



CENTRO FEDERAL DE EDUCAÇÃO
TECNOLÓGICA DE MINAS GERAIS

Diretoria de Pesquisa e Pós-Graduação

Curso de Mestrado em Modelagem
Matemática e Computacional

APLICAÇÃO DAS METAHEURÍSTICAS PSO E GA NA ELABORAÇÃO DE UM MODELO DE CREDIT SCORING

Dissertação de Mestrado, submetida ao Curso de Mestrado em Modelagem Matemática e Computacional, como parte dos requisitos para a obtenção do título de Mestre em Modelagem Matemática e Computacional.

Aluno: Vinícius Marinho

Orientador: Prof. Dr. João Francisco de Almeida Vitor (CEFET/MG)

Co-orientador: Prof. Dra. Elenice Biazzi (CEFET/MG)

Belo Horizonte, Setembro de 2009.

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

MARINHO, Vinícius
M338a Aplicação das Metaheurísticas PSO e GA na
Elaboração de um Modelo de Credit Scoring /Vinícius
Marinho 2009.
84f.

Orientador: João Francisco de Almeida Vitor
Co-orientador: Elenice Biazi

Dissertação (mestrado) - Centro Federal de Educação
Tecnológica de Minas Gerais

1. Otimização Combinatória. 2. Algoritmos Genéticos.
I. Vitor, João Francisco de Almeida. II. Biazi, Elenice
III. Centro Federal de Educação Tecnológica de Minas Gerais.
IV. Título

CDD 519.6

Dedico esse trabalho aos meus pais, Marinho e Branca.
Dedico também a minha querida esposa Narinha.
E por fim dedico ao meu grande amigo e irmão Rômulo.

Agradecimentos

Agradeço aos meus pais, Marinho e Branca, que sempre me apoiaram em todos os momentos da minha vida, me deram o seu exemplo de honestidade e perseverança e estiveram ao meu lado durante toda realização do mestrado.

A minha esposa, Narinha, que nesses quase quatro anos de casamento sempre cuidou bem de mim, sempre me apoiou e fez um grande esforço para que eu conseguisse concluir o mestrado.

Ao meu irmão e melhor amigo, Rômulo, que desde sempre está ao meu lado me ajudando em todas as situações.

A minha grande amiga, Rosiene, que nunca mediu esforços para me ajudar em todos os momentos.

Ao Professor, João Francisco, que foi incansável na realização desse trabalho, sempre confiante, mesmos nos momentos mais difíceis.

A professora Elenice, que me ajudou no mestrado desde a época das disciplinas isoladas.

Ao amigo, Jaime, que me ajudou por diversas vezes, principalmente nos meus primeiros passos na linguagem C.

A minha amiga Mônica Lana, que me incentivou a entrar para o Mestrado e sempre me apoiou com palavras de incentivo.

A minha nova amiga, Adriana, que conheci na disciplina isolada e depois disso nunca mais parou de me ajudar.

Ao meu amigo, Marcos Dutra, que foi incansável em me ajudar, aconselhar e lutou até o fim para conseguir uma base de dados com um número grande de clientes.

Ao meu Amigo, Atacílio, que me informou sobre o Mestrado em Modelagem Matemática e Computacional do CEFET/MG.

A minha amiga, Rosemari, que nesse período torceu por mim e segurou todas as pontas na Escola Margarida Brochado.

Aos meus amigos e familiares, que sempre torceram por mim, me dando forças para ir até o final.

Ao Jurandir e a Cida, do Colégio Minas Gerais, que me apoiaram durante todo esse semestre.

Ao Sinaide, que fez a verificação final dos meus códigos, abriu mão de alguns sábados com a família para me ajudar programar em C.

A Marta, que me orientou nos meus primeiros contatos com o estudo das meta-heurísticas.

A Tatiana, que sempre estava a disposição para ajudar.

Ao Fábio Mourão, que sempre mostrou disposição em ajudar, adiou por mais de

uma vez, a sua volta para casa em Sete Lagoas para me ajudar no meu programa GA.

Ao Rodrigo Leite, que por diversas vezes parou suas atividades para me ajudar em linguagem C e no latex.

Ao Lucas Antônio, que me ajudou com a programação, inclusive me recebendo em sua casa em Itabirito.

Ao Reginaldo Heidler, que ajudou nos testes estatísticos.

Ao professor Sérgio Ricardo, que sempre foi prestativo e atencioso nas diversas vezes que precisei da sua ajuda.

Aos funcionários da secretaria do Mestrado, que sempre foram gentis e eficientes.

Aos colegas da sala de estudo do Mestrado, onde aprendi muitas coisas e conheci muitas pessoas especiais.

A todos vocês meus sinceros agradecimentos.

Resumo

A concessão de crédito a pessoas físicas por parte de agentes financeiros exige uma avaliação do perfil de cada um dos “candidatos a cliente”. Tais agentes necessitam de ferramentas com as quais possam identificar e conceder empréstimos a clientes potencialmente bons pagadores, recusando-o a clientes potencialmente maus pagadores. Como vem crescendo o número de agentes financeiros, autorizados ou não, a efetuar essa modalidade de transação a concorrência entre eles pelo “cliente em potencial” tem se acirrado.

Uma das ferramentas para detecção de um bom cliente e sua imediata aceitação consiste no emprego dos modelos Credit Scoring (Avaliação Objetiva para Concessão de Crédito), que possibilitam, rapidamente, a classificação deste em “bom” ou “mau” cliente.

No presente trabalho, desenvolve-se uma análise comparativa entre os resultados obtidos pelas metaheurísticas Otimização por Enxame de Partículas (Particle Swarm Optimization - PSO) e Algoritmos Genéticos (Genetic Algorithm - GA) na elaboração de um modelo de Credit Scoring.

A metodologia empregada baseia-se na aplicação dessas metaheurísticas sobre um banco de dados com o histórico de clientes, de forma a se definir pela concessão ou não do crédito pretendido. Os testes computacionais realizados tomaram por base três instâncias, para quais, as soluções encontradas com o emprego da metaheurística PSO foram superiores às soluções encontradas com a metaheurística GA.

Palavras-Chave: Modelos de Credit Scoring, Metaheurísticas, Otimização por Enxame de Partículas, Algoritmos Genéticos.

Abstract

The granting of credit to individuals by financial agents requires an assessment of the profile of each of the “prospective client”. Such agents need tools with which they can identify and lend to customers, potentially good payers, denying it to potentially bad payers. As the number of financial agents, authorized or not, making this type of transaction is growing, the competition between them for the potential customer is getting tougher.

One of the tools for the detection of a good customer and his immediate acceptance is based on the use of Credit Scoring models (Objective Evaluation for Granting of Credit), which rapidly provides his classification as a “good” or a “bad” customer.

The present study develops a comparative analysis between the results obtained by the metaheuristic Particle Swarm Optimization - PSO and Genetic Algorithm - GA to design a credit scoring model.

The methodology is based on the application of metaheuristics on a database with a history of customers in order to decide whether or not to grant the desired credit.

The performed computational tests were based on three instances and the solutions found with the PSO metaheuristic were better than those found with the GA metaheuristic.

Keywords: Credit Scoring Models, Metaheuristics, Particle Swarm Optimization, Genetic Algorithms.

Sumário

1	Introdução	1
1.1	Preliminares	1
1.2	Objetivos	2
1.2.1	Objetivo Geral	2
1.2.2	Objetivos Específicos	2
1.2.3	Justificativa para o Trabalho	3
1.2.4	Metodologia Proposta	3
1.3	Organização do Trabalho	3
2	Referencial Teórico	5
2.1	Introdução	5
2.2	Modelos de Credit Scoring	5
2.3	Risco de Crédito	8
2.4	Histórico de Clientes	12
2.4.1	Base de dados I	12
2.4.2	Base de dados II	13
2.4.3	Base de dados III	13
2.5	Modelos de Classificação	15
2.6	Considerações Finais	18
3	Técnicas Metaheurísticas	19
3.1	Introdução	19
3.2	Justificativa do uso das Metaheurísticas	21
3.3	Otimização por Enxame de Partículas - PSO	23
3.3.1	Otimização por Enxame de Partículas Contínuo	24
3.3.2	Otimização por Enxame de Partículas Discreta	28
3.4	Algoritmo Genético - GA	29
3.5	Considerações Finais	33

4	PSO e GA Aplicados na Elaboração de um Modelo de Credit Scoring	34
4.1	Introdução	34
4.2	A Função Objetivo	34
4.2.1	A Função de Avaliação	35
4.3	Preparação das bases de dados	36
4.3.1	Base de dados I	38
4.3.2	Base de dados II	39
4.3.3	Base de dados III	40
4.4	O emprego da Metaheurística PSO	41
4.5	O emprego da Metaheurística GA	44
4.6	Teste de Hipótese	46
4.7	Considerações Finais	48
5	Resultados e Discussões	49
5.1	Introdução	49
5.2	Ajuste dos Parâmetros dos Algoritmos	49
5.2.1	Ajuste dos Parâmetros PSO	50
5.2.2	Ajuste dos Parâmetros GA	50
5.3	Resultados Computacionais PSO e GA	51
5.3.1	Resultados - base de dados I	52
5.3.2	Resultados - base de dados II	55
5.3.3	Resultados - base de dados III	59
5.3.4	Considerações Finais	62
6	Conclusão e Trabalhos Futuros	63
6.1	Conclusão	63
6.2	Trabalhos futuros	64
	Referências	65
A	Os Melhores pesos obtidos na validação dos modelos de credit scoring	70

Lista de Tabelas

3.1	Técnicas utilizadas em modelos de credit scoring	22
4.1	Variável contínua discretizada por decis e risco relativo	38
4.2	Carteira de Clientes Binária	43
5.1	Parâmetros do algoritmo PSO	50
5.2	Parâmetros do algoritmo GA	51
5.3	Treinamento aplicando o Método 1 - Base de Dados I	52
5.4	Validação aplicando o o Método 1 - Base de Dados I	52
5.5	Treinamento aplicando o Método 2 - Base de Dados I	53
5.6	Validação aplicando o Método 2 - Base de Dados I	53
5.7	Comparação aplicando o Método 2 - Base de Dados I	54
5.8	Teste de Hipótese GA - Validação base de dados I	55
5.9	Teste de Hipótese PSO - Validação base de dados I	55
5.10	Treinamento aplicando o Método 1 - Base de Dados II	55
5.11	Validação aplicando o Método 1 - Base de Dados II	56
5.12	Treinamento aplicando o Método 2 - Base de Dados II	57
5.13	Validação aplicando o Método 2 - Base de Dados II	57
5.14	Comparação aplicando o Método 2 - Base de Dados II	58
5.15	Teste de Hipótese GA - Validação base de dados II	59
5.16	Teste de Hipótese PSO - Validação base de dados II	59
5.17	Treinamento aplicando o o Método 2 - Base de Dados III	60
5.18	Validação aplicando o Método 2 - Base de Dados III	60
5.19	Comparação aplicando o Método 2 - Base de Dados III	61
5.20	Teste de Hipótese - Validação base de dados III	61
A.1	Pesos dos melhores resultados PSO e GA - base de dados I - Método 1	70
A.2	Pesos dos melhores resultados PSO e GA - base de dados I - Método 2	71
A.3	Pesos dos melhores resultados PSO e GA - base de dados II - Método 1	71
A.4	Pesos dos melhores resultados PSO e GA - base de dados II - Método 2	72
A.5	Pesos dos melhores resultados PSO e GA - base de dados III - Método 1	72

Lista de Figuras

2.1	Diagrama de um modelo de Credit Scoring, segundo Gonçalves (2005).	7
2.2	Crédito em Relação ao PIB, segundo Dutra (2008).	8
2.3	Variáveis e atributos de um histórico de clientes (base de dados II)	11
2.4	Ponto de corte, segundo Santos e Famá (2007)	11
2.5	Classificação para o risco de crédito	16
3.1	Partícula do PSO em ação, segundo Prado e Saramago (2005).	25
3.2	Esboço do algoritmo PSO	27
3.3	Pseudocódigo Algoritmo Básico PSO.	28
3.4	Estrutura básica do Algoritmo Genético, segundo Silva e Oliveira (2004).	30
3.5	Pseudocódigo Algoritmo Básico GA.	31
3.6	A estrutura de um cromossomo binário.	31
3.7	Probabilidade de seleção de um cromossomo.	32
3.8	Processo de cruzamento de um cromossomo binário.	32
3.9	Processo de Mutação de um cromossomo binário	33
4.1	Esquema básico da elaboração de um modelo de credit scoring.	36
4.2	Código binário para classificação de um cliente	37
4.3	Descriminação das variáveis da base I.	39
4.4	Variáveis categorizadas.	40
4.5	Diagrama Elaboração de um Modelo de Credit Scoring.	41
4.6	Partículas do PSO	42
4.7	Pseudocódigo Algoritmo PSO aplicado ao modelo de credit scoring.	42
4.8	Matriz binária, vetores peso e score	43
4.9	Vetores score e resultado	43
4.10	Cromossomos do GA	44
4.11	Crossover Uniforme.	45
4.12	Pseudocódigo Algoritmo GA aplicado ao modelo de credit scoring.	46
5.1	Conjunto de Validação - Base de dados I.	54

5.2	Conjunto de Validação - Base de dados II.	58
5.3	Conjunto de Validação - Base de dados III.	61

Capítulo 1

Introdução

1.1 Preliminares

A concessão de crédito para pessoas físicas e jurídicas é prática muito comum nas sociedades capitalistas. Esse procedimento aquece a economia podendo influenciar na geração de empregos, no aumento do PIB, nas exportações e importações, entre outros.

Os modelos de avaliação objetiva para concessão de crédito são denominados: *Credit Scoring*. O presente trabalho trata o tema de concessão de crédito sempre como *credit scoring*.

O chamado problema de credit scoring é, basicamente, a busca por um modelo eficiente na avaliação para a concessão do crédito, e na avaliação do risco de crédito. Por esse motivo têm-se aplicado diversas técnicas para a elaboração de um modelo de credit scoring.

Segundo Arnaud et al. (2005), modelos de credit scoring baseados em análise discriminante foram os pioneiros e ainda continuam sendo utilizados. Mas, esse método tem como desvantagem a necessidade de considerar algumas suposições a respeito da base de dados, tais como, linearidade e distribuição normal.

Geralmente, os modelos de credit scoring são elaborados a partir de um histórico de clientes. Posteriormente, esse modelo é empregado na tomada de decisão a respeito de um candidato a cliente.

Gonçalves e Gouvêa (2006) destacam algumas das diversas técnicas empregadas na obtenção de um modelo de credit scoring:

- Regressão Logística
- Árvore de Classificação
- Redes Neurais

- Algoritmos Genéticos
- Análise Discriminante

No presente trabalho utilizou-se as metaheurísticas Otimização por Enxame de Partículas (Particle Swarm Optimization - PSO) e Algoritmos Genéticos (Genetic Algorithm - GA) na obtenção do modelo de credit scoring.

Essas técnicas foram aplicadas em três bases de dados (instâncias) e os resultados foram analisados e comparados com os resultados de outros autores. Observou-se que o emprego das metaheurísticas PSO e GA na construção de um modelo de credit scoring foi satisfatório.

1.2 Objetivos

1.2.1 Objetivo Geral

O presente trabalho tem como objetivo geral a aplicação das metaheurísticas - Otimização Por Enxame de Partículas (PSO) e Algoritmos Genéticos (GA) - na elaboração de um Modelo de Credit Scoring.

1.2.2 Objetivos Específicos

O presente trabalho tem como objetivos específicos:

- Verificar a relevância da aplicação de técnicas metaheurísticas em modelos de credit scoring;
- Analisar o comportamento das Metaheurísticas PSO e GA na elaboração de um modelo de credit scoring;
- Aplicar as metaheurísticas PSO e GA sobre as bases de dados, que foram, anteriormente, solucionadas com o emprego da Análise Discriminante e dos Algoritmos Genéticos, com o objetivo de comparar seus desempenhos;
- Aplicar as metaheurísticas PSO e GA em uma base de dados com grande número de clientes para verificar a eficiência dessas técnicas.

1.2.3 Justificativa para o Trabalho

A concessão de crédito é um procedimento muito utilizado em todo o mundo. No Brasil, essa prática é cada vez mais empregada, e seu impacto na economia é altamente significativo. Os modelos de credit scoring têm auxiliado as instituições financeiras a fornecerem, de forma mais eficiente, crédito às pessoas e empresas.

As instituições financeiras, ao oferecerem crédito a pessoas físicas, necessitam conhecer o risco de crédito envolvido em cada transação (risco de uma futura inadimplência). Por isso, essas instituições têm buscado desenvolver modelos de credit scoring que informem, com boa precisão, se um cliente poderá ou não ser aceito em sua carteira para o crédito.

Diversas técnicas são utilizadas em modelos de credit scoring, e no presente trabalho, utilizou-se o PSO e o GA, uma vez que essas metaheurísticas se mostraram eficazes em diversos outros problemas de otimização.

As metaheurísticas PSO e GA são métodos bem estabelecidos, ainda pouco utilizados na estimação dos modelos de credit scoring.

A elaboração de modelos de credit scoring com o emprego das metaheurísticas PSO e GA, apresentada nesse trabalho, visa comparar o desempenho destes com resultados obtidos por Dutra (2008) e Machado e Neto (2007).

1.2.4 Metodologia Proposta

Procurando atingir os objetivos apresentados, os algoritmos PSO e GA são implementados para a elaboração de um modelo de credit scoring.

As metaheurísticas PSO e GA são empregadas nas bases de dados utilizadas por Dutra (2008), por Machado e Neto (2007), e com uma base de dados de 5000 clientes, advinda de uma instituição financeira sediada em Belo Horizonte, Minas Gerais.

Os resultados são obtidos em duas etapas, o treinamento, onde são construídos os modelos, e a validação, onde os modelos são validados.

1.3 Organização do Trabalho

A estrutura do presente trabalho apresenta no capítulo 2, uma descrição do problema de credit scoring, destacando o histórico dos modelos de credit scoring, suas características e aplicações.

No capítulo 3 é feita uma revisão bibliográfica das metaheurísticas em geral e, posteriormente, uma apresentação das metaheurísticas PSO e AG. É feita uma

revisão bibliográfica das técnicas aplicadas ao problema de credit scoring.

No capítulo 4 são apresentadas as técnicas de preparação utilizadas na elaboração do modelo. Em seguida são apresentadas as adaptações das metaheurísticas PSO e AG na elaboração de um modelo de credit scoring.

Os resultados obtidos são mostrados no capítulo 5, onde são feitas as comparações das metaheurísticas PSO e GA com outras técnicas.

No capítulo 6, encontra-se a conclusão do trabalho e algumas propostas para trabalhos futuros.

Capítulo 2

Referencial Teórico

2.1 Introdução

A prática de concessão de crédito é cada vez mais presente na sociedade contemporânea. A sua utilização está diretamente ligada às relações financeiras do mundo capitalista. Essa prática vem sendo cada vez mais precedida de avaliações por intermédio de modelos de credit scoring, cujo objetivo é conceder o crédito com o menor risco de inadimplência possível.

Segundo Gonçalves (2005), os modelos de credit scoring baseiam-se em dados históricos de clientes de modo a avaliar se um possível futuro cliente será inadimplente. A implementação dessa espécie de modelo torna a avaliação de crédito “online”.

A possibilidade de que uma instituição credora venha a não receber os valores emprestados aos clientes é chamada de risco de crédito. O objetivo da sondagem criteriosa dos dados dos candidatos a crédito é minimizar esse risco, embora não exista um método que possa ser considerado infalível.

Devido à grande concorrência entre as empresas de crédito, existe uma imperiosa necessidade em se reduzir ao máximo o risco de cada uma das operações. Neste contexto, se inscreve a construção dos modelos de credit scoring, nos quais a influência advinda de critérios subjetivos é extremamente reduzida.

2.2 Modelos de Credit Scoring

Os modelos de credit scoring surgiram nos Estados Unidos na primeira metade do século XX. De acordo com Gonçalves (2005), esses modelos difundiram rapidamente pela rede bancária norte-americana. Esses modelos baseiam-se na construção de um vetor de ponderação de cada uma das informações dos clientes (categorias), e no

sistema de pontuação (escore) para a classificação final desse cliente.

No Brasil, o surgimento dos primeiros modelos de credit scoring deram-se no início dos anos 80. Entretanto, somente em meados dos anos 90, com a estabilização da moeda, esses modelos se popularizaram, atingindo, com grande impacto, o mercado financeiro brasileiro.

De acordo com Gonçalves e Gouvêa (2006), após a estabilização da moeda com o Plano Real, surgiram no mercado brasileiro alternativas de investimentos em substituição à rentabilidade do período inflacionário. Em consequência disso, as instituições financeiras aumentaram suas carteiras de crédito pessoal. Entretanto, para liberar o crédito, essas instituições careciam de um sistema avaliativo do candidato, via proposta preenchida com dados pessoais, enviada ao analista de crédito. Esse processo exigia alguns dias para a tomada de decisão, com isso, um candidato a cliente que tardiamente tenha sido avaliado como “bom” pode não retornar à empresa para a conclusão desse contrato. Nesse cenário é que surgem os modelos de concessão de crédito conhecidos como credit scoring.

Segundo Dutra (2008), os modelos de credit scoring tiveram seu emprego difundido com a chegada dos cartões de crédito ao mercado, fato que aumentou consideravelmente o número de clientes, gerando assim a necessidade da automatização do processo. Com o sucesso dos modelos de credit scoring empregados na concessão de cartões de crédito, a partir da década de 80, esses modelos foram também aplicados a outros produtos, como por exemplo, o crédito pessoal.

De acordo com Cheng et al. (2006), a classificação de um cliente por escore tornou-se o método mais empregado para a avaliação de crédito de pessoas físicas em todo o Reino Unido, nas décadas de 80 e 90. Machado e Neto (2007) ressaltam que a maioria dos métodos quantitativos utilizam um sistema de escore (pontuação) e, por esse motivo, são chamados de modelos de credit scoring.

Modelos de análise de crédito (credit scoring) foram adotados com o objetivo de acelerar a avaliação das propostas. Baseiam-se em dados históricos de clientes e avaliam as possibilidades de cada um desses vir a ser um “bom” ou “mau” pagador. Com os modelos de credit scoring, a avaliação de crédito passou a ser efetuada quase que imediatamente.

A figura 2.1 apresenta um modelo de credit scoring aplicado em uma base histórica de clientes.

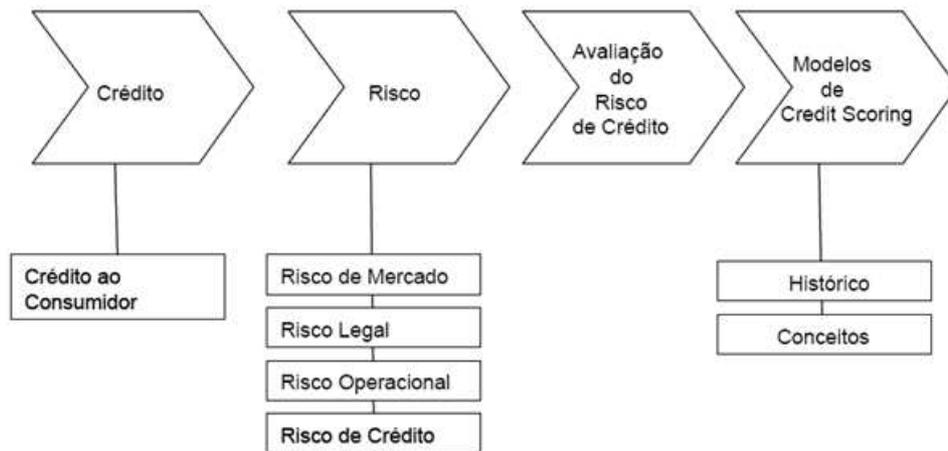


Figura 2.1: Diagrama de um modelo de Credit Scoring, segundo Gonçalves (2005).

Para a elaboração de um modelo de credit scoring os bancos se baseiam em informações cadastrais, atribuindo-lhes pesos ou ponderações, de acordo com a relevância na sua política de crédito. O resultado final do modelo de credit scoring é um sistema de pontuação para avaliar se o cliente será ou não aprovado para a concessão de crédito.

Segundo Santos e Famá (2007), credit scoring é todo modelo de avaliação de crédito que se baseia em uma formulação estatística com base nos dados cadastrais, financeiros, patrimoniais e de idoneidade dos clientes.

Para Dantas e Souza (2008), no Brasil existe uma demanda reprimida por crédito. Um problema preponderante que segura a expansão do crédito refere-se ao ambiente de informação assimétrica do setor bancário. As informações da situação financeira dos clientes, “boa” ou “ruim”, geralmente são de caráter privado, elevando o risco de crédito.

De acordo com Dutra (2008), o crédito para pessoa física no Brasil vem crescendo nos últimos anos. Em 1999, representava 35% do total de crédito concedido. No ano de 2006 esse volume passou para 47%. Os dados do Banco Central mostram que no período entre 1994 e 2007 a participação do crédito no Produto Interno Bruto (PIB) Brasileiro atingiu no máximo 36,5%, conforme observado na figura 2.2. Esse valor é considerado baixo se comparado com os 70%, 120% e 144% de Chile, Alemanha e Estados Unidos respectivamente.

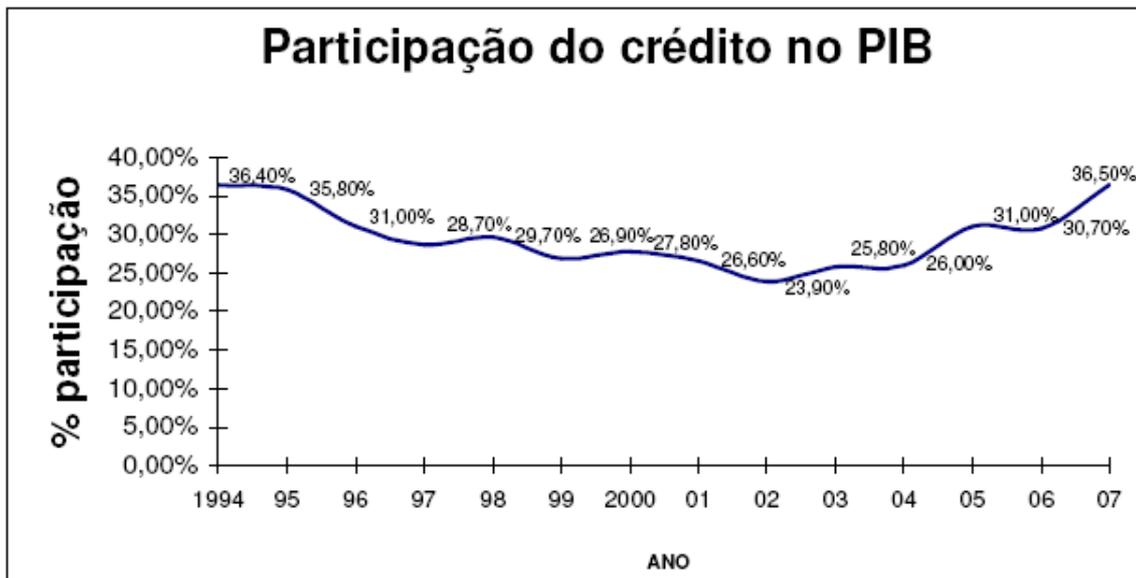


Figura 2.2: Crédito em Relação ao PIB, segundo Dutra (2008).

A constituição de modelos de credit scoring, segundo Gonçalves e Gouvêa (2006), deve seguir as sete etapas a seguir:

1. Levantamento de uma base histórica de clientes.;
2. Classificação dos clientes de acordo com o padrão de comportamento e definição da variável resposta;
3. Seleção de amostra aleatória representativa da base histórica;
4. Análise descritiva e preparação dos dados;
5. Escolha e aplicação das técnicas a serem utilizadas para a construção do modelo;
6. Definição dos critérios de comparação dos modelos;
7. Seleção e implementação do melhor modelo.

2.3 Risco de Crédito

Credit scoring pode ser definido como um método capaz de avaliar o risco de crédito. São utilizados nesse método informações pessoais dos proponentes, os dados históricos de clientes, além do auxílio de técnicas estatísticas (Corrêa e Vellasco, 2008). A avaliação do risco na concessão de crédito é fundamental para que as empresas consigam aumentar a porcentagem de bons clientes em sua carteira de empréstimo.

Segundo Godói et al. (2008), o risco de crédito pode ser definido como o risco que a instituição financeira credora está exposta no caso de uma inadimplência. Apesar da dimensão do problema de risco de crédito, apenas recentemente é que surgiram novas técnicas para auxiliar as instituições financeiras no tratamento desse problema.

Em várias empresas, o risco de crédito é, ainda hoje, efetuado por um profissional de análise de crédito. A partir das informações contidas em um questionário, é responsabilidade desse profissional a tomada de decisão sobre conceder ou não o crédito.

De acordo com Mendes Filho et al. (2005), a avaliação do risco de crédito efetuada por um profissional de análise é resultado de uma avaliação guiada mais por fatores subjetivos - qualitativos - que objetivos - quantitativos. Os possíveis problemas na utilização desse método de concessão de crédito são:

- Baixa eficiência - tempo demorado e necessidade de especialistas;
- Baixa performance - mesmo o emprego de especialistas experientes não permite a determinação da eficiência desse método;
- Pouca flexibilidade - O método é muito sensível a mudanças;
- Inconsistência - especialistas utilizam diferentes critérios para a avaliação.

Segundo Steiner et al. (2007), obtem-se vantagens com o correto uso de uma ferramenta para a concessão de crédito. Pode-se destacar entre as vantagens, o menor número de pessoas na análise de crédito, a agilidade no processamento, a diminuição da subjetividade na tomada de decisão e a redução do risco.

A prática de concessão de crédito, por natureza, envolve sempre algum tipo de risco. Não é possível predizer com absoluta certeza se o empréstimo será quitado. Cada empresa tem uma política específica para tratamento do risco, ao conceder um empréstimo.

No ato da concessão de crédito pessoal, o cliente responde a um questionário elaborado pela instituição financeira. Por meio da análise desse questionário as empresas definem o risco de crédito da operação. As informações solicitadas nesse questionário compõe-se de sexo, idade, profissão, região que reside, aluguel, idade na data de admissão, idade atual, quantidades de contrato na instituição, renda total, dentre outros.

As informações coletadas junto ao candidato a cliente são empregadas pelos modelos de credit scoring na classificação final deste como “bom” ou “mau”. Segundo Chaia (2003) e Queiroz (2006), existem 5 fatores que ajudam a determinar se o crédito será concedido ou não:

1. caráter: refere-se na intenção do cliente vir a efetuar o pagamento, está ligado a tradição de pagamento do cliente em compromissos anteriores;
2. capacidade: refere-se ao desempenho econômico e à geração de recursos para o pagamento, geralmente é associado a renda e aplicações financeiras dos clientes;
3. capital: refere-se à condição financeira e posse de bens;
4. condições: refere-se aos eventos externos, macroeconômicos e suas consequências, normalmente está ligado às políticas governamentais e a situação política financeira, nacional e internacional;
5. colateral: refere-se às garantias colocadas à disposição da empresa pelo cliente, em geral as garantias são maiores que o valor da dívida.

Com níveis elevados de inadimplência nas aplicações de crédito ao varejo, as instituições financeiras foram em busca de sistemas mais apurados para a gerência de risco (Mendes Filho et al., 2005).

Segundo Huang et al. (2005), as vantagens em se utilizar modelos de credit scoring para concessão de crédito são a diminuição do risco e a rapidez do processo.

A utilização de modelos de credit scoring está diretamente associada à avaliação do risco de crédito. O que se consegue com essa técnica é otimizar o tempo de decisão para conceder ou não o crédito e simultaneamente minimizar os riscos.

Para Yang (2007), credit scoring é um método que modela o risco na concessão de crédito. São utilizadas técnicas estatísticas e pontuações pelas instituições financeiras para a avaliação dos clientes em relação ao risco de crédito.

De acordo com Santos e Famá (2007), modelos de credit scoring são baseados em fórmulas estatísticas desenvolvidas a partir dos dados cadastrais dos clientes. Em geral, os modelos de credit scoring apresentam como resultado, pesos que são associados às variáveis do problema e suas categorias. Os pesos têm a função de ponderar cada variável de acordo com o histórico de clientes.

A figura 2.3 apresenta as variáveis e os atributos da base de dados II. As variáveis são discretas ou contínuas, a variável A1 por exemplo pode ser relativa ao sexo do cliente, então possui dois atributos, *a* e *b*, masculino e feminino respectivamente. As variáveis contínuas estão associadas ao quociente de duas informações cadastrais, por exemplo, a idade na data de admissão no emprego pela idade atual do cliente. Na base de dados II os nomes das variáveis não são informados.

Variável	Atributos
A1	b, a
A2	contínua
A3	contínua
A4	u, y, l, t
A5	g, p, gg
A6	c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff
A7	v, h, bb, j, n, z, dd, ff, o
A8	contínua
A9	t, f
A10	t, f
A11	contínua
A12	t, f
A13	g, p, s
A14	contínua
A15	contínua
A16	(+) "bom", (-) "mau"

Figura 2.3: Variáveis e atributos de um histórico de clientes (base de dados II)

Segundo Gonçalves e Gouvêa (2006), as variáveis contínuas são categorizadas de acordo com os seus decis, e, a cada faixa, é verificado o número total de clientes “bons” e “maus”. Decil é um termo utilizado na estatística que divide os valores de uma série em dez partes de mesma frequência. A razão entre o total de “bons” e “maus” clientes é denominada por *risco relativo*, que é utilizado para a categorização da variável. Esse procedimento encontra-se explicado detalhadamente no capítulo 4 do presente trabalho.

Santos e Famá (2007), analisando problemas de credit scoring, verificaram que a soma das pontuações de todos os clientes de uma carteira define a pontuação mínima ou o ponto de corte. As pontuações abaixo desse valor terão o crédito negado e acima dele, aprovado. Para valores muito próximos do ponto de corte, tanto abaixo quanto acima, admite-se um intervalo de dúvida, conforme mostrado na figura 2.4.

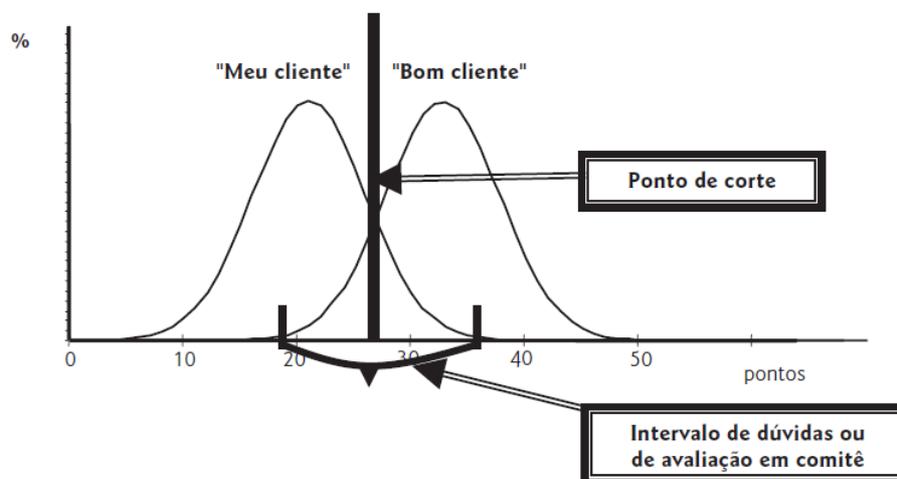


Figura 2.4: Ponto de corte, segundo Santos e Famá (2007)

Segundo Dantas e Souza (2008), avaliar o risco de crédito é verificar a probabilidade de uma futura inadimplência. A decisão da concessão ou não do crédito é tomada após o cálculo do escore de um cliente por um modelo que define um ponto de corte.

Em geral, as pontuações pertencentes à faixa de dúvida indicam que o cliente deverá ser avaliado por um analista de crédito.

2.4 Histórico de Clientes

No presente trabalho foram inicialmente utilizadas duas instâncias da literatura, denominadas base de dados I e II. Essas instâncias foram empregadas por Dutra (2008) e por Machado e Neto (2007).

As comparações dos resultados obtidos por Dutra (2008) e Machado e Neto (2007) estão descritas no capítulo 6 do presente trabalho.

Com o objetivo de verificar a eficiência das metaheurísticas PSO e GA em uma grande cartela de clientes, o presente trabalho utilizou também uma terceira instância, denominada base de dados III, que é composta por uma cartela de 5000 clientes de uma instituição financeira de Belo Horizonte.

2.4.1 Base de dados I

Dutra (2008), em seu trabalho, aplicou a técnica de análise discriminante com a finalidade de obter um modelo de credit scoring. A base de dados empregada é composta por 1151 clientes, dentre os quais, 438 foram considerados “bons”, 466 foram considerados “maus” e 247 foram considerados “intermediários”.

O objetivo do autor foi criar um modelo capaz de classificar com uma boa predição os três grupos por ele propostos. O método aplicado por Dutra (2008) para elaborar e validar o modelo utilizou todos os dados do histórico de clientes, não reservando uma parte da base de dados para a validação.

Na construção do modelo Dutra (2008) utilizou 12 etapas:

1. Preparação da base de dados.
2. Análise discriminante de Fisher.
3. Aplicação do método de stepwise.
4. Verificação da normalidade multivariada.
5. Transformações de Box-Cox.

6. Influência local.
7. Reanálise de Dados.
8. Leverage.
9. Dffits.
10. Teste de box.
11. Função discriminante quadrática.
12. Validação do modelo.

Dutra (2008) obteve com o seu modelo um percentual de acertos de 83,5% na classificação final dos clientes.

2.4.2 Base de dados II

Para a elaboração de um modelo de credit scoring, Machado e Neto (2007) aplicaram a técnica algoritmos genéticos. A base de dados utilizada conta com 653 clientes, dentre os quais, 296 considerados “bons” e 357 considerados “maus”. Para o cálculo do score de cada cliente foi utilizado o método proposto por Picinini et al. (2003).

O cálculo do score foi efetuado por intermédio de uma função objetivo, a qual tem por finalidade classificar cada um dos clientes como “bom” ou “mau”. Posteriormente, essa classificação é comparada com o histórico de clientes, verificando se cada cliente foi classificado corretamente.

Para elaborar e testar o modelo, Machado e Neto (2007) dividiram a base de dados em sete partes, reservando uma das partes para a validação, enquanto as outras eram utilizadas na elaboração do modelo. Com esse método, o algoritmo proposto pelos autores foi executado sete vezes, gerando as sete melhores soluções nas validações.

O resultado final obtido pelos autores foi 85,6% de percentual de acerto. Esse valor é a média dos sete melhores resultados encontrados nas validações.

A base de dados II está disponível em:<ftp://ftp.ics.uci.edu/pub/machine-learning-databases>.

2.4.3 Base de dados III

Uma base de dados deve ter o mesmo número de clientes “bons” e “maus”. Para se obter um resultado considerado robusto, nesse tipo de problema, é necessário que

a base de dados possua pelo menos 1500 clientes considerados “bons” e 1500 considerados “maus” (Gonçalves e Gouvêa, 2006). Devido a essa consideração, empregou-se uma base de dados com 5000 clientes, 2500 considerados “bons” e 2500 considerados “maus” - base de dados III.

Os dados coletados referem-se a operações de créditos concedidas a pessoas físicas, por um banco de Minas Gerais, no período de agosto a setembro de 2008. Os dados foram apurados até o mês de agosto de 2009. No período em que os empréstimos foram concedidos o mundo vivia uma grande crise financeira. Segundo Freitas (2009) a crise internacional atingiu a economia brasileira em um momento de auge. Rumores sobre perdas dos bancos geraram um clima de aversão ao risco de crédito, com isso os bancos diminuíram a concessão de crédito para as pessoas físicas e jurídicas.

A administração desse banco convencionou que créditos com atrasos de até 90 dias, seja possível a sua recuperação, sem que com isso, empresas terceirizadas pelo banco tenham que buscar abrir negociação com clientes, afim de que esses venham de alguma forma completar o seu pagamento. Os clientes que atingirem 91 dias de atraso são automaticamente enviados às assessorias de cobrança. Essas empresas são especializadas em recuperação de crédito (prestadoras de serviço) possuindo uma estrutura organizacional direcionada para atender a essa necessidade.

De acordo com a resolução 2682 de 21 de dezembro de 1999 do Banco Central do Brasil, o saldo devedor do cliente com atraso superior a 180 dias é totalmente provisionado. Provisão é um conceito contábil que corresponde a expectativa de perdas de ativos ou estimativas de valores a desembolsar. Neste caso é vista como expectativa de perda do ativo.

Segundo Fernandes et al. (2008), no sistema financeiro a provisão assume papel fundamental, pois quando corretamente mensurada, revela a qualidade da carteira de crédito, sendo fundamental para o gerenciamento do risco de crédito.

O Banco Central, autorizou com a resolução 2682 as instituições financeiras a classificarem as operações de crédito, em ordem crescente de risco, de acordo com os seguintes níveis:

1. Nível AA;
2. Nível A;
3. Nível B;
4. Nível C;
5. Nível D;

6. Nível E;
7. Nível F;
8. Nível G;
9. Nível H;

Sobre a classificação da operação nos níveis de risco citados anteriormente, deve ser observado mensalmente, por ocasião dos balancetes e balanços, em função de atraso observado na parcela de pagamento, da seguinte maneira:

- atraso entre 15 e 30 dias: risco nível B, no mínimo;
- atraso entre 31 e 60 dias: risco nível C, no mínimo;
- atraso entre 61 e 90 dias: risco nível D, no mínimo;
- atraso entre 91 e 120 dias: risco nível E, no mínimo;
- atraso entre 121 e 150 dias: risco nível F, no mínimo;
- atraso entre 151 e 180 dias: risco nível G, no mínimo;
- atraso superior a 180 dias: risco nível H;

Desta forma, considerou-se o cliente com o status “bom” quando o mesmo atingiu durante o financiamento um atraso inferior a 91 dias, e “mau” quando o cliente deixou de quitar suas obrigações por mais de 180 dias. Não fazem parte do banco de dados III os clientes que saudaram sua dívida entre 91 dias e 180 dias de atraso.

O número de parcelas para pagamento do crédito concedido, no caso dessa instituição foi, em média nove. Sendo assim, até agosto de 2009, tem-se um horizonte temporal suficiente para avaliar qual classificação esses créditos concedidos devem receber.

Dos clientes selecionados foram identificadas as características pessoais e dados econômico-financeiros relativo ao negócio ou atividade econômica dos mesmos, tanto dos clientes “bons” quanto dos “maus”.

2.5 Modelos de Classificação

Uma evolução dos cinco fatores que ajudam a determinar o crédito, são os modelos de classificação, conhecidos como *rating*. As instituições financeiras elaboraram esses

modelos de classificação com o objetivo de verificar o percentual de inadimplência e medir o total da perda esperada (Chaia, 2003).

Segundo Chaia (2003) e Crouhy et al. (2000), em geral se trabalha com dez categorias de classificação dos clientes, sendo, as seis primeiras de alta qualidade e as quatro últimas de baixa qualidade. A figura 2.5 apresenta um exemplo de tabela de classificação.

Classificação		Risco
1	AAA	Mínimo
2	AA	Modesto
3	A	Médio
4	BBB	Aceitável
5	BB	Aceitável com Cautela
6	B	Atenção
7	CCC	Menção Especial
8	CC	Abaixo do Padrão
9	C	Duvidoso
10	D	Prejuízo

Figura 2.5: Classificação para o risco de crédito

A seguir têm-se a descrição das categorias segundo Chaia (2003):

1. AAA: Excelente crédito comercial, qualidade de ativos superior, excelente capacidade de endividamento e cobertura; excelente gestão, com profundidade. A empresa é líder de mercado e tem acesso a mercados de capitais.
2. AA: Bom crédito comercial, qualidade de ativos e liquidez muito boas, forte capacidade de endividamento e cobertura, gestão muito boa em todos os cargos. A empresa goza de reputação muito boa no setor e tem uma fatia de mercado muito forte.
3. A: Crédito comercial médio, dentro dos padrões normais de crédito comercial; qualidade de ativos e liquidez satisfatórias, boa capacidade de endividamento e cobertura; boa gestão em todos os cargos críticos. Empresa de porte e posição médios no setor.
4. BBB: Crédito comercial aceitável, mas com risco maior que a média; qualidade de ativos aceitável, pequeno excesso de liquidez, capacidade de endividamento aceitável. Pode ou não ser altamente ou totalmente alavancada. Requer níveis acima da média de supervisão e atenção ao credor. A empresa não é forte

o bastante para suportar reveses importantes. Empréstimos são transações altamente alavancadas devido a restrições regulamentares.

5. BB: Crédito comercial aceitável, mas com risco considerável; qualidade de ativos aceitável, base de ativos menor e/ou menos diversificada, muito pouca liquidez, capacidade de endividamento limitada. Requer condições estruturadas para assegurar proteção adequada. Pode ou não ser altamente ou totalmente alavancada. Pode ser de porte abaixo da média ou concorrente de segunda linha. Requer supervisão e atenção significativas por parte do credor. A empresa não é forte bastante para suportar grandes reveses. Empréstimos são transações altamente alavancadas devido à situação financeira do devedor.
6. B: Crédito na watch list, merecedora de observação especial da gerência; qualidade de ativos aceitável de forma geral, liquidez um tanto forçada, totalmente alavancada. Alguma fraqueza de gestão. Requer supervisão e atenção contínuas por parte do credor.
7. CCC: Crédito comercial marginalmente aceitável; alguma fraqueza. Negócio de forma geral indesejável que constitui um risco de crédito desnecessário e indevido, mas não a ponto de justificar seu enquadramento como abaixo dos padrões. Embora o ativo esteja atualmente protegido, é potencialmente fraco. Não se prevê perda de juros ou de principal. Fraquezas potenciais poderiam incluir uma condição financeira enfraquecida; um programa de re-pagamento não-realista; fontes inadequadas de recursos financeiros ou falta de garantias reais, informações de crédito ou documentação adequadas. A empresa é medíocre.
8. CC: Crédito comercial inaceitável; re-pagamento normal em risco. Embora não se preveja perda de principal ou de juros, uma fraqueza clara e bem definida coloca em risco o recebimento da dívida. O ativo é inadequadamente protegido pelo atual patrimônio líquido tangível e pela capacidade de pagamento do devedor ou garantia real oferecida. Poderá já ter havido uma perda parcial de juros.
9. C: Re-pagamento total questionável. Existem problemas sérios que indicam a probabilidade de uma perda parcial do principal. As fraquezas são de tal forma pronunciadas que, com base em informações, condições e valores correntes, o recebimento é altamente improvável.
10. D: Perda total esperada. Um ativo incobrável ou de tão pouco valor que não justifica sua classificação como ativo efetivo. Tal ativo, entretanto, poderá ter

algum valor de recuperação, marginal mas, não o ponto em que uma baixa contábil seria postergável. mesmo que uma recuperação parcial possa ocorrer no futuro.

Essas classificações são utilizadas na prática para se medir o risco de crédito principalmente de pessoas jurídicas. O presente trabalho faz uma abordagem somente com modelos de credit scoring na concessão de crédito para pessoas físicas.

2.6 Considerações Finais

Na seção 2.1 deste capítulo é apresentada uma introdução ao problema de credit scoring; na seção 2.2, o modelo de credit scoring é detalhado; na seção 2.3 é efetuado um relato sobre risco de crédito; na seção 2.4, foi feita a verificação de como foram utilizadas as bases de dados I e II por outros autores. Em seguida, encontra-se descrita uma terceira base de dados - base de dados III; e na seção 2.5, foi apresentada os modelos de classificação, rating.

Capítulo 3

Técnicas Metaheurísticas

3.1 Introdução

O termo metaheurística foi utilizado por Glover (1986) e deriva dos radicais gregos: *heurisken* = encontrar; *meta* = além de. De acordo com Osman e Laporte (1996), a metaheurística pode ser definida como um processo de geração iterativo, responsável por gerar heurísticas subordinadas pelas combinações de vários conceitos inteligentes ao se explorar e expandir o espaço de busca. São utilizadas ainda estratégias de aprendizagem para a estrutura das informações com o objetivo de encontrar soluções próximas à solução ótima.

Segundo Souza (2008), as técnicas metaheurísticas apresentam caráter geral e possuem métodos capazes de buscar fugir de ótimos locais. Diferenciam-se umas das outras, basicamente, pela técnica empregada para escapar destes ótimos locais.

As metaheurísticas podem ser divididas em dois grupos:

1. Busca local
2. Busca populacional

Esses grupos se diferenciam pela técnica utilizada na exploração do espaço amostral.

Metaheurísticas baseadas na busca local caracterizam-se pela exploração do espaço de soluções por meio de movimentos, que são aplicados a cada passo sobre a solução corrente, gerando uma outra candidata a solução em sua vizinhança (Souza, 2008).

Os métodos baseados na busca populacional, por sua vez, consistem em manter um conjunto de boas soluções e combiná-las de forma a tentar produzir soluções ainda melhores (Souza, 2008).

De acordo com Bezerra (2008), os métodos que utilizam metaheurísticas apresentam boa solução em tempos computacionais aceitáveis, e, em consequência disso são amplamente utilizados em problemas de otimização.

As principais técnicas metaheurísticas são descritas por Souza (2008):

- Algoritmos meméticos - Os algoritmos meméticos são uma variante do GA. Essa técnica baseia-se no melhoramento dos indivíduos, antes de serem realizados o cruzamento e a mutação.
- Colônia de formigas - Colônia de formigas é uma metaheurística utilizada em problemas de otimização que consiste em simular o comportamento de um enxame de formigas, que cooperem entre si na busca de alimento, por meio do feromônio depositado por cada formiga, durante o seu deslocamento no espaço de busca. Essa prática permite às outras formigas seguirem o rastro deixado como informação na busca pelo seu objetivo.
- Multi-start - A metaheurística multi-start é um método considerado de fácil implementação e tem como característica principal realizar amostragens do espaço de busca, refinando cada solução encontrada para diversificar esse espaço e fugir de ótimos locais.
- Busca tabu - Na busca tabu, o método utilizado para explorar o espaço de busca consiste no movimento de uma solução para outra que seja a melhor num conjunto formado pelos vizinhos mais próximos. Essa técnica possui memória de armazenamento das melhores soluções. Os procedimentos realizados por ela evitam que fique presa em ótimos locais.
- Busca Local - A técnica de busca local define para cada solução encontrada uma vizinhança com soluções próximas. Dessa forma, é efetuada uma busca local com o objetivo de maximizar a solução. Caso não se obtenha sucesso, volta-se para a solução corrente e aplica-se o método novamente. Em caso contrário, a solução encontrada é considerada um ótimo local em relação à vizinhança.
- Busca em Vizinhança Variável - A técnica busca em vizinhança variável, se caracteriza por explorar o domínio em busca de vizinhos considerados distantes da solução atual. A busca em torno de uma nova solução é feita somente se for realizado um movimento de melhora. O método utilizado na exploração do espaço de busca consiste na troca sistemática de estruturas de vizinhanças.
- Simulated Annealing - Simulated annealing é uma metaheurística de busca local probabilística baseada no resfriamento de um conjunto de átomos aquecidos

(recozimento). Seu procedimento começa com uma solução inicial aleatória, a cada iteração um laço de repetição é executado gerando aleatoriamente um único vizinho da solução atual.

- Annealing Microcanônico - Annealing microcanônico é uma variação da metaheurística Simulated Annealing, que baseia-se na variação dos estados de energia constante.
- GRASP - GRASP é uma metaheurística que possui duas fases: 1) construção, quando é gerada uma solução elemento a elemento, e 2) busca local, é pesquisado o ótimo local na vizinhança da solução construída. O resultado será a melhor solução encontrada entre todas as iterações.
- Iterated Local Search - A metaheurística iterated local search (ILS) é um procedimento de busca local que promove perturbações na solução ótima local com a finalidade de gerar uma solução melhor. As perturbações do ILS devem ser fortes o suficiente para escapar de ótimos locais e explorar outras regiões do espaço de busca e, por outro lado, devem ser fracas o suficiente para armazenar as características do ótimo local atual.
- Guided Local Search - Guided local search se baseia na promoção de modificações na função custo com a finalidade de escapar de ótimos locais.
- Scatter Search - Scatter search é uma metaheurística populacional que se baseia na construção de soluções usando a combinação de outras soluções, tendo sua origem na programação inteira.

Para a elaboração de um modelo de credit scoring são utilizados neste trabalho as metaheurísticas evolutivas, Otimização por Enxame de Partículas (PSO) e Algoritmos Genéticos (GA).

3.2 Justificativa do uso das Metaheurísticas

Desde a pioneira análise discriminante até os algoritmos evolutivos, diversas técnicas foram empregadas para a construção de um modelo de credit scoring. Estas técnicas, têm contribuído cada vez mais para a diminuição do risco na concessão de crédito a pessoas físicas. Tendo em vista o caráter probabilístico da maioria desses modelos, não se pode determinar com exatidão qual o método mais eficaz no tratamento desse problema.

Segundo Dutra (2008), a análise discriminante apresenta com frequência, problemas de violações de pressuposições. Mesmo assim, essa técnica é ainda empregada com frequência no desenvolvimento de modelos de credit scoring.

Para Mendes Filho et al. (2005), na avaliação de risco de crédito, a maioria dos gerentes trabalha com uma grande quantidade de informações, advindas das mais diversas fontes. Além disso, muitas dessas fontes podem estar incompletas ou incorretas. Os gerentes, muitas vezes, não conseguem explicar o processo de tomada de decisões. Além disso, os ambientes são dinâmicos e exigem que decisões sejam tomadas rapidamente. As redes neurais bem como as técnicas metaheurísticas, oferecem uma ótima alternativa para o tratamento de informações dessa natureza, já que podem ser modificadas mais rapidamente do que os sistemas convencionais e são mais simples que os métodos convencionais aplicados na modelagem não linear.

A tabela 3.1 apresenta algumas técnicas utilizadas na elaboração de um modelo de credit scoring, dentre elas técnicas metaheurísticas.

Tabela 3.1: Técnicas utilizadas em modelos de credit scoring

Autores	Técnicas utilizadas	Número de clientes na base de dados
Dutra (2008)	Análise Discriminante	1151
Jiang e Yuan (2007)	1- PSO e Regressão Linear ; 2- PSO e Regressão Logística	1057
Machado e Neto (2007)	Algoritmos Genéticos	653
Gao et al. (2006)	1- PSO; 2- Algoritmos Genéticos; 3- Redes Neurais	736
Gonçalves e Gouvêa (2006)	1- Redes Neurais; 2- Algoritmos Genéticos	20000
Arnaud et al. (2005)	Redes Neurais	30000
Mendes Filho et al. (2005)	Redes Neurais	5635

As técnicas PSO e GA apresentam grande similaridade em sua aplicação. Tal fato dá sentido a um dos objetivos deste trabalho, comparar o desempenho dessas duas metaheurísticas na resolução desse tipo de problema.

Neste trabalho, a primeira técnica utilizada foi a otimização por enxame de partículas. Sendo o PSO uma metaheurística recém-desenvolvida, sua aplicação na construção de modelos de credit scoring ainda é reduzida.

Gao et al. (2006) utilizaram uma base de dados composta por 468 clientes “bons” e 268 “maus” para o desenvolvimento de modelos de credit scoring empregando as metaheurísticas PSO, GA e a Rede Neural Back-Propagation. Obtiveram nos testes de acurácia 85,3% de acertos com o PSO, 81,6% de acertos com os algoritmos genéticos e 84,7% com a rede neural.

Jiang e Yuan (2007) empregaram modelos de predição para desenvolver modelos de credit scoring. Baseados em regressão linear e regressão logística construíram um modelo combinado, usando otimização por enxame de partículas (PSO) para

pesquisar os pesos para cada um desses modelos. Obtiveram nos testes de acurácia 83,39% de acertos com a regressão linear - PSO e 91,67% de acertos com a regressão logística - PSO. Foram utilizadas duas bases de dados, uma com 500 clientes e a outra com 1057 clientes.

A segunda técnica utilizada neste trabalho foi o algoritmo genético (GA), que é uma metaheurística clássica, pertencente à computação evolutiva e empregada nos mais diversos problemas de otimização, inclusive na estimação dos modelos de credit scoring.

De acordo com Bezerra (2008), a cada iteração, os algoritmos PSO e GA buscam combinar e manter o conjunto das melhores soluções, fato relevante para problemas de otimização.

3.3 Otimização por Enxame de Partículas - PSO

A Otimização por Enxame de Partículas é uma técnica metaheurística que foi inicialmente desenvolvida por Kennedy e Eberhart (1995). Essa técnica baseia-se no comportamento social de várias espécies, tais como, bando de pássaros, enxame de abelhas e cardume de peixes. Segundo Reynolds (1987), na revoada de pássaros, os resultados adquiridos pelo bando estão ligados à interação entre todos os componentes. Neste trabalho, os termos, pássaro, abelha e peixe serão designados pelo termo “partícula”.

No PSO, em princípio, as partículas deslocam-se cada qual em uma direção, na busca de otimização de seu objetivo. Quando uma partícula localiza uma posição “melhor” do que as anteriores, essa informação passa a influenciar a direção e a velocidade de busca das demais. São registradas a melhor posição de cada partícula (*pbest*) e a melhor posição entre todas as partículas do enxame (*gbest*). Dessa forma, a metaheurística PSO emprega, da mesma forma que os pássaros, os peixes e as abelhas, a interação social na busca por um ninho, ou por uma fonte de alimento.

De acordo com Millonas (1994) cinco princípios básicos são necessários para desenvolver modelos de otimização bio inspirados. O PSO atende a todos, a saber:

1. Princípio da proximidade - capacidade do enxame responder através de seu comportamento às variações dos estímulos ambientais de tempo e espaço;
2. princípio da qualidade - capacidade do enxame responder aos fatores de qualidade do meio;
3. princípio da resposta diversa - capacidade do enxame em alocar seus recursos fora de áreas restritas, prevenindo perturbações do meio;

4. princípio da estabilidade - capacidade do enxame em manter seu comportamento quando ocorrer mudanças no meio;
5. princípio da adaptabilidade - capacidade do enxame em mudar a modalidade do seu comportamento quando essa atitude for benéfica.

O PSO divide-se em duas categorias, PSO Discreto e PSO Contínuo.

3.3.1 Otimização por Enxame de Partículas Contínuo

O PSO é uma metaheurística populacional. Inicialmente é gerado um conjunto aleatório de partículas $P(t)$, que são candidatas à solução do problema de otimização. Cada uma dessas partículas representa um ponto no espaço n-dimensional. Dessa forma, de acordo com Shi e Eberhart (1998a), têm-se que:

- $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$: partícula i , candidata a ser uma possível solução do problema, onde x_{ij} representa uma variável da solução;
- $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$: velocidade ou taxa de variação da posição da partícula i .
- $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$: melhor posição histórica obtida pela partícula i ou a posição com o melhor valor de aptidão já registrado até a presente iteração. Também é chamada de $pbest_i$. A melhor posição dentre todas as posições do enxame é denominada de $gbest$.
- $f(x)$: é a função de avaliação, que determina o valor de aptidão da partícula i .

O algoritmo PSO, proposto por Kennedy e Eberhart (1995), surgiu da experiência de algoritmos modeladores do comportamento social de bandos de pássaros. Esses pássaros, tratados aqui por partículas, durante a busca por alimento ou ninho, utilizam a sua experiência própria e a experiência coletiva do bando. O algoritmo PSO possui um vetor de velocidade e outro de posição. A velocidade de cada partícula é atualizada, e uma nova posição para cada uma das partículas é calculada a cada iteração.

As posições iniciais de cada partícula são determinadas aleatoriamente dentro do espaço de busca, e as velocidades iniciais podem ser aleatórias ou nulas.

O algoritmo PSO gera três matrizes, X, V e B e um vetor G. As matrizes X e V armazenam as posições e as velocidades de cada partícula, respectivamente. A cada iteração as matrizes são atualizadas e as melhores posições registradas por cada partícula até a presente iteração ($pbest$) são armazenadas na matriz B. A melhor posição entre todas as posições de B ($gbest$) é armazenada no vetor G.

No algoritmo PSO cada partícula é um ponto do espaço n-dimensional que ajusta o seu deslocamento baseado na própria experiência ou na experiência coletiva. Então, a cada iteração são atualizadas a posição e a velocidade de cada uma das partículas em todas as dimensões. Tal procedimento faz com que elas se desloquem gradualmente para os melhores valores históricos, $pbest$ e $gbest$.

A figura 3.1 mostra a movimentação de duas partículas no espaço bidimensional, segundo a metaheurística PSO. Em seu deslocamento, no espaço de busca, cada partícula registra a sua melhor posição, que pode ser alterada no transcorrer das iterações até uma outra posição, melhor que a anterior. O objetivo a ser alcançado é o deslocamento de todas as partículas em direção às melhores posições.

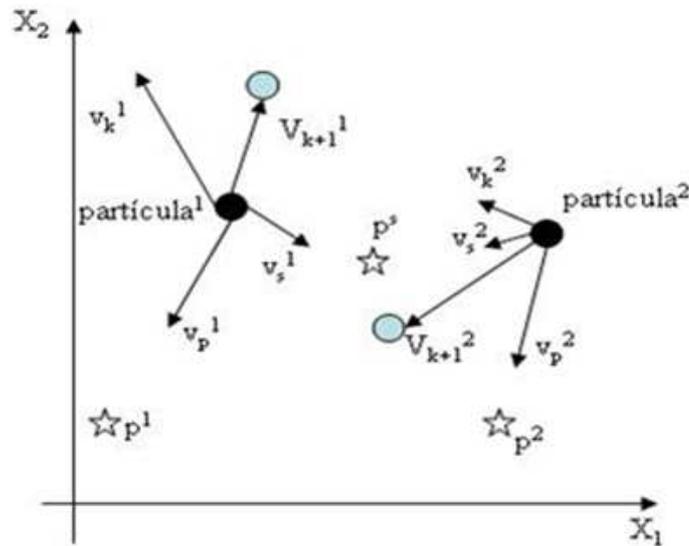


Figura 3.1: Partícula do PSO em ação, segundo Prado e Saramago (2005).

Os símbolos empregados na figura 3.1 representam:

- v_s^i = velocidade em direção ao ótimo atual do enxame;
- v_p^i = velocidade em direção ao ótimo atual da partícula i ;
- p^s = Posição ótima do enxame até a presente iteração;
- p^i = Posição ótima da partícula até a presente iteração;
- $particula^1$ e $particula^2$ = posições atuais;
- v_k^1 e v_k^2 = velocidades atuais;
- V_{k+1}^1 e V_{k+1}^2 = posições na próxima iteração.

Segundo Shi e Eberhart (1998a), a posição e a velocidade de cada uma das partículas, a cada iteração, são recalculadas utilizando-se as equações 3.1 e 3.2, respectivamente.

$$v_{in}^{t+1} = wv_{in}^t + c_1r_1(pbest_{in} - x_{in}^t) + c_2r_2(gbest_n - x_{in}^t) \quad (3.1)$$

$$x_{in}^{t+1} = x_{in}^t + v_{in}^{t+1} \quad (3.2)$$

Nessas expressões são empregados os seguintes termos:

- w : Peso inercial responsável por ponderar a influência da velocidade anterior no cálculo da velocidade atual;
- c_1 e c_2 : constantes positivas, parâmetros responsáveis por ponderar a influência sobre o comportamento cognitivo e social das partículas, respectivamente;
- r_1 e r_2 : números aleatórios distribuídos uniformemente no intervalo $[0-1]$;
- $t = 1, 2, \dots, iter_{max}$: número de iteração corrente, com o limite máximo igual a $iter_{max}$.
- i : índice da partícula que varia de 1 até p , onde p é o tamanho do enxame;
- n : dimensão da partícula.

Segundo Shi e Eberhart (1998a), na equação 3.1:

- O primeiro termo corresponde à velocidade prévia da partícula i , ponderado pelo peso inercial, que permite ampliar a diversidade do espaço de busca do problema. Valores maiores do peso inercial favorecem a busca global. Valores menores favorecem a busca local, relacionada com o refinamento da solução do problema;
- O segundo e o terceiro termos representam o saber individual (herança cognitiva) e a cooperação entre as partículas (herança social), respectivamente, e são utilizados para mudar a velocidade da partícula i . Sem estes termos, a partícula continuaria se deslocando na mesma direção até atingir a região de fronteira;
- Se o primeiro termo da expressão não for utilizado, a velocidade da partícula i durante o deslocamento seria determinada pela tendência de retornar às suas melhores posições, convergindo prematuramente.

A figura 3.2, apresenta um esboço do algoritmo PSO.

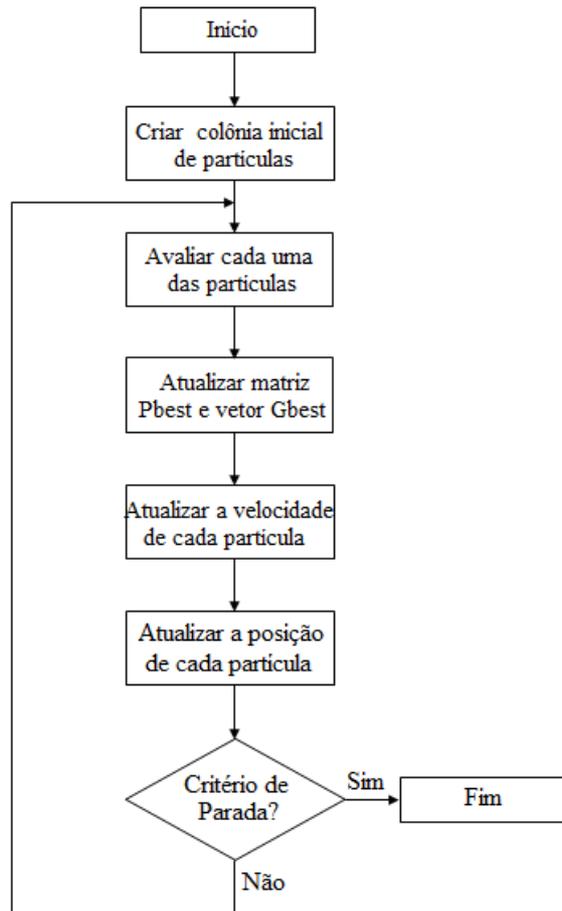


Figura 3.2: Esboço do algoritmo PSO

De acordo com Shi e Eberhart (1998a) e Suganthan (1999), os valores sugeridos para os parâmetros c_1 e c_2 são $c_1 = c_2 = 2$, mantendo o equilíbrio entre o saber individual e o saber coletivo no comportamento atual da partícula. Segundo Prado e Saramago (2005) alguns autores utilizam valores diferentes para c_1 e c_2 , mas deve-se sempre manter $c_1 + c_2 = 4$.

O parâmetro w , tem por objetivo ampliar o espaço de busca das partículas. Kennedy e Eberhart (1997), após uma extensa análise empírica, sugerem que o valor desse parâmetro deva estar contido no intervalo: $0,7 < w < 1,4$. Segundo Shi e Eberhart (1998b), o parâmetro w pode fornecer um equilíbrio ao algoritmo PSO em relação a sua capacidade local e global de exploração.

Verifica-se que o número de partículas utilizadas pelo algoritmo PSO, para problemas de otimização, varia de acordo com a natureza do problema proposto. Segundo Trelea (2003), quanto maior o número de partículas do enxame, menor será o número de iterações necessárias para se obter a melhor solução. Assim, deve-se buscar determinar um número de partículas que garanta uma exploração suficiente

do espaço amostral, sem que isso acarrete um aumento demasiado do custo computacional do algoritmo.

Após a geração aleatória da posição de cada partícula, inicia-se o processo iterativo que será executado até que um dos critérios de parada seja atingido. Esses critérios, em geral, são o número máximo de iterações e o número de iterações sem melhora da solução encontrada pelo algoritmo.

A aptidão de cada partícula do enxame é dada pela função objetivo, e o cálculo da aptidão é efetuado pela comparação entre a aptidão atual e a anterior. O algoritmo PSO atualiza a velocidade e posição de cada partícula i através das equações 3.1 e 3.2 respectivamente. A cada iteração, a melhor posição de cada partícula é atualizada na matriz B e a melhor posição do enxame é atualizada no vetor G .

A Figura 3.3 apresenta um pseudocódigo do algoritmo básico para o PSO, proposto por Suganthan (1999).

Algoritmo PSO
 Iniciar a posição e a velocidade das P partículas da população;
 Faça:
 Avaliar indivíduos através do cálculo da função objetivo: $f(x)$;
 Atualizar $pbest_i$; Se $f(x_i) > f(pbest_i)$
 então $pbest_i = x_i$;
 Atualizar $gbest$; Se $pbest_i > f(gbest)$
 então $gbest = pbest_i$;
 Atualizar a velocidade de cada partícula, conforme equação (3.1);
 Atualizar a posição de cada partícula, conforme equação (3.2);
 Enquanto (critério de parada não atingido)
Fim Algoritmo PSO.

Figura 3.3: Pseudocódigo Algoritmo Básico PSO.

3.3.2 Otimização por Enxame de Partículas Discreta

A Otimização Discreta por Enxame de Partículas (DPSO) foi proposta por Kennedy e Eberhart (1997) para tratar problemas de ordenação e arranjo de elementos discretos e problemas combinatórios. No espaço de busca discreto, uma partícula do enxame pode ocupar um dos vértices de um hiper-cubo, movendo-se suavemente através do ajuste de trajetórias de suas partículas que estão associadas na manipulação de suas coordenadas. O Algoritmo DPSO mantém a estrutura básica do PSO contínuo.

Outra variação do PSO contínuo para forma discreta foi apresentada por García e Pérez (2007) como Otimização por Saltos de Rãs (Jumping Frog Optimization - JFO) e tem sua origem no movimento de um grupo de rãs que saltam de uma

pedra a outra em um determinado intervalo de tempo. Nesse método, as partículas se locomovem no espaço de busca saltando de uma solução para outra. O JFO mantém a estrutura básica do PSO contínuo, guardando suas melhores posições individuais e coletivas, apresentando mudança na atualização da posição uma vez que nesse método a velocidade não é considerada.

Se o espaço de busca do problema de otimização for discreto é necessário a utilização de métodos como o DPSO e JFO, que são mais adequados para esse tipo de situação. O presente trabalho apresenta um domínio contínuo e por isso, emprega unicamente a metaheurística PSO em sua versão contínua.

3.4 Algoritmo Genético - GA

O Algoritmo genético (GA) é uma metaheurística pertencente a uma família de modelos computacionais inspirados na evolução natural das espécies. Proposto por Jonh H. Holland em 1975, o GA tem como objetivo primeiro a resolução de problemas de otimização.

Semelhante ao PSO, O GA é uma metaheurística populacional, inicializada com uma população gerada aleatoriamente, composta por candidatos a soluções do problema, os cromossomos.

Segundo Zuben e Castro (2004), o GA representa uma das técnicas mais empregadas dentre as componentes da Computação Evolutiva.

Os termos empregados pela metaheurística Algoritmos Genéticos podem ser definidos por:

- $P(t) = (x_1^t, x_2^t, \dots, x_n^t)$: população, conjunto de indivíduos (cromossomos) pertencentes ao espaço de busca;
- $x_i^t = (g_1^t, g_2^t, \dots, g_n^t)$: cromossomo, candidato à solução do problema de otimização;
- g_j^t : gene, representa uma variável da solução;
- Alelo: refere-se ao conjunto de valores possíveis que um gene pode assumir;
- Aptidão (*fitness*): mede a adequação de um cromossomo para solução do problema tratado (geralmente associada ao valor da função objetivo específica do problema).

Segundo Mitchell (1997), o GA, assim como os outros métodos computacionais inspirados em sistemas naturais, pode ser classificado de acordo com duas aplicações básicas:

1. Como modelos computacionais de processos evolutivos naturais.
2. Como técnicas para resolver problemas tecnológicos, em especial, como ferramenta de Otimização.

A figura 3.4 apresenta a estrutura básica de um algoritmo GA.

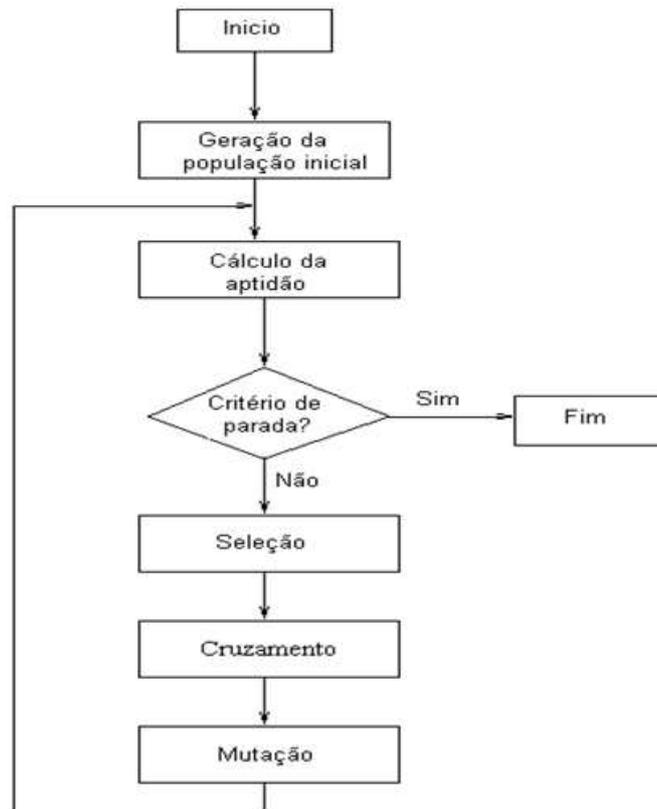


Figura 3.4: Estrutura básica do Algoritmo Genético, segundo Silva e Oliveira (2004).

De acordo com Bezerra (2008) o desempenho de um algoritmo genético está diretamente associado ao ajuste dos seus parâmetros de controle, relativos a:

1. Probabilidade de cruzamento;
2. probabilidade de mutação;
3. tamanho da população.

A estimação criteriosa desses parâmetros é fundamental na busca da melhor solução do problema de otimização. Man et al. (1986) apontam que, a escolha dos parâmetros de controle é extremamente dependente da natureza da função objetivo do problema de otimização em questão.

A figura 3.5 ilustra o Algoritmo para um GA simples.

```
Algoritmo GA  
  Inicie a população;  
  Avalie a população;  
  Enquanto (critério de parada não for atingido)  
    faça:  
      Selecione indivíduos para próxima população;  
      Aplique cruzamento e mutação;  
      Avalie a população;  
  Fim Enquanto;  
Fim Algoritmo.
```

Figura 3.5: Pseudocódigo Algoritmo Básico GA.

Em sua fase inicial, é gerada aleatoriamente uma população inicial de indivíduos, também chamados de cromossomos, que são candidatos a solução do problema. A solução advinda da metaheurística GA é dada por um conjunto de valores reais ou binários, genes. A cadeia dos genes forma o cromossomo, conforme representado na figura 3.6.

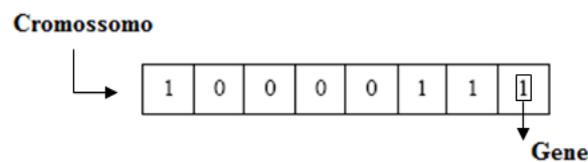


Figura 3.6: A estrutura de um cromossomo binário.

A função de aptidão (fitness) avalia a aptidão de cada cromossomo. A avaliação de um cromossomo consiste em verificar se este cromossomo otimiza a solução do problema. Essa verificação é feita a partir de uma função representativa do problema e seu objetivo é fazer uma medida de aptidão para cada indivíduo (cromossomo). Este procedimento vai se repetir durante toda a execução do GA.

A operação de seleção, feita após a avaliação de aptidão de todos os indivíduos, é responsável pela reserva de uma porcentagem dos cromossomos da geração atual para reprodução na próxima geração, sendo os indivíduos restantes, descartados. A probabilidade de um indivíduo vir a ser selecionado é proporcional à sua aptidão. Na figura 3.7, os números 1, 2 e 3 representam três indivíduos e suas respectivas probabilidades de que venham a ser escolhidos para reprodução. Dessa forma, o indivíduo número 2, devido a sua maior aptidão, terá maior probabilidade de ser selecionado.

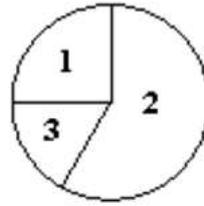


Figura 3.7: Probabilidade de seleção de um cromossomo.

No cruzamento (crossover), a partir de um par de cromossomos “pais” selecionados, são gerados dois cromossomos “filhos”, definidos dentro de um critério específico de cruzamento. Se nenhum cruzamento for efetuado, os “filhos” serão a cópia exata dos “pais”.

Um critério utilizado para o cruzamento é corte do cromossomo em um ponto específico (ponto de corte), esse ponto é escolhido aleatoriamente nos cromossomos “pais” e os divide em duas partes. Os cromossomos “filhos” são formados com uma parte do “pai 1” e outra do “pai 2”. A figura 3.8 ilustra o cruzamento entre cromossomos. Pode existir mais de um ponto de corte (multi-point crossover) no cruzamento.

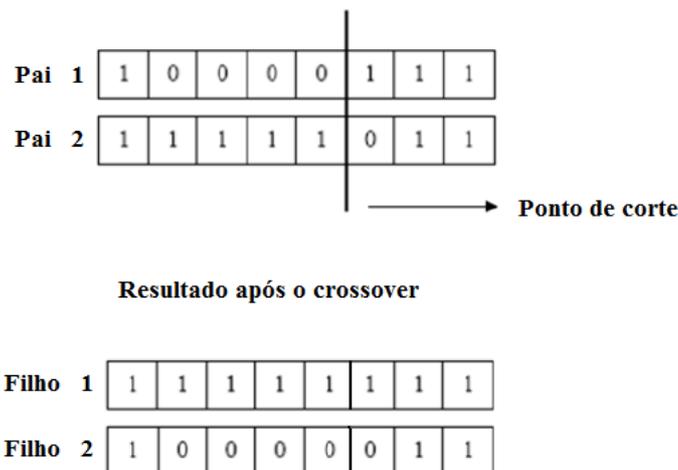


Figura 3.8: Processo de cruzamento de um cromossomo binário.

O processo de mutação ocorre em alguns indivíduos da população. Uma taxa percentual de mutação é atribuída à população, o que significa que haverá alteração em um ou mais genes de um cromossomo. O objetivo da mutação é garantir que a população não convirja prematuramente para um máximo ou mínimo local. A mutação garante também a diversidade nas características dos indivíduos, permitindo assim que informações que não estiveram presentes em nenhum dos indivíduos da população sejam introduzidas conforme ilustrado na figura 3.9. Além disso, ela permite que novos pontos do espaço de busca sejam avaliados.

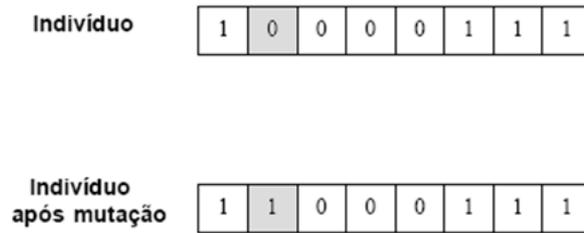


Figura 3.9: Processo de Mutação de um cromossomo binário

Após a mutação, o algoritmo efetua a avaliação dos indivíduos da nova geração e a verificação dos critérios de parada.

3.5 Considerações Finais

Neste capítulo, apresentou-se na seção 3.1 uma introdução sobre as metaheurísticas, listando as principais encontradas na literatura, na seção 3.2 foi feita uma justificativa para o uso de metaheurísticas no presente trabalho, na seção 3.3 fez-se um estudo sobre a metaheurística PSO e na seção 3.4 fez-se um estudo sobre o algoritmo genético. As metaheurísticas são empregadas nos mais diversos problemas de otimização conseguindo bons resultados, por isso o sua utilização está cada vez mais comum em diversas áreas do conhecimento.

Capítulo 4

PSO e GA Aplicados na Elaboração de um Modelo de Credit Scoring

4.1 Introdução

A aplicação de metaheurísticas populacionais PSO e GA na elaboração de um modelo de credit scoring, foi estabelecida utilizando a função objetivo proposta por Picinini et al. (2003). As versões do PSO e do GA implementados neste trabalho alcançaram bons resultados nas validações dos modelos.

4.2 A Função Objetivo

A função objetivo é utilizada para o cálculo do escore (S_j) de cada indivíduo (partícula no PSO e cromossomo no GA). Essa função está representada na equação 4.1:

$$S_j = \sum_{i=1}^n w_i(p_{ij}) \quad (4.1)$$

para:

- S_j é o escore obtido pelo cliente j ;
- w_i é o peso a ser estimado, relativo à categoria i ;
- p_{ij} é o indicador binário, igual a 1, se o cliente j possui a categoria i e 0, caso contrário.

Efetuada a estimativa dos pesos relativos (w), um candidato a cliente é classificado com “bom” ou “mau” de acordo com o valor de seu escore. Se $S_j \geq 0$, o cliente é considerado “bom”. Entretanto, se $S_j < 0$, o cliente é considerado “mau”.

A estratégia empregada neste trabalho, visando verificar a competência das metaheurísticas desenvolvidas, considera um banco de dados histórico com as informações dos clientes e seus respectivos escores, se “bom” ou “mau” pagador. Uma parte dessas informações fica reservada para a estimação dos pesos (w) (treinamento) e a parte restante à validação (teste) do modelo.

4.2.1 A Função de Avaliação

Após o cálculo do escore de cada um dos clientes reservados à validação, é necessário verificar se a classificação de cada cliente foi correta. Esse cálculo é efetuado através da função de avaliação, sendo denominado percentual de acertos.

O percentual de acertos é obtido pelo quociente entre o total de acertos encontrados e o total de clientes reservado, para validação, conforme representado na equação 4.2:

$$\text{Percentual de acertos} = \frac{\text{total de acertos}}{\text{total de clientes}} \quad (4.2)$$

O emprego dos algoritmos PSO e GA visa solucionar um problema de otimização. Com base no histórico de clientes reservado para o treinamento, buscou-se obter um vetor de pesos $W^T = [w_1, w_2, w_3, w_4, \dots, w_n]$, que resulte num critério otimizado de classificação com elevado percentual de acertos, na validação.

Então, a partir de um histórico de clientes, utiliza-se as metaheurísticas PSO e GA na construção dos modelos de credit scoring. O esquema básico das aplicações das metaheurísticas GA e PSO em um modelo de credit scoring está ilustrado na figura 4.1.

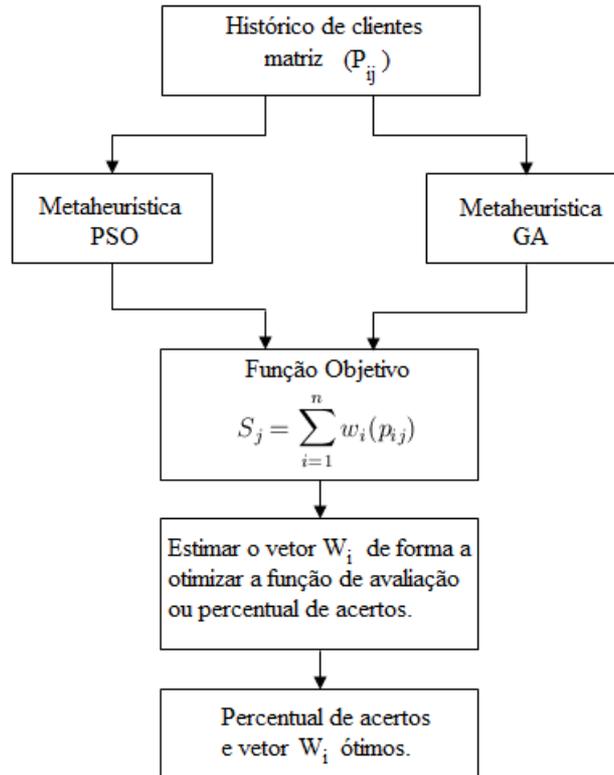


Figura 4.1: Esquema básico da elaboração de um modelo de credit scoring.

4.3 Preparação das bases de dados

No presente trabalho foram utilizadas três bases de dados para crédito pessoal.

1. A primeira instância (base de dados I) é relativa ao histórico de clientes de uma instituição financeira de Belo Horizonte. Essa base contém 904 clientes e foi utilizada por Dutra (2008).
2. A segunda instância (base de dados II) é a mesma utilizada por Machado e Neto (2007). Essa base tem 653 clientes.
3. A terceira instância (base de dados III) é formada por um histórico de 5000 clientes de uma instituição financeira de Belo Horizonte.

Cada instância é dividida em dois grupos, sendo um deles utilizado para a obtenção e outro para a validação de um modelo de credit scoring.

1. Treinamento - onde foram construídos os modelos de credit scoring;
2. Validação - onde os modelos foram avaliados.

Na elaboração dos modelos de credit scoring, foram adotados dois métodos, a saber:

1. O Método 1, que consiste em dividir a base de dados em duas partes: a primeira parte fica reservada para o treinamento, e a segunda para a validação. Esse procedimento foi adotado por Gonçalves (2005).
2. O Método 2, que consiste em dividir a base de dados em q partes. Para o treinamento são utilizadas $q - 1$ partes da base, ficando reservada então, uma parte para a validação. O processo é repetido para cada uma das q partes, gerando assim, q conjuntos de treinamento e validações. Esse procedimento foi adotado por Machado e Neto (2007).

Ao se reservar uma parte da base de dados histórica de clientes para a validação, o que se busca é que a estimativa dos pesos não tenha sido influenciada por esse conjunto de informações. Os testes (validação) são realizados utilizando a função objetivo e a função de avaliação com o melhor resultado encontrado durante a construção do modelo (treinamento).

De acordo com a função objetivo, as informações dos clientes são reclassificadas segundo um código binário, neste, o número 1 (um) representa a presença da característica e o 0 (zero), a sua ausência, conforme ilustrado na figura 4.2. Nessa ilustração de um cliente, as duas primeiras colunas informam a categoria sexo. Esse é masculino se $[1, 0]$ e feminino se $[0, 1]$. A variável região onde reside, é composta por cinco categorias, cada uma dessas apontando uma região. O cliente reside na região quatro, que pode referir-se a um conjunto de bairros etc. A variável salário está categorizada em três faixas, ficando o cliente classificado na segunda faixa que provavelmente é um intervalo entre dois valores em unidade monetária.

Sexo		Região onde reside					Salário		
1	0	0	0	0	1	0	0	1	0

Figura 4.2: Código binário para classificação de um cliente

Uma base de dados de um histórico de credit scoring é dividida em variáveis discretas e contínuas. A categorização das variáveis contínuas pode ser feita determinando os decis dessas variáveis. Nas 10 faixas criadas pelos decis são identificados os clientes “bons” e “maus”; logo após é calculado o risco relativo conforme descrito por Gonçalves e Gouvêa (2006).

O risco relativo é o quociente entre percentual de clientes “bons” e o percentual de clientes “maus”, de cada faixa determinada pelos decis. A categorização de uma variável contínua está descrita na tabela 4.1. O risco relativo pode ser utilizado

para diminuir o número de categorias por variável, uma vez que se pode agrupar categorias com risco relativo próximo.

Tabela 4.1: Variável contínua discretizada por decis e risco relativo

	Bons	Maus	% Bons	% Maus	Risco Relativo	Categorização
Faixa 1	14	30	0,3182	0,6818	0,4667	1
Faixa 2	54	22	0,7105	0,2895	2,4545	2
Faixa 3	20	49	0,2899	0,7101	0,4082	1
Faixa 4	36	21	0,6316	0,3684	1,7143	3
Faixa 5	21	52	0,2877	0,7123	0,4038	1
Faixa 6	71	11	0,8659	0,1341	6,4545	4
Faixa 7	20	43	0,3175	0,6825	0,4651	1
Faixa 8	42	24	0,6364	0,3636	1,7500	3
Faixa 9	50	17	0,7463	0,2537	2,9412	2
Faixa 10	55	9	0,8594	0,1406	6,1111	4

4.3.1 Base de dados I

No presente trabalho, utilizou-se a base de dados I com 904 clientes, sendo 438 considerados “bons” e 466 considerados “maus”. Não se levou em consideração os 247 clientes intermediários propostos por Dutra (2008).

A *base de dados I*, foi re-categorizada em 42 dimensões binárias. A sua representação na função objetivo está de acordo com a equação 4.3

$$S_j = \sum_{i=1}^{42} w_i(p_{ij}) \quad (4.3)$$

Utilizando o *primeiro método* para a obtenção de um modelo de credit scoring, separou-se aleatoriamente, 600 clientes, sendo 300 “bons” e 300 “maus”, esses foram utilizados no treinamento. Os outros 304 clientes ficaram reservados para a validação do modelo.

Num segundo momento, com o emprego do *segundo método*, a base foi separada em nove grupos, sendo oito deles com 100 clientes, e um com 104 clientes.

As variáveis originais da base I, estão descritas na figura 4.3.

Variável	Descrição	Variável	Descrição
X1	Valor Financiado	X7	Renda total
X2	Valor da Prestação	X8	Profissão
X3	Quantidade de Prestação	X9	Região que reside
X4	Idade na data admissão/idade atual	X10	Sexo
X5	Quantidade de contratos na instituição	X11	Aluguel (Sim/Não)
X6	Idade	X12	Resultado (Bom/Mau)

Figura 4.3: Descriminação das variáveis da base I.

4.3.2 Base de dados II

A *base de dados II* possui 653 clientes, sendo 296 considerados “bons” e 357 considerados “maus”. Essa base foi utilizada por Machado e Neto (2007).

A base de dados II, foi re-categorizada em 62 dimensões binárias. Esse número de dimensões foi o mesmo adotado por Machado e Neto (2007) em seu trabalho. A sua representação na função objetivo está de acordo com a equação 4.4.

$$S_j = \sum_{i=1}^{62} w_i(p_{ij}) \quad (4.4)$$

Utilizando o *primeiro método* para a obtenção de um modelo de credit scoring, separou-se aleatoriamente 500 clientes, sendo 250 “bons” e 250 “maus”, esses foram utilizados no treinamento. Os outros 153 clientes ficaram reservados para a validação do modelo.

Com o *segundo método*, a base foi separada em sete grupos, sendo seis deles com 93 clientes, e um com 95 clientes.

A alternativa de se manter os mesmos parâmetros utilizados por Machado e Neto (2007), foi comparar a eficiência dos algoritmo PSO e GA utilizados no presente trabalho, com aquela obtida por esses autores.

A Figura 4.4 apresenta a categorização das variáveis para a base II. Os nomes das variáveis não são informados pelo banco de dados.

Variáveis	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
Categorias	2	5	3	2	2	14	9	4	2	2	4	2	2	4	5

Figura 4.4: Variáveis categorizadas.

4.3.3 Base de dados III

A base de dados III, possui 5000 clientes, sendo 2500 considerados “bons” e 2500 considerados “maus”. Essa base é relativa a um histórico de clientes, de uma instituição financeira de Belo Horizonte. Os empréstimos foram concedidos no período de agosto a setembro de 2008.

Originalmente, essa base era composta por 10 variáveis: sexo, idade, profissão, renda total, região que reside, aluguel, valor financiado, quantidade de prestações, valor da prestação, quantidade de contratos na instituição e idade na data admissão/idade atual, sendo re-categorizadas em 40 dimensões binárias. A sua representação na função objetivo está de acordo com a equação 4.5.

$$S_j = \sum_{i=1}^{40} w_i(p_{ij}) \quad (4.5)$$

Para essa base de dados, foi utilizado somente o segundo método de construção do modelo.

A base foi separada aleatoriamente em 10 partes, cada uma com 500 clientes.

A figura 4.5 mostra um diagrama detalhado sobre a construção de um modelo de credit scoring.

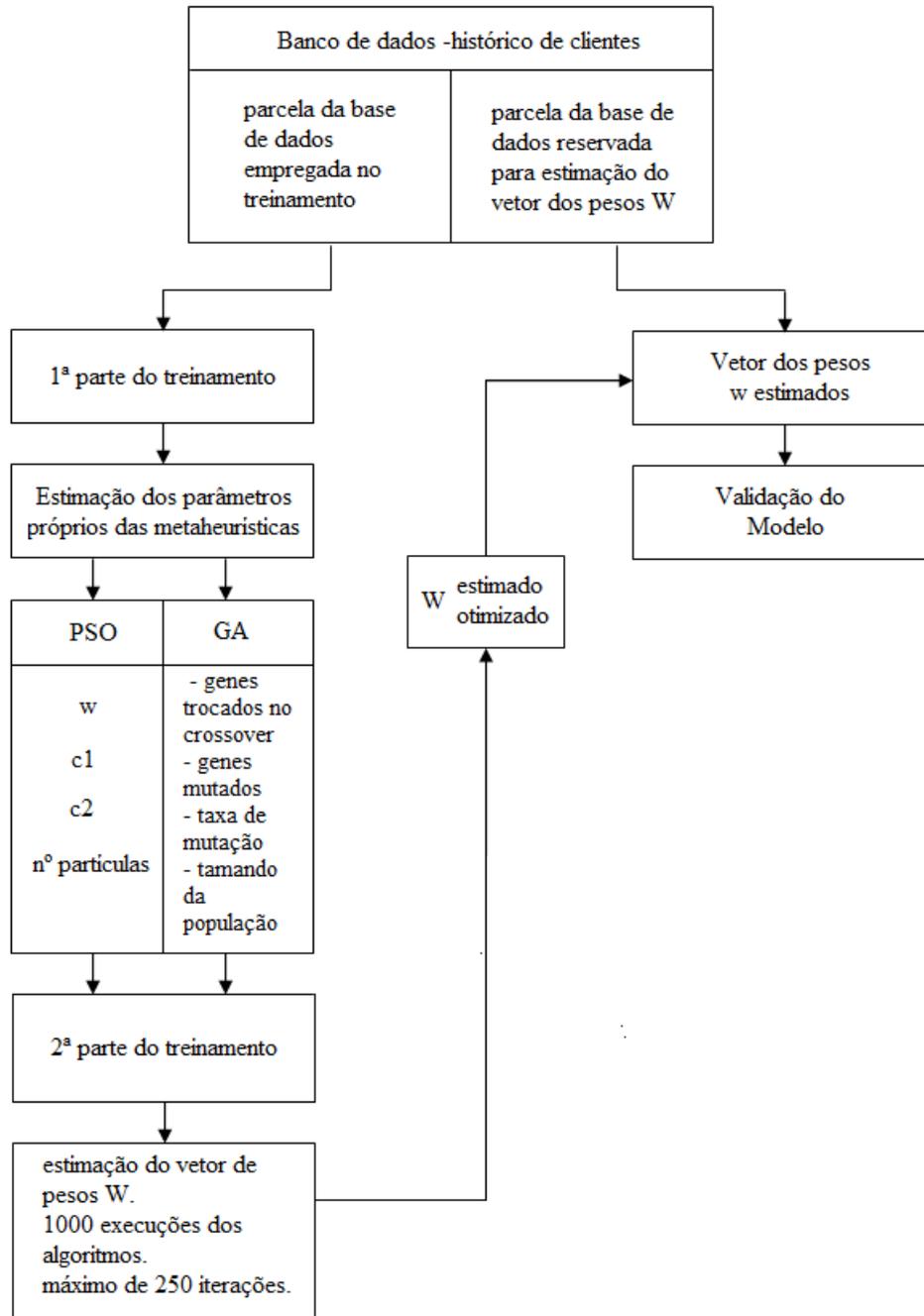


Figura 4.5: Diagrama Elaboração de um Modelo de Credit Scoring.

4.4 O emprego da Metaheurística PSO

Inicialmente é gerada uma população com p partículas no espaço n -dimensional. Para cada uma das n posições de cada partícula são gerados aleatoriamente pesos, escolhidos no intervalo $[-1,1]$, conforme figura 4.6. O domínio de validade dos pesos e a função objetivo empregados nesse trabalho são originalmente propostos por

Picinini et al. (2003).

	1ª	2ª	3ª		n - 1	n
Partícula 1	-0,30	0,18	-0,55	. . .	0,67	0,46
Partícula 2	0,91	-0,12	0,77	. . .	-0,31	0,44
⋮						
⋮						
⋮						
Partícula P	0,22	-0,16	0,35	. . .	0,88	0,14

Figura 4.6: Partículas do PSO

O algoritmo PSO aplicado ao modelo de credit scoring está ilustrado na figura 4.7.

<p>Iniciar Algoritmo PSO Iniciar população; Carregar Clientes; Enquanto ($i < n^o \text{ iterações}$) e ($\text{itera\c{c}ões sem melhora} < 0,1 * n^o \text{ itera\c{c}ões}$) Faça: 1 - Calcular o escore através da equação (4.1) 2 - Calcular o índice de acertos através da equação (4.2) 3 - Atualizar a matriz B($pbest_i$); Se $f(x_i) > f(pbest_i)$, então: $pbest_i = x_i$; Fim Se 4 - Atualizar o vetor G($gbest$); Se $pbest_i > f(gbest)$, então: $gbest = pbest_i$; Fim Se 5 - Atualizar a velocidade de cada partícula, conforme equação (3.1); 6 - Atualizar a posição de cada partícula, conforme equação (3.2); Fim Enquanto Escreva Solução; Fim _ Algoritmo PSO.</p>
--

Figura 4.7: Pseudocódigo Algoritmo PSO aplicado ao modelo de credit scoring.

Na Matriz formada pelas informações cadastrais, as linhas representam os diferentes clientes, e as colunas cada uma das categorias. As posições de uma partícula são iguais ao número de dimensões do problema. O escore de cada cliente é calculado através da equação 4.1. O produto entre a matriz dos clientes e o vetor de uma partícula (peso) resulta num vetor solução, conforme exemplo a seguir.

Um exemplo para uma melhor compreensão do problema de credit scoring será agora apresentado. Seja a carteira de clientes binária apresentada pela tabela 4.2.

Tabela 4.2: Carteira de Clientes Binária

Cliente	Sexo		Região de Residência				Salário			Resultado
1	0	1	1	0	0	0	1	0	0	1
2	1	0	0	1	0	0	0	1	0	2
3	1	0	0	0	1	0	1	0	0	1
4	0	1	0	0	0	1	0	0	1	1
5	0	1	0	1	0	0	1	0	0	1
6	1	0	1	0	0	0	1	0	0	2
7	1	0	0	0	1	0	0	1	0	2
8	0	1	0	0	0	1	0	0	1	1

As categorias envolvidas nesse exemplo são sexo, região de residência, salário e, na última coluna o desfecho de cada um dos contratos. Nessa coluna o número 1 significa que o cliente cumpriu com todos os pagamentos (“bom”) e o número 2 representando os casos nos quais o cliente não cumpriu com o mesmo (“mau”). A figura 4.8 apresenta a matriz dos clientes (A), o vetor dos pesos (W) e o vetor escore (S).

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad W = \begin{bmatrix} -0,76 \\ -0,20 \\ 0,75 \\ 0,20 \\ -0,91 \\ 0,99 \\ 0,99 \\ 0,27 \\ -0,3 \end{bmatrix} \quad \text{e} \quad S = \begin{bmatrix} 1,54 \\ -0,29 \\ -0,68 \\ 0,46 \\ 0,99 \\ 0,98 \\ -1,40 \\ 0,46 \end{bmatrix}$$

Figura 4.8: Matriz binária, vetores peso e escore

$A * W = S$, onde S é o vetor escore, e cada um dos seus componentes S_j pertence ao conjunto dos números reais.

R_j é o vetor resultado relacionado ao escore de cada cliente. Se $S_j < 0 \Rightarrow R_j = 2$, do contrário, se $S_j \geq 0 \Rightarrow R_j = 1$.

$$S = \begin{bmatrix} 1,54 \\ -0,29 \\ -0,68 \\ 0,46 \\ 0,99 \\ 0,98 \\ -1,40 \\ 0,46 \end{bmatrix} \quad \text{e} \quad R = \begin{bmatrix} \mathbf{(1)} \\ \mathbf{(2)} \\ 2 \\ \mathbf{(1)} \\ \mathbf{(1)} \\ 1 \\ \mathbf{(2)} \\ \mathbf{(1)} \end{bmatrix}$$

Figura 4.9: Vetores escore e resultado

Na figura 4.9, o vetor R , quando comparado com a coluna dos resultados da tabela 4.2 apresenta um total de 6 acertos, que foram registrados por parêntesis. O percentual de acertos portanto, definido como a razão entre o total de acertos e o total de clientes, é igual a 75%.

A seguir, os cálculos deverão ser repetidos com um novo vetor de pesos (partícula ou cromossomo) gerados pelas metaheurísticas PSO ou GA, respectivamente, visando otimizar a solução.

Cada partícula é multiplicada pela matriz de clientes, determinando dessa forma o escore de cada cliente. Em seguida é calculado o percentual de acertos de cada partícula. O objetivo do algoritmo é otimizar o valor desse percentual de acertos.

Ao final das iterações, cada partícula estará relacionada a um percentual de acertos. No desenvolvimento do algoritmo são identificadas as matrizes de posição X , velocidade V , das melhores posições de cada partícula B ($pbest$) e do vetor com a melhor posição entre todas as partículas G ($gbest$). Tais matrizes e o vetor são atualizados segundo as equações: 3.1 e 3.2. O processo é repetido até atingir o número máximo de iterações ou até atingir 10% do número de iterações sem melhora do algoritmo.

4.5 O emprego da Metaheurística GA

No GA, similar ao que ocorreu no PSO, inicialmente é gerada uma população com x cromossomos no espaço n -dimensional. Para cada uma das n posições de cada cromossomo são gerados aleatoriamente pesos, escolhidos no intervalo $[-1,1]$, conforme mostrado na figura 4.10. Esse intervalo assim como a função objetivo são propostos por Picinini et al. (2003), ao apresentar um modelo de credit scoring utilizando algoritmos genéticos.

	1ª	2ª	3ª		n - 1	n
Cromossomo 1	-0,30	0,18	-0,55	. . .	0,67	0,46
Cromossomo 2	0,23	-0,46	0,68	. . .	0,21	-0,53
⋮				⋮		
Cromossomo x	0,99	0,45	-0,34	. . .	0,88	0,14

Figura 4.10: Cromossomos do GA

Na matriz de dados, cada linha representa um cliente. Cada categoria de um cliente corresponde a uma dimensão do espaço n -dimensional. O número de genes

de um cromossomo é igual ao número de dimensões do problema. O escore de cada cliente é calculado através da equação 4.1. O processo é feito realizando a multiplicação da matriz de clientes com o vetor de um cromossomo (peso).

Cada cromossomo é multiplicado pela matriz de clientes, calculando o escore de cada cliente. Então, através da equação 4.2, é calculado o percentual de acertos de cada cromossomo. O objetivo do algoritmo é otimizar o valor desse percentual de acertos.

A técnica de otimização gera inicialmente a matriz de clientes e a matriz com os cromossomos. Em seguida é feita a avaliação da população inicial através das equações 4.1 e 4.2.

O próximo passo é a escolha dos pais para efetuarem o cruzamentos (método de seleção). Os pais foram escolhidos aleatoriamente na matriz de cromossomos.

Para o cruzamento foi utilizado o método conhecido como crossover uniforme onde é gerado aleatoriamente um vetor binário, também conhecido como “máscara”. Esse procedimento está ilustrado na figura 4.11.

Pais	Máscara	Filhos
001100010001	100111010101	001100110001
001010100001		001010000001

Figura 4.11: Crossover Uniforme.

Se o primeiro valor da máscara binária for igual a 1 (um), então o filho 1 recebe o primeiro gene do pai 1, logo o filho 2 recebe o primeiro gene do pai 2. Se o próximo valor da máscara for 0 (zero), então o filho 1 recebe o segundo gene do pai 2, logo o filho 2 recebe o segundo gene do pai 1. O processo é repetido até o final da máscara e nesse momento os filhos são formados e armazenados na matriz filhos.

O próximo passo é denominado mutação. O método escolhido aqui consiste na mutação por indivíduo. Se um indivíduo é selecionado para a mutação, então serão substituídos alguns de seus genes, escolhidos aleatoriamente, onde os novos valores dos genes mutados pertencem ao intervalo $[-1,1]$.

A matriz dos cromossomos pais e a matriz dos cromossomos filhos são utilizadas em uma escala de proporcionalidade (método da roleta) onde cada indivíduo será escolhido de acordo com o seu percentual de acertos. Os indivíduos com maiores percentuais de acerto têm maior chance de passarem para uma próxima geração.

Após os processos descritos, a geração dos filhos é então avaliada. Através das equações 4.1 e 4.2 é atualizada a melhor solução encontrada na população e uma nova população é formada.

O processo é repetido sucessivamente, até atingir o número máximo de iterações ou até atingir 10% do número de iterações sem melhora do algoritmo.

O algoritmo GA aplicado ao modelo de credit scoring está ilustrado na figura 4.12.

Iniciar Algoritmo GA
 Iniciar população;
 Carregar clientes ;
 Avaliar população através das equações (4.1) e (4.2)
Enquanto($i < n^{\circ}$ iterações) e ($iterações\ sem\ melhora < 0,1 * n^{\circ}$ iterações)
 Faça:
 1 - Selecionar indivíduos para o cruzamento;
 2 - Aplicar cruzamento;
 3 - Aplicar mutação;
 4 - Selecionar indivíduos para próxima geração;
 5 - Avaliar população através das equações (4.1) e (4.2)
 6 - Atualizar melhor solução
 Se $solução > melhor\ solução$
 então: $melhor\ solução = solução$
 Fim Se
Fim_Enquanto;
 Escreva Solução;
Fim_Algoritmo GA.

Figura 4.12: Pseudocódigo Algoritmo GA aplicado ao modelo de credit scoring.

4.6 Teste de Hipótese

Em diversas situações faz-se necessário realizar generalizações sobre uma população com base em dados amostrais. Tal procedimento exige decidir em aceitar ou rejeitar uma afirmação sobre determinado parâmetro (estimação de parâmetros). Essa afirmação é chamada de hipótese, e a tomada de decisão sobre esse parâmetro é denominada teste de hipótese.

No teste de hipótese H_0 é chamado de hipótese nula, que é formulada sobre o parâmetro populacional. H_1 é a hipótese alternativa a hipótese nula. Para definir a região de aceitação e rejeição do teste hipótese utiliza-se o nível de significância α , geralmente fixado em 5%.

A partir da amostra, o menor nível de significância que conduz a rejeição da hipótese nula é denominado $P - value$.

O procedimento geral para a construção de um teste de hipótese para um parâmetro populacional pode ser interpretado da seguinte maneira. Considera-se uma variável X em uma dada população. Tem-se uma hipótese sobre determinado parâmetro θ dessa população. Então, afirma-se que esse valor é um número θ_0 . A partir de uma amostra aleatória dessa população deseja-se aceitar ou rejeitar tal hipótese.

Então, para a hipótese nula têm-se, $H_0 : \theta = \theta_0$. Caso a hipótese nula seja rejeitada, considera-se então a hipótese alternativa $H_1 : \theta \neq \theta_0$. Dependendo da formulação do problema podem ocorrer variações para a hipótese alternativa, $H_1 : \theta < \theta_0$ ou $H_1 : \theta > \theta_0$.

A partir de uma decisão tomada podem ocorrer erros, que são descritos por:

- Erro do tipo I - rejeitar a hipótese nula quando esta é verdadeira.
- Erro do tipo II - não rejeitar a hipótese nula quando esta é falsa.

Passos para construção de um teste de hipótese, segundo Bussab e Morettin (1987):

1. Fixe qual a hipótese H_0 a ser testada e qual a hipótese alternativa H_1 .
2. Use a teoria estatística e as informações disponíveis para decidir qual estimador será usado para julgar a hipótese nula.
3. Fixe a probabilidade α de cometer um erro de primeira espécie, e use este valor para construir a região crítica.
4. Use as informações fornecidas pela amostra para encontrar o valor da estatística que definirá a decisão.
5. Se o valor da estatística observado na amostra não pertencer à região crítica, aceita-se H_0 , caso contrário, rejeita-se.

De acordo com Magalhães e Lima (2001), quando não se tem informação sobre a variância da variável aleatória que está sendo estudada, precisa-se contornar essa dificuldade. Mantendo a suposição de que a variável aleatória de interesse tem distribuição Normal de média μ e variância σ^2 . O desvio padrão é desconhecido, ele precisa ser estimado.

Pode-se utilizar a tabela t -Student quando o desvio padrão é desconhecido e o tamanho da amostra é menor que 30. A formulação da hipótese: $X : N(\mu, \sigma^2)$, μ e σ^2 desconhecidos. Então, para a hipótese nula têm-se, $H_0 : \mu = \mu_0$. Caso a hipótese nula seja rejeitada, considera-se então a hipótese alternativa $H_1 : \mu \neq \mu_0$. A hipótese alternativa poderia ser $H_1 : \mu > \mu_0$ ou $H_1 : \mu < \mu_0$, o que mudaria apenas a região crítica de bilateral para unilateral.

Pode-se utilizar o $P - value$ para aceitar ou rejeitar a hipótese nula. Se $P - value \leq \alpha$ então rejeita-se a hipótese nula, se $P - value > \alpha$ então não rejeita-se a hipótese nula.

No presente trabalho, são realizados teste de hipótese para os resultados obtidos com o método 2. Para os resultados obtidos com o método 1 o teste de hipótese não é aplicado, uma vez que essa metodologia considera apenas o melhor resultado da validação. Todos os testes de hipótese são realizados no software Minitab 14.1, utilizando a tabela *t*-Student. Os resultados estão descritos no capítulo 5.

4.7 Considerações Finais

A seção 4.1 deste capítulo apresentou uma introdução sobre modelos de credit scoring, na seção 4.2 foram apresentadas a função objetivo e a função de avaliação do modelo, na seção 4.3 verificou-se como foram preparados os dados para a construção do modelo, nas seções 4.4 e 4.5 verificou-se como aplicar a metaheurística PSO e AG respectivamente, em um modelo de credit scoring; e na seção 4.6 fez-se um estudo sobre um teste de hipótese.

Capítulo 5

Resultados e Discussões

5.1 Introdução

Este capítulo apresenta os resultados computacionais encontrados nas simulações de um Modelo de Credit Scoring. As Metaheurísticas empregadas na solução são a Otimização Por Enxame de Partículas (PSO) e Algoritmos Genéticos (GA).

As instâncias utilizadas neste trabalho são as mesmas empregadas por Dutra (2008) (Base de dados I), com o método de Análise Discriminante, e Machado e Neto (2007) (Base de dados II), com a metaheurística Algoritmos Genéticos. Foi utilizada também a base de dados III, referente a uma cartela de 5000 clientes de uma empresa de crédito situada em Belo Horizonte.

Os Algoritmos foram implementados na linguagem de programação C, utilizando o Dev-C++ versão 4.9.9.2. O computador onde foram executados os programas utiliza o Windows Vista Ultimate com Processador Genuine Intel(R) CPU T1400 @1.73GHz 1.73GHz, 3062 MB de Memória RAM e Sistema Operacional de 32 bits.

A estimação do vetor otimizado dos pesos (partícula no PSO e cromossomo no GA) é alcançada através do treinamento. Os testes ou validação são efetuados por intermédio do vetor dos pesos otimizados com o emprego da função objetivo e do conjunto de dados reservado para tal fim. A adoção desse procedimento permite verificar e otimizar a precisão das partículas ou dos cromossomos ao prever o comportamento de um candidato a cliente.

5.2 Ajuste dos Parâmetros dos Algoritmos

O ajuste dos parâmetros internos dos algoritmos PSO e GA é efetivado durante a elaboração do modelo de credit scoring.

Nessa etapa, também chamada de treinamento, obteve-se o conjunto dos melho-

res parâmetros para cada método.

Uma vez determinados esses parâmetros, são realizados os treinamentos finais. Obtido o vetor dos pesos (w) otimizado, são efetuadas as validações do modelo.

Na sequência, estão os ajustes de parâmetros do PSO e do GA.

5.2.1 Ajuste dos Parâmetros PSO

Utilizando a metaheurística PSO, os parâmetros a serem estimados são: w , $c1$, $c2$, relacionados pela equação 3.1, além do número de partículas do enxame. O processo é feito através de uma busca extensiva, na qual todos os coeficientes são mantidos constantes, exceto um, que é testado dentro de seu domínio de validade. Após a determinação do melhor valor para esse parâmetro, passa-se a empregá-lo de forma constante, e a variação de outro parâmetro será avaliada.

Os resultados são obtidos a partir da estratégia em se manter três parâmetros fixos e variar, em um domínio específico, um quarto parâmetro. Assim, é obtido o melhor *gbest* dentre as 1000 execuções, ou seja, a partícula com o maior percentual de acertos. Em seguida é designado um novo parâmetro a ser estimado e os 3 outros mantidos constantes. Ao final, obteve-se um conjunto de valores ótimos para cada um dos parâmetros. A cada nova base de dados, esse conjunto de parâmetros deve ser reavaliado.

Os melhores parâmetros encontrados para as bases de dados I, II e III, estão na tabela 5.1.

Tabela 5.1: Parâmetros do algoritmo PSO

Bases	Parâmetros			
	w	c1	c2	n° Partículas
I	0,95	2	2	200
II	1	2	2	250
III	0,95	2	2	400

5.2.2 Ajuste dos Parâmetros GA

Utilizando a metaheurística GA, os parâmetros a serem estimados são a taxa de mutação, número de genes trocados no crossover, número de genes mutados e número de cromossomos. O processo é semelhante ao utilizado no PSO, onde um dos parâmetros é variado e os outros permanecem fixos. Assim como no PSO, para uma nova base de dados, esse conjunto de parâmetros deve sempre ser reavaliado.

Os melhores parâmetros encontrados para as bases de dados I, II e III, estão na tabela 5.2.

Tabela 5.2: Parâmetros do algoritmo GA

Bases	Parâmetros			
	n° genes trocados no crossover	n° genes mutados	taxa de mutação	n° Cromossomos
I	20	10	2%	400
II	40	15	2%	400
III	20	10	3%	600

5.3 Resultados Computacionais PSO e GA

Numa segunda etapa do treinamento, os resultados obtidos, seja com PSO ou GA, são refinados através de 1000 execuções do algoritmo, com um número máximo de 250 iterações. Esse procedimento foi adotado para as bases de dados I, II e III.

O percentual de acertos calculado com a parte da base dados empregada no treinamento não possui a capacidade de validar o modelo. Esse percentual de acertos encontra-se apresentado no trabalho apenas como informação suplementar para análise de resultados.

A validação do modelo é efetuada ao final da etapa de treinamento, sendo, esta sim, responsável pela verificação das competências empregadas para construção do modelo. Estes testes empregam unicamente o subconjunto da base de dados reservada para tal. Os resultados das duas metaheurísticas estão sendo comparados entre si e também com os resultados de Dutra (2008) e Machado e Neto (2007).

De acordo com Picinini et al. (2003), modelos de credit scoring com percentual de acertos superiores a 65% podem ser considerados adequados, embora, nos dias de hoje, diversos trabalhos apresente valores muito superiores.

5.3.1 Resultados - base de dados I

A base de dados I é composta por 904 clientes, onde 600 clientes são designados aleatoriamente para o treinamento e 304, para avaliação. São feitos treinamentos utilizando o método 1. O melhor resultado do PSO e do GA, encontrado para a base de dados I durante o treinamento, está representado na tabela 5.3.

Tabela 5.3: Treinamento aplicando o Método 1 - Base de Dados I

	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
PSO	90,00%	82,82%	88,33%	78	47	68	4,78	3,14	4,03
GA	86,00%	79,00%	83,05%	157	26	95	12,65	1,95	7,75

Pode-se destacar a boa performance do PSO no ajuste dos dados. Além de obter um bom resultado, manteve uma alta média de acertos em todas as execuções. O número de iterações necessárias do algoritmo PSO é aproximadamente metade do obtido pelo GA.

Na validação do modelo, as duas metaheurísticas apresentaram resultados satisfatórios. PSO e GA atingiram um resultado melhor que o obtido por Dutra (2008), conforme mostrado na tabela 5.4.

Tabela 5.4: Validação aplicando o o Método 1 - Base de Dados I

	Método	Resultado
Dutra	Análise Discriminante	83,50%
Presente Trabalho	PSO	85,62%
Presente Trabalho	GA	83,66%

Ao se empregar o método 2 sobre a base de dados I, esta foi dividida em 9 subconjuntos. O método 2 consiste exatamente em se resolver 9 problemas. O primeiro deles reserva o 1º subconjunto para validação e as demais para treinamento. O segundo problema reserva o 2º subconjunto para validação e os demais para treinamento. Outros sete problemas são ainda resolvidos com os demais subconjuntos reservados para a validação de cada um.

As nove soluções obtidas no treinamento, encontram-se apresentadas na tabela 5.5.

Tabela 5.5: Treinamento aplicando o Método 2 - Base de Dados I

Treinamento PSO	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	90,17%	83,58%	89,00%	63	77	64	4,52	5,64	5,12
2	90,05%	87,06%	88,58%	54	48	53	3,96	3,51	4,00
3	90,17%	83,96%	88,68%	48	38	62	3,40	2,68	4,50
4	90,80%	86,44%	89,33%	126	46	71	9,17	3,32	5,18
5	90,55%	86,69%	88,87%	58	38	55	4,15	2,70	4,00
6	90,92%	86,69%	89,28%	56	67	64	4,11	4,95	4,69
7	90,42%	87,44%	89,28%	45	57	55	3,20	4,13	4,02
8	89,80%	85,32%	88,46%	62	36	55	4,55	2,57	4,21
9	89,75%	84,75%	88,81%	90	53	66	6,92	3,95	5,03
Treinamento GA	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	85,88%	78,63%	82,94%	164	35	92	17,78	3,74	10,13
2	85,38%	77,25%	82,26%	134	27	87	42,63	2,69	12,50
3	84,88%	76,88%	82,56%	90	26	84	9,42	2,62	9,36
4	86,38%	77,25%	82,94%	120	27	87	13,01	2,88	10,32
5	85,00%	78,00%	82,47%	153	28	84	17,56	2,91	9,17
6	86,38%	79,38%	82,89%	152	62	90	16,89	6,29	9,47
7	85,63%	78,00%	82,53%	87	26	82	8,95	2,63	8,78
8	85,38%	78,50%	82,08%	117	35	86	12,33	3,59	9,12
9	86,13%	79,25%	82,67%	198	29	90	20,65	2,92	9,56

Nesse conjunto de treinamento, o algoritmo PSO manteve uma boa média de percentual de acertos, bastante próxima dos melhores resultados. O GA também apresentou um bom percentual médio de acertos. Sua média de iterações e o seu tempo médio foram superiores ao do PSO. O PSO conseguiu atingir o melhor resultado com menos iterações e com um tempo menor que o GA. Os piores resultados obtidos pelo PSO são próximos dos melhores valores advindos, do GA, chegando, em alguns dos casos, a superá-los.

A tabela 5.6 apresenta, os melhores resultados na validação, de cada um dos nove arranjos. Verifica-se que o PSO apresentou melhor percentual de acertos em todos os testes.

Tabela 5.6: Validação aplicando o Método 2 - Base de Dados I

	1	2	3	4	5	6	7	8	9
PSO	87,00%	89,00%	90,00%	83,00%	88,00%	86,00%	88,00%	91,00%	88,46%
GA	78,00%	85,00%	86,00%	82,00%	83,00%	84,00%	82,00%	87,00%	84,62%

A figura 5.1 apresenta, para fins de comparação, o gráfico das performances do PSO e do GA na validação, para a base de dados I.

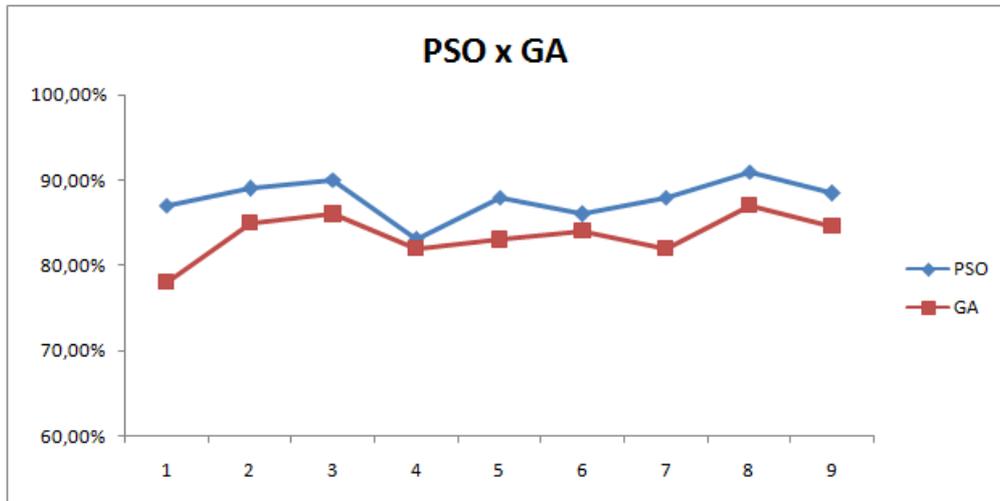


Figura 5.1: Conjunto de Validação - Base de dados I.

A comparação com o resultado obtido por Dutra (2008), deverá ser feita utilizando a média dos nove melhores resultados da validação. O resultado médio do GA é praticamente igual ao resultado obtido por Dutra (2008). O resultado médio do PSO foi superior aquele obtido pela análise discriminante e pelo GA. Os resultados encontram-se apresentados na tabela 5.7.

Tabela 5.7: Comparação aplicando o Método 2 - Base de Dados I

	Método	Resultado
Dutra	Análise Discriminante	83,50%
Presente Trabalho	PSO	87,83%
Presente Trabalho	GA	83,51%

Para esse conjunto de validação são realizados dois testes de hipóteses, um com o PSO e outro com o GA. O resultado obtido por Dutra (2008) é a hipótese nula, e os resultados atingidos com o PSO e o GA a hipótese alternativa.

A tabela 5.8 apresenta o resultado no teste de hipótese para o GA.

Tabela 5.8: Teste de Hipótese GA - Validação base de dados I

Hipótese Nula	Hipótese Alternativa	Intervalo de Confiança (95%)	Nível de significância	P-value
0,8350	0,8351	(0,814570; 0,855697)	0,05	0,988

Verifica-se no teste de hipótese que o P -value é maior que o nível de significância adotado, logo não rejeita-se a hipótese nula. Existe 95% de chances de se encontrar o valor médio dentro do intervalo de confiança, a hipótese alternativa está contida nesse intervalo.

A tabela 5.9 apresenta o resultado no teste de hipótese para o PSO.

Tabela 5.9: Teste de Hipótese PSO - Validação base de dados I

Hipótese Nula	Hipótese Alternativa	Intervalo de Confiança (95%)	Nível de significância	P-value
0,8350	0,8783	(0,860273; 0,896305)	0,05	0,001

Verifica-se no teste de hipótese que o P -value é menor que o nível de significância adotado, logo rejeita-se a hipótese nula e não rejeita-se a hipótese alternativa. Existe 95% de chances de se encontrar o valor médio dentro do intervalo de confiança, a hipótese alternativa está contida nesse intervalo.

5.3.2 Resultados - base de dados II

Nesse caso, a base é composta por 653 clientes. No primeiro método, o subconjunto reservado para treinamento conta com 500 elementos. Os 153 elementos restantes são utilizados na validação. São utilizadas as metaheurísticas PSO e GA com o método 1. Os resultados do treinamento estão dispostos na tabela 5.10.

Tabela 5.10: Treinamento aplicando o Método 1 - Base de Dados II

	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
PSO	92,00%	86,20%	89,57%	76	76	62	5,55	5,86	4,63
GA	89,40%	81,80%	86,70%	154	32	90	13,91	3,10	8,03

O PSO apresentou, novamente, uma melhor performance que o GA. Em geral o GA também apresentou bons resultados no percentual de acertos. Mas o número de iterações e o tempo médio mostraram que o PSO foi mais eficaz nesse método.

Na tabela 5.11 são dispostos os resultados de validação dos modelos. Os resultados finais do PSO e GA apresentam valores muito próximos de percentuais de acertos.

Tabela 5.11: Validação aplicando o Método 1 - Base de Dados II

Método	Resultado
PSO	84,31%
GA	83,66%

A tabela 5.12, mostra os resultados dos sete problemas de treinamento com PSO e com GA, baseado na divisão da base de dados em sete grupos (método 2). Em geral, os resultados do PSO foram melhores do que os do GA. Em ambas as metaheurísticas, o número de iterações médio e o tempo médio apresentam valores muito próximos, embora no PSO estes valores sejam menores que no GA. Pode-se também verificar que os valores do resultado médio do PSO é superior aos valores dos melhores resultados do GA. Dessa forma, o PSO apresenta, de maneira geral, uma maior eficiência quando comparado com o GA.

Tabela 5.12: Treinamento aplicando o Método 2 - Base de Dados II

Treinamento PSO	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	91,07%	83,21%	88,51%	196	106	99	9,95	5,54	5,03
2	90,71%	82,14%	88,04%	56	32	57	3,36	1,92	3,64
3	91,79%	84,11%	88,72%	48	54	57	3,68	4,16	4,30
4	90,89%	85,18%	88,74%	64	39	56	4,66	2,94	4,14
5	91,07%	83,75%	88,44%	55	33	60	3,94	2,36	4,49
6	90,89%	84,29%	88,70%	65	65	53	4,82	4,95	4,09
7	91,40%	83,69%	89,06%	51	35	56	3,68	2,55	4,22
Treinamento GA	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	87,32%	82,68%	85,84%	118	32	85	11,35	2,95	8,06
2	87,14%	80,18%	85,00%	137	28	89	13,30	2,49	8,56
3	88,39%	82,86%	85,79%	125	40	88	11,76	3,75	8,79
4	87,50%	82,68%	85,72%	125	38	94	11,70	3,52	9,66
5	86,96%	81,43%	85,10%	191	26	91	17,79	2,41	8,64
6	87,50%	82,86%	85,83%	124	37	89	12,31	3,43	8,69
7	88,35%	84,05%	86,48%	157	47	93	15,08	4,67	9,09

A tabela 5.13 apresenta os sete resultados da validação para a base de dados II segundo as duas metaheurísticas empregadas.

Tabela 5.13: Validação aplicando o Método 2 - Base de Dados II

	1	2	3	4	5	6	7
PSO	87,10%	89,20%	87,10%	87,10%	85,00%	83,90%	85,30%
GA	83,90%	87,10%	79,60%	90,30%	90,30%	78,50%	83,20%

A figura 5.2 apresenta, para fins de comparação, o gráfico com as performances do PSO e do GA na validação, para a base de dados II.

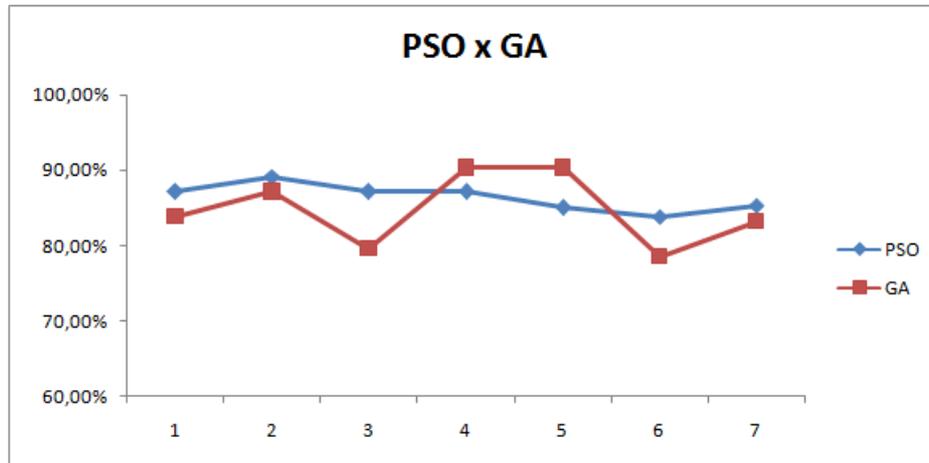


Figura 5.2: Conjunto de Validação - Base de dados II.

A média desses resultados está apresentada na tabela 5.14. O PSO obteve como resultado médio, uma melhor performance, do que aquela obtida pela metaheurística GA desenvolvida nesse trabalho e por Machado e Neto (2007).

Tabela 5.14: Comparação aplicando o Método 2 - Base de Dados II

	Método	Resultado
Machado e Neto	Algoritmos Genéticos	85,59%
Presente Trabalho	PSO	86,39%
Presente Trabalho	GA	84,70%

Para esse conjunto de validação são realizados dois testes de hipóteses, um com o PSO e outro com o GA. O resultado obtido por Machado e Neto (2007) é a hipótese nula, e os resultados atingidos com o PSO e o GA a hipótese alternativa.

A tabela 5.15 apresenta o resultado no teste de hipótese para o GA.

Tabela 5.15: Teste de Hipótese GA - Validação base de dados II

Hipótese Nula	Hipótese Alternativa	Intervalo de Confiança (95%)	Nível de significância	P-value
0,8559	0,8470	(0,802998; 0,891002)	0,05	0,638

Verifica-se no teste de hipótese que o P -value é maior que o nível de significância adotado, logo não rejeita-se a hipótese nula. Existe 95% de chances de se encontrar o valor médio dentro do intervalo de confiança, a hipótese alternativa está contida nesse intervalo.

A tabela 5.16 apresenta o resultado no teste de hipótese para o PSO.

Tabela 5.16: Teste de Hipótese PSO - Validação base de dados II

Hipótese Nula	Hipótese Alternativa	Intervalo de Confiança (95%)	Nível de significância	P-value
0,8559	0,8639	(0,847517; 0,880197)	0,05	0,278

Verifica-se no teste de hipótese que o P -value é maior que o nível de significância adotado, logo não rejeita-se a hipótese nula. Existe 95% de chances de se encontrar o valor médio dentro do intervalo de confiança, a hipótese alternativa está contida nesse intervalo.

5.3.3 Resultados - base de dados III

A base de dados III é formada por 5000 clientes, 2500 “bons” e 2500 “maus”. Desenvolveu-se para essa base de dados modelos de credit scoring empregando, unicamente, o método 2. A base de dados é, neste caso, dividida em 10 subgrupos com 500 indivíduos cada. dessa forma, são desenvolvidos 10 modelos, cada um deles contando com um diferente subgrupo de validação. Ao se empregar o método 1, no qual é resolvido apenas um problema, com uma parte da base de dados reservada para validação, os resultados podem apresentar determinadas tendências, devido às características dos clientes escolhidos para treinamento e teste.

Ao se selecionar aleatoriamente um grupo para treinamento e outro para validação, é possível que na parte escolhida para validação estejam presentes, por exemplo, um número muito maior de pontos influentes do que no subconjunto de treinamento.

Tabela 5.17: Treinamento aplicando o o Método 2 - Base de Dados III

Treinamento PSO	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	88,64%	84,49%	86,48%	239	91	137	140,71	54,08	79,94
2	88,40%	82,82%	85,77%	201	105	108	116,03	60,59	62,50
3	88,40%	86,53%	87,19%	113	142	119	64,86	83,25	69,49
4	88,38%	83,13%	86,13%	82	72	101	46,99	41,48	58,13
5	88,71%	82,91%	87,14%	85	69	130	48,40	40,08	78,00
6	87,91%	77,07%	84,95%	138	39	93	81,50	23,32	54,30
7	87,80%	84,04%	85,85%	70	85	83	45,64	63,27	54,36
8	88,51%	80,33%	86,59%	138	110	138	82,62	69,04	84,54
9	83,33%	82,13%	85,88%	138	43	97	106,08	32,38	65,61
10	87,78%	85,09%	86,64%	76	56	96	55,09	31,88	64,83
Treinamento GA	Melhor Resultado	Pior Resultado	Resultado Médio	Iterações Melhor Resultado	Iterações Pior Resultado	Iterações Médias	Tempo Melhor Resultado (s)	Tempo Pior Resultado (s)	Tempo Médio (s)
1	86,20%	81,51%	82,99%	152	91	106	71,63	43,08	50,24
2	85,67%	79,33%	83,45%	180	60	121	84,55	28,40	57,34
3	86,62%	80,69%	84,02%	196	64	123	95,35	32,60	61,55
4	86,36%	81,53%	84,30%	175	79	129	90,25	39,80	65,76
5	87,00%	79,22%	84,38%	191	56	134	252,88	74,94	90,04
6	86,78%	80,98%	84,00%	222	72	135	109,62	35,05	77,38
7	84,27%	78,73%	82,42%	141	55	100	66,69	26,00	48,10
8	86,20%	78,76%	84,10%	198	50	143	95,44	25,04	70,26
9	85,84%	82,27%	83,87%	250	84	136	119,28	41,61	65,59
10	85,98%	79,38%	83,98%	138	40	136	69,02	18,66	64,14

A tabela 5.17 apresenta os resultados dos dez conjuntos de treinamento. Nesse caso, o PSO novamente apresenta um melhor percentual de acertos em relação ao GA.

Na validação, o PSO obteve percentuais de acertos superiores à aqueles resultantes do GA, que também mostraram-se bastante adequados. A tabela 5.18 mostra o resultado do conjunto de validação.

Tabela 5.18: Validação aplicando o Método 2 - Base de Dados III

	1	2	3	4	5	6	7	8	9	10
PSO	91,00%	87,20%	87,40%	87,60%	82,80%	88,40%	86,00%	88,60%	86,20%	86,80%
GA	87,80%	85,00%	87,60%	86,20%	82,20%	88,80%	81,00%	88,80%	84,40%	85,00%

Os resultados da validação com o PSO e o GA, na base de dados III, estão representado na figura 5.3.

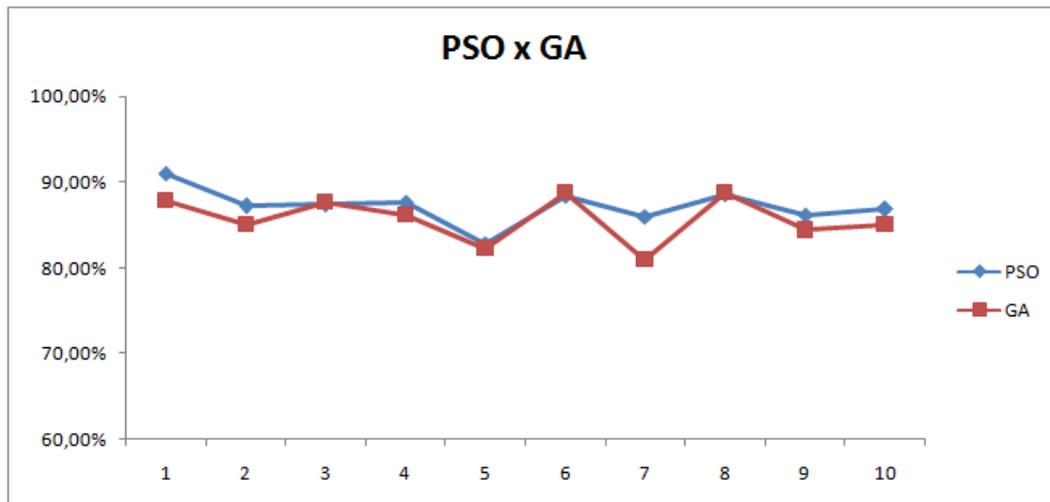


Figura 5.3: Conjunto de Validação - Base de dados III.

O resultado médio obtido pelo conjunto de validação encontra-se disposto na tabela 5.19. O PSO apresentou um valor superior ao GA, embora ambos os métodos tenham mostrado boa capacidade de predição nos problemas de credit scoring.

Tabela 5.19: Comparação aplicando o Método 2 - Base de Dados III

Método	Resultado
PSO	87,20%
GA	85,68%

Para esse conjunto de validação é realizado um teste de hipótese. O resultado obtido com o GA é a hipótese nula, e o resultado atingido com o PSO a hipótese alternativa.

A tabela 5.20 apresenta o resultado no teste de hipótese.

Tabela 5.20: Teste de Hipótese - Validação base de dados III

Hipótese Nula	Hipótese Alternativa	Intervalo de Confiança (95%)	Nível de significância	P-value
0,8568	0,8720	(0,856919; 0,887081)	0,05	0,049

Verifica-se no teste de hipótese que o P -value é menor que o nível de significância adotado, logo rejeita-se a hipótese nula e não rejeita-se a hipótese alternativa. Existe 95% de chances de se encontrar o valor médio dentro do intervalo de confiança, a hipótese alternativa está contida nesse intervalo.

5.3.4 Considerações Finais

Neste capítulo, foram apresentados os resultados obtidos na elaboração de um Modelo de Credit Scoring. Os testes foram separados em treinamento e validação. Utilizou-se, paralelamente, dois métodos da literatura, para a construção e validação do modelo.

A metaheurística PSO atingiu melhores resultados que o GA, obtendo na grande maioria da vezes melhor resultado com um número menor de iterações. O tempo computacional do PSO também quase sempre foi inferior ao GA.

Na validação, tanto PSO quanto GA obtêm performance adequada em relação a bibliografia presente. Nas comparações com Dutra (2008) e Machado e Neto (2007), os resultados do GA ficaram próximos dos resultados desses autores, sendo superados pela metaheurística PSO, que manteve um resultado médio superior aos demais.

Os testes apontam que, ao se utilizar o método 2 (separação da base de dados em q partes e resolução de q problemas), as técnicas metaheurísticas de estimação do vetor de pesos (W) são avaliadas de forma mais adequada e representativa, devido à sua característica média, e por levar em consideração a influência de cada um dos dados, desde os mais até os menos influentes.

No trabalho intitulado: Aplicação das Metaheurísticas PSO e AG na Estimação de Parâmetros de um Modelo de Secagem em Leito Fluidizado, Gomes (2008) obteve melhores resultados com PSO do que com o GA. Bezerra (2008) também obteve melhores resultados com o DPSO do que com o GA, no trabalho: Algoritmos Evolutivos Paralelos Aplicados ao Problema das p -Medianas.

Neste trabalho o PSO mostrou-se superior ao GA na elaboração de um modelo de credit scoring. Mas as duas técnicas obtiveram resultados nos testes considerados bons pela literatura.

Capítulo 6

Conclusão e Trabalhos Futuros

6.1 Conclusão

No presente trabalho foram desenvolvidos modelos de credit scoring utilizando as metaheurísticas PSO e GA. As instâncias onde se aplicou os métodos foram as mesmas utilizadas por Dutra (2008)(Análise Discriminante) e Machado e Neto (2007) (Algoritmos Genéticos). Para verificar a eficiência do método utilizou-se uma terceira instância com uma base de dados mais extensa. Essa instância corresponde a uma grande cartela de clientes de uma instituição financeira de Belo Horizonte.

Os modelos de credit scoring são muito utilizados pelas instituições financeiras, que procuram maneiras de reduzir o risco em uma concessão de crédito. As técnicas metaheurísticas PSO e GA mostraram-se eficazes na elaboração de tais modelos, pois tiveram um grande percentual de acertos.

Quando se define um problema de elaboração de modelos de credit scoring, torna-se necessária a preparação da base de dados (histórico de clientes) e a adequação (refinamento) do método a ser utilizado. Entretanto, se as dimensões do problema forem alteradas com a re-categorização da base de dados, ou até mesmo a sua substituição, torna-se necessário novo ajuste de todos os parâmetros dos algoritmos.

Os modelos de credit scoring necessitam de uma base de dados histórica de clientes. As características sócio-econômicas presentes em uma cartela de clientes podem ser alteradas devido a diversos fatores. Assim, um histórico de clientes pode não representar o comportamento desse grupo em períodos de tempo muito diferentes.

A utilização de duas técnicas metaheurísticas foi importante na realização desse trabalho, pois permitiu a comparação de resultados e das características presentes

nos dois métodos. O PSO e o GA são metaheurísticas populacionais que melhoram seus indivíduos a cada iteração do algoritmo.

Os melhores resultados são obtidos com o PSO, nas três bases de dados. Essa técnica se mostrou adequada para a resolução do problema proposto.

Em relação ao tempo computacional e ao número de iterações necessárias para a obtenção do melhor resultado, o PSO se mostrou melhor que o GA nas três instâncias utilizadas. A metaheurística PSO é mais eficiente e robusta na elaboração de um modelo de credit scoring do que o GA, conforme a maioria dos resultados.

Conclui-se, pois, que a aplicação das metaheurísticas PSO e GA na elaboração de um problema de credit scoring se mostrou eficiente nas instâncias utilizadas, e seu uso deve ser amplamente difundido.

6.2 **Trabalhos futuros**

Como trabalhos futuros para continuidade desse estudo sugere-se a aplicação de outras metaheurísticas na resolução do problema de credit scoring. Pode-se utilizar a técnica ILS (Iterated Local Search), e um método híbrido, utilizando uma metaheurística populacional juntamente com outra não populacional.

Os modelos híbridos são utilizados com sucesso em diversos trabalhos de otimização. A sua aplicação em modelos de credit scoring pode levar a melhores soluções. Por exemplo, sugere-se o emprego da técnica Colônia de Formigas acompanhado da metaheurística ILS.

Outra proposta de trabalho é utilizar a metaheurística PSO com outra função objetivo, para verificar a eficácia dessa técnica em uma nova formulação matemática.

A utilização de outras metaheurísticas para elaboração de um modelo de credit scoring é relevante, uma vez não se ter encontrado na literatura, grande emprego dessas técnicas.

Referências Bibliográficas

Arnaud, A. L.; Cunha, R. C. L. V.; Ludermir, T. B. e Adeodato, P. J. L. (2005). Modelo híbrido com redes neurais artificiais e técnicas não-supervisionadas para o problema de credit scoring. *XXV Congresso da Sociedade Brasileira de Computação, São Leopoldo, RS.*

Banco Central do Brasil, (). Critérios de classificação das operações de crédito e regras para constituição de provisão para créditos de liquidação duvidosa. Banco Central do Brasil. Resolução 2.682. de 21 de dezembro de 1999. Disponível em: < <https://www3.bcb.gov.br/normativo/detalharNormativo.do?N=099294427&method=detalharNormativo>> Acesso em: 20 out. 2009.

Bezerra, Sinaide Nunes. (2008). Algoritmos evolutivos paralelos aplicados ao problema das p-medianas. Master's thesis, Centro Federal de Educação Tecnológica de Minas Gerais.

Bussab, Wilson O e Morettin, Pedro A. (1987). *Estística Básica*. 4. ed. São Paulo. Atual.

Chaia, Alexandre Jorge. (2003). Modelos de gestão de risco de crédito e sua aplicabilidade ao mercado brasileiro. Master's thesis, Universidade de São Paulo.

Cheng, E. W. L.; Ching, Y. H. e Tang, B. S. (2006). Alternative approach to credit scoring by dea: Evaluating borrowers with respect to pfi projects. *School of Management, Queensland University of Technology, Australia. Department of Building and Real Estate, The Polytechnic University of Hong Kong.*

Corrêa, Marcelo França e Vellasco, Marley. (2008). Análise de risco de crédito em correspondentes bancários através de redes neurais. *Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, PUC-Rio.*

Crouhy, Michel; Galai, Dan e Mark, Robert. (2000). A comparative analysis of current credit risk models. *Journal of Banking & Finance*, v. 24, p. 59–117.

- Dantas, Régis Façanha e Souza, Sérgio Aquino De. (2008). Modelo de risco e decisão de crédito baseado em estrutura de capital com informação assimétrica. *Pesquisa Operacional*, v. 28, p. 263–284.
- Dutra, Marcos Santos. (2008). Uma abordagem alternativa de credit scoring usando análise discriminante: Eficiência na concessão de crédito para o segmento de pessoas físicas no Brasil. Master's thesis, Centro Federal de Educação Tecnológica de Minas Gerais.
- Fernandes, Dimas Tadeu Madeira; deMoura, Heber José; Ponte, Vera Maria Rodrigues; deLuca, Márcia Martins Mendes e Oliveira, Marcelle Colares. (2008). Os impactos da resolução n. 2.682 e dos programas de reestruturação do sistema financeiro nacional no nível de provisionamento da carteira de crédito do setor bancário. *R. Cont. Fin. - USP - São Paulo*, v. 19, p. 44–55.
- Freitas, Maria Cristina Penido. (2009). Os efeitos da crise global no brasil: aversão ao risco e preferência pela liquidez no mercado de crédito. *Estudos Avançados n° 66*, v. 23.
- Gao, Liang; Zhou, Chi; Gao, Hai-Bing e Shi, Yong-Ren. (2006). Credit scoring model based on neural network with particle swarm optimization. *Department of Industrial & Manufacturing System Engineering Huazhong Univ. of Sci. & Tech., Wuhan, 430074, China*.
- García, F. J. M. e Pérez, J. A. M. (2007). Optimización por enjambre para la p-mediana continua y discreta. *Quinto Congreso Español de Metaheurísticas, Algoritmos Evolutivos y Bioinspirados. MAEB 2007, Puerto de La Cruz, Tenerife*.
- Glover, F. (1986). Metaheuristics in combinatorial optimization: Overview and conceptual comparison, acm computing surveys. *citado por Blum C., Roli, A., 2003.*, v. 35, No. 3, p. pp. 268–308.
- Godói, André Cadime; Yoshino, Joe Akira e deDeus Oliveira, Rogério. (2008). Risco de crédito e alocação Ótima para uma carteira de debêntures. *Est. econ., São Paulo*, v. 38, p. 349–372.
- Gomes, Maria Marta Ribeiro Costa. (2008). Aplicação das metaheurísticas pso e ag na estimação de parâmetros de um modelo de secagem em leito fluidizado. Master's thesis, Centro Federal de Educação Tecnológica de Minas Gerais.
- Gonçalves, Eric Bacconi. (2005). Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos. Master's thesis, Universidade de São Paulo.

- Gonçalves, Eric Bacconi e Gouvêa, Maria Aparecida. (2006). Análise de risco de crédito com o uso de modelos de redes neurais e algoritmos genéticos. *XI Semead - Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo (FEA USP)*.
- Huang, Jih-Jeng; Tzeng, Gwo-Hshiung e Ong, Chorng-Shyong. (2005). Two-stage genetic programming (2sgp) for the credit scoring model. *Published by Elsevier Inc.*
- Jiang, Minghui e Yuan, Xuchuan. (2007). Combining forecast model based on pso for personal credit scoring. *School of Management, Harbin Institute of Technology, Harbin, 150001, China.*
- Kennedy, J. e Eberhart, R. C. (1995). Particle swarm optimization. *Proc of the IEEE International Conference on Neural Networks, Piscataway, NJ*, p. 1942–1948.
- Kennedy, J. e Eberhart, R.C. (1997). A discrete binary version of the particle swarm algorithm. *In Proceedings of the Conference on Systems, Man, and Cybernetics*, v. 5, p. 4104–4109.
- Machado, Tiago Capobianco e Neto, Luiz Leduino Salles. (2007). Algoritmo genético no auxílio à decisão de concessão de crédito. *X Encontro de Modelagem Computacional - Instituto Politécnico/UERJ - Nova Friburgo - RJ.*
- Magalhães, Marcos Nascimento e Lima, Antônio Carlos Pedroso. (2001). *Noções de Probabilidade e Estatística*. 3. ed. São Paulo. IME - USP.
- Man, K. F.; Tang, K. S. e Kwong, S. (1986). Genetic algorithms: concepts and applications. *IEEE Transactions on Industrial Electronics*, n. 5, v. 43, p. 519–534.
- Mendes Filho, Elson Felix; de Leon Ferreira deCarvalho, André Carlos Ponce e Matias, Alberto Borges. (2005). Utilização de redes neurais artificiais na análise de risco de crédito a pessoas físicas. *Laboratório de Inteligência Computacional / SCE / ICMSC, USP, Centro de Pesquisas em Crédito e Banking / FEA - RP, USP.*
- Millonas, M. M. (1994). Swarm, phase transitions, and collective intelligence. *In C. G. Langton, Ed. Artificial Life III. Addison Wesley, Reading, MA.*
- Mitchell, M. (1997). An introduction to genetic algorithms. *Cambridge: Mit Press*, p. 1–31, 166–178.
- Osman, I. H. e Laporte, G. (1996). Metaheuristics in combinatorial optimization: Overview and conceptual comparison, acm computing surveys. *citado por Blum C., Roli, A., 2003*, v. 35, No. 3, p. pp. 268–308.

- Picinini, R.; Oliveira, G. M. B. e Monteiro, L. H. A. (2003). Mineração de critério de credit scoring utilizando algoritmos genéticos. *Bauru: VI Simpósio Brasileiro de Automação Inteligente*, p. 463–466.
- Prado, J. R. e Saramago, S. F. P. (2005). Otimização por colônia de partículas. *FAMAT em Revista*, v. 4, p. 87–103.
- Queiroz, Renata S. Bernardes. (2006). A importância dos modelos de credit scoring na concessão de crédito ao consumidor no varejo. *Seminários em Administração FEA-USP. Universidade de São Paulo*.
- Reynolds, W. C. (1987). Flocks, herds, and schools: a distributed behavioral model. *Computer Graphics*, v. 21, n° 4, p. 25–34.
- Santos, J. O e Famá, R. (2007). Avaliação da aplicabilidade de um modelo de credit scoring com variáveis sistêmicas e não-sistêmicas em carteiras de crédito bancário rotativo de pessoas físicas. *R. Cont. Fin. Universidade de São Paulo*, p. 105–117.
- Shi, Y. e Eberhart, R. C. (1998a). A modified particle swarm optimizer. *Proceedings of the IEEE Congress on Evolutionary Computation (CEC), Piscataway, NJ.*, p. 69–73.
- Shi, Y. e Eberhart, R. C. (1998b). Parameter selection in particle swarm optimization. *Evolutionary Programming VII: Proceedings of the Seventh Annual Conference on Evolutionary Programming, New York.*, p. 591–600.
- Silva, G. L. e Oliveira, H. F. (2004). Aplicação da metaheurística algoritmos genéticos na solução do problema das n rainhas. *Faculdade de Viçosa - Semana Acadêmica*.
- Souza, Marcone Jamilson Freitas. (2008). Inteligência computacional para otimização. *Notas de aula, 2008/1 DECOM/ICEB/UFOP - Universidade Federal de Ouro Preto*.
- Steiner, Maria Teresinha Arns; Nievola, Júlio Cesar; Soma, Nei Yoshihiro; Shimizu, Tamio e Neto, Pedro José Steiner. (2007). Extração de regras de classificação a partir de redes neurais para auxílio à tomada de decisão na concessão de crédito bancário. *Pesquisa Operacional, n.3*, v. 27, p. 407–426.
- Suganthan, P. N. (1999). Particle swarm optimizer with neighborhood operator. *Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway. NJ.*, p. 1958–1962.

Trelea, I. O. (2003). The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, n. 6, v. 85, p. 317–325.

Yang, Yingxu. (2007). Adaptive credit scoring with kernel learning methods. *European Journal of Operational Research*, v. 183, p. 1521–1536.

Zuben, F. Von e Castro, L. (2004). Computação evolutiva: Uma "nova" forma de resolver problemas. *DCA/FEEC/Unicamp*.

Apêndice A

Os Melhores pesos obtidos na validação dos modelos de credit scoring

A seguir são apresentados os melhores pesos obtidos com as metaheurísticas PSO e GA, na validação dos Modelos de Credit Scoring.

Tabela A.1: Pesos dos melhores resultados PSO e GA - base de dados I - Método 1

PSO		GA	
X1	0.637348	X1	-0.546200
X2	2.649359	X2	-0.595200
X3	0.711369	X3	-0.315600
X4	1.264048	X4	-0.358800
X5	-4.285643	X5	0.066400
X6	5.742401	X6	0.835800
X7	-1.109153	X7	-0.179200
X8	-1.881922	X8	0.074000
X9	6.198573	X9	-0.958200
X10	-4.082839	X10	-0.936600
X11	-1.425005	X11	0.188600
X12	6.362731	X12	0.428200
X13	1.054134	X13	0.484800
X14	0.323827	X14	0.120200
X15	-4.203790	X15	-0.236000
X16	-7.507710	X16	-0.773600
X17	-0.908313	X17	0.069000
X18	-2.686208	X18	-0.162400
X19	-0.626079	X19	-0.183600
X20	-1.246310	X20	-0.260600
X21	-9.749352	X21	-0.926600
X22	9.245595	X22	0.927800
X23	-1.130821	X23	-0.868000
X24	4.694996	X24	0.607200
X25	3.619385	X25	0.206600
X26	2.933765	X26	0.310000
X27	1.038770	X27	-0.033000
X28	3.219619	X28	-0.623000
X29	-1.243681	X29	0.413600
X30	0.138868	X30	-0.605400
X31	2.676125	X31	0.043400
X32	-4.156753	X32	-0.909600
X33	0.154504	X33	-0.272600
X34	-0.079178	X34	-0.434400
X35	0.870602	X35	0.249200
X36	0.006267	X36	-0.256800
X37	2.133820	X37	0.390800
X38	0.311130	X38	0.289000
X39	-1.065542	X39	0.262400
X40	0.298187	X40	0.488200
X41	0.086137	X41	-0.057200
X42	1.921369	X42	0.958200

Tabela A.2: Pesos dos melhores resultados PSO e GA - base de dados I - Método 2

PSO		GA	
X1	0.677034	X22	1.664684
X2	-0.280659	X23	-0.210019
X3	0.017712	X24	0.485297
X4	-0.014059	X25	0.391171
X5	0.253076	X26	0.850338
X6	-0.391382	X27	-0.326566
X7	0.211538	X28	0.141971
X8	0.175931	X29	-0.508984
X9	0.124801	X30	-0.292204
X10	0.749324	X31	-0.482534
X11	0.239974	X32	-0.185674
X12	-0.200559	X33	-0.359715
X13	0.030072	X34	-0.291756
X14	-0.132663	X35	-0.575040
X15	-0.111580	X36	-0.025524
X16	-0.252611	X37	0.591228
X17	0.150514	X38	0.189546
X18	0.365277	X39	-0.009609
X19	0.390000	X40	0.383675
X20	-0.028563	X41	-0.957274
X21	-1.345331	X42	-0.517816
X1	0.300000	X22	0.986400
X2	0.322400	X23	-0.413600
X3	-0.369600	X24	0.019000
X4	0.351400	X25	-0.960000
X5	0.179800	X26	-0.250000
X6	0.475200	X27	-0.506000
X7	-0.199600	X28	0.356400
X8	-0.413200	X29	0.414200
X9	0.143200	X30	-0.030000
X10	0.816800	X31	0.535000
X11	-0.651400	X32	0.023200
X12	0.758200	X33	-0.613800
X13	0.625200	X34	0.604800
X14	0.115600	X35	0.940200
X15	-0.476600	X36	0.084200
X16	-0.379400	X37	0.618200
X17	-0.193000	X38	0.781400
X18	-0.176200	X39	-0.024200
X19	0.207600	X40	0.014200
X20	-0.288200	X41	-0.827800
X21	-0.995400	X42	-0.051200

Tabela A.3: Pesos dos melhores resultados PSO e GA - base de dados II - Método 1

PSO		GA	
X1	0.218291	X32	3.348543
X2	-2.690639	X33	0.487793
X3	-1.564729	X34	-1.392026
X4	-2.949708	X35	0.416218
X5	-1.318824	X36	0.728309
X6	1.969161	X37	-2.429309
X7	-0.258245	X38	0.647076
X8	5.983586	X39	8.805505
X9	-2.377249	X40	14.338549
X10	0.566577	X41	10.650045
X11	-6.315282	X42	-13.868723
X12	-1.871278	X43	12.186925
X13	-7.032800	X44	-5.357217
X14	-10.950037	X45	1.141924
X15	-0.967122	X46	1.504715
X16	-5.320926	X47	-1.853845
X17	9.588826	X48	0.603949
X18	2.411966	X49	1.866857
X19	-0.967093	X50	3.504138
X20	-4.944297	X51	3.841527
X21	-3.429199	X52	0.242510
X22	-0.542960	X53	-1.692419
X23	3.333492	X54	-2.563230
X24	-3.517888	X55	-1.781085
X25	-1.635316	X56	5.267126
X26	-0.317015	X57	-3.567424
X27	7.237076	X58	2.535526
X28	8.067381	X59	-1.197502
X29	-0.168128	X60	3.038611
X30	-4.156605	X61	3.951250
X31	6.545046	X62	0.642552
X1	0.538200	X32	0.237600
X2	0.289200	X33	0.667200
X3	-0.425600	X34	-0.901800
X4	0.113200	X35	-0.543000
X5	-0.971200	X36	0.006800
X6	0.304600	X37	0.234000
X7	0.155200	X38	-0.226600
X8	-0.793200	X39	0.218400
X9	-0.924600	X40	0.265800
X10	-0.606800	X41	0.504400
X11	0.392000	X42	-0.889400
X12	0.677000	X43	0.976200
X13	0.399000	X44	-0.814600
X14	-0.554200	X45	0.744600
X15	-0.480400	X46	-0.409000
X16	0.716200	X47	0.275600
X17	0.563200	X48	0.757800
X18	0.881000	X49	-0.390600
X19	0.269200	X50	-0.349600
X20	-0.695000	X51	-0.086200
X21	-0.047000	X52	-0.574800
X22	-0.653600	X53	-0.296600
X23	-0.987600	X54	0.739000
X24	-0.785800	X55	-0.624400
X25	-0.190600	X56	0.386400
X26	0.840800	X57	-0.159400
X27	0.760800	X58	0.066000
X28	0.741600	X59	-0.915600
X29	0.241800	X60	-0.287800
X30	-0.137400	X61	0.966200
X31	-0.845200	X62	0.591600

Tabela A.4: Pesos dos melhores resultados PSO e GA - base de dados II - Método 2

PSO		GA					
X1	-0.164290	X32	1.088339	X1	0.462400	X32	0.642400
X2	0.163138	X33	-0.249409	X2	0.449000	X33	0.888200
X3	-0.503304	X34	0.507838	X3	-0.133000	X34	0.477400
X4	-0.257569	X35	0.476320	X4	0.128200	X35	-0.954600
X5	-0.163535	X36	-0.179769	X5	-0.918400	X36	-0.537000
X6	0.209054	X37	-0.523009	X6	0.270600	X37	0.042000
X7	0.127047	X38	-0.010136	X7	0.068200	X38	-0.654200
X8	0.962096	X39	0.425988	X8	-0.626000	X39	-0.849000
X9	-1.373046	X40	0.736373	X9	-0.715000	X40	0.138600
X10	-0.000451	X41	1.312756	X10	-0.550200	X41	0.809600
X11	0.178135	X42	-2.353358	X11	-0.960400	X42	-0.943000
X12	-0.065382	X43	2.313549	X12	-0.600800	X43	0.784200
X13	0.890084	X44	-0.642942	X13	0.673400	X44	-0.323400
X14	0.003393	X45	-0.260208	X14	-0.376000	X45	0.833400
X15	0.460996	X46	-1.330881	X15	-0.092800	X46	-0.090800
X16	0.149577	X47	0.155143	X16	0.271800	X47	0.745000
X17	0.963227	X48	-0.203091	X17	0.298400	X48	0.713600
X18	0.327057	X49	0.431086	X18	-0.325200	X49	0.760400
X19	-0.287570	X50	-0.448056	X19	-0.887000	X50	0.379400
X20	1.105548	X51	0.475313	X20	-0.968600	X51	0.040600
X21	-0.152878	X52	-0.110462	X21	0.824800	X52	0.079000
X22	1.033135	X53	-0.195907	X22	-0.413400	X53	0.192400
X23	-0.686231	X54	-0.721940	X23	-0.676200	X54	0.513200
X24	-0.091867	X55	0.750624	X24	-0.915400	X55	0.001000
X25	0.592375	X56	-0.063161	X25	-0.088800	X56	-0.056600
X26	0.230726	X57	-1.137370	X26	-0.849200	X57	-0.514000
X27	1.315123	X58	-0.687987	X27	0.870600	X58	0.099600
X28	-0.773092	X59	-0.885572	X28	0.532800	X59	-0.641800
X29	-0.974443	X60	1.540029	X29	-0.770600	X60	0.166200
X30	0.359771	X61	0.284146	X30	-0.126400	X61	0.458800
X31	-1.100103	X62	1.944801	X31	-0.232600	X62	0.305200

Tabela A.5: Pesos dos melhores resultados PSO e GA - base de dados III - Método 1

PSO		GA					
X1	0.048242	X21	-0.152125	X1	-0.320400	X21	-0.153000
X2	0.236342	X22	-1.245576	X2	0.229800	X22	-0.559600
X3	0.284668	X23	0.091781	X3	-0.476200	X23	-0.041200
X4	0.117657	X24	0.094186	X4	-0.781200	X24	-0.362600
X5	-0.002915	X25	1.623882	X5	-0.459400	X25	0.992800
X6	1.806979	X26	0.913929	X6	0.974600	X26	0.789000
X7	-0.983142	X27	-0.501546	X7	-0.697800	X27	-0.011600
X8	-0.739687	X28	-0.403101	X8	-0.812600	X28	-0.111400
X9	-0.508201	X29	-0.222617	X9	-0.337400	X29	0.039400
X10	1.211121	X30	-1.012773	X10	0.915800	X30	-0.218800
X11	1.503200	X31	0.281852	X11	0.623200	X31	0.458600
X12	0.305223	X32	0.057468	X12	0.483800	X32	0.135200
X13	0.745375	X33	1.971285	X13	0.433600	X33	0.917600
X14	0.402211	X34	0.097055	X14	0.455600	X34	-0.850200
X15	0.249158	X35	-0.034431	X15	0.552000	X35	-0.343400
X16	0.598264	X36	0.151223	X16	0.721800	X36	-0.968400
X17	-1.288532	X37	-0.236888	X17	-0.845800	X37	0.584600
X18	0.073063	X38	-0.032360	X18	0.049800	X38	0.392000
X19	-0.079543	X39	-0.321045	X19	-0.338600	X39	0.349800
X20	-0.432231	X40	-0.187238	X20	-0.441800	X40	0.526400

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)