Universidade Federal de Pernambuco

Centro de Informática

Graduação em Engenharia da Computação

Proposta de Trabalho de Graduação

**Utilização de Técnicas de Mineração de Dados Para a Previsão de Risco de Sinistro no Mercado de Seguros**

**Aluna:** Flávia Porto Costa (fpc@cin.ufpe.br)

**Orientador:** Germano Crispim Vasconcelos (gcv@cin.ufpe.br)

Outubro de 2014

**Índice**

1. Contexto e Motivação ....................................................................................3

2. Objetivos.........................................................................................................4

3. Cronograma....................................................................................................5

4. Referências.....................................................................................................6

5. Possíveis Avaliadores......................................................................................7

6. Assinaturas.....................................................................................................8

1. **Contexto e Motivação**

A realização de previsões é uma atividade de extrema importância para projetos e empresas, pois facilitam o planejamento e gerenciamento de várias áreas, desde planejamento de produção, pessoal, financeiro, operacional, entre outros. Nos últimos anos, a grande maioria de previsões é realizada por métodos de mineração de dados através de algoritmos de aprendizagem.

Hoje em dia, temos ao nosso dispor grandes volumes de dados que podem ser utilizados como um aprendizado para futuras tarefas. A mineração de dados nos permite um processamento dessa grande quantidade de dados de um sistema de forma automática, eliminando a necessidade de tarefas repetitivas e sistemáticas.

A mineração de dados é utilizada para extrair padrões e tendências presentes no banco de dados, tais padrões e tendências são transformados em conhecimento e podem ser utilizados para resolver e simplificar problemas atuais [1].

A previsão de risco de sinistro é uma opção para as seguradoras, que permite aumentar os lucros e diminuir os riscos. Através de informações sobre os veículos dos segurados, será possível fazer uma estimativa do risco de sinistro, e dessa forma, melhorar o gerenciamento e o planejamento da empresa, como sugerido anteriormente.

Nessa pesquisa, serão utilizados dois métodos de aprendizagem de mineração de dados. O primeiro será árvore de decisão e o segundo uma rede neural.

1. Árvore de decisão:

Uma árvore de decisão toma como entrada um objeto ou situação descrito por um conjunto de atributos e retorna uma decisão - o valor de saída previsto, de acordo com a entrada [2].

1. Rede Neural:

As Redes Neurais Artificias (RNAs) são sistemas compostos por unidades de processamento simples, chamadas de nodos ou neurônios, que estão distribuídos em arranjos espaciais composto por uma ou mais camadas e interligados por conexões. Tais conexões são associadas a pesos que armazenam o conhecimento do modelo e servem para criar uma ponderação entre os sinais de entrada dos nodos predecessores [3].

O estudo será uma oportunidade de consolidar e aprofundar o conhecimento nas áreas de Mineração de Dados, Aprendizagem de Máquina e Inteligência Artificial a partir de um problema complexo do mundo real. Acredita-se também, que a realização desta pesquisa trará uma forma automática e eficiente de prever riscos de sinistro beneficiando, de forma geral, as empresas seguradoras.

1. **Objetivo**

O objetivo geral desse estudo é realizar uma análise do desempenho das duas técnicas de aprendizagem sugeridas acima quando aplicadas ao banco de dados para avaliar o risco de sinistro.

O banco de dados utilizado para a pesquisa será proveniente da competição *Claim Prediction Challenge* [http://www.kaggle.com/c/ClaimPredictionChallenge], que utiliza dados de clientes reais, que utilizaram seguros entre 2005 e 2007.

Para alcançar o objetivo geral, temos como objetivos específicos:

* Seleção dos dados disponíveis
* Pré-processamento dos dados obtidos, onde informações julgadas desnecessárias serão removidas.
* Aplicação dos algoritmos de aprendizagem propriamente ditos (árvore de decisão e rede neural)
* Interpretação e avaliação dos resultados obtidos

1. **Cronograma**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Atividades** | Outubro | | | Novembro | | | Dezembro | | | Janeiro | | | Fevereiro | | |
| Levantamento bibliográfico |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Pré-processamento dos dados. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Aplicação dos algoritmos de Árvore de Decisão e Rede Neural |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Análise dos resultados |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Elaboração do relatório final |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |
| Elaboração da apresentação |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |

1. **Referências**

[1] Andrea Dal Pozzolo. Comparison of Data Mining Techniques for Insurance Claim Prediction, 2011.

[2] Juliana Moreira Barbosa, Tiago Garcia de Senna Carneiro, Andrea Iabrudi Tavares. Métodos de Classificação por Árvores de Decisão Disciplina de Projeto e Análise de Algoritmos.

[3] Alvaro Joao Sales Silva, Análise de desempenho dos sistemas híbridos baseados em MLP e PSO para previsão de séries temporais, 2014.

1. **Possíveis Avaliadores**

* Cleber Zanchettin {cz@cin.ufpe.br}

1. **Assinaturas**

Recife, 23 de outubro de 2014.

|  |  |
| --- | --- |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Flavia Porto Costa  (Aluno) | Germano Crispim Vasconcelos  (Orientador) |