemplo Bancos, confiem, de forma cega, nas suas previsões, o que pode ter consequências indesejáveis, como por exemplo, a ocorrência de fenómenos, designados por Taleb (2007), de Cisnes Negros, que são acontecimentos altamente improáveis com um enorme impacto. Taleb (2007) alerta para o facto de que muitas vezes as ferramentas estatísticas são aplicadas de forma cega a outras áreas de estudo, mas muitas vezes estas ferramentas não são compreendidas pelos seus utilizadores.

Assim, é importante perceber que estas ferramentas são excelentes no auxílio da tomada de decisão, mas não devem ser utilizadas sem supervisão e sem, em última instância, existir uma análise que incorpore o fator humano.

Um modelo de classificação pode ser construído a partir de vários métodos quantitativos distintos. Ong et al. (2005) distinguem dois tipos de métodos utilizados na construção destes modelos: i) os métodos estatísticos paramétricos que utilizam, por exemplo, regressão linear e análise discriminante; ii) os métodos estatísticos não paramétricos utilizam redes neuronais artificiais, árvores de decisão, entre outros. A mesma distinção é realizada por Desai et al. (1996) com nomes diferentes. Esses autores dividem os mesmos grupos de

métodos em dois grupos: métodos estatísticos convencionais e métodos estatísticos avançados, respetivamente.

### 2.1.1 Vantagens e Desvantagens

A aplicação de modelos de classificação é cada vez maior em contexto organizacional, como forma de auxílio da tomada de decisão ao nível operacional. Por isso, interessa identificar as vantagens e desvantagens destes modelos, uma vez que se está a propor a aplicação de um deles, num contexto empresarial.

Thomas et al. (2002) verificam que os modelos de classificação apresentam vários benefícios, como por exemplo, a menor quantidade de informação necessária para tomar a decisão, já que o uso dos modelos, pós-construídos, necessita apenas de um determinado número de variáveis com significância estatística, que são aquelas utilizadas pelo modelo. Além disso, os modelos de classificação são construídos em bases de dados com um número de casos muito maior do que aquele que uma pessoa consegue memorizar. Além disto, um ser humano não consegue estabelecer as interações entre um grande número de variáveis e características de um determinado caso, devido ao seu elevado grau de complexidade, enquanto um bom modelo de classificação, devidamente testado, conseguirá fazê-lo.

Chandler & Coffman (1979) também enumeraram um conjunto de benefícios destes modelos. Alguns destes benefícios são o mais eficiente processamento do tempo, a minimização dos custos e esforços no processamento das decisões, o menor número de erros, a menor quantidade de informação necessária para tomar a decisão, entre outros. Mais especificamente, os autores referem que estes métodos têm vantagens quanto à facilidade de exercer controlo sobre as decisões e de prever resultados e elaborar relatórios de gestão. Contudo, os autores também alertam que o método tradicional de julgamento subjetivo, de um ser humano, tem vantagens em casos excecionais. Isto porque, nesses casos, a

capacidade de previsão dos modelos de classificação pode não ser tão boa, pelo facto de não conhecer casos desse género.

Por outro lado, Heffernan (2005) apresenta algumas preocupações relativamente aos modelos de classificação. Critica o facto de os modelos se basearem em informação histórica porque, as variáveis, ou os seus pesos, são assumidos como constantes ao longo do tempo, o que torna o modelo menos preciso, a não ser que seja atualizado frequentemente.

Thomas et al. (2002) (1996) também apresentam alguns inconvenientes dos modelos de classificação, como o problema da falha na classificação e o facto de estes modelos excluïrem certas variáveis (impossível incluir todas), o que os torna estatisticamente incompletos.

Reichert et al. (1983), apesar das dúvidas (com as ferramentas disponíveis à época) sobre a capacidade de previsão destes modelos, apontam-lhes como vantagens a forma consistente, objetiva e eficiente como as previsões são feitas.

De facto, é difícil negar que os modelos de classificação são uma ferramenta que permite tomar decisões de forma controlada, sustentada e rápida. No entanto, estes modelos não devem ser utilizados para tomar decisões automaticamente, mas sim para apoiar a tomada de decisão.

### 2.1.2 Redes Neuronais Artificiais

As RNA, também designadas por Redes Neuronais, surgiram em 1943, pelo trabalho de McCulloch e Pitts e são inspiradas no funcionamento do cérebro humano.

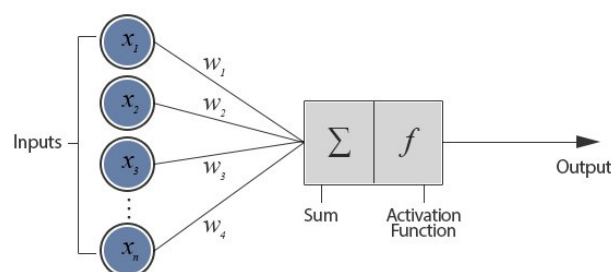
De acordo com Haykin (1999), as Redes Neuronais Artificiais podem ser definidas da seguinte forma: uma rede neuronal artificial é um processador massivamente distribuído em paralelo e constituído por unidades de processamento simples, que tem propensão natural para armazenar conhecimento experimental e torná-lo disponível para ser utilizado. Assemelha-se ao cérebro humano em dois aspetos:

- O conhecimento é adquirido pela rede a partir do meio envolvente, através de um processo de aprendizagem;
- As intensidades das ligações entre neurónios, conhecidos como pesos sinápticos, são usadas para armazenar o conhecimento adquirido.

Lingireddy et al. (2005) definem RNA como uma rede de vários processadores simples (ou unidades ou neurónios), cada um possivelmente tendo uma pequena quantidade de memória local, interconectados por canais de comunicação que transportam a informação numérica. Os mesmos autores indicam várias aplicações possíveis deste tipo de redes, incluindo classificação de informação (ou *grouping*) e aproximação de funções (ou mapeamento).

Lingireddy et al. (2005) apresentam vários modelos de RNA, como por exemplo, o perceptrão e as redes feed forward, nas quais se insere o Perceptrão Multicamada (PMC). Existem outros tipos de redes disponíveis. No entanto, por o PMC ser a rede eleita para este trabalho, este vai ser analisado em pormenor. Esta foi a RNA eleita por se tratar da RNA mais utilizada, ter uma concepção simples e ter capacidade para captar os parâmetros estatísticos do conjunto de dados, segundo Kasabov (1996). Podem-se encontrar referências a outras redes, por exemplo, no estudo de West (2000), que investiga a capacidade de 5 tipos de rede, diferentes, como a rede de base radial.

O perceptrão (Figura 1) é um dos modelos de RNA mais simples e foi inicialmente desenvolvido por Rosenblatt (1958).



**Figura 1:** Perceptrão

A rede recebe vários *inputs* ( $X_1, X_2, \dots$ ) e produz um único resultado, que determina a que classe um padrão de informação pertence. Os nós de *input* e *output* estão conectados por pesos  $W_i$  (ver figura 1). Assim, a fórmula matemática de um perceptron com “n” *inputs* pode ser escrita da seguinte forma:

$$Y = f\left(\sum_{i=1}^n X_i W_i + \theta\right),$$

sendo  $\theta$  uma constante sempre igual a 1.

Minsky & Papert (1972) defendem que este tipo de rede só pode ser utilizado para classificar padrões linearmente separáveis e propuseram a utilização de uma rede feed forward com duas camadas para resolver este problema.

Assim, surgiram as redes feed forward, sendo o PMC um tipo de rede *feed forward*, que permite resolver o problema apontado por Minsky & Papert (1972), ao classificar padrões não linearmente separáveis.

Ong et al. (2005) defendem que a arquitetura de uma RNA *feed forward* pode ser representada, no geral, como um sistema de, pelo menos, três camadas, sendo elas, uma camada de *input*, uma ou mais camadas ocultas e uma camada de *output*. A camada de *input* processa, primeiramente, a informação apresentada à rede para a camada escondida, através de sinais. A camada escondida utiliza essa informação para calcular os pesos de cada variável através de uma função de transferência, como por exemplo, uma função hiperbólica tangente ou logística, enviando a informação para a camada de *output*. Estes autores afirmam que ao combinar um conjunto de neurónios computacionais num sistema altamente interconectado, é possível detetar a complexa relação não linear da informação.

Desai et al. (1996) apresentam o mesmo modelo de rede do PMC (ver figura 2). Para estes autores, cada camada da rede consiste em um ou mais elementos processadores, conhecidos como nós (neurónios). A camada de *input* apresenta p nós, consoante o número de variáveis independentes, sendo que cada elemento processador da camada envia sinais  $X_i$  ( $i=1, \dots, p$ ) para cada um dos elementos processadores das camadas ocultas (neste tipo de rede podem existir várias

camadas ocultas). Nestas camadas, cada nó ( $j=1,\dots,p$ ) produz uma ativação  $a_j = G\left(\sum_i W_{ij}X_i\right)$ , onde  $W_{ij}$  representa o peso associado às conexões entre os  $p$  nós da camada de input e os  $j$  nós das camadas ocultas. Os nós da camada de *output* funcionam de uma forma similar aos nós das camadas ocultas para produzir o resultado da rede.

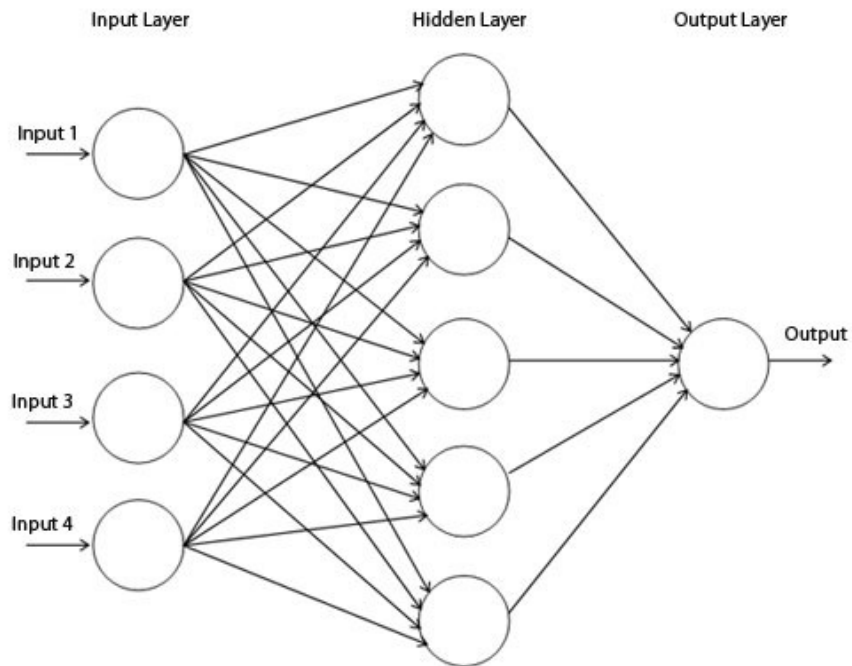


Figura 2: Perceptrão Multicamada

### 2.1.1.1 Funções de Ativação

Uma das grandes vantagens dos modelos PMC é a sua capacidade para reconhecer padrões não linearmente separáveis, como já vimos anteriormente

Como referido no subcapítulo anterior, o modelo PMC utiliza funções de ativação entre as camadas da rede, para processar a informação entre os nós das camadas. As funções mais importantes são as seguintes:

- Função de ativação linear ( $f(x) = x$ ) devolve os valores sem transformações
- Função de ativação Sigmoidal ( $f(x) = \frac{1}{1 + e^{-x}}$ ) transforma os valores para o intervalo de (0,1)

- Função de ativação tangente hiperbólica ( $f(x) = 2 * \frac{1}{1 + e^{-x}} - 1$ ) transforma os valores no intervalo de (-1,1)

O modelo PMC deve incluir as funções mais adequadas, de forma a ter capacidade para aproximar funções não lineares, ou seja, reconhecer padrões não linearmente separáveis, segundo Lingireddy et al. (2005). Então, o modelo PMC deve incluir pelo menos uma função de ativação não linear, para introduzir não linearidade na rede, sendo que é prática comum utilizar a função de ativação linear na camada de output, já que esta não limita os resultados possíveis, uma vez que não converge assintoticamente. Karlic & Olgac (2011) elaboram um estudo para medir o desempenho de várias funções de ativação que podem ser aplicadas às RNA e concluem que a função de ativação tangente hiperbólica é aquela que tem melhor desempenho e que pode ser utilizada na maior parte das aplicações de um PMC.

### 2.1.1.2 *Inputs*

Segundo Lingireddy et al. (2005), a contribuição de um *input* depende da sua variabilidade relativamente a outro *input*. Por isso, é importante escalar a informação, para que a variabilidade dos *inputs* reflita a sua importância, ou que pelo menos, o contrário não suceda. Se um *input* é mais importante que outro, este deve ter uma variabilidade nos seus valores superior. Isto porque se um *input* varia entre 0 e 1, enquanto outro varia entre 0 e 1000000, a rede vai esforçar-se mais a aprender o segundo alvo, possivelmente excluindo o primeiro.

Selecionar os *inputs* é, também, um processo importante na construção de uma RNA. Um modelo com um número de *inputs* otimizado, impede que se gaste tempo a coleccionar informação desnecessária e permite que se eliminem impactos não desejados de informação irrelevante. Para isso existe um processo de três passos referido por Lingireddy et al (2005):

1. Construir uma RNA feed forward com todos os possíveis *inputs*;

2. Reconstruir a rede com um número menor de *inputs* com diferentes combinações;
3. Verificar os efeitos da redução de dimensão e remover as variáveis *input* com contribuição insignificante.

### 2.1.1.3 Treino da Rede

O objetivo das redes neuronais é perceber a relação entre os *inputs* e os *outputs*. Segundo Lingireddy et al. (2005), para tal é necessário que os *inputs* da rede tenham informação objetiva sobre os *outputs*, para que exista uma função que descreva a relação entre as variáveis (dependentes e independentes) e que possa ser aprendida e aproximada pelas RNA.

Na metodologia das RNA, o conjunto de dados disponível é geralmente dividido em três subconjuntos para o treino, teste e validação. O subconjunto de treino é utilizado para a rede aprender a função e definir os pesos sinápticos associados aos nós. O subconjunto de teste é utilizado para avaliar o desempenho da generalização da rede. No caso de a precisão dos resultados do teste não ser a desejada, o treino deve continuar. Se se atingirem bons resultados, o modelo pode ser avaliado com o subconjunto de validação, que confirma a generalização.

O processo de treino da rede é análogo ao processo de minimizar uma função de erro. Nesse sentido, as redes neuronais utilizam algoritmos de otimização para minimizar a sua função erro.

Hoje em dia, existem muitos tipos de algoritmos de otimização diferentes com aplicação às RNA. No entanto, no presente estudo apenas vão ser descritos o algoritmo de Retropropagação e o algoritmo de Descida de Gradiente, por se tratar de dois dos algoritmos mais simples com aplicação às RNA.

O algoritmo da retropropagação foi o primeiro algoritmo bem-sucedido no processo de treino das RNA. Este algoritmo minimiza o erro através do método de procura gradual. O método define pesos para as conexões nos nós para o início do treino e ajusta-os pela sua contribuição para a função de erro. No fundo, este algoritmo permite uma procura gradual local do erro, percorrendo todos os nós

da rede e ajustando os seus pesos, embora não garanta a descoberta do mínimo da função erro. Isto porque esta função pode ter vários mínimos, motivo por que é importante saber quando parar o treino. Lingireddy et al. (2005) referem que o algoritmo da retro propagação é um algoritmo de aprendizagem supervisionada, o que significa que para cada padrão de *input*, existe um *output* correspondente, que funciona como alvo a atingir. Durante a fase de treino, qualquer diferença entre o *output* produzido pela rede e o *output* alvo será tratada como um erro, sendo o objetivo do treino minimizar o erro.

O algoritmo Descida do Gradiente é um dos algoritmos mais simples utilizados no treino de RNA. Para minimizar a função erro, este começa por definir um vector de pesos seleccionado de forma aleatória, na maior parte das vezes, designado por  $W^{(0)}$ . Posteriormente, o algoritmo atualiza os peso deste vector a cada iteração, de forma que, no passo  $r$ , se mova uma pequena distância na direção da maior taxa de diminuição do erro, ou seja, na direção do gradiente negativo, avaliado por  $W^{(r)}$ :

$$\Delta W^{(r)} = -\eta \nabla E|_{W^{(r)}}$$

Numa versão sequencial da descida do gradiente, a função de erro do gradiente é avaliada para cada padrão de cada vez, sendo que o vector de pesos é atualizado através da seguinte forma:

$$\Delta W^{(r)} = -\eta \nabla E^n|_{W^{(r)}}$$

Onde  $n$  representa o número de diferentes objetos disponíveis para o treino. O parâmetro  $\eta$  representa a taxa de aprendizagem, e assumindo um certo valor, é esperado que o valor de  $E$  diminuirá a cada iteração, até chegar ao vector de pesos que satisfaz a restrição de paragem do treino dada ao algoritmo, segundo Bishop (1995). Note-se que o valor de  $\eta$  é dado pela seguinte equação:

$$\eta = -\frac{\Delta W^{(r)}}{\nabla E|_{W^{(r)}}}$$

Isto significa que a taxa de aprendizagem do algoritmo é dada pela relação entre a alteração no vector dos pesos e a alteração no valor do erro produzida pelas alterações no vector de pesos.

Existem outros algoritmos de treino que podem ser aplicados com sucesso às RNA. Por exemplo, Kollias (1989) desenvolve um algoritmo de treino de RNA baseado numa modificação do método dos mínimos quadrados e Moller (1993) desenvolve um algoritmo designado por Gradiente Conjugado em Escala, que se baseia numa modificação do algoritmo de retropropagação.

### 2.1.3 Análise Multivariada Discriminante

A análise multivariada discriminante é um dos métodos mais utilizados em relação a problemas de classificação e foi, primeiramente, proposto por Fisher (1936) como um método de discriminação e classificação.

O método AMD é uma técnica estatística paramétrica cujo propósito é definir uma função que permita discriminar um objeto perante dois ou mais grupos, baseando-se num conjunto de características desse objeto.

Este sistema assume um conjunto de pressupostos que por vezes pode limitar a sua aplicação a certo tipo de problemas de classificação. É assumida uma distribuição multivariada normal para variáveis independentes e homogeneidade nas matrizes de variância e covariância de cada grupo, segundo Poulsen & French (2008). Além disso, este método é bastante sensível a *outliers* e devem ser utilizadas variáveis independentes, ou seja, variáveis sem correlação.

Como refere Altman (1968), aAMD é uma técnica estatística utilizada para classificar uma observação em dois ou mais grupos, a priori definidos, com base nas características do individuo (observação), sendo que, no primeiro passo, há que definir as características que distinguem os dois (ou mais) grupos. Esta técnica é muito utilizada em problemas de classificação em que a variável dependente tem carácter qualitativo, como homem ou mulher.

Assim, a AMD procura estabelecer uma combinação linear entre estas características que melhor discriminam entre os grupos, definindo um conjunto de coeficientes discriminantes, com base na correlação entre as variáveis, que quando aplicados aos atributos (valores de uma característica) de um objeto (observação), permite classificá-lo em um dos dois grupos.

Segundo Desai et. al. (1996), o objetivo de um modelo de análise discriminante linear é definir uma função de forma a transformar os valores das variáveis dos objetos num resultado discriminante geral,  $Z$ , tal como proposto por Altman (1968):

$$Z = W_1X_1 + W_2X_2 + \dots + W_pX_p$$

Onde  $X$  representa as variáveis independentes e  $W$  representa o peso (coeficiente) de cada variável no modelo e  $p$  representa o número de variáveis independentes.

Greene (1992), define um indivíduo  $i$  com um vetor de atributos  $X_i$  e define as variáveis independentes como características e os valores dessas características, como atributos. Assim, este autor defende que a técnica da análise discriminante passa por distinguir um indivíduo  $i$  perante duas populações caracterizadas por uma distribuição normal multivariada de atributos  $X$ . Desta forma, um elemento com um vetor de atributos  $X_i$ , pode ser classificado numa dessas duas populações.

Este método pode-se considerar similar ao perceptrão, utilizado nas RNA, na medida em que o modelo utiliza um conjunto de variáveis independentes, também chamadas de características na literatura, combinando-as numa função de transformação que estabelece o resultado. Trata-se de um método que se afigura intuitivo e que está provavelmente dotado de adequada eficiência, mas tal como o perceptrão (das RNA), não permite distinguir padrões não linearmente separáveis, na medida em que o seu *output* é produzido por uma função linear. No entanto, este é ainda um dos métodos mais aplicados neste tipo

de problema e tem muitas outras possíveis aplicações, como por exemplo, na área de estudo da biologia e ciências comportamentais.

### 2.1.4 Regressão Linear Múltipla

Embora este estudo não tenha como propósito aplicar a regressão linear múltipla (RLM), esta é apresentada para contextualizar a regressão logística, mais á frente introduzida.

Conforme Guimarães & Cabral (2007), o método de regressão linear múltipla permite descrever a relação entre um conjunto de variáveis quantitativas independentes,  $X_i$  ( $i=1,2,\dots,m$ ), e uma variável dependente quantitativa,  $Y$ , através da seguinte equação:

$$Y_n = B_0 + \beta_1 X_{1j} + B_2 X_{2j} + \dots + B_m X_{mj} + e_j$$

Onde,  $m$  é o número de variáveis independentes  $X_1, \dots, X_m$ ;  $j$  representa o número de observações das variáveis  $X_i$ ,  $e_j$  é o erro aleatório associado ao valor observado  $Y_j$  e  $B_m$  são parâmetros fixos a estimar, da relação entre as variáveis  $X_1$  a  $X_m$  e  $Y$ .

A teoria da regressão, na qual o modelo se fundamenta, assume que os erros, que incidem diretamente sobre os valores observados de  $Y$ , têm um valor esperado nulo e variância constante, são mutuamente independentes e seguem uma distribuição normal. Se tais hipóteses se verificarem, os valores da variável dependente também vão ser independentes, seguindo uma distribuição normal, com variância  $\sigma^2$  e valor esperado  $\mu Y_n$ .

O modelo permite estimar os parâmetros  $B_0, \dots, B_m$  ao recorrer ao método dos mínimos quadrados, que, a partir de um conjunto de observações, minimiza o erro que decorre da diferença entre os valores de  $Y$  previstos pelo modelo e os observados.

## 2.1.5 Regressão Logística

A Regressão logística é um caso particular dos modelos de regressão linear, em que a variável dependente é binária e não uma variável contínua, como no caso anterior. Este é um tipo de método que pode servir vários tipos de funções de distribuição, como por exemplo, a distribuição normal e a distribuição de Poisson.

Desai et al. (1996) definem a RL através seguinte equação:

$$Z = \frac{1}{1 + e^{-z}} ,$$

- Z é a probabilidade de uma determinada classe de resultado

- $$z = \sum_{j=1}^m B_i X_{ij} + e_j$$

Em que i representa o número de variáveis independentes (i=1,2,3...m) e j é o número de observações.

Este método assume uma relação linear entre as variáveis dependentes, expressas numa escala logarítmica, e as independentes, sendo também, um dos métodos estatísticos mais utilizado na resolução de problemas de classificação.

Segundo Ong et al. (2005), apesar de a regressão logística ter bom desempenho em várias aplicações, quando as relações no sistema são não lineares, a precisão deste modelo diminui, sendo os modelos estatísticos não paramétricos uma solução para este tipo de problema.

## 2.1.6 Modelos de Classificação de Crédito

Pode-se considerar que o fenómeno de conceder crédito surge cerca de 3000 a.C., com base em cereais, na Babilónia, uma zona rica em agricultura, por essa altura, conforme Homer & Sylla (1996). De facto, o crédito é quase tão antigo como a atividade económica. A palavra crédito deriva da palavra do latim "credo", que significa "Eu acredito", segundo Anderson (2007). Este conceito consiste em alguém ceder a um outro agente, uma parte do seu património, com

a esperança de o obter de volta, numa data futura. Assim, se para conceder crédito é necessário confiar, uma relação de crédito tem risco inerente. Então, surge o conceito de risco de crédito que, segundo James Lam (2014), é o risco de um cliente, contraparte ou fornecedor, falhar em cumprir com as suas obrigações.

Os MCC são abordagens que procuram classificar um cliente quanto ao risco de crédito, com base em métodos estatísticos aplicados ao problema de conceder crédito.

Gup et al. (2001) definem classificação de crédito como o uso de modelos estatísticos para determinar a probabilidade de um candidato de crédito entrar em *default* com a sua obrigação, enquanto Thomas et al. (2002) definem classificação de crédito como o conjunto de modelos de decisão e as respetivas técnicas que ajudam as instituições financeiras no processo de conceder crédito.

Então, pode-se concluir que estes modelos têm como objetivo resolver um problema que é a distinção de candidatos de crédito face ao seu risco de crédito, ou seja, estes modelos utilizam métodos estatísticos para distinguir um novo candidato a obter crédito perante padrões já definidos, tratando-se, portanto, de um problema de classificação. Este problema passa por classificar um candidato a obter crédito como pertencente a um grupo, com base nas suas características, sendo que existem pelo menos dois grupos, maus pagadores e bons pagadores. Em suma, as diferentes categorias de pagador em que o candidato pode ser classificado constituem a variável dependente do modelo, enquanto as características dos candidatos, utilizadas pelo modelo para os distinguir, são as variáveis independentes.

A abordagem mais tradicional para tomar a decisão de conceder crédito baseia-se numa avaliação humana subjetiva. Este tipo de solução utiliza experiências passadas de crédito para tentar prever se um candidato a crédito vai ser bom ou mau pagador. No entanto, com o crescimento da procura de crédito e a evolução das tecnologias informáticas, foi-se generalizando o acesso a novos

modelos estatísticos mais sofisticados, cujo objetivo é ajudar a tomada de decisão de conceder crédito.

Durand (1941), foi o primeiro a mencionar um sistema para avaliação de candidatos a crédito, tendo procurado identificar que parâmetros eram importantes para quem concedia crédito e quais as características mais significantes a nível estatístico.

Nesse período, as empresas começavam a ter dificuldades em gerir a questão do crédito, porque muitos dos analistas financeiros habitualmente envolvidos na tomada de decisões sobre a concessão de crédito, foram destacados para a guerra e por isso, era difícil encontrar pessoas que detivessem a experiência para orientar este tipo de decisão. Desta forma, começou a surgir a necessidade de mecanizar a decisão. Aliás, muitas empresas pediram a esses analistas para elaborar em uma lista, o conjunto de regras que eles utilizavam para tomar a decisão, antes de partirem para as frentes de combate, conforme referido por Thomas (2000).

Contudo, os modelos de classificação só se desenvolvem e propagam a várias utilizações, na década de 60, com estudos desenvolvidos por Beaver (1966) e Altman (1968), sendo que Altman (1968) foi o primeiro a elaborar um modelo de previsão de falência operacional.

Assim, é a partir da década de 70 que se verifica uma aplicação crescente deste tipo de modelos, tal como é apontado por Chandler & Coffman (1977). Também, West (2000) defende que os modelos de classificação de crédito são cada vez mais utilizados pelas instituições financeiras na tomada de decisões de conceder crédito.

Este fenómeno surge associado ao aparecimento dos cartões de crédito no final da década de 60 e com o Ato de Igual Oportunidade de Crédito, concretizado nos Estados Unidos da América (EUA), em 1974, e que tinha como objetivo impedir a discriminação no acesso ao crédito, com base no sexo ou estado de casamento,

e que mais tarde foi alterado (1976), de forma a impedir discriminações pela raça, cor, religião, entre outros aspetos.

Com o crescimento da indústria de crédito, ao longo das últimas 5 décadas, vários métodos de classificação foram estudados e otimizados, de forma a aumentar a sua precisão de classificação. Isto porque, uma melhoria na precisão de um modelo, de uma fração de uma percentagem, pode traduzir-se em poupanças de milhões de euros para uma instituição financeira, conforme referem Ong et al. (2005).

No entanto, a presente revisão de literatura também utiliza estudos com propósitos diferentes de prever a situação de crédito de um cliente. Isto porque estes modelos recorrem a métodos estatísticos que podem ser utilizados para outros fins e que por isso, também devem ser analisados (por exemplo, Altman (1968) aplica um modelo de classificação para previsão de falência de empresas). Com efeito, apesar dos recentes estudos e evoluções dos modelos de classificação, provocados pelo crescimento da indústria de crédito ao consumidor, não foi encontrada literatura sobre aplicação destes modelos ao acto de conceder vendas a crédito a clientes, por parte de empresas comerciais. Admite-se que isto está relacionado com o facto de muitos destes estudos serem confidenciais, já que criam vantagens competitivas para as empresas e por isso, não existe incentivo à sua divulgação. Por exemplo, Silva (2011) apresenta a análise de crédito do Banco Santander Consumer Portugal, Banco especialista no financiamento ao consumo, onde evidencia a utilização de um modelo de classificação de crédito por parte do Banco na sua avaliação de candidatos a crédito, mas não entra em pormenores sobre o funcionamento do mesmo.

## 2.1.7 Exemplos de aplicação de modelos de classificação de crédito

Os métodos de classificação estudados têm inúmeras aplicações em diversas áreas de estudo, como por exemplo, medicina, biologia, ou economia. Aliás, um dos primeiros modelos de classificação aplicados, foi o de Fisher (1936), que introduziu a ideia de discriminar perante dois grupos de uma população, na estatística, através da análise discriminante. Esse autor pretendia distinguir duas variedades de iris com base no tamanho das plantas e distinguir a origem dos crânios através das suas medidas físicas.

Este sub-capítulo pretende introduzir alguns exemplos de utilização destes modelos de classificação em situações de classificação de crédito. No entanto, como foi difícil encontrar exemplos concretos de aplicação de modelos de crédito em Bancos e, muito menos, em empresas comerciais, alargou-se o campo de pesquisa, à aplicação de modelos de classificação a situações de previsão de falência, que muitas vezes são propostos como medidas de classificação de crédito. Behr & Guttler (2007) referem, até, que os bancos não têm incentivos a divulgar o seu processo de medida de risco de crédito.

Um dos exemplos mais conhecidos da aplicação de um modelo de classificação, à previsão de falência, é o desenvolvido por Altman (1968), que aplicou a AMD à classificação de empresas, tendo em conta a sua situação financeira. Este algoritmo tem como objetivo prever a possibilidade de uma empresa entrar em falência no futuro (sendo essa a variável dependente do seu modelo). Todavia, o modelo não permite prever quando uma empresa vai falir, ou seja, o método indica a possibilidade de uma empresa estar insegura financeiramente, com base na sua semelhança com empresas que entraram em falência, estudadas pelo modelo. No seu primeiro teste, percebeu-se que o modelo tem 72% de precisão em prever falência com dois anos de antecedência em relação ao evento. No entanto, em testes posteriores, descobriu-se que o

modelo tem uma precisão entre 80-90%, para o mesmo efeito, com um ano de antecedência. Também, Deakin (1972) aplica a AMD à previsão de falência de empresas.

Por outro lado, Ohlson (1980) utiliza a RL para criar um modelo de classificação que tem como objetivo prever a falência de empresas e Shumway (2001), também, utiliza a RL para o mesmo fim. Portanto, verifica-se que, com o passar do tempo, a RL foi sendo cada vez mais utilizada, em contexto de previsão de falência de empresas, por contrapartida da AMD que, segundo Ohlson (1980) apresenta várias desvantagens quando comparada com a regressão logística. No entanto, é também, nesta altura, que as RNA começam a aparecer como modelos a ser utilizados para problemas de classificação deste género.

Charitou et al. (2004) criam um modelo de previsão de falências de empresas, comparando o poder de previsão das RNA com o da RL, concluindo que ambos os métodos têm bons resultados, embora as RNA tenham, no geral, uma melhor taxa de classificação.

As RNA têm aplicação em muitas áreas de estudo como a engenharia, as ciências comportamentais, a biologia, etc. A sua aplicação a modelos de decisão de crédito é evidenciada por estudos desenvolvidos, por exemplo, por West (2000), que investiga a precisão de classificação de crédito de cinco RNA. Desai et al. (1996) também desenvolvem um estudo sobre a precisão de modelos de classificação de redes neuronais por comparação com outros modelos. Desai et al. (1996) aplicaram três métodos de classificação (RL, AMD, RNA) a três uniões de crédito, de forma a comparar os resultados de cada método. Os autores concluíram que as RNA têm melhores resultados que os outros métodos, quando a medida de desempenho são os casos de maus pagadores, bem classificados. No entanto, quando a medida de desempenho utilizada são os casos de bons pagadores e maus pagadores, bem classificados, a RL equipara-se às RNA.

Quanto a exemplos de aplicação de modelos de classificação a medidas de risco de crédito em bancos ou organizações comerciais, foi mais difícil encontrá-

los, tal como explicado anteriormente. No entanto, Viganò (1983) elabora um MCC para calcular a probabilidade de reembolso do crédito concedido, através da AMD, tendo como objetivo aumentar a qualidade do portfólio de crédito, através de uma melhor análise de crédito. Este modelo foi testado no Banco Nacional de Crédito Agrícola de Burkina Faso.

## 2.2 Problema de alocação de recursos

No sentido de resolver o problema de indicar um montante de crédito a conceder a cada cliente, foi necessário desenvolver alguma literatura sobre problemas de alocação de recursos e ferramentas que o permitem fazer. Assim, neste ponto da revisão de literatura faz-se uma introdução sobre o que é otimização, dá-se uma noção do que é a teoria da seleção de portfólio e introduz-se os algoritmos genéticos (AG), que servem de base a ferramentas de otimização.

Aranha & Iba (2007) descrevem o problema de otimização como um problema de alocação de recursos, ou seja, para um montante de recursos finito e um conjunto de possíveis aplicações para alocar esses recursos, existe uma alocação que maximiza uma dada função objetivo.

O problema de otimização de recursos está presente em muitas áreas de estudo. Por exemplo, Lin et al. (2013) aplicam um modelo de otimização, utilizando AG, para determinar o nível de recursos ótimo nos serviços cirúrgicos. Shangguan et al. (2002) propõe um modelo de otimização dos recursos regionais de irrigação de água, utilizando programação dinâmica. Também Senouci & Eldin (2004) aplicam um modelo de otimização de recursos à programação de projetos de construção, utilizando AG.

No caso de um portfólio financeiro, o processo passa por escolher uma alocação de recursos monetários para um conjunto de ativos, tendo como objetivo maximizar o retorno esperado, para um dado nível de risco, sendo que o retorno esperado é a função objetivo.

Markowitz (1952) propõe o método da média e variância para a otimização do problema de seleção do portfólio. No fundo, o retorno esperado do portfólio é a média ponderada do retorno esperado de cada ativo que o compõe e o risco esperado do portfólio é dado pela variância do seu retorno. O objetivo é maximizar o retorno esperado, para um dado nível de risco. Embora, o problema do presente trabalho não seja a seleção de um portfólio, este conceito será interessante numa fase mais adiantada no estudo.

Por outro lado, Venkatesan & Kumar (2004) desenvolvem um modelo de otimização de recursos de marketing, tendo como base a maximização da “Customer Lifetime Value” (CLV) dos clientes. Os autores partem de uma equação de CLV, que é função da frequência de compra esperada, da margem de contribuição esperada e dos custos de marketing, para definir uma alocação de custos de marketing que maximize a CLV. Utilizando AG, fizeram variar o nível de contacto a cada cliente, que tem repercursões nas variáveis da equação da CLV, de forma a maximizar a CLV de cada um e por isso, da sua carteira de clientes. Embora, este não seja o problema que se pretende resolver, o problema de definir um montante de crédito a conceder a um candidato de crédito do Grupo Nors passa por maximizar o retorno que se espera obter de cada cliente, minimizando os custos associados ao serviço de conceder crédito.

Dickinson (2001) também desenvolve um modelo para otimização do Valor Atualizado Líquido (VAL) de um portfólio de projetos interdependentes da empresa *The Boeing Company* utilizando um software derivado do *solver* do *excel*.

Existem várias técnicas para resolução de problemas de otimização de recursos, como por exemplo, programação linear, RNA, e algoritmos genéticos, de acordo com Senouci & Eldin (2004). No entanto, o presente trabalho apenas vai utilizar os algoritmos genéticos, para a resolução do problema em questão.

De acordo com Filho et al. (1994), “AG utilizam teorias evolucionarias biológicas para resolver problemas de otimização”. Segundo estas teorias, os elementos de uma população mais adequados são aqueles com maior

probabilidade de sobreviver e gerar descendência, transmitindo a sua biologia hereditária às novas gerações. Whitley (1994), afirma que os AG funcionam como funções otimizadoras, tendo uma grande variedade de aplicações.

Então, segundo Filho et al (1994), um AG compreende um conjunto de elementos individuais de uma população e um conjunto de operadores, biologicamente inspirados, definidos sobre a população. A nível informático, o AG mapeia um problema num conjunto de linhas, sendo que cada uma destas linhas representa uma potencial solução do problema. Assim, Filho et al. (1994), indicam que o AG funciona através de um ciclo de fases:

1. Criação de uma população de linhas
2. Avaliação de cada linha
3. Seleção das melhores linhas
4. Manipulação genética para criar uma nova população de linhas melhorada

Inicialmente, é definida uma população de linhas como um ponto de partida de busca da melhor solução. Cada elemento da população é codificado numa linha (cromossoma), para ser manipulado pelos operadores genéticos. Em segundo lugar, cada linha é avaliada, de acordo com o seu desempenho em respeitar as restrições impostas pelo problema. Depois, um mecanismo de seleção seleciona as linhas (cromossomas) com melhor desempenho para a manipulação genética, sendo esse o mecanismo responsável por garantir a sobrevivência dos melhores cromossomas. No fundo, o processo de manipulação genética passa por processar a informação genética dos genes dos “pais” da população atual, para produzir uma nova população de linhas. Esta pode recorrer a duas operações: *crossover* e mutação. *Crossover* é uma recombinação da informação genética de uma população. O processo de seleção associado ao *crossover* garante que estruturas genéticas especiais são retidas para gerações futuras. Estas estruturas especiais, chamadas de *building blocks*, representam as linhas de uma população com melhor desempenho. No entanto, o processo de

recombinação não garante que algumas destas estruturas não desapareçam da população, em detrimento de outras estruturas genéticas, o que pode conduzir a função a encontrar mínimos locais, nem consegue explorar secções de espaço de procura que não estejam representados na estrutura genética da população. Então, o processo de mutação funciona como solução para estes problemas, na medida em que o seu operador introduz nova informação genética na população, ao modificar aleatoriamente alguns dos seus *building blocks*, ajudando o algoritmo de procura a ultrapassar mínimos locais. Este processo cria novas estruturas, que não descendem de informação genética anterior, e por isso, representam novas secções do espaço de procura do algoritmo. Em suma, o processo de recombinação cruza informação genética de dois cromossomas para produzir novos cromossomas. Enquanto a mutação é um processo que, aleatoriamente, altera uma parte da informação genética de um cromossoma. Assim, cada ciclo produz uma nova população de soluções para ser avaliada. Estes ciclos são realizados até a solução ótima ser encontrada. Taleb (2012), discute um aspeto interessante, que se relaciona com este assunto, referindo que a evolução beneficia da aleatoriedade para provocar mudanças nas características das gerações seguintes, no sentido em que tudo o que está vivo ou é orgânico acaba por morrer, mas só depois de gerar descendentes com um código genético de alguma forma diferente do dos seus pais.

# Capítulo 3

## Metodologia

Neste capítulo, procura-se explicar o estudo e respetiva pesquisa desenvolvidos no presente trabalho e a metodologia adotada, bem como o critério de escolha dos casos investigados e os aspetos relacionados com a recolha e a análise de dados.

Este estudo tem em vista criar um modelo de classificação de crédito que permita auxiliar a área de CaR, do Norshare, na sua tomada de decisão de crédito e criar uma ferramenta para indicar o montante de crédito a conceder a cada cliente que realize uma candidatura de crédito. Nesse sentido, o presente trabalho pretende responder às seguintes perguntas: “Qual o método quantitativo que poderia ser usado como técnica de apoio à decisão de classificação de crédito no Grupo Nors e qual o seu desempenho?” e “Que modelo se pode aplicar para definir um montante de crédito a conceder a um candidato de crédito?”.

O presente trabalho incide sobre o caso das CaR do Grupo Nors porque no âmbito do estágio realizado foi apresentada a possibilidade de ficar encarregue de um projeto com vista a criar um instrumento de classificação de crédito e indicação de um montante de crédito a conceder.

Assim, o trabalho assenta numa metodologia de análise quantitativa que recorre aos métodos de classificação da RL, RNA e AMD, explicados na revisão de literatura, sendo a sua aplicação apresentada no capítulo de proposta do modelo. Também se utilizaram conceitos gerais sobre alocação de recursos, introduzidos na literatura, para desenvolver o modelo de definição do montante de crédito a conceder. No entanto, também se desenvolveu um capítulo destinado a esclarecer a situação atual da empresa, em termos de classificação de crédito, através de informação recolhida na fase inicial do estágio realizado na

empresa. Esta componente empírica, do presente estudo, pretende clarificar o funcionamento da análise de crédito das CaR, aproveitar o seu conhecimento e confrontar as suas práticas com a literatura existente, de forma a identificar oportunidades de melhoria e propôr soluções para esse efeito, sendo que, o fim último deste estudo, era, precisamente, criar uma solução para otimizar as decisões da análise de crédito e o montante de crédito concedido.

### 3.1 Procedimento de recolha de dados

No início do estágio e cumprindo o respetivo plano delineado pela empresa, concretizou-se a oportunidade de passar por quase todas as áreas do Norshare, nomeadamente, Contas a Pagar, Compras, Serviço ao Cliente, Contabilidade, Tesouraria e Contas a Receber, podendo observar as atividades desenvolvidas em cada área. Foi-me sendo, assim, explicado o trabalho desenvolvido, obtendo uma visão integrada do funcionamento do CSP, bem como das suas atividades mais recorrentes. Durante a passagem pelas CaR, foi possível tomar conhecimento dos processos desenvolvidos pela área, destacando-se a análise de clientes para concessão de crédito, a aprovação dos diversos tipos de negócios, e a previsão de imparidades de clientes. Esta fase permitiu recolher muita informação sobre o funcionamento da área, sendo que se procurou desde o início, detalhar todos os procedimentos incluídos na análise de clientes orientada para a concessão de crédito. Por outro lado, foram disponibilizados pelo supervisor do estágio, em ficheiro Excel, todos os registos de decisões de crédito tomadas pela área no período entre 1 de Julho de 2015 e 23 de Setembro de 2016 (data de início do projeto no âmbito do estágio, que começou pela análise das decisões tomadas no passado). Os dados compreendiam informação sobre cada cliente, como o nome e o seu código (atribuído pela empresa), data do pedido de crédito, o tipo de pedido, a empresa do grupo em que o cliente pretendia crédito, o *plafond* pretendido, o parecer das CaR e a sua justificação.

Foram também utilizadas outras fontes de informação secundárias como apresentações e documentos disponibilizados pela empresa.

## 3.2 Base de Dados

Tal como é descrito no capítulo 2, para desenvolver um modelo de classificação, é primeiro necessário recolher informação e prepará-la. Assim, criou-se uma base de dados com o objetivo de caracterizar os clientes de crédito num conjunto de variáveis que descrevessem adequadamente o problema de análise de crédito, para depois testar os diferentes métodos propostos. No processo de levantamento da informação, recolheram-se valores de muitas potenciais variáveis para cada cliente, para evitar repetir o processo mais tarde. No fundo recolheram-se 28 potenciais variáveis independentes, sendo que algumas correspondem às variações dos indicadores económicos e financeiros.

Os indicadores económico-financeiros podem assumir a forma de indicadores absolutos ou relativos. No entanto, é consensual entre os analistas financeiros que os indicadores absolutos apresentam limitações, devido ao efeito de escala, enquanto os indicadores relativos (rácios) constituem uma forma mais fácil de sintetizar a informação e de a utilizar para comparação de casos. Com efeito, apesar das limitações dos rácios ao nível de informação, sendo que estes apenas apresentam informação momentânea sobre a realidade das empresas, não descrevendo, por exemplo, a tendência de evolução, estes constituem uma técnica de análise fácil e rápida. Assim, muitos dos indicadores escolhidos para o presente trabalho assumem a forma de rácios ou quocientes, embora também se utilizem indicadores absolutos e a sua variação entre dois anos, numa tentativa de compreender a sua evolução de curto prazo.

È, também, importante referir que os indicadores económico-financeiros podem ser baseados em informação contabilística ou em informação do mercado de capitais. Segundo Taffler & Agarwal (2008), os rácios baseados em informação

do mercado de capitais têm maior poder de previsão de uma empresa entrar em incumprimento com uma obrigação ou entrar em falência. Isto porque, em mercados eficientes, os rácios baseados no mercado refletem toda a informação contabilística e não contabilística, além de que estes rácios refletem expectativas futuras de cash flow e, por isso, têm um maior poder para propósitos de previsão. Por outro lado, os rácios contabilísticos baseiam-se em informação passada e podem ou não ter capacidade adequada a prever o futuro. Além disso, os rácios contabilísticos são mais suscetíveis a manipulação por parte da gestão da respetiva empresa.

Mensah (1984) também sugere que as características da envolvente económica das empresas alvo do estudo podem influenciar o estado financeiro das mesmas, sendo que esta envolvente está em constante alteração. Além disso, o autor conclui que a precisão dos MCC varia consoante diferenças no ambiente económico, o que pode ser ultrapassado atualizando os modelos frequentemente, e que alguns modelos de classificação são mais apropriados para determinados setores de indústria.

No entanto, no caso do presente estudo, a maior parte das empresas que constituem a amostra são de pequena dimensão e por isso, não têm acesso ao mercado de capitais. Desta forma, não seria possível colecionar informação sobre rácios baseados no mercado de capitais para grande parte das empresas, pelo que se usaram apenas rácios contabilísticos. Além disso, muitos dos modelos aplicados anteriormente apenas utilizam variáveis contabilísticas e obtiveram bons resultados, como por exemplo, o Z-Score apresentado por Altman et al. (1977), para empresas não negociadas publicamente no mercado de capitais.

De forma a caracterizar a situação económica e financeira dos clientes da empresa Nors, candidatos a crédito, para perceber o seu comportamento enquanto pagadores, escolheu-se um conjunto de indicadores económicos e financeiros. Para o presente trabalho optou-se por levantar um conjunto de variáveis, de forma a poder testar a sua significância para a variável que se

pretende prever e testar diferentes combinações de variáveis nos modelos a aplicar.

Assim, na seleção dos indicadores a incluir no presente estudo, tentou-se incluir indicadores variados que caracterizem a empresa de pontos de vista diferentes, indicadores já utilizados na análise de crédito da empresa ou sugeridos pelos profissionais que realizam a análise de crédito nas CaR da empresa e indicadores estudados pelo autor do presente trabalho, ao longo da parte curricular do mestrado.

Procurou-se, então, caracterizar todos os clientes que pediram crédito com base no seguinte conjunto de potenciais variáveis independentes (para os modelos), sendo que, para todos os indicadores económico-financeiros, foi retirada informação dos dois anos presentes nos relatórios financeiros, bem como a sua variação:

- Autonomia Financeira
- Grau de Rotação do Ativo
- Rentabilidade Operacional do Ativo
- EBITDA
- Fundo de Maneio
- Necessidades de Fundo de Maneio
- Tesouraria Liquida
- Capital Próprio
- Ativo Total
- Volume de Negócios
- Resultado Operacional
- Data da informação financeira
- Idade da empresa
- Histórico interno de pagamentos do cliente
- Montante de processos jurídicos abertos contra o cliente (últimos 5anos)

- Relação entre o montante de processos jurídicos abertos contra o cliente e o Capital Próprio (Montante processos jurídicos/ Capital Próprio)
- Duração da relação comercial com cliente

Desta forma, tentou-se estabelecer um conjunto variado de indicadores inseridos em categorias diferentes, de forma a diversificar o foco de análise.

O indicador Autonomia Financeira (AF) representa a capacidade de endividamento da empresa e pode ser visto como indicador do seu grau de independência financeira. Foi selecionado por ser um dos rácios a que as CaR dão mais importância, aquando da decisão de conceder crédito e por ser importante que as empresas tenham capacidade de endividamento de reserva para fazer face a situações mais complicadas.

O Grau de Rotação do Ativo (GRA) representa a capacidade do capital investido gerar volume de negócios, ou seja, indica a capacidade produtiva da empresa para aproveitar o seu ativo para gerar atividade económica. Desta forma, o GRA foi selecionado por se tratar de um importante indicador da gestão de ativos de uma empresa, sendo um dos indicadores sugeridos pelo aluno.

A Rentabilidade Operacional do Ativo (ROA) representa o resultado operacional por unidade de capital investido, ou seja, é uma medida da eficiência da empresa de gerir e afetar o seu ativo para gerar atividade económica. É diferente do GRA, no sentido em que inclui na sua análise a estrutura de custos da empresa. Foi selecionado por se tratar de um importante indicador de rentabilidade, independente de decisões de financiamento, já que não diferencia entre capitais próprios e alheios e por se tratar de um dos indicadores mais importantes para comparação de desempenho económico de empresas do mesmo setor, tendo sido, também, sugerido pelo aluno.

A Tesouraria Líquida (TL) é um indicador de liquidez que se traduz pela diferença entre o Fundo de Maneio (FM) e as Necessidades de Fundo de Maneio (NFM). O FM consiste na diferença entre o ativo circulante e o passivo circulante, ou seja, corresponde a uma margem de segurança que garante que a empresa

tem capacidade de liquidez para fazer face aos compromissos de curto prazo. As NFM revelam as necessidades de financiamento suscitadas pelo ciclo de exploração, isto é, são as necessidades de financiamento que decorrem da necessidade de realizar pagamentos (Fornecedores, Estado, Pessoal, etc) antes de se receber dos clientes. Assim, a TL corresponde à posição de tesouraria uma vez satisfeitas as necessidades de fundo de maneiio, sendo que a sua utilização foi sugerida pela empresa. As NFM e o FM também se introduzem como indicadores a utilizar, por também se tratarem de indicadores de liquidez e serem utilizados pela empresa aquando da sua análise.

Os indicadores absolutos EBITDA, Capital Próprio (CP), Ativo (A), Resultado Operacional (RO) e Volume de Negócios (VN) foram incluídos como indicadores complementares aos já apresentados e por serem indicadores importantes para a análise de crédito da Nors. Além disso, o A é um dos indicadores mais utilizados em estudos de previsão de falência através de métodos de classificação, conforme Jo et al. (1997).

Por outro lado, como a informação dos relatórios financeiros das empresas é referente a anos diferentes, porque a maior parte das empresas não coloca os novos relatórios financeiros ao mesmo tempo, procurou-se incluir esse desfasamento temporal no estudo, através de uma variável categórica, de forma a testar a sua significância na variável a prever.

As variáveis idade da empresa e duração da relação comercial da empresa com o Grupo Nors também foram escolhidas por serem fatores aos quais a análise de crédito das CaR dá atenção, aquando da tomada de decisão.

Importa ainda justificar a escolha da variável histórico interno de pagamentos, que é um indicador importante na análise de crédito, segundo a experiência da empresa, bem como o montante de processos jurídicos em aberto contra a entidade em relação ao capital próprio, que também é um indicador tido em conta nos pareceres internos da empresa. A variável histórico interno teve de ser categorizada, de forma a ser processada pelos métodos utilizados. Então definiu-

se que esta variável tem três categorias que refletem o comportamento de pagamentos passado do cliente:

- 1 – histórico positivo
- 2 – sem histórico
- 3 – histórico negativo

O critério utilizado para atribuir uma categoria de histórico a cada cliente foi o mesmo que o utilizado para definir a variável dependente (resultado do crédito), ou seja, se paga até 30 dias (pós data de vencimento) tem categoria 1, se não tem histórico pertence à categoria 2, se paga pós 30 dias tem categoria 3. Este critério será explicado mais à frente.

Por último, escolheu-se a variável montante monetário envolvido nos processos jurídicos abertos contra a entidade (que está a pedir crédito), por ser uma das variáveis mais importantes para as CaR, sendo que, como na sua análise comparam este montante com o CP, também se utilizou a variável montante envolvido em processos jurídicos abertos contra a entidade em relação ao Capital Próprio. Esta variável será designada por “Montante/CP” nos modelos de classificação apresentados no capítulo 4.

Uma vez feito o levantamento de todas as potenciais variáveis independentes para “alimentar os modelos”, foi necessário definir o resultado que se pretendia que os modelos a desenvolver estimassem, ou seja, qual o *output* ou *outputs* pretendidos (variável dependente). Note-se que os modelos de classificação a aplicar são modelos de aprendizagem supervisionada. Assim, sendo o objetivo do estudo classificar os clientes, de acordo com a sua capacidade de pagamento de crédito, nas empresas do grupo, definiu-se como *output* a qualidade do cliente enquanto pagador, através de duas categorias, ou seja, se este virá a ser um bom pagador ou mau pagador (nos modelos apresentados no capítulo 5, esta variável será designada como “resultado de crédito”). Para o efeito, incluiu-se a visão de alguns técnicos das CaR, que lidam com as decisões de crédito, sobre o que é um bom pagador e um mau pagador, no sentido de definir um critério que os

distinguisse. A área entende que geralmente um bom pagador é um cliente que paga até 30 dias pós data de vencimento. Então procurou-se definir a variável dependente (*output*) com base no comportamento de pagamentos, pós a data da decisão de crédito, de forma a perceber a sua categoria de pagador.

Com base nesta informação foram testados vários métodos e criaram-se vários modelos de classificação, com resultados diferentes, de forma a apresentá-los á empresa, para esta analisar a sua potencial utilidade e discutir os resultados dos diferentes modelos.

Posteriormente, contudo, a empresa considerou interessante criar um modelo de classificação em 3 categorias de pagador, ou seja, distinguir os clientes enquanto bons pagadores, razoáveis ou maus pagadores. No entanto, esta parte do trabalho insere-se em análise exploratória, porque, apenas, as RNA permitem uma classificação em três categorias, enquanto os restante métodos propostos não, sendo que o presente trabalho tem como objetivo comparar o desempenho dos três métodos propostos. Para esse efeito definiu-se o *output* com base no comportamento do cliente pós data da decisão de crédito, sendo que os critérios utilizados foram os seguintes:

- Bom pagador – paga até um máximo de 20 dias, após a data de vencimento
- Razoável - paga entre 20 e 80 dias após data de vencimento
- Mau – paga atrasado em mais de 80 dias relativamente à data de vencimento

Revela-se, no entanto, que em alguns casos não se adotou um rigor excessivo nesta classificação dos dados, quando se tratava de apenas 1 ou 2 dias de diferença.

No decorrer do processo de preparação dos dados, muitos clientes foram excluídos por não terem a informação pretendida. No fundo, cerca de 55% dos pedidos de crédito que ocorreram, durante o período analisado, foram excluídos, sendo que, durante este período foram realizados cerca de 2300 pedidos de crédito. Por exemplo, muitos dos clientes da empresa não têm informação financeira pública e por isso, não puderam ser alvo do estudo. Além disso, optou-

se por excluir todos os clientes com CP superior a 100 milhões, porque muitas vezes estes clientes, apesar dos seus bons indicadores financeiros, acabam por pagar atrasado dado que são muito grandes e a Nors não lhes rejeita crédito, o que se traduz num efeito que não é pretendido incluir nos modelos de classificação.

### 3.3 Outras considerações

Após a análise da informação, procurou-se aplicar os modelos estudados na revisão de literatura para apoiar as decisões de crédito e montantes de crédito a conceder, sendo que esses são os problemas que este trabalho pretende resolver.

Assim, foram utilizados programas estatísticos que foram aplicados aos registos de crédito da empresa, de forma a medir a sua capacidade para resolução do problema, comparar os seus resultados e aplicá-los ao contexto real. Desta forma, foram utilizados dois programas diferentes.

O *DecisionTools Suite* da *Palisade* engloba várias ferramentas de análise de risco e tomada de decisão, como por exemplo, o *NeuralTools*, que desenvolve RNA, o *StatTools*, que permite a realização de testes estatísticos como a AMD e RL, o *@Risk* e o *Evolver*, que utilizam AG para resolução de problemas de otimização, entre outros. O “*Evolver*” é uma aplicação para otimização de problemas determinísticos, enquanto o “*RiskOptimizer*” do programa *@Risk* se destina a otimizações que envolvem a simulação de *Monte Carlo* de distribuições probabilísticas.

Também foi utilizado o programa *SPSS Statistics* da *IBM* que engloba várias ferramentas de análise estatística e previsão, como por exemplo, as RNA.

Os modelos de classificação de crédito de RNA, apresentados no capítulo 4, são elaborados a partir do programa *SPSS*. Isto porque o programa *SPSS*, além de prever a categoria de pagador a que um cliente pertence, também tem como resultado a probabilidade do cliente pertencer a uma determinada categoria de

pagador à escolha. Isto permitiu que se integrasse no modelo de indicação do montante de crédito a conceder, a análise das RNA, como se verá no capítulo de proposta do modelo. Além disso, o programa *SPSS* inclui mais opções para o desenvolvimento das RNA, por comparação com o programa da *Palisade*, nomeadamente as funções de ativação a utilizar, o tipo de treino da rede e algoritmo de treino a utilizar. No entanto, também se elaboraram RNA, a partir do programa da *Palisade*, embora não sejam apresentadas no presente trabalho, sendo que o desempenho dos dois programas é semelhante.

### 3.4 Limitações do Método

Em primeiro lugar, é necessário referir que as empresas alvo do estudo foram escolhidas por conveniência, dado que foram as empresas que fizeram pedidos de crédito no último ano e que apresentaram a informação financeira necessária para serem alvo do estudo.

Em segundo lugar, como os modelos foram construídos com informação disponibilizada em relatórios financeiros da Informa D&B, estes apenas podem ser aplicados a clientes com a mesma informação disponível, nomeadamente, informação sobre as variáveis utilizadas pelos modelos. Por exemplo, os modelos não puderam ser construídos com clientes de países estrangeiros, porque os relatórios financeiros da Informa D&B de empresas de países estrangeiros não possuem a mesma informação que os das empresas portuguesas e por isso, os modelos de classificação propostos apenas poderão ser aplicados a empresas portuguesas.

Por último, todos os modelos de classificação de crédito enfrentam o problema da inferência dos rejeitados. Este problema prende-se com o facto de que só as decisões de crédito aprovadas e concretizadas no passado, permitem que o cliente seja classificado como bom ou mau pagador, tendo em consideração o seu comportamento de pagador, enquanto as decisões rejeitadas não permitem

classificar o cliente como bom ou mau pagador, uma vez que não existe registo do seu comportamento de pagador. Portanto, se os clientes rejeitados forem retirados da amostra, sobre a qual os modelos vão ser desenvolvidos, irá ser criado um “*bias*”, quando os modelos forem utilizados para classificar um cliente anteriormente rejeitado. Assim, existem várias técnicas para ultrapassar esse problema, como por exemplo, a extrapolação e a “*augmentation*”. No entanto, no presente trabalho, optou-se por utilizar uma das técnicas mais simples, muitas vezes criticada, que consiste em assumir os rejeitados como maus pagadores. Procedeu-se da seguinte forma, porque os casos rejeitados representam o típico caso de cliente ao qual o Grupo não pretende dar crédito sendo que, como o modelo está a ser desenvolvido para este Grupo, em particular, faz sentido proceder de acordo com as suas preferências.

# Capítulo 4

## O Grupo Nors

Neste capítulo pretende-se descrever, brevemente, a história do grupo Nors, a organização dos Serviços Partilhados, e por fim, descrever a análise de crédito das CaR, de forma a contextualizar a realidade em que se insere o presente trabalho.

### 4.1 História do Grupo

A história do Grupo Nors começa em 1933, quando Luís Jervell inicia a representação da Volvo, em Portugal. No entanto, só em 1949, nasce a empresa Auto Sueco, fundada por Luís Jervell e Yngvar Poppe Jensen, como representante e importadora dos automóveis e camiões da marca Volvo. Ao longo do tempo, a história da empresa fica marcada por vários episódios, sendo importante destacar alguns, para contextualizar a realidade da organização, nos dias de hoje.

Em 1959, é fundada a Auto Sueco Coimbra, atualmente conhecida como Grupo Ascendum, como representante da marca em seis distritos, do centro do país. Todavia, só em 1965, a entidade passa a representar a marca Volvo em todo o país, ao abrir a sua primeira filial em Lisboa.

Por outro lado, a internacionalização do grupo dá-se em 1991, com a criação da Auto Sueco Angola. Também, nessa década, o Grupo estende-se a Espanha e a partir de 2000, começa a expandir a sua atividade para África e Brasil.

Hoje em dia, o grupo Nors é um grupo com um volume de negócios de 563.263 milhares de euros (2015), que tem como visão ser um dos líderes mundiais em soluções de transporte e equipamentos de construção e está presente em 3 continentes e 15 países, sendo que cada um destes países tem necessidades diferentes e requer um conhecimento específico. Assim, a Nors opta por dividir

o seu negócio geograficamente em 6 áreas organizacionais (ver Figura 3), que servem cada um destes mercados e funcionam de forma diferente:

Região Ibéria - engloba Portugal e Espanha e comercializa Camiões, Autocarros, Equipamentos Industriais e Peças

Região Angola (Angola) – destinada a Angola e comercializa Camiões, Autocarros, Automóveis, Equipamentos Industriais e de Construção, Peças e Vidro de Construção

Região Brasil – engloba o Brasil e comercializa Camiões, Autocarros e Máquinas Agrícolas

Região África – engloba Namíbia, Tanzânia, Botswana, Quênia, Uganda e Moçambique e comercializa Camiões, Autocarros, Equipamentos Industriais e de Construção e Peças

*Ventures* – destinada a Portugal e comercializa Automóveis, Indústria, Seguros e Centros de Inspeção

*Ascendum* – engloba Portugal, E.U.A., México, Turquia, Espanha, Europa Central (República Checa, Roménia e Áustria) e comercializa Equipamentos Industriais e de Construção, Camiões, Automóveis, Máquinas Agrícolas, Equipamentos para Aeroportos, Ferrovias, Estruturas Portuárias e Peças.

Neste caso, é fácil perceber que apenas a área *Ventures* e *Ascendum* servem mercados comuns ao de outras áreas. A área *Ventures* oferece serviços que a área Ibéria não oferece ao mercado português, como por exemplo, seguros. A área *Ascendum*, representa o grupo *Ascendum* e a sua área operacional que está estabelecida em diversos países, oferecendo soluções diferentes como, por exemplo, equipamentos para aeroportos.



**Figura 3:** Organização Funcional Nors  
**Fonte:** Documento Interno da Empresa

Em 2013, o mercado com um maior volume de vendas foi o mercado do Brasil, com uma percentagem de 29% do volume total de negócios. Foi seguido do mercado português, com uma percentagem de 24%, e do angolano, com uma percentagem de 14%.

Quanto ao seu negócio, a Nors opta por dividi-lo em 4 áreas: *Original Equipment Solutions*; *Integrated Aftermarket Solutions*; *Recycling Solutions*; e *Safekeeping Solutions*.

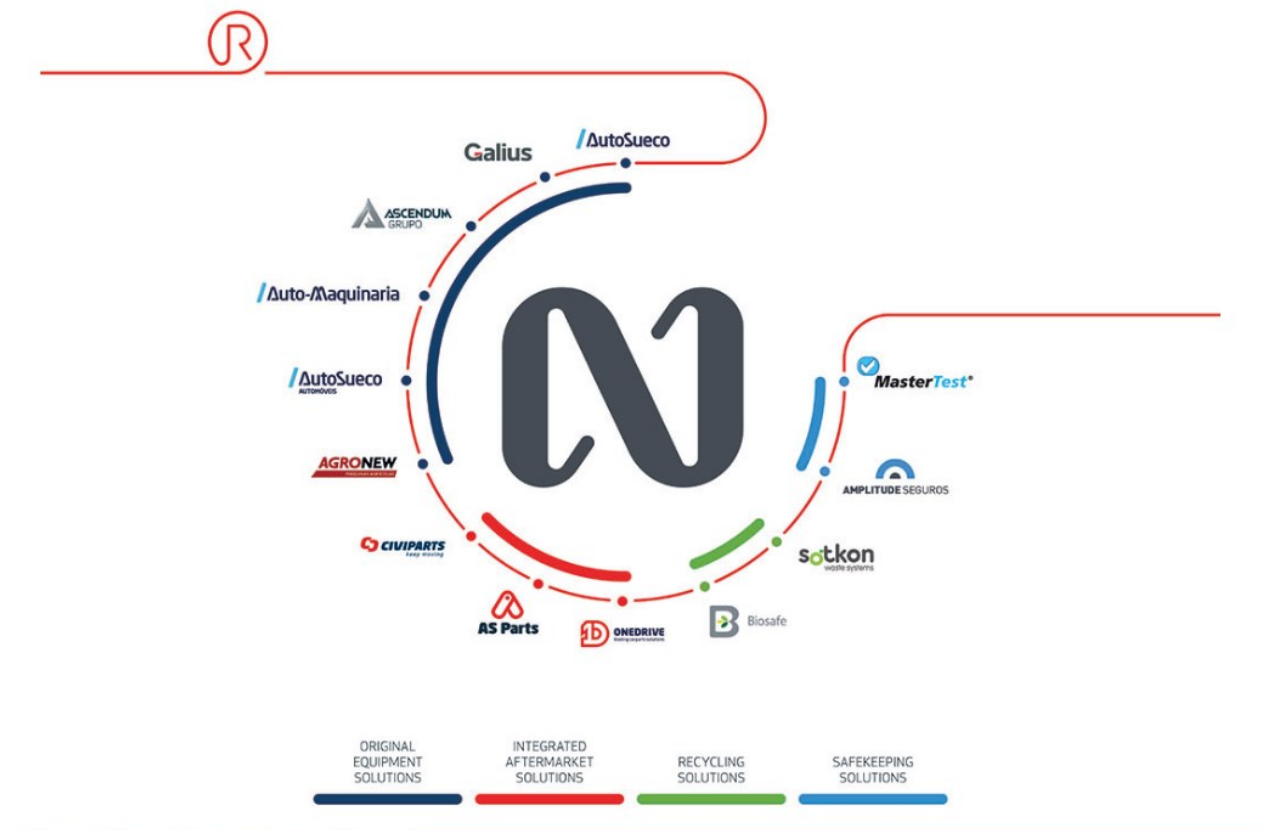
Assim, a área *Original Equipment Solutions* representa a venda de equipamentos originais de camiões, equipamentos de construção, automóveis, autocarros, motores marítimos e indústrias, geradores e equipamentos agrícolas, e oferece soluções de marcas como a Volvo, a Renault Trucks, a Mazda e a SDMO.

A área *Integrated Aftermarket Solutions* pretende oferecer soluções pós-venda aos clientes, sendo que a área se subdivide em pós venda de pesados e ligeiros. Desta forma, a AS Parts e a One Drive integram as soluções de ligeiros e a CiviParts integra a área de soluções para pesados.

A área *Recycling Solutions* oferece soluções de reciclagem e equipamentos de reciclagem, tendo como objetivo um compromisso com o futuro. Esta área conta com a marca Biosafe que oferece soluções de reciclagem, como o granulado de borracha e a Sotkon que oferece equipamentos de reciclagem de limpeza urbana.

Por último, a área *Safekeeping Solutions* oferece serviços de seguros e de inspeção de veículos, através das marcas Amplitude Seguros e Mastertest, respetivamente.

A Figura 4 evidencia a organização da Nors por área de negócio e as empresas do grupo, presentes em cada área.



**Figura 4:** Organização do Negócio Nors  
**Fonte:** Documento Interno da Empresa

Desta forma, o Grupo consegue oferecer soluções aos seus clientes que se complementam, visto que, a empresa atua na venda de peças originais, de serviço pós venda e ainda na venda de seguros e inspeções, além das soluções sustentáveis de reciclagem.

Quanto às vendas por produto, em 2013, os produtos com maior percentagem no volume de negócios do Grupo foram os equipamentos de construção, com uma percentagem de 41%, seguidos dos camiões, com uma percentagem de 40% e dos componentes, com uma percentagem de 7%.

## 4.2 Serviços Partilhados da Nors

Quinn et. al.(2000) definem Serviços Partilhados como “a prática pela qual as unidades de negócio de empresas e organizações decidem partilhar um conjunto de serviços, de forma a servir as várias unidades de negócio da mesma maneira e impedir a duplicação de uma série de funções”.

Para Bergeron (2002) um centro de serviços partilhados é uma estratégia que transforma um subconjunto de funções de negócio existentes numa nova unidade de negócio semiautónoma. Esta tem uma estrutura de gestão desenhada para promover eficiência, criação de valor, redução de custos e melhorar a qualidade do serviço prestado aos clientes internos da organização, atuando de forma independente, competindo no mercado.

Bergeron (2002) enumera um conjunto de potenciais benefícios dos CSP. Este autor divide-os em benefícios na perspectiva da organização mãe e na perspectiva da própria unidade de serviços partilhados.

Assim, os potenciais benefícios, na perspectiva da empresa mãe, são a redução de custos, a melhoria do serviço, o menor número de distrações, que permite um maior foco nas competências e atividades *core* e o potencial de criar um centro externo focado na lucratividade.

Já na perspectiva da própria unidade de suporte, os potenciais benefícios são o aumento da eficiência que se traduz numa melhor qualidade de serviço, a preços comparáveis ou menores, a padronização de processos e aplicação das corretas tecnologias, o menor número de pessoas para desenvolver o mesmo serviço, devido à capacidade de concentrar e focar recursos em determinados propósitos

e o aumento de economias de escala, devido à concentração de recursos especializados.

De facto, apesar de existirem várias definições sobre Serviços Partilhados, todas elas têm em comum a prática de consolidação de funções e processos, comuns a toda a organização, numa unidade de suporte que apoia as restantes unidades de negócio, evitando a duplicação de atividades e processos. Além disso, as principais características deste modelo de gestão são a agregação de processos como forma de reduzir custos, a especialização nas atividades para aumentar a qualidade do serviço e a gestão orientada para os clientes. Os CSP podem abranger várias funções como, por exemplo, finanças, contabilidade, recursos humanos, auditoria, jurídico e compras.

O CSP do Grupo Nors tem como objetivo concentrar atividades que são comuns a várias empresas do Grupo, como forma de especializar recursos e reduzir custos. Este foi criado em 2007 e emprega cerca de 100 pessoas, estando sediado no Porto. Neste momento, o CSP concentra atividades financeiras (Contabilidade, Tesouraria, Contas a Receber, Contas a Pagar e Serviço ao Cliente) e atividades de suprimento (Compras, Gestão de Pessoal e Qualidade e Ambiente e SHST)

Desta forma, cada área exerce um conjunto de atividades para todas as empresas do grupo, que se inserem em geografias diferentes e que, do ponto de vista do Norshare, são chamadas de clientes.

### 4.3 Análise de Crédito das Contas a Receber

A Política de Finanças do Grupo Nors é da responsabilidade do CFO ao qual respondem, funcional e hierarquicamente, a Direção Financeira e o Norshare. Um dos tópicos abordados na Política de Finanças do grupo é a Gestão do Fundo de Maneio. O Fundo de Maneio está diretamente relacionado com o equilíbrio financeiro da empresa, na medida em que implica uma gestão integrada e

eficiente ao nível de clientes, inventários e fornecedores, de modo a garantir que o grupo gera cash flow necessário para financiar as suas atividades, minimizando as necessidades de financiamento externo. O presente estudo está relacionado com a gestão de clientes e por isso, é importante refletir sobre as normas do Grupo, no que toca a esta matéria.

De acordo com o relatório de Política de Finanças da empresa, o Grupo Nors privilegia todas as formas de pagamento que não constituem risco de crédito, ou seja, a concessão de crédito a clientes deve ser encarada como extraordinária, especialmente na venda de chassi/equipamentos. A gestão de clientes deve garantir a mitigação do risco de crédito a que a Nors está exposta, garantindo que os prazos de recebimento estabelecidos dos clientes são cumpridos, não afetando o equilíbrio financeiro do grupo.

As seguintes iniciativas devem ser realizadas no sentido de monitorizar o risco de crédito:

- Gestão dos prazos de recebimento;
- Atribuição de classes de risco a clientes de acordo com a informação interna e externa disponível;
- Definição de limites de crédito concedidos a clientes;
- Monitorização da evolução do nível de crédito concedido;
- Obtenção de garantias de crédito de clientes, sempre que o risco inerente e a situação financeira de determinado cliente assim o recomende;
- Realização de análise de imparidade das contas a receber.

Quanto à análise de crédito desenvolvida pelas Contas a Receber, esta pode surgir em 3 situações diferentes.

Em primeiro lugar, devem ser submetidos à aprovação de crédito todos os negócios de venda de prestação de serviços e componentes (pós venda) que recorram a financiamento interno, ou seja, todos os negócios do negócio pós venda que tenham associado um limite de crédito atribuído.

Em segundo lugar, devem ser submetidos à aprovação de crédito todos os negócios de venda de chassi ou equipamentos que recorram a financiamento interno, ficando isentos aqueles que recorram a financiamento externo ou pronto pagamento.

Em terceiro lugar, devem ser submetidas à análise de crédito as propostas de concurso público.

O presente estudo vai apenas analisar as normas da análise de crédito do pós venda, porque é nesse contexto que se pretende construir um Modelo de Classificação de Crédito. Além disso, vão ser apresentados outros mecanismos do Grupo, mais especificamente das Contas a Receber, no sentido de mitigação do risco de crédito e monitorização deste.

O processo de atribuição de crédito a clientes é iniciado pelo preenchimento de uma Ficha de Avaliação de Crédito por parte das Unidades de Negócio (UN) com as informações fornecidas pelos clientes, sendo esta enviada posteriormente para as Contas a Receber do Norshare, para realização de uma análise ao cliente. Estas fichas devem ser asseguradas pelas Contas a Receber, bem como a sua revisão e atualização, e devem prever campos como a identificação e caracterização do cliente, condições de pagamento pretendidas, entre outros. A análise de crédito ao cliente deve recorrer a informação segura e fidedigna, tanto a nível interno, como a nível externo, de forma a providenciar um parecer quanto ao crédito a conceder.

Desta forma, as Contas a Receber podem receber pedidos de crédito de vários tipos, sendo eles: pré-análise, recomendação, reanálise, abertura e reabertura. Todos os pedidos são, geralmente, submetidos pelos comerciais das UN's. Grande parte dos pedidos são do tipo abertura ou reabertura, sendo que, os pedidos do tipo abertura são pedidos para um cliente ter crédito numa determinada empresa do grupo, pela primeira vez, enquanto, os pedidos reabertura são, como o próprio nome indica, pedidos de clientes reincidentes, sendo que o encerramento do seu crédito pode-se dar por falta de atividade de

compra ou por falha de pagamentos, ou seja, estes são pedidos de clientes que já tiveram crédito na empresa do Grupo à qual estão a submeter o pedido de crédito.

Assim, a análise de crédito tem como objetivo perceber se é viável conceder, a um determinado cliente, de uma certa empresa do grupo, a possibilidade de comprar a crédito, com base nas suas características.

A análise de crédito desenvolvida pelas Contas a Receber é do tipo tradicional, basendo-se numa avaliação humana subjetiva, de acordo com a literatura. Isto é, os analistas financeiros da área tomam a decisão de conceder crédito a um cliente com base na sua experiência e num conjunto de critérios, mas não recorrendo a nenhum modelo que determine ou apoie a sua decisão final. Isto significa que, apesar do analista recorrer a critérios previamente definidos, tem liberdade para fazer uma avaliação subjetiva em cada caso e tomar a decisão. Além disso, este conjunto de critérios, previamente definidos, permite que o analista esteja habituado a avaliar os clientes com os mesmos indicadores, o que permite poupar tempo de análise e uniformizar as suas decisões.

A análise desenvolvida pelos analistas financeiros é constituída por duas fases. Primeiramente, o analista financeiro analisa o cliente recorrendo a informação externa, utilizando um conjunto de indicadores, alguns do tipo económico-financeiro e outros mais associados com o histórico judicial da empresa em questão. A área recorre, geralmente, a relatórios financeiros elaborados pela Informa D&B (empresa especializada em informação sobre empresas), podendo recorrer a outras fontes de informação, sempre que considerar apropriado.

Os indicadores mais frequentemente utilizados na análise são a autonomia financeira, o capital próprio, o grau de rotação do ativo, o volume de negócios, EBITDA, solvabilidade e indicadores de risco elaborados pela própria Informa D&B, incluídos nos relatórios (no entanto, não se conhece o processo de cálculo destes indicadores de risco). No entanto, em certos casos menos comuns, os

analistas recorrem a outros indicadores. Além disso, são analisados os incidentes judiciais em aberto contra o cliente (montante do incidente é comparado com o seu Capital Próprio) e também, os administradores do cliente e suas ligações, de forma a perceber se existem ligações com outras empresas com processos judiciais em aberto ou processos de insolvência.

Após esta análise, é analisada a informação interna do cliente, ou seja, o histórico interno de pagamentos do cliente. Desta forma, o analista consegue perceber se o cliente já comprou a crédito, se teve episódios de cheques devolvidos e se, em algum momento, já se atrasou nos pagamentos. Assim, a decisão de conceder crédito passa pela situação financeira da empresa e pelo seu histórico de pagamentos, sendo que, a última costuma ter um grande peso na decisão final.

A Tabela 1 sintetiza a responsabilidade da atribuição e aprovação dos limites de crédito na venda de prestação de serviços e componentes (ver quadro 4):

Limites de crédito	Atribuição de Limite de Crédito	Prazo de Renegociação da dívida	
		< Z dias	>Z dias
<X	Norshare	Norshare	CA da Região
>X	CA da Região	CA da Região	CA da Região

**Tabela 1:** Responsabilidade de atribuição de crédito

**Fonte:** Documento Interno da Empresa

Este X e Z devem ser definidos pelo CFO para cada Região.

Em qualquer processo de renegociação não deve ser permitido que o cliente usufrua do crédito, salvo em situações excepcionais e de acordo com a autorização das Contas a Receber do Norshare.

No caso de o Diretor da UN não concordar com a decisão de aprovação/rejeição do Norshare, pode levar o caso ao Conselho de Administração (CA) da região, acompanhado do parecer das CaR, independentemente do limite de crédito.

É, também, da responsabilidade das CaR categorizar os clientes em relação ao risco de crédito, de acordo com informação interna e externa. A tabela 2 sumariza as diferentes classes de risco:

<b>Designação</b>	<b>Crédito</b>	<b>Descrição</b>
Normal	Sim	Limitado ao <i>plafond</i> atribuído e aprovado
Excecional	Sim	Crédito ilimitado, sem controlo de <i>plafond</i>
Acompanhado	Sim	Implica anterior suspensão e reabertura recente de crédito com <i>plafond</i> mais baixo
Vendas PP	Não	Cliente sem histórico de crédito. Código com o qual os clientes “nascem”
Suspensão	Não	Atribuído a clientes com histórico de atraso de pagamentos
Cortado	Não	Atribuído a clientes com histórico de cheques devolvidos
Pré Contencioso	Não	Cliente interpelado pelo advogado, extrajudicialmente. Este código é temporário, com duração máxima de 30 dias
Contencioso	Não	Cliente com processo judicial em curso

**Tabela 2:** Classes de Risco de Crédito  
**Fonte:** Documento Interno da Empresa

A revisão das classes de risco e dos limites de crédito atribuídos aos clientes podem acontecer no decorrer da atividade operacional, dependendo do comportamento dos clientes com os seus compromissos com o grupo, bem como, quando nova informação, interna ou externa, indica essa necessidade. Além disso, é da responsabilidade das Contas a Receber a revisão periódica, de dois em dois anos, das classes de risco e limites de crédito, sendo que apenas esta área deve ter acesso para alteração dos códigos de risco dos clientes.

Os clientes excecionais surgem devido à complexidade e impacto das relações comerciais com determinados clientes, sendo que é o CFO que define quais os clientes excecionais:

- Empresas do Grupo;
- Entidades Financeiras;
- Companhias de Seguros;
- Empresas detidas pelos Sócios;

- Fornecedores Core;
- Clientes estratégicos.

A definição dos clientes estratégicos é da responsabilidade do CA da Região que deve privilegiar os clientes transversais a todas as regiões do grupo e com volume de compras mais significativos. Além disso, é da responsabilidade do CA da Região assegurar a comunicação, de quais os clientes estratégicos, às Contas a Receber, bem como qualquer decisão de risco, crédito ou cobrança com estes.

As Contas a Receber desenvolvem ainda outras atividades no sentido de minimização do risco de crédito, nomeadamente, o processo de cobranças e a realização de reuniões de crédito.

Quanto ao processo de cobrança da Nors, este segue quatro fases.

Em primeiro lugar, a equipa de cobranças das Contas a Receber deve contactar e registar todos os contactos com os clientes que apresentem dívidas vencidas.

Em segundo lugar, as Contas a Receber devem proceder à suspensão do crédito do cliente, a partir de determinado tempo de mora definido no Regulamento da Região pelo CA da Região. Nesta fase, caso se proceda à suspensão do crédito, o cliente deve ser alertado formalmente.

Em terceiro lugar, as Contas a Receber devem entrar em contacto com o cliente para analisar a sua situação de crédito e perceber os motivos do incumprimento. No caso de se tratar de uma situação de indisponibilidade financeira temporária ou definitiva, as Contas a Receber devem negociar um novo plano de pagamentos com o cliente, respeitando os prazos de renegociação de dívida, definidos na tabela 7, acima apresentada. Esta fase designa-se de pré contencioso.

Por último, no caso de não se chegar a acordo com o cliente na fase de pré contencioso, as Contas a Receber, em colaboração com o diretor da UN, devem dar início à preparação do processo de contencioso. Este processo deve incluir o histórico dos vários momentos de interação com o cliente nas diferentes fases de cobrança, a análise da globalidade da dívida do cliente, incluindo as suas relações com outras UN, e o parecer sobre as potenciais causas que levaram ao

incumprimento. Uma vez completo este processo, deve ser enviado para os assessores jurídicos externos a documentação associada ao processo.

Além disto, a área de Contas a Receber do Norshare promove a realização de reuniões de crédito, no sentido de monitorização e reporte do risco de crédito. Estas reuniões devem ser realizadas por Região e por UN com a presença do diretor da UN e do seu Responsável Comercial, além das Contas a Receber. Nestas reuniões devem ser analisadas a etária de antiguidade de saldos de clientes, as ações de cobrança realizadas, as dívidas com maior antiguidade, a identificação dos clientes pelas diferentes fases de cobrança, bem como os procedimentos para a redução do risco de crédito e análise do cálculo das perdas por imparidade. É da responsabilidade do Gestor das Contas a Receber calendarizar estas reuniões, que devem ser realizadas no mínimo seis vezes por ano, preparar a documentação de suporte das mesmas e notificar os vários intervenientes.

#### 4.4 Desempenho da Análise de Crédito das Contas a Receber

Para propor soluções de melhoria ao processo de análise de crédito das CaR, foi, primeiro, necessário avaliar o seu desempenho, com base na análise dos registos das decisões tomadas no passado e comparação com o desfecho dessas decisões, ou seja, verificar se os e os clientes se revelaram bons ou maus pagadores. Desta forma, foi possível perceber qual o desempenho das decisões tomadas e identificar as oportunidades de melhoria do processo de decisão, caso existissem. Esta análise das decisões de crédito não tem em conta as diferentes empresas do grupo, analisando o desempenho das decisões para o Grupo todo, visto que o modelo de classificação que se pretende criar não vai distinguir empresas do grupo. Além disso, é importante referir que os quadros e respetivos valores não representam todos os registos de pedidos de crédito entre o período

de Julho de 2015 e Setembro de 2016. Isto porque, como já foi referido, muitos dos pedidos de crédito de empresas foram eliminados por falta de informação financeira, para serem incluídos no estudo.

A Tabela 3 demonstra as decisões tomadas pelas CaR, sendo que os casos aprovados e rejeitados representam essas decisões. Então, percebe-se que a análise de crédito desenvolvida apresenta uma taxa de reprovação de cerca 19% do total de pedidos de crédito.

<b>Decisão</b>	<b>Nº de casos</b>
<b>Aprovado</b>	879
<b>Rejeitados</b>	206
<b>Total</b>	1085

**Tabela 3:** Decisão de Crédito

Num momento posterior á decisão de crédito, os casos foram analisados e classificados como bons ou maus pagadores, como já foi referido. No quadro seguinte (Tabela 4) apresenta-se o desfecho das decisões de crédito aprovadas pelas CaR, apresentadas anteriormente.

<b>Casos</b>	<b>Nº de Casos</b>
<b>Aprovados</b>	
<b>Bons pagadores</b>	593
<b>Maus pagadores</b>	286
<b>Total</b>	879

**Tabela 4:** Desfecho das decisões de crédito

Este quadro evidencia que cerca de 33% dos casos aprovados pelas CaR se revelam maus pagadores. No fundo, isto significa que a análise de crédito das CaR tem uma taxa de classificação de crédito correta de 67% para os casos aprovados. No entanto, esta taxa não engloba os casos rejeitados. Isto porque, nos casos rejeitados não é possível avaliá-los num momento futuro, de forma a perceber se a decisão foi correta ou não. Tal como referido na metodologia, este problema prende-se com a inferência dos rejeitados, sendo que este trabalho considera que as decisões de rejeição são corretas. Por isso, as CaR apresentam uma taxa de classificação de crédito correta de 67%.

A Tabela 5 resume a classificação dos casos analisados pelas CaR, durante o período considerado:

<b>Decisão das CaR</b>	<b>Bom pagador</b>	<b>Mau pagador</b>	<b>Total</b>
<b>Aprovado</b>	593(67%)	286 (33%)	879 (81%)
<b>Rejeitado</b>	Não aplicável		206 (19%)
<b>Total</b>			1085

**Tabela 5:** Matriz de classificação das decisões de crédito das CaR

Posteriormente, foi analisado o impacto da variável histórico interno nas decisões das CaR e qual a sua relação com o comportamento dos clientes enquanto pagadores.

Os quadros seguintes ilustram o histórico interno dos casos aprovados e dos casos rejeitados (ver Tabela 6 e 7).

<b>Casos aprovados</b>	<b>N ° de casos</b>
<b>Histórico positivo</b>	264
<b>Sem histórico</b>	259
<b>Histórico negativo</b>	356
<b>Total</b>	879

**Tabela 6:** Histórico interno dos casos aprovados

<b>Casos rejeitados</b>	<b>Nº de casos</b>
<b>Histórico positivo</b>	8
<b>Sem histórico</b>	56
<b>Histórico negativo</b>	142
<b>Total</b>	206

**Tabela 7:** Histórico interno dos casos rejeitados

Com base nestes quadros percebe-se que cerca de 21% dos casos aprovados apresentam um histórico negativo na empresa, enquanto os restantes 79% são representados por casos de clientes com histórico positivo ou sem histórico. Por contrapartida, 67% dos casos rejeitados apresentam histórico negativo e cerca de 30% dos casos rejeitados não têm histórico. Portanto, percebe-se que esta variável tem um peso considerável nas decisões das CaR, porque um cliente com histórico positivo muito dificilmente será privado de crédito, enquanto no caso de um cliente com histórico negativo a análise é mais difícil. Por um lado, a empresa confia nos clientes com histórico positivo, cedendo-lhes quase sempre crédito, sendo a sua análise fácil nestes casos. Por outro lado, as CaR têm de distinguir os clientes que apresentam histórico interno negativo enquanto clientes que têm capacidade para ter crédito e clientes que não devem ter crédito, sendo esta análise muito mais difícil. Os casos de clientes, sem histórico, são os casos mais problemáticos para a análise das CaR, por não se conhecer o comportamento do cliente e por isso, os indicadores económicos e financeiros têm um maior peso nessas decisões.

Os quadros seguintes (ver Tabelas 8 e 9) ilustram a distribuição da variável histórico interno nos pedidos do tipo abertura aprovados e pedidos do tipo reabertura aprovados, sendo, também, considerado o resultado dessas decisões.

<b>Pedidos Abertura Aprovados</b>	<b>Nº de Casos</b>
<b>Bom pagador</b>	<b>435</b>
• Histórico positivo	159
• Sem histórico	168
• Histórico negativo	108
<b>Mau pagador</b>	<b>192</b>
• Histórico positivo	64
• Sem histórico	78
• Histórico negativo	50
<b>Total</b>	<b>627</b>

**Tabela 8:** Desfecho das decisões de crédito de pedidos abertura

<b>Pedidos Reabertura Aprovados</b>	<b>Nº de Casos</b>
<b>Bom pagador</b>	<b>150</b>
• Histórico positivo	26
• Sem histórico	6
• Histórico negativo	118
<b>Mau pagador</b>	<b>87</b>
• Histórico positivo	5
• Sem histórico	4
• Histórico negativo	78
<b>Total</b>	<b>237</b>

**Tabela 9:** Desfecho das decisões de crédito de pedidos reabertura

Então, facilmente se verifica que nos casos de pedidos abertura a taxa de classificação de crédito correta da empresa é de apenas 31%, enquanto no caso dos pedidos reabertura esta taxa é de 36%, sendo que, conforme explicado anteriormente, apenas estão considerados os casos aprovados nesta taxa. Isto porque, como se pode ver nos quadros, no caso dos pedidos reabertura existem muitos casos com histórico negativo, cerca de 82%. Desta forma, percebe-se que

a empresa dispõe crédito a clientes com histórico negativo, porque o ato de ceder crédito fomenta a faturação do cliente, funcionando como uma forma de fidelizar os clientes ao Grupo.

Também importa referir que, por um lado, verifica-se que existem pedidos abertura com histórico negativo. Isto porque pode haver pedidos abertura de clientes, com histórico negativo, que têm crédito noutra empresa do grupo. Por outro lado, existem pedidos reabertura de clientes sem histórico interno de pagamentos porque são pedidos de clientes que já tiveram crédito, que deve ter sido encerrado por falta de atividade.

Por último, importa quantificar os montantes monetários envolvidos na atividade de conceder crédito do grupo. A Tabela 10 quantifica o montante monetário envolvido neste conjunto de clientes (961 casos).

<b>Resultado do crédito</b>	<b>Nº de casos</b>	<b>Montante</b>
<b>Bom pagador</b>	593	2.203.775€
<b>Mau pagador</b>	286	1.405.750€
<b>Total</b>	879	3.609.525€

**Tabela 10:** Montantes de crédito concedido

Assim, verifica-se que cerca de 39% do montante de crédito concedido encontra-se em crédito mal parado, ou seja, está à disposição de clientes maus pagadores. Note-se que neste quadro apenas estão representados os casos aprovados, dado que não existe crédito disponibilizado aos casos rejeitados. No entanto, este montante de crédito representado corresponde apenas aos 1085 pedidos de crédito que compõe os casos analisados neste trabalho. Na realidade, o Grupo tem cerca de 30 milhões de euros de crédito concedido a clientes do pós-venda. Desta forma, percebe-se que o problema de classificação de crédito tem uma dimensão bastante importante e por isso, é importante criar ferramentas que permitam melhorar as decisões ou pelo menos mensurar o risco envolvido em cada operação.

Após a análise do desempenho das decisões de crédito das CaR, percebe-se que existe a possibilidade de melhorar as decisões tomadas. Assim, procurou-se utilizar um modelo de classificação, de forma a procurar obter melhores resultados do que os apresentados, sendo que a sua aplicação pode constituir uma ferramenta de apoio na tomada de decisão de crédito, importante. Desta forma, utilizaram-se vários modelos de previsão e comparou-se os seus resultados.

# Capítulo 5

## Resultados do desenvolvimento dos modelos de crédito

No presente trabalho, começou por se testar os três métodos propostos para a resolução do problema de classificação do crédito do Grupo Nors, nomeadamente, as RNA, a RL e a AMD. O objetivo era construir um modelo que distinguísse os candidatos a crédito em duas categorias: bons pagadores e maus pagadores.

Antes de se explicar o processo de construção dos modelos, faz sentido incluir uma análise sobre o conjunto de dados. Na figura 5 é apresentado um quadro com as correlações lineares entre a maior parte das variáveis utilizadas. Neste quadro, verifica-se que certas variáveis têm uma correlação significativa com outras. Por exemplo, o ROA tem uma correlação positiva forte com a Autonomia Financeira, bem como o Ativo e o Capital Próprio, o Ativo e a Tesouraria Líquida, a Tesouraria Líquida e o Capital Próprio e o Capital Próprio e a Autonomia Financeira, como seria de esperar. Por outro lado, certas variáveis apresentam uma correlação forte negativa como o Resultado Operacional e a Tesouraria Líquida.

Correlação	AF	ROA	GRA	EBITDA	FM	TL	CP	A	Montante/Ci	HI	VN	RO	RC
AF	1												
ROA	0,306565	1											
GRA	-0,159269	0,079138	1										
EBITDA	0,051473	0,058708	-0,091593	1									
FM	0,029953	0,00312	0,00129	0,344391	1								
TL	1,71E-05	-0,006763	0,048413	0,148902	0,850804	1							
CP	0,125242	0,023787	-0,124359	0,614662	0,50982	0,236811	1						
A	0,017099	0,004564	-0,076955	0,652649	0,830021	0,662375	0,73865	1					
Montante/CP	-0,021812	-0,006024	-0,020157	-0,009638	-0,006557	0,000799	-0,015767	-0,006503	1				
HI	-0,083066	-0,034109	-0,097562	0,027056	0,021087	0,013547	-0,011166	0,023423	0,020896	1			
VN	0,004352	-0,004514	0,07488	0,10153	0,667981	0,623581	0,472367	0,638699	-0,004275	-0,002497	1		
RO	0,043291	0,058845	-0,082186	0,656551	-0,252507	-0,410647	0,143248	-0,00813	-0,005504	0,009619	-0,639261	1	
RC	0,224734	0,115443	0,048567	0,041424	0,028022	0,012224	0,042498	0,013295	-0,057484	-0,148677	-0,008409	0,048185	1

12

Figura 5: Matriz de correlação das variáveis

<sup>1</sup> RC – Resultado Crédito (variável dependente); HI – Histórico Interno

<sup>2</sup> Note-se que não estão presentes todas as variáveis independentes, por exemplo, as de variáveis de variação (outras foram excluídas em teste, por exemplo, a idade)

Esta relação negativa evidencia o *trade-off* que as empresas enfrentam entre resultados e equilíbrio financeiro. Esta relação é extremamente importante, até porque este trabalho incide sobre a gestão de clientes do grupo Nors, o que tem implicações diretas na sua tesouraria líquida, sendo que uma gestão mais conservadora, dando prioridade ao equilíbrio financeiro, pode, muitas vezes, significar uma deterioração dos resultados, pela diminuição do volume de negócios. No caso do seu *portfólio* de clientes com crédito, isto pode-se traduzir, por exemplo, em situações em estes poderão facilitar nos prazos de recebimentos dos seus clientes clientes, em troca de um maior volume de negócios e também, optarem por pagar aos seus fornecedores por pronto pagamento, no sentido de obter certo tipo de descontos, o que se traduz em melhores resultados, por via de uma estrutura de custos menos pesada, mas que deterioram o equilíbrio financeiro da empresa. Além disso, foram incluídas, nestes quadros, as variáveis independentes utilizadas para construir os modelos, nomeadamente, a variável “resultado de crédito concedido” com três categorias de pagador (bom, mau, razoável) e a variável resultado com duas categorias de pagador. O quadro evidencia que as variáveis às quais os modelos dão mais importância, como será verãõ mais á frente, são aquelas que têm uma correlação mais forte com as variáveis explicadas. Exemplo destas variáveis são a Autonomia Financeira, o ROA e a Categoria de Histórico de pagamentos interno.

No anexo 1 apresentam-se gráficos de dispersão, de certas variáveis independentes por comparação com a variável dependente, “resultado do crédito concedido”, de forma a perceber como as variáveis estão representadas na variável dependente a explicar (neste caso, o programa apenas permite a introdução de variáveis numéricas e por isso, 0 significa mau pagador e 1 significa bom pagador). Por exemplo, no caso da autonomia financeira, uma das variáveis com maior correlação com a variável resultado (duas categorias), verifica-se que a maior parte dos casos que têm uma autonomia financeira

negativa e portanto, capital próprio negativo, se revelam maus pagadores. No entanto, também, existem vários *outliers* que apresentam uma autonomia financeira negativa e se revelam bons pagadores. No entanto, este fenómeno pode estar associado com o facto de a AC rejeitar a maior parte dos casos que apresentam Capital Próprio negativo e por isso, autonomia financeira negativa, sendo que por isso, se assumem como casos de maus pagadores para os modelos, embora, na realidade, possa não ter influência na capacidade de pagamento de uma empresa. Nesse sentido, vamos analisar a dispersão da variável ROA por comparação com a variável dependente. Neste caso, também se verifica que, no geral, os casos de clientes com ROA negativo se revelam maus pagadores, ainda que existam muitos *outliers*, com ROA negativo e que se revelam bons pagadores.

## 5.1 Modelos de Classificação

Quanto ao processo de construção dos modelos, considerou-se a metodologia proposta por Lingireddy et al. (2005) para a construção de uma RNA, descrito na revisão de literatura. Embora este processo seja proposto para a construção de uma RNA, o mesmo pode-se aplicar a outros modelos de classificação, porque a chave para o sucesso destes modelos, passa por otimizar as variáveis a utilizar, sendo este o propósito dessa metodologia. Então, vamos descrever o processo de construção das RNA, porque este foi idêntico para os outros métodos. Além disso, o processo de construção das RNA é o mais complexo, de entre os métodos testados, porque também inclui a escolha do número de nós, número de camadas, tipo de treino e algoritmo de treino.

Além disto, interessa referir que, como o presente trabalho tem como objetivo comparar os resultados dos três métodos propostos, para depois sugerir a aplicação de um deles, é importante definir uma medida de desempenho dos modelos. Conforme Abdou & Pointon (2011), a matriz de classificação é uma das medidas de desempenho mais utilizadas nas áreas de contabilidade e finanças

para medir o desempenho de aplicações de classificação de crédito e por isso, é esta medida que vai ser utilizada no presente trabalho.

Portanto, o processo de construção de uma RNA começa por perceber a relação entre as variáveis, construindo inicialmente redes com um grande número de nós. Posteriormente, deverá-se restringir o tamanho da RNA, à medida que se vão descartando variáveis independentes com pouco poder explicativo. No entanto, é necessário testar muitas combinações diferentes de variáveis independentes, para ter a certeza de que uma variável pode ser descartada. Note-se também que a análise das correlações entre as variáveis, apresentada no início do capítulo, já dá uma indicação de quais as variáveis com maior poder explicativo para o resultado. Após esta fase, já se deve ter uma ideia das variáveis mais importantes.

Depois, é necessário otimizar o número de variáveis independentes a utilizar, sendo que a arquitectura (número de nós e camadas) da rede depende do número de *inputs* utilizado e deve estar desenhada de modo que não absorva ruído (informação não desejada) e produza bons resultados. Este fenómeno é descrito por Jain et al. (1999) como o “Peaking Phenomenon”. Os autores sustentam que o aumento de *inputs* a um modelo classificador de padrões, pode piorar a sua capacidade de classificação, caso os casos de treino da amostra sejam poucos face ao número de *inputs*. Esta parte do processo é comum a todos os métodos empregados, no presente trabalho, para a resolução do problema, ou seja, para todos os modelos construídos foi realizado um processo de seleção das variáveis mais importantes e de otimização de quais variáveis utilizar. No entanto, no caso das RNA o processo de construção é mais complexo e por isso, de seguida, explicam-se as restantes fases de construção das RNA.

No fundo, o processo de otimização do número de variáveis independentes ocorre em simultâneo ao processo de definição de arquitectura da rede. Ou seja, à medida que se vão descartando variáveis e testando diferentes combinações, a arquitectura da rede deve ser ajustada. A definição da arquitectura da rede é um

dos aspetos mais importantes para o sucesso da RNA, mas é desenvolvido a partir de um processo de tentativa-erro. Não existe nenhuma regra a aplicar para definir o número de nós ou camadas a utilizar, sendo que estes derivam do problema a resolver.

Nas RNA apresentadas mais à frente, utilizaram-se duas camadas ocultas.

- Primeiro, porque com duas camadas ocultas todas as RNA testadas obtiveram melhores resultados, do que com uma camada.
- Segundo, a capacidade de aproximação de uma RNA não é função do número de nós da RNA, mas sim do número de ligações entre eles, o que muitas vezes induz a premissa de que o aumento do número de nós, leva a uma maior capacidade de aprendizagem, o que pode não ser totalmente verdade. Ou seja, com duas camadas ocultas, uma RNA tem mais ligações sinápticas, para o mesmo número de nós.
- Quanto ao número de nós, este foi definido por um processo de tentativa-erro, sendo restringido à medida que o número de variáveis independentes diminui. Esta fase da construção da RNA é extremamente importante, porque se a rede for demasiado grande para o número de inputs que recebe, ou seja, se existirem demasiados nós, nas camadas ocultas, face ao número de inputs, a rede vai armazenar informação não desejada durante o processo de treino, o que depois vai influenciar os seus resultados, negativamente. Assim, percebe-se que a arquitectura da rede é ajustada conforme a otimização do número de variáveis a utilizar se vai desenrolando, sendo que são função uma da outra e por isso, o processo é complexo.
- Quanto á função de ativação entre os nós, inicialmente utilizou-se a função de ativação tangente hiperbólica nas camadas ocultas, de forma a introduzir não lineariedade na rede, porque conforme descrito na revisão de literatura é uma das funções com melhor desempenho. No entanto, também se testou a função de ativação sigmoidal nas camadas ocultas, que muitas vezes obteve melhores resultados. Desta forma, algumas RNA apresentadas foram

construídas com a função de ativação sigmoïdal nas camadas ocultas enquanto outras foram construídas a partir da função de ativação tangente hiperbólica. Na camada de saída utilizou-se a função de ativação linear porque ser a mais frequentemente utilizada na camada de saída.

- Quanto ao tipo de treino, este, também, foi resultado de um processo tentativa erro. Segundo o manual do SPSS Statistics de RNA, o tipo de treino online é o mais indicado para conjuntos de dados grandes. O tipo de treino lote (batch) é o mais indicado para bases de dados pequenas e o tipo de treino mini-batch é o mais indicado para bases de dados médias. O algoritmo de treino utilizado está associado ao tipo de treino. Se o treino for do tipo lote, pode-se escolher entre dois algoritmos de treino: gradiente conjugado em escala e descida de gradiente. Se o tipo de treino for online ou mini-batch, só está disponível a opção descida de gradiente. Por isso, para cada RNA construída testaram-se vários tipos de treino e algoritmos de treino.
- Quanto à divisão dos grupos de treino e teste, optou-se por seguir a sugestão do programa SPSS, isto é, 70% dos casos foram utilizados para treino da rede e 30% utilizados para teste sendo, os casos selecionados aleatoriamente para cada grupo.
- É, também, importante referir que todas as variáveis, excepto a variável categórica histórico interno, foram alvo de redimensionamento. Tal como descrito na revisão de literatura, a escala dos inputs tem importância nos resultados das RNA, o que se verificou nos testes realizados. Neste trabalho utilizaram-se duas formas de redimensionamento, o redimensionamento padronizado e o normalizado ajustado.

Assim, foram realizados vários testes com RNA, RL e AMD, no sentido de construir vários modelos de classificação e comparar os seus resultados, para propor às CaR o mais indicado. De seguida serão apresentados os modelos de

classificação apresentados às CaR, construídos por cada método, com propósitos de classificação diferentes, para se poder comparar os seus resultados.

### Primeiros modelos apresentados às CaR

Em primeiro lugar, vão ser apresentados os métodos de classificação propostos às CaR, numa reunião realizada em Dezembro com os responsáveis pelas CaR, que distinguem um cliente em bom ou mau pagador.

O modelo de classificação seguinte foi construído por uma RNA, com 2 camadas ocultas, 6 nós na primeira camada e 4 na segunda, funções de ativação do tipo sigmoide nas camadas ocultas e identidade na camada de *output* e tipo de treino mini batch, com algoritmo de descida do gradiente, sendo que se pode encontrar no Anexo 2 mais detalhe sobre a RNA\_1. Vamos designar esta RNA por RNA\_1. Note-se, também, que a matriz de classificação da RNA\_1 (ver Tabela11) inclui, apenas, os resultados da fase de teste.

	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
Observado bom pagador		140	26	88,6%
Observado mau pagador		52	104	51,6%
% Global		59,6%	40,4%	75,8%

Tabela 11: Matriz de Classificação RNA\_1

As variáveis utilizadas e, respetiva, importância foram as seguintes (ver Tabela 12):

<b>Variável Independente</b>	<b>Importância</b>	<b>Importância Normalizada</b>
<b>Autonomia Financeira</b>	0,170	95,3%
<b>Montante/CP</b>	0,186	90%
<b>Histórico Interno</b>	0,197	95,3%
<b>ROA</b>	0,076	36,9%
<b>Tesouraria Líquida</b>	0,093	44,9%
<b>Variação Tesouraria Líquida</b>	0,022	10,4%
<b>Variação ROA</b>	0,207	100%
<b>Capital Próprio</b>	0,050	24%

**Tabela 12:** Importância variáveis independentes RNA\_1

O segundo modelo apresentado foi construído a partir da análise discriminante e os seus resultados e respetivas variáveis constam nos quadros seguintes. Este modelo designa-se por AMD\_1 (ver Tabela 13 e 14).

<b>Função discriminante</b>	<b>Coefficiente</b>
<b>Autonomia Financeira</b>	1,66307E-05
<b>ROA</b>	-0,006554012
<b>% Variação ROA</b>	-0,003399695
<b>GRA</b>	-0,000845203
<b>EBITDA</b>	-3,20972E-09
<b>Tesouraria Liquida</b>	2,87868E-09
<b>Capital Próprio</b>	-9,48508E-09
<b>Montante/CP</b>	0,027406045
<b>Histórico Interno</b>	-0,914383242

**Tabela 13:** Função Discriminante do AMD\_1

<b>Observado</b>	<b>Predito</b>			
	<b>Matriz de classificação</b>	<b>bom pagador</b>	<b>mau pagador</b>	<b>% Correto</b>
	<b>bom pagador</b>	499	85	85,4%
	<b>mau pagador</b>	227	274	54,7%
	<b>% Global</b>	66,9%	33,1%	71,2%

**Tabela 14:** Matriz de classificação do modelo AMD\_1

Por último, o seguinte modelo foi construído com recurso ao método de regressão logística e os seus resultados constam nas Tabelas 15 e 16. O seguinte modelo designa-se por RL\_1.

<b>Função de Regressão</b>	<b>Coefficiente</b>	<b>P value</b>
• <b>Constante</b>	-0,391518099	0,0009
• <b>Autonomia Financeira</b>	-2,64758E-05	0,5302
• <b>ROA</b>	0,019012167	0,0002
• <b>% Variação ROA</b>	0,016252969	0,0564
• <b>GRA</b>	0,000844685	0,1755
• <b>Tesouraria Liquida</b>	-2,58166E-09	0,7604
• <b>Capital Próprio</b>	8,88212E-09	0,3326
• <b>Montante/CP</b>	-0,334208041	0,0625
• <b>Histórico Interno</b>	0,828744224	< 0.0001

**Tabela 15:** Função do modelo RL\_1

Observado	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
	bom pagador	491	93	84,08%
	mau pagador	227	274	54,69%
	% Global	66,2%	33,8%	70,51%

**Tabela 16:** Matriz de classificação do modelo RL\_1

Estes modelos foram criados de forma independente, diminuindo o número de variáveis à medida que se ia percebendo o nível de significância que cada método dava às variáveis independentes. No entanto, não deixa de ser interessante referir que os três métodos têm muitas variáveis em comum, o que significa que, os métodos estão em acordo quanto às variáveis mais importantes para explicar a variável dependente.

Além disso, todos os métodos utilizados tiveram mais dificuldade em classificar de forma correta os casos de maus pagadores, o que pode ser justificado, pelo facto de existirem menos casos de maus pagadores, face aos bons pagadores, na amostra.

Por um lado, é interessante, referir que a variável independente com maior importância é o Histórico Interno. No caso da RNA, isto pode-se verificar no quadro que indica a importância de cada variável. Na RL, pode-se verificar pelo *p value* de cada variável, sendo que quanto menor o *p value*, maior o poder explicativo da variável. No caso da AMD, o poder explicativo verifica-se pelo coeficiente da variável na função discriminante, sendo que quanto maior o coeficiente, maior peso a variável terá no modelo.

Por outro lado, a autonomia financeira tem pouca importância para o modelo de RL\_1 e AMD\_1, embora seja uma das variáveis com maior importância, no

caso da RNA. Enquanto o capital próprio tem pouca importância em todos os modelos.

Além disto, também é pertinente que a variável Montante de Processos Jurídicos em aberto face ao Capital Próprio seja uma das variáveis com maior significado estatístico para todos os modelos apresentados. Esta é uma das variáveis com maior peso na decisão de crédito das CaR, porque se espera que um cliente que tem processos jurídicos em aberto contra si seja problemático, sendo que a decisão costuma ser rejeitar, se este montante for muito elevado face ao capital próprio. Isto porque caso a empresa cliente perca o processo/s jurídico e este fôr de montante muito elevado face ao Capital próprio, a empresa poderá-se encontrar em situação de insolvência.

Por último, interessa referir que os modelos de RL\_1 e AMD\_1 necessitam de mais variáveis independentes para caracterizar as categorias de pagador. Verifica-se que ambos os modelos recorrem a nove variáveis explicativas, mas, mesmo assim, não conseguem atingir o grau de precisão na previsão das RNA. De facto, ao longo dos testes, verificou-se que os modelos das RNA não necessitam de tantos *inputs*, tanto que, o *peaking phenomenon* no modelo das RNA inclui um menor número de variáveis. Ou seja, nos modelos lineares é muito mais notório o facto de estes métodos serem estatisticamente incompletos, porque sempre que se retiravam variáveis, com pouco poder explicativo, os modelos tinham piores resultados, enquanto as RNA não.

Quanto aos resultados, sendo a medida de desempenho dos modelos, a percentagem global de classificações corretas, as RNA têm claramente melhor desempenho. Se se avaliar o desempenho dos modelos, pela percentagem de casos de bons pagadores corretamente classificados, as RNA, também, têm melhor desempenho. No entanto, se se avaliar o desempenho dos modelos pela percentagem de casos de maus pagadores corretamente classificados, a AMD é o método que obtém melhores resultados, seguida da RL. Assim, pode-se

considerar que o modelo da RNA\_1 é o que tem melhor desempenho e por isso, foi este o modelo proposto às CaR a aplicar à sua classificação de crédito.

O melhor desempenho das RNA deve-se a vários fatores.

- Em primeiro lugar, com já foi referido, as RNA são mais adequadas para resolver problemas de classificação de padrões não linearmente separáveis, o que justifica a sua maior precisão quando comparadas com a AMD e a RL, o que pode ser o caso.
- Em segundo lugar, o problema do presente trabalho viola princípios do modelo de AMD, porque certas variáveis independentes não seguem uma distribuição multivariada normal, como por exemplo, a variável histórico interno, as matrizes de dispersão dos dois grupos (categorias de pagador) não são homogéneas e existe correlação entre variáveis independentes, o que justifica os melhores resultados das RNA.
- Por último, já se esperava que as RNA tivessem melhores resultados porque, conforme referido por outros autores, como Desai et al. (1996), as RNA têm melhor desempenho na resolução de problemas de classificação.

### **Segundos modelos apresentados às CaR**

No final da reunião, decidiu-se, em parceria com os responsáveis das CaR, que seria interessante incluir no presente trabalho os resultados dos modelos de classificação sem a variável categórica do Histórico Interno, para perceber qual o seu impacto e se, de facto, essa é uma informação pertinente ou não. No anexo 3, estão presentes os resultados de classificação dos três métodos, sem recurso à variável Histórico Interno. O modelo construído a partir de uma RNA será designado RNA\_2, o construído a partir da AMD será designado por AMD\_2 e o construído a partir da RL será designado RL\_2. Verifica-se que os resultados se deterioram quando os modelos não utilizam a variável histórico interno.

De facto, este fenómeno vai de encontro aos erros tipo I e II, que a empresa verifica com os seus clientes. Ou seja, o Grupo Nors tem muitos clientes com crédito com bons indicadores financeiros e económicos, que se revelam maus

pagadores, porque são clientes que dão a relação por garantida. Isto pode-se dever a dois fatores. Ou são clientes que sabem ser demasiado grandes, a nível financeiro e económico, para lhes ser rejeitado crédito (por esta razão se excluíram os casos de clientes com Capital Próprio superior a 100 milhões de euros) ou por terem uma relação com o Grupo há muito tempo. Estes são os casos de erro tipo II, clientes que se esperam ser bons pagadores que se revelam maus pagadores. Por outro lado, o grupo tem muitos clientes com crédito com maus indicadores económicos e financeiros que se revelam bons pagadores, porque a atividade do cliente está dependente do fornecimento dos produtos do Grupo e por isso, dão extrema importância à relação com a Nors. Além disso, estes clientes dão mais importância ao crédito obtido, por saberem, que se não cumprirem os prazos de pagamentos, será muito difícil readquirir o crédito, uma vez perdido, dada a sua situação económica e financeira. Estes são os casos de erro tipo I.

No fundo, verificam-se que existem muitos clientes que deveriam ser bons pagadores pelas suas características, mas revelam-se maus pagadores e vice-versa. Por estas razões a variável histórico interno tem tanta importância para a decisão de crédito das CaR e também, para os modelos de classificação.

Com a elaboração destes três modelos de classificação, comparação dos seus resultados e proposta de aplicação, um dos objetivos do presente trabalho ficou realizado.

No entanto, os pedidos de crédito de clientes sem histórico de pagamentos, no Grupo, são um dos maiores desafios na decisão de crédito das CaR. Isto porque, como explicado anteriormente, a informação proveniente do histórico de pagamentos do cliente é extremamente valiosa para a decisão de crédito e por isso, para pedidos de crédito de clientes novos ou sem histórico, a decisão de crédito torna-se mais difícil por não existir evidência empírica do comportamento de pagamentos do cliente.

### Terceiros modelos apresentados às CaR

Desta forma, decidiu-se criar um modelo de classificação de crédito a aplicar, exclusivamente, a pedidos de crédito de novos clientes. Por um lado, este modelo evita a utilização da variável histórico interno e é obrigado a classificar os clientes utilizando apenas os seus indicadores. Por outro lado, este modelo revela-se uma boa ferramenta de auxílio numa das decisões mais problemática da análise de crédito das CaR. Portanto, este modelo foi treinado e testado, única e exclusivamente, com os casos de clientes sem histórico, sendo que o número de casos de treino e teste é menor.

Para a construção deste modelo utilizaram-se as RNA, a RL e a AMD, mas apenas vão ser apresentados os resultados das RNA, enquanto os resultados dos outros métodos e mais especificações da RNA\_3, apresentam-se no Anexo 4. Este modelo vai-se designar RNA\_3. O modelo construído pela RL designar-se-á RL\_3 e o modelo construído pela AMD designar-se-á AMD\_3.

Os quadros seguintes (Tabelas 17 e 18) ilustram os resultados de classificação do modelo RNA\_3, na fase de teste e os *inputs* utilizados por este:

	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
Observado bom pagador		56	5	91,8%
mau pagador		21	19	47,5%
% Global		76,2%	23,8%	74,3%

Tabela 17: Matriz de classificação do modelo RNA\_3

<b>Variável Independente</b>	<b>Importância</b>	<b>Importância Normalizada</b>
<b>Capital Próprio</b>	0,284	100%
<b>Ativo</b>	0,161	56,5%
<b>Resultado Operacional</b>	0,153	53,7%
<b>ROA</b>	0,041	14,5%
<b>Variação ROA</b>	0,076	26,9%
<b>Autonomia Financeira</b>	0,132	46,6%
<b>Tesouraria Liquida</b>	0,089	31,3%
<b>Montante jurídico aberto/Capital Próprio</b>	0,063	22,3%

**Tabela 18:** Importância das variáveis independentes do modelo RNA\_3

Este modelo não apresenta tão bons resultados como o modelo RNA\_1, porque foi treinado com um menor número de casos, o que deteriora os resultados, por haver menos casos para caracterizar cada categoria de pagador. No entanto, o modelo apresenta bons resultados para uma RNA treinada e testada com apenas 314 casos. Além disso, é importante referir que este modelo, ao contrário dos outros, dá muita importância ao Capital Próprio, Ativo e Resultado Operacional. Isto pode-se dever ao facto de muitos dos maus pagadores presentes nos 314 casos, serem pedidos de crédito rejeitados pelas CaR, sendo que os clientes aos quais as CaR rejeitam crédito são, normalmente, clientes com maus indicadores a nível de Capitais Próprios.

O facto de este modelo incluir clientes de crédito novos, exclusivamente, tem influência nos resultados do modelo, como se verifica. Note-se que todos os

clientes presentes no treino deste modelo, estavam a começar uma nova relação comercial, aquando do recebimento do crédito. Então, aqueles que se revelam bons pagadores são, de facto, clientes com capacidade financeira para cumprir com os pagamentos e por isso, para ter crédito. Por outro lado, aqueles que se revelam maus pagadores, são clientes que não têm capacidade para cumprir com os pagamentos, porque, pouco tempo depois de iniciarem as compras a crédito, já não conseguem cumprir com os pagamentos. No fundo, este modelo permite isolar, aquilo a que se vai designar neste trabalho, de efeito confiança. Para explicar melhor este efeito, vamos assumir que nenhum cliente pede crédito com o intuito de não cumprir com os prazos de pagamento, até porque falhar pagamentos não é bem visto em nenhum tipo de atividade. Todavia, certos clientes acabam por falhar. Quando se começa uma nova relação comercial, como de qualquer outro tipo, não existe confiança, sendo que à medida que o tempo passa a confiança vai aumentando, para o bem ou mal. Então, enquanto nos dados dos modelos apresentados anteriormente, estão presentes muitos casos de clientes que se atrasam porque já têm uma relação duradoura com a Nors e por isso, já não têm rigor no cumprimento dos prazos. O novo modelo apresentado apenas inclui clientes de crédito novos em que não existe confiança na relação comercial. Então, aqueles que vão falhar nos prazos de pagamento, são clientes que não têm uma situação financeira que lhes dê liquidez para cumprir com os pagamentos e por muito que tentem contrariar, vão entrar em incumprimento. Enquanto aqueles que cumprem são, de facto, clientes com capacidade financeira para cumprir com os pagamentos e ter crédito no grupo. Desta forma, isola-se o efeito confiança, incluindo apenas os casos em que esta não existe. Por exemplo, na base de dados descrita no capítulo três, estão presentes 36 clientes com *plafond* de crédito superior a 10000 euros, sendo que, no Grupo Nors, um *plafond* de crédito superior a este montante é evidência de uma relação já duradoura, em que existe confiança no cliente, porque a empresa não concede estes montantes a clientes que não conhece. Destes 36 clientes, 19 são maus pagadores, sendo estes

os clientes mais importantes para o Grupo, em termos comerciais, nesta base de dados. Assim, verifica-se o efeito confiança, ou seja, dos 36 clientes em que a empresa tem mais confiança e que se esperam ser cumpridores, quase 50% são maus pagadores, precisamente, por contarem com a confiança da parte da Nors. Isto vai de encontro, à justificação dos vários casos de erro tipo I e II. Havendo confiança, os clientes dão, por vezes, a relação como garantida, resultando em erro tipo II. Por outro lado, a empresa tem sempre menos confiança nos clientes com maus indicadores, por muito que a relação não seja recente, e por isso, muitos destes clientes, que se esperam entrar em incumprimento, revelam-se bons pagadores, evidenciando o erro tipo I.

Assim, apesar de a RNA\_3 ser treinada e testada com um menor número de casos, o que levaria a piores resultados, em condições normais, porque existem menos casos para caracterizar cada categoria de pagador, o isolamento do efeito descrito permite que a RNA\_3 tenha bons resultados.

Apesar de se ter concluído um dos objetivos deste trabalho ao criar estes modelos de classificação, percebeu-se que um modelo de classificação de crédito que distinguísse os clientes em três categorias, poderia ser mais útil para a área de CaR do Grupo e por isso, procurou-se criar um novo modelo de classificação de crédito.

#### **Quartos modelos de classificação apresentados à CaR**

Então, conforme já explicado na metodologia, classificou-se os clientes presentes na amostra, de acordo com a variável dependente do modelo, ou seja, enquanto bons, maus ou razoáveis pagadores, de forma a recomençar o processo de construção do modelo de classificação. Esta parte do trabalho é uma análise exploratória, porque este modelo vai ser construído, exclusivamente, a partir de RNA. Isto porque os outros métodos sugeridos, neste trabalho, requerem variáveis dependentes binárias e por isso, não podem ser aplicados a este problema.

Desta forma, procurou-se criar um novo modelo de classificação de crédito com mais utilidade, focando a atenção nas RNA. Para esse feito, recomeçou-se o processo de otimização da RNA, começando com uma rede grande em termos de nós e variáveis, para depois otimizar a sua arquitetura e *inputs*.

O modelo de classificação de crédito em três categorias, proposto à área Contas a Receber obteve os seguintes resultados de classificação (Tabela19):

	Predito				
	Matriz de classificação	bom pagador	razoável	mau pagador	% Correto
Observado	bom pagador	98	6	2	92,5%
	razoável	20	68	14	66,7%
	mau pagador	10	35	41	47,7%
	% Global	43,5%	37,1%	19,4%	70,4%

Tabela 19: Matriz de classificação do modelo RNA\_4

Mais uma vez, note-se que o modelo teve mais dificuldade em classificar, de forma correta, os maus pagadores sendo que, neste caso, ainda, existem menos casos, na amostra, por categorias de pagadores. Este facto explica, também, o maior número de casos mal classificados pelo modelo, no geral, face aos modelos de classificação de crédito em duas categorias, ou seja, este modelo tem uma precisão de classificação menor, porque o seu treino baseou-se em amostras mais pequenas, para cada categoria de pagador. Isto significa que o modelo teve de caracterizar cada categoria de pagador, isto é, cada padrão possível de classificação, com um menor número de casos, o que diminui a sua capacidade de previsão e o nível de confiança das suas classificações. Contudo, quanto aos *inputs* utilizados pelo modelo, estes voltam a estar em sintonia com os modelos construídos anteriormente, ou seja, mesmo para um modelo de classificação de

crédito com um propósito diferente (classificação em três categorias), as variáveis mais importantes voltam a coincidir, em grande parte, com aquelas utilizadas nos modelos anteriores. Tal pode ser evidenciado na Tabela 20, que representa a importância de cada *input*, para esta RNA:

<b>Variável Independente</b>	<b>Importância</b>	<b>Importância Normalizada</b>
<b>Capital Próprio</b>	0,084	50,9%
<b>Ativo</b>	0,072	43,5%
<b>Resultado Operacional</b>	0,088	53,2%
<b>ROA</b>	0,128	77,1%
<b>Variação Capital Próprio</b>	0,118	71,0%
<b>Autonomia Financeira</b>	0,091	54,9%
<b>Tesouraria Líquida</b>	0,125	75,3%
<b>Montante jurídico aberto/Capital Próprio</b>	0,166	100%
<b>Histórico Interno</b>	0,128	77,1%

**Tabela 20:** Importância das variáveis independentes do modelo RNA\_4

Podem-se encontrar mais especificações desta rede, no Anexo 5.

Portanto, os modelos propostos às Car foram os seguintes: o modelo RNA\_1, que classifica os clientes em bons ou maus pagadores; o modelo RNA\_3, a aplicar a casos sem histórico interno, que classifica os clientes em bons ou maus pagadores; e o modelo RNA\_4 que classifica os clientes em bons, razoáveis ou maus pagadores.

Como conclusão deste capítulo, pode-se dizer que as RNA foram o método com melhores resultados, nos vários modelos de classificação criados, e que todos os métodos utilizados se revelaram dependentes da variável categórica histórico interno para obter bons resultados de classificação.

## 5.2 Modelo de Definição de Montante de Crédito

Após a elaboração dos modelos de classificação de crédito, procurou-se encontrar uma solução para o problema de definir o montante de crédito a conceder a cada candidato de crédito. Este é um problema de alocação de recursos, tal como descrito na Revisão de Literatura.

Portanto, Markowitz (1952) propõe um modelo de seleção de portfólios que consiste em escolher aqueles que maximizam o retorno esperado, para um dado nível de risco, medido pela variância do retorno. No entanto, o problema do presente trabalho não é escolher um portfólio.

Suponha-se que o conjunto de clientes que têm crédito numa determinada empresa do Grupo Nors (por exemplo, CiviParts) é um portfólio, em que cada cliente representa uma opção de investimento. No entanto, todos os clientes presentes neste portfólio já foram submetidos à decisão de crédito, sendo que o problema, neste caso, é determinar o montante a investir, e não, propriamente, a decisão de crédito. Note-se que todos os clientes que estão neste portfólio têm de receber crédito, já que a decisão de crédito foi de aprovação.

Inicialmente, procurou-se adaptar a solução de Markowitz a este portfólio, ou seja, definindo o *plafond* de crédito a conceder a cada cliente, de forma a maximizar o retorno esperado, para um dado nível de risco associado. No entanto, o modelo de Markowitz (1952) mede o retorno médio dos ativos e a sua variância através de séries temporais. No caso do Grupo Nors, não foi possível estabelecer uma série temporal para o retorno médio dos ativos e do portfólio, bem como da sua variância, porque não existem registos da evolução do *plafond*,

ao longo do tempo. Logo, a solução de Markowitz (1952) não pode ser aplicada ao problema em questão, no entanto o princípio de maximizar o retorno, minimizando o risco, pode ser aplicado. Foi, então, necessário desenvolver uma solução para medir o retorno de cada opção de investimento do portfólio.

Então, optou-se por recorrer a uma medida financeira de avaliação de investimentos, que geralmente é utilizada para auxiliar a tomada de decisão, em investimentos. No fundo, tentou-se adaptar o conceito de valor atualizado líquido, conhecido por VAL, para avaliar os investimentos do portfólio de clientes com crédito, do Grupo, no sentido de aplicar os algoritmos genéticos a maximizar uma medida de criação de valor, tal como no exemplo demonstrado na Revisão de Literatura, em que se maximiza a CLV (capítulo 2.2).

O VAL é um indicador absoluto que auxilia na fundamentação da tomada de decisão de investimento, ao criar uma medida de criação de valor, comparando o valor do investimento, dado pela soma dos *cash flows* futuros do investimento atualizados para o presente, e o investimento necessário para o desenvolvimento do projeto, sendo que se o VAL é positivo, o projeto tem potencial para criação de valor.

Seja o Valor do Investimento (VI) dado pela seguinte forma:

$$VI = \sum_{t=1}^n \frac{CF_t}{(1+r)^t} + \frac{VR_n}{(1+r)^n} ,$$

Onde:

- $CF^t$  - cash flow gerado pelo investimento no período t
- r - taxa de desconto dos cash flows futuros
- n - número de anos de previsão de cash flows periódicos futuros
- $VR^n$  - valor residual do investimento no momento n

O VAL será dado pela seguinte fórmula:

$$VAL = VI - I_0$$

Onde,  $I_0$  representa o preço do investimento, ou seja, o montante de capital necessário para o desenvolvimento do investimento.

No entanto, não é possível medir o VAL para cada cliente de crédito, a começar pelo facto de que, o montante de crédito concedido, a cada cliente, não tem subjacente uma duração definida e por isso, não é possível atualizar o VI para o presente. Além disso, seria extremamente difícil medir o valor residual de cada cliente.

Então, considerou-se que o investimento em cada cliente tem uma duração de, apenas, um período, um ano, e que não existe taxa de desconto dos cash flows futuros, que representa o custo de oportunidade do capital investido. Desta forma, o Valor líquido do investimento (VL) será dado pela seguinte forma:

$$VL = VI_t - I_0,$$

Onde:

- VI representa o cash flow gerado pelo investimento durante o período de um ano
- $I^0$  representa o preço do investimento.

Note-se que se o VI consiste no *Cash Flow* que o investimento pode gerar durante um ano, o Investimento Inicial é considerado para 1 ano também, até porque as Contas a Receber do Grupo Nors acabam por rever o *plafond* de cada cliente com uma periodicidade inferior a um ano, com o decorrer da sua atividade.

Então, foi necessário definir uma forma de medir o investimento inicial e o retorno desse investimento. O Investimento Inicial foi definido como o montante de crédito multiplicado pelo WACC<sup>3</sup> (custo médio ponderado do capital), da empresa à qual o pedido de crédito foi submetido, isto porque, apesar de a empresa não entregar o montante disponibilizado aos clientes, este tem um custo

---

<sup>3</sup> WACC – em inglês, Wheighted average cost of capital

associado, que pode ser medido desta forma, já que se está a arriscar capital da empresa, que, por sua vez, tem um custo subjacente, o WACC. Desta forma, o investimento inicial representa os custos do empréstimo em que a empresa está a incorrer, ao conceder o montante de crédito.

Por outro lado, optou-se por medir o Valor do Investimento, isto é, o *cash flow* esperado do investimento, sendo que este tem apenas a duração de um ano, como a faturação esperada, durante um ano, do cliente, multiplicada pela margem bruta, da respetiva empresa do grupo, e deduzida das perdas por imparidade esperadas desse cliente. Isto porque o retorno que a empresa obtém com o cliente não é a sua faturação, mas sim a margem associada a essa faturação. Caso contrário, estaríamos a incluir os custos da atividade das empresas, no retorno que estas obtêm dos clientes. Além disso, deve-se ter em atenção o risco associado a cada cliente, que se pode traduzir em perdas por imparidade. Tal como descrito no capítulo de 4, a área Contas a Receber é responsável pelas perdas por imparidade, sendo que à medida que o cliente se atrasa, as perdas por imparidade vão sendo registadas, o que não significa que esse valor não seja obtido, posteriormente. No entanto, para o presente modelo, considerou-se que não existe possibilidade de reobter as imparidas registadas, numa data futura, porque esse é o factor que permite que o modelo distinga os clientes pelo risco, como se verá mais à frente. Contudo, foi, também, necessário definir uma forma de medir as imparidades esperadas para cada cliente, candidato a crédito.

Tendo em atenção que o processo de conceder um montante de crédito, a um candidato de crédito, é posterior à decisão de conceder e que, na primeira parte deste estudo, se desenvolveu um modelo de classificação de crédito para auxiliar a tomada de decisão de conceder crédito, que tem como resultado a categoria de pagador, a que um determinado cliente pertence, mas também a probabilidade de um cliente pertencer a cada categoria de resultado possível, pode-se utilizar essa probabilidade como forma de medir as imparidades previstas e distinguir os clientes. Por isso se apresentam, exclusivamente, modelos elaborados pelo

programa SPSS, sendo que o programa da Palisade não permite obter esta probabilidade, no caso do PMC, tipo de rede utilizada neste trabalho. Ou seja, utilizando a probabilidade de um cliente ser mau pagador para medir as imparidades previstas, o modelo estará a incluir uma taxa de risco que resulta da análise das RNA e que permitirá distinguir os clientes com base nos seus indicadores, sendo que as imparidades vão ser dadas pela multiplicação da taxa de risco do cliente e a faturação esperada deste.

Além disso, foi necessário definir uma forma de calcular a faturação esperada. Nesse sentido, calculou-se a relação entre a faturação do ano anterior e o montante de crédito disponível, de cada cliente. Desta forma, utilizando a aplicação *bestfit* do *@Risk* foi possível calcular a distribuição que a relação entre o *plafond* e a faturação seguem, no portfólio de clientes de crédito, de forma a aplicar essa distribuição a novos pedidos de crédito, para calcular a sua faturação esperada. Note-se que se assume que a disponibilização de crédito se traduz num aumento da faturação. Isto porque, se não fosse assim, a empresa não disponibilizaria crédito aos seus clientes. Assim, percebeu-se que esta relação segue uma distribuição LogNormal<sup>4</sup>. Desta forma, nos casos de novos candidatos de crédito sem histórico na empresa, pode-se aplicar a razão entre a faturação esperada e o montante de crédito, do portfólio, enquanto para clientes que já têm histórico de crédito, pode-se aplicar a razão entre a sua faturação do ano anterior e o montante de crédito de que dispunha.

Então, a fórmula que avalia o valor líquido do investimento no cliente será dada pela seguinte expressão:

$$\begin{aligned}
 VL &= -I_0 + VI \Leftrightarrow VL = -CE + CF_t \\
 &\Leftrightarrow VL = -(Plafond * WACC) + ((FE * MB) - \text{Imparidades}) \Leftrightarrow \\
 &\Leftrightarrow VL = -(Plafond * WACC) + (Plafond * Pr\text{oporção} * MB) - (txrisco * FE)
 \end{aligned}$$

Onde,

- VL – valor líquido do investimento

---

<sup>4</sup> Distribuição de uma variável aleatória cujo logaritmo assume uma distribuição normal

- $I_0$  - Investimento Inicial
- CE – Custos do “Empréstimo” (custos de conceder o crédito)
- $CF_t$  - Cash Flow esperado do investimento no período t
- t – período de duração do investimento (1 ano)
- Plafond – montante de crédito disponibilizado
- WACC – custo médio ponderado do capital
- FE – Faturação esperada do cliente para o período de um ano
- MB – Margem Bruta
- Imparidades – Imparidades previstas
- Proporção – distribuição da relação entre plafond e FE calculada com o bestfit
- Tx risco – probabilidade predita pela análise das RNA de cliente ser mau pagador

De destacar que a MB deveria ser substituída pela Margem de Contribuição de cada empresa, mas para não estar a pedir que a empresa divulgasse as suas margens, neste trabalho, utiliza-se uma MB representativa de 27,3%. O mesmo se passa com o WACC (custo médio ponderado do capital), sendo que se assumiu um WACC de 7,5%. Isto não tem importância para a questão da definição do montante de crédito a conceder, porque estes valores são iguais para todos os clientes. No entanto, aconselha-se que a empresa utilize os valores reais, dado que isso permitirá que os valores líquidos do investimento em cada cliente, bem como do portfólio total, sejam mais corretos.

Desta forma, foi possível formular o problema de forma que, tanto o VI e o  $I_0$  estejam dependentes do montante de crédito disponibilizado, o que permitiu aplicar os algoritmos genéticos a resolver um problema de otimização de recursos, cujo objetivo é maximizar o valor líquido do portfólio de crédito de cada empresa do grupo. Note-se que este problema só se coloca para clientes que façam uma candidatura a crédito, uma vez que não se pode alterar o montante

de crédito disponibilizado para clientes que já têm um montante atribuído, salvo em situações em estes revelem atrasos de pagamentos ou não usem o crédito disponibilizado. Assim, o modelo foi criado para ser aplicado a um portfólio de clientes de crédito de uma determinada empresa do Grupo. Isto porque os programas utilizados permitem ter em conta o valor total do portfólio de clientes com crédito da empresa, para otimizar o montante de crédito dos novos pedidos de crédito.

Utilizou-se o *RiskOptimizer*, aplicação do *software DecisionTools* da *Palisade*, para resolver o problema, de forma a maximizar o *Value at Risk* do VL do Portfólio (5% da distribuição à esquerda do valor líquido total), ajustando o montante de crédito a conceder aos novos clientes do portfólio (aqueles que pediram o crédito), o que influencia o VI e o  $I_0$ . A aplicação *RiskOptimizer*, tal como já referido no capítulo 3.3, utiliza algoritmos genéticos e permite introduzir incerteza no modelo através da simulação de *Monte Carlo*<sup>5</sup>, já que está inserido na aplicação @Risk, que permite assumir distribuições para as variáveis incertas.

Uma vez desenvolvidos os modelos de otimização e de classificação de crédito, foi decidido que o próximo passo seria testar os modelos em novos casos. Assim, apresenta-se, no Anexo 6, os resultados de classificação de crédito do modelo RNA\_4 para três clientes da CiviParts que pediram crédito, sendo que se utilizou o modelo de classificação em três categorias, por ser o modelo escolhido pelas CaR para esta fase de teste integrado com o modelo de decisão do montante de crédito. Para não divulgar o nome dos clientes, atribuiu-se um número a cada um, para facilitar a sua identificação. No Anexo 6, também, se apresentam os dados utilizados pelo modelo RNA\_4 para classificar os clientes, a classificação do modelo (*Predicted Value*), e a probabilidade de ser mau pagador (*Selected Probability*). Verifica-se que o modelo classificou o cliente 1 e 3 como bons pagadores e o cliente 2 como mau pagador, sendo que esta classificação foi de

---

<sup>5</sup> Método heurístico que utiliza a aleatoriedade para introduzir incerteza nos seus resultados

encontro à análise das Contas a Receber e também, ao indicador de risco presente nos relatórios financeiros, da Informa D&B.

O modelo RNA\_1 também foi testado em pedidos de crédito reais, em parceria com as CaR. Por isso, inserem-se, no Anexo 9, as classificações de crédito atribuídas pelo modelo RNA\_1, para os pedidos de crédito dos clientes 4, 5, 6, 7 e 8.

Posteriormente, aplicou-se o modelo de otimização do montante de crédito a conceder aos clientes para ajustar o montante de crédito a conceder a esses três clientes, tendo em conta todo o portfólio de crédito de clientes da CiviParts.

Em primeiro lugar, foi necessário calcular o valor líquido do portfólio da CiviParts. É importante referir que não se utilizou o modelo RNA\_4, ou qualquer outro, para definir uma taxa de risco para os clientes do Portfólio que já tinham crédito, por várias razões. Primeiro, o processo de recolha de informação para as variáveis a ser utilizadas pelo modelo RNA\_4 seria demasiado moroso, bem como para os restantes modelos de classificação apresentados anteriormente, dado que este portfólio conta com mais de 1000 clientes. Segundo, muitos dos clientes não poderiam ser alvo da classificação da RNA, por falta de informação disponível e por isso, teriam de ser excluídos, sendo que o valor do portfólio não estaria tão bem caracterizado.

Portanto, para calcular uma taxa de risco para o modelo, utilizaram-se as imparidades registadas pelo Banco de Portugal por sector de atividade, para calcular uma taxa de risco por comparação das imparidades registadas e do volume de negócios, por setor de atividade. Desta forma, calculou-se uma taxa média ponderada com base na representação de cada sector de atividade, no volume de negócios do portfólio de crédito da CiviParts. Para esta taxa, utilizou-se o @Risk para assumir uma distribuição Pert<sup>6</sup>, de forma a introduzir alguma incerteza no valor desses investimentos. Além disso, para os clientes com crédito,

---

<sup>6</sup> Distribuição com uma curva semelhante à distribuição normal, com resultados limitados entre certos valores

já presentes no portfólio, assumiu-se uma Faturação Esperada igual à faturação do ano anterior.

Em segundo lugar, introduziram-se os novos pedidos de crédito no portfólio, bem como alguns dados necessários para a aplicação do modelo. Então, para os três novos clientes, utilizou-se uma FE calculada a partir da proporção, explicada anteriormente e deu-se uma distribuição Pert à taxa de risco das RNA de cada cliente, para introduzir incerteza no modelo. Além disso, para as imparidades previstas utilizou-se uma distribuição de risco cumulativa<sup>7</sup> do @Risk, que pode variar entre 0 e o valor da multiplicação da taxa de risco e a FE, de forma a introduzir, no modelo, a noção de que com o aumento da FE, aumenta a probabilidade de perdas por imparidades, sendo que esta distribuição está também dependente da taxa de risco de cada cliente. Isto significa que um cliente com uma taxa de risco maior que a taxa de risco de outro, vai ter um máximo de imparidades previstas maior e as suas imparidades previstas vão aumentar à medida que a FE aumenta e vão aumentar a um ritmo maior que as imparidades de um cliente com uma taxa de risco menor.

Por último, utilizou-se o *RiskOptimizer* para maximizar o VaR do Valor Líquido Total do Portfólio, ajustando, apenas, os valores dos montantes de crédito a conceder aos três novos clientes, utilizando as seguintes restrições, definidas em parceria com as Contas a Receber:

- o montante de crédito a conceder a cada cliente varia entre 1000 euros e 15000 euros;
- o montante de crédito a distribuir pelos três clientes é de 20000 euros;
- cada cliente tem de ter um montante de crédito inferior a 50% do seu Capital Próprio;
- Cada cliente tem de ter um montante de crédito inferior à autonomia financeira multiplicada pelo Capital Próprio

---

<sup>7</sup> Função que indica a probabilidade de uma variável aleatória ser menor ou igual à variável independente da função

Estas restrições podem ser ajustadas consoante os clientes a avaliar, as disponibilidades financeiras da empresa, o risco que empresa está disposta a incorrer, etc. No entanto, tem de haver sempre um montante a distribuir pelos candidatos de crédito, caso contrário o modelo não apresentará bons resultados, sendo que este montante, também, pode ser ajustado consoante o número de clientes a analisar, o seu nível de risco, as disponibilidades financeiras. Isto porque, se este é um problema de alocação de recursos, tem de se definir os recursos a alocar. Este pressuposto do modelo não vai ao encontro da situação do grupo Nors, que não tem um limite de crédito definido a distribuir pelo portfólio total ou por um conjunto de clientes. Todavia, as disponibilidades monetárias não devem ser consideradas ilimitadas. Por isso, este modelo não apresenta soluções que se possam considerar ótimas do ponto de vista do Grupo Nors, mas encontra soluções que se podem considerar aproximações dos pontos ótimos. Além disso, cada empresa do Grupo poderia fazer depender este montante a distribuir pelos novos clientes do seu portfólio de crédito consoante as suas NFM ou TL. Aliás, poderia-se avaliar as disponibilidades, de cada empresa, semanalmente e definir o montante a distribuir tendo em conta as disponibilidades financeiras da semana anterior, definindo os montantes de crédito de cada cliente, por exemplo, à segunda-feira.

Desta forma, definindo que o @Risk pode fazer 1000 iterações para as distribuições assumidas, o RiskOptimizer atribuiu o crédito da seguinte forma:

- 2586 euros ao cliente 1;
- 2414 euros ao cliente 2;
- 15000 euros ao cliente 3.

Com estes *plafonds* de crédito atribuídos a estes três novos clientes, o VAR do portfólio de crédito da CiviParts seria de 5.258.599 de euros, aproximadamente.

Estes resultados, bem como as distribuições assumidas pelas variáveis incertas, estão presentes no Anexo 7.

Posteriormente, aplicou-se o mesmo modelo de otimização aos clientes 1 e 2, clientes com as taxas de risco das RNA mais elevadas (38% e 34%) dos três clientes analisados, de forma a perceber o desempenho do modelo em clientes com bastante risco, exclusivamente. Tratando-se de clientes com bastante risco, optou-se por alterar ligeiramente as restrições. Definiu-se que o montante a distribuir continuaria a ser de 20000 euros, sendo que cada um poderia ter no máximo 10000 euros de crédito. Neste caso, os resultados foram os seguintes:

- 2192 euros ao cliente 2;
- 2808 euros ao cliente 1.

Com estes *plafonds* de crédito atribuídos a estes dois clientes, o VaR do portfólio de crédito da CiviParts seria de 5.258.596 euros, sendo que estes resultados e as distribuições das variáveis incertas estão presentes no Anexo 8.

Por fim, pode-se dizer que os testes dos modelos de classificação de crédito em casos reais e os montantes de crédito a atribuir a esse casos, indicados pelo modelo de definição de montantede de crédito, apresentados neste capítulo, foram de encontro ao parecer da análise de crédito das CaR.

### 5.3 Limitações dos resultados

Como referido na literatura, os modelos de classificação de crédito têm inúmeras vantagens, contudo, é também importante analisar algumas desvantagens que devem ser tidas em conta aquando da sua utilização.

Primeiro, é necessário ter em consideração que estes modelos são, sempre, estatisticamente incompletos, porque não é possível incluir todas as variáveis num modelo deste género, segundo Thomas et al. (1996)

Segundo, conforme referido por Eisenbeis (1978), estes modelos não têm em consideração que certos empréstimos podem ser lucrativos, ainda que o cliente entre em incumprimento em certo momento.

Terceiro, como referido por Heffernan (2005), estes modelos baseiam-se em informação histórica, utilizando variáveis e pesos associados assumidos como constantes ao longo do tempo. Por isso, é importante atualizar os modelos de classificação de crédito, constituindo a única forma de combater este problema. Por um lado, ao atualizar os modelos, podem-se incluir mais registos de crédito, construindo modelos com melhor precisão. Por outro lado, ao atualizar o modelo, consegue-se acompanhar as tendências macroeconómicas, que têm implicações nos indicadores financeiros e económicos das empresas candidatas de crédito. Heffernan (2005) também considera que o facto de estes modelos terem, muitas vezes, apenas dois resultados possíveis constitui uma fragilidade. Nesse sentido, contruiu-se e propôs-se o modelo RNA\_4, por este incluir mais categorias de classificação de crédito possíveis.

Quanto ao modelo de decisão de crédito, este constitui uma ferramenta que apresenta boas soluções para o problema de definição do montante de crédito a conceder, ainda que estas não sejam ótimas, para o presente caso, do grupo Nors. Este pode ser encarado como um indicador a ter em conta aquando da definição do montante de crédito a conceder a um cliente, que tem em atenção a análise das RNA e o valor total e variância da distribuição do portfólio de crédito.

Além disso, não se encontrou literatura sobre aplicação de modelos de classificação de crédito e de definição de *plafond* de crédito a organizações comerciais e por isso, espera-se que os modelos apresentados, neste trabalho, possam ser interessantes para outras organizações deste género.

Neste sentido, os objetivos propostos no início do presente trabalho foram cumpridos, ao propor vários modelos de classificação de crédito e ao criar um modelo de decisão de *plafond* de crédito.

Contudo, seria moralmente incorreto não alertar para certas consequências que podem decorrer do uso inapropriado de modelos deste género.

As organizações, como Levy (1994) defende, inserem-se em ambientes que são caracterizados por sistemas complexos, dinâmicos e não lineares e por isso,

caóticos. Nestes sistemas, pequenos distúrbios multiplicam-se ao longo do tempo, pela natureza do sistema e por isso, são dependentes das condições iniciais e altamente imprevisíveis, conforme Levy (1994). O planeamento surge como forma para combater a incerteza que caracteriza estes sistemas, conforme Thietar & Forgues (1995). Segundo estes autores, ao formalizar o processo de decisão, os gestores estão a criar ilhas de certeza, o que lhes permite proceder de maneira racional. A formalização do processo de decisão coloca o tomador da decisão numa situação que já conhece, isolada do ambiente caótico em que esta muitas vezes se insere, criando uma situação de certeza artificial, onde se pode proceder de forma racional e confortável, recorrendo a mecanismos explicados por Simon (1990). Todavia, Levy (1994) defende que os sistemas caóticos nunca passam pelo mesmo estado duas vezes, sendo que estes processos de tomada de decisão assumem que as situações se repetem. Embora, estes processos sejam importantes para combater a incerteza, devem ser analisados de forma cuidadosa, porque uma organização demasiado presa à formalização dos processos, pode perder capacidade para gerar, aquilo a que Nonaka (1988) designa por, conflitos creativos, o que permite que a organização evolua continuamente. Ao criar este tipo de conflitos, a organização está constantemente a rever a sua forma de operar, o que permite desenvolver melhores soluções, o que só é possível se houve liberdade nas unidades que compõe a organização.

Os modelos apresentados no presente trabalho são, precisamente, modelos que permitem isolar um conjunto de variáveis para prever uma variável dependente, não tendo refletindo, diretamente, sobre o sistema complexo em que as empresas se inserem. É, por isso, importante salientar que estes devem ser utilizados como auxiliares na tomada de decisão e que os seus resultados devem ser sempre questionados. Isto porque, questionar os seus resultados, permite gerar conflitos creativos, que levam a melhores soluções. Por exemplo, estes modelos negligenciam o contexto macroeconómico e por isso, é importante

que sejam atualizados. Mas o problema de atualização do modelo só se coloca, se os seus resultados forem questionados.

Tal como Box (1979) refere, todos os modelos estão errados, mas alguns são úteis.

# Capítulo 6

## Conclusões

O presente trabalho procurou comparar vários modelos de classificação, construídos a partir de três métodos diferentes (RL, AMD e RNA), para propôr a aplicação de um modelo de classificação de crédito, às CaR da Nors e criar um modelo que indique um montante de crédito a conceder aos clientes.

Constatou-se que as RNA são o melhor método para a classificação de crédito do Grupo Nors, por ser o método que apresentou melhores resultados, ao longo dos vários testes realizados. Nesse sentido, os modelos de classificação propostos às CaR foram o modelo RNA\_1, o modelo RNA\_3 e o modelo RNA\_4. Isto porque cada um destes modelos tem um propósito de classificação diferente que pode ser interessante para a classificação de crédito do negócio pós venda do Grupo Nors. No entanto, o confronto dos resultados dos três métodos testados, no presente trabalho, foi bastante importante para conferir robustez e confiança aos resultados dos modelos propostos.

Os melhores resultados das RNA são explicados pelo facto destas terem capacidade para resolver problemas de classificação não linearmente separáveis, sendo que, por exemplo, o PMC tem capacidade para aproximar qualquer função contínua. Enquanto os métodos de RL e AMD não têm essa capacidade e por isso, não obtêm tão bons resultados, embora também constituam excelentes ferramentas de classificação.

Este trabalho tem um importante contributo académico, no sentido em que propõe e testa vários modelos de classificação, construídos a partir de três métodos diferentes, comparando o seu desempenho, e aplica-os a um caso de uma empresa comercial, o que possibilitou testar os modelos, em casos reais. Além disso, este trabalho propõe um modelo de definição do montante de crédito a conceder a cada cliente, também devidamente testado em casos reais, sendo

que não se encontrou literatura sobre modelos de classificação de crédito aplicados a empresas comerciais, nem literatura sobre modelos que indiquem o montante de crédito a conceder a clientes por instituições financeiras ou comerciais.

Este trabalho tem, também, um contributo valioso para o Grupo Nors porque propõe vários modelos de classificação de crédito, com aplicações diferentes e com melhor desempenho do que os atuais resultados de classificação da análise de crédito das CaR. Além disto, este trabalho propõe um modelo que indica o montante de crédito a conceder a cada cliente, que constitui um dos maiores problemas das CaR aquando da sua decisão de crédito, até porque, até então, não tinham nenhum indicador para definir o montante de crédito a conceder aos seus clientes. Este trabalho vai de encontro aos objetivos do CSP do Grupo Nors, no sentido em que os modelos propostos permitem tomar decisões de forma mais rápida e sustentada, traduzindo-se num menor custo para as empresas do Grupo que recorrem aos serviços do CSP, gerando valor para o Grupo. Todavia, os resultados do presente trabalho podem ser aplicados a outras indústrias, embora seja necessário ajustar os modelos a esses casos.

Assim, em resposta à questão de investigação, o método quantitativo com melhores desempenho de classificação que poderia ser aplicado ao caso de classificação de crédito do Grupo Nors são as RNA e o modelo que se pode aplicar para definir o montante de crédito a conceder a um cliente de crédito passa por maximizar o retorno esperado com esse cliente, minimizando os custos associados ao serviço de conceder crédito.

## 6.1 Desenvolvimentos futuros

Ao longo deste trabalho foram testados vários métodos de classificação, que foram aplicados à classificação de crédito do Grupo Nors. No entanto, constata-se que existem outros métodos com potencial para resolução deste tipo de problema.

Nesse sentido, avança-se, desde já, uma hipótese de investigação futura. Tal como se referiu ao longo do trabalho, os modelos de RNA foram construídos, testando dois tipos de algoritmos de otimização diferentes. No entanto, seria interessante procurar construir um modelo de RNA a partir de algoritmos de otimização diferentes dos utilizados, de forma a testar outros algoritmos e comparar os seus resultados.

Além disto, também seria interessante testar a capacidade de classificação de outros métodos de classificação. Assim, sugere-se que se teste o método de árvores de decisão para resolução do problema de classificação de crédito do Grupo Nors, de forma a comparar os seus resultados com os apresentados neste trabalho.

# Bibliografia

- Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18(2-3), 59-88.
- Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32(8), 1541-1551.
- Allen, L., DeLong, G., & Saunders, A. (2004). Issues in the credit risk modeling of retail markets. *Journal of Banking & Finance*, 28(4), 727-752.
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609.
- Altman, E. I. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models. *Stern School of Business, New York University*, 9-12.
- Altman, E. I., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, 21(11/12).
- Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. *Journal of banking & finance*, 1(1), 29-54.
- Anderson, R. (2007). *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford University Press.
- Aranha, C., & Iba, H. (2007, September). Modelling cost into a genetic algorithm-based portfolio optimization system by seeding and objective sharing. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on* (pp. 196-203). IEEE.
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the operational research society*, 54(6), 627-635.
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111.
- Behr, P., & Güttler, A. (2007). Credit risk assessment and relationship lending: An empirical analysis of German small and medium-sized enterprises. *Journal of Small Business Management*, 45(2), 194-213
- Bergeron, B. (2002). *Essentials of shared services* (Vol. 26). John Wiley & Sons.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with applications*, 40(1), 356-364.
- Blum, M. P. (1974). The failing company doctrine. *BC Indus. & Com. L. Rev.*, 16, 75
- Box, G. E. (1979). Robustness in the strategy of scientific model building. *Robustness in statistics*, 1, 201-236.
- Chandler, G. G., & Coffman, J. Y. (1977). Using credit scoring to improve the quality of consumer receivables: legal and statistical implications. In *Financial Management Association meetings, Seattle, Washington* (Vol. 85)
- Chandler, G. G., & Coffman, J. Y. (1979). A comparative analysis of empirical vs. judgmental credit evaluation. *Financial Review*, 14(4), 23-23.

- Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3), 465-497
- Coffman, J. Y., & Chandler, G. G. (1983). *Applications of performance scoring to accounts receivable management in consumer credit*. Credit Research center, Krannert Graduate School of Management, Purdue University.
- Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3), 1447-1465.
- Deakin, E. B. (1972). A discriminant analysis of predictors of business failure. *Journal of accounting Research*, 167-179
- Desai, V. S., Conway, D. G., Crook, J. N., & OVERSTREET, G. A. (1997). Credit-scoring models in the credit-union environment using neural networks and genetic algorithms. *IMA Journal of Management Mathematics*, 8(4), 323-346.
- Desai, V. S., Crook, J. N., & Overstreet, G. A. (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research*, 95(1), 24-37.
- Dickinson, M. W., Thornton, A. C., & Graves, S. (2001). Technology portfolio management: optimizing interdependent projects over multiple time periods. *IEEE Transactions on engineering management*, 48(4), 518-527.
- Dimitras, A. I., Zanakis, S. H., & Zopounidis, C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, 90(3), 487-513.
- Durand, D. (1941). Risk elements in consumer instalment financing. *NBER Books*
- Eisenbeis, R. A. (1978). Problems in applying discriminant analysis in credit scoring models. *Journal of Banking & Finance*, 2(3), 205-219.
- Elazouni, A. M., & Metwally, F. G. (2007). Expanding finance-based scheduling to devise overall-optimized project schedules. *Journal of Construction Engineering and Management*, 133(1), 86-90.
- Ericsson, J., & Renault, O. (2006). Liquidity and credit risk. *The Journal of Finance*, 61(5), 2219-2250.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.
- Fraser, D., Gup, B., & Kolari, J. (2001). *Commercial Banking: The Management of Risk (South)*
- Greene, W. (1998). Sample selection in credit-scoring models. *Japan and the world economy*, 10(3), 299-316.
- Greene, W. H. (1992). A statistical model for credit scoring.
- Guimarães, R. C., & Cabral, J. A. S. (2007). *Estatística*. McGraw-Hill, Interamericana de España.
- Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.
- Haykin, S. (1999). *Neural networks a comprehensive introduction*.
- Haykin, S., & Network, N. (2004). A comprehensive foundation. *Neural Networks*, 2 (2004)
- Heffernan, S. (2005). *Modern banking*. John Wiley & Sons.
- Homer, S., & Sylla, R. E. (1996). *A history of interest rates*. Rutgers University Press.
- Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1), 4-37.

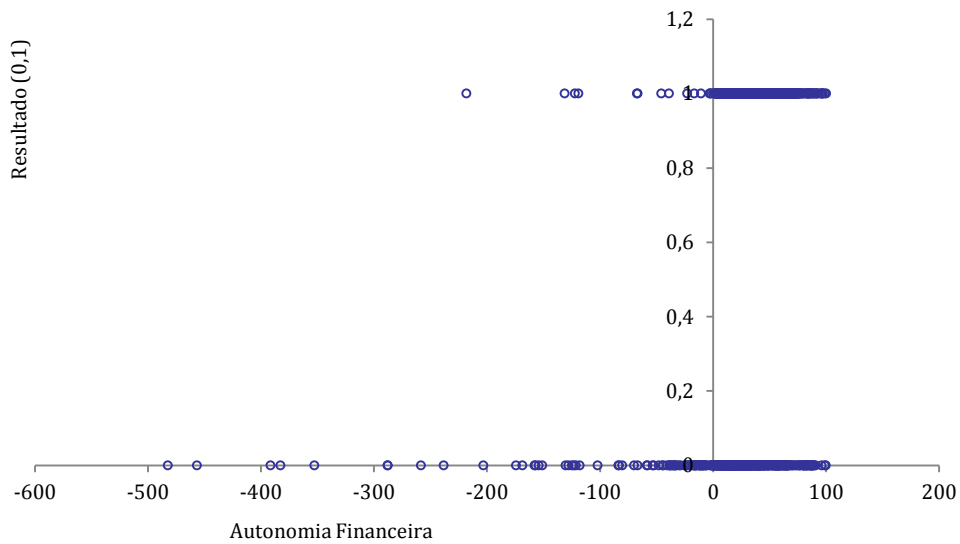
- Jo, H., Han, I., & Lee, H. (1997). Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. *Expert Systems with Applications*, 13(2), 97-108.
- Jorion, P. (2009). Risk management lessons from the credit crisis. *European Financial Management*, 15(5), 923-933.
- Karlik, B., & Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4), 111-122
- Kasabov, N. K. (1996). *Foundations of neural networks, fuzzy systems, and knowledge engineering*. Marcel Alencar.
- Kasabov, N. K. (1996). Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems. *Fuzzy sets and Systems*, 82(2), 135-149.
- Kollias, S., & Anastassiou, D. (1989). An adaptive least squares algorithm for the efficient training of artificial neural networks. *IEEE Transactions on Circuits and Systems*, 36(8), 1092-1101.
- Laguna, M. (1997, April). Metaheuristic optimization with evolver, genocop and optquest. In *EURO/INFORMS Joint International Meeting, Plenaries and Tutorials* (pp. 141-150).
- Laguna, M., & Marti, R. (2002). Neural network prediction in a system for optimizing simulations. *IIE Transactions*, 34(3), 273-282.
- Lam, J. (2014). *Enterprise risk management: from incentives to controls*. John Wiley & Sons.
- Levy, D. (1994). Chaos theory and strategy: Theory, application, and managerial implications. *Strategic management journal*, 167-178.
- Lin, R. C., Sir, M. Y., & Pasupathy, K. S. (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega*, 41(5), 881-892.
- Lingireddy, S., & Brion, G. M. (Eds.). (2005). *Artificial neural networks in water supply engineering*. ASCE Publications.
- MacCrimmon, K. R., Wehrung, D., & Stanbury, W. T. (1988). *Taking risks*. Simon and Schuster.
- Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance*, 46(2), 469-477.
- Markowitz, Harry. "Portfolio selection." *The journal of finance* 7.1 (1952): 77-91.
- Mays, F. E., & Lynas, N. (2004). *Credit scoring for risk managers: The handbook for lenders*. Thomson/South-Western.
- Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research*, 380-395.
- Mester, L. J. (1997). What's the point of credit scoring?. *Business review*, 3, 3-16.
- Minsky, M., & Papert, S. (1972). Artificial intelligence progress report.
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural networks*, 6(4), 525-533.
- Nonaka, I. (1988). Creating organizational order out of chaos: Self-renewal in Japanese firms. *California management review*, 30(3), 57-73.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.
- Ong, C. S., Huang, J. J., & Tzeng, G. H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications*, 29(1), 41-47.

- Poulsen, J., & French, A. (2008). Discriminant function analysis. *San Francisco State University: San Francisco, CA*. <http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf>.
- Quinn, B., Cooke, R., & Kris, A. (2000). *Shared services: mining for corporate gold*. Financial Times, Prentice Hall.
- Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business & Economic Statistics*, 1(2), 101-114.
- Ribeiro Filho, J. L., Treleaven, P. C., & Alippi, C. (1994). Genetic-algorithm programming environments. *Computer*, 27(6), 28-43.
- Rosenblatt, F. (1958). *Two theorems of statistical separability in the perceptron*. United States Department of Commerce.
- Sabzevari, H., Soleymani, M., & Noorbakhsh, E. (2007). A comparison between statistical and data mining methods for credit scoring in case of limited available data. In *Proceedings of the 3rd CRC Credit Scoring Conference, Edinburgh, UK*.
- Schreiner, M. (2000). Credit scoring for microfinance: Can it work?. *Journal of Microfinance/ESR Review*, 2(2), 105-118.
- Senouci, A. B., & Eldin, N. N. (2004). Use of genetic algorithms in resource scheduling of construction projects. *Journal of Construction Engineering and Management*, 130(6), 869-877.
- Shangquan, Z., Shao, M., Horton, R., Lei, T., Qin, L., & Ma, J. (2002). A model for regional optimal allocation of irrigation water resources under deficit irrigation and its applications. *Agricultural Water Management*, 52(2), 139-154
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101-124
- Silva, L. P. D. (2011). *Relatório de estágio curricular no Banco Santander Consumer Portugal, SA: Os Indicadores Económico-Financeiros na Análise de Crédito* (Doctoral dissertation, Instituto Superior de Economia e Gestão)
- Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, 41(1), 1-20.
- Sola, J., & Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3), 1464-1468.
- Sundaram, R. (2001). Multiple discriminant analysis. *Department of Electrical and Computer Engineering Mississippi State University Mississippi State*, [http://www.isip.msstate.edu/publications/courses/ece\\_8443/papers/2001/mmda/p02\\_paper\\_v0.pdf](http://www.isip.msstate.edu/publications/courses/ece_8443/papers/2001/mmda/p02_paper_v0.pdf).
- Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (Vol. 2). Random house.
- Taleb, N. N. (2012). *Antifragile: Things that gain from disorder* (Vol. 3). Random House Incorporated.
- The Economist. 2015. <http://geloookahead.economist.com/storage/>
- Thietart, R. A., & Forgues, B. (1995). Chaos theory and organization. *Organization science*, 6(1), 19-31.
- Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2), 149-172.
- Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit scoring and its applications*. Siam.
- Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, 34(4), 2639-2649.

- Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of marketing*, 68(4), 106-125
- Viganò, L. (1993). A CREDIT SCORING MODEL FOR DEVELOPMENT BANKS: AN AFRICAN CASE STUDY/UN MODELE DE" CREDIT SCORING" POUR LES BANQUES DE DEVELOPPEMENT: UNE ETUDE DE CAS AFRICAIN. *Savings and development*, 441-482
- Watanabe, S. (1985). *Pattern recognition: human and mechanical*. John Wiley & Sons, Inc..
- West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11), 1131-1152.
- Whitley, D. (1994). A genetic algorithm tutorial. *Statistics and computing*, 4(2), 65-85.
- Yin, R. K. (1981). The case study as a serious research strategy. *Science communication*, 3(1), 97-114.
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting research*, 59-82.

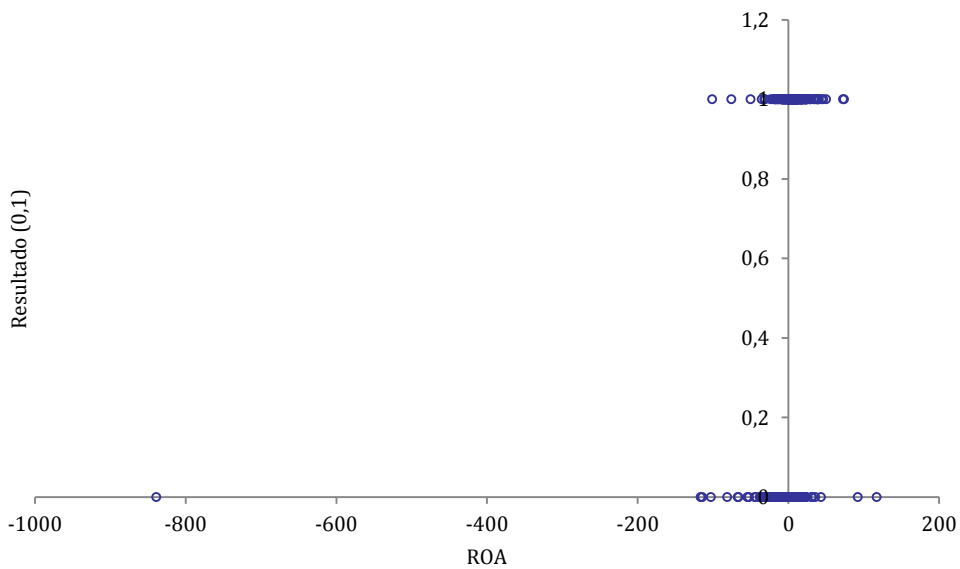
# Anexo 1. Gráficos de Dispersão

**Gráfico de dispersão de Resultado (0,1) comparado a Autonomia Financeira**



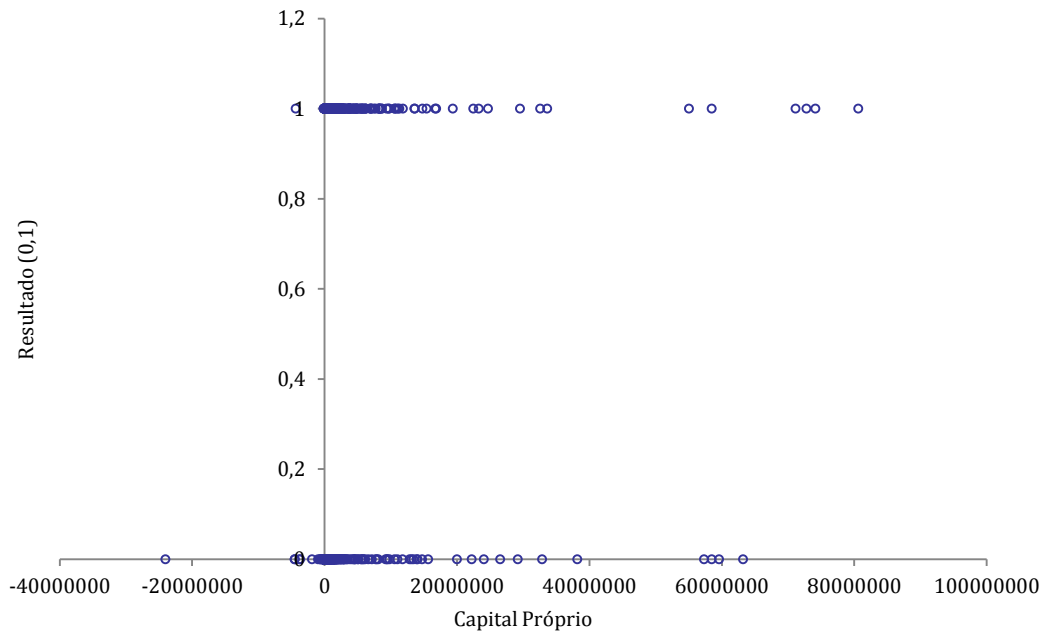
**Figura 6:** Gráfico de Dispersão entre Resultado e Autonomia Financeira

**Gráfico de dispersão de Resultado (0,1) comparado a ROA**



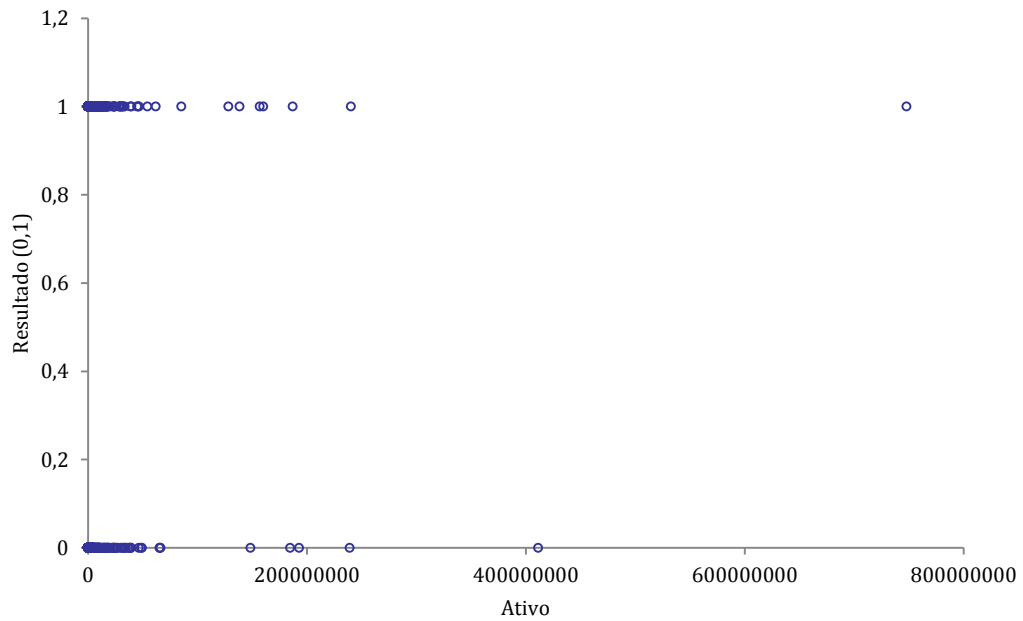
**Figura 7:** Gráfico de Dispersão entre resultado e ROA

**Gráfico de dispersão de Resultado (0,1) comparado a Capital Próprio**



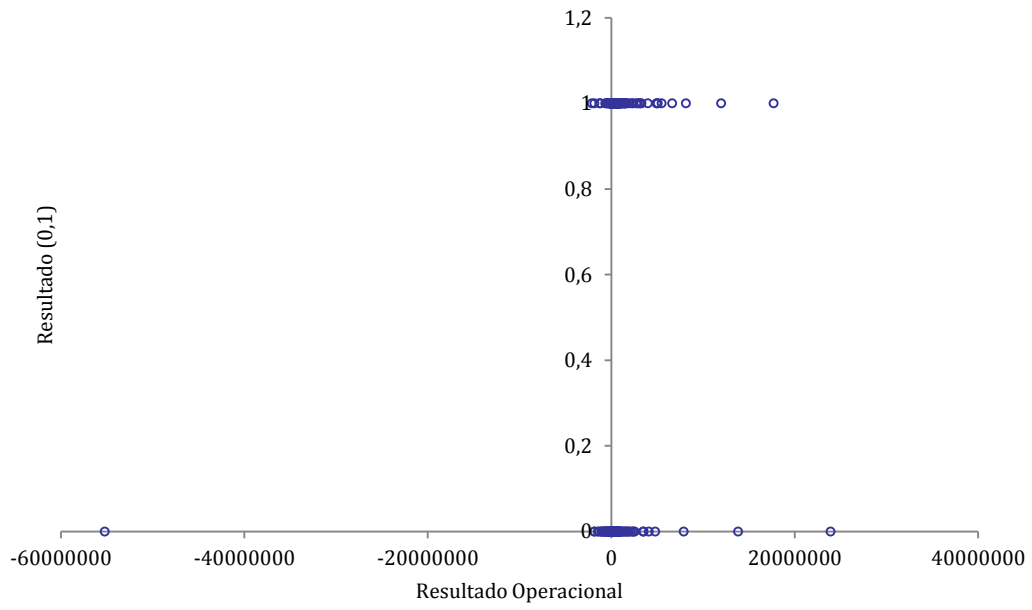
**Figura 8: Gráfico de Dispersão entre Resultado e Capital Próprio**

**Gráfico de dispersão de Resultado (0,1) comparado a Ativo**



**Figura 9: Gráfico de Dispersão entre Resultado e Ativo**

### Gráfico de dispersão de Resultado (0,1) comparado Resultado Operacional



**Figura 6:** Gráfico de Dispersão entre Resultado e Resultado Operacional

## Anexo 2. RNA\_1

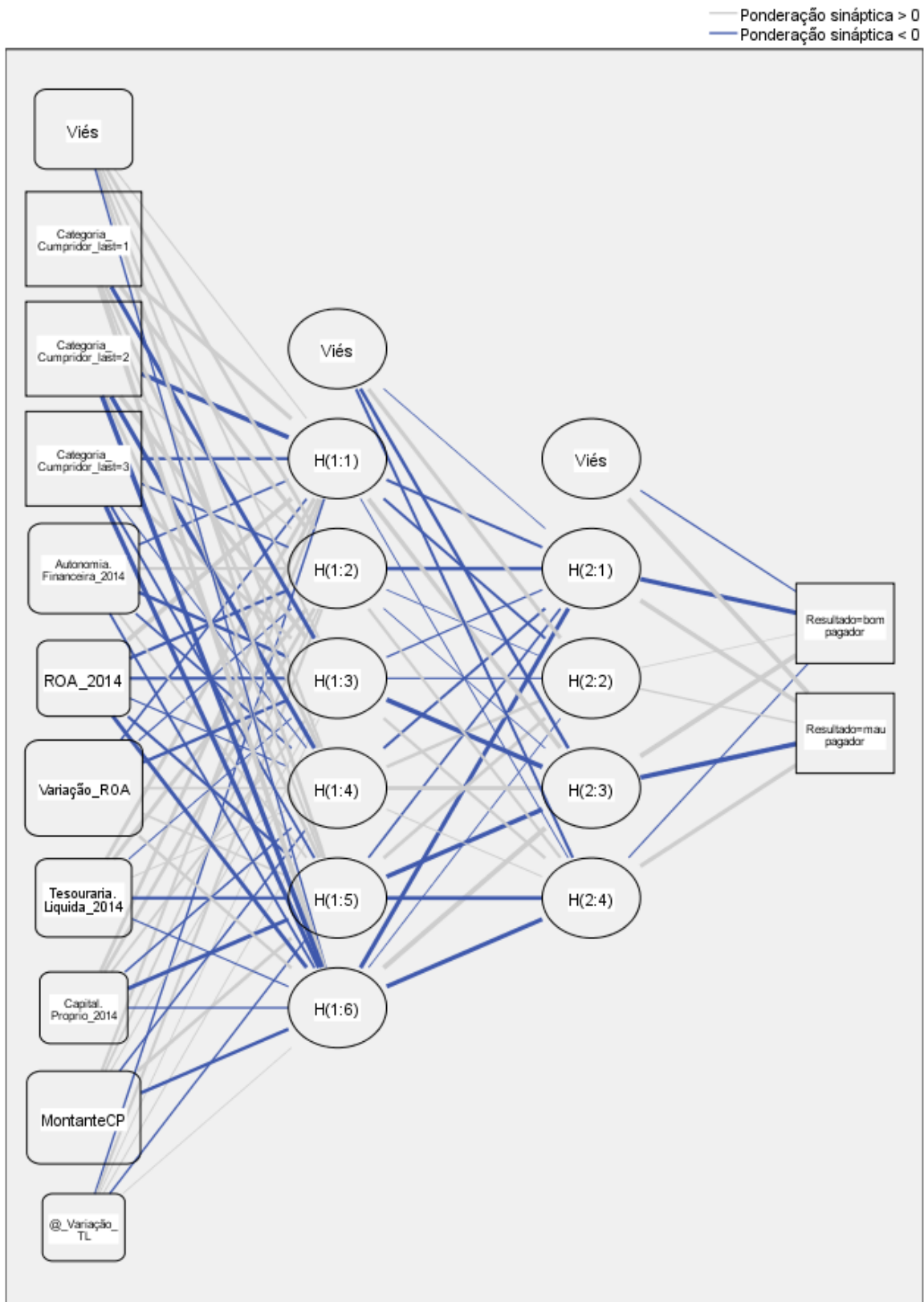
### Informações de rede

Camada de entrada	Fatores	1	Categoria_Cu mpridor_last
	Covariáveis	1	Autonomia. Financeira_20 14
		2	ROA_2014
		3	% Variação_ROA
		4	Tesouraria. Liquida_2014
		5	Capital. Proprio_2014
		6	Montante/CP
		7	% _Variação_TL
	Número de unidades <sup>a</sup>	10	
	Método de reescalonamento para covariáveis	Padronizado	
Camadas ocultas	Número de camadas ocultas	2	
	Número de Unidades na Camada Oculta 1 <sup>a</sup>	6	
	Número de Unidades na Camada Oculta 2 <sup>a</sup>	4	
	Função de ativação	Curva sigmoide	
Camada de saída	Variáveis dependentes	1	Resultado
	Número de unidades	2	
	Função de ativação	Identidade	
	Função de erro	Soma dos Quadrados	

a. Excluindo a unidade de viés

**Figura 7:** Informações RNA\_1

Os quadrados à direita na Figura 8, 12 (Anexo 4) e 16 (Anexo 5) representam a variável dependente (categoria de pagador do cliente) dos modelos de classificação construídos a partir das RNA, designada por “Resultado do Crédito” nestes modelos.



Função de ativação de camada oculta: Curva sigmoide

Função de ativação de camada de saída: Identidade

**Figura 8:** Arquitectura RNA\_1

### Sumarização do modelo

Treinamento	Soma dos erros quadráticos	135,831
	Porcentagem de predições incorretas	24,3%
	Regra de parada usada	Tempo máximo de treinamento (25 minutos) excedido
	Tempo de treinamento	0:25:00,42
Testes	Soma dos erros quadráticos	58,041
	Porcentagem de predições incorretas	24,2%

Variável Dependente: Resultado

**Figura 9:** Resumo do treino da RNA\_1

### Classificação

Amostra	Observado	Predito		Porcentagem Correta
		bom pagador	mau pagador	
Treinamento	bom pagador	363	57	86,4%
	mau pagador	125	204	62,0%
	Porcentagem Global	65,2%	34,8%	75,7%
Testes	bom pagador	140	26	84,3%
	mau pagador	52	104	66,7%
	Porcentagem Global	59,6%	40,4%	75,8%

Variável Dependente: Resultado

**Figura 10:** Matriz de Classificação RNA\_1

## Anexo 3. Modelos de Classificação sem variável Histórico Interno

O modelo RNA\_2 foi construído com duas camadas ocultas, a primeira com 5 nós e a segunda com 3 nós, utilizando funções de ativação tangente hiperbólica nas camadas ocultas e a função identidade na camada de *output*. O tipo de treino foi o *online*, com o algoritmo de otimização descida do gradiente. Os seguintes quadros (ver Tabelas 21 e 22) representam os resultados do modelo RNA\_2.

Observado	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
bom pagador		155	19	89,1%
mau pagador		112	39	25,8%
% Global		82,2%	17,8%	59,7%

Tabela 21: Matriz de classificação RNA\_2

Variáveis	Importância	Importância Normalizada
Autonomia Financeira	0,149	62,9%
ROA	0,207	87,4%
% Variação ROA	0,237	100%
Tesouraria Líquida	0,107	45,1%
Capital Próprio	0,119	50,4%
Montante/CP	0,180	76%

Tabela 22: Importância das variáveis RNA\_2

Os quadros seguintes ilustram os resultados do modelo de classificação RL\_2 (Tabelas 23 e 24).

	<b>Predito</b>			
	<b>Matriz de classificação</b>	<b>bom pagador</b>	<b>mau pagador</b>	<b>% Correto</b>
<b>Observado bom pagador</b>		507	77	86,82%
<b>mau pagador</b>		342	159	31,75%
<b>% Global</b>		78,2%	21,8%	61,38%

Tabela 23: Matriz de classificação do modelo RL\_2

<b>Função de Regressão</b>	<b>Coefficiente</b>	<b>P value</b>
<b>Constante</b>	-0,095128325	0,3752
<b>Autonomia.Financeira_2014</b>	-2,33666E-05	0,5402
<b>ROA_2014</b>	0,025308174	< 0.0001
<b>GRA_2014</b>	0,001251503	0,0387
<b>EBITDA_2014</b>	-4,88724E-09	0,9091
<b>Tesouraria.Liquida_2014</b>	-5,38252E-10	0,9508
<b>Capital.Proprio_2014</b>	1,65003E-08	0,1237
<b>Montante/CP</b>	-0,349389537	0,0412

Tabela 24: Importância das variáveis do modelo RL\_2

Os quadros seguintes indicam os resultados do modelo AMD\_2 (Tabelas 25 e 26).

Observado	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
	bom pagador	314	270	56,7%
	mau pagador	202	299	53,8%
	% Global	47,5%	52,5%	56,5%

Tabela 25: Matriz de classificação AMD\_2

Função Discriminante	Coefficiente
Autonomia.Financeira_2014	1,29167E-05
ROA_2014	-0,007467978
% Variação_ROA	-0,003487427
GRA_2014	-0,001275704
EBITDA_2014	-1,1711E-08
Capital.Proprio_2014	-1,416E-08
%_Variação_CP	0,00622265
Montante/CP	0,029508363

Tabela 26: Importância das variáveis do modelo AMD\_2

Neste caso, verifica-se que o modelo RL\_2 é o que tem melhores resultados.

## Anexo 4. Modelos de Classificação (RNA\_3; RL\_3; AMD\_3)

As figuras seguintes (Figuras 11, 12, 13 e 14) representam os resultados e processos de treino do modelo RNA\_3. Este modelo foi com duas camadas ocultas, a primeira com 8 nós e a segunda com 4 nós, utilizando a função de ativação tangente hiperbólica nas camadas ocultas. O treino utilizado foi o *online*, com o algoritmo descida do gradiente.

### Informações de rede

Camada de entrada	Covariáveis	1	Autonomia. Financeira_2014
		2	% Variação_ROA
		3	Tesouraria. Liquida_2014
		4	Capital. Proprio_2014
		5	Ativo_2014
		6	Montante/CP
		7	Resultado Operacional 2014
		8	ROA_2014
	Número de unidades <sup>a</sup>		8
	Método de reescalonamento para covariáveis		Normalizado ajustado
Camadas ocultas	Número de camadas ocultas		2
	Número de Unidades na Camada Oculta 1 <sup>a</sup>		8
	Número de Unidades na Camada Oculta 2 <sup>a</sup>		4
	Função de ativação		Tangente hiperbólica
Camada de saída	Variáveis dependentes	1	Resultado 2 CAT
	Número de unidades		2
	Função de ativação		Identidade
	Função de erro		Soma dos Quadrados

a. Excluindo a unidade de viés

**Figura 11:** Informações do modelo RNA\_3



Função de ativação de camada oculta: Tangente hiperbólica

Função de ativação de camada de saída: Identidade

**Figura 12:** Arquitectura da rede RNA\_3

### Sumarização do modelo

Treinamento	Soma dos erros quadráticos	36,949
	Porcentagem de predições incorretas	27,6%
	Regra de parada usada	Tempo máximo de treinamento (30 minutos) excedido
	Tempo de treinamento	0:30:00,06
Testes	Soma dos erros quadráticos	18,699
	Porcentagem de predições incorretas	25,7%

Variável Dependente: Resultado 2 CAT

**Figura 13:** Resumo do treino do modelo RNA\_3

### Classificação

Amostra	Observado	Predito		Porcentagem Correta
		bom pagador	mau pagador	
Treinamento	bom pagador	105	8	92,9%
	mau pagador	51	50	49,5%
	Porcentagem Global	72,9%	27,1%	72,4%
Testes	bom pagador	56	5	91,8%
	mau pagador	21	19	47,5%
	Porcentagem Global	76,2%	23,8%	74,3%

Variável Dependente: Resultado 2 CAT

**Figura 14:** Matriz de Classificação do modelo RNA\_3

Os quadros seguintes ilustram os resultados do modelo AMD\_3 (Tabelas 27 e 28).

Observado	Predito			
	Matriz de classificação	bom pagador	mau pagador	% Correto
	bom pagador	152	22	87,4%
	mau pagador	72	69	48,9%
	% Global	71,1%	28,8%	70,2%

Tabela 27: Matriz de classificação do modelo AMD\_3

Função Discriminante	Coefficiente
Autonomia.Financeira_2014	-0,008681381
ROA_2014	-0,002250758
% Variação_ROA	-0,003499648
GRA_2014	0,000417261
Capital.Proprio_2014	1,43887E-08
%_Variação_CP	0,044561578
Ativo_2014	3,20654E-06
Montante/CP	0,107501202
Vendas 2014	-2,18169E-06
Resultado Operacional 2014	-3,54201E-07

Tabela 28: Função discriminante do modelo AMD\_3

Os quadros seguintes representam os resultados do modelo RL\_3. (Tabelas 29 e 30)

<b>Observado</b>	<b>Predito</b>			
	<b>Matriz de classificação</b>	<b>bom pagador</b>	<b>mau pagador</b>	<b>% Correto</b>
	<b>bom pagador</b>	161	13	92,53%
	<b>mau pagador</b>	79	62	43,97%
	<b>% Global</b>	76,2%	23,8%	70,79%

Tabela 29: Matriz de classificação do modelo RL\_3

<b>Função de Regressão</b>	<b>Coefficiente</b>	<b>P value</b>
<b>Constante</b>	-0,014243906	0,9546
<b>Autonomia Financeira</b>	0,014028657	0,0002
<b>ROA</b>	0,005706033	0,5588
<b>Varição ROA</b>	0,019876579	0,1268
<b>GRA</b>	-0,000148253	0,8935
<b>Tesouraria Líquida</b>	-1,21918E-08	0,8248
<b>Capital Proprio</b>	-2,49923E-08	0,3710
<b>Varição CP</b>	-0,057337008	0,1730
<b>Ativo</b>	-3,59624E-06	0,0286
<b>Montante/CP</b>	-0,46014468	0,1988
<b>Vendas</b>	2,53883E-06	0,0442
<b>Resultado Operacional</b>	4,07638E-07	0,0348

Tabela 30: Função de regressão do modelo RL\_3

Mais uma vez, verifica-se que as RNA têm melhores resultados de classificação. No entanto, neste caso, a RL tem melhores resultados de classificação que a AD. O modelo RL\_3 foi construído a partir de um maior

número de variáveis independentes, sendo que o aumento do número de variáveis levou à melhoria dos resultados de classificação, ao tornar o modelo, estatisticamente, mais completo. No entanto, no caso da AD, o aumento do número de variáveis, a partir do número que o AMD\_3 apresenta, não levou à melhoria dos resultados de classificação. Por outro lado, o modelo RNA\_3 é o que apresenta um menor número de variáveis independentes e melhores resultados de classificação. Mais uma vez, verifica-se que as RNA têm um peaking phenomenon com um número de variáveis menor, sendo a RL o método que necessita de mais variáveis para obter bons resultados, neste caso.

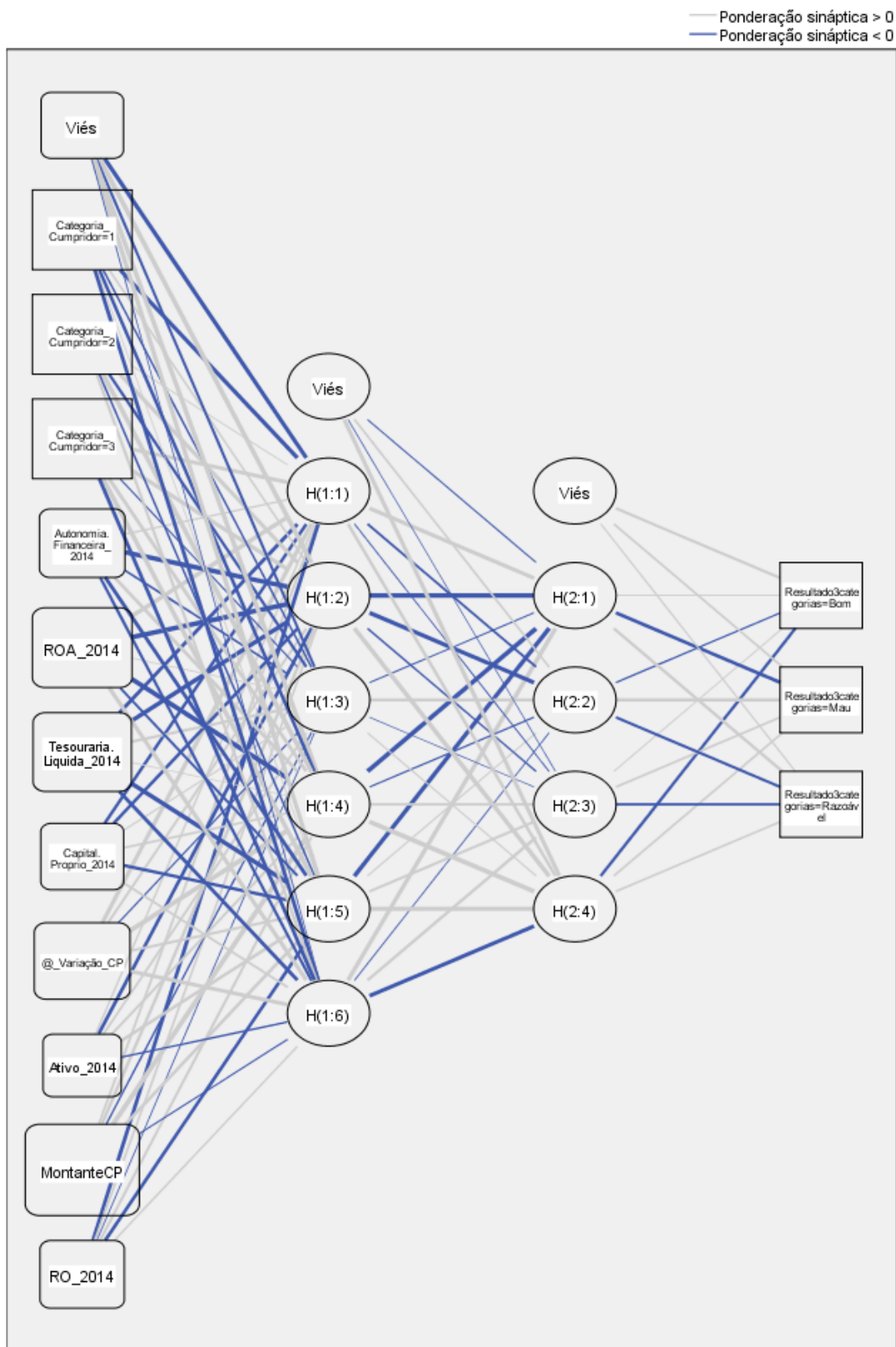
## Anexo 5. RNA\_4

As Figuras 15, 16, 17, 18 e 19 apresentam os resultados e processos de treino do modelo RNA\_4. Este modelo foi construído com duas camadas ocultas, 6 nós na primeira e 4 nós na segunda, com função de ativação tangente hiperbólica, nas camadas ocultas. O tipo de treino foi o *online*, com algoritmo descida de gradiente. Neste caso, os dados foram normalizados ajustados.

Informações de rede			
Camada de entrada	Fatores	1	Categoria_Cu mpridor
	Covariáveis	1	Autonomia. Financeira_20 14
		2	ROA_2014
		3	Tesouraria. Liquida_2014
		4	Capital. Proprio_2014
		5	% _Variação_CP
		6	Ativo_2014
		7	Montante/CP
		8	RO_2014
	Número de unidades <sup>a</sup>		11
Método de reescalonamento para covariáveis		Normalizado ajustado	
Camadas ocultas	Número de camadas ocultas		2
	Número de Unidades na Camada Oculta 1 <sup>a</sup>		6
	Número de Unidades na Camada Oculta 2 <sup>a</sup>		4
	Função de ativação		Tangente hiperbólica
Camada de saída	Variáveis dependentes	1	Resultado 3 categorias
	Número de unidades		3
	Função de ativação		Identidade
	Função de erro		Soma dos Quadrados

a. Excluindo a unidade de viés

**Figura 15:** Informações do modelo RNA\_4



Função de ativação de camada oculta: Tangente hiperbólica

Função de ativação de camada de saída: Identidade

**Figura 16:** Arquitectura do modelo RNA\_4

### Sumarização do modelo

Treinamento	Soma dos erros quadráticos	140,632
	Porcentagem de predições incorretas	33,0%
	Regra de parada usada	Tempo máximo de treinamento (35 minutos) excedido
	Tempo de treinamento	0:35:00,53
Testes	Soma dos erros quadráticos	59,902
	Porcentagem de predições incorretas	29,6%

Variável Dependente: Resultado 3 categorias

Figura 17: Resumo do treino do modelo RNA\_4

### Classificação

Amostra	Observado	Predito			Porcentagem Correta
		Bom	Mau	Razoável	
Treinamento	Bom	207	5	19	89,6%
	Mau	20	86	92	43,4%
	Razoável	53	31	154	64,7%
	Porcentagem Global	42,0%	18,3%	39,7%	67,0%
Testes	Bom	98	2	6	92,5%
	Mau	10	41	35	47,7%
	Razoável	20	14	68	66,7%
	Porcentagem Global	43,5%	19,4%	37,1%	70,4%

Variável Dependente: Resultado 3 categorias

Figura 18: Matriz de classificação do modelo RNA\_4

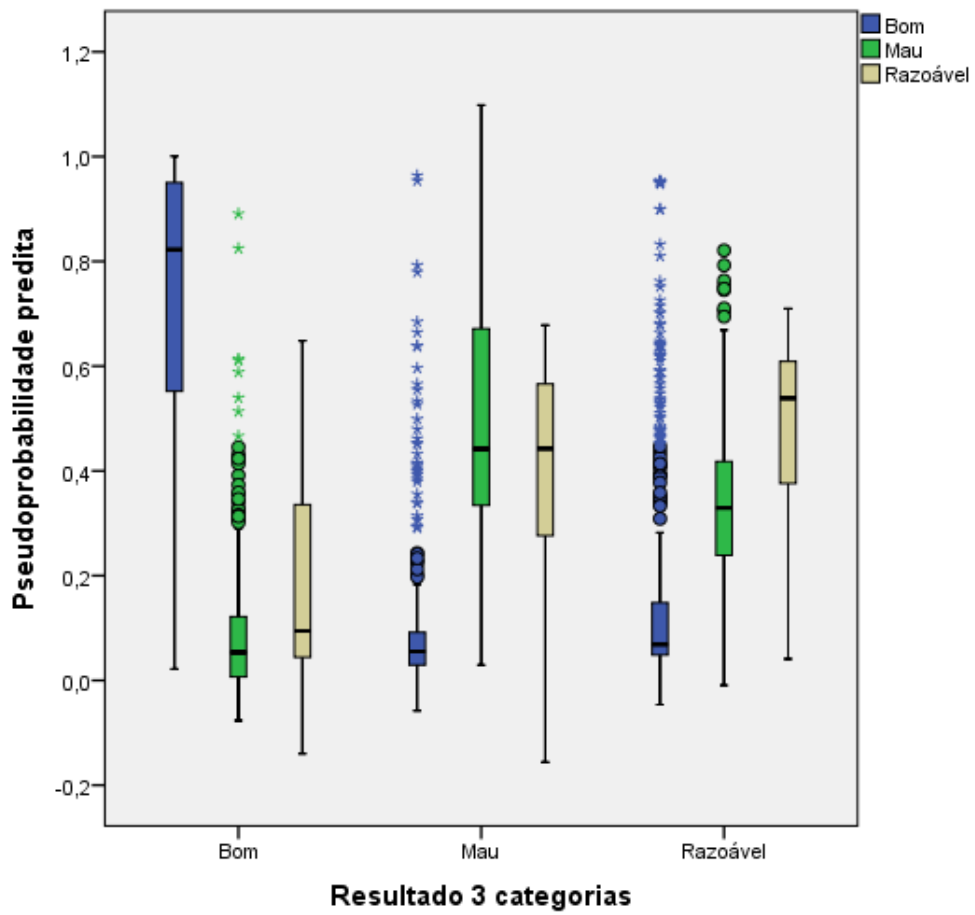


Figura 19: Probabilidade predita de classificação do modelo RNA\_4

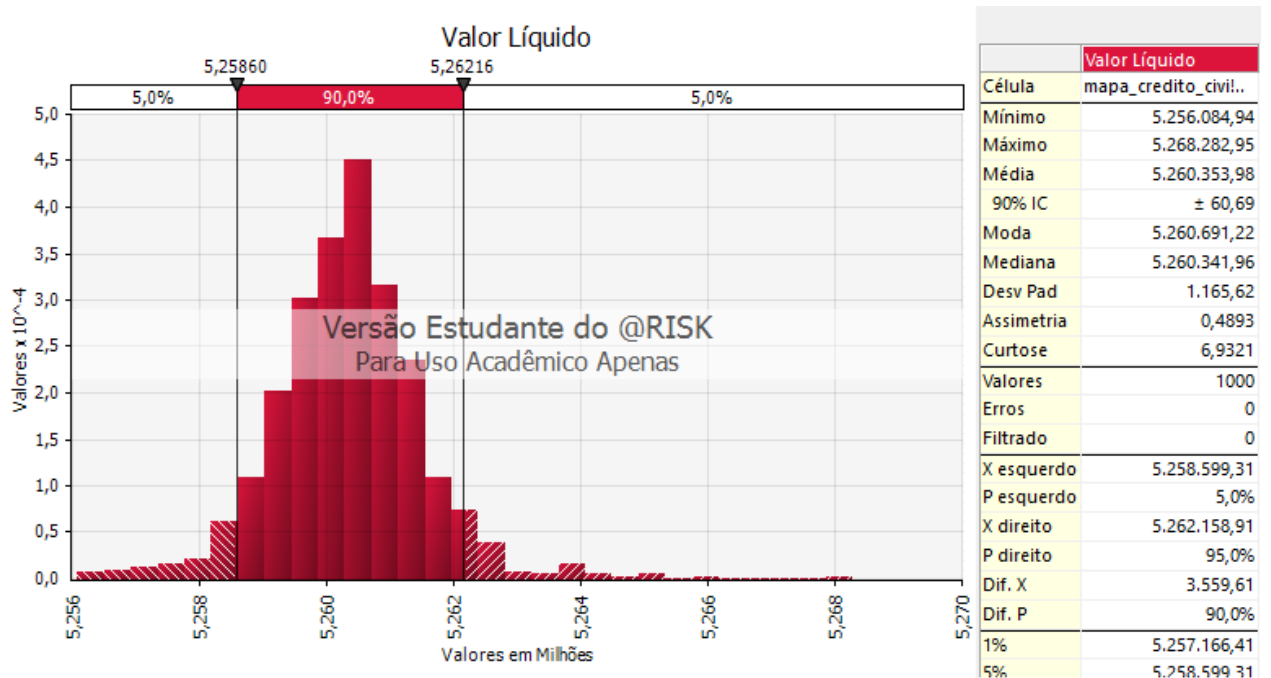
## Anexo 6. Classificação de Crédito do modelo RNA\_4

Ciente	AF	ROA	TL	RO	A	CP	CP_variação	Montante/CP	CAT_histórico	SelectedProbability	PredictedValue
1	29,8	-4,04	8185,2	-32391,4	802489,7	238763,2	-11,9%	0,0	2	0,34	Bom
2	47,0	-6,70	-29818,7	-6884,0	102589,9	48216,0	97,2%	0,1	2	0,38	Mau
3	69,9	12,14	260875,9	46496,5	383139,9	267920,8	15,2%	0,0	2	0,14	Bom

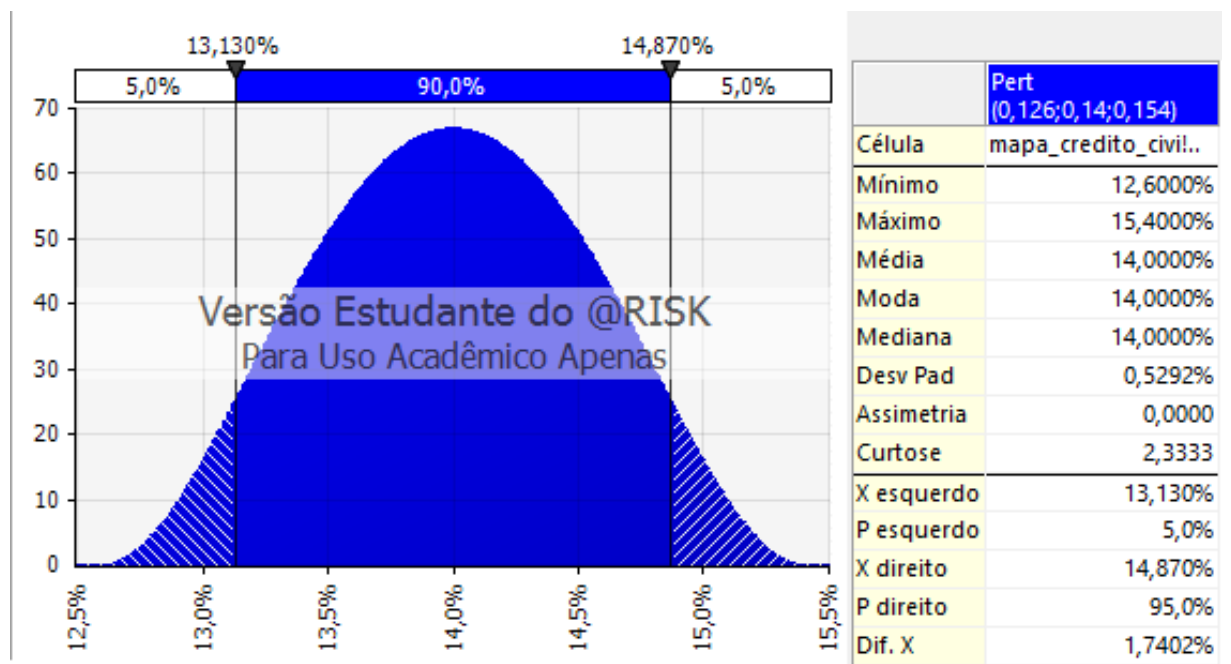
**Figura 20:** Classificações de crédito do modelo RNA\_4

<sup>8</sup> CAT\_Histórico refere-se à variável categórica Histórico Interno

## Anexo 7. Resultados do RiskOptimizer a atribuir crédito aos clientes 1, 2 e 3



**Figura 21:** Distribuição do Valor líquido do Portfólio da CiviParts pós atuação do RiskOptimizer para conceder crédito aos 3 clientes



**Figura 22:** Distribuição Pert da tx de risco do Cliente 3

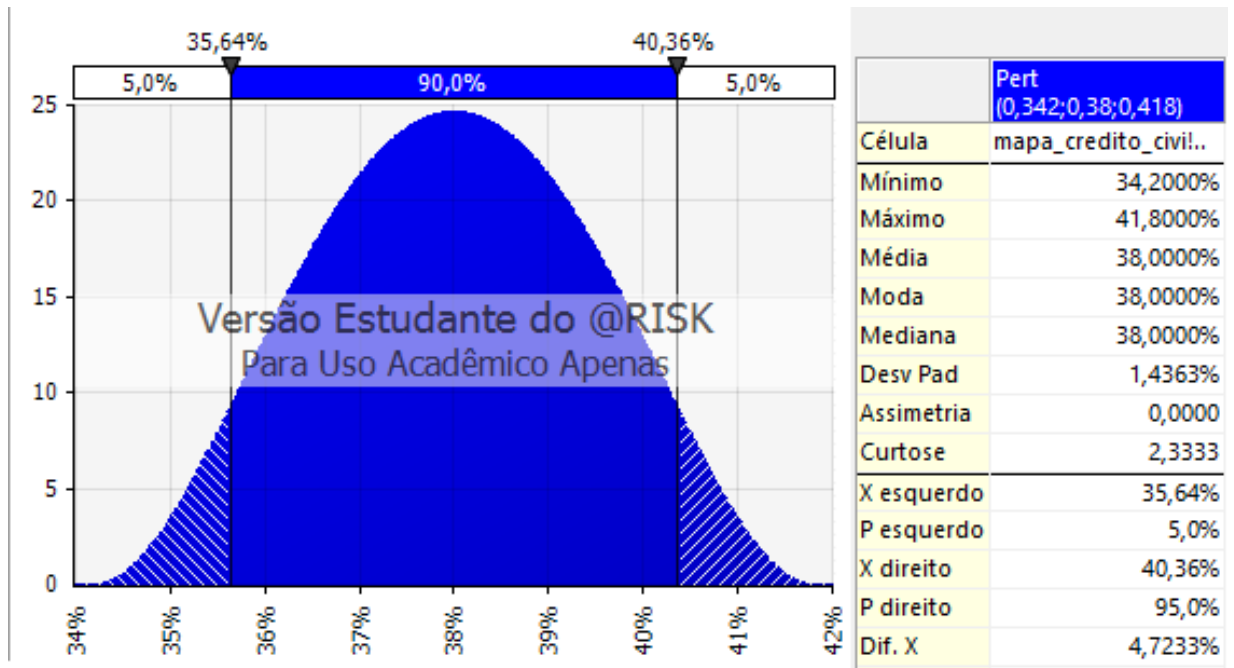


Figura 23: Distribuição Pert da tx de risco do Cliente 2

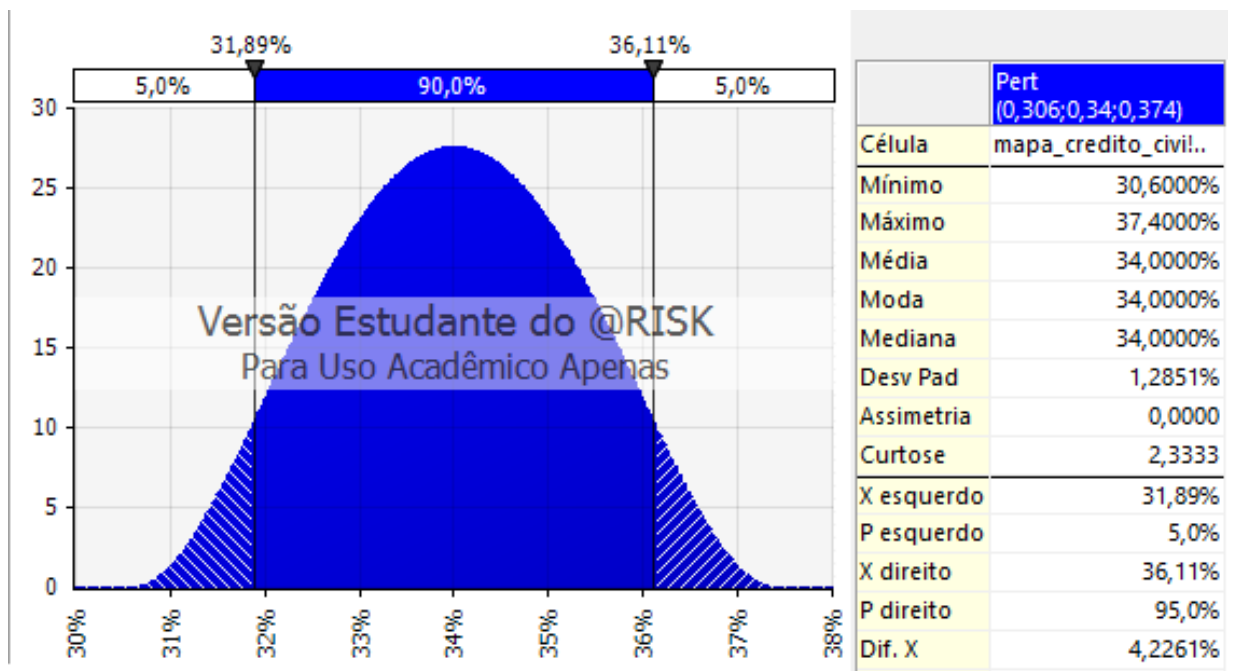


Figura 24: Distribuição Pert da tx de risco do Cliente 1

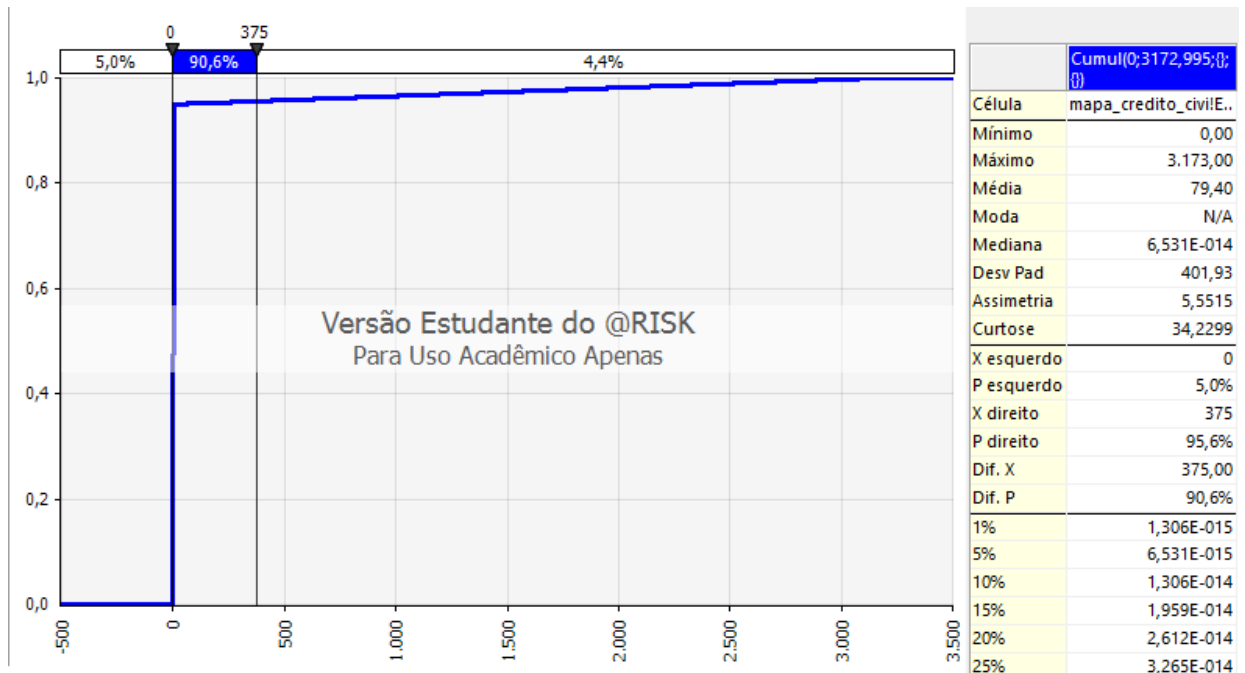


Figura 25: Distribuição Cumulativa das Imparidades Previstas do Cliente 3

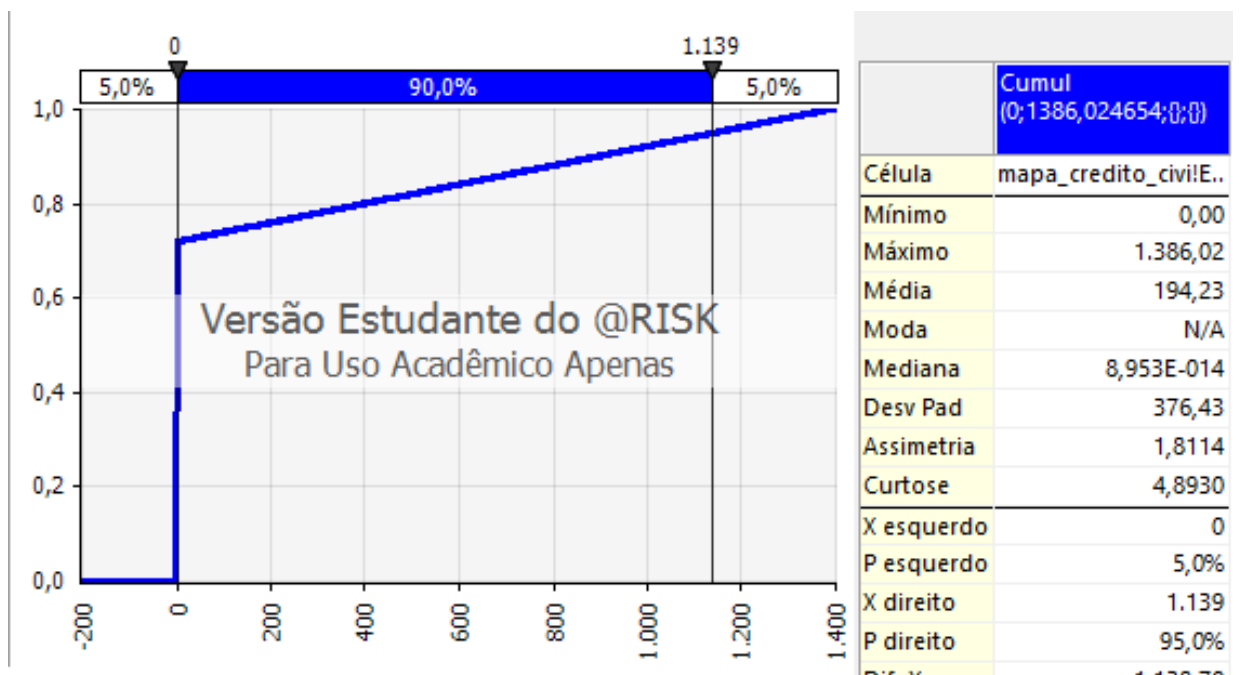


Figura 26: Distribuição Cumulativa das Imparidades Previstas do Cliente 2

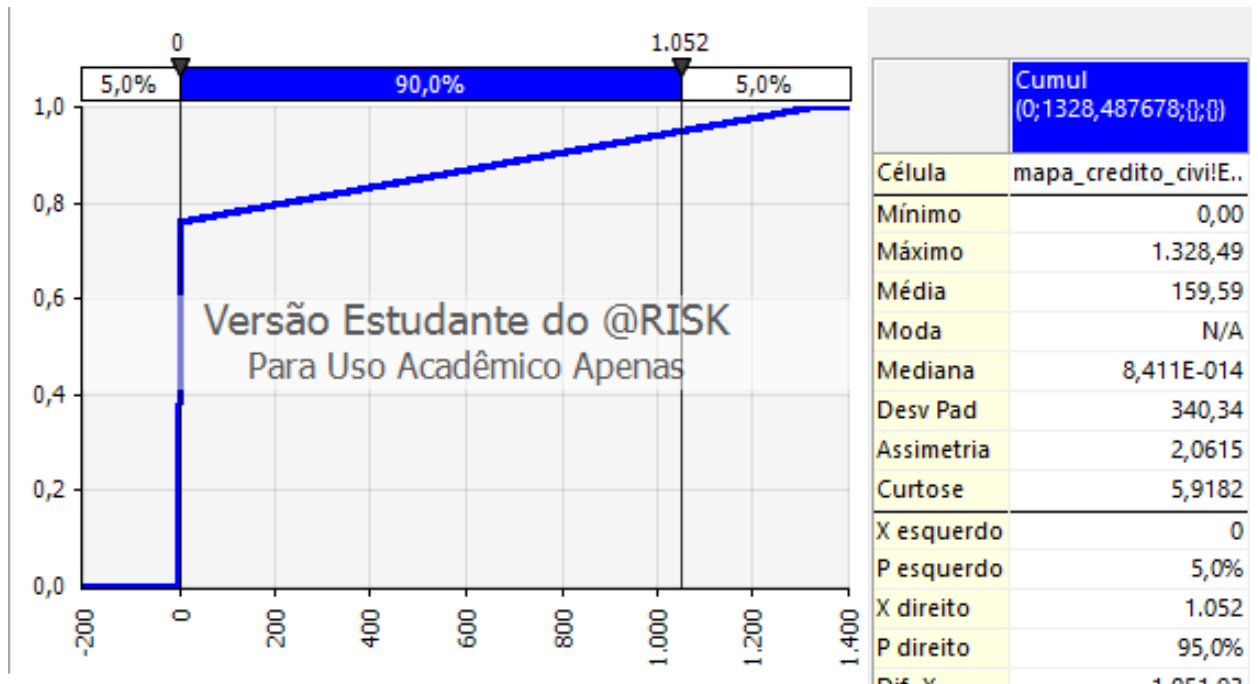


Figura 27: Distribuição Cumulativa das Imparidades Previstas do Cliente 1

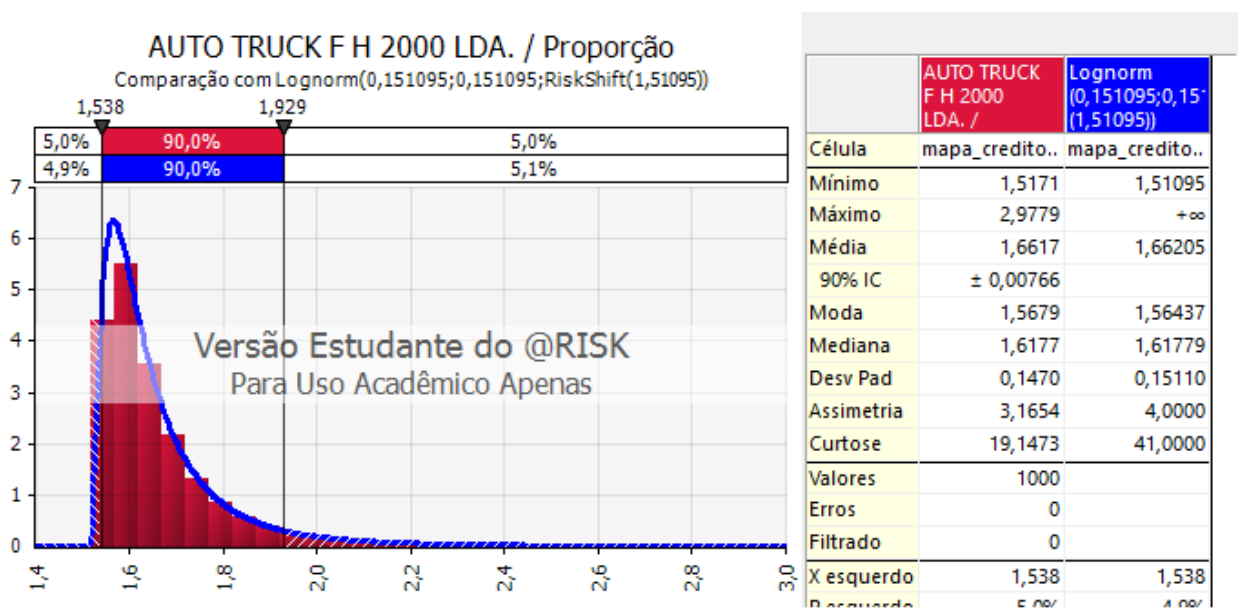


Figura 28: Distribuição da Proporção entre FE e Plafond disponibilizado

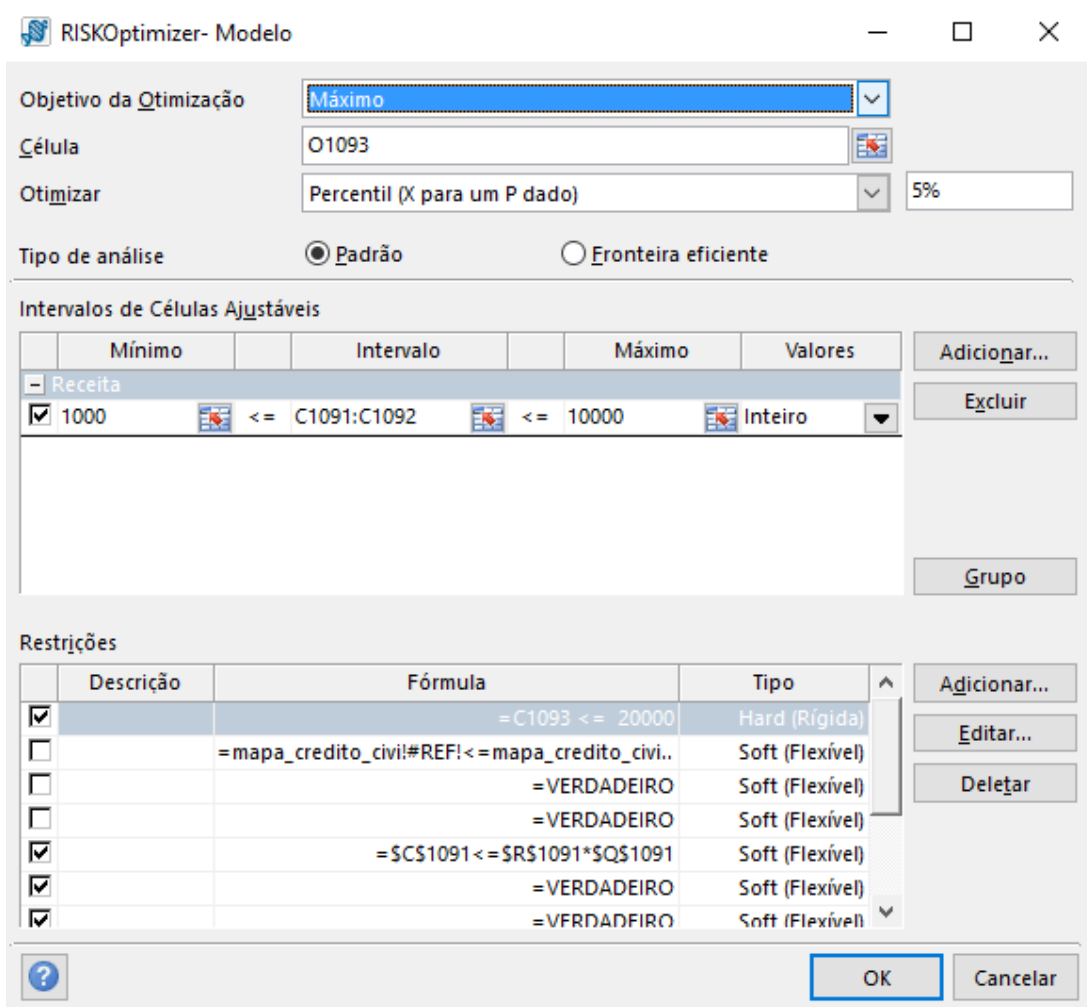


Figura 29: Modelo RiskOptimizer com restrições definidas

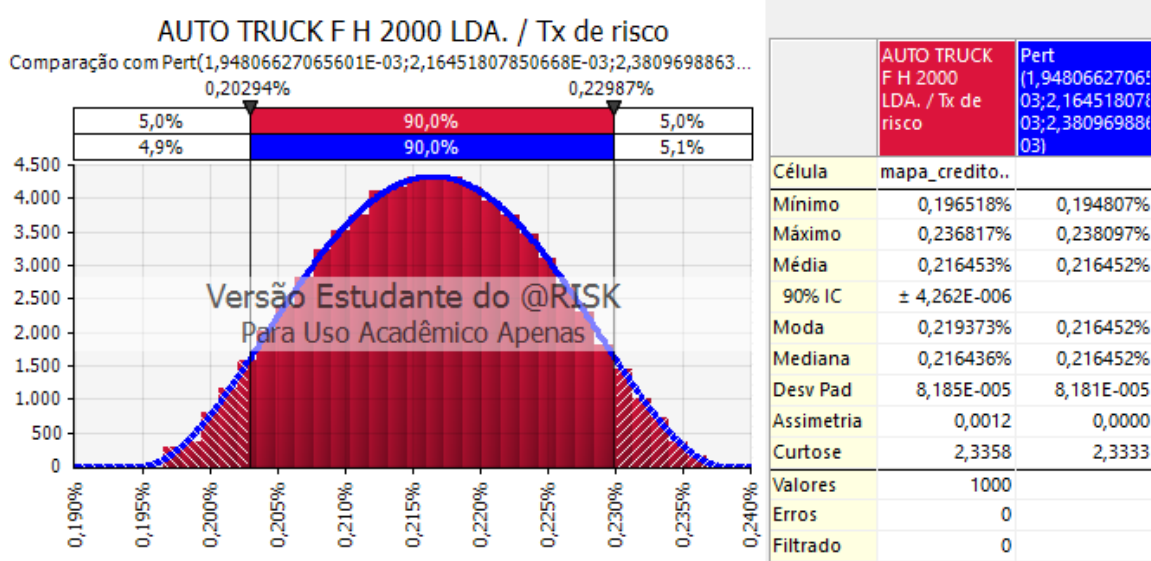
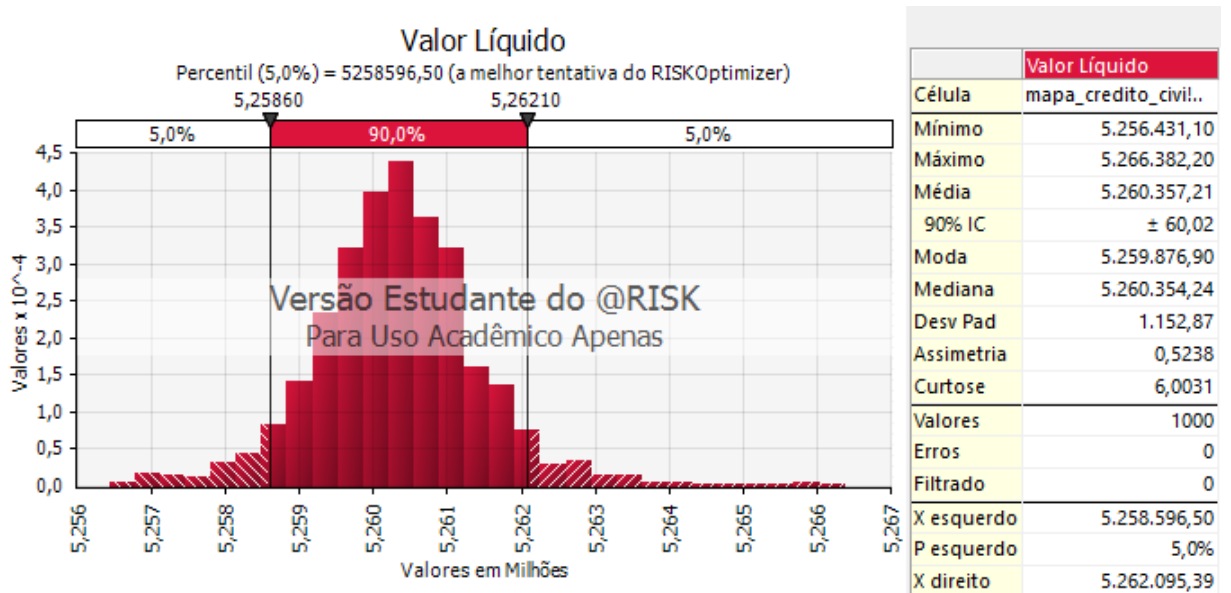
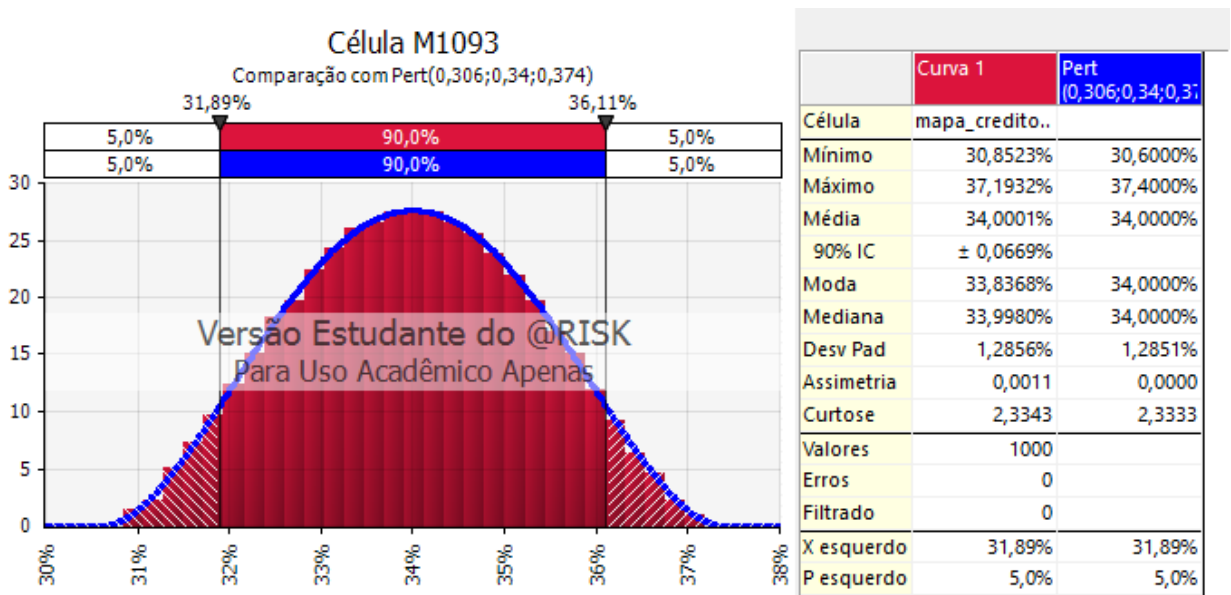


Figura 30: Distribuição Pert da tx de risco ponderada utilizada para calcular as imparidades previstas dos clientes já existentes no Portfólio da CiviParts

## Anexo 8. Resultados do RiskOptimizer a atribuir crédito aos clientes 1 e 2



**Figura 31:** Distribuição do Valor Líquido do Portfólio após a intervenção do RiskOptimizer para conceder crédito aos clientes 1 e 2



**Figura 32:** Distribuição Pert da tx de risco do Cliente 1

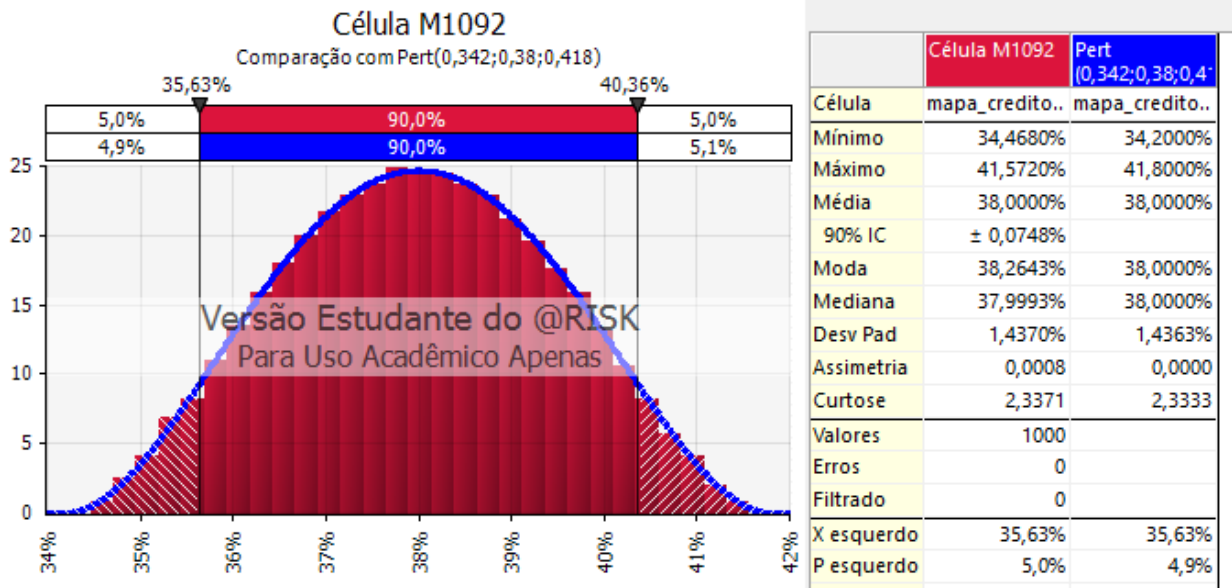


Figura 33: Distribuição Pert da tx de risco do Cliente 2

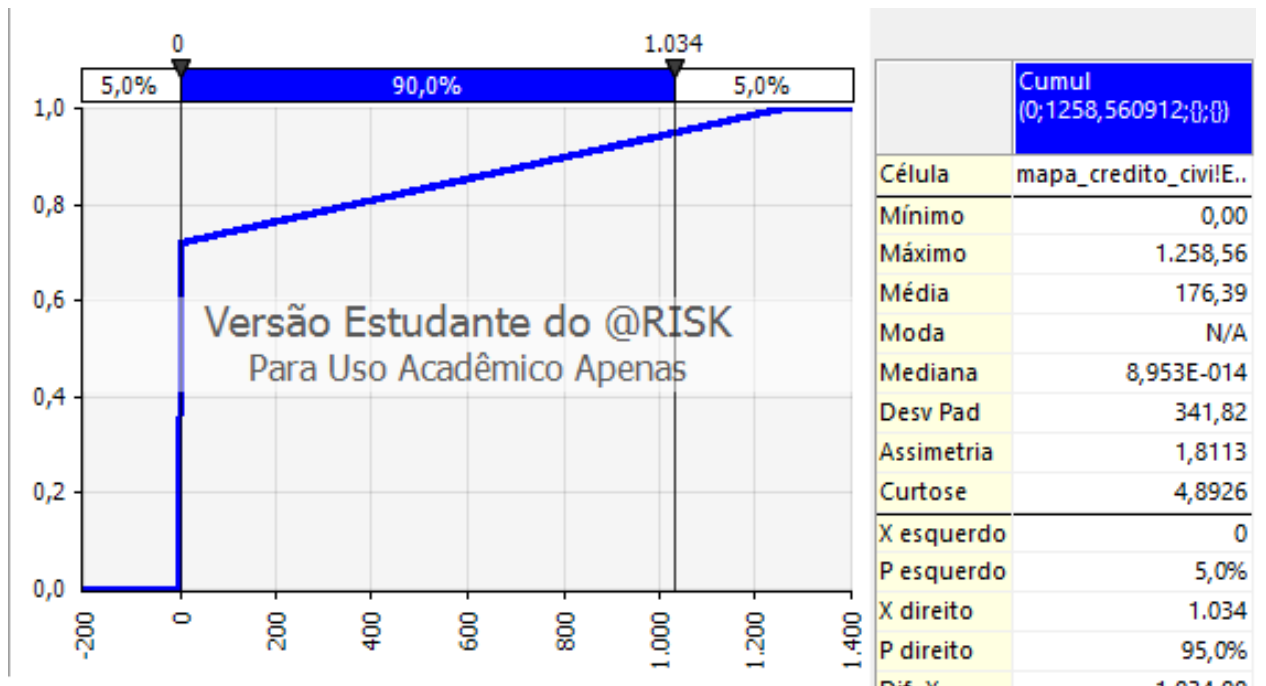


Figura 34: Distribuição Cumulativa das Imparidades Previstas do Cliente 2

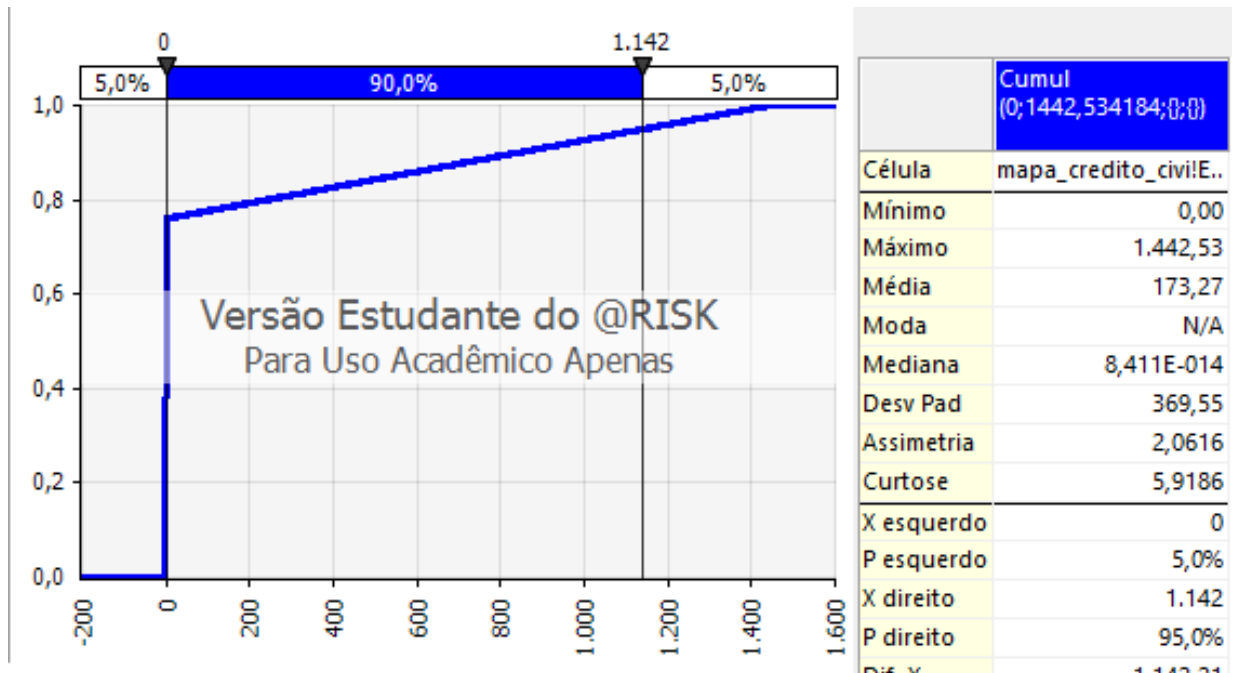


Figura 35: Distribuição Cumulativa das Imparidades Previstas do Cliente 1

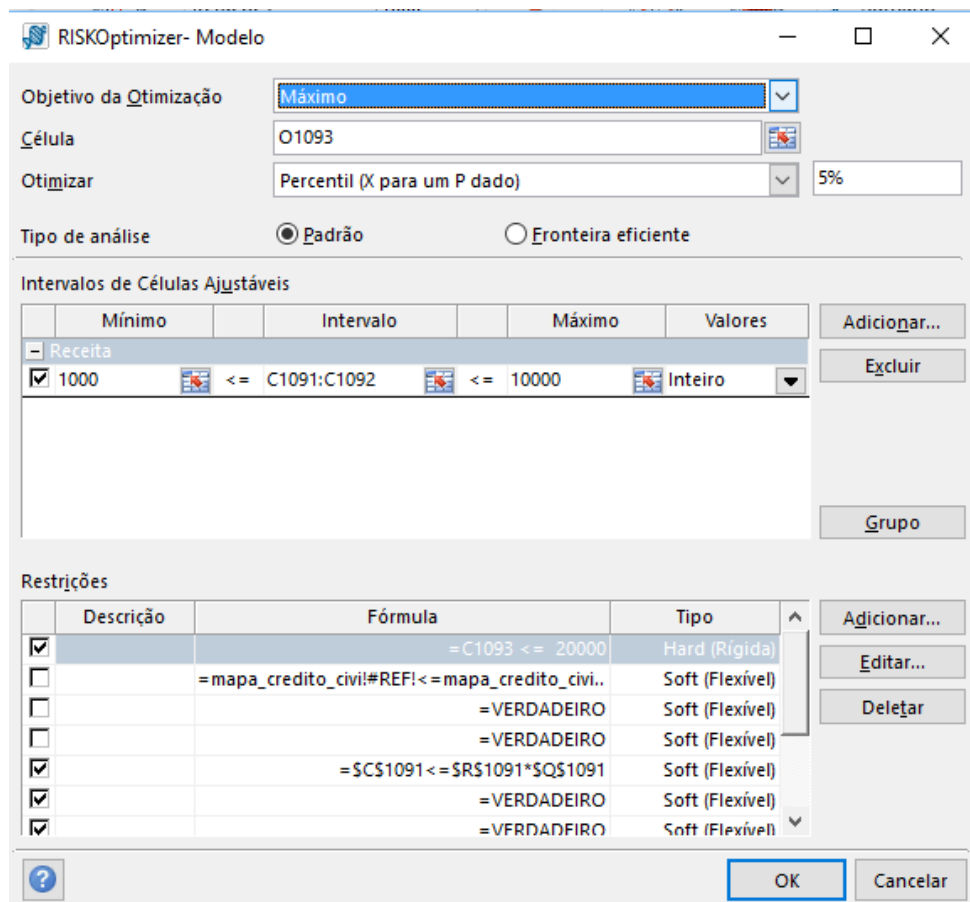


Figura 36: Modelo RiskOptimizer com restrições definidas

## Anexo 9. Classificações de crédito dos clientes 4, 5, 6, 7 e 8, pelo modelo RNA\_1

Cliente	CP	CP	AF	ROA	Variação_ROA	TL	TK_variação	RO	Montante/CP	Categoria Histórico	SelectedProbability	PredictedValue
4	1679442,1	0,25	46,17	12,42	-0,33	47065,77	-0,96	451805,00	0,00%	3	0,77	mau pagador
5	65804,5	0,07	53,59	4,41	2,80	1891,45	-0,86	5416,15	0,00%	3	0,77	mau pagador
6	-178670,6	-0,44	-742,68	-225,21	0,46	-92024,26	-0,22	-54179,98	0,00%	2	0,53	mau pagador
7	13562283,9	-0,02	59,34	-0,08	-1,03	-1530294,44	0,16	-17758,03	1,68%	3	0,80	mau pagador
8	478310,3	0,02	29,43	2,37	-0,29	63894,61	0,06	38473,20	0,78%	1	0,26	bom pagador

Figura 37: Classificações de crédito do modelo RNA\_1

