

## Estudo de Caso - Checkpoint 4SI

Sua empresa de consultoria especializada em tecnologia foi contratada para fornecer suporte à **ABX Tecnologia**. Esta empresa atua no segmento de importação e revenda de produtos, atendendo a uma variedade de clientes que inclui desde pequenas lojas até grandes redes de supermercados e armazéns.

## Entendendo o problema:

- A maioria dos clientes da empresa atualmente solicita prazos de pagamento superiores a 30 dias após a entrega dos produtos. Para gerenciar essas solicitações, a empresa possui um time especializado em análise de crédito, responsável por determinar o limite máximo de crédito que cada cliente pode obter para compras a prazo.
- No entanto, atualmente não há uma separação clara dentro da equipe de analistas de crédito. Dentro do grupo, existem especialistas em microcrédito, focados em pequenas operações, e especialistas em concessão de crédito para grandes negócios. Isso indica que a análise de crédito não é uniforme e varia de acordo com o segmento do cliente.

Entendendo o problema: Análise externa de  
Risco e Crédito




## Definição do problema:

1. A empresa gostaria de saber se é possível traçar alguma estratégia comum para cada grupo de cliente, no entanto essa análise nunca foi feita e ninguém, até o momento, sabe ao certo qual seria a melhor forma de conseguir realizar esse tipo de análise. Hoje, no conjunto de dados, não existem informações claras relativas ao segmento do cliente ou tamanho, apenas dados relativos ao faturamento, ano de fundação, endividamento e etc.
  - Isso ajudaria muito os times de marketing e análise de crédito;



## Definição do problema:

2. Um problema já mapeado é o que fazer após a geração desses grupos quando chegar um novo cliente solicitando um novo crédito?
- Como seria possível indicar o grupo desse cliente sem ter que refazer todos os grupos?
- 

## Definição do problema:

3. A empresa deseja, entre outras coisas, criar um “robô automatizado” (na linguagem da própria empresa), que realize uma recomendação de crédito para que os analistas se baseiam nessa recomendação para conceder o limite máximo de crédito.
  - Vale salientar que por se tratar de valores monetários é importante que este modelo tenha um bom desempenho;

Definição do problema:

O que mais seria possível ser feito???

## O QUE É ESPERADO COMO ENTREGA?

Solução **implementada** utilizando Deep Analytics e modelos de Machine para os 3 problemas e outras soluções **diferenciadas** e **inovadores** (opcional) que possam ser aplicadas ao conjunto de dados proposto.

- Devem ser entregues em formato de Jupyter Notebook;

Apresentação em formato de **Pitch de 5 minutos** a ser realizada por cada um dos grupos no último dia de desafio.

- O que **não** é esperado para a apresentação: Apenas slides com métricas e resultados;
- O que é esperado para a apresentação: Métricas e resultados em um formato entendível para uma área de negócio e o que esses resultados podem representar para o negócio em si em termos práticos:



## DESCRIÇÃO DO CONJUNTO DE DADOS:

O conjunto de dados é formado por 8974 solicitações de crédito de diversos clientes, um mesmo cliente pode ter diversas solicitações de crédito com valores distintos.

## DESCRIÇÃO DO CONJUNTO DE DADOS:

Abaixo temos a descrição dos 38 campos que formam o conjunto de dados:

- numero\_solicitacao: Número da solicitação;
- razaoSocial: Razão Social Anonimizada;
- nomeFantasia: Nome Fantasia Anonimizado;
- cnpjSemTraco: CNPJ Anonimizado;
- maiorAtraso: Maior atraso de pagamento em dias;
- margemBrutaAcumulada: Margem bruta acumulada;
- percentualProtestos: Percentual protestos;
- primeiraCompra: Data da primeira compra na empresa;
- prazoMedioRecebimentoVendas: prazo médio do recebimento de vendas do cliente;
- titulosEmAberto: valor total de títulos em aberto;
- valorSolicitado: valor de crédito solicitado;

# DESCRIÇÃO DO CONJUNTO DE DADOS:

- status: Status da solicitação:
  - AprovadoAnalista: aprovado por um analista
  - AprovadoComite: aprovado por um comitê, normalmente para valores mais expressivos
  - AguardandoAprovacao: aguardando aprovação
  - DocumentacaoReprovada: documentação reprovada
  - EmAnaliseDocumentacao: em Análise da documentação
  - ReprovadoAnalista: reprovado por um analista
  - ReprovadoComite: reprovado por um comitê
- definicaoRisco: categorização de risco;
- percentualRisco: onde 0 é baixo e 1 é alto;
- diferencaPercentualRisco:  $1 - \text{percentualRisco}$ ;
- dashboardCorrelacao: uma correlação interna de risco, mas que os analistas não conseguiram explicar nas reuniões realizadas;
- **valorAprovado**: Valor que foi aprovado pelos analistas;
- dataAprovadoEmNivelAnalista: data em que a solicitação de crédito foi aprovada por um analista de crédito;
- dataAprovadoEmComite: data em que a solicitação de crédito foi aprovada por um comitê;
- periodoBalanco: Período do balanço informado na documentação da empresa;
- ativoCirculante: Ativo circulante informado na documentação da empresa;

## DESCRIÇÃO DO CONJUNTO DE DADOS:

- passivoCirculante: Passivo circulante informado na documentação da empresa;
- totalAtivo: Ativo informado na documentação da empresa;
- totalPatrimonioLiquido: Patrimonio líquido informado na documentação da empresa;
- endividamento: Endividamento informado na documentação da empresa;
- duplicatasARreceber: Duplicatas a receber informado na documentação da empresa;
- estoque: estoque Informado na documentação da empresa;
- faturamentoBruto: Faturamento bruto informado na documentação da empresa;
- margemBruta: Margem bruta informado na documentação da empresa;
- periodoDemonstrativoEmMeses: Período do demonstrativo informado na documentação da empresa;
- custos : Custos informado na documentação da empresa;
- limiteEmpresaAnaliseCredito: limite de crédito fornecido por uma empresa externa de análise de crédito;
- anoFundacao: Ano de fundação da empresa;
- intervaloFundacao: Categoria do ano de fundação;
- capitalSocial: Capital social informado na documentação da empresa;
- restricoes: Flag informando se existem restrições relacionadas ao cliente;
- empresa\_MeEppMei: Flag informando se o cliente é um pequeno negócio;
- scorePontualidade: score de pontualidade entre 0 e 1, onde 0 significa que o cliente é pontual;

## REGRAS E PREMISSAS:

- Grupos com até 5 estudantes;
- É permitido ferramentas ou API externas para realizar as tarefas de pré-processamento de dados e treinamento de modelos (KMINE ou Orange por exemplo);
- É sugerido que os notebooks sejam implementados em python;
- É permitida a utilização de outros algoritmos e técnicas não vistos em sala de aula para pré-processamento, seleção de características, clusterização, classificação, regressão e etc.;
- Algumas aulas serão destinadas apenas ao checkpoint, mas o desafio pode ser desenvolvido também fora do horário das aulas;
- As dúvidas devem ser sanadas apenas nas aulas do desafio;

## DICAS E TRUQUES:

- Faça um bom entendimento dos problemas:
  - O que é esperado em cada um deles?
  - O que é necessário e o que não é necessário para resolver cada um dos problemas?
- Faça um bom entendimento dos dados, verifique os valores contidos em cada coluna e o significado dos mesmos para o problema (nem sempre isso estará claro).
- eventualmente o Excel pode auxiliar nesse processo;
- Faça um bom pré-processamento dos dados:
  - Verifique dados nulos;
  - Normalização;
  - Seleção de características;

## DICAS E TRUQUES:

- Saiba avaliar os modelos de Machine Learning e Deep Learning:
  - Este conjunto de dados é um conjunto de dados real, logo não é esperado que as taxas de acerto e métricas sejam extremamente altas;
- Os exemplos feitos em sala de aula podem auxiliar no desenvolvimento da solução.
- Pensem o problema de forma prática, as vezes é necessário pensar o problema em pedaços;

## Calendário:

	4SIA Segunda-feira	4SIA Quarta-feira	4SIR Quinta-feira	LABORATÓRIO	CONTEÚDO
semana 1	05/08	07/08	08/08	intro e lab1	Introdução à Data Science e CRISP-DM Visualização de Dados e Estatística Descritiva
semana 2	12/08	14/08	15/08	lab 2	Machine Learning Supervisionado - Classificação
semana 3	19/08	21/08	22/08	lab 3	Machine Learning Supervisionado - Regressão
semana 4	26/08	28/08	29/08	lab 4 e lab 5	Dicas e truques
semana 5	02/09	04/09	05/09	lab 6 (novo)	Machine Learning Não Supervisionado
semana 6	09/09	11/09	12/09	Aula estúdio	Aula estúdio para o desenvolvimento do CP
semana 7	16/09	18/09	19/09	CP	Projeto Final e Apresentação