

## 6

### Conclusão

No presente trabalho, foram comparadas as seguintes metodologias utilizadas para classificação: *Generalized Additive Models (GAM)*, *Regressão Logística*, *Redes Neurais*, *Análise Discriminante*, *k-Vizinhos mais Próximos (k-NN)* e *Classification and Regression Trees (CART)*, com o modelo STLR-Tree, o qual é uma adaptação do modelo STR-Tree aplicado na classificação de variáveis binárias.

Para este fim, algumas alterações foram realizadas em sua forma estrutural e na estimação. A parte linear passou a ser estimada por Máxima Verossimilhança, através do método iterativo de Newton-Raphson, tal como em uma Regressão Logística e seus parâmetros não-lineares são estimados através de uma procura em *grid*. Os testes Multiplicadores de Lagrange, que determinam a inclusão de novos regimes no modelo, eram realizados através de estatísticas de teste onde se calculava a soma dos quadrados dos resíduos no caso Gaussiano, passou a ser feita a ser feito com o cálculo da função desvio, que é o equivalente para o caso não Gaussiano, onde as variáveis dependentes pertençam à família exponencial. Essas mudanças se devem ao fato do modelo STLR-Tree não apresentar uma estrutura sistemática do erro, dado a distribuição de sua variável dependente.

Com as duas bases de dados utilizadas, selecionou-se dois conjuntos de variáveis explicativas de cada uma delas. Esses conjuntos foram escolhidos pois, na avaliação das tabelas de classificação, apresentaram os melhores resultados. Sendo assim, os demais métodos foram aplicados a esses quatro conjuntos de variáveis.

Os resultados obtidos pelo STLR-Tree foram bastante satisfatórios, sendo ele detentor da segunda melhor taxa de acerto total nas aplicações: *Spam* e *Consumo de Energia*, na análise *in sample*. Ambos provavelmente conseguiram captar uma relação não-linear entre as variáveis que os demais não fizeram. Na aplicação *DCAS* o modelo ficou atrás apenas atrás de GAM e CART para a mesma taxa. Atentamos para o fato dos resultados do exemplo *Fraude no Consumo de Energia Elétrica* terem sido trabalhados e ajustados especificamente em um estudo sobre Redes Neurais.

Na análise *out-of-sample*, o modelo se manteve na mesma posição em comparação aos demais métodos nas aplicações *Spam* e *DCAS*, porém, para no *Fraude no Consumo de Energia Elétrica* acabou sendo ultrapassado poucos pontos percentuais por Análise Discriminante.

Além disso, o modelo se apresentou bastante parcimonioso nos dois primeiros exemplos, não passando de dois nós terminais na estrutura das árvores. Como comparação, temos o CART que se estruturou com 13 nós terminais para os exemplos da base *Spam* e na base *DCAS*. Já no último exemplo, *Fraude no Consumo de Energia Elétrica*, o número de nós terminais do STLR-Tree foi maior do que o CART, 5 e 4 nós respectivamente. Entretanto o primeiro foi bastante superior em duas das três taxas de acertos avaliadas (total, para 1 (sucesso), obtendo um acerto total de 68.48% contra 59.92%.

Concluimos com isto, que o modelo se adaptou bem as alterações realizadas, mostrando-se uma alternativa em análises não-lineares a ser utilizada para classificação. Porém deve ainda ter seu desempenho computacional melhorado, objetivando minimizar o tempo de duração de suas aplicações e, além disso, ser adaptado para classificação não apenas binária. Neste caso a distribuição da variável dependente não é Binomial, e sim *Multinomial*.