

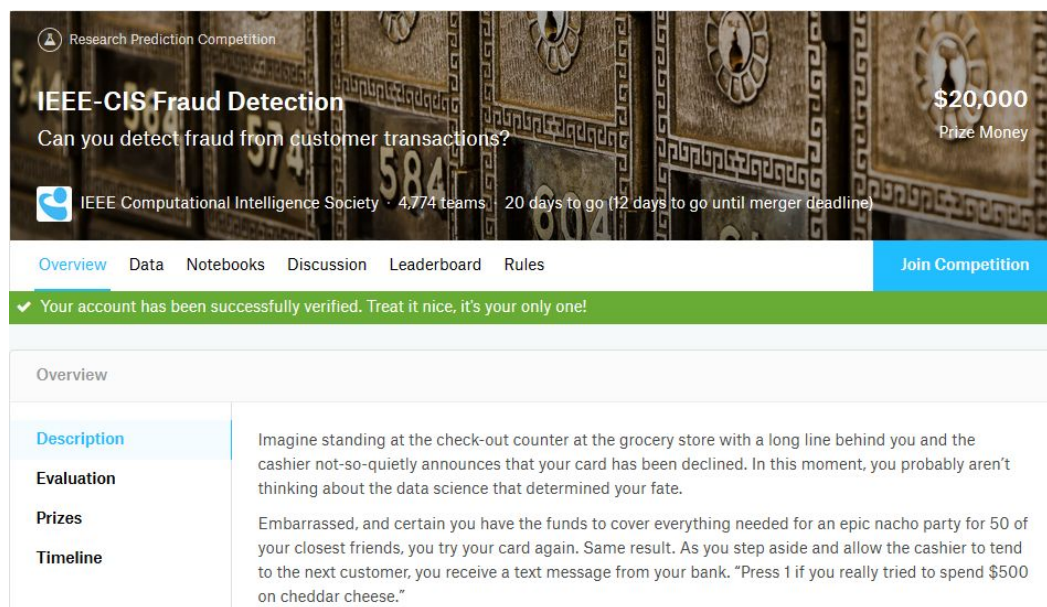
**Bacharelado em Ciência da Computação - UFRPE/UAG**  
**Reconhecimento de Padrões 2019.2**  
**Prof. Dr. Luis Filipe**

**Projeto de Disciplina 01**

**INFORMAÇÕES INICIAIS**

O projeto proposto consiste na participação em uma competição internacional de Aprendizagem de Máquina. A competição *IEEE-CIS Fraud Detection* está hospedada no portal de competições [www.kaggle.com](http://www.kaggle.com). O prazo para primeira submissão na competição, bem como para assinar o *External Data Disclosure*, encerra-se em **24/09/2019**. O prazo final para submissão de soluções encerra-se em **01/10/2019**.

Entrega do relatório de projeto para a disciplina: **07/10/2019**.



**Figura 1** - Página da competição *IEEE-CIS Fraud Detection* no sistema kaggle.com.

Durante toda a competição, soluções parciais dos competidores podem ser enviadas para avaliação em uma base de testes secreta. Usando as taxas de acerto de cada algoritmo, forma-se um *Leaderboard* como ilustrado na **Figura 2**. Ao final do período de submissões, os três primeiros colocados do *Leaderboard* serão premiados com **US\$ 10.000,00**, **US\$ 7.000,00**, e **US\$ 3.000,00**, respectivamente.

Public Leaderboard

Private Leaderboard

This leaderboard is calculated with approximately 20% of the test data.

The final results will be based on the other 80%, so the final standings may be different.

[Raw Data](#)

[Refresh](#)

In the money

Gold

Silver

Bronze

#	Team Name	Notebook	Team Members	Score 🏆	Entries	Last
1	alijs			0.9642	63	2d
2	Konstantin Yakovlev			0.9640	166	5h
3	ML Keksika			0.9638	115	5h
4	YuyaYamamoto			0.9628	71	3h
5	conundrum.ai			0.9623	48	1h
6	Fatih Ozturk			0.9623	89	1h
7	Jie Wu			0.9611	135	15h
8	[ka.kr] boradori			0.9611	211	6h
9	クソザコやねん			0.9602	100	13h
10	Tomoyo			0.9598	88	20h

Figura 2 - Leaderboard da competição em 11/09/2019.

## TEMA DO PROJETO

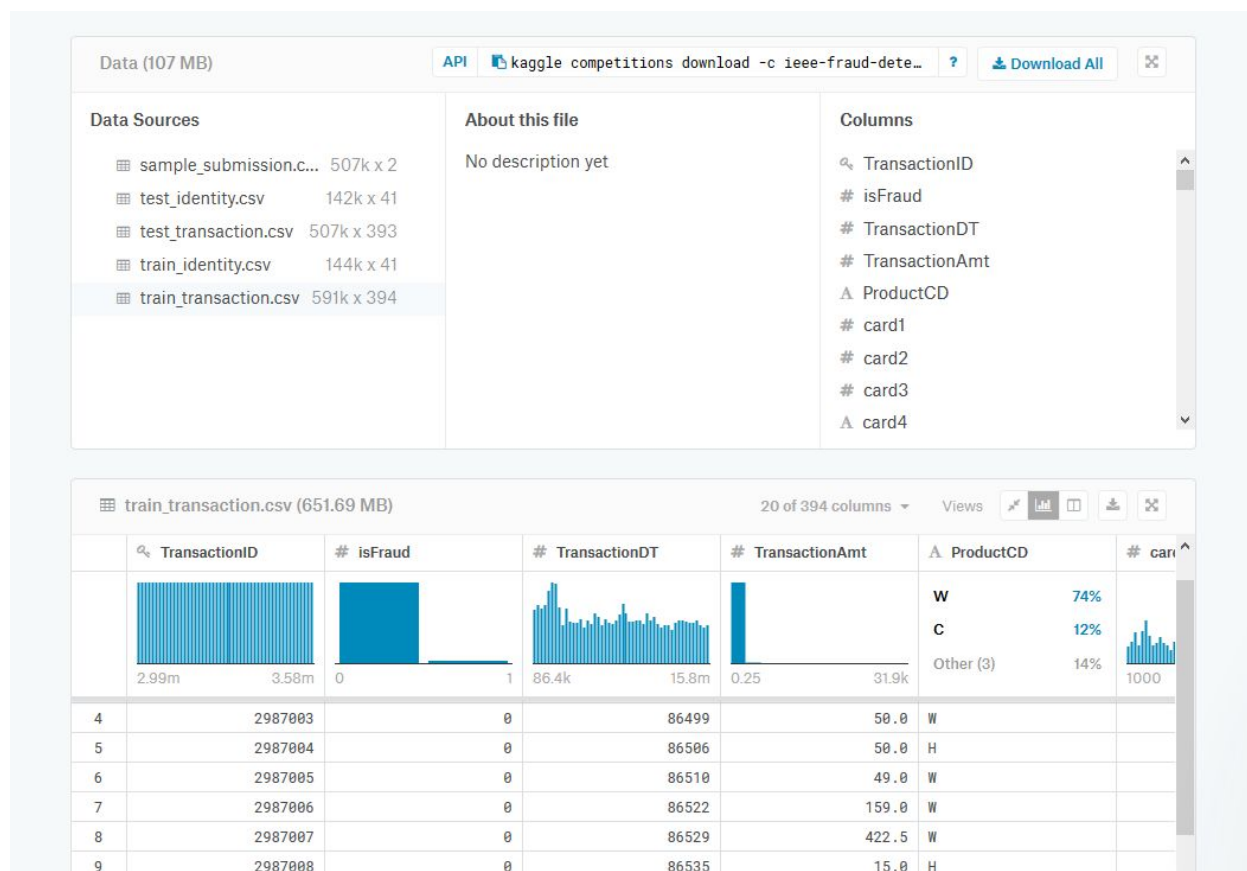
Essa competição é patrocinada pela Vesta Corporation com o objetivo de alcançar o melhor algoritmo para detecção de fraudes em compras com cartões de crédito.

## BASE DE DADOS

A **Figura 3** apresenta uma visualização dos dados disponibilizados para competição.

A base de treino disponível é formada pelos arquivos *train\_transactions.csv* e *train\_identity.csv* (cada padrão é unido pela chave TransactionID). A informação de classe está no atributo isFraud. Informações sobre o significado de cada característica da base pode ser encontrada aqui: <https://www.kaggle.com/c/ieee-fraud-detection/discussion/101203>.

Os arquivos *test\_transactions.csv* e *test\_identity.csv* contém os padrões dos conjunto de teste sem a informação de classe isFraud. Por fim, o arquivo *sample\_submission.csv* ilustra como os dados devem ser organizados para correção automática visando o *Leaderboard*.



**Figura 3** - Visualização dos dados disponibilizados.

## CRITÉRIOS PARA PROJETO DA DISCIPLINA

O trabalho realizado deve atender os seguintes critérios para ser considerado como solução válida para projeto da disciplina.

- Realização de um estudo preliminar e exploratório do poder de classificação de cada característica ou conjunto de características. Por exemplo, usando a métrica de verossimilhança definida por Bayes;
- Definição de três hipóteses de arquitetura com base do estudo preliminar anterior. Em que cada arquitetura deve-se combinar ao menos dois dos seguintes classificadores: Naive Bayes, Árvore de Decisão, Rede Neural;
- Avaliação das hipóteses através de validação cruzada na base de treinamento e testes estatísticos;
- Uso de quaisquer recursos das bibliotecas sickit-learn e pandas.

## RELATÓRIO

O relatório deverá ser escrito utilizando o template *“Example of Elsevier article template with dummy text”* disponível no sistema overleaf. O documento deve conter:

- Título, Resumo, Introdução e Conclusão coerentes com um **trabalho de nível superior**;
- Descrição do método proposto contendo **TODAS** as justificativas para as decisões de projetos tomadas;
- Apresentação e explicação dos **trechos de códigos** mais significativos para a construção da solução;
- Detalhamento dos **resultados intermediários** obtidos ao longo da construção da solução.