**FIAP - FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA**

**RODRIGO DE CARO**

**VINICIUS MATOS ANDRADE**

**WILLIAM RODRIGUES DA SILVA**

**SCSIA – SISTEMA DE CORRETAGEM DE SEGUROS BASEADO EM INTELIGÊNCIA ARTIFICIAL**

**SÃO PAULO - SP**

**2011**

**FIAP - FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA**

**RODRIGO DE CARO**

**VINICIUS MATOS ANDRADE**

**WILLIAM RODRIGUES DA SILVA**

**SCSIA – SISTEMA DE CORRETAGEM DE SEGUROS BASEADO EM INTELIGÊNCIA ARTIFICIAL**

|  |  |
| --- | --- |
|  | Monografia apresentada para obtenção do Título de bacharel pelo Curso de Graduação de Sistemas de Informação da FIAP. |

**Orientador: Prof. Dr. César da Costa**

**SÃO PAULO - SP**

**2011**

Dedicamos este trabalho à nossa família por nos ter apoiado o período do curso, pela paciência infinda, compreensão, incentivo, energia e confiança, sem os quais o desafio de escrever este trabalho, não seria por nós superados.

**AGRADECIMENTOS**

Este trabalho estaria incompleto sem a menção a pessoas, cujo apoio e incentivo influenciaram-nos de vários modos durante a sua elaboração.

Ao professor Dr. César Augusto, coordenador do curso de Sistemas de Informação da FIAP, por colocar à disposição a estrutura de laboratórios da FIAP para o desenvolvimento do trabalho.

Ao Professor Dr. César da Costa pela orientação e apoio fornecido durante a elaboração deste trabalho.

Ao Sr. Mario Loschiavo, corretor de seguros da Loschiavo Corretora Seguros LTDA, pelo apoio com informações sobre seguros.

Ao colega de classe Bruno Carletti, pelas informações sobre atuais sistemas de seguradoras.

Aos nossos pais e irmãos pelo amor e dedicação em nossa formação.

Aos amigos e companheiros de graduação que, de uma forma direta ou indireta, contribuíram para que este trabalho se concretizasse.

**RESUMO**

Este trabalho apresenta uma proposta para a implementação de um sistema para a corretagem de seguros baseado em Inteligência Artificial em seguradoras, o qual utiliza na sua arquitetura uma rede neural perceptron simples, um banco de dados e processos para qualificar os dados. A utilização de Inteligência Artificial visa melhorar a definição de grau de risco do segurado. Para validar a proposta apresentada neste trabalho, foi implementado um sistema capaz de determinar o grau de risco que o perfil desejado irá apresentar, bem como diminuir a manutenção periódica no sistema e consequentemente gerando menos gastos para a seguradora. Os resultados satisfatórios obtidos nos ensaios práticos realizados com esse sistema, associados com as metodologias empregadas no desenvolvimento da rede neural e estrutura do sistema, mostram que essa proposta é viável e consistente, principalmente no que concerne a implementação de um sistema que pode ser treinado ao longo de sua utilização para situações não previstas e quanto maior for o histórico de clientes, maior será a precisão obtida em seus resultados. Outra vantagem obtida foi a diminuição das manutenções complexas e frequentes que os sistemas atuais de seguradoras necessitam.

Palavras-chave: *Inteligência Artificial,* corretagem de seguros e Redes Neurais artificiais.

**ABSTRACT**

This paper presents a proposal to implement a system for the insurance brokerage based on Artificial Intelligence in insurance, which in its architecture uses a simple perceptron neural network, a database and processes the data to qualify. The use of Artificial Intelligence aims to improve the definition of the riskiness of the insured. To validate the proposal presented in this paper, we implemented a system capable of determining the degree of risk that will provide the desired profile, and decrease the regular maintenance on the system and consequently generating less expense to the insurer. The satisfactory results obtained in field trials conducted with this system, associated with the methodologies employed in the development of neural network and system structure, show that this proposal is viable and consistent, especially regarding the implementation of a system that learns about himself and greater the customer history, the greater the accuracy achieved in its results. Another advantage was obtained by the reduction of complex and frequent maintenance than current systems require insurers.

Keywords: Artificial Intelligence, insurance brokerage and artificial neural networks.**LISTA DE FIGURAS**

[Figura 1 - Neurônio Biológico 23](#_Toc311146332)

[Figura 2 - Neurônio Artificial 25](#_Toc311146333)

[Figura 3 - Processo de KDD (Descoberta de Conhecimento do Banco de Dados) 30](#_Toc311146336)

[Figura 4 - Arquitetura do sistema 36](#_Toc311146337)

[Figura 5 - Estrutura das tabelas do sistema 37](#_Toc311146338)

[Figura 6 - Treinamento da rede 45](#_Toc311146351)

[Figura 7 - Cotação do segurado 47](#_Toc311146352)

[Figura 8 - Definição do grau de risco dos perfis 48](#_Toc311146353)

[Figura 9 - Definição do peso para os atributos 48](#_Toc311146354)

[Figura 10 - Resultado da cotação 49](#_Toc311146355)

[Figura 11 - Cotação de Vinicius Matos Andrade 50](#_Toc311146356)

[Figura 12 - Resultado de Vinicius Matos Andrade 51](#_Toc311146357)

[Figura 13 - Cotação de William Rodrigues da Silva 52](#_Toc311146358)

[Figura 14 - Resultado de William Rodrigues da Silva 52](#_Toc311146359)

[Figura 15 - Cotação de Rodrigo de Caro 53](#_Toc311146360)

[Figura 16 - Resultado de Rodrigo de Caro 54](#_Toc311146361)

[Figura 17 - Cotação de Demétrio dos Santos Andrade 55](#_Toc311146362)

[Figura 18 - Resultado de Demétrio dos Santos Andrade 55](#_Toc311146363)

[Figura 19 - Cotação de Moises Rodrigues da Silva 56](#_Toc311146364)

[Figura 20 - Resultado de Moises Rodrigues da Silva 57](#_Toc311146365)

[Figura 21 - Cotação de Sidneia Oliva de Almeida 58](#_Toc311146366)

[Figura 22 - Resultado de Sidneia Oliva de Almeida 58](#_Toc311146368)

[Figura 23 - Cotação de Érika Matos Andrade 59](#_Toc311146369)

[Figura 24 - Resultado de Érika Matos Andrade 60](#_Toc311146370)

[Figura 25 - Cotação de Marlene Rosa de Matos Andrade 61](#_Toc311146371)

[Figura 26 - Resultado de Marlene Rosa de Matos Andrade 61](#_Toc311146372)

[Figura 27 - Cotação de Euclides dos Santos Andrade 62](#_Toc311146373)

[Figura 28 - Resultado de Euclides dos Santos Andrade 63](#_Toc311146374)

[Figura 29 - Cotação de Beatriz Valera Capitina 64](#_Toc311146375)

[Figura 30 - Resultado de Beatriz Valera Capitina 64](#_Toc311146376)

**LISTA DE EQUAÇÕES**

[Equação 1 – Equação de McCulloch e Pitts 27](#_Toc311145967)

[Equação 2 – Função de ativação 27](#_Toc311145968)

**LISTA DE TABELAS**

[Tabela 1 – Aplicações de Data Mining 35](#_Toc311146181)

[Tabela 2 - Dados extraídos para o banco de dados 38](file:///C:\Users\vmandr\Desktop\%5bTCC%5d%20-%20Documentação%20-%20Monografia(Revisão).docx#_Toc311146182)

[Tabela 3 - Exemplo de codificação dos dados de acordo com a figura 39](#_Toc311146183)

[Tabela 4 - Tabela de idade 40](#_Toc311146184)

[Tabela 5 – Tabela do valor do carro 40](#_Toc311146185)

[Tabela 6 – Tabela de bairros. 41](#_Toc311146186)

[Tabela 7 - Tabela de classificação de sinistros 41](#_Toc311146187)

[Tabela 8 - Tabela de tempo de habilitação 42](#_Toc311146188)

[Tabela 9 - Tabela de classificação de profissão 42](#_Toc311146189)

[Tabela 10 - Tabela de grau de instrução 43](#_Toc311146190)

[Tabela 11 - Tabela de classificações de carros 43](#_Toc311146191)

[Tabela 12 - Tabela de marcas de carros do banco de dados 44](#_Toc311146192)

[Tabela 13 - Tabela de cursos de formação 44](#_Toc311146193)

**SUMÁRIO**

[CAPÍTULO 1 - INTRODUÇÃO 12](#_Toc349257164)

[1.1 Considerações Gerais 12](#_Toc349257165)

[1.2 Objetivos Gerais 13](#_Toc349257166)

[1.2.1 Objetivos Específicos 14](#_Toc349257167)

[1.3 Justificativa 14](#_Toc349257168)

[1.4 Metodologia 16](#_Toc349257169)

[1.5 Estrutura do Trabalho 17](#_Toc349257170)

[CAPÍTULO 2 - FUNDAMENTAÇÃO TEÓRICA 19](#_Toc349257171)

[2.1 Ambiente empresarial na era da informação 19](#_Toc349257172)

[2.2 A história das redes neurais 20](#_Toc349257173)

[2.2.1 Aplicação de redes neurais 22](#_Toc349257174)

[2.2.2 Redes Neurais Biológicas X Redes Neurais 23](#_Toc349257175)

[2.2.2.1 O funcionamento do neurônio artificial 25](#_Toc349257176)

[2.2.2.2 Perceptron Multicamadas 28](#_Toc349257177)

[2.3 Data Mining 28](#_Toc349257178)

[2.3.1 Definição 28](#_Toc349257179)

[2.3.2 Weka 33](#_Toc349257180)

[2.3.3 Aplicações 35](#_Toc349257181)

[CAPÍTULO 3 - DESENVOLVIMENTO DO TRABALHO 36](#_Toc349257182)

[3.1 Sistema SCSIA 36](#_Toc349257183)

[3.1.1 Módulo de banco de dados 37](#_Toc349257184)

[3.1.1.1 Estrutura do banco de dados 37](#_Toc349257185)

[3.1.2 Tratamento dos dados utilizados para entrada da rede 38](#_Toc349257186)

[3.1.2.1 Codificação dos dados 39](#_Toc349257187)

[3.1.3 Módulo de redes neurais 44](#_Toc349257188)

[3.1.4 Descritivo operacional do sistema 45](#_Toc349257189)

[3.1.4.1 Treinamento da rede 45](#_Toc349257190)

[3.1.4.2 Cotação do segurado 46](#_Toc349257191)

[3.1.4.3 Definição do grau de risco dos perfis 47](#_Toc349257192)

[3.1.4.4 Definição do peso dos atributos 48](#_Toc349257193)

[3.1.4.5 Resultado da cotação 49](#_Toc349257194)

[CAPÍTULO 4 - ENSAIOS PRÁTICOS 50](#_Toc349257195)

[4.1 Discussão dos resultados 65](#_Toc349257196)

[CAPÍTULO 5 - CONCLUSÕES 66](#_Toc349257197)

[5.1 Trabalhos futuros 67](#_Toc349257198)

[REFERÊNCIAS BIBLIOGRÁFICAS 68](#_Toc349257199)

[APENDICE 69](#_Toc349257200)

[Manual do sistema 69](#_Toc349257201)

# INTRODUÇÃO

Este capítulo é constituído pelas considerações gerais, pela proposta sugerida pelo SCSIA como objetivos gerais, pelos objetivos específicos do trabalho, por uma justificativa e a metodologia utilizada para o desenvolvimento do trabalho. Nas considerações gerais, é apresentada a atual forma de classificação do grau de risco de sinistro do perfil de segurados. Nos objetivos gerais, é apresentada uma nova solução que realizará esta classificação dos perfis baseada em Inteligência Artificial. No item relativo aos objetivos específicos são apresentadas as funções que o sistema deverá disponibilizar para os usuários. Na parte reservada à justificativa, os motivos pela realização deste trabalho e suas vantagens se implantado nas seguradoras. Na metodologia, é apresentada a forma de desenvolvimento do trabalho. No item referente à estrutura do trabalho é mostrado de forma resumida o conteúdo de cada capítulo.

## Considerações Gerais

No século XXI, com a atual tecnologia disponível no mercado, as empresas de diversos setores se tornaram muito competitivas. A inovação de sistemas, e aperfeiçoamento dos processos operacionais e internos das empresas está criando uma cultura de que a Tecnologia da Informação é realmente importante para o negócio e finalmente começando a tratá-la como um centro de lucros. Em outros países, empresas já consideram a Tecnologia da Informação como um centro de lucros para a empresa e acabam investindo para que o negócio da empresa seja mais lucrativo, disponível e seguro.

As empresas utilizam a Tecnologia da Informação para capturar, armazenar e organizar a grande massa de dados que são resultados de suas operações diárias de todos os departamentos. Porém esta grande massa de dados ainda não está disponibilizando todas as informações necessárias para que a equipe estratégica da empresa possa prever movimentações de mercado, ajustar os processos da empresa e planejar a estratégia que será aplicada em determinado local ou tempo.

Atualmente por exemplo, as seguradoras não possuem recursos suficientes para determinar com alta precisão o grau de risco em que cada perfil de usuário se enquadra. A forma que estas seguradoras utilizam hoje para determinar o grau de risco é baseada apenas em estatísticas simples e pesquisas de mercado, muitas vezes não levando em consideração atributos que poderiam influenciar neste cálculo. Desta forma, estas seguradoras possuem sistemas que realizam apenas contas baseada em fórmulas e consequentemente acabam tornando a cotação de seguro “tabelada”.

A tecnologia atual oferece suporte para que o cálculo de seguros seja mais coerente. Os sistemas baseados em Inteligência Artificial, Data Mining e outras tecnologias que simulam o cérebro humano seriam mais precisas e suscetíveis a atualizações em tempo real. Sistemas baseados nestas tecnologias fornecem suporte para que o cálculo do seguro seja mais flexível e preciso. Usando Inteligência Artificial, o sistema passará por processos de aprendizagem em sua rede para que adquira cada vez mais conhecimento e ofereça ao usuário informações mais precisas, conforme as informações inseridas para a definição do grau de risco do segurado e o histórico de clientes cadastrados no sistema.

## Objetivos Gerais

O objetivo deste projeto é desenvolver um sistema que irá auxiliar a corretagem de seguros, possibilitando por meio de técnicas de Inteligência Artificial, definir o perfil de risco do cliente segurado. A partir dos atributos inseridos no sistema com seu respectivo peso, o sistema irá realizar um cálculo que resultará em uma porcentagem, que indicará a chance de ocorrência de sinistro para determinado perfil. Este resultado será produzido baseado nas entradas que o usuário inserir no sistema a partir do histórico de cotações anteriores. Sendo assim, o valor desta porcentagem irá variar de acordo com os atributos e pesos inseridos pelo usuário.

### Objetivos Específicos

Como objetivo específico busca-se através deste projeto de pesquisa:

* Estabelecer os principais atributos que deve ser utilizado pelas seguradoras na cotação de seguros, como idade, modelo do carro, local de residência, tempo de carta, incidentes ou ocorrências anteriores;
* Identificar o perfil do segurado que possui maior probabilidade de acionar o seguro.
* Identificar os benefícios que esta solução trará às seguradoras.
* Identificar novos perfis baseados no histórico de clientes anteriores;
* Disponibilizar a probabilidade de determinado cliente acionar o seguro.

## Justificativa

Tendo como base os objetivos apresentados, este projeto, se implantado em seguradoras de automóveis, contribuirá para que os custos da seguradora relacionados a sinistros sejam minimizados, pois o valor cobrado do segurado será condizente com os custos que o mesmo irá gerar para a seguradora.

Existem casos em que a seguradora é prejudicada por não conseguir identificar com exatidão o perfil do segurado, sendo assim, é cobrado um valor que não supre os gastos gerados pelo segurado. Em outros casos a seguradora deixa de conquistar novos clientes por não possuir parâmetros para definir se estes irão beneficiar a seguradora, tanto em termos financeiros como agregação na base de conhecimento.

Em alguns sistemas de seguradoras, o cálculo da cotação do seguro é baseado em tabelas que são atualizadas periodicamente. Existe uma fórmula que calcula os atributos do segurado, multiplicando estes atributos por pesos pré-definidos em tabelas, e esta fórmula resultará em uma variável final que será a base de cálculo para o preço do seguro.

A proposta é que o SCSIA faça com que os pesos inseridos pelo usuário, juntamente com os valores dos atributos, resulte em uma variável que indique o grau de risco de abertura de sinistro. Estes pesos serão atualizados conforme a rede neural for treinada, sendo assim não será necessária a manutenção periódica do sistema.

O desenvolvimento deste trabalho é voltado para empresas que realizam este tipo de serviço, estimulando-as a ter mais clientes e diminuir os seus custos, proporcionando um valor mais adequado para cada tipo de perfil dos clientes. O cliente por sua vez, verificando que o seguro de seu carro está realmente de acordo com seu perfil, definirá se é de interesse ou não realizar o seguro do carro.

O momento para a implantação deste projeto é oportuno, tendo em vista o aumento nas compras de automóveis devido a facilidade de crédito para financiamentos e a entrada de montadoras estrangeiras no país, sendo assim, quanto maior o número de automóveis em circulação, maior será a procura destes dos serviços oferecidos pelas seguradoras.

Uma vez simulando o cérebro humano, será extremamente interessante utilizar cálculos mais precisos para que através deste meio, as seguradoras possam fidelizar mais clientes e provavelmente ter menores custos em casos de abertura de sinistro. A maioria das seguradoras não possui um sistema baseado em Inteligência Artificial que irá simular o cérebro humano conforme o aprendizado do sistema.

É de interesse das seguradoras saber qual a chance do cliente acionar o seguro para propor um valor adequado que não gere prejuízos futuramente, sendo que uma vez ocorrido este cenário, ambas as partes serão beneficiadas.

## Metodologia

O estudo baseia-se em pesquisa qualitativa com estudos do perfil de cada segurado, por meio de análise de uma base de dados contendo as informações históricas de clientes da seguradora, dados que foram levantados no momento do cadastro para ser realizada a cotação do seguro.

Por meio dos dados coletados buscam-se elencar estratégias mais adequadas para encontrar um valor mais justo para a cotação e atrair um público maior para este tipo de consumo, e consequentemente estruturar os objetivos e planejar as ações para expansão do mercado.

É necessária a pesquisa de soluções similares, como os atuais sistemas de seguradoras baseados em estatísticas que atuam no mercado da corretagem de veículos automotores. Para isso foram realizadas pesquisas com colegas de faculdade que no momento trabalham com soluções para sistemas de seguradoras, pessoas que atuam trabalhando com corretagem de seguros para veículos.

Para atingir o objetivo do trabalho, a identificação das necessidades para o desenvolvimento e implantação do SCSIA, a técnica para a coleta de dados foi a entrevista individual, realizada com alguns corretores de seguros, usuários de softwares de corretoras de seguros, e desenvolvedores destes sistemas, para que pudessem contribuir com suas respectivas experiências na área de corretagem de seguros para veículos.

Para adquirir o conhecimento teórico para o desenvolvimento do sistema, foram compreendidos os conceitos através de livros, artigos de internet e aulas na faculdade no curso de Sistemas de Informação e também para o conhecimento de tecnologias aplicadas ao sistema, buscando orientações de pessoas que trabalham com as tecnologias usadas no sistema além de nossa experiência com as mesmas.

## Estrutura do Trabalho

O **capítulo 1** é composto por: considerações gerais no ambiente das seguradoras de veículos no Brasil explicando a atual deficiência dos sistemas baseados em estatísticas e tabelas; proposta do SCSIA e seus objetivos; justificativa do desenvolvimento do trabalho e a metodologia descrita de acordo com a forma de desenvolvimento deste trabalho apontando os métodos utilizados para a realização do mesmo.

No **capítulo 2**, é abordado sobre as empresas na era da informação, seguido pela definição de rede neural, árvore de decisão e concluindo com Data Mining. Estes tópicos têm como objetivo dar uma breve introdução para que o leitor consiga compreender os termos e conceitos utilizados no desenvolvimento do trabalho e para o entendimento do algoritmo do sistema.

No **capítulo 3**, é relatado sobre o desenvolvimento do trabalho, mostrando a solução que o sistema trará, e imagens de telas de funcionamento do sistema, características que o sistema terá e as funcionalidades, também é mostrado o algoritmos usados no código fonte, detalhado os módulos do sistema mostrando suas respectivas funções e o funcionamento de cada módulo como, por exemplo, o funcionamento do módulo de treinamento da rede, detalhando a operação do sistema.

O **capítulo 4** apresenta o ensaio prático do sistema com alguns cadastros em seu banco de dados mostrando os resultados obtidos através do sistema, estes testes serão baseados em perfis reais de segurados. O objetivo dessa fase é verificar o comportamento de todo o sistema para validar a implantação de Inteligência Artificial nos sistemas de seguradoras de veículos.

No **capítulo 5** são mostradas as conclusões do trabalho e algumas indicações para trabalhos futuros.

# FUNDAMENTAÇÃO TEÓRICA

Este capítulo é dividido em tópicos que compõem a revisão bibliográfica sobre os principais assuntos do tema deste trabalho, está dividido em três partes.

Primeiro contendo o ambiente empresarial na era da informação, seguido da definição de rede neural e data mining.

## Ambiente empresarial na era da informação

Segundo TURBAN (2003) a era da informação tornou o mercado competitivo repleto de desafios que esta em constante mudança, tornando um ambiente complexo, globalizado e voltado para o cliente, dificultando assim a concorrência de uma empresa sem uso da tecnologia, isso devido às empresas necessitarem reagir rapidamente aos problemas e oportunidades que surge nesse ambiente empresarial moderno, os fatores que tem maior influencia sobre essas variações são de origem social, jurídico, econômico, físico e político. A maior pressão que as empresas sofrem é em produzir mais utilizando menos recursos, ou seja, aumentar sua receita diminuindo seu investimento.

De acordo o mesmo autor, com esse ambiente cada vez mais veloz, as empresas necessitam diminuir o intervalo entre a ocorrência de um evento comercial até o momento em que as informações obtidas sobre este cheguem aos responsáveis pelas tomadas de decisão. Essa diminuição de tempo é obtida através de tecnologias de telecomunicação e sistemas de apoio a decisão cada vez mais avançadas. Os processos e transações lentos, que envolviam relatórios em papel e correio tradicional, já não fazem mais parte do cenário da maioria das empresas.

Ainda para TURBAN (2003), conforme o tempo passa, os clientes se tornam mais exigentes, buscando mais detalhes sobre o produto, como características, garantia oferecida, entre outras, e desejam tudo isso de forma muito rápida, com esse cenário as empresas necessitam oferecer informações rápidas e de qualidade ou correm o risco de perder para a concorrência.

Outro fator importante é a questão da personalização dos produtos buscando atender as exigências e necessidade de cada cliente, e associado à boa qualidade, preço baixo e de forma rápida, para isso as empresas buscam avanços tecnológicos para conseguir certa vantagem em relação à concorrência.

## A história das redes neurais

Os estudos avançados sobre o tema Redes Neurais foram inicialmente conduzidos por McCulloch e Pitts.

O psiquiatra e neuroanatomista interpretou o neurônio como sendo entradas binarias por somas ponderadas, produzindo uma entrada efetiva, e contou com a união do matemático Pitts, no ano de 1942.

“Em um artigo descrito na época por McCulloch e Pitts descreviam um cálculo lógico das redes neurais que unificava os estudos de neurofisiologia e da lógica matemática. Eles assumiam que o modelo formal de um neurônio seguia a lei do “tudo ou nada”. Como um número suficiente dessas unidades simples em conexões sinápticas ajustadas apropriadamente e operando de forma síncrona,

McCulloch e Pitts mostraram que uma rede assim construída realizaria, a princípio, a computação de qualquer função computável. “Este era um resultado significativo e com ele é geralmente aceito o nascimento das disciplinas de redes neurais e inteligência artificial”. (HAYKIN, 2001, ed. 2, p. 63)

Segundo Hebbs, a conectividade do cérebro é continuamente modificada conforme um organismo vai aprendendo tarefas funcionais diferentes e que os agrupamentos neurais são criados por tais modificações.

Frank Rosenbalt conseguiu reproduzir o perceptron de única camada com conceito de aprendizagem, porém logo após Minsky e Papert com o auxílio da Matemática provaram que redes em única camada continham falhas.

O recomeço da segunda e mais avançada pesquisa a respeito das redes neurais, deu-se por conta de uma rápida paralisação até os anos 80, esta teve como principal motivo a escassez de recursos para o desenvolvimento de novas rotinas e filtros.

John Hopfield um renomado cientista conhecido mundialmente, contribuiu em 1982 com sua pesquisa neuronal que ficou conhecida como rede de Hopfield, esta pesquisa teve como principal fundamento a armazenagem de informações em redes profundas e estáveis, o qual teve uma vasta contribuição para o mundo da neurocomputação artificial.

Um fator que incentivou futuras pesquisas e implementações na área de redes neurais foi a publicação do artigo de Kohonen na década de 80, que mencionava os mapas auto-organizáveis que tinham como referências modelos bidimensionais ou unidimensionais, que proporcionou uma maior relevância por parte da comunidade cientifica da época.

Com a publicação do livro de Braitenberg em 1984 (Vehicles: Experiments in Synthetic Psychology), foi possível direcionar um melhor entendimento sobre o complexo processo o qual defenda o conceito do desempenho auto-organizado, já para Linkser o principio da auto-organização deve-se a uma rede perceptiva.

O método do algoritmo de retro propagação foi desenvolvido por Rumelhart, Hinton e Williams em 1986.

Já na década de 90 pode-se observar que os novos neurocientistas intensificaram seus esforços, para que se fosse possível alcançar resultados cada vez mais próximos do neurônio biológico, aproximando dos resultados dos dias atuais.

### Aplicação de redes neurais

A presença cada vez mais efetiva das redes neurais em nossas vidas se deve ao grande avanço tecnológico, ajudando o homem a atingir seus objetivos com resultados muito melhores e com mais eficiência e rapidez, e estão sendo aplicadas em diversos setores.

Respostas complexas são obtidas de forma imediata através dos sistemas que utilizam redes neurais, utilizando um mapa topológico do cliente é possível rastrear suas informações e analisar seu perfil, sendo mais fácil para rastrear saques de criminosos (saques que fujam do padrão do cliente), conhecer mais o cliente e gerar novas promoções especializadas para cada cliente.

Por gerar informações redundantes devido ao grande número de entradas nos sistemas, e pelo tempo que duraria o ciclo das análises da informação é que os seres humanos que avaliavam estes perfis estão sendo substituídos pelos sistemas de rede neural.

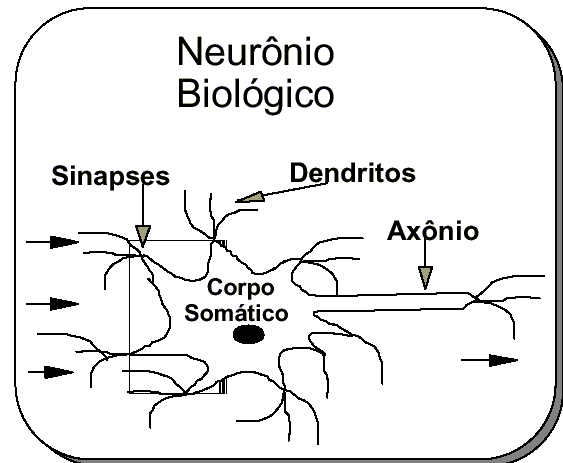
Outro caso muito delicado é o da saúde onde cada segundo pode ser vital para ser salva uma vida, para isso existem softwares que tem a capacidade de identificar e mapear doenças com exatidão, e a uma velocidade muito superior do que a de um ser humano na mesma função.

Mas esses sistemas não substituem um especialista, e sim auxilia na tomada de decisão já que é o especialista que indica como será o padrão usado na identificação, através do treinamento da rede neural, em um futuro breve esses sistemas serão capazes de auxiliar o mapeamento de doenças em tempo real.

### Redes Neurais Biológicas X Redes Neurais

Uma das maiores criações da natureza foi o desenvolvimento do cérebro humano, pois embora haja uma complexidade, existe também uma eficiência natural e dinâmica.

A rede neural biológica é constituída por milhões de unidades de grande importância nomeadas neurônios (conforme detalhado na Figura 1). Os neurônios são compostos por um corpo celular ou soma, com inúmeras ramificações denominadas dendritos, também possui uma longa fibra chamada de axônio, que se ramifica em filamentos e subfilamentos os quais se conectam com dendritos de outros corpos celulares formando assim uma enorme rede de comunicação.



1. Neurônio Biológico

Por meio das reações eletroquímicas, os sinais se disseminam podendo causar a diminuição ou aumento do sinal que é transmitido. Para que a rede se alimente e forme nossos sentidos, é preciso que o sinal atinja um específico limiar de forma que o axônio chegue à extremidade onde existam ligações através de dendritos que se comunicam de modo dinâmico com outros neurônios.

O processamento do cérebro humano é totalmente paralelo, são em média 10 bilhões de neurônios interligados, onde cada um deles desenvolve entre mil e dez mil conexões com neurônios adjacentes.

Como a rede neural artificial projeta de maneira lógica e matemática o funcionamento de um neurônio humano, pode-se dizer que a rede neural artificial não é muito distinta da rede neuronal biológica.

McCulloch e Pitts (MCP) em 1943 sugeriram um modelo neural que simplificava o que já se sabia sobre o neurônio biológico naquele momento.

O conceito matemático que os autores propuseram teve como resultado um modelo matemático com N terminais de entrada (***X1, X2, X3,..., Xn****)*os quais representavam os dendritos, com apenas um terminal de saída Y representando o axônio.

Os terminais de entrada do neurônio possuem pesos acoplados (***W1, W2, W3,..., Wn****)* com valores que podem variar de forma positiva ou negativa, isso pode ocorrer se as sinapses correspondentes forem inibitórias ou excitatórias.

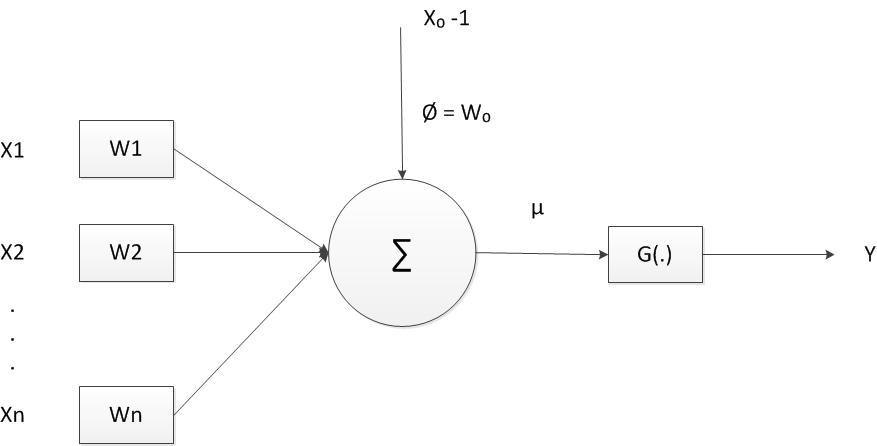
A causa de uma sinapse particular no neurônio pós-sináptico é dada por: ***Xi*** e ***Wi***, dessa forma os pesos irão determinar em que grau o neurônio deve considerar sinais de disparo que ocorrem naquela conexão.

#### O funcionamento do neurônio artificial

O cérebro humano e os modelos conhecidos de sistema nervoso biológico serviram como base para que fosse criada a estrutura das redes neurais artificiais, tomando como base um modelo simplificado dos neurônios biológicos foi possível desenvolver as unidades processadoras e os elementos computacionais, que são denominadas de neurônios artificiais.

A inspiração para esses modelos foi à análise da geração e propagação de impulsos elétricos pela membrana celular dos neurônios biológicos, já os artificiais, utilizados em redes neurais artificiais, são denominados como não lineares, as saídas quase que em sua maioria são contínuas, e utilizam funções simples, como capturar os sinais disponibilizados para suas entradas, agrupá-los de acordo com sua função operacional e produzir uma resposta, levando em consideração sua função de ativação inerente.

Proposto por McCulloch e Pitts (1943) é o modelo de neurônio mais utilizado nas diferentes arquiteturas de redes neurais artificiais conforme apresentado na Figura 2, devido ao seu modelo mais simples e por englobar os fundamentos básicos de uma rede neural biológica que é ter o conceito de paralelismo e de alta conectividade. Esse modelo representado graficamente:



1. Neurônio Artificial

O neurônio artificial de McCulloch e Pitt é constituído por sete elementos básicos:

1 – Sinais de entrada (***X1, X2, X3,..., Xn***).

São entradas de sistemas, que podem ser medidas ou sinais, gerados pelos meios externos que representam os valores atribuídos as variáveis de uma determinada aplicação, geralmente esses sinais são normalizados, visando facilitar e dar mais eficácia computacional aos algoritmos de aprendizagem.

2 – Pesos Sinápticos ***(W1, W2, W3,..., Wx)***.

São valores que servirão para ponderar cada uma das variáveis de entrada da rede, permitindo-se quantificar as suas relevâncias em relação à funcionalidade do respectivo neurônio.

3 – Combinador linear ***(∑)***

Sua função é agregar todos os sinais de entrada que foram ponderados pelos respectivos pesos sinápticos a fim de produzir um valor de potencial de ativação.

4 – Limiar de ativação ***(Ø)***

É uma variável que especifica qual será o patamar apropriado para que o resultado produzido pelo combinador linear possa gerar um valor de disparo em direção à saída do neurônio.

5 – Potencial de ativação ***(µ)***

É o resultado produzido pela diferença do valor produzido entre o combinador linear e o limiar de ativação. Se tal valor é positivo, ou seja, se µ ≥ 0 então o neurônio produz um potencial excitatório, caso contrário, o potencial será inibitório.

6 – Função de ativação ***(G)***

Seu objetivo é limitar a saída do neurônio dentro de um intervalo de valores razoáveis a serem assumidos pela sua própria imagem funcional.

7 – Sinal de saída ***(Y)***

Consiste do valor final produzido pelo neurônio em relação a um determinado conjunto de sinais de entrada, podendo ser também utilizado por outros neurônios que estão sequencialmente interligados.

As duas expressões seguintes sintetizam o resultado produzido pelo neurônio artificial proposto por McCulloch e Pitts.

Equação 1 – Equação de McCulloch e Pitts

Equação 2 – Função de ativação

Assim, resume-se o funcionamento de um neurônio artificial através dos seguintes passos:

* Apresentação de um conjunto de valores, que representam as variáveis de entrada do neurônio.
* Multiplicação de cada entrada do neurônio pelo seu respectivo peso sináptico.
* Obtenção do potencial de ativação produzido pela soma ponderada dos sinais de entrada, subtraindo-se o limiar de ativação.
* Aplicação de uma função apropriada, tendo-se como objetivo limitar a saída do neurônio.
* Compilação da saída a partir da aplicação da função de ativação neural em relação ao seu potencial de ativação.

#### Perceptron Multicamadas

De acordo com Haykin (2001) de uma maneira mais superficial um perceptron multicamadas é constituído de neurônios interconectados dispostos em múltiplas camadas, em que os neurônios de camadas diferentes têm conexões diretas com o neurônio da camada seguinte. Completando Haykin, Braga (2000) cita que esse tipo de rede possui um altíssimo poder computacional já que recebem dados linearmente não separáveis.

Ainda segundo Braga a interligação dos processamentos gerados pelos neurônios das camadas precedentes conectados a um neurônio específico é o que define o processamento realizado por ele.

Para resolver problemas com resoluções muito difíceis o perceptron de multicamadas é um dos mais indicados, através da técnica de treinamento supervisionado a partir de um algoritmo muito conhecido, retro propagação de erro. Denominado como uma espécie de outro algoritmo conhecido como mínimo quadrado médio, a retro propagação de erro baseia-se na regra de aprendizagem por correção de erro.

## Data Mining

### Definição

Nos dias de hoje, com o grande fluxo de dados nas empresas, temos acesso à grande quantidade de informações nos bancos de dados que no dia-a-dia podem não fazer sentido. A verdade é que toda essa quantia de dados tem um valor enorme se conseguirmos trabalhar com a mineração destes dados, tornamos estes bancos de dados, uma fonte de lucros para as empresas. Apesar de serem muito valiosas, as empresas não conseguem aproveitar o máximo destes dados.

Estas informações podem ser utilizadas para melhorar processos que não estão sendo realizados da melhor maneira, consequentemente tendo uma melhor visão das tendências do mercado, características especificas dos clientes e da própria empresa. Do ponto de vista estratégico, temos a probabilidade de ter melhores resultados, uma vez que detectado alguma tendência no mercado, podemos agir com mais precisão e mais estratégia para que tenhamos melhores resultados.

GIGABYTES e TERABYTES são as medições que é constantemente encontrada nos bancos de dados de empresas de diversos setores do mercado. Analisar estas informações manualmente é praticamente impossível, ultrapassa as habilidades técnicas e a capacidade humana. Técnicas computacionais foram criadas para ao menos auxiliar os analistas a encontrar uma forma das informações passarem conhecimento às empresas.

Data Mining é uma tecnologia que se baseia em grandes massas de dados, conseguindo tratar as informações, e tirar desta grande quantidade de dados, informações úteis à empresa.

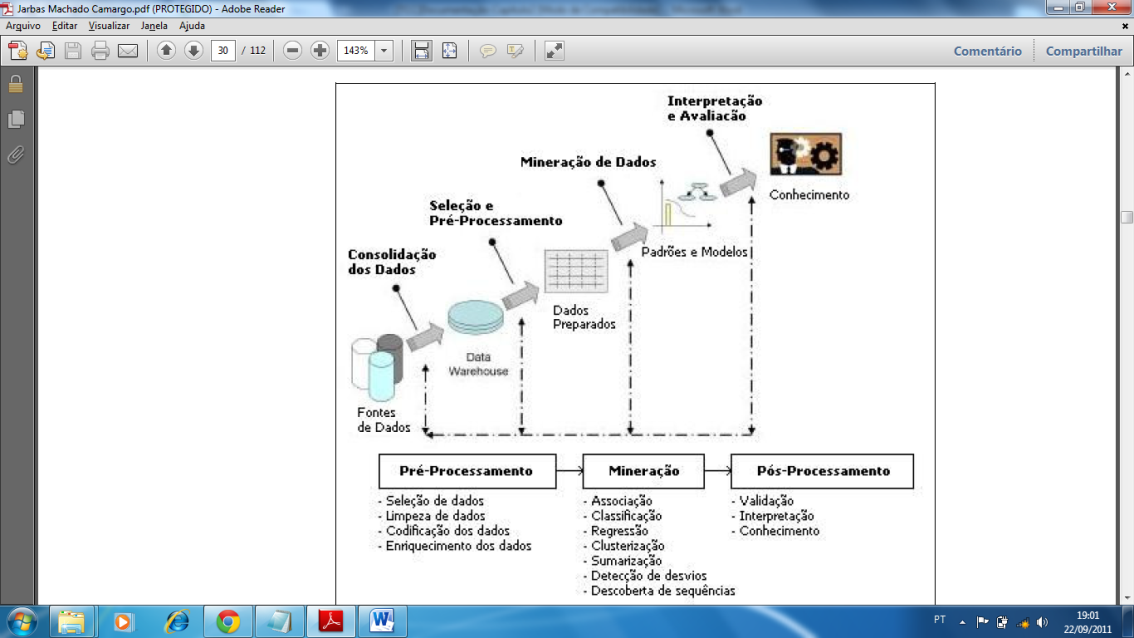
Segundo O’BRIEN (2004), o Data Mining tem como objetivo buscar o conhecimento e auxiliar com base nessas informações a tomada de decisão. Não existe uma área especifica que esta técnica pode ser aplicada. Alguns exemplos de áreas em que aplicamos o Data Mining são: análise de riscos, controle de qualidade, analise de dados científicos. Os colaboradores da parte estratégica das empresas utilizam o Data Mining para tomar decisões nas operações da empresa, visando obter vantagem competitiva no mercado.

Segundo OLIVEIRA (2002), o Data Mining é o processo automatizado de captura e análise de enormes conjuntos de dados, para extrair uma informação ou significado.

Segundo Fayyad ET al. (1996, apud GOLDSCHMID, 2005), Data Mining é “o processo não trivial, interativo e iterativo, com a função de identificar, em dados, padrões validos, novos, potencialmente úteis e ultimamente compreensíveis”.

Conforme citado no trecho acima, a palavra “iterativo” é caracterizada como a repetição do processo de KDD (Descoberta de Conhecimento do Banco de Dados – Knowledge-Discovery in Databases) coforme exibido na Figura 3, em busca de qualidade de resultados utilizando-se de refinamentos sucessivos. A palavra “interativa” indica a necessidade da presença do homem para controlar o processo. A expressão “Não trivial” reforça a dificuldade no processo de KDD. (GOLDSCHMIDT, 2005).

O processo de KDD usa diversos algoritmos para encontrar os padrões “válidos, novos e valiosos”, mas isso não é o suficiente para ter uma solução concreta para determinar padrões valiosos, o que demanda a interação de homens.



1. Processo de KDD (Descoberta de Conhecimento do Banco de Dados)

São realizadas buscas em diversas fontes de dados, como bancos de dados, relatórios, logs que são tratados por diversos processos, como por exemplo, ETL’s, para que sejam removidas todas as partes não aproveitáveis dos dados, e também para formatar os dados que irão ser inseridos em outros tipos de bancos de dados.

Data Warehouse e Data Mart, são exemplos de bancos de dados que contém apenas informações tratadas por processos que realizaram a “preparação” dos dados para que estes dados possam ser consolidados, adequados e selecionados para o processo de mineração e enfim transformar os dados de um banco de dados, em informações extremamente importantes para as empresas. Algoritmos de classificação, estatística, entre outros, são usados para desfrutar o máximo possível destes dados. O homem como analista tem por seu dever, fazer com que de forma interativa, busque possibilidades até que os padrões apareçam. Este processo de encontrar padrões necessita que os dados brutos sejam sistematicamente simplificados para desconsiderar o que é específico e privilegiar o que é genérico.

Segundo GOLDSCHMIDT (2005), a mineração de dados é composta por cinco funções:

* Classificação: características que definem um grupo. Exemplo: Produtos que são de alguma promoção ou liquidação.
* Agrupamento: grupos de itens que possuem uma mesma característica. Exemplo: Grupo 1: Itens vendidos em uma data específica. A diferença do agrupamento e da classificação, é que no agrupamento não existe uma característica pré-definida.
* Associação: relaciona eventos ocorridos em uma determinada ocasião. Exemplo: Quantidade de produtos vendidos em uma promoção.
* Sequenciamento: parecida com associação, porém a relação é com base em um período de tempo. Exemplo: Promoções repetidas em uma loja.
* Previsão: com base em massas de dados existentes, calcula-se uma estimativa. Para isso algumas técnicas são utilizadas como a descoberta de associação, classificação, regressão, clusterização, sumarização, detecção de desvios, descoberta de sequencias. Abaixo será explicado um pouco de cada técnica:
* Descoberta de associação: busca de itens que transacionam frequentemente e que ocorram de forma simultânea pelos bancos de dados. Algoritmos utilizados: Apriori, GSP, DHP.
* Classificação: busca encontrar métodos para classificar itens para que novos itens sejam classificados previamente. Algoritmos utilizados: Redes Neurais, Lógica Indutiva.
* Regressão: busca por função para mapear valores de banco de dados em valores reais, porém é restrita a números. Técnicas utilizadas: Estatística.
* Clusterização: separar registros em clusters onde podem ser compartilhados. Algoritmos utilizados: K-Means, K-Modes, K-Medoids.
* Sumarização: busca por mostrar conteúdos diferentes com as mesmas características. Algoritmos utilizados: Lógica indutiva, algoritmo genético.
* Detecção de desvios: detecta registros fora do padrão. Técnica utilizada: Estatística.
* Descoberta de sequencias: descobre associações pela ordem dos eventos

Existem diversas ferramentas especificas para realizar a fase de mineração de dados. Neste trabalho, será adotado o Weka como ferramenta.

### Weka

O Weka é uma ferramenta aberta com o código aberto, aplicável em KDD, desenvolvido na linguagem de programação JAVA, pelos alunos do curso de Ciência da Computação da Universidade de Waikato na Nova Zelândia em 1993. Teve o intuito de agregar algoritmos para a mineração de dados na área de Inteligência Artificial. O Software é licenciado pela General Public License, sendo possível a alteração do seu código fonte.

Possui uma série de heurísticas para mineração de dados, relacionadas à regressão, clusterização, classificação, regras de associação e visualização, entre elas: NaiveBayes, Bagging, LogistBoot, Linear Regression, IB1, Part, Ridor, ID3 e LMT.

A visualização das Árvores de Decisões já ocorre com a poda, e a Matriz de Confusão é apresentada apontando os erros e acertos considerados pelo sistema.

O processo de Extração do Conhecimento através desta ferramenta é composto por cinco etapas:

* Seleção: define a escolha da base de dados a ser analisada.
* Pré-processamento: etapa de “limpeza” dos dados.
* Transformação: são modificados ou transformados em formatos apropriados à mineração, que pode ser por agregação, generalização, normalização, construção de atributos ou redução de dados.
* Mineração de dados: etapa de utilização de técnicas de algoritmos.
* Interpretação: etapa de análise dos resultados obtidos por meio da mineração dos dados, e a partir dos quais se adquire o conhecimento.

A utilização da ferramenta pode ser realizada de diversas maneiras. Possui quatro diferentes interfaces permitindo que todos os seus algoritmos sejam chamados diretamente via código Java.

As interfaces são:

* Simple Client: nesta interface, a integração do usuário com o Weka ocorre por meio de linhas de comando. Requer um profundo conhecimento do programa, porém é flexível e ágil para usuários avançados.
* Explorer: trata-se da interface de utilização mais comum, e enquadra separadamente as etapas de pré-processamento, mineração de dados e pós-processamento.
* Experimenter: constitui um ambiente de experimentação em que testes estatísticos podem ser conduzidos a fim de avaliar o desempenho de diferentes algoritmos de aprendizado.
* Knowledgeflow: é uma ferramenta gráfica que permite o planejamento de ações na construção de um fluxo de processos de KDD.

O Weka possui diversos métodos de associação, classificação e clusterização. A inclusão ou remoção de novos métodos pode ser realizada possibilitando uma customização da ferramenta.

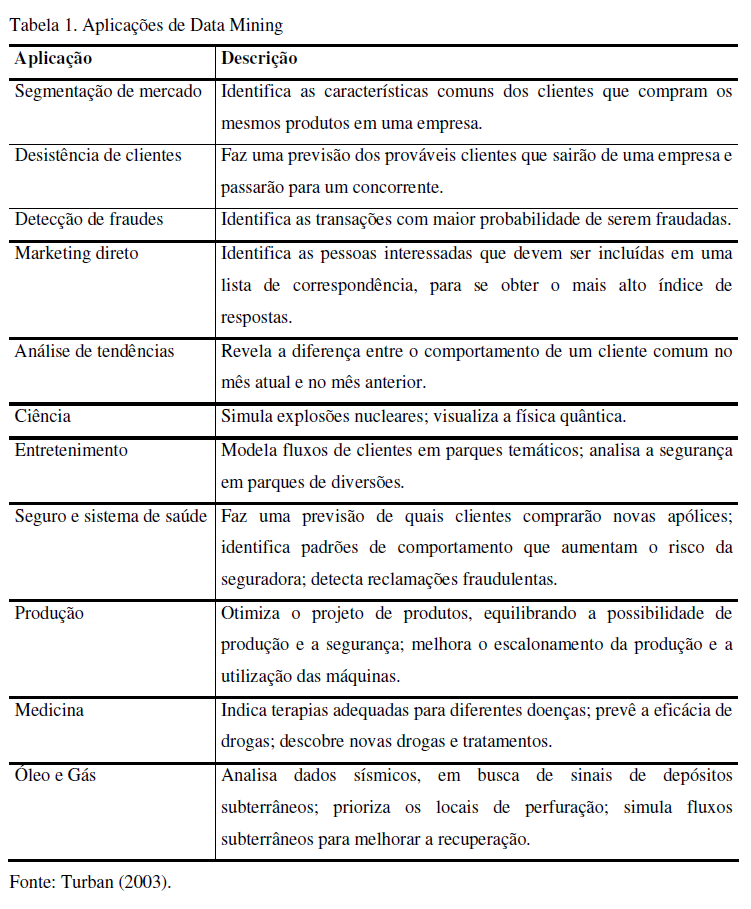
Esta ferramenta suporta a abertura direta de arquivos ARFF, CSV, C45. Porém, apenas consegue manipular os ARFF. Este é um arquivo usado para definir atributos e seus valores.

### Aplicações

O Data Mining é utilizado para diversas aplicações, como relações com os clientes, venda cruzada e venda de atualizações, complementos, aperfeiçoamentos, gerenciamento de campanhas, analise de mercados, canais e preços, e analise de segmentação de clientes.

Abaixo, a tabela 1, mostra algumas aplicações de Data Mining no mercado:

Tabela 1 – Aplicações de Data Mining



# DESENVOLVIMENTO DO TRABALHO

Este capítulo apresenta uma proposta para implantação do Sistema de Corretagem de Seguros baseado em Inteligência Artificial – SCSIA, o qual utiliza na sua arquitetura redes neurais Perceptron Simples. Dessa arquitetura é apresentado o modo de tratamento de entradas e saídas e a sequência de execução do programa de aplicação, o ambiente de desenvolvimento, as ferramentas utilizadas para programação do sistema e o modo de extração, transformação e carga dos dados.

## Sistema SCSIA

O sistema será constituído por três módulos conforme figura 5. O primeiro será a fonte para a base de conhecimento do sistema, ou seja, um banco de dados de segurados. O segundo módulo se encarrega de garantir a qualidade dos dados, sendo composto por tratamento e codificação dos dados. E por último, um módulo de rede neural, que será o *core* do sistema, este será encarregado de interligar os demais módulos e interagir com o usuário a partir de uma Interface Homem Máquina (IHM) que permitirá a emissão de relatórios, o treinamento da rede, e a qualificação de perfis.



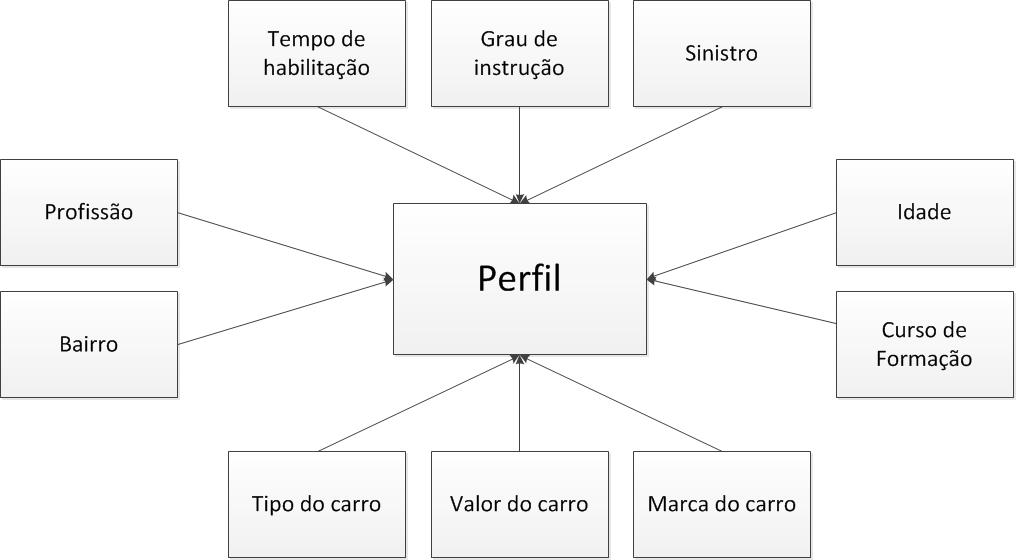
1. Arquitetura do sistema

### Módulo de banco de dados

A técnica de mineração de dados será utilizada para descobrir os conhecimentos guardados no banco de dados vide figura 6, para isso, organização e capacidade de escolher as características relevantes, são características fundamentais para que o aprendizado possa acontecer.

Para uma melhor solução de um grande problema, através da técnica de redes neurais, será decomposto em subproblemas de menor proporção e mais simplificados, tornando a análise mais fácil e com maior capacidade de “aprender” a executar tarefas por meio de treinamentos, para aplicações sobre um volume de dados de teste para obtenção de resultados na busca de conhecimento.

#### Estrutura do banco de dados



1. Estrutura das tabelas do sistema

### Tratamento dos dados utilizados para entrada da rede

O conhecimento do perfil dos segurados é o alvo desse estudo. O setor de corretagem de seguros tem uma enorme necessidade de um respaldo de informações que possibilitam melhorar a qualidade e confiabilidade das informações sobre o perfil de cada segurado disponível.

Extraindo as informações sobre o perfil dos segurados disponíveis no banco de dados da seguradora, pode-se descobrir e montar regras de classificação que permitam a definição do perfil ideal para determinado cliente da seguradora.

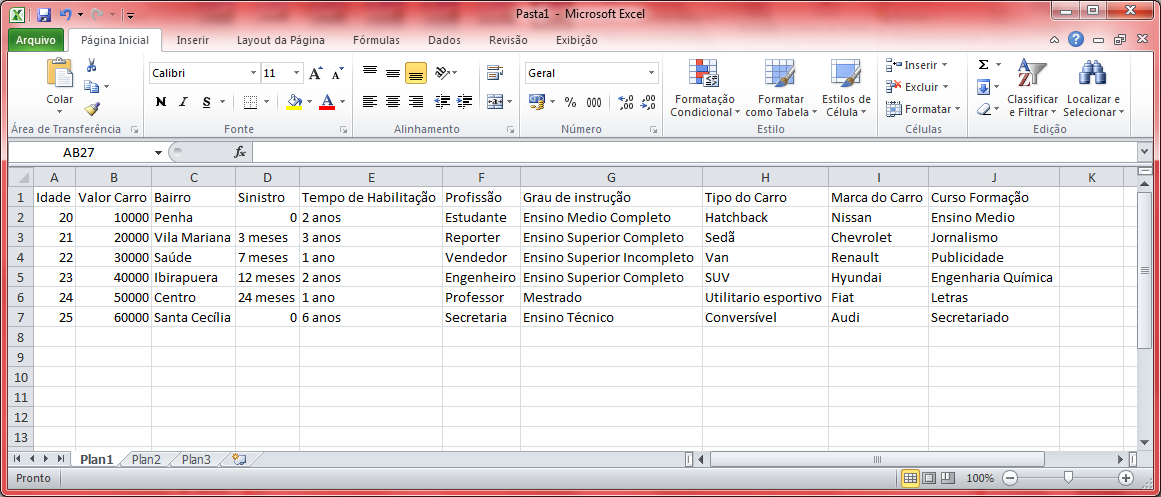
Os dados de entrada (amostras) utilizados para realização dos testes são extraídos de uma planilha Microsoft Excel 2007© (contendo as colunas que representam os campos da tabela), para um SGBD Microsoft SQL Server 2008 R2 Enterprise. Durante o tratamento das informações da planilha os dados sofrem uma codificação que os enriquecem e os preparam para o processo de descoberta. A tabela 2 apresenta o layout da planilha que será extraída para o banco de dados.

Tabela - Dados extraídos para o banco de dados

Para realização dos testes foi utilizado um microcomputador HP Pavilion com sistema operacional Windows 7 Professional com 3 Gb de memória RAM, 127 Gb de espaço em disco e processador AMD Turion 64 X2 2.00GHz, para analisar o perfil de 100 clientes da seguradora, com os seguintes perfis: (I) Perfil de alto risco e (II) Perfil de baixo risco. Os seguintes dados foram escolhidos como atributos, que indicam o perfil do segurado:

1. Idade
2. Valor do carro
3. Bairro
4. Sinistro
5. Tempo de habilitação
6. Profissão
7. Grau de instrução
8. Tipo do carro
9. Marca do carro
10. Curso de formação

Esta escolha dos atributos foi baseada em suposições, ao submeter o sistema à um cenário real estes atributos deverão ser definidos por um especialista da área de seguros.

#### Codificação dos dados

Após a tratamento dos dados é necessário realizar a codificação dos mesmos, ou seja, deve-se criar padrões para representar os dados. A tabela 3 exemplifica a codificação dos dados

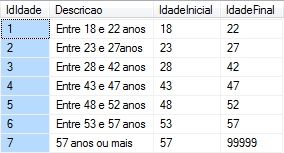
Tabela 3 - Exemplo de codificação dos dados de acordo com a figura



Os atributos foram classificados conforme segue:

Idade: Esta coluna apresenta a faixa de idade do solicitante do seguro. Foi dividida em 7 classificações conforme a tabela 4.

Tabela 4 - Tabela de idade



Valor do carro: Esta coluna apresenta o valor do carro do solicitante do seguro.

Foi dividido em 8 classificações conforme mostra a tabela 5.

Tabela 5 – Tabela do valor do carro



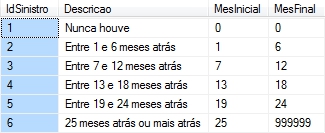
Bairro: Esta coluna apresenta o grau de risco do bairro de residência do solicitante do seguro. Foi dividido em 5 classificações conforme a tabela 6.

Tabela 6 – Tabela de bairros.



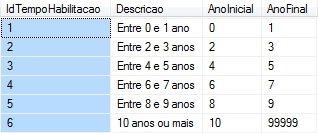
Sinistro: Esta coluna apresenta qual foi a ultima vez que o solicitante de seguro teve alguma abertura de sinistro. Foi dividido em 6 classificações conforme a tabela 7.

Tabela 7 - Tabela de classificação de sinistros



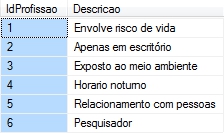
Tempo de habilitação: Esta coluna determina quanto tempo o solicitante do seguro é habilitado, para determinar sua experiência com direção. Foi dividido em 6 classificações conforme a tabela 8.

Tabela 8 - Tabela de tempo de habilitação



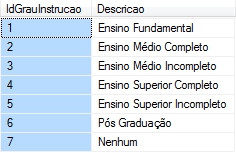
Profissão: Esta coluna classifica qual o tipo de profissão o solicitante do seguro trabalha, para saber qual o perfil profissional do segurado. Foi dividido em 6 classificações conforme a tabela 9.

Tabela 9 - Tabela de classificação de profissão



Grau de instrução: Esta coluna classifica o grau de instrução do solicitante do seguro, para saber qual o nível escolar do segurado. Foi dividido em 7 classificações conforme a tabela 10.

Tabela 10 - Tabela de grau de instrução



Tipo do carro: Esta coluna classifica o tipo do carro que está sendo segurado. Foi dividido em 9 classificações conforme a tabela 11.

Tabela 11 - Tabela de classificações de carros



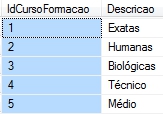
Marca do carro: Esta coluna classifica a marca do carro que está sendo segurado. Foi dividida em 5 classificações conforme a tabela 12.

Tabela 12 - Tabela de marcas de carros do banco de dados



Curso de formação: Esta coluna classifica qual a ultima formação acadêmica do segurado. Foi dividida em 5 classificações conforme a tabela 13.

Tabela 13 - Tabela de cursos de formação



### Módulo de redes neurais

O módulo de rede neural utilizado é baseado na técnica de rede neural conhecida como rede perceptron criada em 1958, pelo cientista Frank Rosenbalt. A rede é constituída de uma única camada, sobre arquitetura *“Feed forward”* que processa o fluxo de informações no sentido da camada de entrada em direção à camada neural de saída, inexistindo qualquer tipo de realimentação.

O módulo é constituído de dois submódulos:

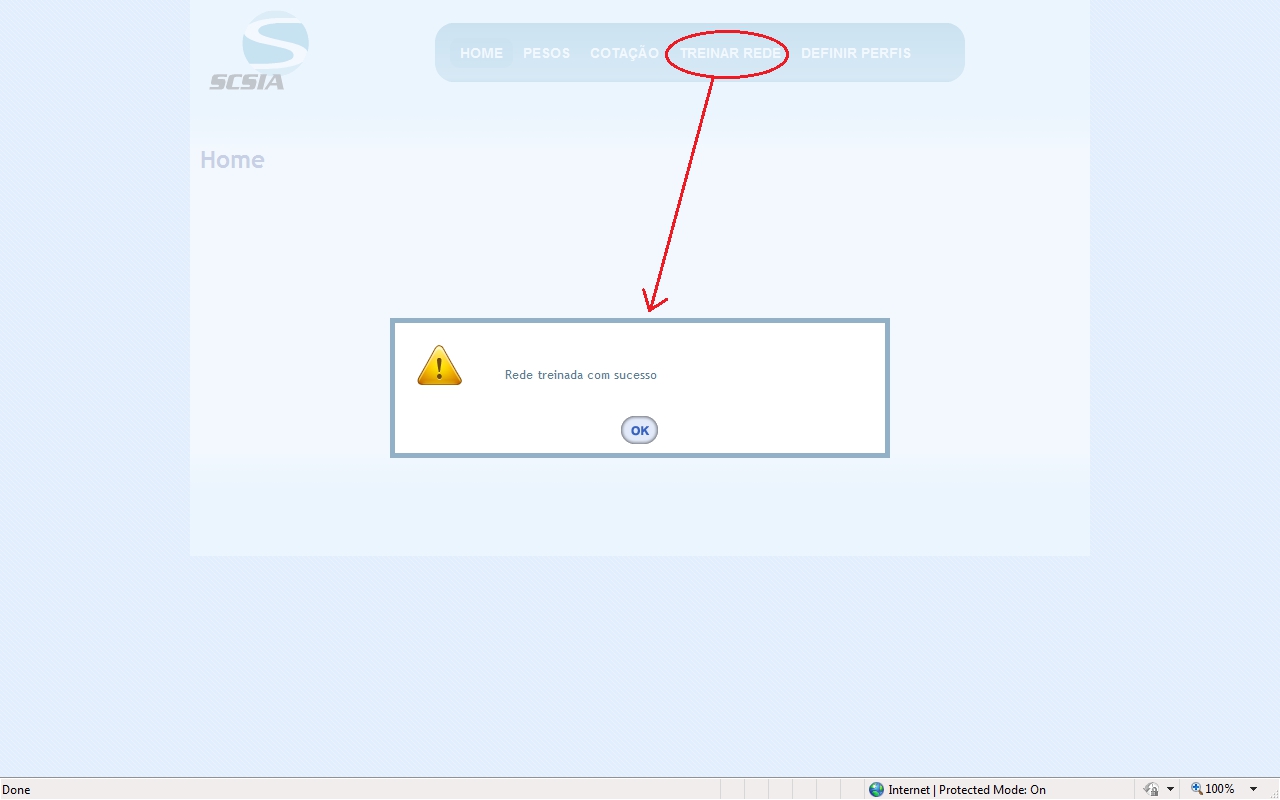
* Treinamento
* Execução

Os modelos foram desenvolvidos em linguagem C#. O algoritmo de aprendizado utilizado foi o de Hebb.

### Descritivo operacional do sistema

#### Treinamento da rede

Para a realização do treinamento da rede neural, é necessário clicar no botão “Treinar rede” e aguardar até que a mensagem “Rede treinada com sucesso” apareça conforme figura 6.



1. Treinamento da rede

#### Cotação do segurado

Para realizar a cotação de um novo segurado, é necessário clicar em “Cotação” no menu principal e preencher corretamente os dados no formulário que é apresentado. As informações devem ser preenchidas da seguinte maneira.

* Idade: Neste campo deve ser preenchido a idade do perfil à ser testado somente com números.
* Valor do carro: Neste campo deve ser preenchido o valor do carro à ser segurado somente com número.
* Tipo do carro: Neste campo deve ser selecionado o tipo do carro à ser segurado.
* Grau de instrução: Neste campo deve ser selecionado o grau de instrução do perfil à ser testado.
* Curso formação: Neste campo deve ser selecionado o curso de sua última formação.
* Tempo habilitação: Neste campo deve ser preenchido a quantidade em anos que o segurado possui habilitação.
* Ocorrência de sinistro anterior: Neste campo deve ser selecionado de acordo com o intervalo de meses que o segurado teve sua ultima abertura de sinistro.
* Profissão: Neste campo deve ser selecionado a atual profissão do segurado.
* Marca do carro: Neste campo deve ser selecionado a marca do carro à ser segurado.
* Bairro: Neste campo deve ser selecionado de acordo com o bairro de residência atual do segurado.

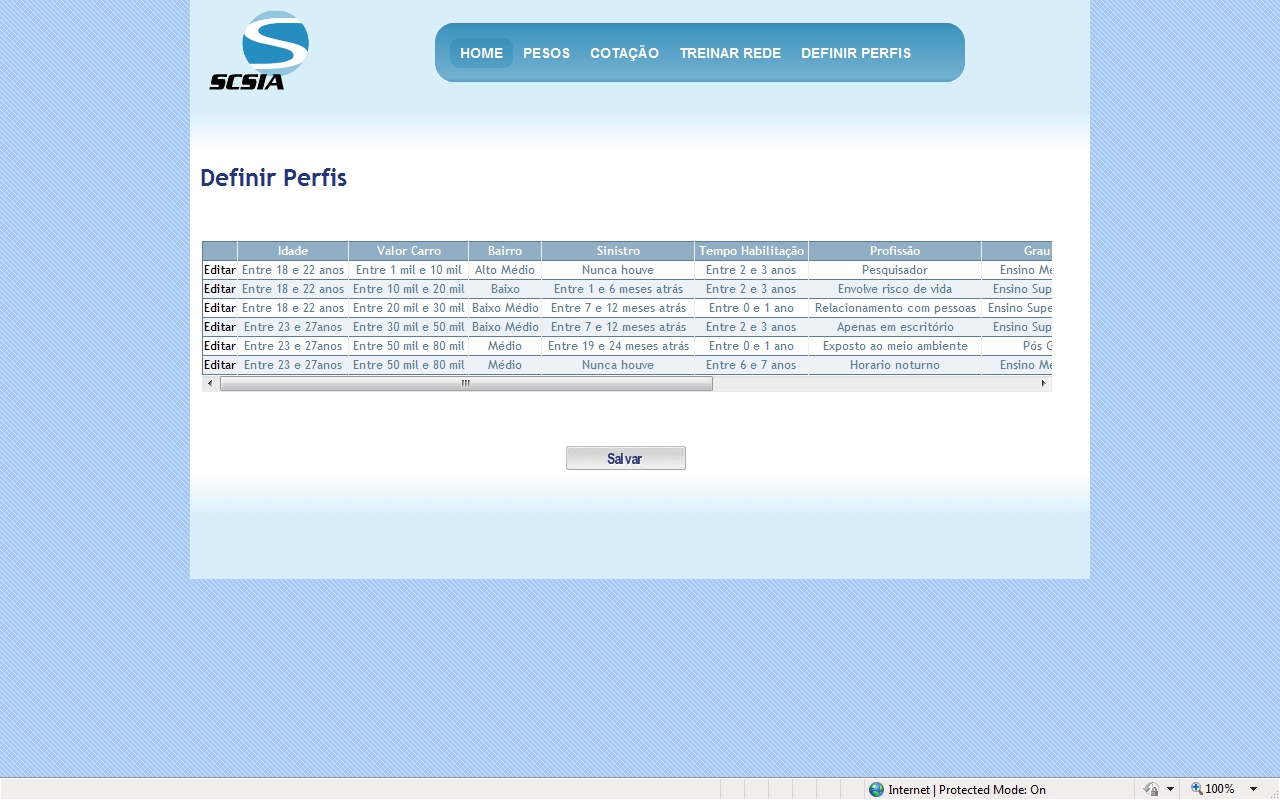
Após preencher todos os atributos da tela (figura 7) é necessário clicar em “Consultar” e aguardar o resultado.



1. Cotação do segurado

#### Definição do grau de risco dos perfis

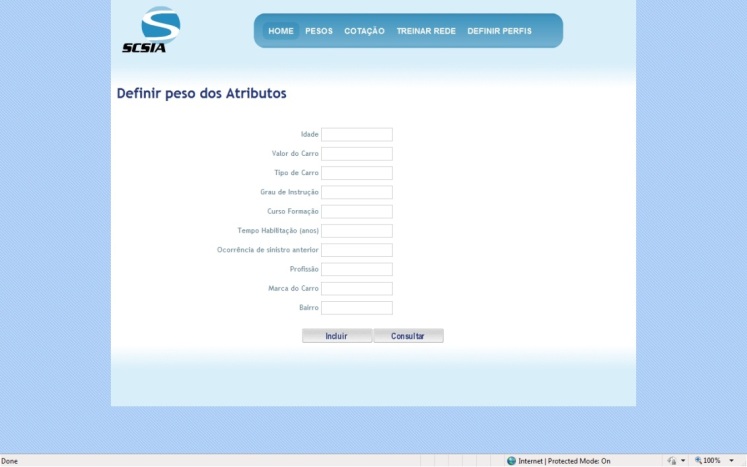
Nesta tela, deve-se definir o grau de risco dos perfis que foram carregados no sistema pelo ETL. Para isso, deve-se clicar em “Definir perfis” e aguardar até que os perfis carregados no sistema sejam apresentados. Após a apresentação dos mesmos, deve-se clicar em “Editar” em cada perfil e definir qual o grau de risco. Clicar em “Salvar” após editar todos os perfis, conforme figura 8.



1. Definição do grau de risco dos perfis

#### Definição do peso dos atributos

Nesta tela, deve-se definir o peso para cada atributo do perfil. Para isso, clicar em “Pesos”, definir os pesos desejados em cada atributo e clicar em “Salvar” para que os pesos sejam alterados. O valor de cada peso deverá ser determinado por um especialista da área de seguros após realizada uma pesquisa de mercado na qual será definido o grau de importância de cada atributo, conforme figura 9.



1. Definição do peso para os atributos

#### Resultado da cotação

Esta tela é apresentada após clicar no botão “Consultar” na tela de Cotação. Ela irá indicar se o perfil consultado é de alto ou baixo risco, conforme figura 10.



1. Resultado da cotação

# ENSAIOS PRÁTICOS

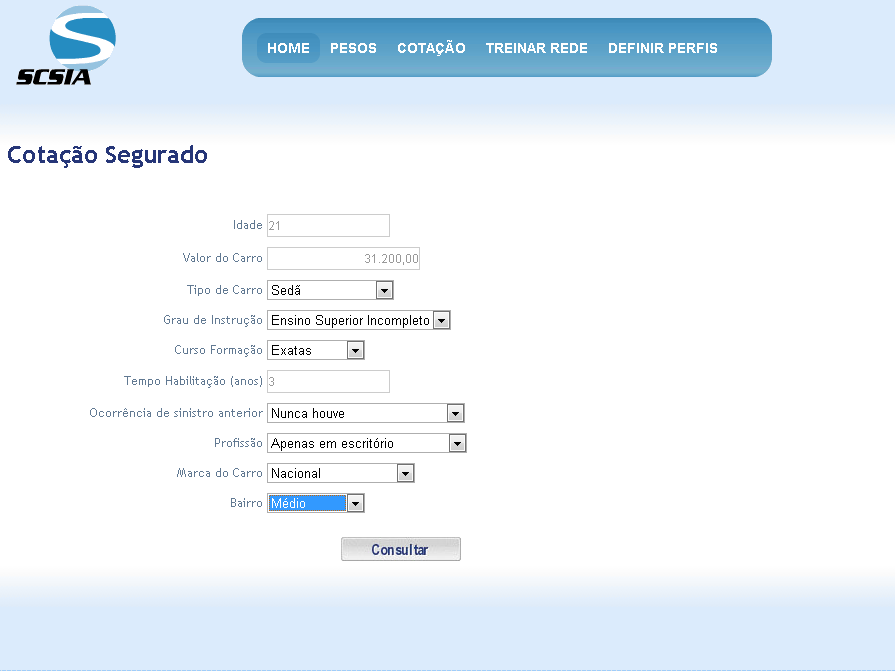
Para ensaio prático do sistema foram realizados 10 testes conforme descrito a seguir.

**Teste 1**

Foi realizada uma cotação para o segurado VINICIUS MATOS ANDRADE, estabeleceu-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 11.



1. Cotação de Vinicius Matos Andrade

*Passo 2*

O resultado da cotação é apresentado conforme a figura 12.



1. Resultado de Vinicius Matos Andrade

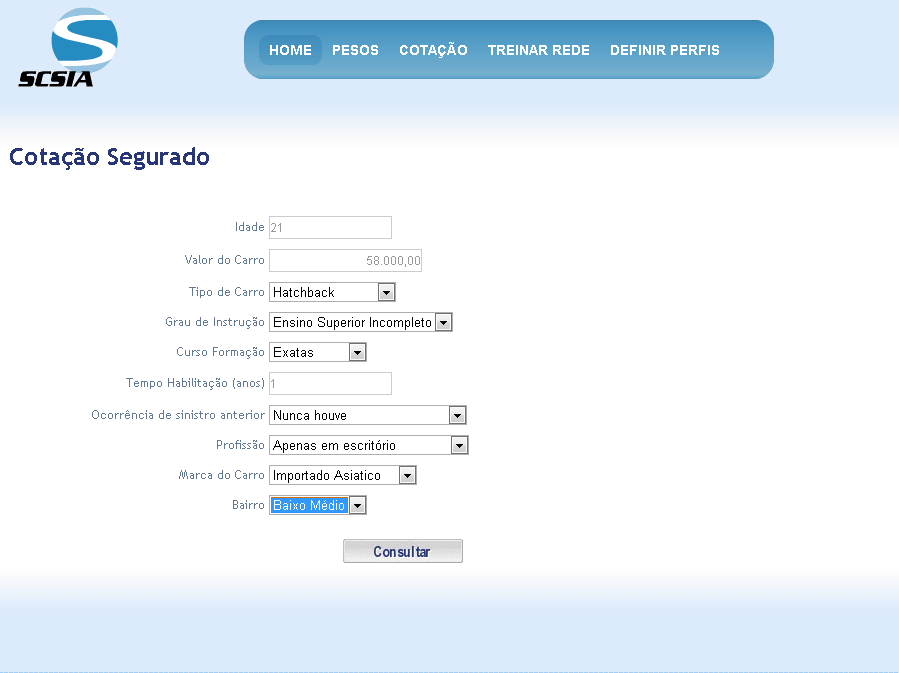
Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

**Teste 2**

Foi realizada uma cotação para o segurado WILLIAM RODRIGUES DA SILVA, estabeleceu-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 13.



1. Cotação de William Rodrigues da Silva

*Passo 2*

O resultado da cotação é apresentado conforme a figura 14.



1. Resultado de William Rodrigues da Silva

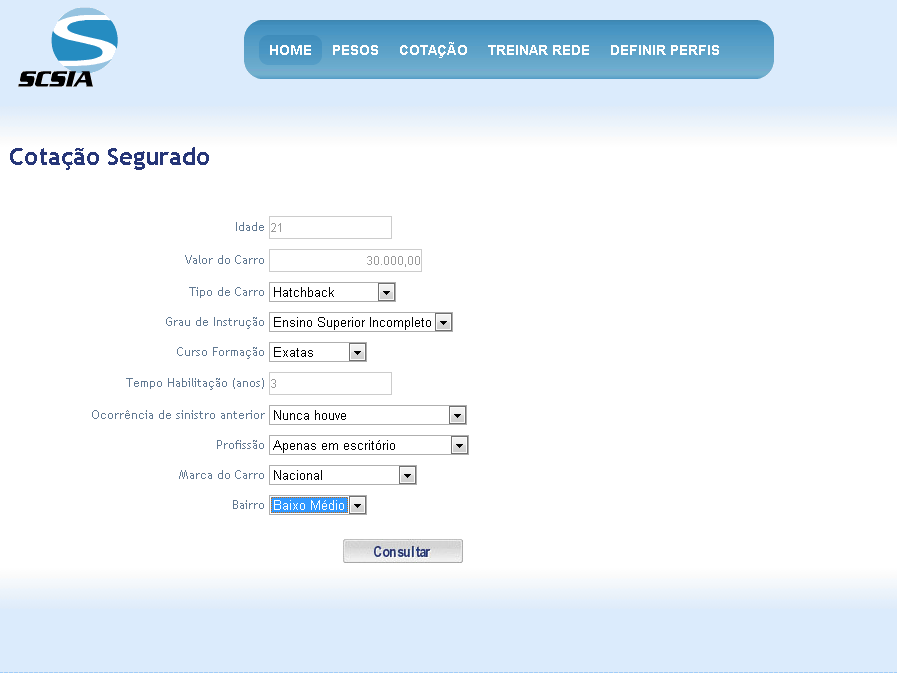
Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

**Teste 3**

Foi realizada uma cotação para o segurado RODRIGO DE CARO, estabeleceu-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 15.



1. Cotação de Rodrigo de Caro

*Passo 2*

O resultado da cotação é apresentado conforme a figura 16.



1. Resultado de Rodrigo de Caro

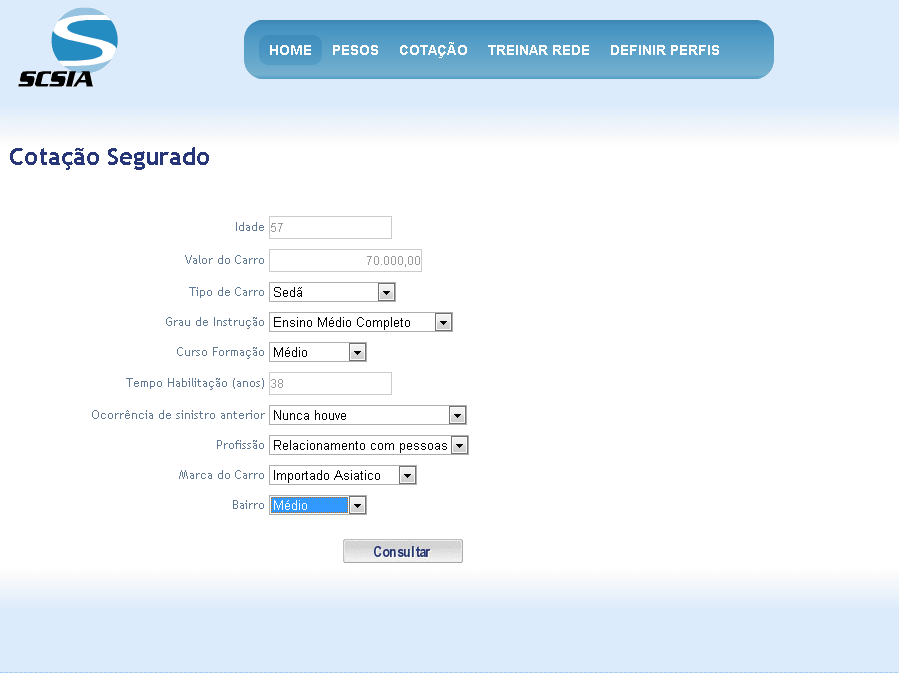
Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

**Teste 4**

Foi realizada uma cotação para o segurado DEMÉTRIO DOS SANTOS ANDRADE, estabeleceu-se os seguintes passos.

*Passo 1*

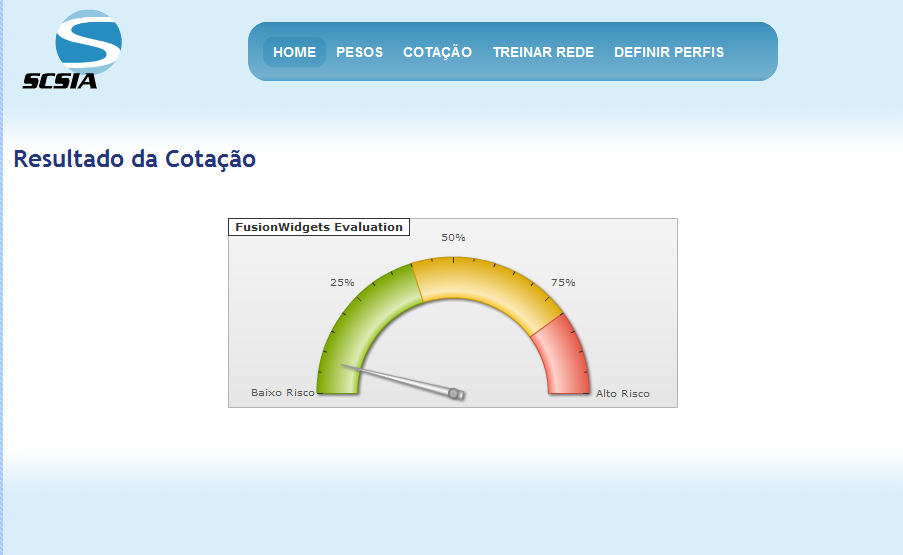
Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 17.



1. Cotação de Demétrio dos Santos Andrade

*Passo 2*

O resultado da cotação é apresentado conforme a figura 18.



1. Resultado de Demétrio dos Santos Andrade

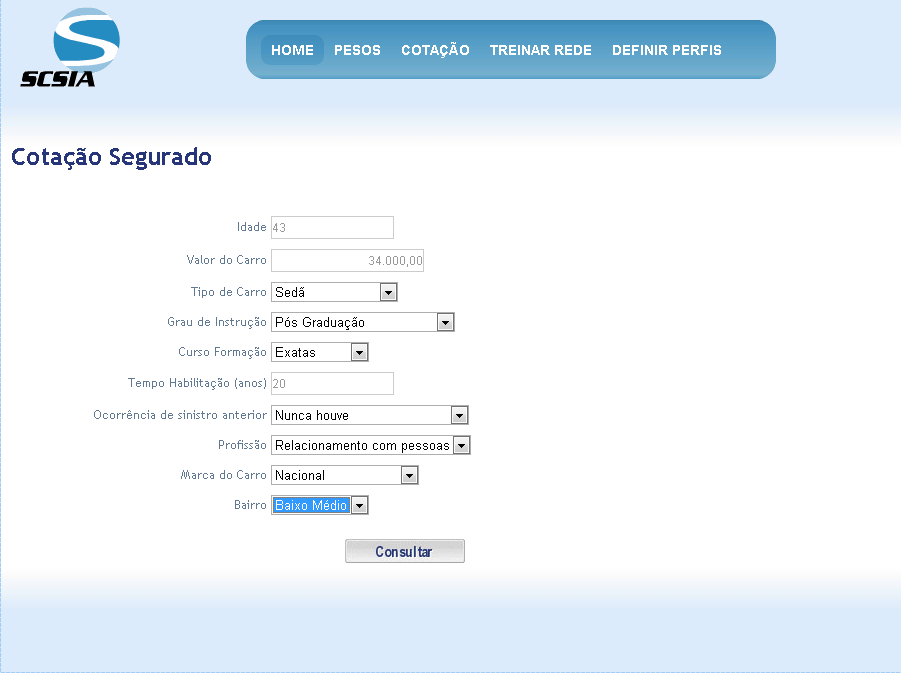
Neste teste, o perfil que foi consultado apresenta baixo risco de sinistro.

**Teste 5**

Foi realizada uma cotação para o segurado MOISES RODRIGUES DA SILVA, estabeleceu-se os seguintes passos.

*Passo 1*

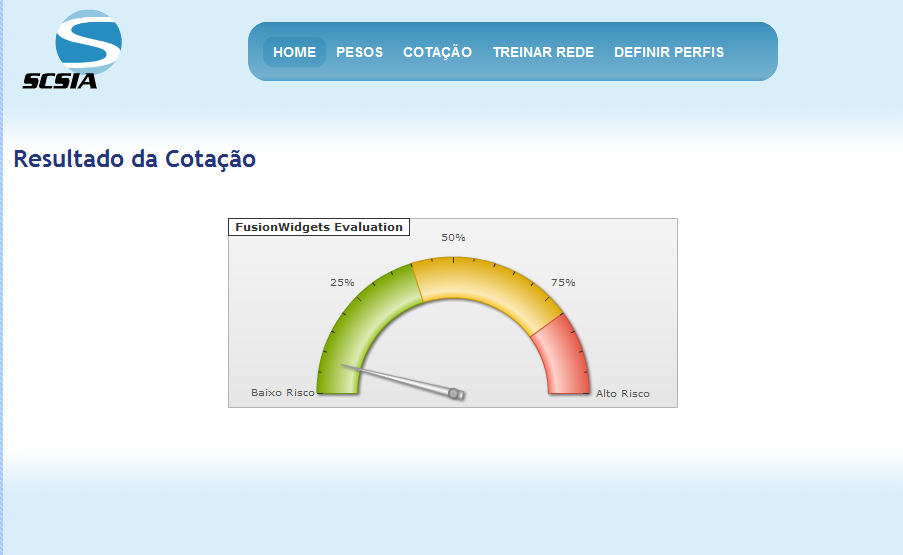
Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 19.



1. Cotação de Moises Rodrigues da Silva

*Passo 2*

O resultado da cotação é apresentado conforme a figura 20.



1. Resultado de Moises Rodrigues da Silva

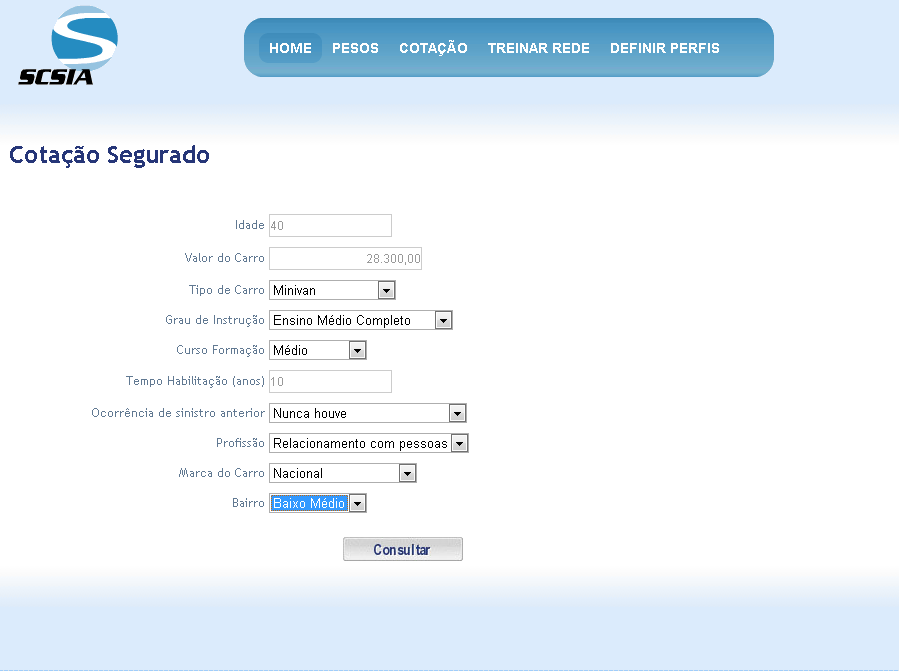
Neste teste, o perfil que foi consultado apresenta baixo risco de sinistro.

**Teste 6**

Foi realizada uma cotação para o segurado SIDINEIA OLIVA DE ALMEIDA, estabeleceu-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 21.



1. Cotação de Sidneia Oliva de Almeida

*Passo 2*

O resultado da cotação é apresentado conforme a figura 22.



1. Resultado de Sidneia Oliva de Almeida

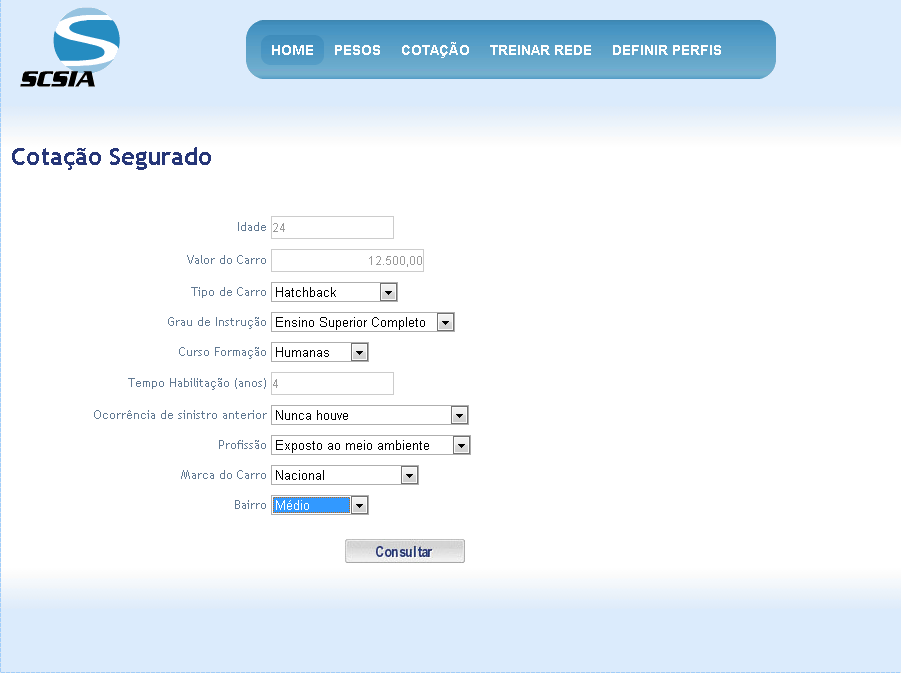
Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

**Teste 7**

Foi realizada uma cotação para o segurado ÉRIKA MATOS ANDRADE, estabeleceu-se os seguintes passos.

*Passo 1*

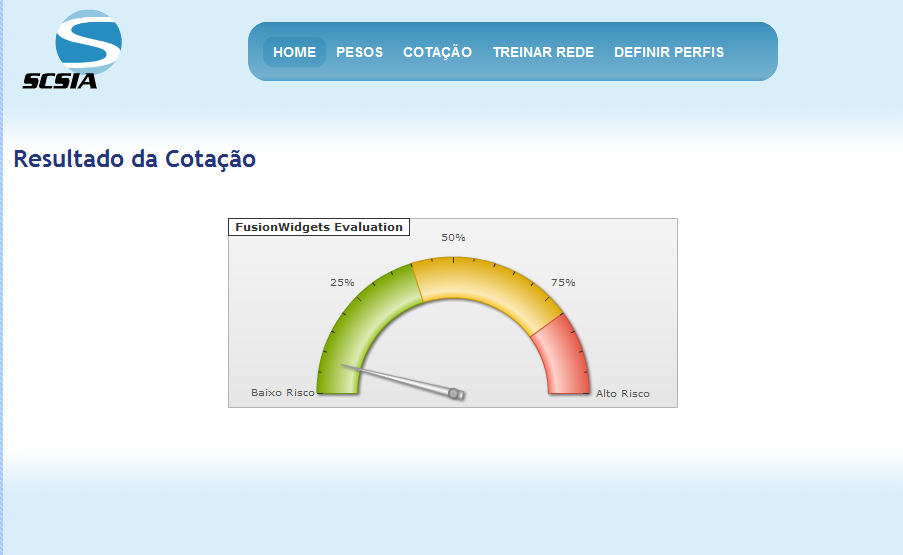
Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 23.



1. Cotação de Érika Matos Andrade

*Passo 2*

O resultado da cotação é apresentado conforme a figura 24.



1. Resultado de Érika Matos Andrade

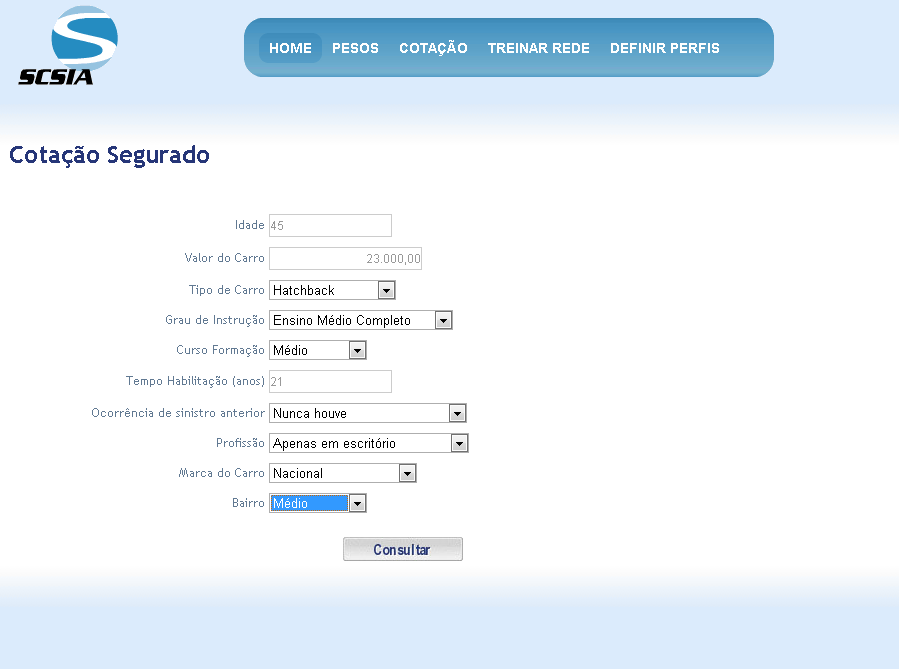
Neste teste, o perfil que foi consultado apresenta baixo risco de sinistro.

**Teste 8**

Foi realizada uma cotação para o segurado MARLENE ROSA DE MATOS ANDRADE, estabeleceu-se os seguintes passos.

*Passo 1*

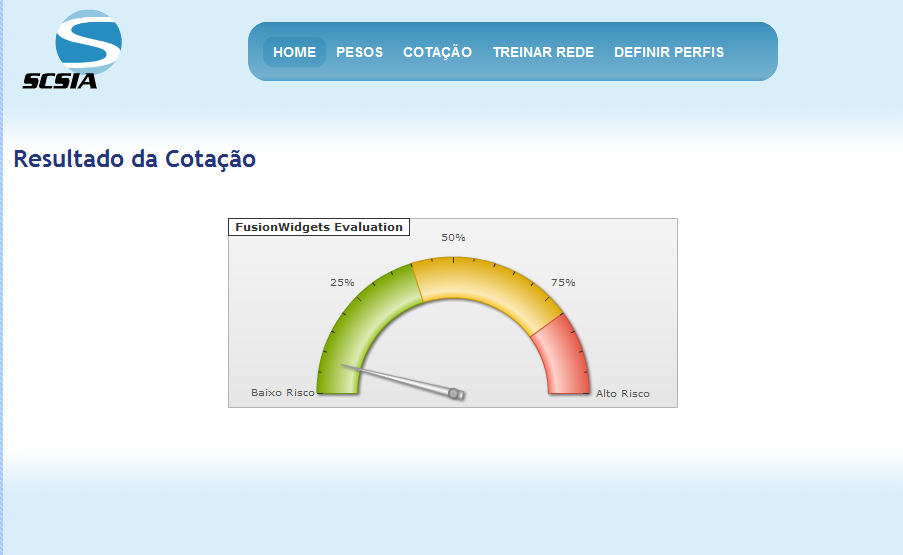
Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 25.



1. Cotação de Marlene Rosa de Matos Andrade

*Passo 2*

O resultado da cotação é apresentado conforme a figura 26.



1. Resultado de Marlene Rosa de Matos Andrade

Neste teste, o perfil que foi consultado apresenta baixo risco de sinistro.

**Teste 9**

Foi realizada uma cotação para o segurado EUCLIDES DOS SANTOS ANDRADE, estabeleceu-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 27.



1. Cotação de Euclides dos Santos Andrade

*Passo 2*

O resultado da cotação é apresentado conforme a figura 28.



1. Resultado de Euclides dos Santos Andrade

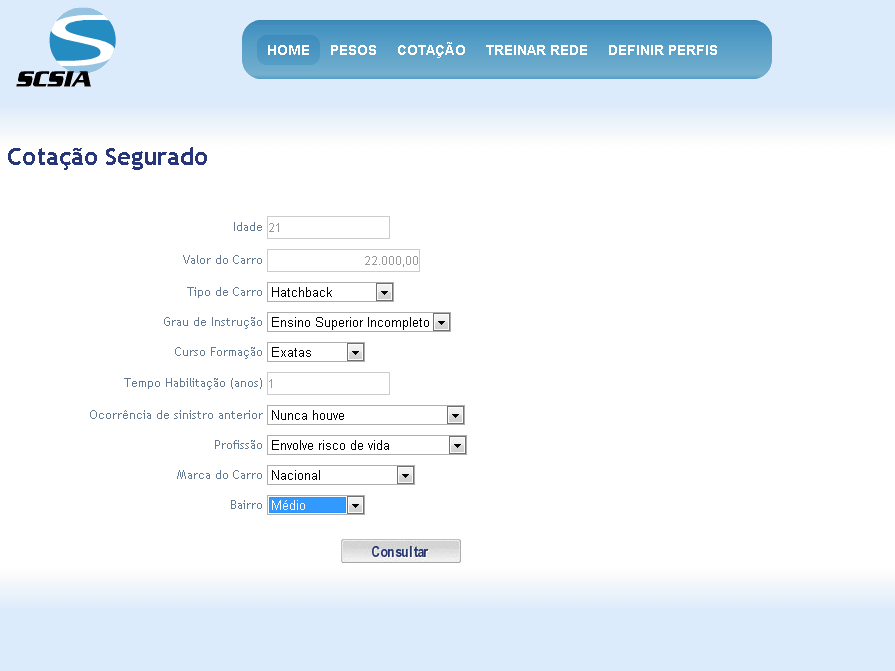
Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

**Teste 10**

Foi realizada uma cotação para o segurado BEATRIZ VALERA CAPITINA, estabeleceram-se os seguintes passos.

*Passo 1*

Seleciona-se no menu principal a opção “Cotação”, preenchendo-se o formulário conforme a figura 29.



1. Cotação de Beatriz Valera Capitina

*Passo 2*

O resultado da cotação é apresentado conforme a figura 30.



1. Resultado de Beatriz Valera Capitina

Neste teste, o perfil que foi consultado apresenta alto risco de sinistro.

## Discussão dos resultados

O treinamento da rede sempre será de acordo com os perfis do segurados definidos pela seguradora. Nos testes realizados foram definidos os seguintes pesos para os atributos para o treinamento da rede (lembrando que o valor de cada peso deverá ser determinado por um especialista da área de seguros após realizada uma pesquisa de mercado na qual será definido o grau de importância de cada atributo).

* Idade: 5
* Valor do carro: 4
* Tipo do carro: 2
* Grau de instrução: 1
* Curso formação: 1
* Tempo habilitação: 3
* Ocorrência de sinistro anterior: 3
* Profissão: 2
* Marca do carro: 2
* Bairro: 4

# CONCLUSÕES

No ensaio do sistema, utilizou-se uma base de dados baseada em dados fictícios, por meio da simulação do programa que a classificação dos perfis tendeu para o perfil de alto risco devido aos perfis utilizados para o teste. No sistema apresentado neste trabalho obteve-se em 10 perfis a classificação de 4 perfis sendo como de baixo risco e 6 perfis de alto risco.

O sistema SCSIA foi testado conforme descrito no capítulo 4, e apresentou os resultados esperados, ou seja, mostrou-se eficiente como uma ferramenta de apoio a corretagem de seguros.

Não foi possível testá-lo em uma corretora de grande porte, como por exemplo, Porto Seguro, Liberty Seguros, em virtude da impossibilidade por razões de segurança em seu banco de dados.

Um dos limites observados no desenvolvimento do trabalho está relacionado com a dependência das atuais aplicações utilizadas nas seguradoras. É importante ressaltar que o SCSIA não realizará cotações de seguros e sim auxiliará os sistemas utilizados pelas seguradoras. Porém esta dependência tornará o cálculo feito pelos sistemas atuais mais confiáveis.

## Trabalhos futuros

Como sugestões para trabalhos futuros, já que o enfoque principal deste trabalho é a disponibilização do grau de risco mais preciso, são:

1. Alteração de rede neural perceptron simples para a rede neural multicamadas, que permitirá a disponibilização da porcentagem de risco que cada segurado apresentará.
2. Definir o grau de risco dos perfis de clientes automaticamente.
3. Interface gráfica para o cadastro de novos atributos pelo usuário final, que resultará conforme o histórico de segurados, em uma analise mais precisa do grau de risco.
4. Disponibilização de um WebService para que outras seguradoras possam consumir este serviço e ter resultados com maior precisão.

REFERÊNCIAS BIBLIOGRÁFICAS

**TURBAN, Efraim. Administração de Tecnologia da Informação /** Efraim Turban, R. Kelly Rainer, Jr., Richard R. Potter; Tradução de Teresa Cristina Felix de Souza. Rio de Janeiro: Campus, 2003.

**O’BRIEN, James A. Sistemas de Informação:** e as decisões gerenciais na era da internet. 2. Ed. São Paulo: Saraiva, 2004.

**OLIVEIRA, Wilson Jose de. Data Warehouse.** Florianópolis: Visual Books, 2002.

**GOLDSCHIMIDT, Ronaldo; PASSOS, Emanuel. Data Mining;**  um guia pratico. Rio de Janeiro: Elsevier, 2005.

**SILVA, Ivan nunes da; SPATTI, Danilo Hernane; FLAUZINO, Rogério Andrade; Redes Neurais Artificiais:** Para engenharia e ciências aplicadas. Art Liber.

**HAYKIN, Simon. Redes Neurais:** Princípios e praticas. 2. Ed. Porto Alegre: Bookmann, 2001.

**HEBB, Donald. O. The Organization of Behavior.** New York: Wiley, 1949. P. 60-65

**Redes Competitivas de KOHONEN. Acessado em: 10.jul.2011. Disponível em http://www.inf.ufpr.br/~aurora/tutoriais/kohonen.pdf**

**BRAGA, Antônio de Pádua; LAUDERMIR, Teresa Bernarda; CARVALHO, André Carlos Ponce de Leon Ferreira. Redes Neurais Artificiais: teoria e aplicação:** Livros Técnicos e Científicos Editora S.A., 2000.

APENDICE

Manual do sistema

*Bem vindo ao SCSIA!*

O projeto SCISIA surgiu no final do ano de 2011 como um trabalho de conclusão do curso Sistemas de Informação da FIAP – Faculdade de Informática e Administração Paulista.

Foi desenvolvido por Rodrigo de Caro, Vinicius Matos Andrade e William Rodrigues da Silva, com a orientação do Prof. Dr. Cézar da Costa.

Abaixo, iremos demonstrar um breve tutorial para a utilização do sistema.

Primeiro funcionamento

Em seu primeiro funcionamento iremos realizar um processo para a carga de dados de outros bancos de dados para o SCISA. Importamos um arquivo com a extensão “.txt” para o sistema com o agendamento de uma tarefa no Windows. Ao realizar este processo, um processo de ETL trata as informações contidas no arquivo “.txt” e realiza uma carga para o banco de dados do SCSIA para que possamos dar inicio à nossas analises e testes de perfis.

Passo 1 – Definir grau de risco dos perfis carregados pelo ETL

Feito a carga de dados teremos que definir o grau de risco dos perfis inseridos pela carga do ETL. Para isso, clique em “Definir Perfis”. Ao abrir a pagina com os perfis carregados, iremos clicar em “Editar” para definir o grau de risco dos perfis.



Figura 1 – Definição do grau de risco dos perfis

Passo 2 – Definir pesos para os atributos

Neste passo iremos definir os pesos iniciais para cada atributo do perfil do candidato. Para isso, clique em “Pesos”, preencha os pesos conforme desejado em cada atributo e clique em “Salvar”.



Figura 2 – Definição dos pesos dos atributos

Passo 3 – Treinamento da rede neural

Este procedimento deve ser realizado toda vez que os pesos dos atributos forem alterados. O treinamento da rede neural irá ajustar os pesos dos atributos de acordo com o conhecimento da base de dados do SCSIA. Para realizar este procedimento, clique em “Treinar rede” e aguarde até que uma mensagem “Rede treinada com sucesso” seja apresentada.



Figura 3 – Treinamento da rede neural

Passo 4 – Consulta do grau de risco do perfil

Para verificar o grau de risco do perfil desejado, iremos realizar uma consulta no sistema. Nesta consulta, os seguintes dados devem ser inseridos:

Idade: Neste campo deve ser preenchido a idade do perfil à ser testado somente com números.

Valor do carro: Neste campo deve ser preenchido o valor do carro à ser segurado somente com números.

Tipo do carro: Neste campo deve ser selecionado o tipo do carro à ser segurado.

Grau de instrução: Neste campo deve ser selecionado o grau de instrução do perfil à ser testado.

Curso formação: Neste campo deve ser selecionado o curso de sua ultima formação.

Tempo habilitação: Neste campo deve ser preenchido a quantidade em anos que o segurado possui habilitação.

Ocorrência de sinistro anterior: Neste campo deve ser selecionado de acordo com o intervalo de meses que o segurado teve sua ultima abertura de sinistro.

Profissão: Neste campo deve ser selecionado a atual profissão do segurado.

Marca do carro: Neste campo deve ser selecionado a marca do carro à ser segurado.

Bairro: Neste campo deve ser selecionado de acordo com o bairro de residência atual do segurado.



Figura 4 – Inserção dos atributos do perfil à ser testado

Após preencher todas as informações necessárias, clique no botão “Consultar” e aguarde até que a tela com o resultado da consulta seja apresentada.

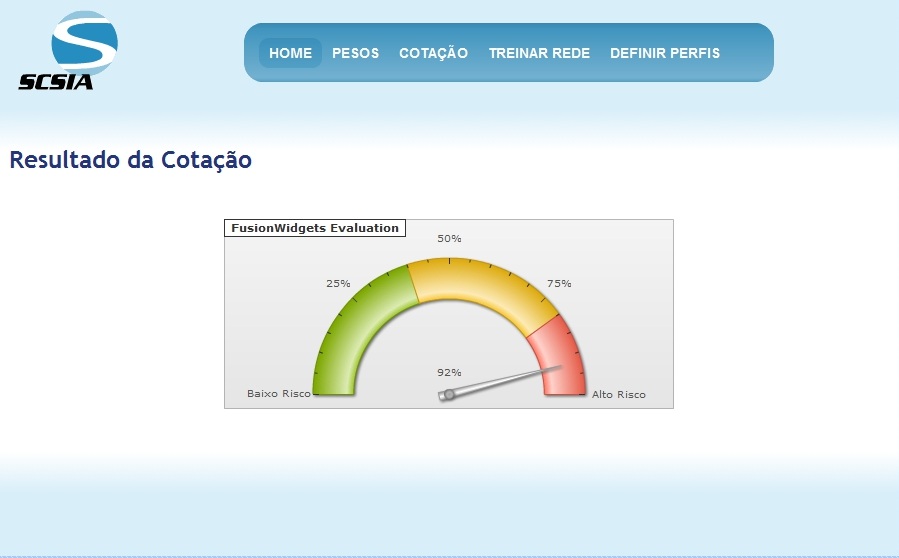


Figura 5 – Resultado da cotação do perfil