



FIAP





DESAFIO MACHINE E DEEP LEARNING



PARABÉNS!

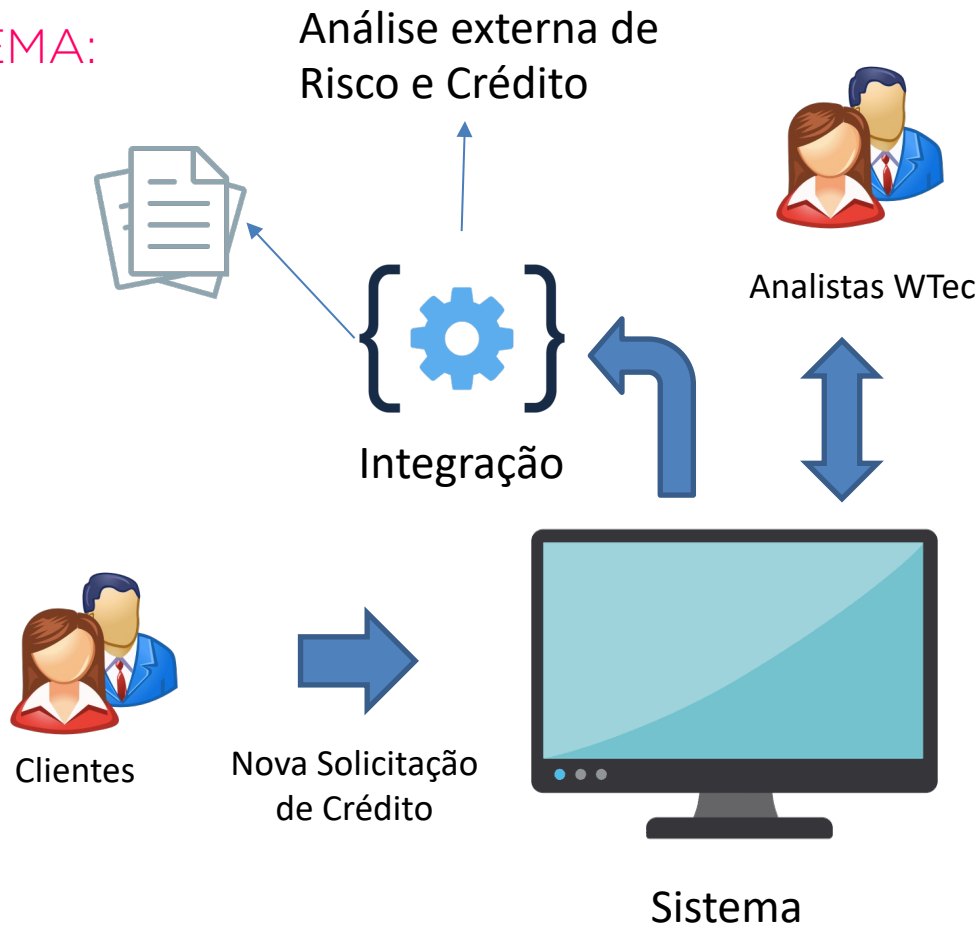
Seu equipe foi selecionada para resolver o principal problema da empresa **Wtec Suprimentos**, importadora de diversos suprimentos de informática, como cabos, cartuchos de impressoras, mouses, teclados, câmeras e etc.

Seu negócio consiste em importar os produtos e realizar a revenda para pequenas lojas, redes de armazéns e grandes redes de supermercado.

O PROBLEMA:

- Atualmente a imensa maioria dos clientes ao realizar uma compra solicita um prazo de pagamento superior a 30 dias após a entrega dos produtos.
- Hoje existe um time especializado em análise de crédito que determina o limite máximo de crédito que esse cliente pode comprar a prazo.
- Atualmente não existe uma separação bem definida dentro do time de analistas de crédito, mas sabe-se que existem dentro do time, pessoas que são especialistas em micro-crédito e outras que são especialistas em concessão de crédito para grandes negócios. Em outras palavras, a análise de crédito não é homogênea, varia de acordo com o segmento.

O PROBLEMA:



O 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;

O PROBLEMA:

2: 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 baseiem 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;

O PROBLEMA:

O que mais seria possível ser feito???

O QUE É ESPERADO COMO ENTREGA?

A solução **implementada** utilizando modelos de Machine e Deep Learning para os 2 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;

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.

Estará disponível no Portal Fiap.

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 Wtec;
- 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:

- Não é permitido ferramentas ou API externas para realizar as tarefas de pré-processamento de dados e treinamento de modelos (KMINE ou Orange por exemplo);
- Os notebooks devem ser 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.;
- É permitido eventualmente “transformar” alguma coluna do conjunto de dados como parte do pré-processamento, mas não é permitido nenhum tipo de enriquecimento dos dados;

REGRAS E PREMISSAS:

- O desafio pode ser desenvolvido também fora do horário das aulas alocadas para o mesmo;
- 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). A ferramenta Dbeaver e o Excel podem 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;
- Em caso de dúvidas procurem o professor 😊



OBRIGADO

FIAP

Copyright © 2022 | Professor Felipe Gustavo Silva Teodoro

Todos os direitos reservados. A reprodução ou divulgação total ou parcial deste documento é expressamente proibida sem consentimento formal, por escrito, do professor(a)/autor(a).





FIAP

