

Evento: III Encontro Nacional de Estudantes de Atuária (ENEAT)

Proposta de Minicurso: **Modelagem Preditiva - Uma Abordagem de *Statistical Machine Learning***

Carga Horária: 4 horas

Palestrante: Lucas Pereira Lopes – Mestrando em Estatística no programa interinstitucional ICMC – USP e UFSCar e graduado em Ciências Atuariais pela Universidade Federal de Alfenas – MG. Suas principais linhas de pesquisa são: modelos preditivos (*Machine Learning*), riscos financeiros, séries temporais e análise de dependência por meio de cópulas - <http://lattes.cnpq.br/5968999804623198>

Objetivo: Este minicurso tem por objetivo de introduzir aos interessados a algumas ideias e técnicas de aprendizado de máquina com o objetivo em modelagem preditiva. Os tópicos abordados serão: regressão (técnicas onde a variável resposta é contínua) e classificação (variável resposta é qualitativa).

Ao final deste minicurso, o aluno deverá:

- Saber descrever um problema de predição matematicamente;
- Saber comparar a performance de diferentes algoritmos preditivos;
- Conseguir implementar algoritmos básicos de predição;
- Saber as potencialidades dessas técnicas em suas respectivas áreas de pesquisa.

Conteúdo:

1º Parte = Regressão

1. Introdução
 - a. As duas culturas de modelagem de dados
 - b. Seleção de modelos = *underfitting* e *overfitting*
 - c. *Tuning Parameters*
2. Métodos Paramétricos
3. Métodos Paramétricos em Alta Dimensão
 - a. Regressão *Stepwise*
 - b. Regressão LASSO
 - c. Regressão *Ridge*
4. Métodos Não-paramétricos
 - a. K-NN
 - b. *Nadaraya-Watson*
 - c. *Reproducing Kernel Hilbert Spaces*
 - d. Árvores de Regressão
 - e. *Bagging*
 - f. Florestas Aleatórias

2º Parte = Classificação

1. Introdução
 - a. Função de risco; seleção de modelos; balanço entre viés-variância; medidas de qualidade
2. Métodos de classificação
 - a. Classificadores *Plug-in*
 - i. Métodos de regressão

- ii. Regressão Logística
 - iii. *Naive-Bayes*
 - iv. Análise Discriminante
- b. *Support Vector Machine (SVM)*
- c. Árvores de Classificação
- d. *Bagging* e Florestas Aleatórias
- e. KNN
- 3. Outros Aspectos de Classificação
 - a. Conjunto de dados desbalanceados
 - b. Combinando classificadores

Bibliografia:

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. **The elements of statistical learning**. New York: Springer series in statistics, 2001.

IZBICKI, R. Machine Learning sob a ótica estatística, (2016), rizbicki.wordpress.com/teaching/

JAMES, G. et al. **An introduction to statistical learning**. New York: springer, 2013.