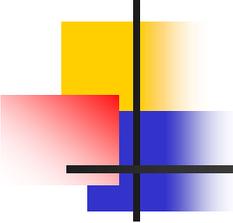


# Tópicos do curso

---

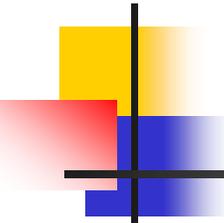
- Aprendizado de Máquina
- Algoritmos de indução de árvores de decisão
- Algoritmos probabilísticos
- Redes neurais
- SVM e Deep Learning
- Agrupamento de dados



# Objetivo

---

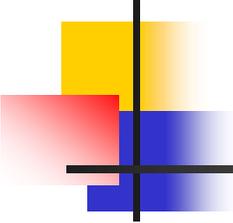
- Apresentar os aspectos fundamentais e principais algoritmos de aprendizado de máquina, que investiga técnicas para desenvolver algoritmos capazes de aprender, ou melhorar seu desempenho, utilizando exemplos de situações previamente observadas.



# Apresentação e expectativas

---

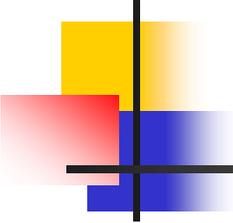
- Qual minha formação
- O que espero do curso?
- No que ele vai ajudar minha dissertação / tese?



# Ementa

---

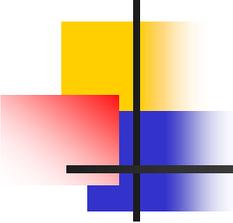
- Aspectos básicos de Aprendizado de Máquina
- Tarefas de aprendizado
- Viés indutivo
- Aprendizado descritivo
- Aprendizado preditivo
- Algoritmos de Aprendizado de Máquina
- Algoritmos que seguem diferentes paradigmas, incluindo algoritmos baseados em procura (algoritmos de indução de arvores de decisão e de conjuntos de regras, redes neurais artificiais (perceptron e multilayer perceptron) e modelos probabilísticos (regressão logística e naive Bayes)
- Medidas de avaliação; Aplicações de Aprendizado de Máquina.



# Exercícios

---

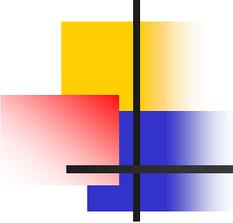
- Por em prática o que for visto durante o curso
  - Preparação de dados
  - Implementação
  - Realização de experimentos
  - Análise de resultados
  - Bem escrito



# Projeto

---

- Utilizar algoritmos de AM vistos em aula para resolver problema real
  - Ligado a dissertação ou tese
  - Dados públicos
  - Artigo científico formato LNCS
    - 8 a 10 páginas, coluna simples
    - Mestrado: português
    - Doutorado: inglês



# Avaliação (PCCMC)

---

- $NF = (5 * N_{PV} + 4 * N_{Pj} + 1 * N_R) / 10$

- Onde:

- $N_{PV}$ : nota da prova

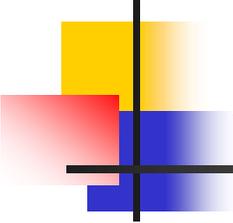
- $N_{Pj}$ : nota do projeto

- $N_R$ : nota dos relatórios

- Se algumas das notas  $< 5$

- MF = menor valor entre as notas

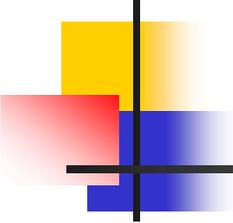
- **Não haverá prova substitutiva nem recuperação**



# Avaliação (MECAI)

---

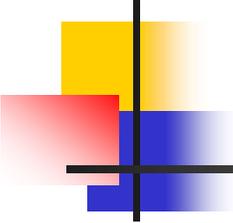
- $NF = (5 * N_{pV} + 5 * N_{pJ}) / 10$
- Onde:
  - $N_{pV}$ : nota da prova
  - $N_{pJ}$ : nota do projeto
  - Se algumas das notas  $< 5$ 
    - MF = menor valor entre as notas
- **Não haverá prova substitutiva nem recuperação**



# Práticas

---

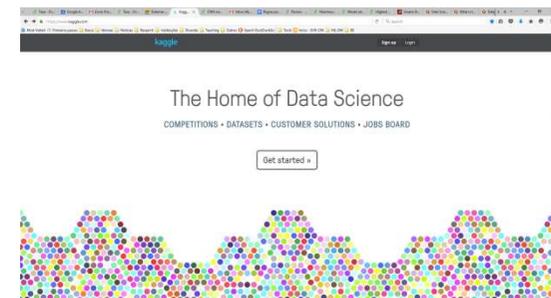
- Grupos de até 5 pessoas
  - Grupos com pelo menos um aluno de mestrado acadêmico/doutorado
    - Usar R ou Python
    - Aula pratica
  - Grupos apenas com alunos do MECAI
    - Utilizar ferramentas para mineração de dados
    - R, Python, WEKA, KNIME ou Rapid Miner
  - Cada grupo deve ter um coordenador

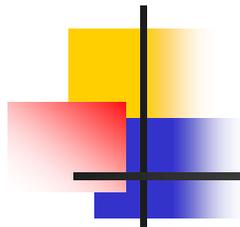


# Práticas (PCCMC)

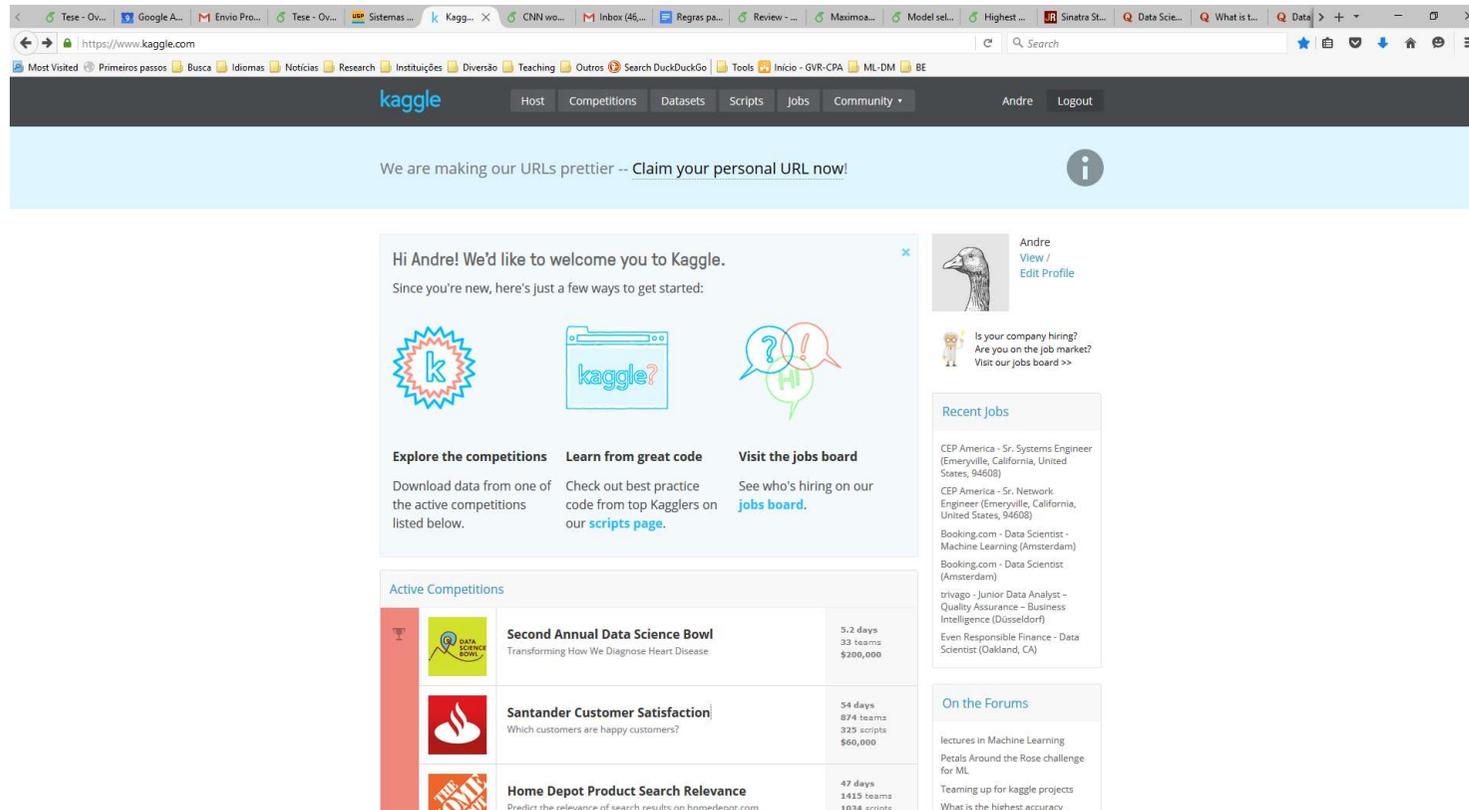
---

- Aplicar conceitos vistos em conjunto de dados do Kaggle
  - <https://www.kaggle.com/>
  - Problema de classificação
  - Práticas das 10:00 as 12:00
  - Relatórios semanais
    - Até duas páginas (pdf)
    - Falar o que foi feito





# Projeto - Kaggle



The screenshot shows the Kaggle website interface. At the top, there's a navigation bar with the Kaggle logo and links for Host, Competitions, Datasets, Scripts, Jobs, and Community. A user profile for 'Andre' is visible with a 'Logout' button. A light blue banner below the navigation bar says 'We are making our URLs prettier -- Claim your personal URL now!' with an information icon.

A welcome message for Andre is displayed: 'Hi Andre! We'd like to welcome you to Kaggle. Since you're new, here's just a few ways to get started:'. Below this are three icons with corresponding text:

- Explore the competitions**: Download data from one of the active competitions listed below.
- Learn from great code**: Check out best practice code from top Kagglers on our [scripts page](#).
- Visit the jobs board**: See who's hiring on our [jobs board](#).

On the right side, there's a user profile for Andre with a 'View / Edit Profile' link. Below the profile is a 'Recent Jobs' section listing several job openings, including 'CEP America - Sr. Systems Engineer' and 'Booking.com - Data Scientist - Machine Learning (Amsterdam)'. There is also a 'Jobs board' link with the text 'Is your company hiring? Are you on the job market? Visit our jobs board >>'. At the bottom right, there's a section titled 'On the Forums' with links to 'lectures in Machine Learning', 'Petals Around the Rose challenge for ML', and 'Teaming up for kaggle projects'.

The 'Active Competitions' section is a table with the following data:

Competition	Description	Days	Teams	Scripts
 <b>Second Annual Data Science Bowl</b>	Transforming How We Diagnose Heart Disease	5.2 days	33 teams	\$200,000
 <b>Santander Customer Satisfaction</b>	Which customers are happy customers?	54 days	874 teams	325 scripts
 <b>Home Depot Product Search Relevance</b>	Predict the relevance of search results on homedepot.com	47 days	1415 teams	1034 scripts

# Projeto

- Competição selecionada
  - Direção segura
    - <https://www.kaggle.com/c/porto-seguro-safe-driver-prediction/>



The screenshot shows the Kaggle competition page for "Porto Seguro's Safe Driver Prediction". The page features a header with the competition title, a prize money amount of \$25,000, and a description: "Predict if a driver will file an insurance claim next year." Below the header, there is a navigation menu with options: Overview, Data, Kernels, Discussion, Leaderboard, Rules, Team, My Submissions, and Late Submission. The "Data Description" section is visible, providing details about the competition's goal and data features.

Featured Prediction Competition

**Porto Seguro's Safe Driver Prediction** \$25,000  
Prize Money

Predict if a driver will file an insurance claim next year.

Porto Seguro · 5,169 teams · a year ago

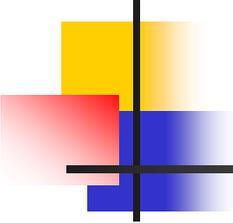
Overview Data Kernels Discussion Leaderboard Rules Team My Submissions **Late Submission**

Data Description

**Data Description**

In this competition, you will predict the probability that an auto insurance policy holder files a claim.

In the train and test data, features that belong to similar groupings are tagged as such in the feature names (e.g., `ind . reg . car . ca1c`). In addition, feature names include the postfix `bin` to indicate binary features and `cat` to indicate categorical features. Features without these designations are either continuous or ordinal. Values of `-1` indicate that the feature was missing from the observation. The `target` column signifies whether or not a claim was filed for that policy holder.

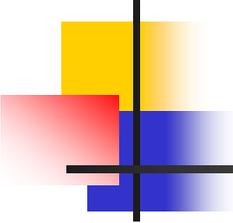


# Calendário

---

- 17/05 Aula
- 24/05 Aula
- 31/05 Aula
- 07/06 Aula
- 14/06 **Viagem**
- 21/06 Aula
- 28/06 Entrega de trabalho e Prova

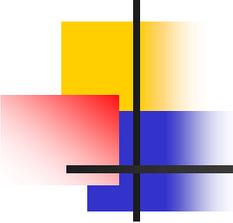
Eventuais alterações serão comunicadas no site do curso



# Conteúdos das aulas

---

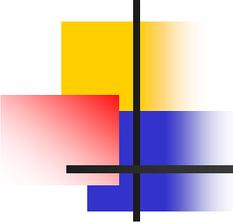
- Aula 1
  - Apresentação
  - Aprendizado de máquina
  - Algoritmos de aprendizado de máquina
- Aula 2
  - Algoritmos de indução de árvores de decisão
  - Algoritmos probabilísticos



# Conteúdos das aulas

---

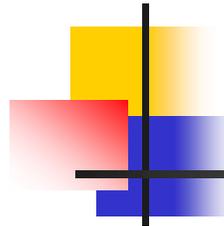
- Aula 3
  - Redes neurais
  - SVMs
  - Deep learning
- Aula 4
  - Agrupamento de dados
  - Ensembles
- Aula 5: prova



# Bibliografia

---

- Faceli, K., Lorena, A., Gama, J. e de Carvalho, A., Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina, LTC, 2011
- Flach, P., Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012
- Alpaydin, E., Machine Learning: The New AI (MIT Press Essential Knowledge series), 2016
- Moreira, J., de Carvalho, A. e Horvath, T., General Introduction to Data Analytics, Wiley, 2018
- Alpaydin, E. Introduction to Machine Learning, MIT Press, 2004
- Witten, I., Frank, E., Hall, M., Pal, C., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016
- Mitchell, T. M. Machine Learning. McGraw-Hill, 1997.



# Perguntas

---

