

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística



Anais de Resumos

XI SEMANÍSTICA

XI Semana Acadêmica do Departamento de Estatística
da UFRGS

<http://www.ufrgs.br/semanistica>

Porto Alegre - 27, 28 e 29 de outubro de 2021

Mini Palestra 8:

Estudo comparativo entre abordagens de Machine Learning em modelos de Credit Scoring

Cinthia Becker, Lisiane Priscila Roldão Selau

Resumo: Com o crescimento da demanda e popularização do mercado de crédito no Brasil, as empresas estão buscando maneiras de aprimorar a assertividade na hora de conceder crédito. Estudos recentes mostram que os métodos de Inteligência Artificial têm alcançado melhor desempenho que os métodos estatísticos tradicionais, sendo assim, este trabalho introduz técnicas de Machine Learning ainda pouco estudadas em crédito (Árvore de Decisão, Random Forest, Bagging, Adaboost e Support Vector Machine), a fim de fornecer um modelo com melhor poder explicativo. Para fins de comparação, adotou-se a abordagem tradicional de Regressão Logística. Todos os modelos foram desenvolvidos em uma base de dados real e foram avaliados sob o mesmo conjunto de validação, com base em três indicadores: percentual de acerto, área abaixo da curva ROC e teste KS. O modelo que apresentou melhor desempenho nos três indicadores avaliados e em ambas amostras de estudo foi o Adaboost, sendo esta uma técnica a ser levada em consideração na hora da criação de um modelo de Credit Scoring. No entanto, a superioridade encontrada na técnica mencionada pode ser considerada pouco significativa, o que sugere que devido a dificuldade de interpretação e implementação, pode não valer a pena usá-la quando comparada com a técnica padrão de Regressão Logística.

Mini Palestra 9:

Transformações de atributos: estudo do impacto de diferentes combinações de técnicas na predição de modelos de Credit Scoring

Cinthia Becker, Henry Cagnini

Resumo: Algoritmos de classificação são amplamente utilizados no contexto de crédito, porém, para garantir um bom percentual de acerto, é necessário que a preparação dos dados seja a mais completa e consistente possível. Com a finalidade de automatizar o processo de transformação de atributos, este trabalho visa investigar diversas combinações de técnicas, além de analisar quais os impactos que elas desencadeiam nos modelos de crédito. Foram criadas três combinações de transformações diferentes a partir das técnicas: Discretização, Numérico para Nominal, Normalização e One-hot Encoding.