

ISCTE – Instituto Universitário de Lisboa
Departamento de Ciências e Tecnologias da Informação

**CREDIT SCORING E A PREVISÃO DE FALÊNCIAS NO
CONTEXTO DE BASILEIA II**

Fernanda Maria Esteves Romão

Tese submetida como requisito parcial para obtenção do grau de

Mestre em Sistemas Integrados de Apoio à Decisão
Especialidade em Análise de Informação

Orientador:

Professor Doutor Duarte Trigueiros, Catedrático,
Universidade do Algarve, Faculdade de Economia

Co-orientadora:

Professora Doutora Maria José Trigueiros, Professora Auxiliar,
ISCTE – Instituto Universitário de Lisboa

[Setembro, 2009]

Agradecimentos

A realização de uma dissertação de mestrado é um trabalho solitário, comparado a uma prova de maratona. O atleta da maratona tem que percorrer os longos 42 km, mais de 2 horas de corrida solitária, todavia conta com a presença do treinador, que o orientou na preparação física, psicológica e estratégica. Ao longo da corrida o atleta sabe que existem as bancas de descanso, onde poderá recolher bebidas para reforçar o seu metabolismo, vive ainda o apoio do público que, espalhado pelo percurso, o incentiva.

Para muitos atletas, o objectivo é a vitória, serem os primeiros a cruzar a linha de meta, para outros o desafio é superarem-se a si próprios, para outros ainda o simples facto de estarem a disputar a prova é já um desafio e uma vitória, o seu objectivo é apenas conseguir cruzar a linha de meta. Em suma, encontram-se inúmeras motivações para a conquista de um desafio.

Também eu percorri a minha maratona, não em quilómetros mas em páginas escritas, e também eu tive o meu treinador, na pessoa do Professor Duarte Trigueiros, cuja intervenção foi decisiva na elaboração do presente documento. Sem a sua orientação não teria chegado ao fim, não teria cruzado a meta. Quero pois expressar-lhe o meu agradecimento pela disponibilidade e ensinamentos que me transmitiu.

À minha co-orientadora, Professora Maria José Trigueiros, agradeço a orientação inicial, criando os alicerces que iriam suportar todo o trabalho de investigação e redacção do texto.

Aos meus amigos, eles sabem quem são, tenho a agradecer o facto de se terem cruzado na minha vida, e terem permanecido a meu lado, são um público maravilhoso!

Do meu Pai e minha irmã Bela, valeu a compreensão, pelas minhas frequentes ausências.

As minhas bancas de descanso foram, sem dúvida, geridas pela minha Mãe. Realmente, o seu apoio logístico foi fundamental. Quero pois expressar-lhe o meu bem-haja pelo cuidado, amor e paciência que tem tido para comigo.

Resumo

A dissertação estuda metodologias de estimação do Credit Scoring e Previsão de Falências e a sua aplicação às directivas de Basileia II no que respeita ao risco de crédito. Neste contexto desenvolveu-se e validou-se um modelo preditivo para o cálculo do PD (Probability of Default), uma das componentes do risco de crédito.

Pretende-se com este trabalho dar a conhecer o estado da arte dos modelos preditivos de Credit Scoring e Previsão de Falências; desenvolver um modelo de estimação da PD no âmbito das disposições regulatórias de Basileia II e com aplicabilidade a Instituições Bancárias, ao abrigo da IRB e da AIRB e que faça uso de dados de fácil acesso e elevada aplicabilidade.

Palavras-chave: Sistemas de Apoio à Decisão, Credit Scoring, Previsão de Falências, Risco de Crédito, Modelos Previsionais, Basileia II.

Abstract

This thesis aims to studies methods of estimation of Credit Scoring and Forecasting Bankruptcy and its application to the Basel II directives with regard to the credit risk. In this context, was developed and validated a predictive model for the calculation of the PD (Probability of Default), one of the components of credit risk.

The intend of this work, was to inform the state of the art of predictive models of Credit Scoring and Predicting Bankruptcy and develop a model for the estimation of PD under the regulatory requirements of Basel II and its applicability to banks under the IRB and AIRB methods and to make use of data easily accessible and high applicability.

Keywords: Decision Support Systems, Credit Scoring, Probability of Default, Credit Risk, Predictive Models, Basel II.

Índice

Agradecimentos	i
Resumo	ii
Abstract	iii
Índice	iv
Índice de Tabelas	vi
Índice de Figuras	vii
Lista de abreviaturas	viii
1. Introdução	1
1.1 Enquadramento do problema	1
1.2 O Problema.....	3
1.3 Motivação.....	4
1.4 Pressupostos	4
1.5 Objectivos da dissertação	4
1.6 Metodologia	5
1.7 Estrutura da Dissertação.....	7
2. Enquadramento Teórico	8
2.1 Disposições Reguladoras de Basileia II para o Risco de Crédito.....	8
2.1.1 O Capital Mínimo Regulatório.....	12
2.1.2 O risco de crédito em Basileia II.....	13
2.2 Credit Scoring	18
2.2.1 Apontamento Histórico	18
2.2.2 Os modelos de Credit Scoring.....	21
2.3 Os modelos estatísticos, de Data Mining e Basileia II.....	30
2.4 Previsão de Incumprimento.....	31
2.4.1 Modelos Preditivos com dados da empresa	32
3. Desenvolvimento de um Modelo de Previsão de Falências	36
3.1 Preparação dos dados e construção das Amostras.....	37
3.2 Modelação dos dados	41

3.3	Simulação do Modelo e cálculo da Probability of Default (PD).....	43
3.4	Análise de Resultados	49
4.	Conclusões do Estudo	51
5.	Bibliografia	53
Anexo A		56
Anexo B		57
Anexo C		60
Anexo D		66

Índice de Tabelas

Tabela 1 - Tabela identificadora das variáveis da Compustat utilizadas na amostra.....	56
--	----

Índice de Figuras

Figura 1 – Enquadramento e decomposição das componentes do risco de crédito no âmbito de Basileia II e identificação do alvo de estudo.....	2
Figura 2 – Processo Interactivo CRISP-DM.....	5
Figura 3 – Estrutura do Acordo de Basileia II.....	11
Figura 4 – Fórmula de cálculo do Capital Mínimo Regulatório.....	12
Figura 5 – Tabela de Ponderadores de Risco (fonte?).....	14
Figura 6 – Os componentes da Expected Loss e respectivas abordagens ao risco.....	15
Figura 7 – Ponderadores de Risco na Abordagem Standard.....	15
Figura 8 – Árvore de Decisão Simples.....	28
Figura 9 – Gráficos representativos da variável Data 6 – Activos Totais, Antes e Após a transformação logaritmica.....	40
Figura 10 – Composição das Amostras.....	40
Figura 11 – Sumário das características do modelo.....	42
Figura 12 – Representação da rede neuronal.....	42
Figura 13 – Importância explicativa relativa das variáveis seleccionadas para o modelo de rede neuronal.....	43
Figura 14 – Esquema de Modelo de Simulação.....	44
Figura 15 – Fragmento do Output do Modelo de Simulação.....	45
Figura 16 – Variáveis estandardizadas e níveis de confiança[11].....	46
Figura 17 – Taxas de Incumprimento acumuladas (em percentagem) 1920-2002 da Moody's[11].....	47
Figura 18 – Função Student, outra aproximação usada para estimar o VAR[11].....	48
Figura 19 – O VAR num histograma da distribuição de ganhos e perdas[11].....	49
Figura 20 – Histograma com a distribuição das Expected Loss (EL).....	50

Lista de abreviaturas

AIRB – Advanced Internal Ratings-Based

BCBS - Basel Committee on Banking Supervision

BD – Base de Dados

BdP – Banco de Portugal

BIS – Bank for International Settlements

EAD – Exposure to Default

ECAI - External Credit Assessment Institution

EL – Expected Loss

INE – Instituto Nacional de Estatística

IRB – Internal Ratings-Based

ISCTE – Instituto de Ciências do Trabalho e da Empresa

LGD –Lost Given Default

MSIAD – Mestrado em Sistemas Integrados de Apoio à Decisão

PD – Probability of Default

1. Introdução

“Risk Management” is a beautiful lie. Beautiful lies are lies that we know are not true but desperately want to believe in. Risk, it seems, is the ultimate unknown.

Satyajit Das

1.1 Enquadramento do problema

O risco faz parte da condição humana, do seu ciclo de vida, e do ciclo de vida de todas as organizações que coexistem na sociedade. A origem do risco é multidimensional, multivariada, e só existe em ambientes de incerteza, decorrente de variáveis exógenas e endógenas que não se conseguem prever com exactidão.

As instituições bancárias, dado o cariz da sua actividade, coexistem com níveis de risco elevados. Poder-se-á afirmar que o risco é o seu “core business”. Todavia, os riscos assumidos pelas instituições bancárias diferem de outros tipos de risco, dado o seu cariz sistémico. Eles repercutem-se em riscos que indirectamente são assumidos pelos próprios credores do banco, seus clientes, accionistas e gestores [1].

Para salvaguardar os interesses dos clientes, credores, assim como de toda a economia, as instituições bancárias regem-se por um conjunto rigoroso de orientações, regulamentos, legislação e supervisão, que grosso modo assegurem que a actividade bancária não excedeu um determinado nível de risco, salvaguardando a sobrevivência da instituição bancária, bem como os interesses dos seus stakeholders e da própria economia envolvente.

Neste contexto, foi criado o Comité de Supervisão de Basileia (Basel Committee on Banking Supervision), cuja função visa o fortalecimento e solidez dos sistemas financeiros aderentes.

No âmbito da sua actividade reguladora, o Comité de Supervisão de Basileia emitiu, no ano de 2004, um documento genericamente designado de Basileia II. Este, após rectificação e publicação nas instâncias da União Europeia e em outras economias aderentes, consubstancia as linhas orientadoras da actividade bancária e financeira ao nível internacional. Neste contexto, cada entidade oficial supervisora da actividade bancária e

financeira do respectivo país, bem como as entidades bancárias a exercer actividade nesse mesmo país, são obrigadas a seguir as orientações de Basileia II.

O normativo Basileia II encontra-se decomposto em 3 pilares: Pilar 1 - determinação de requisitos mínimos de fundos próprios para cobertura dos riscos de crédito, de mercado e operacional; Pilar 2 - processo de supervisão e Pilar 3 - prestação de informação sobre a situação financeira e a solvabilidade das instituições.

Enquadrado no primeiro Pilar de Basileia II encontra-se o risco de crédito e, nesse âmbito, um conjunto de regras e métodos para o seu cálculo. Dá a possibilidade de as instituições bancárias recorrerem a modelos próprios ou, em alternativa, a indicadores fornecidos por empresas da especialidade (empresas de Ratings).

Assim, em resultado das exigências normativas, nomeadamente Basileia II, da própria evolução do mercado de capitais e da economia mundial, as instituições bancárias têm vindo a privilegiar o desenvolvimento das metodologias de cálculo do risco de crédito, sobretudo ao nível dos três grandes parâmetros PD (Probability of Default), LGD (Loss given Default) e EAD (Exposure at Default).

O alvo de análise e estudo da presente dissertação é a estimação da PD (Probability of Default), excluindo outros componentes necessários para a estimação do *Value at Risk*, e a sua aplicabilidade irá restringir-se aos métodos IRB (Internal Ratings' Based Approach) e AIRB (Advanced Internal Ratings' Based Approach) de Basileia II.

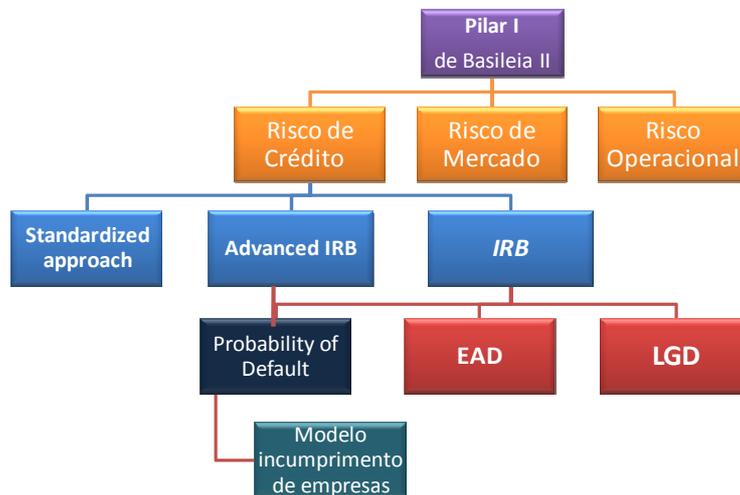


Figura 1 – Enquadramento e decomposição das componentes do risco de crédito no âmbito de Basileia II e identificação do alvo de estudo.

Da estimação da PD apenas serão estudadas as metodologias que utilizam variáveis exógenas (disponíveis a qualquer pessoa) e que se apliquem ao risco de incumprimento de empresas em relação às dívidas, de médio e longo prazo, contraídas junto de bancos.

Finalmente, de entre as variáveis exógenas, a dissertação privilegia a metodologia que usa dados contabilísticos, não aquelas que usam dados fornecidos por mercados de capitais. A escolha das metodologias que aplicam variáveis contabilísticas prende-se com o facto das mesmas terem sido as primeiras concebidas para a previsão de falências, e as mais utilizadas na estimação da PD no âmbito de Basileia II. A figura 1 procura ilustrar a incidência mais relevante da dissertação.

1.2 O Problema

O mercado livreiro transborda de bibliografia que versa sobre a temática do risco, nas suas mais variadas vertentes, nomeadamente na gestão de risco e em metodologias de estimação de risco. Encontramos fortemente desenvolvidas metodologias e ferramentas de cálculo de risco, dirigidos a vários sectores de actividade económica e financeira, em que são aplicados diferentes modelos, em diferentes cenários de actuação.

Desde o século passado existem empresas especializadas em risco, quantificação do risco aplicada nas várias vertentes do risco, nomeadamente, risco de crédito, risco operacional e de mercado.

O desenvolvimento das tecnologias de informação representou, sem margem para dúvidas, um veículo impulsionador de todo um negócio relacionado com o risco, nomeadamente em termos de desenvolvimento de software específico, na criação de modelos de apoio à decisão – e aqui destacam-se os modelos para Credit Scoring, direccionados sobretudo aos sectores de investimento, bancário e segurador.

Mas a verdade é que o risco e a incerteza coexistem na sociedade de uma forma holística, não apenas junto das entidades que operam em actividades de risco. Pese embora toda a investigação, avanço tecnológico e desenvolvimento de todo um sector de actividade consultivo relacionado com a gestão de risco, em boa verdade as crises financeiras acontecem em ciclos cada vez mais frequentes e os resultados dessas crises figuram-se cada vez mais globais em termos geográficos, mais transversais no que respeita a sectores da economia e avassaladores em termos de impacto na sociedade.

Vive-se hoje uma crise financeira da qual não há memória. Vários bancos mundiais faliram ou tiveram que ser resgatados pelos Governos dos seus países, colocando em exposição e em questão todo um processo de decisão ligado às opções estratégicas e de negócio da maioria das instituições financeiras.

Estando esta temática na ordem do dia, há certamente um bom fundamento para o estudo de metodologias de previsão do risco, consubstanciando o âmbito do presente estudo, cuja centralidade assenta em modelos de previsão de falências e de Credit Scoring aplicados ao risco de crédito.

Na perspectiva de uma instituição bancária, o risco de crédito reveste-se de maior importância face aos outros riscos, como o dos mercados ou o operacional. Pois o primeiro decorre da própria actividade bancária; representa valores monetários mais elevados (comparados com o risco de mercado ou risco operacional) e acarreta o perigo de insolvência para a instituição.

Porque a atribuição de crédito é a actividade principal da maioria das instituições bancárias, a aplicabilidade dos resultados do estudo será sem dúvida um dos factores motivadores do nosso projecto.

1.3 Motivação

A realização deste estudo será o veículo essencial à atribuição do Grau de Mestre em Sistemas Integrados de Apoio à Decisão à Mestranda Fernanda Romão. Também motivante será todo o trabalho de desenvolvimento de um modelo predictivo com aplicabilidade empresarial e o aprofundamento do conhecimento científico sobre as matérias versadas na tese, nomeadamente finanças, regulação bancária, análise de dados, Data Mining, concepção de metodologias de previsão de falências e Credit Scoring.

1.4 Pressupostos

- A alteração da actual legislação aplicada ao nível da regulação bancária, nomeadamente os normativos instrutórios de Basileia II;
- O desenvolvimento do Modelo Predictivo, recorrendo a uma Base de Dados de empresas dos EUA;

1.5 Objectivos da dissertação

A presente dissertação encontra-se dirigida para a análise das metodologias de estimação do Credit Scoring e previsão de falências consideradas mais relevantes pela mestranda e respectivo orientador, e a sua aplicação às directivas de Basileia II, no que respeita ao risco de crédito.

Neste contexto será desenvolvido e validado um modelo que incidirá particularmente, e com uma perspectiva inovadora, no cálculo do PD (Probability of Default), um dos três componentes do risco de crédito.

Pretende-se com a realização deste trabalho:

- Dar a conhecer o actual estado da arte em termos de modelos de Credit Scoring e previsão de falências;
- Desenvolver um modelo de Estimação da PD aplicado a Basileia II, que possa vir a ser usado em instituições bancárias, ao abrigo das metodologias IRB ou AIRB;
- Desenvolver um modelo assente em dados de fácil acesso e elevada aplicabilidade;

1.6 Metodologia

A metodologia de trabalho a aplicar no desenvolvimento do presente estudo terá como referência a metodologia CRISP-DM¹, aplicada em processos de Data Mining, onde a vertente do negócio a modelar percorre as diversas fases da metodologia, sendo ainda de considerar relacionamentos cíclicos entre as fases [2].

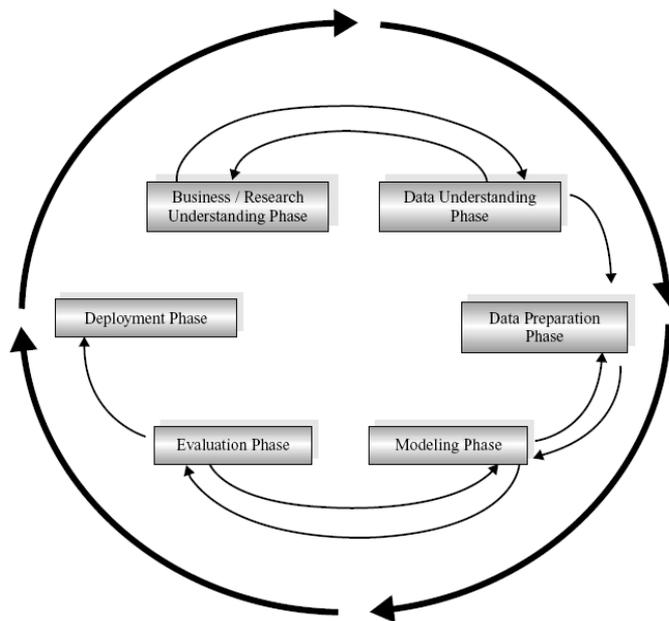


Figura 2 – Processo Interactivo CRISP-DM

¹ Cross-Industry Standard Process for Data Mining – CRISP - DM

As fases da metodologia são as seguintes:

- **Business Understanding**

A primeira fase do CRISP-DM consiste no entendimento e investigação do problema, onde são clarificados os objectivos e requisitos em termos de negócio. Nesta fase são também definidas as linhas orientadores para a concretização dos objectivos.

- **Data Understanding**

Nesta fase procede-se à recolha dos dados necessários para modelação, sendo para isso necessário a compreensão da informação, de modo a verificar se a mesma é adequada ao problema formulado. A importância da compreensão dos dados reside no facto de ser o ponto de partida efectiva para o desenvolvimento do projecto. Atendendo à qualidade dos dados que existem para modelar, é nesta fase que se percebe se existem dados para a formulação do problema em termos de Data Mining.

- **Data Preparation**

Trata-se de uma fase trabalhosa que consiste na preparação dos dados para modelação e compreende a selecção de variáveis, e a limpeza, integração e formatação dos dados.

- **Modeling**

Reunidos os *DataSet*, inicia-se o processo de escolha das técnicas de modelação, a esquematização dos vários modelos e a sua avaliação em termos de performance, ou seja, a orientação dos modelos obtidos face ao problema enunciado.

- **Evaluation**

A fase de validação e avaliação dos modelos está estreitamente ligada à primeira fase de formulação do problema, pois a análise e avaliação dos resultados é sempre confrontada com os objectivos delineados e a capacidade do modelo para a geração de soluções. Eventualmente será necessária uma consequente revisão do processo, caso o modelo não atinja os resultados pretendidos.

- **Deployment**

A última fase agrega a elaboração dos planos de implementação, monitorização e manutenção da solução, assim como a definição dos relatórios e verificação final do projecto.

1.7 Estrutura da Dissertação

Seguindo uma disposição tradicional, a presente dissertação encontra-se estruturada em quadro partes ou capítulos:

O primeiro capítulo é reservado à introdução às temáticas abordadas, formulação do problema, esquematização da metodologia e motivações.

No capítulo segundo está reflectido o estado da arte no que respeita a Metodologias/ Modelos de Credit Scoring e Previsão de Falências. Será ainda explanada a questão de Basileia II e o enquadramento do Credit Scoring e da previsão de falências nos instrumentos preconizados no referido normativo.

O terceiro capítulo abrange a comparação de metodologias e interpretação de resultados do desenvolvimento do modelo de estimação da PD (Probability of Default). O Modelo assenta em variáveis contabilísticas e outras variáveis exógenas à empresa, e a sua interpretação será realizada à luz das directivas de Basileia II, designadamente no âmbito do I pilar do instrumento regulador.

Importa referir que o desenvolvimento do modelo seguirá a metodologia CRISP-DM, com a particularidade dos dados terem origem na importante base de dados financeiros COMPUSTAT da Thomson (US).

O último capítulo apresenta uma súmula do trabalho realizado nos anteriores capítulos e extrai as mais importantes conclusões.

2. Enquadramento Teórico

*We have the duty of formulating, of summarizing,
and of communicating our conclusions, in
intelligible form, in recognition of the right of
other free minds to utilize them in making their
own decisions.*

R.A.Fisher

2.1 Disposições Reguladoras de Basileia II para o Risco de Crédito

O Banco de Compensações Internacionais é a mais antiga Organização Financeira Internacional do Mundo, tendo sido criado no ano de 1930 no contexto do chamado Plano Young e, na altura, vem assumir as funções anteriormente exercidas pelo Agente Geral de Reparções, em Berlim.

Inicialmente, as funções do BCI abraçavam principalmente a supervisão do pagamento das indemnizações impostas à Alemanha pelo Tratado de Versalhes, após a Primeira Guerra Mundial. A própria designação do Banco deriva do seu papel original, que abarcava competências de administração do fundo Dawes e Young (empréstimos internacionais emitidos para financiar as indemnizações) e também para promover a cooperação em geral entre bancos centrais.

Actualmente, o Banco de Compensações Internacionais (em inglês: *Bank for International Settlements*) é a organização internacional responsável pela supervisão bancária, visando "promover a cooperação entre os bancos centrais e outras agências na busca de estabilidade monetária e financeira". Sediado em Basileia, na Suíça, reúne 55 bancos centrais de todo o mundo [3].

O BIS é responsável por publicar estatísticas e relatórios sobre os bancos centrais e sistema financeiro global. Essas publicações são apresentadas em “workshops” e seminários em que também são consagrados outros serviços diferenciados às instituições bancárias. Na sua sede, estão abrigadas varias secretárias de comités como o Comité dos Mercados; o Comité do Sistema Financeiro Global e o Comité de Supervisão Bancária de Basileia, fundados pelo G-10, em 1962, 1971 e 1974, respectivamente.

O Comité de Supervisão Bancária de Basileia (BCBS, sigla de *Basel Committee on Banking Supervision* em inglês) é uma organização que congrega autoridades de supervisão bancária, visando o fortalecimento e a solidez dos sistemas financeiros. Foi estabelecido em 1975 pelos presidentes dos bancos centrais dos países do Grupo dos Dez (G-10) e, normalmente, reúne-se no Banco de Compensações Internacionais, em Basileia, Suíça, local onde se situa a sua Secretaria permanente.

Desde o seu surgimento, o BCBS é um fórum de discussão para o melhoramento das práticas de supervisão bancária, buscando aperfeiçoar as ferramentas de fiscalização internacionalmente. Apesar de não ter autoridade para fazer cumprir as suas recomendações, a maioria dos países, membros ou não, tendem a implementar as políticas ditadas pelo Comité.

O Comité de Basileia é constituído por representantes de autoridades de supervisão bancária dos bancos centrais de Bélgica, Canadá, França, Alemanha, Itália, Japão, Países Baixos, Suécia, Suíça, Reino Unido e Estados Unidos, além de Luxemburgo e Espanha, que não são membros do G-10.

Ao abrigo das suas competências, o Comité de Basileia firmou, no ano de 1988, o acordo designado de Acordo de Capital de Basileia, oficialmente denominado em inglês de *International Convergence of Capital Measurement and Capital Standards*, também designado por Acordo de Basileia ou Basileia I, tendo sido o mesmo ratificado por mais de 100 países.

Este acordo teve como objectivo criar exigências mínimas de capital, que deveriam ser respeitadas por bancos comerciais, como precaução contra o risco de crédito, prevendo também mecanismos de estimação de risco de crédito e o estabelecimento de exigências de um padrão mínimo de capital.

Segundo Goodhart, Hofman e Segoviano (2004: 596), citados por Silva [4] a liberalização dos mercados financeiros conduziu a um aumento da competição e à redução das margens de lucro. A inexperiência na gestão do risco e a entrada em novos mercados e clientes, assim como a concessão de empréstimos de forma inapropriada, aumentaram o risco da carteira de crédito dos bancos e os “write-offs” (crédito abatido ao activo). Tal situação conduziu a uma redução do rácio de capital e à fragilidade do sistema financeiro.

Assim a génese de Basileia I ² encontra-se nas diversas crises que assolaram os mercados financeiros internacionais nos anos 70 e 80, as quais culminaram na falência de várias instituições bancárias e no enfraquecimento do sector financeiro de alguns países, nomeadamente do México, Brasil, Rússia e outros.

² Importa referir que, até á implementação de Basileia I, o requerimento de capital era baseado na fixação de índices máximos de alavancagem. Os Bancos somente poderiam emprestar 12 vezes o seu capital e reservas, ponderados pelo risco de crédito.

Conforme descrito por Silva [4], em Janeiro de 1996, foi publicado o “*Amendment to the Capital Accord to Incorporate Market Risks*”, primeira adenda ao acordo original, alargando os requerimentos de capital também ao risco de mercado³. O que abriu caminho para uma revisão mais ampla do acordo de 1988. Basileia I, apesar de ter sido adoptado em mais de 100 países, não conseguiu alcançar o objectivo principal, o de conseguir tornar o sistema financeiro estável.

Acontecimentos futuros, tais como a crise asiática de 1997-1998; a liberalização dos mercados financeiros (que culminou numa maior variação do preço dos activos); o crescimento da competição devido aos desafios da globalização; a consolidação industrial (que motivou as empresas a mudanças contínuas, quer nas suas relações externas quer nas operações internas); a volatilidade dos mercados (implicando a redução dos proveitos e aumento dos riscos); em suma uma evolução da economia global, provaram que Basileia I não respondia aos objectivos para os quais tinha sido delineada. Havia que avançar para uma solução, um novo acordo.

A primeira versão do texto do novo Acordo é publicada 1999, e pela primeira vez é feita a alusão ao risco operacional⁴, que, com a implementação definitiva de Basileia II, passaria a ser parte integrante no cálculo dos requisitos de capital.

Com a publicação, em 14 de Junho de 2006, da Directiva Bancária Codificada (2006/48/CE) [5] e da Directiva de Adequação de Fundos Próprios (2006/49/CE) [6] foi acolhida, na ordem jurídica comunitária, a revisão do Acordo de Capital (concluída em Novembro de 2005), hoje conhecida por «Basileia II».

O novo regime prudencial foi adoptado em Portugal através da publicação do Decreto-Lei n.º 103/2007 [7] e do Decreto-Lei n.º 104/2007 [8], ambos de 3 de Abril, e da emissão de um conjunto de Avisos e de Instruções do Banco de Portugal, visando a regulamentação das disposições constantes daqueles Decretos-Lei .

De forma genérica, o regime prudencial proposto por Basileia II encontra-se apoiado por uma estrutura que comporta os chamados três pilares:

- Pilar 1 - determinação de requisitos mínimos de fundos próprios para cobertura dos riscos de crédito⁵, de mercado e operacional;

³ Representa a possibilidade de perda, num determinado período de tempo, devido a movimentos aleatórios nas taxas de juro, nas taxas de câmbio, nos preços dos activos financeiros e nas *commodities*, e a mudanças na correlação (interacção) entre eles e nas suas volatilidades (Banco Itaú, 2003) citado por Silva (2006).

⁴ O Basel Committee on Banking define risco operacional como "the risk of direct or indirect losses resulting from inadequate or failed internal processes, people and systems or from external events". Um exemplo base de risco operacional é a fraude [10].

⁵ O risco de crédito encontra-se estritamente associado à PD (Probabilidade de Incumprimento) dos clientes em relação ao reembolso dos créditos, originando a perda total ou parcial destes. Este risco refere-se ainda à possibilidade de redução da capacidade de cumprimento dos créditos e de reputação dos clientes. Os

- Pilar 2 - processo de supervisão;
- Pilar 3 - prestação de informação sobre a situação financeira e a solvabilidade das instituições;

Face a Basileia I, o actual Acordo apresenta como alterações mais significativas, o incentivo à aproximação dos requisitos de fundos próprios ao risco, passando a reconhecer-se, para efeitos regulamentares e desde que cumpridas determinadas condições, os sistemas de gestão e as medidas de risco das instituições, e a autonomização do risco operacional.

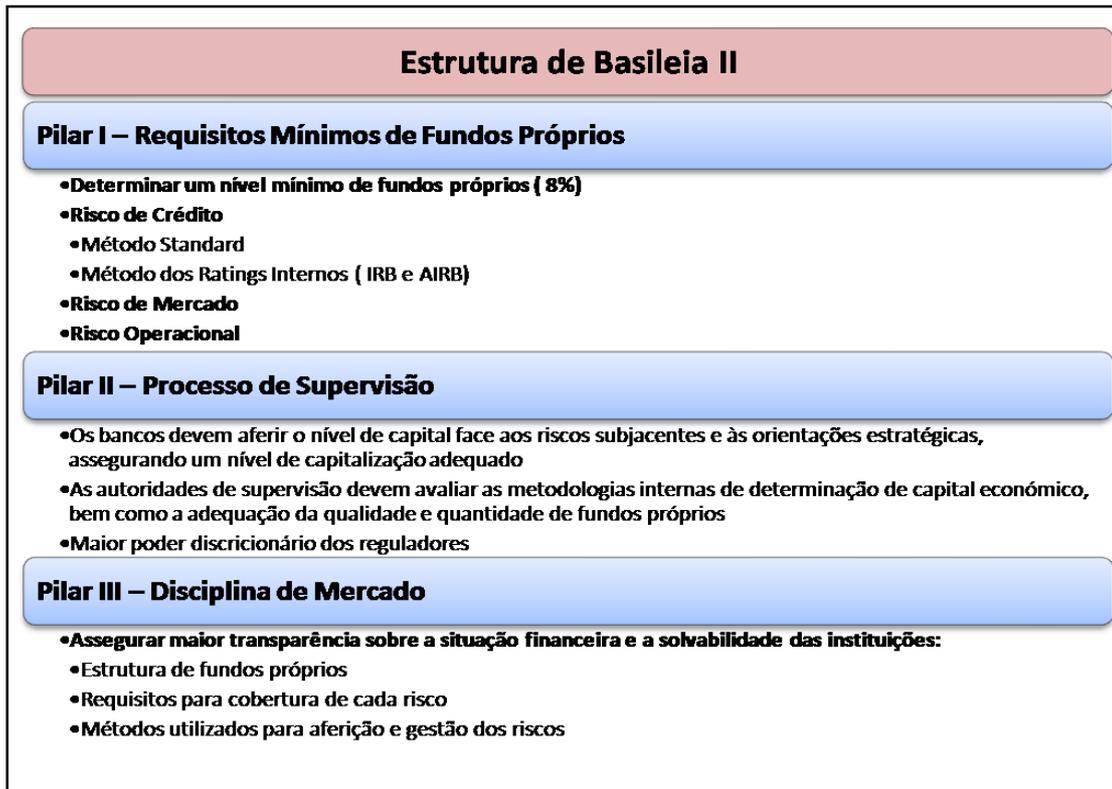


Figura 3 – Estrutura do Acordo de Basileia II

mercados financeiros penalizam a redução da capacidade de reembolso de um mutuário, quer pela via do aumento das taxas de juro, quer através da redução do valor das cotações das acções, ou ainda pela alteração da notação de Rating. Trata-se do principal risco enfrentado pelos bancos, sendo decorrente de transacções que originam direitos, contingentes ou potenciais contra uma contraparte (Banco Itaú, 2003). O risco de crédito pode ser dividido em:

- De incumprimento, quando o devedor não cumpre com uma determinada cláusula contratual;
- País ou soberano, que se verifica quando um país não honra uma dívida por questões económicas ou políticas, ou um conjunto de entidades de um país são incapazes de pagar os empréstimos nas datas contratadas. Pode ser considerado, em primeiro lugar, como soberano quando o devedor é uma entidade estatal ou entidades por ele garantidas, podendo ser ineficazes as acções legais para o cumprimento atempado da obrigação, alegando razões de soberania. Em segundo lugar, como de transferência, quando os devedores residentes apresentam uma incapacidade generalizada de liquidar as suas obrigações, nomeadamente face à incapacidade de aceder a divisas para o efeito;
- De liquidação (ou de clearing), quando a troca efectiva de fundos ou outro activo é inviabilizada [4].

2.1.1 O Capital Mínimo Regulatório

Basileia II, à semelhança do preconizado por Basileia I, aponta para uma percentagem superior a 8% de capital capaz de cobrir a exposição aos riscos. Isto, como mínimo requerido para os Bancos, de modo a poderem funcionar regularmente e com riscos controlados. Todavia, a grande evolução face a Basileia I está no facto de o risco operacional ser uma variável autónoma face ao risco de crédito e de mercado.

A fórmula apontada no novo Acordo para o cálculo do Capital Mínimo é:

$$\text{Capital Mínimo} = \frac{\text{Total Capita (TIER I + TIER II)}}{\text{Risco de Crédito+ Risco Mercado+Risco Operacional}} \geq 8\%$$

Figura 4 – Fórmula de cálculo do Capital Mínimo Regulatório

Por forma a completar a caracterização do Capital Mínimo Regulatório imposto por Basileia II, importa descrever as três formas de capital reconhecidas em Basileia II [9], conforme descritos no texto de Basileia II [10], e adaptado de Jorion [11], e Silva [4], ou dito por outras palavras os três níveis de qualidade de capital. Assim temos:

- **TIER I ou core capital** – Corresponde ao capital que não está alocado/ associado a outros compromissos, inclui capital social e resultados retidos. Capital realizado, prémios de emissão, reservas, lucros, fundo para riscos bancários gerais e diferenças negativas de primeira consolidação e reavaliação (equivalência patrimonial) e insuficiências de provisões para encargos com reformas. Note-se que 50% do Capital Mínimo Regulatório deverá ter origem na TIER I.
- **TIER II ou capital suplementar** – Considera-se o capital que se cativou de um compromisso de longo prazo, nomeadamente passivos subordinados com vencimento indeterminado, reservas de reavaliação, títulos de participação e acções preferenciais remíveis. Ainda títulos de participação, empréstimos subordinados (maturidade superior a 5 anos e não exceda 50% da Tier I), acções preferenciais, provisões para riscos gerais de crédito, com o limite de 1,25% do total dos activos ponderados pelo risco.
- **TIER III** – São provisões, dívidas e outras origens de curto prazo, nomeadamente lucros líquidos da carteira de negociação e empréstimos subordinados de curto

prazo. Sendo considerado capital de pior qualidade, a sua aplicabilidade restringe-se a cobrir os riscos de mercado.

2.1.2 O risco de crédito em Basileia II

Funções dos Ponderadores de Risco⁶ na IRB

No negócio bancário, as perdas financeiras ocorrem com alguma frequência. Existem sempre mutuários que não cumprem com as suas obrigações, sendo que as perdas sofridas por uma instituição bancária variam de ano para ano e dependem da conjuntura económica e outros factores. Em suma, não existe estabilidade e consistência quanto à distribuição das perdas de uma instituição bancária ao longo dos anos.

Embora não seja possível saber de antemão as perdas devido a incumprimento no crédito num determinado ano, pode-se, no entanto, estimar esse valor, designado por Expected Loss, que indica a perda média de uma determinada carteira. As instituições financeiras vêem a EL como uma componente do custo do negócio e fazem a sua gestão através de uma taxação diferenciada da exposição do crédito aos riscos (maior risco maior taxa) e também recorrendo a provisões.

Uma das responsabilidades dos bancos consiste em assegurar fundos para proteger o próprio banco de catástrofes financeiras pontuais, mas que podem exceder os níveis expectáveis. Muito embora perdas com estas características não ocorram anualmente, quando acontecem podem originar danos consideráveis ao banco e até a todo o sistema financeiro.

Assim, perdas caracterizadas por serem superiores aos níveis esperados pelos bancos, e que acontecem em espaços temporais também eles imprevisíveis, são designadas de Perdas Inesperadas (Unexpected Losses – UL). Taxas de juro, prémios de risco, encargos com as exposições de crédito podem absorver algumas das perdas caracterizadas como sendo Unexpected Losses. Todavia, o mercado não irá com certeza cobrir na totalidade as perdas inesperadas. É necessário capital disponível para cobrir os riscos deste tipo de perdas, logo o Capital disponível tem uma função importante como absorção de perdas inesperadas.

O pior cenário que se poderia imaginar, seria um banco perder todo o seu portfólio de crédito num dado ano. Este cenário, embora improvável,⁷ acontece. Situações realmente

⁶ Risk Weight Funcions

⁷ Mas já verificados no ano de 2008 com a falência de vários bancos na Europa e na América. Em Portugal temos os casos do BPN (salvo pelo Estado, através da injeção de capital que levou à sua nacionalização) e do Banco Privado, ainda em processo de negociações, que chegaram à falência técnica.

más têm a capacidade de afectar todos os bancos, pois levam a uma perda de confiança generalizada. Uma reserva de capital capaz de fazer face a uma situação destas seria economicamente inviável pois inibiria a actividade principal do banco, emprestar dinheiro.

A Abordagem IRB adoptada em Basileia II [12] focaliza-se na frequência da insolvência do banco derivada das perdas devido ao incumprimento de crédito que as entidades supervisoras estão dispostas a aceitar. A alocação de capital existe para assegurar que as perdas inesperadas excedem os níveis de capital apenas com uma probabilidade mínima.

A Expected Loss de um portfólio assume-se por ser igual à proporção dos créditos que podem incorrer em incumprimento num determinado espaço temporal (1 ano no contexto de Basileia II), multiplicado pelas exposições ao incumprimento por saldar, multiplicado pela percentagem de exposição que não será recuperada.

É evidente que os bancos não dispõem dos dados exactos para realizarem o cálculo acima mencionado. Mas o cálculo pode ser feito quando os factores multiplicativos identificados no parágrafo anterior correspondem aos parâmetros de risco preconizados em Basileia II pela Abordagem IRB.

A PD (Probability of Default), probabilidade de incumprimento, é expressa em percentagem e considera-se o maior dos riscos associados ao crédito. Caracteriza-se por representar a possibilidade de um mutuário não cumprir com os pagamentos de um empréstimo. Em Basileia II é calculada por classe de Rating no espaço temporal de 1 ano.

Probability of Default	Corporate	Residential Mortgage	Other Retail
0.03%	14.44%	4.15%	4.45%
0.10%	29.65%	10.69%	11.16%
0.25%	49.47%	21.30%	21.15%
0.50%	69.61%	35.08%	32.36%
0.75%	82.78%	46.46%	40.10%
1.00%	92.32%	56.40%	45.77%
2.00%	114.86%	87.94%	57.99%
3.00%	128.44%	111.99%	62.79%
4.00%	139.58%	131.63%	65.01%
5.00%	149.86%	148.22%	66.42%
10.00%	193.09%	204.41%	75.54%
20.00%	238.23%	253.12%	100.28%

Note: Illustrative weights for LGD = 45%, maturity of 2.5 years, and large corporate exposures (firms with turnover greater than 50 million euros).

Figura 5 – Tabela de Ponderadores de Risco [11]

A EAD (Exposure at Default) representa uma estimativa do montante por saldar no caso do incumprimento de um mutuário, ou seja, é o valor exposto ao incumprimento de um determinado crédito.

A LGD (Loss Given Default) representa a perda em caso de incumprimento em termos de percentagem da EAD, i.e, depois de tratados os efeitos colaterais em dívida: é o que é irrecuperável em percentagem.

A Expected Loss (EL) descreve-se em termos económicos segundo a fórmula identificada na figura abaixo. Também são apresentadas na figura as diferentes Abordagens para o cálculo do risco de crédito.

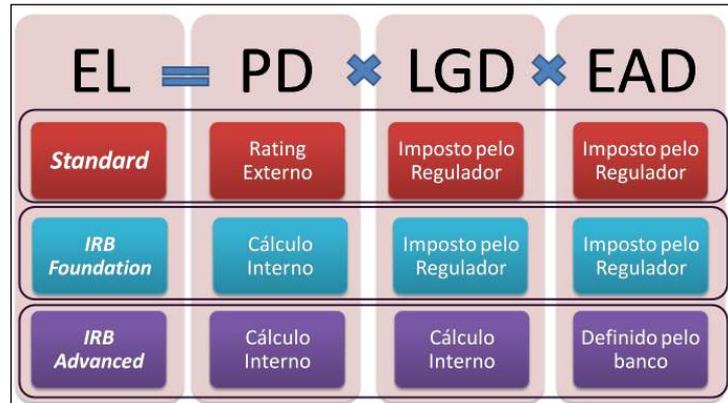


Figura 6 – Os componentes da Expected Loss e respectivas abordagens ao risco

Conforme já mencionado anteriormente, Basileia II define para efeitos de adequação de capital três tipos de riscos (mercado, crédito e operacional), que caracteriza em termos de componentes, aplicabilidade, limites e métodos de cálculo no Pilar I do Novo Acordo. Face ao âmbito da presente tese, apenas iremos debruçar-nos com mais pormenor sobre o risco de crédito.

O Comité de Basileia permite aos bancos escolher entre duas metodologias, no que toca ao cálculo dos requisitos de capital para o risco de crédito, a Abordagem Standard e a Rating Interno (IRB) [13].

Claim	Credit Rating					
	AAA/ AA-	A+/ A-	BBB+/ BBB-	BB+/ B-	Below BB-	Unrated
Sovereign	0%	20%	50%	100%	150%	100%
Banks—option 1	20%	50%	100%	100%	150%	100%
Banks—option 2	20%	50%	50%	100%	150%	50%
Short-term	20%	20%	20%	50%	150%	20%
Claim	AAA/ AA-	A+/ A-	BBB+/ BB-	Below BB- Unrated		
Corporates	20%	50%	100%	150%	100%	

Note: Under option 1, the bank rating is based on the sovereign country in which it is incorporated. Under option 2, the bank rating is based on an external credit assessment. Short-term claims are defined as having an original maturity less than three months.

Figura 7 – Ponderadores de Risco na Abordagem Standard [14]

A vantagem da aplicação da Abordagem Standard relaciona-se com o facto de os ponderadores de risco serem definidos por agências de Rating⁸ reconhecidas a nível internacional. A adopção da referida metodologia apresenta-se vantajosa no caso de bancos de pequena dimensão, que não disponham de dados históricos, nem de departamentos com capacidade analítica que promovam a criação das suas próprias metodologias.

A aplicação do método de Rating Interno na definição de ponderadores de risco de crédito assume-se como uma inovação face a Basileia I. Esta aproximação deve sempre ser aplicada, claro está, sob vigilância de entidades de supervisão (BIS) e permite aos bancos desenvolver as suas próprias notações de Rating, conduzindo a que classifiquem os seus clientes com base nos seus próprios ponderadores.

A vantagem notória da aplicação da Abordagem IRB relaciona-se com a aplicação dos fundos próprios, ou seja, se determinado cliente não for considerado de risco elevado, a necessidade de capital a alocar a esse compromisso é menor, logo existe a chamada “poupança de capital”.

A adopção da IRB está condicionada em Basileia II por alguns critérios [3], nomeadamente:

- a) o reconhecimento pelo supervisor dos métodos e processos internos de atribuição de Rating e de quantificação do risco de crédito (BIS, 2004: 62,86,92,94-95) [10],
- b) os bancos têm de demonstrar que estão a usar sistemas de crédito durante, pelo menos, os três anos anteriores à adopção do IRB;
- c) o desenho do sistema de Rating deverá estar devidamente documentado (incluindo a indicação da periodicidade das revisões de Rating atribuído);
- d) o período de dados usados para estimar a PD deve ser de, pelo menos, cinco anos (BIS, 2004:30; relacionando a PD e a notação de risco BBB por país);
- e) Existência de registos históricos das notações atribuídas a dívidas soberanas, outros bancos e empresas;
- f) Avaliação independente da atribuição de Rating, realizada no mínimo anualmente, mesmo que efectuada internamente pelo banco;

⁸ Empresas de Rating devem ser aprovadas pela entidade reguladora nacional, podemos mencionar a Standard & Poor's e a Moody's como referência.

- g) Implementação de procedimentos de simulação da adequação de capital face a situações limite (os chamados Stress Tests⁹), como cenários de crise económica, sectorial entre outras (BIS, 2004: 89) [10].

O Acordo permite, mediante autorização do supervisor, que uma instituição utilize uma Abordagem para um conjunto de carteiras de crédito e uma Abordagem diferente para outras, devendo, no entanto, evoluir para a utilização generalizada da Abordagem mais avançada. É esperado que um banco acabe por adoptar o método IRB na totalidade da sua carteira de créditos, e não só em parte dos seus investimentos, sendo que a sua adopção poderá ser efectuada de forma faseada (BIS, 2004: 57) [10].

Entende-se que esta medida permite assegurar que as necessidades mínimas de capital se apliquem a todo o grupo bancário, independentemente da sua localização geográfica. Esta medida impede motivações para a escolha da melhor arbitragem, em cenários geográficos menos rigorosos, obrigando também a uma evolução no processo de transição do método Standard para o IRB, nas diversas instituições de um mesmo grupo.

A adopção do método IRB tem um carácter de continuidade, pelo que o retorno ao Standard apenas é autorizado em situações como o desinvestimento num conjunto significativo de negócios inerentes ao crédito, mediante autorização do supervisor (BIS, 2004: 57) [10].

Em relação à classificação do risco de crédito, é de referir que as exposições a crédito de retalho¹⁰ (que não fazem parte desta tese) devem ser orientadas para o risco do tomador e da transacção, devendo os bancos demonstrar que o processo fornece uma diferenciação significativa do risco, mediante um grupo de exposições homogéneas (BIS, 2004: 83-84) [15].

No que concerne ao segmento de dívidas de empresas, crédito soberano e crédito concedido a outros bancos, deverá existir uma classificação separada para os mutuários e garantes (e.g. avalistas), e cada exposição terá inerente uma linha de crédito.

Conforme referido anteriormente, na análise ao risco de crédito promovida em Basileia II, a vantagem na aplicação dos métodos IRB e AIRB, em detrimento da Abordagem Standard, resume-se pois ao facto do desenvolvimento de métodos internos por parte das instituições bancárias poder originar poupanças de capital aquando da afectação do mesmo aos créditos, i.e., conhecendo os clientes, tendo histórico, informação, com mais equidade se avalia os riscos associados a um determinado crédito, o nível de Rating a atribuir, logo a alocação de capital (em termos de requisitos mínimos) é menor.

⁹ Definir Stress testing – relaciona-se com a validação dos modelos previsionais, que deverão ser submetidos a cenários “catastróficos”, ou mesmo de limite, de modo a aferir o seu comportamento

¹⁰ Caem fora do âmbito da presente tese.

2.2 Credit Scoring

2.2.1 Apontamento Histórico

A concessão de crédito e respectivas taxas de juro associadas estão documentadas como existentes desde as antigas civilizações, nomeadamente na Babilónia 2000 anos AC, conforme refere Lewis [16]. Na Roma antiga, o recurso ao crédito era visto como uma actividade de risco para a integridade pessoal do devedor e da respectiva família. Contribuíam para este facto as altas taxas de juro e as penalidades para os incumpridores: o incumprimento era equiparado a roubo ou fraude e punido da mesma forma.

O não pagamento de uma dívida podia levar à prisão, à morte ou à escravatura (a possibilidade do credor vender o incumpridor e a sua família para liquidação da dívida, Anderson [17]).

Dado que as penalidades para quem não cumpria com o crédito eram bastante rigorosas, tornou-se compreensível que a actividade de crédito não fosse vista com bons olhos por parte de pensadores religiosos como Buda, Jesus, Mohammed, Aristoteles e São Tomás de Aquino.

A atribuição de crédito para efeitos de comércio foi implementada a partir de 1100 DC, com a multiplicação de feiras de comércio por toda a Europa. Também por essa altura foram instituídas as primeiras casas de penhoras, que, por mais estranho que pareça, não cobravam qualquer interesse na actividade.

Por volta de 1300, e dado o desenvolvimento da actividade das casas de penhoras que por esta altura já cobravam juros bastantes elevados, a Igreja, alegando a imoralidade do acto de cobrança de juros, banuiu da Europa as casas de penhoras.

Mas, com a evolução do comércio a ideia do juro não morreu; em 1516 a cobrança de juros estava difundida e plenamente aceite pela sociedade regente da altura.

A legislação relativa à falência foi instituída primeiramente nas cidades Estado da Itália; na Inglaterra de Henrique VIII apareceram as primeiras leis da falência. O código comercial de Napoleão de 1807 influenciou as leis relativas às falências, não só em França mas também na Espanha e em Portugal.

O sistema de hipotecas conhecido desde a antiga Roma, também se foi alastrando por toda a Europa no pós império Romano. Os grandes impulsionadores desta prática são os judeus, que já no séc. XII aceitavam hipotecas da nobreza que necessitava de financiamento para as suas Cruzadas.

Quanto aos descobertos bancários e ao cheque, Anderson [17] refere que antes de 1800 só os muito ricos tinham acesso a estes instrumentos. Mas o advento da Revolução Industrial e a necessidade de financiamento levou a que na Inglaterra se flexibilizasse a atribuição de crédito e se liberalizasse a taxa de juro. Por esta altura, os cheques estavam solidamente implementados por toda a Europa e muitos comerciantes e industriais preferiam-nos ao dinheiro vivo.

Até meados do séc. XX, a concessão de crédito assentava na experiência do banqueiro e na capacidade e vontade aparente do devedor para pagar a sua dívida, ou seja, a atribuição de um crédito estava associado a um “rating” assente na opinião que o credor tinha do devedor e na experiência que esse mesmo credor tinha do mercado [18].

A primeira publicação mundial sobre informação financeira de um conjunto de empresas americanas surgiu em 1860, quando Henry Varnum Poor lançou a sua “*History of Railroads and canals in the United States*”. Com o sucesso desta publicação aparece a *Poor’s Publications*, cuja finalidade era informar os investidores sobre a solidez da dívida emitida por empresas.

Em 1900 John Moody and Co, publica “*Moody’s Manual of Industrial and Miscellaneous securities*”, cujo objectivo era fornecer aos investidores informação financeira e estatística das empresas que emitiram dívida.

Em 1914 foi criada a empresa de serviços financeiros Moody’s, que semanalmente desde então tem publicado “*Moody’s credit perspectives*”, alegando que foi o primeiro a atribuir “ratings” através de um sistema de letras.

A empresa Standard Statistics também se dedicou à análise e estudo de empresas do sector público e privado, e respectivo comportamento financeiro. Em 1941, a Poor’s Publishing e a Standard Statistics fundiram-se, dando lugar à empresa conhecida por Standard and Poor’s.

Na actualidade, os sistemas de Rating¹¹ são utilizados pelas instituições financeiras e preconizados em Basileia II mais no sentido de valorizarem o risco em entidades soberanas e grandes empresas, pois ao nível do crédito pessoal e PME’s esses sistemas não se manifestam eficientes.

Com o desenvolvimento industrial e comercial, resultante da produção em série e na institucionalização de lojas de retalho, a atribuição de crédito estendeu-se à classe média. Surgiu um novo segmento de mercado, o dos pequenos empréstimos, e foram criados os primeiros bancos comerciais.

O início deste processo dá-se em larga escala em 1920, com a possibilidade de nos Estados Unidos a população poder recorrer ao crédito para a aquisição de automóvel.

¹¹ Indicador da solvabilidade de um devedor. Segundo a Moody’s o Rating de crédito é uma opinião da qualidade de um devedor em termos de cumprimento e solvabilidade.

O aumento do número de pedidos de crédito obrigou os bancos a estandardizarem produtos e automatizarem o processo de atribuição de crédito. A primeira aproximação ao Credit Scoring deu-se com a aplicação de sistemas periciais não automáticos. Observado um conjunto de atributos de bons e maus devedores, os mesmos foram mensurados e classificados numericamente face a uma escala previamente definida.

Após a 2ª Guerra Mundial deu-se uma explosão na economia e foram então desenvolvidas políticas económicas que incentivavam o consumo por parte das famílias, mas ao mesmo tempo escasseavam os peritos em atribuição de crédito.

Neste contexto, esses mesmos peritos foram convidados a escrever as regras que seguiam para a atribuição de um crédito. O resultado foi um sistema híbrido de Credit Scoring, caracterizado por um algoritmo que continha as condições resultantes da experiência dos peritos em análise de crédito, e a opinião do analista de crédito.

Em 1936 o R. A. Fisher desenvolveu uma técnica estatística designada de Análise Discriminante que, face às suas especificidades, pode considerar-se o background matemático do Credit Scoring.

Esta técnica identifica um conjunto de atributos mensuráveis de uma população, para o seu posterior agrupamento face a características semelhantes. Fisher desenvolveu a sua técnica aplicada à investigação em genética, mas cedo outros investigadores aplicaram a metodologia em outras áreas de investigação. Assim, o modelo foi aproveitado em 1941 por D. Durand no reconhecimento de bons e maus empréstimos [19].

A aliança entre a automação e o modelo discriminante de Fisher ou outros semelhantes, fez percorrer um caminho natural para os modelos de Credit Scoring.

Muitas das regras subjectivas na análise de crédito foram eliminadas, o que inicialmente provocou bastante resistência por parte dos analistas de crédito, pois consideravam as técnicas de Credit Scoring uma afronta ao conhecimento convencional dos bancários.

A perseverança, a eficiência dos processos, e o crescimento económico levaram ao refinamento das técnicas, quer em termos de automação como de eficácia dos modelos. Fundamental para o desenvolvimento do Credit Scoring foi sem margem para dúvidas a invenção das máquinas de calcular e dos computadores.

A introdução do cartão de crédito em 1960 impulsionou o já extenso uso de técnicas de Credit Scoring, não apenas por instituições bancárias, mas também por retalhistas. Todavia em muitos países foi decretada ilegal a utilização de variáveis como o sexo, a religião ou a raça como atributos nos sistemas de Scoring.

De particular interesse foi o trabalho desenvolvido por Altman em 1968 [20], com a aplicação do Credit Scoring na previsão do risco de falências de empresas. O método de Altman designado de Z-score será revisto no ponto 2.4.1 do presente documento.

2.2.2 Os modelos de Credit Scoring

O Credit Scoring é um método utilizado para estimar a probabilidade de um crédito deixar de ser amortizado ou pago. Esta definição está estreitamente ligada à actividade bancária, pois conforme descrito nos parágrafos anteriores foi neste sector de actividade que se implementaram os primeiros modelos de Credit Scoring.

Na actualidade os modelos de Credit Scoring encontram-se amplamente difundidos para estimar e minorar o risco de crédito. Empresas retalhistas que se apoiam no crédito para vender os seus produtos, instituições financeiras, empresas de Marketing e obviamente os bancos recorrem a métodos de Scoring para classificar os seus clientes.

Face à comprovada fiabilidade dos métodos de Credit Scoring, os bancos centrais de vários países, assim como o BCBS através de Basileia II, incentivam as instituições bancárias e de crédito a adoptarem estas metodologias.

Em termos de teorias e métodos, e dependendo dos autores, os modelos de Credit Scoring dividem-se em vários grupos. Se a Abordagem estiver relacionada a distribuição dos dados então temos os modelos paramétricos e os modelos ou heurísticas não paramétricas [21]. Todavia se a Abordagem estiver relacionada com os métodos de scoring utilizados, então existem os métodos tradicionais e os métodos de Data Mining.

Outro aspecto importante de referir respeita à aplicabilidade do modelo. Nestes termos as instituições bancárias desenvolveram diferentes grupos de modelos de Scoring:

Modelos Comportamentais - baseiam-se nas informações recolhidas face ao comportamento dos clientes individualmente aquando da quitação de dívidas e outros pagamentos numa base mensal, este tipo de Scoring é dinâmico e é usado pelos bancos para a segmentação de clientes em termos de risco associado, assim como para campanhas de Cross Selling.

Modelos aplicativos de risco - assentam na premissa que o cálculo do Scoring é realizado apenas uma única vez e o seu resultado é estático, i.e. os dados sobre os pagamentos efectuados pelo cliente são esporádicos e envolvem apenas a informação que é recolhida no formulário de requisição de crédito [22].

Um apontamento importante a fazer, antecedendo a caracterização dos modelos de Credit Scoring, tem a ver com as variáveis utilizadas nos modelos, uma vez que a selecção das variáveis e a distribuição das mesmas tem um impacto fulcral na escolha do modelo a aplicar. A discussão apresentada neste estudo aquando da elaboração do modelo é ilustrativa do problema que a selecção e manipulação das variáveis representa.

Caracterização dos Modelos de Credit Scoring

Optou-se para efeitos de caracterização dos modelos por subdividi-los em dois grupos: Paramétricos e Não Paramétricos. Nas páginas seguintes será realizada uma explanação dos modelos mais representativos de cada grupo, bem como a sua aplicabilidade no âmbito do Credit Scoring.

Métodos Paramétricos¹² ou tradicionais

- *Modelo Probabilístico Linear*

A regressão linear simples quando aplicada à dicotomia da variável dependente é chamada de modelo probabilístico linear. O valor a prever é interpretado como evento probabilístico e a probabilidade assume-se dependente das variáveis independentes, que poderão adoptar o papel de características individuais significativas para o modelo.

$$y_i = x_i\beta + \varepsilon_i \quad (1)$$

y_i é a variável aleatória a prever e que pode assumir o valor 1 ou 0

x_i é o vector das variáveis independentes

ε_i distribuição dos erros

β coeficiente do vector

Pressupostos assumidos pelo Modelo de Probabilidade Linear:

- ✓ Linearidade
- ✓ Homocedasticidade
- ✓ Distribuição normal dos erros amostrais
- ✓ Independência dos erros
- ✓ Aditividade
- ✓ Inclusão no modelo das variáveis significativas e exclusão das irrelevantes para o modelo

¹² Assentam na premissa que a amostra apresenta uma distribuição de frequências dos erros amostrais normal, as variâncias são homogéneas (nomeadamente homeoscedásticas), os efeitos dos factores de variação são aditivos e os erros independentes, sendo que os parâmetros que definem as populações são a média e o desvio padrão.

Os modelos lineares apresentam várias limitações, nomeadamente a existência de uma relação muito sensível relativamente aos valores assumidos pelas variáveis independentes. Por vezes, a variável dependente assume valores fora do intervalo $[0,1]$, em modelos com linearidade fraca, os resultados por vezes são erróneos e produzem uma variância inflacionada.

- Modelo Probabilístico Não Linear - Regressão Logística e Probit

Probit

Nem sempre é possível modelar os dados por forma a aplicar a regressão probabilística linear, uma vez que as variáveis não cumprem os requisitos aplicáveis à regressão linear e referidos no ponto anterior. Então há que transformar os dados e, se após a transformação dos dados, estivermos perante uma distribuição probabilística acumulada aproximada à Normal standardizada, então aplica-se um modelo Probit [23].

$$P_i = F(\alpha + X_i) \quad (2)$$

em que $F(\alpha + X_i)$ é a função de distribuição de probabilidade normal Standardizada acumulada onde s é uma variável aleatória normalmente distribuída com média zero e variância unitária.

$$P(Y_i) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{y_i} e^{-s^2/2} \quad (3)$$

Desta forma, se os dados puderem ser transformados usando $p^{-1}(Y_i)$ então tornar-se-ão numa função linear e poder-se-á aplicar OLS. Assim estamos perante um modelo de probabilidade Probit, ou, como algumas vezes é designado, Normit.

$$Y_i = p^{-1}(Y_i) = \alpha + \beta X_i \quad (4)$$

Apesar de o modelo Probit ser uma alternativa atractiva ao modelo probabilístico linear, apresenta dificuldades de estimação originadas pelo facto de $p^{-1}(Y_i)$ não ser fácil de obter, assim, uma forma funcional alternativa é usualmente especificada - a logística.

Logit ou Regressão Logística

A regressão logística é, entre as técnicas de classificação, uma das mais extensivamente utilizadas. A logística binária, binomial ou multinomial aplica-se quando a variável dependente assume valores binários, do tipo Sim/Não, ou quando poderá assumir valores ordinais com mais classes.

Na regressão logística não se modela directamente a variável dependente, a variável a prever é transformada logaritmicamente¹³ e utilizado o método do vizinho mais próximo (MLE¹⁴) para estimar os coeficientes da variável dependente.

Se considerarmos um modelo idêntico ao Probit no facto de $P_i = F(\alpha + X_i)$, mas em que $F(\alpha + X_i)$ é igual à função de distribuição de probabilidade acumulada logística, ou seja,

$$P(Y_i) = \frac{1}{1+e^{-Y_i}} \quad (5)$$

temos o modelo de probabilidade Logit. A função de distribuição da logística aproxima-se bastante da Normal acumulada. Tal como esta, também é simétrica com valor médio zero, mas a variância é igual a $\pi^2/3$. No entanto, a sua inversa é fácil de calcular:

$$Y_i = \ln\left(\frac{P_i}{1-P_i}\right) \quad (6)$$

Donde:

$$\ln\left(\frac{P_i}{1-P_i}\right) = F(\alpha + \beta X_i) \quad (7)$$

Muito embora a regressão logística seja utilizada para a modelação de variáveis nominais, o seu output é uma função contínua em forma de S, que representa a probabilidade associada ao facto de pertencer a um determinado valor assumido pela variável dependente, podendo aumentar ou diminuir a amplitude da curva.

A regressão logística não assume a linearidade de relações entre as variáveis independentes e a variável dependente também não obriga à distribuição normal das variáveis, não assume a homocedasticidade, em suma apresenta menos restrições que a regressão linear.

Pressupostos

- Os erros estocásticos são independentes
- Não existe multicolieridade entre as variáveis independentes
- A variável dependente é quantitativa
- Inclusão no modelo das variáveis relevantes e a exclusão das pouco significativas.

Na análise predictiva, a regressão logística é a mais popular das técnicas, pois consegue trabalhar as variáveis de uma forma linear, muito embora esteja presente em variáveis não lineares. É um modelo robusto, menos rígido no que toca aos pressupostos da sua utilização, de fácil compreensão e aplicação.

¹³ $\ln(p/(1-p))$

¹⁴ MLE – Maximum likelihood estimation

- Análise Discriminante

O modelo de análise discriminante de Fisher tem como objectivo, partindo da existência de grupos mutuamente exclusivos de indivíduos, "descobrir as características que distinguem os membros de um grupo dos de outro, de modo que, conhecidas as características de um novo indivíduo, se possa prever a que grupo pertence..." , Reis, 1992 citada por Camões [23].

Esta "descoberta" é levada a cabo através da estimação de combinações dessas características (variáveis) que discriminem entre os grupos previamente definidos, de tal modo que seja minimizada a probabilidade de erro de incorrecta classificação a posteriori.

A análise discriminante utiliza-se na classificação em dois ou mais grupos exclusivos, recorrendo para o efeito à informação provida por um conjunto de variáveis preditivas (semelhantes às variáveis independentes na regressão), quando não existe nenhuma ordem natural entre os grupos, são mutuamente exclusivos.

A função discriminante é uma variável que resulta da combinação linear de um conjunto de variáveis independentes:

$$L = b_1x_1 + b_2x_2 + \dots + b_nx_n + c \quad (8)$$

- b_1 coeficiente discriminante
- x_1 variável independente
- c é a constante

Os coeficientes discriminantes b_1 são calculados de forma a maximizar a distância entre as médias das variáveis dependentes. A função discriminante pode ser calculada recorrendo ao método dos mínimos quadrados ou da MLE, resultando um tipo de função discriminante logística.

Em suma, a análise discriminante pretende encontrar uma regra que separe ao máximo os agrupamentos de dados (Clusters) porventura existentes e classifique novas ocorrências dentro desses Clusters.

Pressupostos da Análise Discriminante

- Linearidade e aditividade (e ausência de heteroscedasticidade)
- Erros estão distribuídos aleatoriamente
- Não existe multicolieridade

Um ponto fraco da análise discriminante é a sua extrema dependência relativamente ao pressuposto de que os grupos têm equivalente número de membros e variâncias semelhantes. Se dentro da população existe um grupo substancialmente maior que os outros grupos, a análise discriminante pode classificar as observações unicamente nesse grupo. Outra restrição à utilização da análise discriminante tem a ver com o facto da mesma não aceitar variáveis qualitativas independentes.

Métodos Não Paramétricos¹⁵ ou de Data Mining

- *Redes Neurais*

As redes neuronais são algoritmos cuja estrutura de funcionamento é pretensamente inspirada no funcionamento biológico do cérebro, portanto onde existem “neurónios” ou “neurões” e “sinapses” que os ligam entre si.

Uma rede neuronal consiste numa rede interconectada entre as variáveis de input e de output, compreendendo um conjunto não linear de nódulos chamados de neurões que, através de camadas subsequentes, estão interligados através da referida rede. O neurónio é o elemento basilar da rede neuronal, recebe os inputs e transformando-os, recorrendo a um factor multiplicado chamado “peso”, e emite um output que será transmitido para um outro neurónio ou representará o output final. A transformação que acontece no interior do neurão poderá representar nomeadamente uma função logística ou gaussiana, entre outras.

O desafio em modelar uma rede neuronal está na escolha da função transformadora, do tamanho e arquitectura da rede e na calibragem dos pesos, sendo finalidade destes ajustamentos o incremento da exactidão de classificação do modelo.

Normalmente os modelos de redes neuronais são alimentados por quatro sub-amostras de dados:

- Dados de treino
- Dados de validação e teste do modelo
- Dados de produção

Uma rede neuronal aprende com os dados de treino, ou seja, o desenho estrutural da rede em termos de nódulos, camadas e pesos é definido através da aprendizagem que a rede efectua com os dados de treino, i.e. os dados de treino são os dados da modelação.

Um dos problemas das redes neuronais está no “overfitting”. Devido à grande capacidade de aprendizagem que uma rede neuronal pode ter, a mesma encontra facilmente um

¹⁵ Não existe associada à amostra qualquer pretensão sobre a sua distribuição estatística, daí que a sua utilização é ideal para modelar dados com relações não lineares complexas. A fraqueza destes modelos encontra-se na sua fraca robustez face aos modelos paramétricos.

modelo que responda detalhadamente à classificação dos dados de treino. Todavia, quando aplicado o mesmo modelo aos dados de validação, dados desconhecidos para a rede, o mesmo apresenta-se bastante fraco em termos de previsão.

Os dados de teste são dados que não foram utilizados para efeitos de construção do modelo, e através dos dados de teste é atestada a capacidade predictiva do modelo. Por fim, temos os dados de produção, i.e. dados para os quais a variável de output é desconhecida, e será sobre estes dados que o modelo irá trabalhar, apoiando a tomada de decisão.

As redes neuronais permitem relações de não linearidade e interações complexas entre as variáveis preditivas, e assim apresentam-se com vantagem face aos modelos paramétricos na modelação complexa.

A fraqueza das redes neuronais está presente no facto de o modelo não poder ser descrito. São autênticas caixas negras que apresentam os resultados, não associando ao mesmo qualquer poder explicativo.

As redes neuronais são também particularmente adequadas quando não existe necessidade de justificar o modelo, ou seja, por parte dos decisores que não têm expectativas ou comportamentos teóricos quanto à importância e significância das variáveis e do modelo obtido.

- *Árvores de Decisão*

As árvores de decisão, também designadas de árvores de classificação, são mais uma categoria dos métodos não paramétricos. Os modelos em árvores consistem num conjunto sequencial de regras do tipo “Se...Então”, que, devido à sua estrutura embricada (hierárquica) originam uma ou mais subdivisões ou ramos.

Com a aplicação destes métodos a amostra é dividida em grupos, de acordo com as co-variáveis (atributos das variáveis). No caso de árvores de classificação binárias, cada nó da árvore é assignado a regras de decisão que, além de caracterizarem a amostra, subdividem a mesma em dois subgrupos. Uma nova observação é processada de acordo com a regra associada a esse ramo da árvore, outras subdivisões são efectuadas até que se atinja o resultado final, que consiste na classificação da observação.

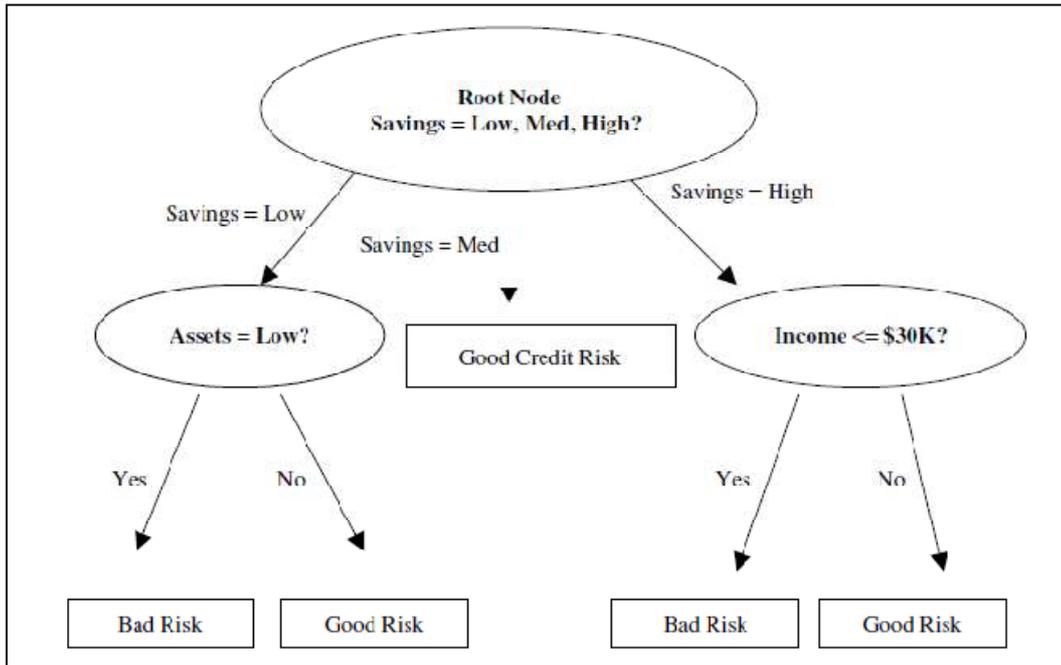


Figura 8 – Árvore de Decisão Simples¹⁶

Característica importante destes modelos é o facto de as variáveis serem tratadas como qualitativas, sendo que a importância de uma variável ou atributo depende da categoria de variável que a antecede a um nível superior em termos de hierarquia [24].

Os algoritmos mais importantes de construção de árvores de decisão são C&RT (Classification and Regression Trees), popularizado pelos trabalhos de Breiman [25] e o CUAID (Chi-square Automatic Interaction Detector) ver Kass [26]. Ambos os algoritmos usam diferentes critérios para identificar e otimizar a separação dos dados em subsets e abater os atributos que não são relevantes para o modelo.

Na generalidade os pontos fracos e fortes das árvores de decisão são:

- Através da classificação, as relações de não linearidade entre as variáveis são facilmente modeladas;
- Tal como as redes neuronais, as árvores de decisão são livres de fazer as associações que considerarem relevantes, i.e., são autónomas na construção do modelo;
- O output é de fácil compreensão;
- As probabilidades associadas aos outputs têm que ser calculadas num passo separado;

¹⁶ Imagem retirada do livro *Discovering Knowledge in Data: Na Introduction to Data Mining* – Daniel T. Larose, 2005 John Wiley & Sons Inc Publication

- O output consiste num reduzido número de categorias e não numa classificação contínua de variáveis. Consequentemente as árvores de decisão só calculam probabilidades no último ramo da árvore e não para cada caso.
- Comparada com outros modelos, as árvores comportam poucas variáveis e categorias, dada a especificidade do modelo, pois o mesmo vai desagregando a amostra em ramos sucessivamente particionados, logo diminuindo a amostra e aumentando a complexidade da árvore.
- A estabilidade e robustez do modelo, por ser não paramétrico, não pode ser validada com procedimentos estatísticos. A estratégia é trabalhar-se com uma amostra de treino para a construção do modelo e outra para a validação do modelo.

Em conclusão, as árvores de decisão estão particularmente adaptadas quando a amostra se caracteriza por um número limitado de variáveis interactivas.

- *Multivariate Adaptive Regression Splines (MARS)*

O Multivariate Adaptive Regression Splines - MARS foi dado a conhecer no ano de 1991, por Jerome H. Friedman, professor da Universidade de Stanford. Sendo um modelo completamente automático, o procedimento de modelagem inspirou-se nas técnicas de particionamento recursivo que governam os modelos de árvores de decisão, ou seja, na máxima “dividir para conquistar”. O resultado é um modelo que identifica a transformação de variáveis e interacções óptimas entre variáveis [27].

A palavra “*multivariate*” indica que o modelo MARS tem capacidade de tratar informação multi-dimensional, examinar características individuais das variáveis e as interacções entre as mesmas. “*Adaptive*” significa simplesmente selectivo, o MARS automaticamente rejeita um certo número de variáveis, caso não contribuam o suficiente para o desempenho final do modelo.

O modelo MARS aplica uma função de regressão, “*Regression*”, que na sua essência é equivalente ao termo normalmente utilizado na estatística, onde existe uma variável dependente, uma constante e as variáveis independentes.

Por último, a palavra “*Splines*” advém de uma ampla classe de funções que são utilizadas para a interpolação da mesma forma que a recta é usada na regressão.¹⁷ Estas funções suavizam (smooth) o modelo, evitando que se torne demasiado detalhado. Para desenvolver um “*Spline*”, o espaço ocupado pelos dados é dividido num número convencional de regiões. As fronteiras entre as regiões são designadas de “*knots*”, que correspondem, numa perspectiva espacial de exposição dos dados, ao ponto de inflexão das curvas.

¹⁷ Sig. Construção de uma fórmula empírica

As estruturas de dados complexas podem esconder uma elevada dimensionalidade, que, ao ser modelada eficazmente, coloca a descoberto importantes padrões e relacionamentos entre os dados. O modelo MARS dispõe das características necessárias para a identificação de informação complexa.

O modelo MARS, na sua essência, não é muito diferente da aplicação de um modelo estatístico não-linear a regiões sucessivas. Dispõe de um algoritmo *forward stepwise* cuja finalidade é seleccionar as características (variáveis, transformações, etc) do modelo e identificar os “knots” apropriados, seguido de um procedimento *backward* para suprimir (prune) do modelo as características indesejáveis.

Segundo Di [27], os modelos MARS não são naturalmente classificadores, muito embora a sua inspiração esteja no particionamento de intervalos dos dados. À semelhança das árvores de decisão, os modelos MARS assentam na construção de múltiplas funções de regressão, assemelhando-se assim mais a esses modelos.

2.3 Os modelos estatísticos, de Data Mining e Basileia II

A questão que se coloca nesta fase do estudo tem a ver com a adaptabilidade dos modelos mencionados no capítulo anterior às Abordagens IRB e AIRB de Basileia II. Segundo a estrutura de Basileia II [28], um sistema de Rating “deve compreender todos os métodos, processos, sistemas de controlo e sistemas de recolha de dados e infra-estruturas tecnológicas, que suportam a avaliação do risco de crédito, a consignação de Ratings Internos e a quantificação de estimadores de incumprimento e perdas”¹⁸.

Por outro lado, os requisitos mínimos para os sistemas de Rating Internos, dispostos na II parte, secção III, H do novo Acordo, respeitam à avaliação de Ratings Internos e seus requisitos. Grosso modo, especificam as características a que as Bases de Dados e respectivos modelos devem obedecer [22], nomeadamente:

- No mínimo 7 classes de Rating para mutuários cumpridores (non-defaulted);
- Nenhuma indevida ou excessiva concentração numa específica classe de Rating;
- Uma diferenciação significativa do risco entre as classes;
- A informação de input deverá ser plausível, intuitiva e actual;
- Toda a informação considerada relevante deverá ser tida em consideração.

¹⁸ Tradução própria de “ *comprises all the methods, processes, controls, and data collection and IT systems that support the assessment of credit risk, the assignment of internal Ratings, and the quantification of default and loss estimates*”

Como se poderá constatar, os requisitos para a IRB não revelam nenhuma preferência por um determinado método. É exactamente uma ideia fulcral da Abordagem IRB, o facto das instituições bancárias serem livres de escolher o método que melhor lhes convier. Deste modo, todos os modelos identificados no capítulo anterior são candidatos a serem usados na Abordagem IRB.

Consequentemente, as características dos modelos têm que ser vistas na perspectiva das especificações de Basileia II, sejam elas apontadas como pontos fracos ou pontos fortes. Por exemplo, os modelos Logit prestam-se mais aos testes de “stress” (Stress Testing, requerido nos itens 434, 345 do acordo), atendendo a que contêm uma característica de série temporal. Os métodos paramétricos permitem testar características individuais das variáveis de input, reforçando assim a plausibilidade dessas variáveis (requerido no item 410 do acordo).

Por outro lado, nenhuma das especificações de Basileia excluiu um determinado modelo, por exemplo, um banco pode ter preferência por um modelo de regressão linear, que não cumpre os requisitos, mas então terá que procurar uma via alternativa para cumprir os requisitos de aplicabilidade do modelo (item 410 do Acordo).

2.4 Previsão de Incumprimento

Compreender os motivos pelos quais algumas empresas vão à falência e outras empresas mantêm-se solventes é uma questão importante aquando da avaliação de uma empresa, seja pelo motivo de investimento nessa empresa, ou na perspectiva do Banco, que está a analisar o risco associado à atribuição de crédito a essa mesma empresa.

Quer os bancos, quer outras entidades financeiras, procuram maximizar os seus lucros, o que requer precisão na avaliação do seu portfolio de créditos e o risco associado ao mesmo. Conforme já mencionado, Basileia II impõe regras precisas sobre a avaliação do risco, e também preconiza a aplicação de Ratings de classificação de carteiras de clientes e o estabelecimento de reservas de capital canalizadas para cobrir possíveis perdas financeiras por parte das entidades bancárias. Previnem-se assim situações que coloquem em causa a estabilidade financeira da instituição e do próprio sector.

Nestes termos, ter uma visão clara do motivo que conduz à solvência de certas empresas, nomeadamente ao nível dos motivos condutores a essa situação, é deveras vital para as instituições bancárias, aquando da decisão da atribuição de um crédito.

Serão os factores macroeconómicos suficientes para a rotura de uma empresa? Certamente que não. Caso fossem, perante condições macroeconómicas muito adversas, assistir-se-ia à falência de economias inteiras. E quanto a crises sectoriais? Essas sim, podem acabar por

inviabilizar um negócio. Mesmo assim, há empresas que resistem mais tempo do que outras e algumas que nem são derrubadas.

Posto o referido acima, conclui-se que a informação financeira da empresa, nomeadamente os seus rácios contabilísticos, é determinante na análise da sua robustez financeira. Sendo certo que essa robustez varia consoante o ciclo económico e a pujança do sector, ou por outro lado, será uma combinação de ambos os factores macro e micro-económicos, que possibilitará um modelo suficientemente fidedigno, na estimação da insolvência de uma empresa.

2.4.1 Modelos Preditivos com dados da empresa

Desde as investigações de Beaver [29] e Altman [20] que a previsão de falências com recurso a dados da empresa tem sido estudada activamente por académicos e profissionais. Este campo da gestão do risco continua bastante activo devido às exigências normativas, como Basileia II, bem como pela própria evolução do sector financeiro.

Em termos de literatura, muito se tem escrito nestas ultimas décadas sobre previsão de falências. Desenvolveram-se e testaram-se novos modelos, aperfeiçoaram-se outros já existentes, resultando por vezes em conclusões contraditórias.

Existem dois tipos de modelos de previsão de incumprimento que são geralmente referidos na literatura [30],

- Modelos baseados em rácios contabilísticos, que aplicam modelos de Credit Scoring, entre os quais a Análise Discriminante e a Regressão Logísticas. Podemos enunciar algumas limitações destes modelos [18]:
 - A informação contabilística pode não ser fidedigna; pode ser manipulada;
 - A própria informação é limitada intrinsecamente, está desactualizada;
 - A informação contabilística é incompleta quando comparada com toda a informação que existe na empresa;
 - Não contempla factores externos macroeconómicos nem a influência da indústria;
 -
- Modelos baseados no mercado, nomeadamente os modelos de Merton ou de Black-Scholes (BSM) [31], caracterizados por se basearem no valor da empresa no mercado., i.e. estes modelos obrigam a que as empresas estudadas estejam cotadas em bolsa, o que nem sempre acontece.

Modelos baseados em rácios contabilísticos

No trabalho realizado por Beaver [29], o autor conclui pela consistência e utilidade da informação contabilística, quando tratada em termos de rácios contabilísticos na previsão de falência das empresas. No mesmo estudo o autor afirma ainda que a informação contabilística também seria útil na resolução de outras questões do foro da gestão, designadamente na eficácia da capitalização dos leasings.

O modelo preditivo desenvolvido por Beaver caracterizou-se por ser uma análise discriminante univariada, ou seja, estudou a capacidade preditiva dos rácios financeiros, porém um rácio de cada vez.

Embora o autor não colocasse de parte a possibilidade de desenvolver um modelo discriminante multivariado, tendo realizado alguns esforços nesse sentido, não avançou por essa via. Na altura constatou que a capacidade preditiva de um só rácio era equivalente à capacidade preditiva de um modelo de multi-rácios.

No ano de 1968 Eduard Altman publica um artigo no “Journal of Finance” intitulado Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, que se tornou uma referência na análise de informação e previsão de falências. Com este artigo o autor pretendia demonstrar a importância da capacidade analítica e preditiva dos rácios contabilísticos, sendo exemplo ilustrativo a capacidade preditiva dos rácios contabilísticos no que respeita à previsão de falências.

A técnica estatística utilizada por Altman [20] foi a Análise Discriminante Multivariada, em detrimento das técnicas de regressão. O autor justifica a escolha referindo a sua aplicabilidade em áreas como a biologia e as ciências comportamentais, mas sobretudo devido à sua aplicação na avaliação do crédito ao consumo e também em processos de classificação de empresas em categorias de investimentos.

Os dois grupos de amostras¹⁹ utilizadas por Altman eram compostas por 33 empresas cada, sendo o primeiro grupo composto por empresas industriais que faliram entre os anos de 1946 e 1965, e o grupo dois composto por empresas seleccionada aleatoriamente e que à data do estudo ainda se mantinham em funcionamento.

Os rácios utilizados por Altman foram:

$$X_1 = \text{Fundo de Maneio} / \text{Activo Total}$$

$$X_2 = \text{Resultados Retidos} / \text{Activo Total}$$

¹⁹ Altman decide retirar da amostra empresas de grande e pequena dimensão, porque o grupo 1 de teste não continha empresas com essas características. Assumiu, na altura, que não era comum empresas de larga dimensão falirem, e recorreu a uma característica dos rácios que consiste na eliminação da dimensionalidade.

$$X_3 = EBIT^{20} / \text{Activo Total}$$

$$X_4 = \text{Valor em Bolsa do capital Próprio} / \text{Valor Contabilístico do Passivo}$$

$$X_5 = \text{Vendas} / \text{Activo Total}$$

A função discriminante resultante do estudo de Altman, também designada de Z-Score (Z abaixo) apresenta na equação que se enuncia:

$$Z = 0,012X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,999X_5 \quad (9)$$

As conclusões de Altman exemplificam a precisão da análise dos rácios contabilísticos para a previsão de incumprimento, a aplicação do modelo para a avaliação de crédito, controlo interno das organizações e Leasing.

Outro autor que marcou a investigação da previsão de incumprimento foi Ohlson (1980), apontado como o primeiro autor a desenvolver um modelo probabilístico de falências, recorrendo à Regressão Logística. Utilizando uma amostra de 105 empresas que faliram e 2.058 que não faliram, considera-se que também foi pioneiro no recurso a uma amostra representativa da população. Em termos de poder preditivo, o autor assume que os resultados obtidos ficam aquém de modelos desenvolvidos anteriormente.

Em 1982, Richard Taffler [32] conduziu um estudo de previsão de falências de empresas do Reino Unido, aplicando rácios contabilísticos e a Análise Discriminante, já sobejamente conhecido e estudado por outros autores. Todavia, a inovação da sua investigação residiu na combinação dos seguintes factores: em termos de dados, a amostra de empresas que não faliram incluía empresas com características semelhantes às empresas que faliram, o que poderia moderar os erros do tipo II (classificar erradamente uma empresa não falida numa empresa falida) e a análise aos custos inerentes à previsibilidade do modelo.

Do seu trabalho, Taffler quis realçar principalmente a eficácia e aplicabilidade da ferramenta analítica que se encontra ao dispor dos decisores, revestindo a mesma de um carácter comercial.

Modelos baseados no mercado

É sobejamente reconhecido o potencial informativo que os Mercados de Valores transmitem relativamente à probabilidade de incumprimento das empresas cotadas em Bolsa. Neste entendimento está subjacente a ideia de que os mercados bolsistas dispõem de

²⁰ Earnings Before Interest and Taxes – Resultados antes de Juros e Impostos

todas as informações necessárias e impulsionadoras da actividade, flutuações dos títulos cotados [33].

O conceito de valorização de uma empresa face ao seu desempenho em bolsa está na génese do modelo de Black-Scholes, sendo posteriormente adaptado por Merton, no que respeita à valorização de activos do mercado financeiro.

O valor das acções de uma empresa num determinado momento T possui um valor superior ao da sua dívida. Então, através da venda de acções, a empresa tem capacidade de pagar as suas dívidas, retendo a diferença entre essa dívida e o valor das acções. Por sua vez, se o valor dessas mesmas acções num determinado momento T for igual ou inferior ao da sua dívida, então a venda das acções não será suficiente para satisfazer os compromissos com os credores. Ora sendo assim, o valor da empresa para os seus donos é nulo ou mesmo negativo e a melhor alternativa que os accionistas têm é que a empresa entre em incumprimento.

Pode, portanto, estimar-se a probabilidade de incumprimento observando a “distância” de uma cotação ao valor da respectiva dívida e a dispersão dessa cotação. A maior distância, menor probabilidade de incumprimento. A maior dispersão, maior probabilidade.

Basicamente os modelos baseados no mercado regem-se por uma filosofia bastante aproximada à descrita no parágrafo anterior. Existem, no entanto, uma série de inconvenientes neste modelos [34], nomeadamente o facto de muitas empresas cotadas em Bolsa com valores de cotação abaixo do seu valor contabilístico não falirem. O modelo apenas se aplica a empresas cotadas e muito embora os modelos baseados no mercado não sejam bons preditores de incumprimento, os mesmos são muito utilizados para o posicionamento em termos de Ranking de risco das empresas²¹.

Sobre as matérias versadas na presente secção, existe alguma investigação publicada em Portugal – Camões [23] e Bonfim [34], e outros. Apraz referir que na sua investigação esses autores utilizaram bases de dados relativas a empresas nacionais. Ao nível internacional existe um manancial de aplicações e derivações dos modelos mencionadas acima, desde a comparação entre modelos contabilísticos e os modelos de mercado (ver Carttkou, Chigrinov, Mchedlishvili [35]), à aplicação dos modelos em vários cenários geográficos, (Bernhardsen, [36]), (Pongsatat, Ramage Lawrence, [33]), ao estudo dos erros de estimação dos modelos (Begley, Ming, Watts, [37]).

²¹ Modelo Moody's KMV www.moodyskmv.com

3. Desenvolvimento de um Modelo de Previsão de Falências

“In God we trust, all others must bring data.”

W. Edwards Deming

Esta secção ilustra um caso de criação de um modelo capaz de classificar, à luz da informação contabilística das empresas, e num curto prazo, a probabilidade de essa empresa continuar solvente ou tornar-se insolvente, associando a cada um desses dois estados uma probabilidade.

Pretendeu-se que o modelo preditivo suportasse o menor número de variáveis possível, por forma a torná-lo mais explicativo e robusto, sem minorar a sua performance no que respeita à fiabilidade em termos de classificação das ocorrências.

Para o desenvolvimento do caso em estudo, foi aplicada a metodologia CRISP-DM, caracterizada na secção 1.6 do presente documento. A metodologia CRISP-DM recomenda, com a fase primeira – Business Understanding –, a compreensão do negócio. Significa o entendimento da importância do risco de crédito no âmbito da actividade bancária, o papel fundamental que impõe a probabilidade de falência como parte integrante das variáveis do risco de crédito, os normativos legais que regulam a actividade bancária, nomeadamente, Basileia II. Instrumento obrigatório no que toca à regulamentação bancária.

A 2ª fase da metodologia consiste na Data Understanding – a compreensão e a qualidade dos dados são garantia da fiabilidade do modelo. Conforme foi referido na secção 2.4, os modelos preditivos que assentam em dados contabilísticos foram pioneiros na previsão de falências, sendo objecto da tese recuperar essa importância, em face dos novos métodos preditivos, de Data Mining, e ainda às disposições de Basileia II, que permitem a sua aplicação em termos comerciais.

Assim, na secção seguinte é explanada a construção da amostra, também é evidente a compreensão dos dados, pois será a partir dessa compreensão que serão construídos os data sets (Amostras).

A 3ª fase da metodologia consiste na Data Preparation, desenvolvida na secção 3.1.. Quanto à fase de Modelação (Modeling Phase) é apresentada na secção 3.2. A fase Evaluation encontra-se desenvolvida nas secções 3.3 e Conclusões. A última fase de Deployment não será concretizada, uma vez que esta fase carece de uma aplicabilidade aproximada do comercial.

3.1 Preparação dos dados e construção das Amostras

Importa lembrar que, no presente estudo, serão utilizados dados contabilísticos de empresas comerciais representativas de uma maioria dos sectores económicos e segmentos de mercados, quando analisados na perspectiva das instituições de crédito.

Muito embora existam modelos que aliam dados de diferentes origens²² como fontes para a construção das variáveis, no presente estudo só serão tidas em conta variáveis cuja origem decorra de dados contabilísticos de empresas, publicamente disponíveis.

As fontes de dados utilizadas no presente estudo são a base de dados Compustat da Thompson (de longe a mais importante fonte de dados financeiros nos Estados Unidos e outras economias), e também um ficheiro cedido pelo Professor Edward Altman, contendo uma lista de empresas falidas dos Estados Unidos, desde 1976, com dimensão relevante (todas as cotadas e também outras não cotadas, mas de dimensão relevante) e respectivos anos de falência. Tais empresas são representativas dos principais sectores de actividade dos EUA.

A base de dados Compustat foi propriedade da empresa Standard & Poors e existe desde o ano de 1962. Contém informação contabilística, financeira e estatística de cerca de 16.775 empresas, activas e inactivas, sediadas em vários pontos do Globo. No que respeita aos EUA a base de dados é particularmente extensa.

O ficheiro cedido pelo Professor Eduard Altman identifica um total de 2025 empresas dos EUA que faliram entre os anos de 1976 e 2006.

A construção das amostras de treino, teste e validação, passou por várias fases, até apresentar os dados e as características necessárias à elaboração do Modelo predictivo.

Numa 1ª fase, foi retirada da Compustat a informação contabilística das empresas falidas (identificadas no ficheiro de Altman). Em termos de dados anuais foram apenas seleccionados os dados referentes aos 2 penúltimos anos de actividade dessas mesmas empresas. Houve o cuidado de excluir os casos em que a empresa depois de declarar falência continuou com as suas actividades, após um processo de reabilitação, sem ter sofrido desmembramento e liquidação.

Na 2ª fase, também da Compustat foram seleccionados dados de empresas solventes para emparelhar com as falidas. Assim, esta sub-amostra tem a particularidade de:

1. conter empresas pertencentes ao mesmo sector de actividade das falidas;
2. conter empresas com a mesma dimensão aproximada da das falidas;
3. encontrarem-se activas e saudáveis (com lucro), nos mesmos anos em que as empresas insolventes faliram.

²² Variáveis macroeconómicas, de mercados financeiros, de instituições de Rating

Em suma, pretendeu-se construir uma amostra de empresas solventes e outra amostra de empresas insolventes para emparelhar com as anteriores, portanto oriundas dos mesmos sectores de actividade, com a mesma dimensão, e em actividade para a mesma sequência de 2 a 3 anos.

A 3ª fase consistiu numa análise aos dados financeiros dos sectores de actividade disponíveis. Decidiu-se retirar das amostras todas as empresas não cotadas, bem como as respeitantes a sectores de actividade Bancária/Financeira, Electricidade e Gás e sector Aeronáutico. Isto por se verificar que, em termos contabilísticos, estes sectores apresentam resultados atípicos face aos restantes sectores de actividade, ou seja, os resultados contabilísticos não seguem uma evolução padronizada ao longo dos anos, podendo haver um ano de lucros e logo o ano seguinte corresponder a um ano de prejuízos significativos.

4ª fase²³ - Selecção de variáveis. A Compustat apresenta um total de 464 variáveis contabilísticas/financeiras; todavia, para a análise em apreço e atendendo aos estudos realizados anteriormente por autores de referência como Altman [38] e Taffler [32], retirou-se da amostra as variáveis que não haviam sido seleccionadas por estes autores, pois tais dados muito provavelmente não trariam qualquer informação útil para o modelo que se pretende obter. Após terem sido removidas tais variáveis, restaram 39 variáveis.

5ª fase - Transformação de variáveis. Atendendo a que um dos objectivos deste trabalho é aplicar diferentes metodologias, foi necessária a criação de variáveis compostas, ou seja, rácios contabilísticos não existentes à partida na Compustat. Designadamente, as variáveis utilizadas por Altman [38] e Taffler [32] nos respectivos modelos de previsão de falências são todos rácios. Foram ainda consideradas em termos individuais as variáveis que integram cada um dos rácios.

Assim foram utilizadas as seguintes variáveis:

Modelo Altman:

$$X1 = \frac{\text{Fundo de Maneio}}{\text{Activo Total}} \quad (10)$$

$$X2 = \frac{\text{Resultados Retidos}}{\text{Activo Total}} \quad (11)$$

$$X3 = \frac{\text{EBIT}^{24}}{\text{Activo Total}} \quad (12)$$

²³ Ver Anexo 1 – Listagem variáveis utilizadas nas amostras

²⁴ Resultados antes de Juros e Impostos

$$X4 = \frac{\text{Capital Próprio}}{\text{Passivo Total}} \quad (13)$$

$$X5 = \frac{\text{Vendas}}{\text{Activo Total}} \quad (14)$$

Modelo Taffler:

$$CF_{TTA} = \frac{\text{Cash Flow}}{\text{Total Activos}} \quad (15)$$

$$S_{TL} = \frac{\text{Vendas}}{\text{Passivo Total}} \quad (16)$$

$$STL_{NCE} = \frac{\text{Passivo Curto Prazo}}{\text{Fundos Permanentes}^{25}} \quad (17)$$

$$WC_{TTA} = \frac{\text{Fundo Maneio}}{\text{Activo Total}} \quad (18)$$

6ª fase - Transformação das variáveis contabilísticas da Compustat em variáveis Logarítmicas²⁶. Conforme mencionado por Trigueiros [39], “*os dados contabilísticos não são imprevisíveis... obviamente obedecem a regras bem definidas*”. Variáveis como Vendas, Activos, Resultados Líquidos, apresentam características multiplicativas, não permitindo a sua modelação através de modelos paramétricos, gaussianos²⁷ aos quais apenas variáveis aditivas podem responder.

²⁵ Capital Próprio + Dívida de Longo Prazo

²⁶ Antecedendo a transformação logarítmica, procedeu-se à conversão dos dados de milhões para unidade (multiplicação por 10⁶)

²⁷ Distribuição normal

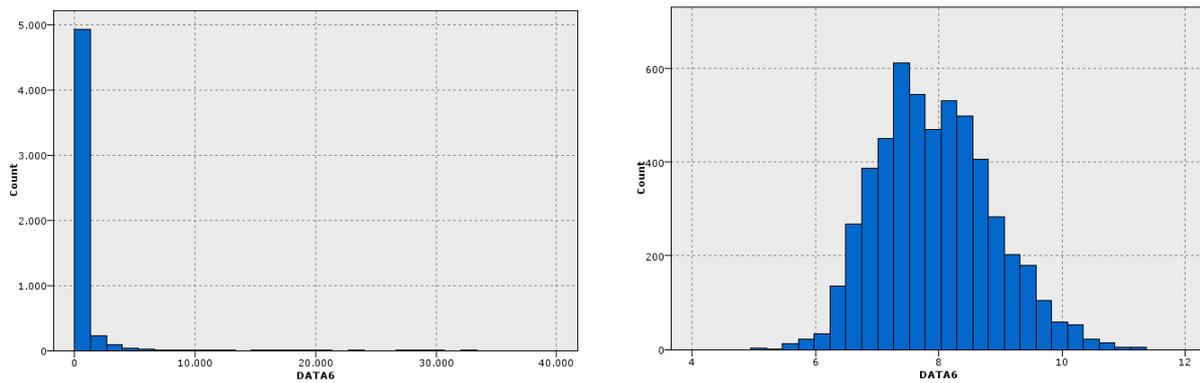


Figura 9 – Gráficos representativos da variável Data 6 – Activos Totais, Antes e Após a transformação logarítmica

Assim, tratando-se de variáveis multiplicativas ou Log-Normais, Trigueiros [40] propõe a sua transformação em variáveis Normais, atribuindo às variáveis um comportamento aditivo próximo da Distribuição Normal. Consoante os dados sejam negativos, positivos ou nulos, a transformação logarítmica aplicada seguirá as seguintes fórmulas:

$$\text{Se } var > 0 \Rightarrow \text{Log}(var) = \log_{10}(var) \quad (20)$$

$$\text{Se } var < 0 \Rightarrow \text{Log}(var) = -\log_{10}(-var) \quad (21)$$

$$\text{Se } var = 0 \Rightarrow \text{Log}(var) = 0 \quad (22)$$

7ª fase – Construção das amostras de treino e de teste/validação (escolhidas ao acaso de entre o total da amostra) e definição dos dois grupos de variáveis a usar, designadamente:

- Grupo das variáveis simples, apenas com transformação logarítmica
- Grupo dos rácios de Altman e Taffler a partir de variáveis simples não transformadas.

Ambas as amostras se caracterizam por conter casos de empresas solventes e insolventes. Em suma, contamos com quatro sub-amostras, duas de treino e duas de teste/validação.

	Variáveis Logarítmicas		Rácios Altman e Taffler	
	Amostra Treino	Amostra Validação	Amostra Treino	Amostra Validação
Nº Empresas Solventes	1789	873	1789	873
Nº Empresas Insolventes	733	506	733	506
Nº Registos Solventes	2993	2887	2993	2887
Nº Registos Insolventes	2300	2675	2300	2675
Nº Total Registos	5293	5562	5293	5562
Nº Sector Actividade	166	209	166	209

Figura 10 – Composição das Amostras

3.2 *Modelação dos dados*

Na modelação dos dados foram usados apenas dois dos algoritmos descritos atrás. Em representação dos modelos paramétricos optou-se pela Regressão Logística; e em representação das heurísticas e modelos não paramétricos recorreu-se à Rede Neuronal, designadamente a uma arquitectura de rede de aprendizagem dirigida, um Multi-Layer Perceptron (Perceptrão de camadas múltiplas)²⁸.

A regressão logística é adequada aos requisitos de Basileia II para o IRB ou AIRB. Quanto à rede neuronal, ela pode violar o requisito da transparência e da robustez, devido à complexidade que pode atingir. Porém, como se verá abaixo, este problema foi contornado com a imposição de uma arquitectura parcimoniosa e a renúncia ao uso de métodos herméticos na preparação e pós-processamento de dados.

As metodologias aplicadas na tese diferem também em relação ao que é habitual, devido ao tipo de atributos utilizados: conforme foi referido na secção anterior, há atributos que são variáveis contabilísticas simples com transformação logaritmica e também outros que são rácios. Deste modo poder-se-á afirmar que a forma como os dados estão transformados irá ditar o modelo futuro. Por si só, esta faceta do presente estudo constitui uma metodologia de trabalho diferenciada em relação à literatura existente.

Recorrendo a uma ferramenta de Data Mining, procedeu-se à elaboração dos modelos previsionais por forma a seleccionar aquele que, com o menor número de variáveis, tenha capacidade para prever a solvência/insolvência das empresa.

Depois de aplicados às amostras em apreço modelos de Regressão Logística e Redes Neurais, obtiveram-se os resultados que se encontram no Anexo B.

O modelo final ou que acabou por ser seleccionado²⁹, por ser parcimonioso no uso de variáveis mas sem sacrificar a precisão, apresenta as seguintes características:

1. usa tanto rácios como variáveis simples transformadas;
2. resulta da aplicação de uma Rede Neuronal e, conforme ilustrado na figura abaixo, apresenta uma percentagem de classificações correctas que ronda os 97% na amostra de validação, sendo também razoavelmente equilibrado na proporção de erros tipo 1 e 2;

²⁸ MLP é uma rede neuronal de aprendizagem supervisionada. Topologicamente as camadas têm a configuração feed-forward (camada de input para camada de output), os nós/ neurões estão organizados em camadas, sendo que o output do nó anterior está conectado com o input do nó posterior. A aprendizagem é supervisionada, atendendo ao facto de o operador poder decidir no numero de nós e camadas, ou dar indicações à rede de outras alternativas possíveis, tal como o exaustive pruning (à medida que a rede vai aprendendo vão sendo retiradas as variáveis cuja contribuição para o modelo tem pouca relevância).

²⁹ Ver Anexo C, output do Modelo

3. requer apenas 8 variáveis (8 neurões na camada de Input) e usa 2 neurões na camada escondida e 1 neurão na camada de saída ou output da Rede Neuronal. Trata-se portanto de um modelo de pequena complexidade.

Eis a sua descrição:

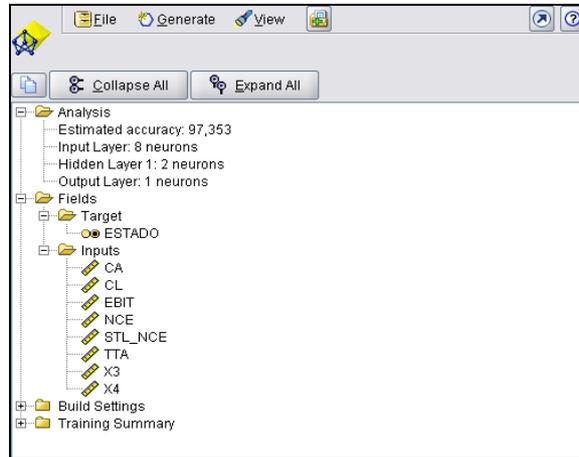


Figura 11 – Sumário das características do modelo

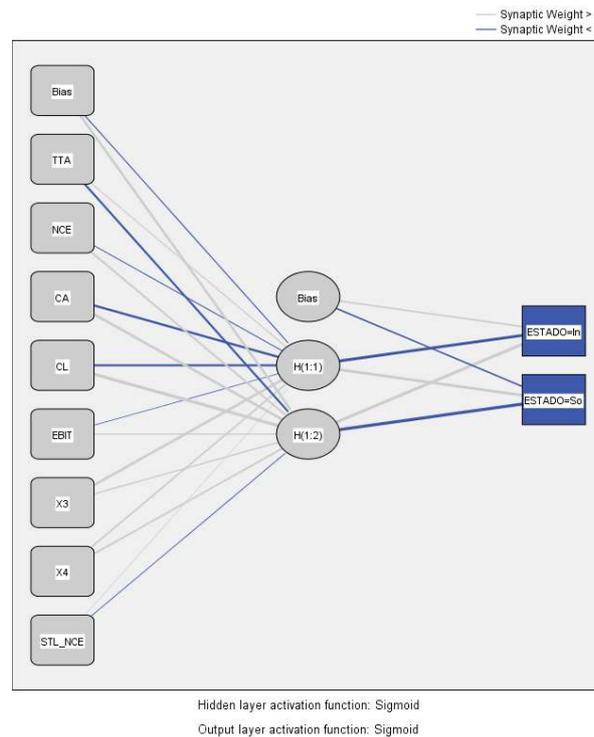


Figura 12 – Representação da rede neuronal

A lista abaixo representa as variáveis e Rácios Altman / Taffler seleccionadas pelo modelo, bem como a sua importância para os resultados obtidos:

- CA – Logaritmo de Activos Circulantes
- CL - Logaritmo de Passivo de Curto Prazo

- TTA – Logaritmo de Total Activos
- STL_NCE – Rácio T3 de Taffler: Passivo de Curto Prazo/ Capital disponível
- EBIT – Logaritmo de Resultados antes de Juros e Impostos
- X3 – Rácio X3 de Altman: EBIT / Total Activos
- X4 – Rácio X4 de Altman: Capital Próprio/ Total Passivo

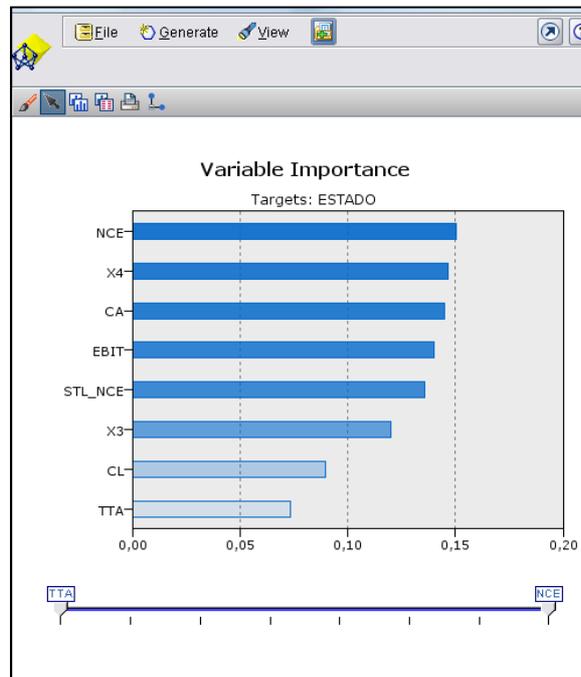


Figura 13 – Importância explicativa relativa das variáveis seleccionadas para o modelo de rede neuronal

Resultados detalhados por sub-amostra podem também ser consultados no Anexo B. De notar a pequena diferença entre poder explicativo observado entre a amostra de treino e a de validação.

3.3 Simulação do Modelo e cálculo da Probability of Default (PD)

Para a simulação do uso comercial do modelo obtido, será criado um cenário hipotético onde se supõe que estamos perante uma instituição de crédito que, para efeitos de cálculo do risco de crédito associado aos empréstimos bancários que atribuiu, e ao abrigo das regras de Basileia II, pretende utilizar o método IRB³⁰.

³⁰ Consultar secção 2.1.2, designadamente Funções dos Ponderadores de Risco na IRB

Neste âmbito, a referida instituição de crédito, baseando-se na sua carteira de clientes e a informação que dispõe, irá desenvolver um algoritmo de cálculo do ponderador de risco PD (Probabilidade de falência) e que, combinado com os restantes ponderadores de risco (EAD e LGD), irá permitir apurar a EL (Perdas Esperadas). Estes mesmos ponderadores permitem também obter a distribuição das EL (Perdas Esperadas) e o VAR (Value at Risk) da sua hipotética carteira de clientes.

Essa hipotética carteira de clientes será o resultado de uma nova amostragem³¹, retirada da Compustat, apenas com dados de empresas activas referentes ao ano de 2004³², num total de 3633 registos.

O primeiro passo da simulação consistiu em submeter os dados da nova amostra ao modelo de Rede Neuronal (MLP), desenvolvido na secção anterior. Como resultado, foi obtido um par de valores associados a cada empresa.

O primeiro valor, entre 0 e 1, designado por \$NC-Estado,³³ representa uma pseudo-probabilidade (condicional), apurada pela Rede Neuronal, de a empresa vir a ser insolvente ou solvente sob certas condições: no caso do outro valor de output chamado \$N-Estado identificar a empresa como solvente, então a variável \$NC – Estado contém a pseudo probabilidade de tal ocorrência, condicional nos valores de input e nas proporções (observadas na amostra) entre estes dois estados; no caso de \$N-Estado apontar para a insolvência, então \$NC-Estado conterá a pseudo probabilidade condicional de insolvência.

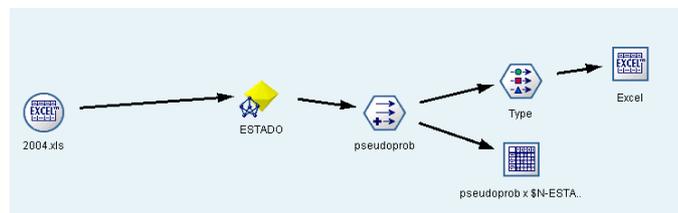


Figura 14 – Esquema de Modelo de Simulação

³¹ A amostra hipotética da carteira de clientes foi trabalhada por forma a assumir as variáveis construídas para a amostra de teste Altman e Taffler.

³² Foram seleccionados dados referentes ao ano de 2004, por se tratar de um ano mais actual e corresponder a um ano considerado não recessivo (ver gráfico Anexo D)

³³ Variável produzida pela rede neuronal

	TTA	NCE	CA	CL	EBIT	X3	X4	STL_NCE	\$N-ESTA	\$NC-ESTADO	pseudoprob
1	24520000	-30000000	205300000	189400000	172000000	0.070	-0.070	-631.333	Insolvente	0.999	0.999
2	160024000	-1790000	809510000	87439000000	120230000	0.075	-0.214	-488.486	Solvente	0.957	0.043
3	193800000	-40000000	172800000	194100000	25490000	-1.315	-0.002	-485.250	Insolvente	1.000	1.000
4	437860000	-42000000	157910000	19288000000	130000000	0.000	-0.209	-459.238	Insolvente	1.000	1.000
5	141100000	-40000000	111900000	932000000	-37900000	-0.269	-0.005	-233.000	Insolvente	0.999	0.999
6	382800000	-15000000	275200000	3469000000	-13050000	-0.332	-0.011	-231.267	Insolvente	1.000	1.000
7	255750000	-56000000	115730000	12206000000	270000000	0.106	-0.202	-217.964	Insolvente	1.000	1.000
8	1386930000	-6370000	799920000	74169000000	458400000	0.033	-0.272	-116.435	Insolvente	1.000	1.000
9	314020000	-15450000	978750000	102793000000	336550000	0.107	-0.441	-66.533	Insolvente	1.000	1.000
10	1233390000	-24100000	146390000	146240000000	676700000	0.055	-0.609	-60.880	Solvente	0.957	0.043
11	458200000	-81000000	438500000	44460000000	208000000	0.045	-0.030	-54.889	Insolvente	1.000	1.000
12	728000000	-40000000	352000000	19000000000	-312000000	-0.429	-0.041	-47.500	Insolvente	0.999	0.999
13	566000000	-95000000	282300000	41550000000	401000000	0.071	-0.474	-43.737	Insolvente	1.000	1.000
14	391000000	-80000000	314000000	32900000000	-375000000	-0.959	-0.024	-41.125	Insolvente	1.000	1.000
15	288040000	-55400000	256880000	207200000000	250500000	0.087	-0.027	-37.401	Insolvente	1.000	1.000
16	149908000	-15760000	240390000	571200000000	357190000	0.238	-0.636	-36.244	Insolvente	1.000	1.000
17	133310000	-16000000	758200000	578800000000	-800000000	-0.006	-0.028	-36.175	Insolvente	1.000	1.000
18	340000000	-100000000	240000000	300000000000	-416000000	-12.235	-0.816	-30.000	Insolvente	1.000	1.000
19	278190000	-47800000	143280000	1311100000000	376400000	0.135	-0.822	-27.429	Insolvente	1.000	1.000
20	113300000	-45000000	838000000	1048000000000	-328000000	-0.289	-0.073	-23.289	Insolvente	0.999	0.999
21	210000000	-100000000	130000000	2200000000000	-210000000	-10.000	-0.045	-22.000	Insolvente	1.000	1.000
22	713000000	-190000000	140000000	4120000000000	-348000000	-0.488	-0.058	-21.684	Insolvente	0.999	0.999
23	600000000	-300000000	470000000	6400000000000	-160000000	-0.267	-0.847	-21.333	Insolvente	0.969	0.969
24	532000000	-120000000	495000000	2470000000000	-431000000	-0.810	-0.049	-20.583	Insolvente	1.000	1.000
25	716700000	-135000000	441500000	2673000000000	921000000	0.129	-0.475	-19.800	Insolvente	1.000	1.000
26	385740000	-106300000	227370000	2058200000000	348800000	0.090	-0.572	-19.362	Insolvente	1.000	1.000
27	313400000	-590000000	126900000	10770000000000	240000000	0.077	-0.250	-18.254	Insolvente	0.997	0.997
28	800000000	-500000000	790000000	85000000000000	-250000000	-0.312	-0.059	-17.000	Insolvente	0.994	0.994
29	394000000	-160000000	261000000	25800000000000	-550000000	-0.140	-0.151	-16.125	Insolvente	0.969	0.969
30	180660000	-272000000	305500000	41840000000000	201000000	0.111	-0.536	-15.382	Solvente	0.952	0.048
31	870000000	-600000000	140000000	92000000000000	-590000000	-0.678	-0.232	-15.333	Insolvente	0.999	0.999
32	147800000	-102000000	122400000	155000000000000	-455000000	-0.308	-0.066	-15.196	Insolvente	0.999	0.999

Figura 15 – Fragmento do Output do Modelo de Simulação

O cálculo da PD implica a normalização e a aplicação da Regra de Bayes³⁴ [14] a essa probabilidade condicional de insolvência, tal como se pode deduzir dos outputs da Rede Neuronal.

Neste estudo, a variável \$NC-Estado obtida como output da Rede Neuronal pode assumir qualquer valor entre 0 e 1, pois é o resultado (manipulado) da aplicação de uma função logística à soma de todos os nós da camada intermédia da rede. A variável \$NC-Estado não pode ser utilizada directamente. Em primeiro lugar, é preciso transformá-la na pseudo-probabilidade de insolvência apenas (pseudo-prob), com o uso da seguinte instrução:

```
IF $NC-Estado = 'solvente'
THEN pseudo-prob = 1 - $NC-Estado
ELSE pseudo-prob = $NC-Estado
ENDIF
```

³⁴O Teorema de Bayes postula que é possível alterar uma análise à luz de novas evidências, assim, através de uma regra matemática, é possível combinar dados existentes com novos dados, por forma a atingir novos resultados mais precisos. A regra de Bayes formula-se em termos probabilísticos como:

$$P(A|B) = \frac{P(B|A) * P(A)}{\sum_{n=0}^k P(B|A_n) * P(A_n)}$$

Simplificando para uma hipótese e uma evidencia:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

- P(A | B) a probabilidade de que a hipótese A seja verdadeira dada a evidencia B, ou a probabilidade à posteriori de um caso exibir uma colecção de atributos B pertencer a A.
- P(B | A) a probabilidade que a evidencia B será observada se a hipótese A for verdadeira, ou por outras palavras, probabilidade condicional de caso que pertence a A, ter a colecção de atributos B.
- P(A) a probabilidade a priori que a hipótese A é verdadeira na ausência de qualquer evidencia específica.
- k o numero de hipóteses possíveis

Depois aplica-se a esta pseudo-probabilidade a transformação inversa da logística, obtendo-se um índice que pode variar entre menos e mais infinito. Este passo é necessário apenas por ser impossível, no caso presente, aceder aos output gerados pela Rede Neuronal antes da transformação logística final.

Este índice (o output antes da última função logística) é então estandardizado (ou mesmo normalizado quando necessário, sem que a ordem relativa de cada caso seja modificada) e procede-se depois à transformação inversa da anterior, portanto uma transformação logística. Obtém-se uma variável que, como antes, apenas pode assumir valores entre 0 e 1, mas agora com um comportamento normalizado.

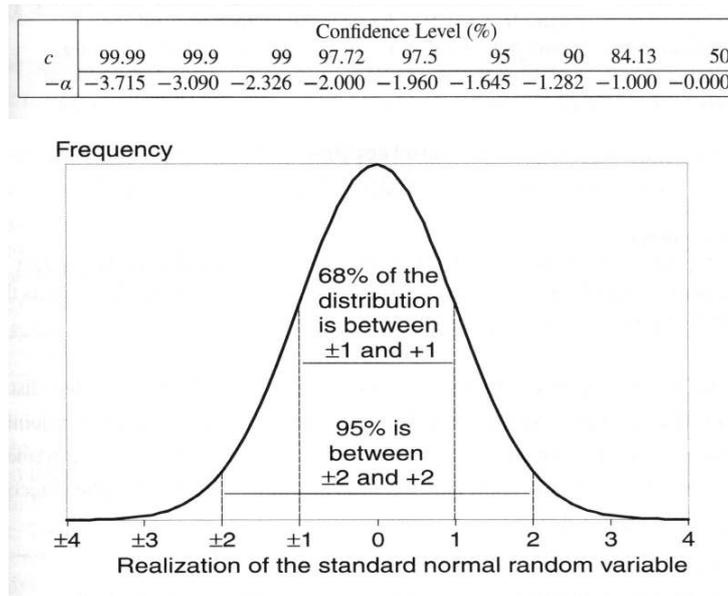


Figura 16 – Variáveis estandardizadas e níveis de confiança [11]

A seguir, com o conhecimento que temos da probabilidade a priori de insolvência e das proporções existentes na amostra de treino entre o número de casos de solvência e de insolvência, é possível obter a probabilidade a posteriori ou real de insolvência, correspondendo à identificada como PD.

Para aceder às probabilidades a priori pode recorrer-se, por exemplo, à tabela infra, que dá probabilidades médias de falência por “Rating” atribuído, ou a outras tabelas semelhantes por sector industrial, por período, por variação no Produto, etc. No caso concreto desta simulação e para simplificar, assumiu-se uma probabilidade a priori de falência média constante de 7% para toda a amostra. O valor é credível.

Rating	Year									
	1	2	3	4	5	6	7	8	9	10
Aaa	0.00	0.00	0.02	0.09	0.19	0.29	0.41	0.59	0.78	1.02
Aa	0.07	0.22	0.36	0.54	0.85	1.21	1.60	2.01	2.37	2.78
A	0.08	0.27	0.57	0.92	1.28	1.67	2.09	2.48	2.93	3.42
Baa	0.34	0.99	1.79	2.69	3.59	4.51	5.39	6.25	7.16	7.99
Ba	1.42	3.43	5.60	7.89	10.16	12.28	14.14	15.99	17.63	19.42
B	4.79	10.31	15.59	20.14	23.99	27.12	30.00	32.36	34.37	36.10
Caa-C	14.74	23.95	30.57	35.32	38.83	41.94	44.23	46.44	48.42	50.19
Inv.	0.17	0.50	0.93	1.41	1.93	2.48	3.03	3.57	4.14	4.71
Spec.	3.83	7.75	11.41	14.69	17.58	20.09	22.28	24.30	26.05	27.80
All	1.50	3.09	4.62	6.02	7.28	8.41	9.43	10.38	11.27	12.14

Rating	Year									
	11	12	13	14	15	16	17	18	19	20
Aaa	1.24	1.40	1.61	1.70	1.75	1.85	1.96	2.02	2.14	2.20
Aa	3.24	3.77	4.29	4.82	5.23	5.51	5.75	5.98	6.30	6.54
A	3.95	4.47	4.94	5.40	5.88	6.35	6.63	6.94	7.23	7.54
Baa	8.81	9.62	10.41	11.12	11.74	12.33	12.95	13.49	13.93	14.39
Ba	21.06	22.65	24.23	25.61	26.83	27.96	29.13	30.24	31.14	32.05
B	37.79	39.37	40.85	42.33	43.62	44.94	45.91	46.68	47.32	47.60
Caa-C	52.30	54.4	56.24	58.22	60.08	61.78	63.27	64.81	66.25	67.59
Inv.	5.30	5.90	6.46	7.00	7.48	7.92	8.30	8.65	8.99	9.32
Spec.	29.47	31.08	32.64	34.07	35.36	36.58	37.72	38.78	39.67	40.46
All	13.01	13.85	14.66	15.40	16.07	16.69	17.24	17.75	18.21	18.64

Figura 17 – Taxas de Incumprimento acumuladas (em percentagem) 1920-2002 da Moody's [11]

Em suma, através deste processo de aferição, o qual usa normalização e a regra de Bayes para estimar, sobre as conhecidas probabilidades a priori e condicionais de falência e a probabilidade a posteriori, resulta a PD³⁵ (Probability of Default).

Calculada a PD, e conhecendo-se os valores para a LGD e EAD, podem calcular-se perdas esperadas, EL, para cada uma das empresas que integram a amostra de simulação. A soma de todas essas EL dará o valor das perdas esperadas na carteira.

Assim, recapitulando em termos de formula da **Expected Loss (EL)**

$$EL = PD \times LGD \times EAD \quad (23)$$

LGD - No nosso estudo, foi assumido que o LGD é a dívida a longo prazo da empresa, e corresponde ao valor que o banco credor perde quando uma empresa vai à falência sem reembolsar os montantes em dívida.

EAD – No nosso estudo este factor multiplicador representa a ideia de que, em geral, numa falência o banco perde quase tudo, mas não perde tudo porque fica com o direito a vender os bens da empresa. Assim, vamos dar a este factor EAD o valor de 1 menos um quinto do

³⁵ Assume-se que as PD calculadas correspondem às PD de cada empresa

rácio dos activos fixos a dividir pelos activos totais, o que tenta representar 1 - o que o banco consegue aproveitar em caso de falência.

Após o cálculo da EL para cada empresa, resulta uma colecção de EL correspondendo a cada uma das ocorrências da amostra e é possível, a partir desta distribuição, estimar também a dispersão dessas perdas esperadas.

Quanto à eventual estimação do VAR, o procedimento seguido é o de Monte Carlo: para cada empresa, cria-se uma variável aleatória F que pode assumir apenas dois valores, 0 e 1. A ocorrência desses valores obedece a um mecanismo de Bernulli (Binomial) onde a probabilidade de ocorrência do 1 é dada pela PD já estimada acima para essa empresa.

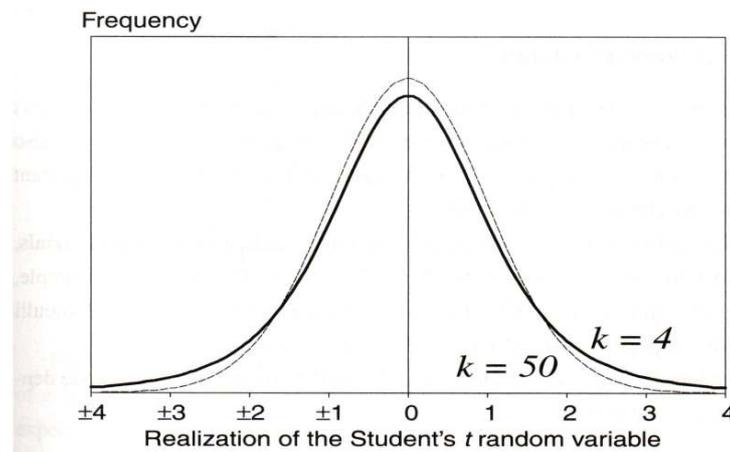


Figura 18 – Função Student, outra aproximação usada para estimar o VAR [11]

A seguir procede-se ao produto, para cada empresa, de LGD por EAD (sem entrar em conta com a PD), mas depois somam-se apenas os produtos onde se verifica que $F = 1$. Pode assim obter-se um valor simulado de uma perda real, possível, na carteira de créditos detida pelo banco, por oposição a uma perda esperada ou media a que a carteira está exposta.

Repetindo o procedimento muitas vezes para valores de F obtidos aleatoriamente obtém-se uma colecção de perdas reais ou possíveis. Claro que também se poderia calcular a perda real da carteira somando as perdas de todas as combinações de falência/solvência possíveis, o que se torna impraticável a partir de carteiras com uns 4 créditos ou mais.

A distribuição das perdas reais é a colecção de perdas reais obtida por sucessivas simulações. Construindo uma série ordenada da menor para a maior dessas perdas reais e identificando a perda correspondente ao percentil 5, é possível apurar o VAR: a maior perda possível para um grau de verosimelhança de 95%. Deve notar-se porém que este VAL não foi aferido ainda para um horizonte temporal concreto.

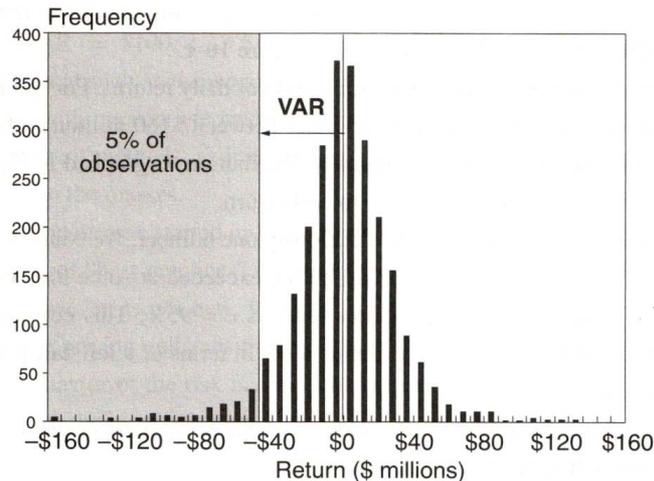


Figura 19 – O VAR num histograma da distribuição de ganhos e perdas [11]

3.4 *Análise de Resultados*

Os resultados da secção anterior foram desenvolvidos tendo presente a aplicação das redes neuronais no apuramento dos PD para uma dada carteira de créditos. A ideia subjacente a todo este trabalho residiu em canalizar para o modelo todas as variáveis que podem ter poder explicativo e em encontrar o modelo mais fiável e robusto para o cálculo da PD; mas também em exemplificar uma forma bastante simples de realizar o cálculo do VAR, obedecendo às orientações de Basileia II.

As redes neuronais não são a solução para todos os problemas de análise de dados, pois como é sabido, em problemas complexos as redes neuronais comportam-se com verdadeiras caixas negras. Atendendo a que os resultados obtidos não são interpretáveis, nem tão pouco transpostos para situações precisas [40]. Assim, as redes violam requisitos das metodologias IRB.

Todavia, no caso de investigações com dados financeiros e contabilísticos, pelo facto dos problemas serem pouco complexos, as redes têm vantagens devido à sua capacidade para retirar conhecimento dos relacionamentos não lineares existentes entre variáveis e dados. Conforme foi exemplificado na secção anterior é possível criar um modelo simples, mas fiável se for utilizada uma arquitectura parcimoniosa de Rede Neuronal (MLP), ao invés da tradicional regressão logística.

Na secção 2.1.2. foram mencionadas as abordagens ao risco de crédito, recapitulando: Abordagem Standard, Abordagem de Ratings Internos e a Abordagem de Ratings Internos Avançados. Sendo que, na secção anterior foi calculada a PD, a Expected Loss e a VAR de uma carteira de clientes, no âmbito da Abordagem de Ratings Internos Avançados.

Importa agora na presente secção validar alguns aspectos da metodologia desenvolvida na secção anterior, ao abrigo dos princípios estabelecidos pela AIGV³⁶, e verificar a conformidade e aplicabilidade do presente modelo, face aos desígnios da entidade supervisora³⁷.

A AIGV desenvolveu seis princípios aplicados à Validação das abordagens Baseadas em Ratings Internos [41]. Assim o Princípio 1 versa sobre o objectivo da validação; a responsabilidade pela Validação é mencionada no Princípio 2; as técnicas de Validação são abordadas nos Princípios 3, 4 e 5 e o Princípio 6 preconiza a sujeição do processo de Validação a uma supervisão independente.

Os Princípios 3, 4 e 5 são os mais relevantes para a validação do modelo desenvolvido, pois será obedecendo a estes princípios que os sistemas de Rating Interno das entidades bancárias serão directamente validados.

Assim, aspectos como:

1. Disponibilidade de dados de qualidade,
2. Procedimentos quantitativos e a influência humana na aplicação dos dados,
3. Transparência, clareza e verosimilhança do modelo que foi produzido
4. E avaliação qualitativa do modelo por parte de equipas de implementação,

são analisados e estudados aquando da Validação do sistema de Rating [41]. Como já referido, o modelo ora desenvolvido obedece a todos estes critérios.

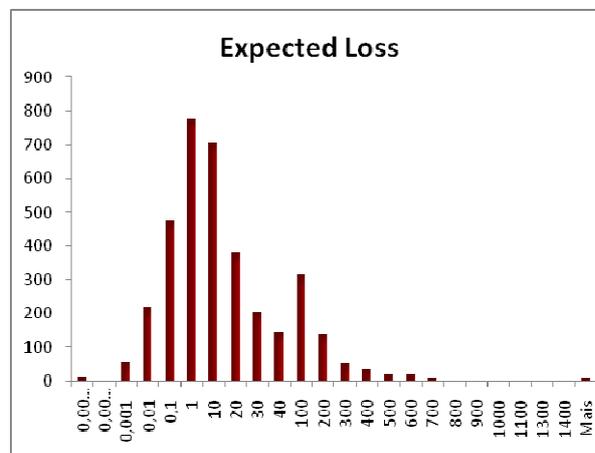


Figura 20 – Histograma com a distribuição das Expected Loss (EL)

As Expected Loss (EL) estão expressas em milhões, sendo que para a amostra em apreço os valores máximos caem entre os 1 e 10 milhões.

³⁶ AIGV – Accord Implementation Group – Sub-group on Validation, traduzido para português como Grupo de Implementação do Acordo – Sub-grupo de Validação

³⁷ No caso de Portugal, a entidade supervisora é o Banco de Portugal

4. Conclusões do Estudo

“Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.”

Thomas H. Davenport

No capítulo de Introdução da tese focaram-se os objectivos a desenvolver com o presente trabalho, toda a problemática gerada em volta do risco, as recentes crises económicas que levaram à falência de várias instituições bancárias e colocaram em grave situação muitas das empresas, com as respectivas repercussões sociais.

O risco nem sempre pode ser previsto, mas face ao aumento da frequência das crises financeiras ao longo do século XX e XXI, as instituições financeiras têm desenvolvido mecanismos que de alguma forma possam estimar a incerteza, designadamente o risco de crédito. Aplica-se o ditado “Mais vale prevenir do que remediar”.

Também as entidades supervisoras têm contribuído para a regularização e supervisão da actividade bancária. Tornou-se de vital importância criar instrumentos que validem e tornem mais fidedignos todos os fluxos financeiros decorrentes da abertura de mercados, do aumento na complexidade dos produtos financeiros e também devido à própria evolução da actividade financeira, que conduziu à criação de verdadeiras redes de negócios. Actualmente, a falência de uma instituição bancária cria uma crise financeira sistémica, devido à já mencionada rede financeira global que liga a maioria das instituições bancárias.

Julgamos ter ficado demonstrado que as Redes Neurais são uma alternativa aos tradicionais modelos estatísticos, são menos exigentes em termos de tratamento de dados, são mais eficazes e também produzem informação útil. À luz dos normativos legais, podemos calcular as Probability of Default, e os VAR, baseando-nos em outputs das Redes Neurais.

Em termos de adequação às exigências de Basileia II, também julgamos ter conseguido mostrar que é possível recorrendo a métodos sobejamente conhecidos, nomeadamente redes neurais, a regra de Bayes, as simulações de Monte Carlo, desenvolver modelos simples de fácil aplicação pelas instituições bancárias. Julgamos ser este o resultado mais relevante do presente trabalho.

As disposições regulatórias de Basileia II apresentam-se tangíveis no que respeita à sua aplicabilidade, permitindo às Instituições Bancárias desenvolver os seus próprios modelos, adequá-los à sua realidade e gerar mecanismos que solidifiquem a actividade bancária da Instituição.

Quanto ao trabalho futuro, o mesmo poderá passar por uma aplicação do Modelo a uma carteira de clientes nacionais de um banco nacional. Julgo ser interessante, quer em termos académicos como comerciais. Aguardam-se contactos!

5. Bibliografia

- [1] **Das, Satyajit.** Perfect Storms”–Beautiful & True Lies In Risk Management. [Online] 2007.
[http://www.wilmott.com/blogs/satyajitdas/enclosures/perfectstorms\(may2007\)1.pdf](http://www.wilmott.com/blogs/satyajitdas/enclosures/perfectstorms(may2007)1.pdf).
- [2] *CRoss Industry Standard Process.* [Online] <http://www.crisp-dm.org/>.
- [3] *Bank of Internacional Settlements.* [Online] <http://www.bis.org/about/history.htm> .
- [4] **Silva, Irménio Ferreira da.** Dissertação de Mestrado em Contabilidade e Administração. 2006.
- [5] Directiva 2006/48/CE do Parlamento Europeu e do Conselho de 14 de Junho de 2006,relativa ao acesso à actividade das instituições de crédito e ao seu exercício (reformulação). *Banco de Portugal.* [Online] 2006.
http://www.bportugal.pt/bank/superv/supervisory_disclosure/documentos/Directiva2006_48_p.
- [6] Directiva 2006/49/CE do Parlamento Europeu e do Conselho de 14 de Junho de 2006 relativa à adequação dos fundos próprios das empresas de investimento e das instituições de crédito. *Banco de Portugal.* [Online] 2008.
http://www.bportugal.pt/bank/superv/supervisory_disclosure/documentos/Directiva2006_49_p.pdf.
- [7] Decreto-Lei nº 103/2007 de 3 de Abril. *Banco de Portugal* . [Online]
http://www.bportugal.pt/bank/superv/supervisory_disclosure/documentos/DL_103_2007_p.pdf.
- [8] Decreto-Lei nº 104/2007 de 3 de Abril. *Banco de Portugal.* [Online] 2007.
http://www.bportugal.pt/bank/superv/supervisory_disclosure/documentos/DL_104_2007_p.pdf.
- [9] Basel II: Consultative Document - Range of practices and issues in economic capital modeling. *Bank for International Settlements.* [Online] 2008.
<http://www.bis.org/publ/bcbs143.pdf?noframes=1>.
- [10] Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. *Bank for International Settlements.* [Online] <http://www.bis.org/publ/bcbs128.htm>.
- [11] **Jorian, Philippe.** Financial Risk Manager Handbook, 3rd Edition. s.l. : Garp & Wiley Finance, 2007.
- [12] Credit risk. *De Nederlandsche Bank.* [Online]
<http://www.dnb.nl/openboek/extern/id/en/bo/41-117007.html>-.

- [13] O novo acordo de Capital – Basileia II, Uma realidade Irreversível. *PARTNERtoPARTNER*. [Online] Junho 2005.
http://www.p2p.com.pt/_pdf/O%20Novo%20Acordo%20de%20Capital%20-%20Basileia%20II_BES.pdf.
- [14] Basel II: Impact Studies (QIS): Overview of documents. *Bank for International Settlements*. [Online] <http://www.bis.org/bcbs/qis/index.htm>.
- [15] **Lewis, Eduard M.** *Introduction to Credit Scoring*. s.l. : Athena Press, 1990.
- [16] **Anderson, Raymond.** *The Credit Scoring Toolkit – Theory and Practice for Retail Credit Risk Management and Decision Automation*. s.l. : Oxford University Press, 2007.
- [17] **Marquez, Javier.** *An Introduction to Credit Scoring for Small and Medium Size Enterprises*. [Online] 2007.
<http://siteresources.worldbank.org/EXTLACOFFICEOFCE/Resources/870892-1206537144004/MarquezIntroductionCreditScoring.pdf>.
- [18] **Komorád, Karel.** *On Credit Scoring Estimation, Masters Thesis*. 2002.
- [19] **Altman, Edward.** *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*. n.º4 Vol. XXIII. s.l. : The Journal of Finance, 1968.
- [20] **Sabzevari, Hassan, Soleymani, Mehdi, Noorbakhsh, Eaman.** *A comparison between statistical and Data Mining methods for Credit Scoring in case of limited available data*. [Online] 2007.
<http://www.crc.man.ed.ac.uk/conference/archive/2007/papers/sabzevari-et-al.pdf>.
- [21] **Englemann, Brend, Rauhmeier, Robert.** *The Basel II Risk Parameters, Estimation, Validation and Stress Testing*. s.l. : Springer, 2006.
- [22] **Camões, Francisco.** *Dissertação de Mestrado em Ciências Empresariais*. s.l. : ISCTE, 1992.
- [23] **Larose, Daniel T.** *Discovering Knowledge in Data: An Introduction to Data Mining*. s.l. : Wiley, 2004.
- [24] **Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.** *Classification and Regression Trees*. s.l. : Wadsworth & Brooks and Cole Advanced Books & Software, 1984.
- [25] **V., Kass G.** *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *Journal of Applied Statistics*. Vol. 29, No. 2, 1980.
- [26] **Di, Wang.** *Long Term Fixed Mortgage Rate Prediction Using Multivariate Adaptive Regression Splines*. s.l. : School of Computer Engineering, Nanyang Technological University, 2006.
- [27] *An Explanatory Note on Basel II IRB Weight Functions*. *Bank for International Settlements*. [Online] Julho 2005.
<http://www.bis.org/bcbs/irbriskweight.pdf?noframes=1>.

- [28] **Beaver, William H.** Financial Ratios as Predictors of Failure. 1966.
- [29] **Berg, Daniel.** *Bankruptcy Prediction by Generalized Additive Models*. s.l. : University of Oslo Statistical Research Report No. 1 ISSN 0806–3842, 2005.
- [30] *Moody's KMV*. [Online] 2009. <http://www.moodyskmv.com/>.
- [31] **Taffler, R.J.** Forecasting Company Failure in the UK using Discriminant Analysis and Financial Ratio Data. s.l. : Royal Statistical Society, 1982.
- [32] **Pongsatit, S., Ramage J., Lawrence, H.,** Bankruptcy Prediction for Large and Small Firms in Asia: A comparison of Olshon and Altman. 2004.
- [33] **Bonfim, Diana.** *Credit Risk Drivers: Evaluating the contribution of firm level information and of macroeconomics dynamics*. s.l. : Banco de Portugal, 2007.
- [34] **Chartkou A., Chigrinov e., Mchedlishvili T.,** *Assessing Probability of Bankruptcy: comparing accounting and Black-Scholes-Merton models*. s.l. : University of Gothenburg, 2006.
- [35] **Bernhardsen, E.** *A model of Bankruptcy Prediction*. s.l. : Norges Bank, 2001.
- [36] **Begley, J., Ming J., Watts, S.** Bankruptcy Classification Errors in the 1980's: An Empirical Analysis of Altman's and Ohson's Models,. 1996.
- [37] **Altman, Eduard.** The Success of Business Failure Prediction Models, An International Survey. s.l. : Elsevier Science Publishers B.V., , 1984.
- [38] **Trigueiros, Duarte.** The Modeling of Accounting Data. 2006.
- [39] **Trigueiros, Duarte.** Neural Networks in Finance and Investing, Accounting and Financial Research. 2006.
- [40] *Raciocínio com Incerteza*. [Online]
http://hermes.ucs.br/carvi/cent/dpei/jltsilva/JLT_Aula-Incerteza.pdf.
- [41] **Blochwitz, Stefan, Hohl, Stefan.** *Validation of Banks' Internal Rating Systems – A Supervisory Perspective*. s.l. : Deutsche Bundesbank and BIS, 2008.

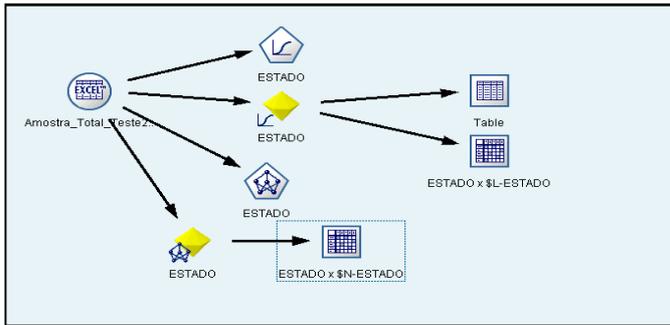
Anexo A

Tabela 1 - Tabela identificadora das variáveis da Compustat utilizadas na amostra.

NAME	TYPE	WIDTH	DECIMAL	LABEL	COLUMNS	MEASURE
DNUM	Numeric	4	0	Industry Classification Code	6	Scale
NAICS	String	6	0	N. American Industry Classification Sys.	8	Nominal
FYR	Numeric	2	0	Fiscal Yearend Month of Data	5	Nominal
UCODE	Numeric	1	0	Update Code	7	Nominal
CONAME	String	28	0	Company Name	26	Nominal
INAME	String	28	0	Industry Name	26	Nominal
GVKEY	Numeric	6	0	SPC Permanent Number	8	Scale
yeara	Numeric	4	0	Fiscal Year	7	Nominal
active	Numeric	1	0	1 = Active; 0 = Inactive	8	Nominal
DATA1	Numeric	10	2	Cash and Short-Term Investments (MMS)	12	Scale
DATA2	Numeric	10	2	Receivables - Total (MMS)	12	Scale
DATA3	Numeric	10	2	Inventories - Total (MMS)	12	Scale
DATA4	Numeric	10	2	Current Assets - Total (MMS)	12	Scale
DATA5	Numeric	10	2	Current Liabilities - Total (MMS)	12	Scale
DATA6	Numeric	10	2	Assets - Total (MMS)	12	Scale
DATA9	Numeric	9	2	Long-Term Debt - Total (MMS)	11	Scale
DATA11	Numeric	10	2	Common Equity - Tangible (MMS)	12	Scale
DATA12	Numeric	10	2	Sales (Net) (MMS)	12	Scale
DATA13	Numeric	9	2	Operating Income Before Deprec. (MMS)	11	Scale
DATA14	Numeric	9	2	Depreciation and Amortization (MMS)	11	Scale
DATA15	Numeric	9	2	Interest Expense (MMS)	11	Scale
DATA16	Numeric	9	2	Income Taxes - Total (MMS)	11	Scale
DATA18	Numeric	9	2	Income Before Extraordinary Items (MMS)	11	Scale
DATA20	Numeric	9	2	Income Before EI- Adj for Common...(MMS)	11	Scale
DATA26	Numeric	6	2	Dividends Per Share - Ex-Date (\$&c)	8	Scale
DATA29	Numeric	7	2	Employees (M)	9	Scale
DATA30	Numeric	9	2	Property, Plant & Equip- Capital Ex(MMS)	11	Scale
DATA34	Numeric	10	2	Debt in Current Liabilities (MMS)	12	Scale
DATA36	Numeric	10	2	Retained Earnings (MMS)	12	Scale
DATA101	Numeric	9	2	Interest Expense on Long-Term Debt (MMS)	11	Scale
DATA162	Numeric	9	2	Cash (MMS)	11	Scale
DATA172	Numeric	9	2	Net Income (Loss) (MMS)	11	Scale
DATA178	Numeric	9	2	Operating Income After Deprec. (MMS)	11	Scale
DATA179	Numeric	9	2	Working Capital (Balance Sheet) (MMS)	11	Scale
DATA180	Numeric	8	2	Working Capital Change - Total(SoC)(MMS)	10	Scale
DATA181	Numeric	11	2	Liabilities - Total (MMS)	13	Scale
DATA193	Numeric	9	2	Short-Term Investments (MMS)	11	Scale
DATA216	Numeric	10	2	Stockholders' Equity - Total (MMS)	12	Scale

Anexo B

Modelos com variáveis transformadas em logaritmos



Resultados da Regressão Logística

\$L-ESTADO			
ESTADO	\$null\$	Insolvente	Solvente
Insolvente	127	853	1320
Solvente	315	454	2224

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 347,587, df = 2, probability = 0

Equation For Insolvente
+ 0,00000000000000000000000000000000

Equation For Solvente

$$0,28 * \text{yeara} + 0,000001493 * \text{NAICS} + 0,8392 * \text{DATA6} + 0,1203 * \text{DATA12} + -0,1312 * \text{DATA15} + 0,1465 * \text{DATA34} + 0,1878 * \text{DATA162} + 0,08842 * \text{DATA172} + -1,373 * \text{DATA181} + 0,09902 * \text{DATA193} + -557,6$$

Aplicação da Rede Neuronal

\$N-ESTADO		
ESTADO	Insolvente	Solvente
Insolvente	1652	648
Solvente	942	2051

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 847,438, df = 1, probability = 0

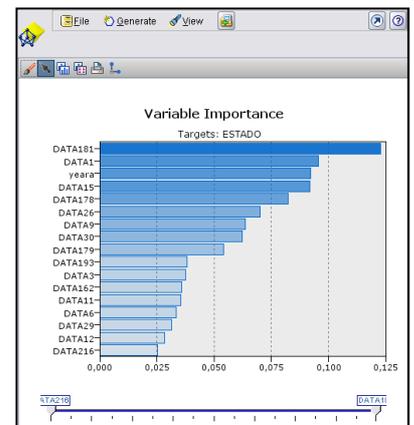
Matrix Appearance Annotations

Analysis

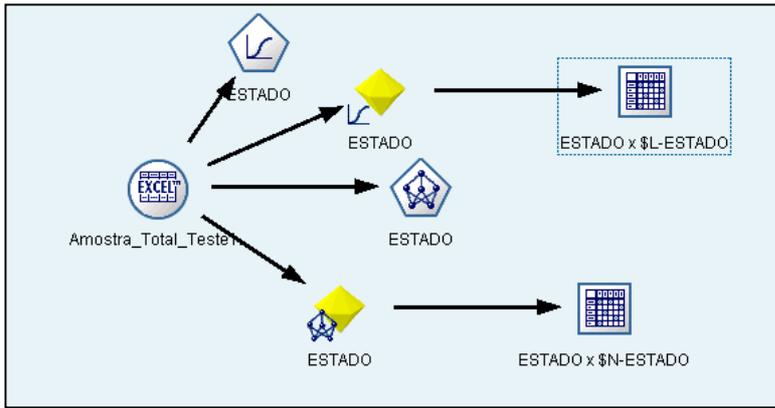
- Estimated accuracy: 69,854
- Input Layer: 17 neurons
- Hidden Layer 1: 4 neurons
- Output Layer: 1 neurons

Fields

- Build Settings
- Training Summary



Modelos com rácios Altman e Taffler



Resultados da aplicação da Regressão Logística

\$L-ESTADO			
ESTADO	\$null\$	Insolvente	Solvente
Insolvente	101	2193	6
Solvente	197	1	2795

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 4.992,834, df = 2, probability = 0

Equation For Insolvente
Base category
+ 0,00000000000000000000

Equation For Solvente
0,0000002181 * TTA +
-0,0000004442 * NCE +
-0,000003701 * CA +
-0,000005562 * CL +
0,000002519 * EBIT +
1,578 * X3 +
0,02331 * X4 +
0,00000004715 * STL_NCE +
+ 3,409

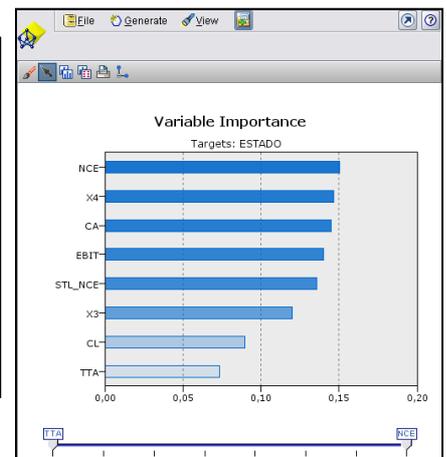
Aplicação da Rede Neuronal

\$N-ESTADO		
ESTADO	Insolvente	Solvente
Insolvente	2237	63
Solvente	61	2932

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 4.800,332, df = 1, probability = 0

Analysis
Estimated accuracy: 97,353
Input Layer: 8 neurons
Hidden Layer 1: 2 neurons
Output Layer: 1 neurons

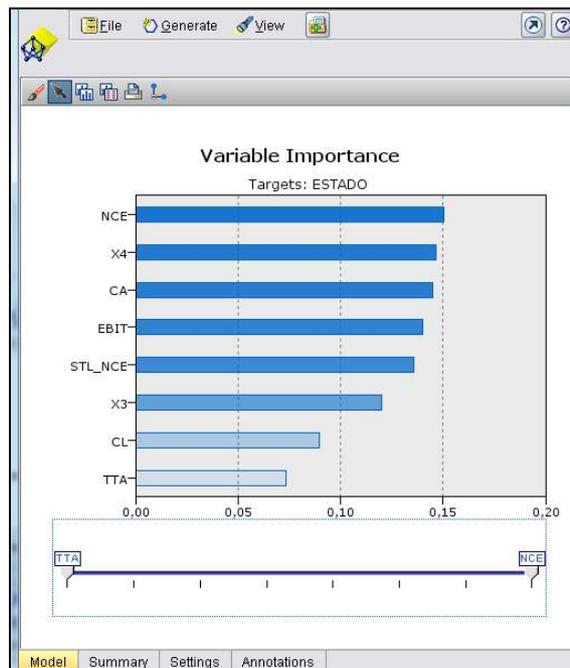
Fields
Build Settings
Training Summary



Resultados da Aplicação da Rede Neuronal à Amostra Validação Rácios Altman e Taffler

\$N-ESTADO		
ESTADO	Insolvente	Solvente
Insolvente	2599	76
Solvente	68	2819

Cells contain: cross-tabulation of fields (including missing values)
Chi-square = 5.000,099, df = 1, probability = 0



File Generate View

Collapse All Expand All

- Analysis
 - Estimated accuracy: 97,353
 - Input Layer: 8 neurons
 - Hidden Layer 1: 2 neurons
 - Output Layer: 1 neurons
- Fields
- Build Settings
- Training Summary

Anexo C

*Multilayer Perceptron Network.

```
MLP ESTADO (MLEVEL=N) WITH TTA NCE CA CL EBIT X3 X4 STL_NCE
/RESCALE COVARIATE=STANDARDIZED
/PARTITION TRAINING=7 TESTING=3 HOLDOUT=0
/ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=1 (NUMUNITS=2) HIDDENFUNCTION=SIGMOID OUTPUTFUNCTION=SIGMOID
/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.0000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET
=0.5
/PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION SOLUTION
/PLOT NETWORK
/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.0010
/MISSING USERMISSING=EXCLUDE .
```

Multilayer Perceptron

Notes

Output Created		
Comments		
Input	Data	C:\Documents and Settings\Duarte\Os meus documentos\Fernanda\Output.sav
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	5293
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Syntax	<pre>MLP ESTADO (MLEVEL=N) WITH TTA NCE CA CL EBIT X3 X4 STL_NCE /RESCALE COVARIATE=STANDARDIZED /PARTITION TRAINING=7 TESTING=3 HOLDOUT=0 /ARCHITECTURE AUTOMATIC=NO HIDDENLAYERS=1 (NUMUNITS=2) HIDDENFUNCTION=SIGMOID OUTPUTFUNCTION=SIGMOID /CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.000005 SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.5 /PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION SOLUTION /PLOT NETWORK /STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15) MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.0010 /MISSING USERMISSING=EXCLUDE .</pre>	
Resources	Processor Time	00:00:00,797
	Elapsed Time	00:00:00,736

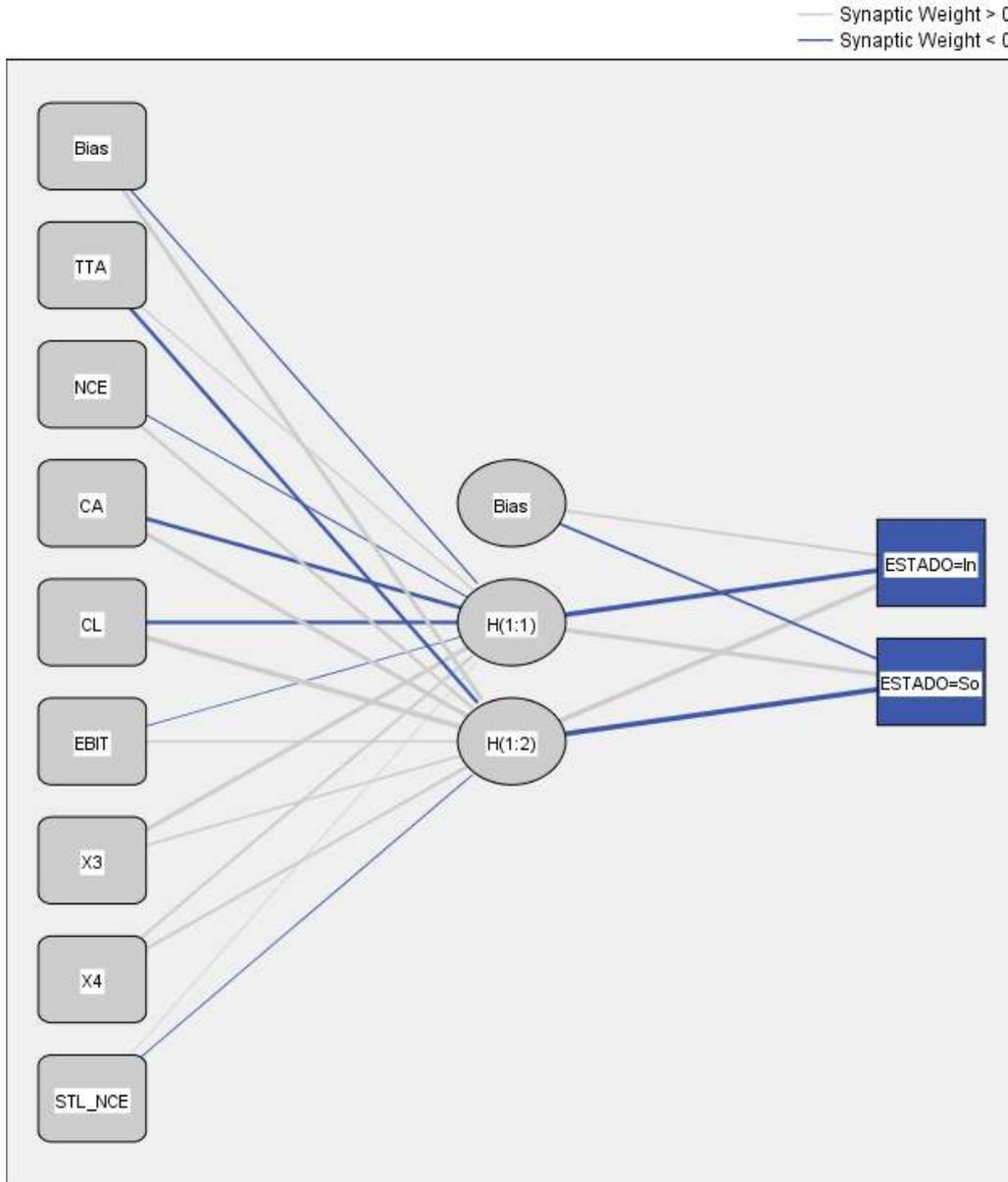
Case Processing Summary

		N	Percent
Sample	Training	3512	70,3%
	Testing	1483	29,7%
Valid		4995	100,0%
Excluded		298	
Total		5293	

Network Information

Input Layer	Covariates	1	TTA	
		2	NCE	
		3	CA	
		4	CL	
		5	EBIT	
		6	X3	
		7	X4	
		8	STL_NCE	
	Number of Units ^a			8
	Rescaling Method for Covariates		Standardized	
Hidden Layer(s)	Number of Hidden Layers			1
	Number of Units in Hidden Layer 1 ^a			2
	Activation Function		Sigmoid	
Output Layer	Dependent Variables	1	ESTADO	
	Number of Units			2
	Activation Function		Sigmoid	
	Error Function		Sum of Squares	

a. Excluding the bias unit



Hidden layer activation function: Sigmoid

Output layer activation function: Sigmoid

Model Summary

Training	Sum of Squares Error	19,678
	Percent Incorrect Predictions	,6%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error ^a
	Training Time	00:00:00,172

Testing	Sum of Squares Error	5,577
	Percent Incorrect Predictions	,3%

Dependent Variable: ESTADO

a. Error computations are based on the testing sample.

Parameter Estimates

Predictor		Predicted			
		Hidden Layer 1		Output Layer	
		H(1:1)	H(1:2)	[ESTADO=In]	[ESTADO=So]
Input Layer	(Bias)	-2,177	7,631		
	TTA	2,536	-7,029		
	NCE	-2,032	6,695		
	CA	-7,215	16,292		
	CL	-7,208	18,907		
	EBIT	-,725	3,135		
	X3	13,696	4,524		
	X4	5,181	6,202		
	STL_NCE	,031	-,959		
Hidden Layer 1	(Bias)			5,015	-4,798
	H(1:1)			-18,118	17,820
	H(1:2)			18,351	-18,405

Classification

Sample	Observed	Predicted
--------	----------	-----------

		In	So	Percent Correct
Training	In	1542	20	98,7%
	So	1	1949	99,9%
	Overall Percent	43,9%	56,1%	99,4%
Testing	In	632	5	99,2%
	So	0	846	100,0%
	Overall Percent	42,6%	57,4%	99,7%

Dependent Variable: ESTADO

Anexo D

O gráfico infra ilustra a influência da macroeconomia no número de falências nos Estados Unidos. Representa a evolução da taxa de incumprimento entre os anos de 1980 e 2004, sendo ainda de notar a sombreado os anos considerados de recessão.

