Business Intelligence

André Rodrigues dos Santos

Modelo de previsão de Inadimplência usando Redes Neurais

Monografia de Final de Curso

31/08/2015

Monografia apresentada ao Departamento de Engenharia Elétrica da PUC/Rio como parte dos requisitos para a obtenção do título de Especialização em Business Intelligence.

Orientador:

Juan Lazo Lazo

AGRADECIMENTO

Dedico esse trabalho a minha família que foi a grande alavanca de vontade e inspiração de mudança na minha vida. Sem dúvida a busca de vencer partiu de uma necessidade e tornou-se uma realidade positiva na minha vida que proporcionou o avanço de conhecimento e profissional na minha vida. Ressalto também a gratidão do amigo Andrei que juntos incentivamos um ao outro por toda a jornada do curso de BI Master e finalmente toda a gratidão pelo apoio do professor Juan Lazo Lazo que apesar de não lecionar tantas disciplinas, me auxiliou neste trabalho, que gentilmente ofereceu todo apoio.

RESUMO

O volume de dados armazenados nas organizações tende a ser cada vez maior. Este fato ocorre devido às organizações terem a necessidade de armazenarem o histórico de suas atividades, bem como os resultados dos processos desempenhados por ela mesma e também pela interação com seus clientes e fornecedores, além de eventuais parcerias com outras organizações. Outro fato que favorece este crescimento no volume de dados pode ser verificado na queda do custo dos dispositivos de armazenamento e avanços na área de Tecnologia da Informação (TI).

Os seres humanos, de uma forma geral, possuem pouca habilidade para analisar manualmente tamanha quantidade de dados e, assim, muitas informações, possivelmente úteis, são desperdiçadas, ficando ocultas dentro das bases de dados das organizações.

Em consequência disto, com a expansão do volume de dados, cresce também a necessidade de desenvolver novas ferramentas e técnicas de extração de conhecimento a partir de dados armazenados. Estas ferramentas e técnicas têm se mostrado cada vez mais indispensáveis.

Com essa motivação, este trabalho visa apresentar uma técnica de avançada de classificação denominada Redes Neurais sob carteira de contratos de créditos bancários inadimplentes, para que se possa extrair padrões nos dados para que auxiliem os gestores na análise da concessão de créditos, visando diminuição da inadimplência.

ABSTRACT

The volume of data stored in organizations tends to be increasing. This is due to the organizations have the need to store the history of your activities and results of procedures performed by herself and also by interacting with customers and suppliers, and potential partnerships with other organizations. Another factor that favors this growth in data volume can be seen in falling cost of storage devices and advances in Information Technology(IT).

Humans, in general, have little ability to manually analyze so much data and so much information, possibly useful, is wasted, being hidden within the databases of organizations.

As a result, expanding the volume of data also increases the need to develop new tools and techniques for extracting knowledge from data stored. These tools and techniques have been shown to be increasingly essential. With this motivation, this paper presents advanced technique classification named Neural Networks under contract portfolio of bank loans in default, so you can extract patterns in the data to assist managers in analyzing the lending in order to decrease the default.

Sumário

1. INTRODUÇÃO

A realidade atual das empresas está inserida no contexto da globalização. Desta forma, torna-se importante que cada empresa busque estratégias para conseguir alcançar alguma vantagem competitiva. Conforme PORTER (1989), vantagem competitiva é o valor que uma empresa consegue criar para seus clientes. O valor é o montante que os compradores estão dispostos a pagar por aquilo que a empresa lhes fornece. Para que a vantagem competitiva seja efetiva, ela precisa ser difícil de imitar, única, sustentável, superior à competição e aplicável em múltiplas situações.

A vantagem competitiva no contexto deste trabalho refere-se a aplicação de melhores taxas, no momento da contratação de créditos, em relação a seus concorrentes. Para que se possa haver uma redução de taxa é necessário que se tenha cada vez menos inadimplência por parte dos clientes, pois do contrário, o risco aumenta e consequentemente as taxas também.

Para aumentar a vantagem competitiva de uma empresa em um ambiente de constantes mudanças, os seus gestores devem tomar as decisões corretas nos momentos certos, utilizando as informações disponíveis. Desta forma, o sucesso poderá ser alcançado a partir das decisões tomadas, desde que seja realizada uma exploração eficaz do relacionamento existente entre os elementos que compõem a realidade de atuação da empresa. Este relacionamento pode ser obtido através da organização e processamento de grandes bases de dados, atualmente disponíveis nas empresas devido aos constantes avanços na área da Tecnologia da Informação, gerando, assim, conhecimento a partir destes dados.

Este trabalho terá como base o processo de KDD, tomando como base a metodologia CRISP-DM (Cross-Industry Standard Process for Data Mining), que foi concebida em 1996 pelo consórcio composto por NCR Systems Engineering Copenhagen, DaimlerChrysler, SPSS (IBM SPSS) e

(OHRA Verzekeringen en Bank Groep B.V) , ultimo banco holandês. O ciclo de vida proposto de um projeto de KDD, segundo a metodologia CRISP-DM, é dividido em seis grandes fases denominadas: Entendimento do Negócio, Entendimento dos Dados, Preparação dos Dados, Modelagem, Avaliação e Implantação, demonstradas na Figura 1.1.

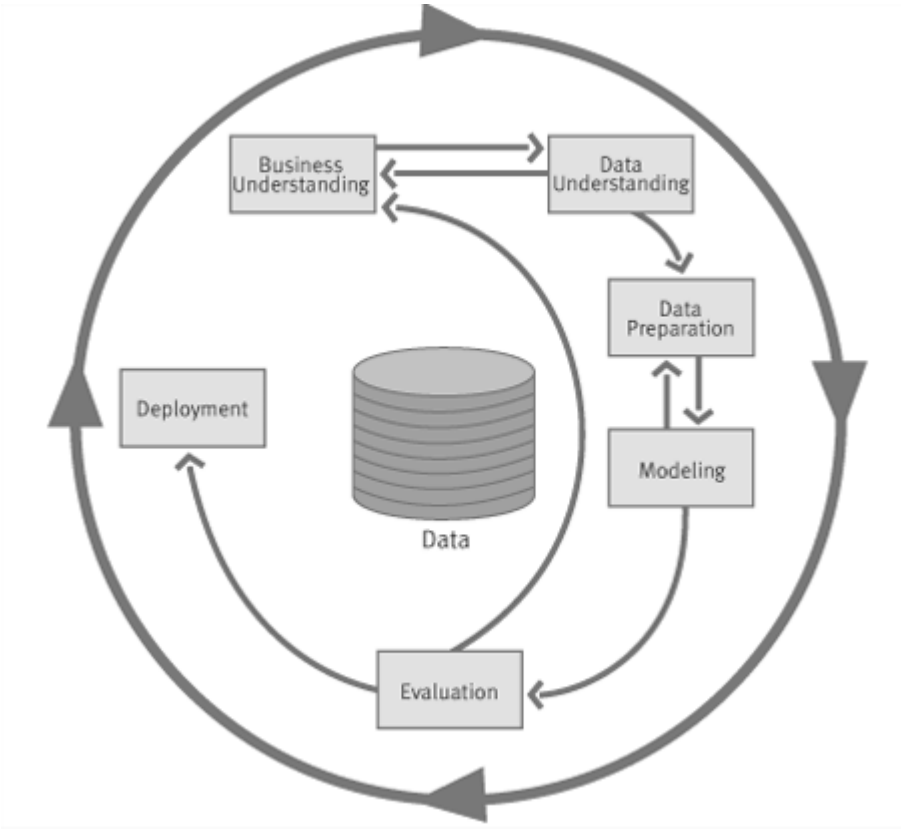

Figura 1.1: Processo de KDD - CRISP-DM (CHAPMAN et al., 2000).

A escolha desta metodologia como base deste trabalho foi motivada pela metodologia sem bem documentada e também pela clara divisão das fases do processo de KDD.

Como etapa preliminar deste trabalho, houve um estudo para identificar e caracterizar o negócio, cuja uma base de dados contendo informações de concessão de créditos será o objeto do estudo desse trabalho. Este estudo preliminar é importante para se conhecer e entender o domínio e escopo do objeto de estudo. Posteriormente será realizada a etapa de Análise da Base de Dados, que definirá o escopo dos dados alvos. Estes dados serão transformados, se necessário, e carregados na ferramenta WEKA, onde os dados serão processados.

1.1. MOTIVAÇÃO

Tomando o consumo das pessoas como um elemento efetivo para a sobrevivência do comércio pode-se dizer que uma das chaves para o desenvolvimento de uma estratégia eficiente e eficaz é a compreensão do comportamento do consumidor. Contudo, conhecer os consumidores e seus hábitos de compras com a maior assertividade possível demanda tecnologia avançada e informação estruturada e disponível.

As empresas precisam investir cada vez mais em sistemas de armazenamentos estruturados, pesquisas sistemas de transmissão de informações e programas de treinamento visando conhecer melhor seus clientes e desenvolver estratégias para obter vantagens competitivas sustentáveis perante seus concorrentes. Nesse contexto, acredita-se que a tecnologia de descoberta de conhecimento em bases de dados seja uma importante ferramenta de auxílio à tomada de decisão gerencial e estratégica, com o intuito de prover uma reação mais rápida e objetiva perante as tendências do mercado.

A correta tomada de decisão em conceder ou não crédito é essencial para a sobrevivência das instituições bancárias. Muitas vezes, o prejuízo causado pelo falta de conhecimento do perfil do cliente, no qual é dado o crédito, pode acarretar em prejuízos significativos ao credor, onde este passa a ter um contencioso passivo cada vez maior, minimizando o lucro obtido em operações de créditos bem sucedidas.

1.2. OBJETIVOS DO TRABALHO

A proposta deste trabalho é apresentar uma aplicação da tecnologia aos estudos sobre a concessão de crédito visando a minimização da inadimplência, sob uma base de dados bancária destinada a créditos a pessoas físicas e jurídicas, que será capaz de realizar classificações e descobrir eventuais padrões e perfis de pessoas boas pagadoras ou não, de forma a dar subsídios para a tomada de decisão gerencial no momento da concessão do crédito.

O objetivo principal deste trabalho é o desenvolvimento de uma rede neural artificial capaz de auxiliar a tomada de decisão no momento da concessão ou não do crédito para pessoas físicas e jurídicas, onde a rede neural será capaz de classificar se o pretendente será um bom ou mau pagador.

Os objetivos específicos propostos foram:

A) realizar um estudo de caso através baseando-se no processo de KDD e utilização da ferramenta WEKA;

B) fornecer informações relevantes ao final do processo alvo do estudo de caso;

C) dar o primeiro passo inicial para a criação de uma ferramenta de pesquisa em concessões de créditos utilizando redes neurais artificiais.

1.3. ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em seis capítulos. Dessa forma, cada capítulo fará a analogia a etapa do processo de KDD. O Capítulo II mostra conceitos e as principais características e a descrição do problema. O Capitulo III descreve o entendimento do negócio, a fonte dos dados e a qualidade destes dados, o que representa a fase Entendimento de Negócio e Dados , preparação dos dados , preparação dos Dados, etapa no qual realiza a formatação dos dados para que os mesmo estejam aptos para serem carregados na ferramenta WEKA . O Capítulo IV mostra as

conclusões retiradas do modelo implementado e explicitação dos resultados alcançados. Capítulo V considerações finais do trabalho e trabalhos futuros. O Capítulo VI esta a referencia bibligrafica.

2. DESCRIÇÃO DO PROBLEMA

Tendo em vista o cenário da globalização e consequentemente a concorrência entre as empresas, definir novas estratégia e alcançar a vantagem competitiva sustentável torna-se uma tarefa cada vez mais desafiadora. Para que uma empresa alcance seus objetivos estratégicos é preciso, muita das vezes, verificar quais informações internas ou externas estão ligadas a esta necessidade, ou seja, quais são os dados relevantes, se eles existem e como estão formatados. Para tanto, torna-se necessário um conhecimento aprofundado do domínio e dos dados em questão.

Se buscarmos na literatura financeira, veremos que o principal objetivo do administrador é a maximização dos lucros SCHERR (1989). Logo, a otimização dos recursos das empresas e a diminuição do risco possível assume papel essencial na gestão financeira.

Seguindo essa ótica, a concessão de crédito representa uma das principais formas de se ter um retorno financeiro alto, entretanto, os riscos associados na concessão de crédito são eminentes na natureza deste tipo de operação.

Em função dos altos riscos atrelados a empréstimos financeiros, principalmente riscos relacionados a inadimplência, as empresas veem continuamente obrigadas a procurara a adotar ferramentas mais eficientes para análise e controle destes riscos.

Essa necessidade aumenta a cada dia, pois cada dia mais crédito são oferecidos no mercado, logo melhoria na gestão dos riscos de crédito representa um dos fatores mais preponderantes na administração financeira.

Um dos fatores que impulsionam a alta dos juros no momento da concessão de créditos está relacionada a probabilidade da inadimplência por parte do consumidor. Isso faz com que os credores, em sua maioria baseias-se em instituições bancárias, investem em soluções de análise de crédito, para minimizar os riscos e consequentemente diminuir o índice de inadimplência.

Essa preocupação em relação ao crédito vem a ser uma vantagem para os bancos, visto que com a diminuição da inadimplência, obtém-se um lucro maior, uma vez que, pelo menos a metade dos ativos dos bancos são operações de crédito. Assim, o principal objetivo deste trabalho é a modelagem de uma rede neural artificial capaz de reconhecer padrões de comportamentos dos consumidores de crédito e discriminar pessoas físicas e jurídicas solventes ou insolventes.

3. ENTENDIMENTO DO NEGÓCIO E DOS DADOS

O crédito é um mecanismo fundamental para manutenção e crescimento econômico dos países. Assim, o crédito se torna importante na vida das empresas e pessoas. Entretanto, o uso inadequado do crédito, em escalas elevadas, poderá levar uma empresa à falência ou um indivíduo à insolvência SILVA (1993).

As empresas, de um modo geral, buscam metodologias, técnicas e ferramentas que possam agregar alguma vantagem competitiva perante seus concorrentes e proporcionar uma maior rentabilidade e segurança nas suas relações comerciais. No caso dos bancos, a realidade também é a mesma, principalmente após o processo de estabilização da moeda brasileira, onde os bancos brasileiros tiveram uma significativa redução dos lucros oriundos dos ganhos inflacionários. Esta perda dos lucros fez com que os bancos aumentassem seus volumes de crédito no mercado ALMEIDA et al. (1997). Segundo STEINER (1999), a prática de concessão de crédito é essencial para a sobrevivência das empresas bancárias. Mas, qualquer erro na decisão de concessão de crédito pode significar que em uma única transação haja o risco da perda do ganho obtido em dezenas de outras concessões bem sucedidas. Isso indica que é necessário, que antes de conceder o crédito, realizar uma analise e comparar o custo de conceder com o custo de negar a operação para cada operação, com o objetivo de minimizar as perdas pelas inadimplências.

Surge assim, a necessidade de se ter ferramentas que possam auxiliar nas decisões de conceder ou não o crédito. Ao se fazer o correto uso de ferramentas na análise de crédito, as empresas estão minimizando seus riscos e obtendo melhor direcionamento do crédito.

Neste trabalho utilizaremos uma base de dados de recuperação de credito, ou seja, os dados aqui presente representam créditos concedidos nos quais os clientes não honraram suas dívidas.

Nesta base o crédito concedido se dá através de uma diversificação de produtos, onde cada produto possuem particularidades distintas. Essa base de dados é considerada um dos maiores repositórios de dados de recuperação de crédito, contudo, por questões legais e por se tratar de um tema sensível, os dados foram transformados para que não se tenha nenhuma possibilidade de identificação dos devedores contidos nesta base.

3.1. Características dos Dados

A base de dados objeto deste estudo possui dados de contratos de concessão de crédito a partir do ano de 2002 até o ano de 20010, sendo que, os dados fornecidos estão organizados em três tabelas distintas. A Figura 3.1.1 representa o modelo de entidades e relacionamento dos dados alvos selecionados para este trabalho.

Contrato – Tabela principal que todos contratos que serão classificados nesse trabalho. Basicamente de relevante nessa modelagem temos o produto, segmentos e subsegmentos e o valor do contrato firmado.

Parte (Cliente) – A tabela de parte são dados referentes aos clientes que firmaram o contrato com a instituição financeira. Temos nessa tabela descrição sobre região (Uf , Cidade) , idade que são dados relevantes que defini padrões de aprendizagem no modelo de rede Neural.

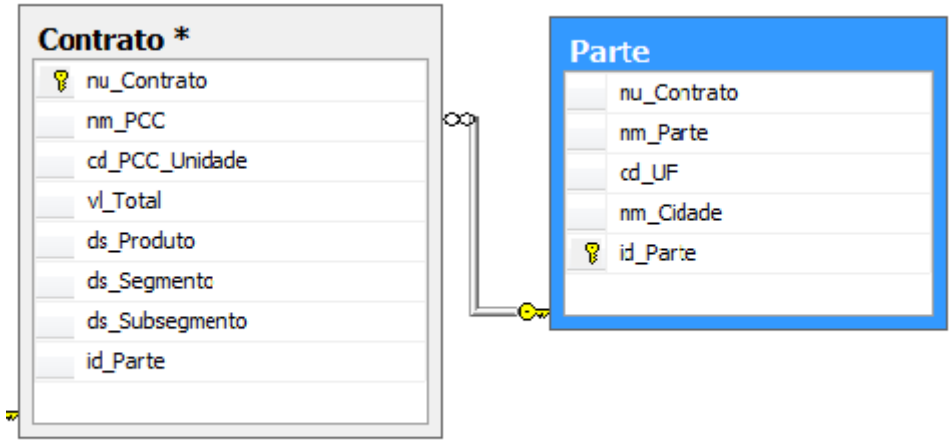

Figura 2.1.1: Modelo de Entidades e Relacionamento

As Tabelas 3.1.2, 3.1.2 representam o dicionário de dados do modelo de entidades e relacionamento representado na Figura 3.1.1.

contrato			
Coluna	Tipo Dados	Permite Nulo	Descriçao
nu_Contrato	Varchar(25)	Não	Identificador único de cada contrato cedido pela instituição Financeira
vl_Total_Contratato	money	Não	Valor Integral do Contrato cedido pela Instituição Financeira
ds_Produto	Varchar(100)	Não	Produto adquirido pelo cliente
ds_Segmento	Varchar(100)	Não	Segmento do Produto do contrato negociado
ds_Sub_Segmento	Varchar(100)	Não	SubSegmento do segmento do Produto do contrato negociado

Tabela 3.1.1: Dicionário de Dados da tabela Contrato.

Devedor			
Coluna	Tipo Dados	Permite Nulo	Descricao
id_Parte	Int	Não	Identificador único do Cliente asociado ao contrato no Banco
nu_Contrato	Varchar(25)	Sim	Identificador de Contrato relacionado a Parte
nm_Parte	Varchar(175)	Sim	Nome do Cliente asociado ao contratos no Banco
UF	char(2)	Sim	Uf relacionado ao cliente
Estado	Varchar(100)	Sim	Estado relacionado ao cliente
Idade	Int	Sim	Idade do Cliente

Tabela 3.1.2: Dicionário de Dados da tabela Cliente.

O trabalho de limpeza e seleção de registros válidos baseou-se na experiência e orientações do administrador de banco de dados que geria os dados em questão. Os dados foram extraídos da base de dados transacional e previamente saneados e formatados para após ser utilizados na classificação no software Weka.

3.2. Preparação dos dados

Segundo a metodologia CRISP-DM, a fase de preparação dos dados geralmente, é realizada diversas vezes e, após a conclusão da desta fase, os dados estarão formatados de tal forma, que estarão aptos para serem aplicado no algoritmo, que no caso deste trabalho será aplicado na ferramenta WEKA. Nesta fase é realizado o trabalho de seleção dos dados-alvos, a limpeza e transformação dos dados. Contudo, não é realizada qualquer alteração no significado dos dados, salvo dados considerados sigilosos, que foram transformados, por questões legais para que não se tenha nenhuma possibilidade de identificação dos devedores contidos nesta base.

Segundo VIANA (2004), essa etapa pode levar até 80% do tempo do processo de KDD e é considerada uma das etapas mais importantes para o sucesso do processo como um todo.

GOLDSCHMIDT e PASSOS (2005) indica que nas bases de dados das organizações é possível encontrar diversos dados incompletos, com ruídos ou inconsistentes. Entretanto, vale ressaltar que o sistema, que gerencia toda a fonte dos dados é normalizado e trabalha apenas com registros íntegros e válidos, desta forma, não foi necessário realizar nenhum tipo de método ou técnica para limpeza dos dados.

Nesta etapa de preparação, limpeza e transformação dos dados foi construído um script SQL apenas para transformar os dados para

impossibilitar a identificação dos clientes devedores, contidos na base de dados objeto de estudo deste trabalho.

3.3. Script de Transformação

O script foi desenvolvido em Transact SQL e utilizado para a transformação do dado original para camuflagem dos dados e também para preparar os dados no formato de entrada utilizado pela ferramenta WEKA.

A Tabela 3.1.3 representa o script utilizado para mascarar os dados do tipo de produto do contrato e para mascarar os nomes dos devedores que também é representado como Parte . As funções SQL reverse e Substring conforme utilizado abaixo proporciona a inverter a cadeia de caracteres e utilizar somente uma parte da mesma.

```
Update Con

   Set ds_Produto     = Reverse(REVERSE(Substring(ds_Produto,1,5))),
       ds_Segmento    = REVERSE(REVERSE(Substring(ds_Segmento,1,5))),
       ds_Subsegmento = EVERSE(REVERSE(Substring(ds_Subsegmento,1,5)))

 from Contrato Con

Update Dev

   Set nm_Parte     = Reverse(REVERSE(Substring(nm_Parte,1,5)))

 from Parte Dev
```

Figura 3.3.1: Script para mascarar os dados.

A ferramenta WEKA requer uma entrada de dados padronizado, ou seja, os dados devem ser convertidos para string e concatenados por virgulas. Sendo assim para tal, foi desenvolvido o script indicado na Tabela 3.3.2 para que os dados fossem transformados conforme padronização da ferramenta WEKA.

```
Select
      Distinct
      nu_Contrato + ',' +
      Convert(Varchar(30),idade) + ',' +
      ds_Nome + ',' +
      UF  + ',' +
      NM_ESTADO + ',' +
      Convert(Varchar(30),vl_Total_Contrato) + ',' +
      Convert(Varchar(30),dt_Inicio_Contrato) + ',' +
      Convert(Varchar(30),dt_Cadastramento)+ ',' +
      ds_Produto + ',' +
      Segmento + ',' +
      SUBSEGMENTO + ',' +
      Convert(Varchar(30),nu_Parcela) + ',' +
      Convert(Varchar(30),dt_Vencimento) + ',' +
      Convert(Varchar(30),dt_Pagamento)+ ',' +
      Convert(Varchar(30),vl_Parcela)
From tb_Load
```

Figura 3.3.2: Script para adequar os dados para a ferramenta WEKA.

A estratégia adotada nesse trabalho especificamente nesse script e concatenar os campos da tabela de carga para gerar um arquivo separado por ‘,’ para ser carregado pelo Weka. O Software Weka consegue carregar todos os dados quando separados por virgula pois tem um recurso para converter o arquivo para o formato nativo (.arff)

3.4. Modelagem da Rede Neural

A etapa de modelagem do processo de KDD, segundo o modelo CRISP-DM, é onde é definida a tarefa de mineração de dados e o algoritmo a ser utilizado. Geralmente, em um mesmo processo de KDD, podem ser utilizadas técnicas de modelagem distintas, além da associação entre as mesmas. Entretanto, devido às características de cada técnica de modelagem, o formato dos dados pode possuir características específicas. Portanto, retornar à fase de preparação de dados é frequentemente necessário.

Existem diferentes tarefas de mineração que podem ser executadas sobre uma base de dados. As tarefas correspondem aos problemas que podem ser tratados pela mineração de dados de uma forma mais ampla. As tarefas mais comuns são:

• classificação: corresponde à descoberta de um conjunto de regras de decisão que permitem classificar novas instâncias a partir de modelos obtidos dos dados já existentes. Para a classificação é necessário um prévio conhecimento das classes das instâncias disponíveis para que possa ser obtido um modelo que seja capaz de classificar novas instâncias;

• agregação: também chamada de “clusterização”, refere-se ao procedimento de agrupar as instâncias de acordo com suas características, ou atributos. Assim, deseja-se que instâncias com valores similares para os atributos fiquem em um mesmo grupo e instâncias com atributos muito diferentes sejam colocadas em grupos distintos;

• associação: procura-se, com esta tarefa, identificar associações entre valores de atributos de instâncias na base de dados. A aplicação mais conhecida para a tarefa de associação é a obtenção de regras de associação a partir de uma base de dados de vendas para tratar o problema da “análise da cesta de compras”.

Este trabalho foi utilizado a tarefa de classificação, utilizando técnicas de Redes Neurais Artificias.

3.5. Teorias de Redes Neurais Artificiais

O trabalho pioneiro sobre redes neurais foi desenvolvido, na década de 40, pelo neurobiologista Warren McCulloch e pelo matemático Walter Pitts, onde eles fizeram uma analogia entre células nervosas do organismo e o processo eletrônico em um trabalho que foi publicado como neurônios formais. O trabalho demonstrava em um modelo de resistores variáveis e amplificadores simulando conexões sinápticas de um neurônio biológico.

Os primeiros modelos de redes neurais introduzidos surgiram na década de 40, onde era realizado a simulação de máquinas. Este modelo, considerado posteriormente um modelo de redes neurais básico.

Desde então, houveram diversos trabalhos com o propósito de aperfeiçoar o modelo até então nunca evoluído. Algumas destas propostas tendem a aperfeiçoar o modelo existente de rede neural para aplicação na indústria e negócios, outras buscavam maior semelhança com os modelos biológicos originais.

Os estudos e desenvolvimento do algoritmo de treinamento backpropagation, por Rumelhart, Hinton e Williams em 1986, precedido por propostas semelhantes ocorridas nos anos 70 e 80, mostrou que é possível treinar eficientemente redes com camadas intermediárias, resultando no modelo de Redes Neurais Artificiais que são mais utilizados atualmente. Este trabalho utilizará adota a técnica de backpropagation para reconhecimento de padrões.

O backpropagation é um algoritmo de aprendizado supervisionado, que faz o treinamento de redes do tipo Perceptron Multicamadas (MLP). Durante o treinamento com o algoritmo backpropagation, a rede opera em uma sequência de dois passos. Primeiro, um padrão é apresentado à camada de entrada da rede, assim,

a atividade resultante flui através da rede, camada por camada, até que a resposta seja produzida pela camada de saída. No segundo passo, a saída obtida é comparada à saída desejada para esse padrão particular. Se esta não estiver correta, é feito o calculo do o erro. Em seguida, o erro é propagado a partir da camada de saída até a camada de entrada, e os pesos das conexões das unidades das camadas internas vão sendo modificados conforme o erro é retropropagado. Daí se deu o nome backpropagation.

Segundo KOHONEN (1998), Redes Neurais é um modelo baseado no cérebro humano para processar informações e adquirir, utilizar e evoluir o conhecimento adquirido, com uma da semelhança do cérebro humano.

3.6. Modelo proposto

A técnica utilizada na classificação de clientes inadimplentes por redes neurais foi a MultiLayer Perceptron, pelo fato da retro propagação dos erros para camadas anteriores que permite a implementação de uma rede mais estável com menor taxa de erros na saída. A saída dessa rede informa se deve ou não conceder o crédito ao proponente, dado certas características do próprio sobre a base de dados disponível objeto de estudo deste trabalho. A seguir será apresentado a ferramenta WEKA que foi utilizada para este a criação da rede.

3.7. Ferramenta WEKA

O WEKA foi desenvolvido na Universidade de Waikato, Nova Zelândia, utilizando a tecnologia Java. O WEKA é uma ferramenta para mineração de dados, que agrega o pré-processamento, e um conjunto de

algoritmos de classificação, regras de associação, regressão, e clusterização, todos implementados dentro da ferramenta FRANK et. Al, (2000). Nele, encontram-se intrinsecamente, diversos algoritmos implementados que são agrupados de acordo com as tarefas de mineração de dados disponíveis. Além dos algoritmos existentes internamente, ele pode acomodar novos algoritmos desenvolvidos pelos próprios usuários.

O WEKA requer um formato especial de arquivos textos denominados arff (attribute relation file format) e além deste input de dados, a ferramenta possibilita o acesso dos dados diretamente nas bases de dados, via interface do JDBC (Java Data Base Connectivity). No arquivo de entrada deve-se declarar no cabeçalho os atributos e os seus respectivos tipos de dados. Caso o atributo seja de valor nominal ele deve ser declarados distintamente, Após a declaração do cabeçalho vêm os registros a serem minerados.

3.8. Variáveis de Entrada da Rede Neural

As variáveis de entrada da Rede Neural estão representadas a seguir:

id_Contrato : Representa o contrato feito pelo cliente ;

idade : Idade do Cliente ;

UF – Uf da residência do cliente ;

Estado – Estado da residência do cliente ;

Vl_Total_Contrato - Valor Total do contrato ;

Produto - Produto contratado do cliente ;

Segmento - Segmento a qual pertence o produto ;

Subsegmento - Subsegmento a qual pertence o segmento

Inadimplência - Informação do Histórico que diz se o cliente pagou o não contrato

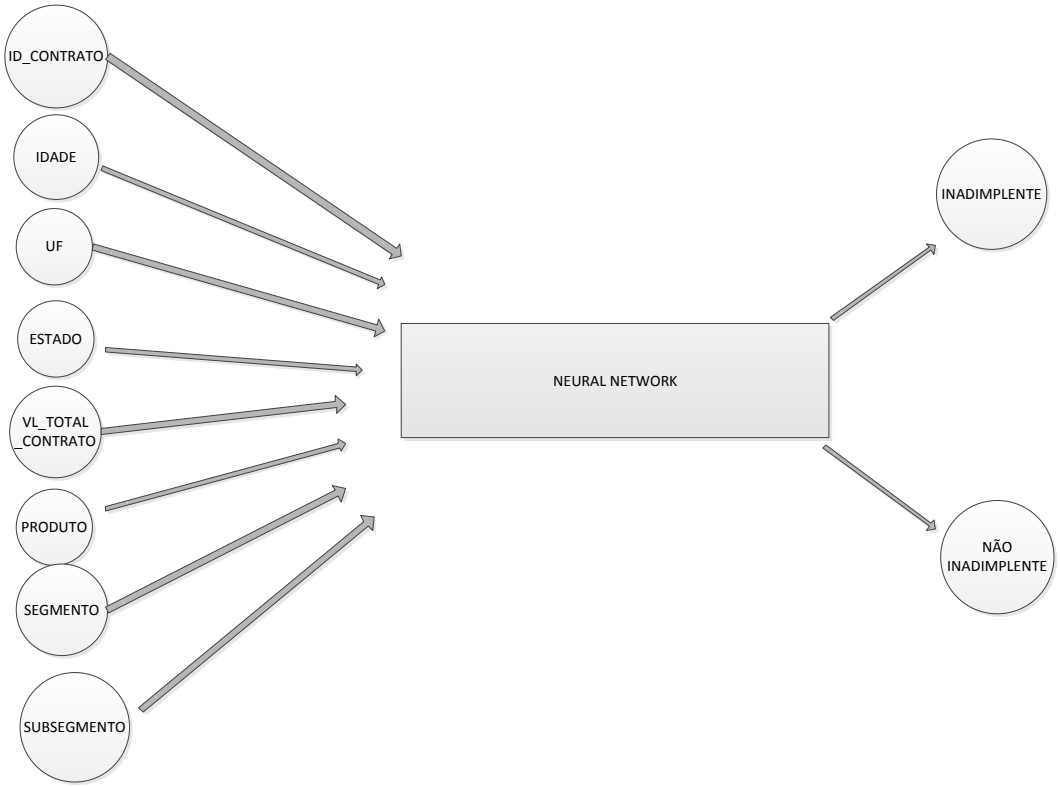

Figura 3.3.4: Entradas e Saídas da Rede Neural

A ferramenta utilizada para aplicar a classificação chamada Weka permite a normalização dos atributos, onde atributos categóricos como UF, produto, Segmento, Subsegmento, são normalizados entre (0 – 1) para efetuar a classificação sobre um modelo de rede neural . No Software Weka quando escolhemos uma classificação por Redes Neurais, nos possibilita usar um recurso chamado (**NormalizeToBinary**) que um flag (True ou False) que implicitamente o software separa distintamente todas as entradas de todos os atributos e defini um valor entre 0 e 1, independente do atributo ser categórico ou não.

3.9. Tabela de Testes

Abaixo iremos explicitar as tabela de validação e testes sobre os modelos de treinados de nossa rede neural sobre as métricas de que permitem calcular o percentual de erro na saída, acertividade e sensibilidade.

Métricas

Erro de classificação Saída : (a+d) / (b+c)

Acertividade : a /(a+b)

Sensibilidade : a/ (a+c)

As colunas (a,b,c,d) são as colunas da matriz de confusão

Exemplo Matriz

		Clientes Avaliados	
		Irregular	Normal
Teste de Redes Neurais	Irregular	a	b
	Normal	c	d

Figura 3.3.4: Matriz de classificação

A seguir vamos mostraremos os resultados de treinamentos de validações configurando a rede com camadas ocultas adequada com o resultado e tempo de treinamento aceitável. A Matrix de confusão explicita a acurácia da classificação dados apresentados na rede, ou seja, em nosso caso

Tabela 3.9.1 – Treinamento

Ex: 2\ 4 - Duas camadas e quatro neurônios na camada intermediaria.

Ciclo	*Camadas x Neurônios*	*Total de Instancias*	*ciclos*	*Learning Rate*	*Momentum*	*a*	*b*	*c*	*d*	*Erro na Saída*	*Acertividade*	*Sensibilidade*
1	1 / 5	545	500	0.8	0.2	451	29	60	5	5,12	0,93	0,88
2	1 /6	545	500	0.8	0.2	452	28	61	4	5,12	0,94	0,88
3	1 /5	545	700	0.4	0.2	438	42	56	9	4,56	0,91	0,88
4	1/4	545	100	0.3	0.2	466	14	56	9	6,8	0,97	0,89
5	1/1	545	1000	0.4	0.2	468	12	0	68	533	0,98	1
6	1/6	545	1500	0.3	0.2	480	0	1	64	544	1	0,98
7	1/7	545	1500	0.4	0.2	478	2	5	60	76,6	1	0,99
8	2/2	545	500	0.4	0.2	468	13	5	30	23,3	0,97	0,98
9	2/4	545	500	0.7	0.2	480	0	65	0	7,38	1	0,88
10	2/6	545	700	0.5	0.2	475	5	5	60	53,5	0,99	0,99
11	2/8	545	700	0.8	0.2	480	0	65	0	7,38	1	0,88

Tabela 3.9.2 - Validação

Ciclo	*Camadas x Neuronios*	*Total de Instancias*	*ciclos*	*Learning Rate*	*Momentum*	*a*	*b*	*c*	*d*	*Erro na Saída*	*Acertividade*	*Sensibilidade*
1	1 / 5	200	500	0.8	0.2	169	4	10	16	13,21	0,97	0,94
2	1 /6	200	500	0.8	0.2	160	14	5	21	9,5	0,91	0,96
3	1 /5	200	500	0.4	0.2	148	26	7	19	5,06	0,85	0,95
4	1/4	200	100	0.3	0.2	168	6	9	17	12,3	0,96	0,94
5	1/1	200	1000	0.4	0.2	168	8	8	16	11,6	0,95	0,95
6	1/6	200	1500	0.3	0.2	171	3	12	14	12,3	0,96	0,94
7	1/7	200	1500	0.4	0.2	170	4	12	14	11,5	0,98	0,93
8	2/2	200	500	0.4	0.2	166	8	8	18	11,5	0,95	0,95
9	2/4	545	500	0.7	0.2	174	0	26	0	6,69	1	0,87
10	2/6	200	700	0.5	0.2	169	5	11	15	11,5	0,98	0,93
11	2/8	200	100	0.8	0.2	480	0	65	0	7,38	1	0,88

Melhor modelo
Learning Rate 0.8
Term Momentum - 0.2
Hidden Layers - 1

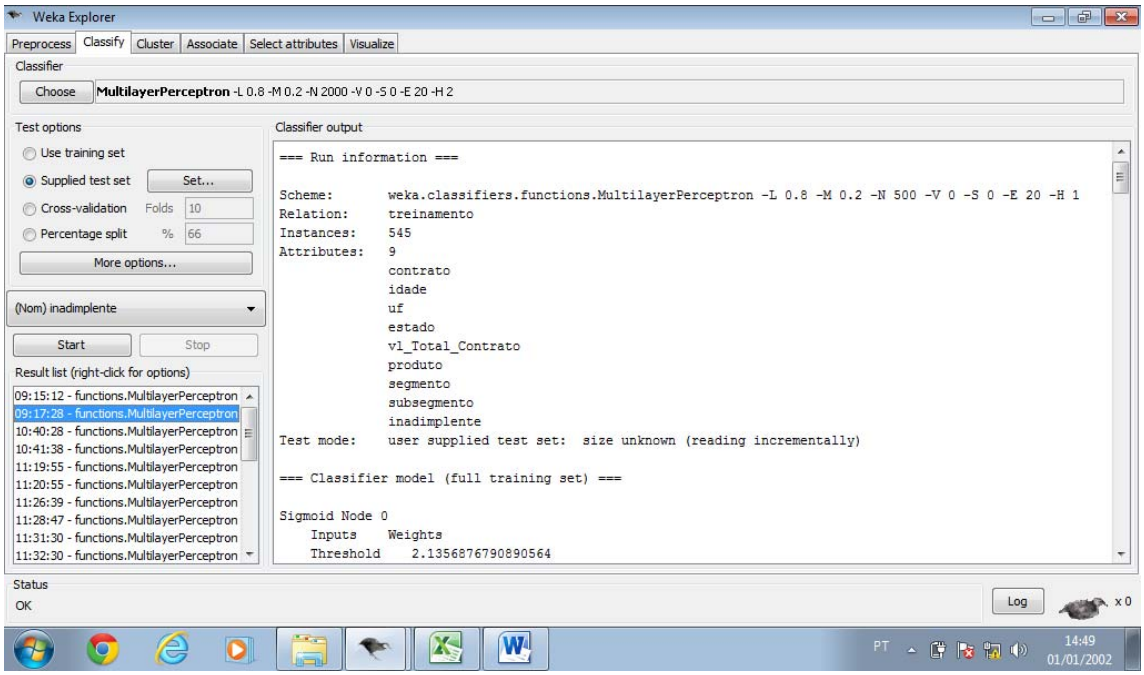

Figura 3.3.6: Sistema Weka

Com o resultado da tabela acima identificamos a melhor modelagem e executamos um Deploy de um arquivo de testes nunca apresentado a rede para validar o poder de acurácia classificação dessa técnica.

Tabela 3.9.3 - Testes

Ciclo	*camadas*	*Total de Instancias*	*ciclos*	*Learning Rate*	*Momentum*	*a*	*b*	*c*	*d*	*Erro na Saída*	*Acertividade*	*Sensibilidade*
1	1	200	500	0.7	0.2	129	0	21	0	6,14	1	0,86

4. CONCLUSÕES

Conforme demonstrado na tabela 3.9.2, mesmo com a base de dados limitada, por se tratar de uma base de dados restrita em tamanho e com conteúdo com mascara de dados, ao uso por completo, o resultado do estudo foi satisfatório, entretanto, nas diversas parametrizações realizadas na ferramenta, a o modelo Multi LayerPerceptron com 2 camadas ocultas, termo de momento configurado 0.2, a taxa de aprendizado configurado com 0.4 e com a seleção de apenas uma camada com oito neurônios foi a configuração que obteve a menor taxa de erro.

Foram realizadas diversas tentativas de configuração no modelo, onde era alterado os parâmetros da Redes na busca do melhor resultado. Alguns resultados foram melhores que outros, mas nenhum modelo se aproximou da parametrização supracitada.

5. TRABALHOS FUTUROS

A principal motivação da realização deste trabalho, bem como a escolha do tema, se dá pelo fato do crescimento gradativo da inadimplência no mercado de concessão de créditos e também pelo fato das organizações e instituições armazenarem cada vez mais informações de seus clientes em suas bases de dados. Com isso, a quantidade de informação armazenada em bancos de dados aumenta a cada dia, fazendo com que a habilidade técnica e a capacidade humana de interpretação dessa informação sejam limitadas ou até mesmo proibitivas.

Com a técnica avançada de análise de dados e a obtenção e entendimento dos padrões intrínsecos nas bases de dados obtidos, além de esconderem muitas informações valiosas, acabam trazendo alguma novidade, o que acaba sendo interessante para os gestores e os pesquisadores. A extração de informações valiosas mostra-se promissora para qualquer domínio de aplicação. Deve-se, entretanto, tomar cuidado com problemas que possam ser encontrados, como a existência de dados incompletos ou inconsistentes, que podem causar distorções nos resultados obtidos.

Com o intuito de minimizar o contencioso jurídico e consequentemente minimizar os risco de instituição financeira, este tipo de análise de dados se torna imprescindível, através de desenvolvimento de modelos capazes de dar subsídios suficientes para que os gestores tomem decisões em conceder ou não o crédito para determinado proponente ao crédito.

Como proposta para trabalhos futuros está a pesquisa de novas técnicas para aperfeiçoar o modelo da Rede Neural e a possibilidade de hibridização do com outras técnicas, como por exemplo, a utilização de Algoritmo Genético.

6. REFERENCIA BIBLIOGRAFICA

ALMEIDA, F. C. de; SIQUEIRA, J. de O. Comparação entre regressão logística e redes neurais na previsão de falência de bancos brasileiros. 3o Congresso Brasileiro de Redes Neurais, 4. Florianópolis, p. 1-6, 1997.

CHAPMAN, P; CLINTON, J; KERBER, R; KHABAZA, T; RAINARTZ, T; SHEARER, C; WIRTH, R. CRIPS-DM 1.0 Step-by-step data mining guide. 2000.

FRANK, A., Ryu, D., Jones, T. W., & Noriega-Crespo A. , ApJ, 494, L79, 1998

GOLDSCHMIDT, R.; PASSOS, E. Data Mining Um Guia Prático. Campus, 2005.

Kohonen, T. . Self-organization of very large document collections: State of the art. In Niklasson, L., Bod´en, M., and Ziemke, T., editors, Proceedings of ICANN98, the 8th International Conference
on Artificial Neural Networks, volume 1, pages 65–74, London.Springer, 1998

PORTER, M. E. A Vantagem competitiva das nações. Rio de Janeiro: Campus, 1989

SCHERR, Frederick C. Modern Working Capital Management. Prentice-Hall, 1989.

SILVA, José Pereira da. Análise e Deceisão de Crédito. 2s. ed. São Paulo: Atlas, 1993

STEINER, M. T. A. Redes Neurais. Universidade Federal do Paraná. Métodos Numéricos em Engenharia - Pesquisa Operacional, 1999.

VIANA, R. Mineração de Dados. Introdução e Aplicações. SQL Magazine, Ed. 10, Ano 1, 2004.