**1) Nome e número USP.**

Raphael Donaire Albino – 8240823.

**2) Título do artigo escolhido, breve descrição (problema de pesquisa e objetivo) e técnica utilizada.**

**Artigo**

* **Título:** *Customer churn prediction in telecommunications.*
* **Ano de publicação:** 2012.
* **Revista:** [Expert Systems with Applications](http://dx.doi.org/10.1016/j.eswa.2011.08.024).
* **Fator de impacto:** 2,981.
* **Classificação Capes:** A1.

**Descrição:**

A partir da problemática que as empresas de serviço de telecomunicação sofrem com a evasão dos seus clientes, o artigo teve como objetivo apresentar um conjunto de variáveis que auxiliarão na identificação das chances dos clientes romperem os contratos de serviço.

Para alcançar o objetivo, os autores avaliaram as variáveis apresentadas na Tabela 1 a partir de 7 técnicas de modelagem: regressão logística; Naive Bayes; classificação linear; árvore de decisão; redes neurais artificiais multicamada; máquinas de vetores de suporte; algoritmo de mineração de dados evolucionária.

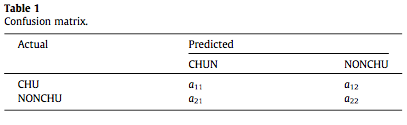
Tabela - Variáveis utilizadas nos modelos

|  |  |
| --- | --- |
| **Variáveis** | **Descrição** |
| 2 six-month Henley segmentation | Divisão dos clientes a partir de suas características, necessidades e valor comercial. |
| Precise 4- month call details | Duração das ligações, número de ligações recebidas, número de mudanças na duração das ligações nos últimos quinze dias. |
| Information of grants | Subsídio de terceiro no pagamento das contas. |
| Line information | Tipos de pacotes, indicadores de controle de crédito, data de primeiro uso do serviço, aluguéis de equipamentos, cobranças pagas, etc. |
| Bill and payment information | Frequência de pagamento das contas. |
| Account information | Número de linhas de telefone, tipos de linha, código do distrito. |
| Demographic profiles | Idade, classe social, gênero, etc. |
| Service orders that are extracted from existing limited information | Quantidade de serviços solicitados pelos clientes, taxas de aluguel e data de aprovação dos serviços. |

Vale ressaltar que no artigo os autores descrevem em detalhes quais foram as inovações quanto as definições das variáveis mencionadas acima. Outro aspecto metodológico relevante mencionado pelos autores foi o trabalho de normalização aplicado nas variáveis para que os modelos fossem executados.

Foram utilizados como critérios de avaliação da qualidade dos modelos a taxa de evasão verdadeira (TP) e a taxa de evasão falsa (FP). TP é resultado da razão entre o número de predições corretas de evasão (a11) dividido pela soma do número de predições corretas (a11) e incorretas (a12) de evasão. Já FP foi calculado a partir da razão entre o número de predições incorretas de não evasão (a21) pela soma do número de predições incorretas (a21) e corretas de não evasão (a22).





Para a análise das taxas foi utilizada a técnica da curva ROC a fim de encontrar os pares das taxas de predição. Para calcular os valores acima da curva ROC foi utilizada a seguinte fórmula:



Onde S0 é a soma dos rankings dos testes padrão da classe 0 (evasão), n0 o número de testes padrão no teste que pertencem a classe 0 (evasão) e n1 o número de testes padrão no teste que pertencem a classe 1 (não evasão).

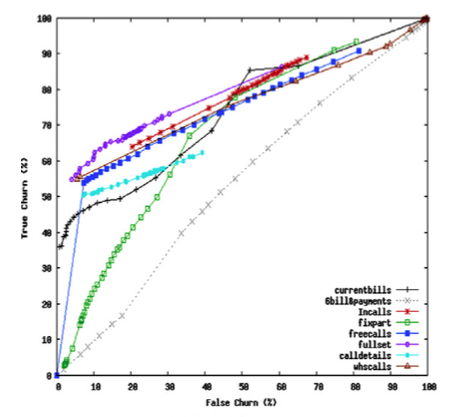
Para a execução dos experimentos os autores utilizaram uma amostra de 827.124 clientes selecionados randomicamente de uma base de dados provida por empresas de telecomunicação da Irlanda. Na base de dados de treinamento, os autores utilizaram 13.562 clientes que romperem o contrato de serviço (evasão) e 400.000 que mantiveram o contrato de serviço. Para a base de teste foram utilizados os mesmos números de clientes usados na base de treinamento. Cada cliente foi representado por um conjunto de 738 características oriundas das variáveis mencionadas anteriormente.

**3) Para a técnica em análise, identificar no artigo os passos desenvolvidos e utilizados da técnica, comparando com o que foi aprendido na disciplina.**

Assim como aprendido na disciplina, os autores utilizaram o Naive Bayes como forma de encontrar a probabilidade de um cliente parar ou não de utilizar o serviço de telefonia. O objetivo dos autores era alcançar a maior taxa de evasão verdadeira com a menor taxa de evasão falsa.

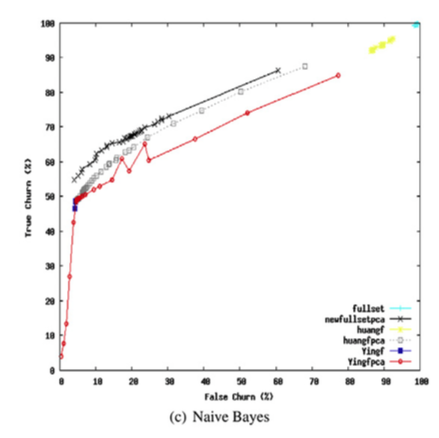
Os autores descreveram a qualidade do modelo a partir da acurácia das predições, porém, não mencionaram o uso do estimador de Laplace. Além disso, conforme ilustrado na Figura 1, os autores utilizaram a curva ROC para analisar a precisão das predições para cada uma das novas variáveis propostas.

Figura - Curva ROC da predição das novas variáveis propostas para a previsão de evasão



Outro ponto interessante identificado no artigo foi uma comparação da qualidade de predição das novas variáveis frente a outros estudos. As linhas em azul e vermelho na Figura 2 representam predições realizadas a partir de modelos construídos por outros autores. Conforme pode ser observado, os resultados das predições obtidas na base de dados estudada a partir do novo conjunto de variáveis propostos pelos autores é melhor do que os modelos que foram utilizados como referência.

Figura – Curva ROC comparando o modelo de predição proposto pelos autores e outros modelos desenvolvidos



A partir de uma comparação das predições entre os métodos, os autores concluíram que o método Naive Bayes obteve o pior resultado de predição por conta do alto número de variáveis.

**4) Descrever os pontos forte e fracos do artigo.**

Tabela - Pontos fortes e fracos

|  |  |
| --- | --- |
| **Pontos fortes** | **Pontos fracos** |
| - Embasamento teórico para a construção das variáveis aplicadas no modelo.  - Comparação entre diferentes técnicas para medir a qualidade das predições.  - Apresentação de quais respostas esperar de cada técnica utilizada para avaliar a evasão dos clientes.  - Os resultados apresentados geraram implicações teóricas e práticas. | - Falta de explicação em profundidade sobre os modelos criados a partir das técnicas. |

**5) Considere que você seja o revisor do artigo; que críticas ou sugestões você poderia considerar (caso existam).**

**Sugestões:**

* Detalharem melhor a composição das variáveis que foram utilizadas. Os autores falaram de forma macro.
* Explicar em mais detalhes as aplicações das técnicas. Os autores poderiam quebrar o assunto em um conjunto de artigos mais aprofundados explicando os modelos.

**6) Faça uma auto avaliação do que aprendeu (ou não) com essa atividade.**

Das técnicas que aprendi até aqui na disciplina, redes bayseanas foi aquela que tive mais dificuldade em acompanhar, talvez por ser a que eu não tinha nenhum tipo de conhecimento prévio. Pelo fato de ter perdido uma aula sobre a técnica, avalio que tive meu aprendizado prejudicado.

Ao ler este artigo aprendi como aplicar o conceito sobre redes bayseanas e como analisar a qualidade das predições.