

# PROJETO DE PESQUISA

## Regressão Logística com Seleção de Amostra State-Dependent Aplicada a Dados de Credit Scoring

Aluno: Katia Sousa

Orientador: Paulo Silva

### 1. Justificativa e importância

(justificativas e importância do projeto de pesquisa em termos de relevância para a área)

A classificação de clientes é de fundamental importância para se determinar uma possível concessão de crédito. De acordo com Diniz e Louzada (2012), modelos estatísticos de classificação têm sido utilizados por instituições financeiras como a principal ferramenta para auxílio na concessão de crédito a clientes. Modelos de classificação em credit scoring são construídos a partir de bases de dados contendo informações relevantes dos clientes (tais como sexo, idade, estado civil e renda, entre outras), com a performance financeira dos clientes, avaliada a partir do tempo em que o relacionamento cliente-empresa começou, como uma classificação dicotômica.

O objetivo de modelos de Credit Scoring consiste, então, em classificar e prever clientes como bons ou maus pagadores. Neste contexto, várias técnicas de classificação, tais como análise discriminante, árvores de classificação, regressão logística, support vector machine, redes Bayesianas e redes neurais, entre outras, são comumente utilizadas.

Na presença de dados desbalanceados, em que a amostra considerada para o desenvolvimento de um modelo de classificação é desbalanceada, com uma das classes contendo apenas uma pequena parcela dos indivíduos, modelos de classificação com seleção de amostra state-dependent podem ser considerados. Em Credit Scoring, por exemplo, espera-se que o grupo de bons pagadores seja o grupo predominante. De acordo com Ferreira, Louzada, Diniz (2015), esta técnica promove uma correção na probabilidade de inadimplência estimada a partir de um modelo de classificação qualquer.

### 2. Objetivos

(Objetivos gerais e específicos do Projeto de Pesquisa)

Visto que na literatura, seleção de amostra state-dependent tem sido considerada para os modelos de regressão logística, neste projeto de pesquisa propõe-se aplicar esta técnica corretiva considerando, em particular, dados de credit scoring de um determinado portfólio de um banco brasileiro.

Objetivos específicos:

- Estudar os modelos de classificação mais comumente empregados em credit scoring, com enfoque em modelos de regressão logística;
- Comparar a performance dos modelos regressão logística usual com modelos de regressão logística com seleção de amostra state-dependent, por meio de estudos de simulação considerando diferentes cenários de desbalanceamento, utilizando um software apropriado, como, por exemplo, o Python, com o uso de medidas de avaliação da capacidade preditiva, como sensibilidade, especificidade, acurácia, valores preditivos positivo e negativo, entre outras;
- Aplicar a metodologia estudada em um conjunto de dados real de um portfólio de um banco brasileiro.

### 3. Metodologia

(Metodologia a ser utilizada no Projeto de Pesquisa)

Primeiramente, será feito um levantamento bibliográfico acerca das principais técnicas de classificação empregadas em problemas de Credit Scoring, bem como referências sobre a técnica corretiva proposta. Em seguida, mediante o uso de simulação, será feita uma comparação entre as probabilidades de inadimplência e medidas preditivas dos modelos de classificação sem e com state-dependent. Na sequência, a metodologia será aplicada em um conjunto de dados reais extraído de um portfólio de um banco brasileiro.

#### 4. Cronograma

(Relação itemizada das atividades previstas, em ordem sequencial e temporal, de acordo com os objetivos traçados no projeto e dentro do período de um ano)

O tempo previsto para a realização do trabalho é de 11 meses, sendo que o plano de trabalho é composto pelas 4 etapas a seguir:

- a) Estudo sobre técnicas de classificação e levantamento bibliográfico (2 meses);
- b) Comparação entre as principais técnicas estudadas via simulação (3 meses);
- c) Aplicação dos modelos em dados reais (3 meses);
- d) Discussão dos resultados obtidos e confecção do TCC (3 meses).

#### 5. Resultados e Impactos Esperados

(Relação dos resultados ou produtos que se espera obter após o término da pesquisa)

Dentre os resultados esperados, destacamos:

- a) O desenvolvimento de um modelo de credit scoring com seleção de amostra state-dependent para o portfólio do banco brasileiro em questão.
- b) A implementação do modelo desenvolvido.

#### 6. Referências Bibliográficas

(Relação itemizada das referências que subsidiam a proposta de pesquisa em ordem alfabética, com no máximo 10 referências)

DINIZ, C.; LOUZADA, F. **Modelagem Estatística para Risco de Crédito**. São Paulo: ABE, 2012. 178p. (Minicurso – 20º SINAPE).

FERREIRA, P. H.; LOUZADA, F.; DINIZ, C. Credit Scoring modeling with state-dependent sample selection: a comparison study with the usual logistic modeling. **Pesquisa Operacional**, v. 35, n. 1, p. 39-56, 2015.

LOUZADA, F.; FERREIRA-SILVA, P. H.; DINIZ, C. A. R. On the impact of disproportional samples in credit scoring models: an application to a Brazilian bank data. **Expert Systems with Applications**, v. 39, n. 9, p. 8071-8078, 2012.