

Predição de Pagamentos Atrasados Através de Algoritmos Baseados em Árvore de Decisão

Arthur Flor de Sousa Neto
Bruno César de Andrade
Glauber Nascimento de Oliveira
José Fernando Guilhermino da Silva

13 de maio de 2021

Conteúdo

1. Introdução
 2. Conjunto de Dados
 3. Modelagem
 4. Experimento
 5. Resultados e Discussões
 6. Conclusão
-

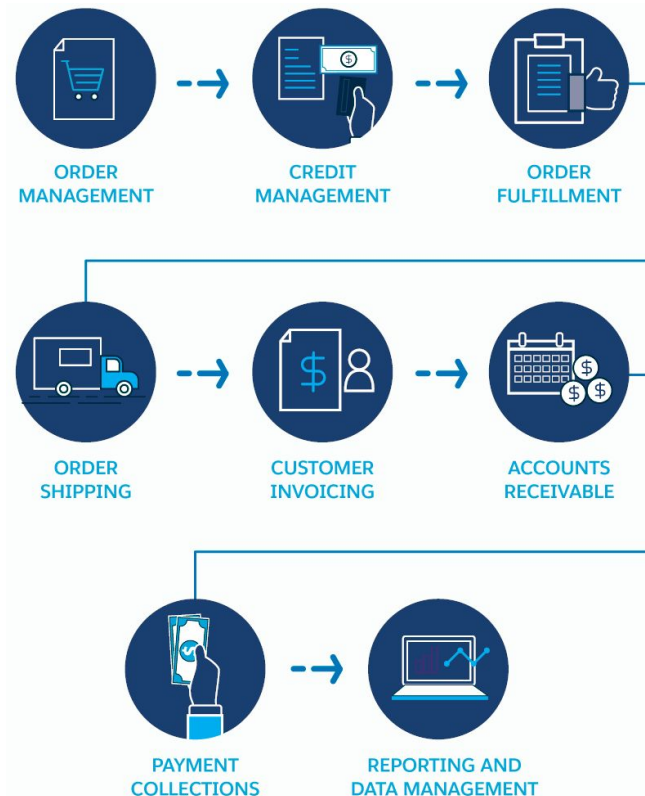
Introdução

Invoice-to-Cash

Atividades do setor financeiro dentro do processo Order-to-Cash.

Tem como principais funções:

- Lidar com a priorização de contas;
- Resolução de disputas;
- Contato com o cliente;
- Chamadas de cobrança.

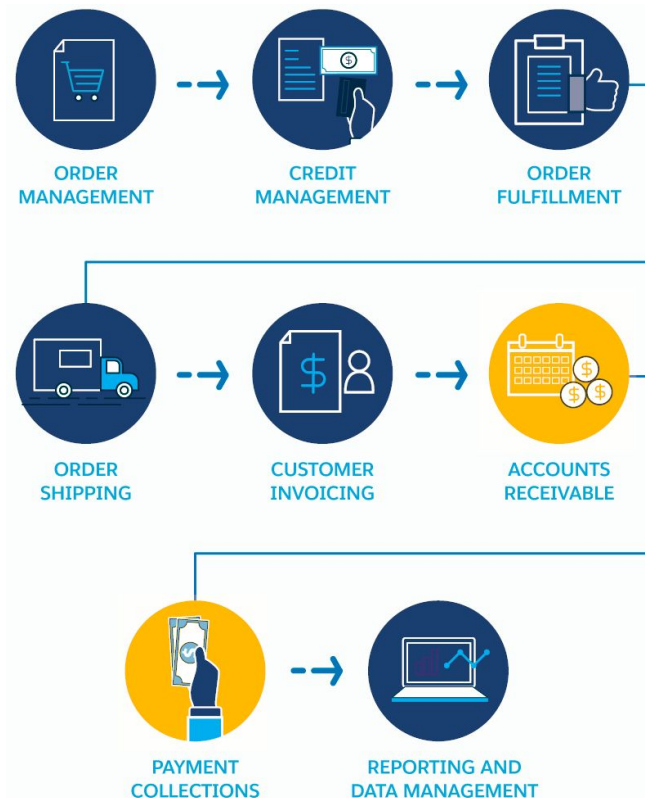


Fonte: <https://www.salesforce.com/products/cpq/resources/what-to-know-about-order-to-cash-process>

Introdução

Apesar da importância para qualquer negócio, a etapa de cobrança é geralmente processada manualmente, desse modo:

- **Todos** os clientes são contatados em intervalos fixos, embora alguns sempre tenham pago em dia;
- Quanto **mais tarde** um cliente é contatado, **menor a probabilidade** das faturas serem pagas a tempo.

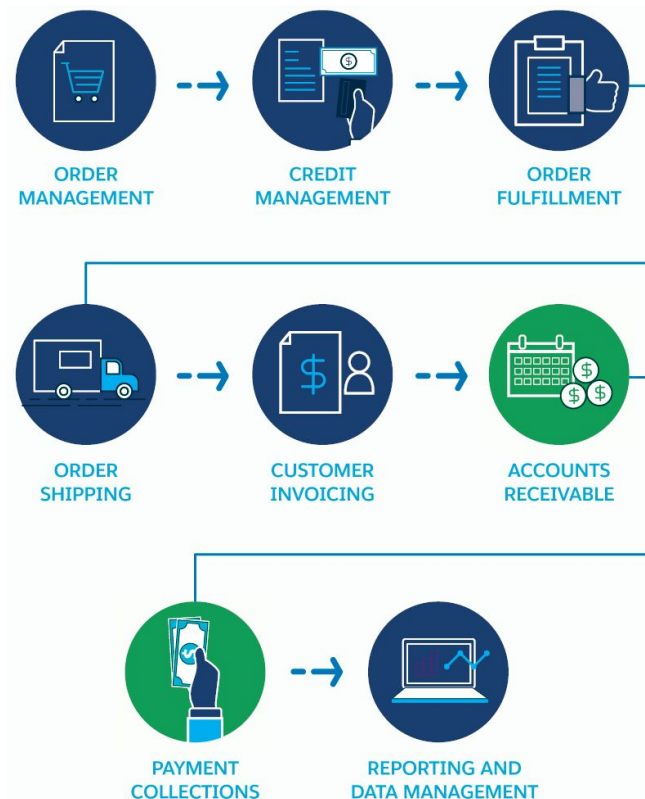


Fonte: <https://www.salesforce.com/products/cpq/resources/what-to-know-about-order-to-cash-process>

Introdução

Este trabalho propõe a exploração de técnicas de mineração de dados com machine learning, visando otimizar o processo de coleta de contas a receber.

Para este cenário, são analisados modelos baseados em árvore de decisão, no qual oferecem informações relevantes no auxílio do processo.



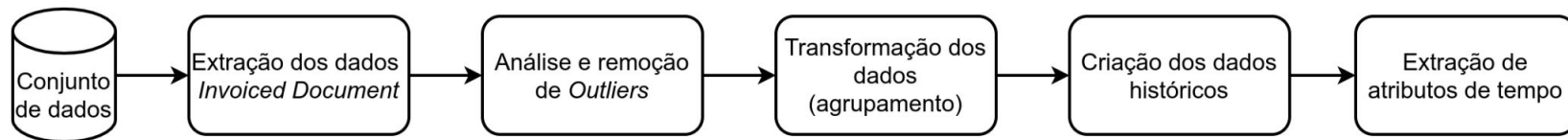
Fonte: <https://www.salesforce.com/products/cpq/resources/what-to-know-about-order-to-cash-process>

Conjunto de Dados

O conjunto de dados utilizado foi fornecido por uma empresa privada, contendo registros de todo o processo financeiro interno, no período de janeiro de 2018 até fevereiro de 2021.

Pré-processamento realizado:

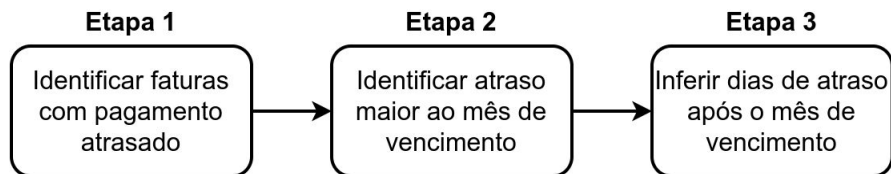
- Invoiced Document;
- Outliers;
- Agrupamento de faturas de mesmo cliente e no mesmo dia;
- Atributos de histórico por cliente;
- Atributos de tempo referente a fatura.



Modelagem

Para alcançar o objetivo do trabalho, o problema foi modelado em três etapas:

1. Identificar as faturas que serão pagas no prazo ou com atraso;
2. Identificar se as faturas com atraso, serão pagas ainda no mês de vencimento ou não;
3. Inferir o número de dias de atraso, para as faturas que serão atrasadas além do mês de vencimento.

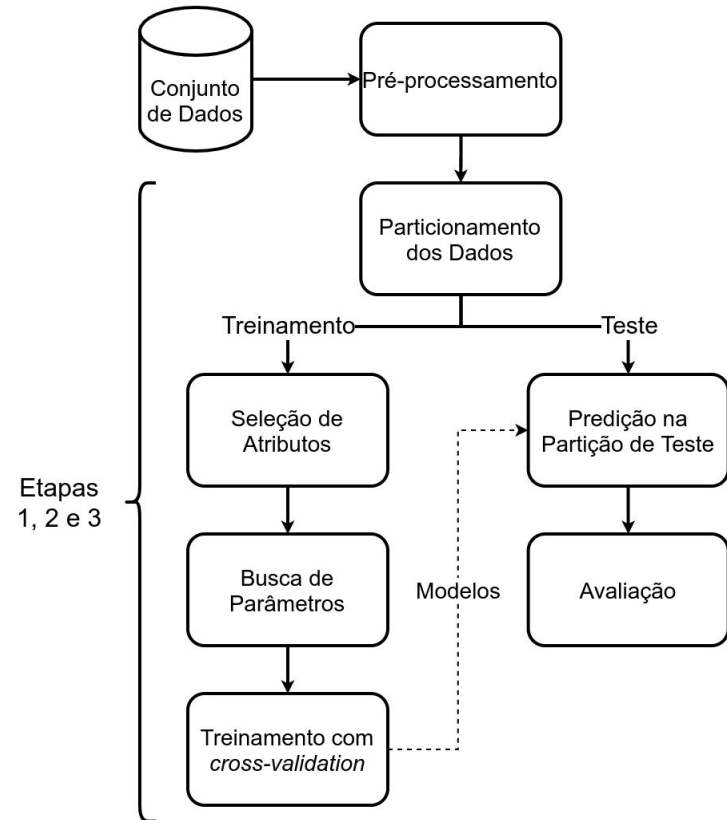


Experimento

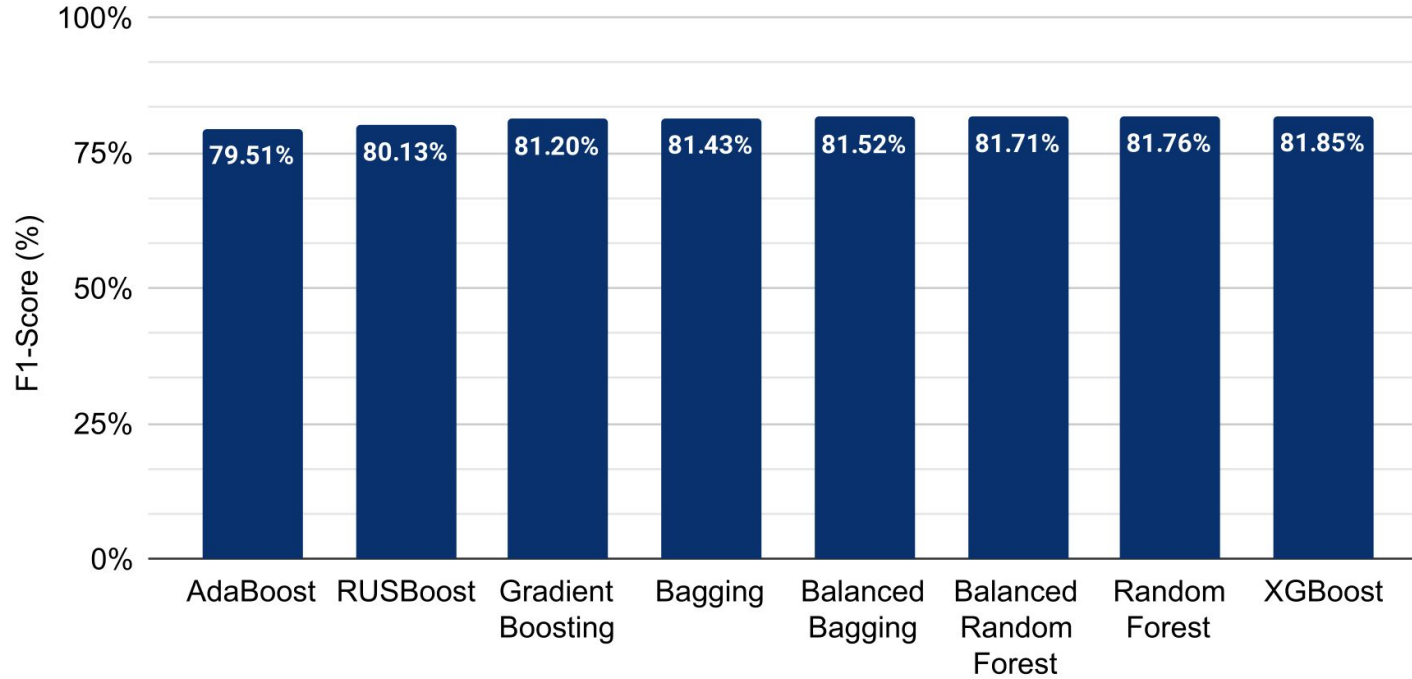
O experimento foi realizado através de oito algoritmos de árvore de decisão, variando entre três abordagens de construção (bagging, random e boost).

Para a execução e validação dos modelos, o conjunto de dados foi particionado em dois:

- Treino: faturas com data de vencimento anterior a fevereiro de 2021;
- Teste: faturas com data de vencimento em fevereiro de 2021.



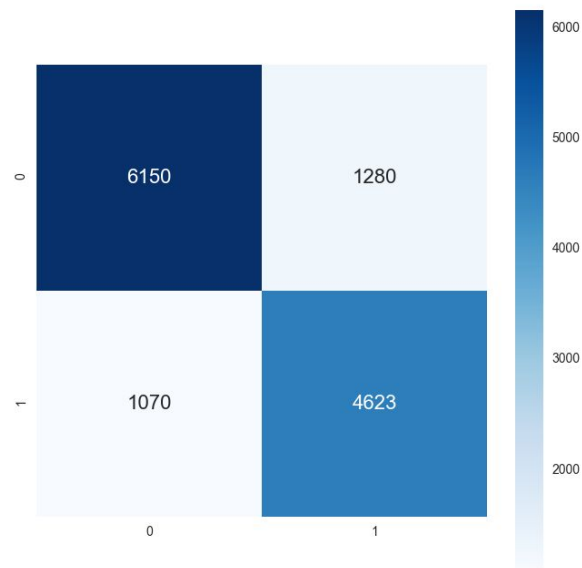
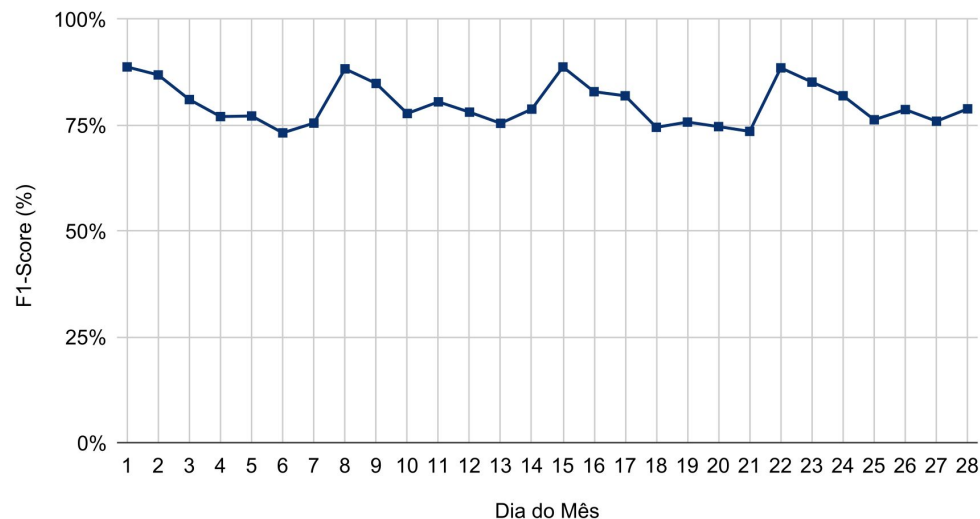
Resultados (Etapa 1)



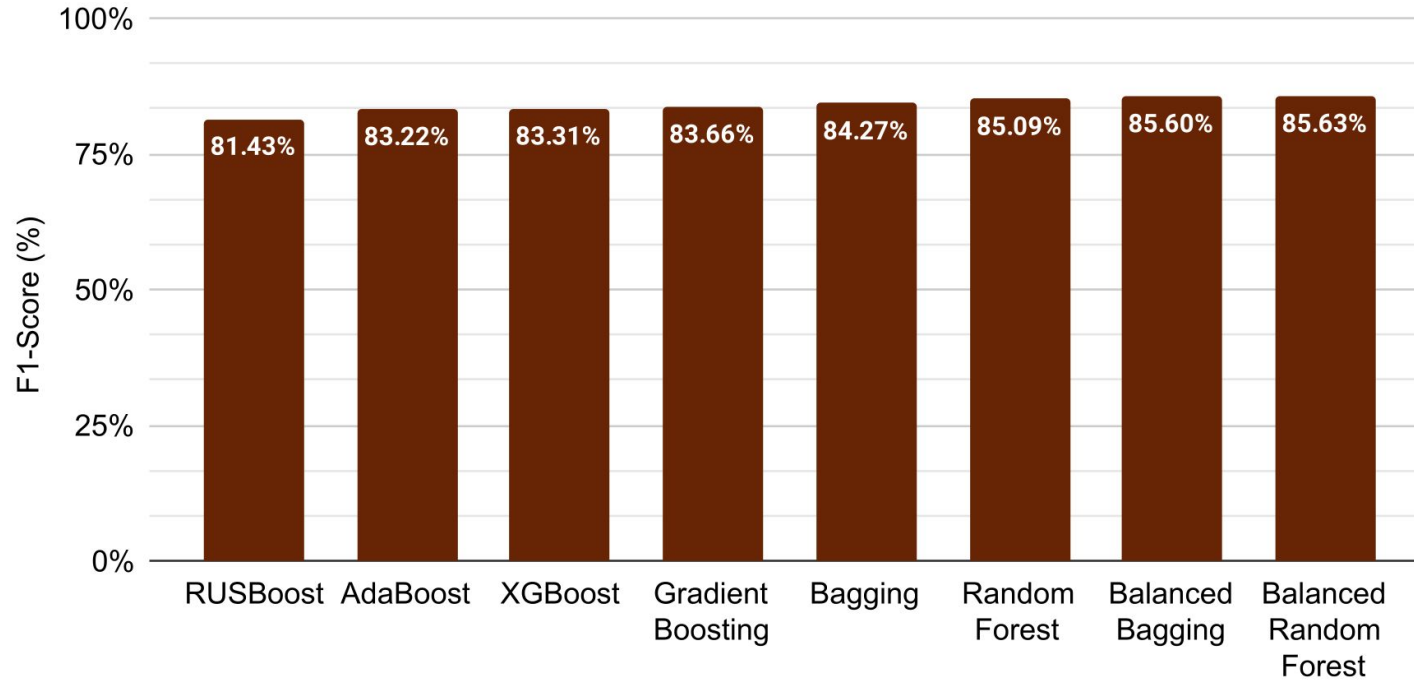
Resultados (Etapa 1)

Melhor resultado alcançado com XGBoost:

- **82,09%** dos dados foram reconhecidos corretamente;
- **81,20%** das faturas atrasadas foram reconhecidas.



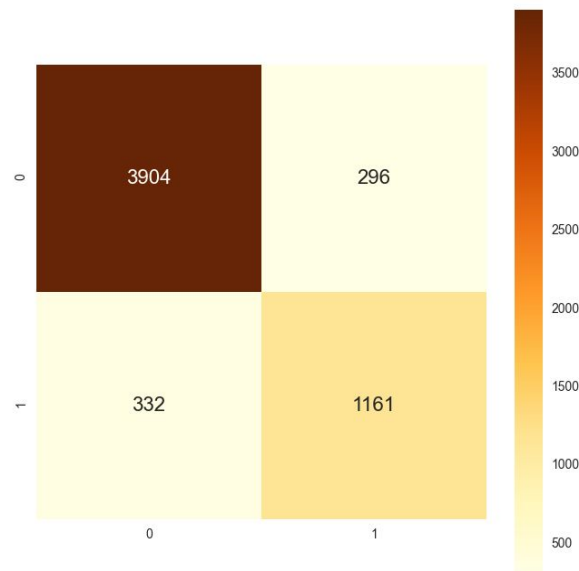
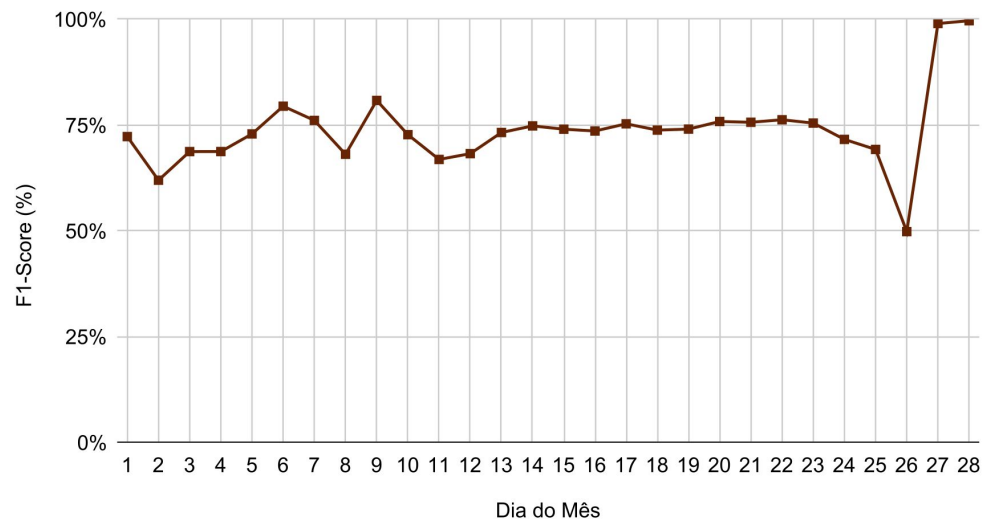
Resultados (Etapa 2)



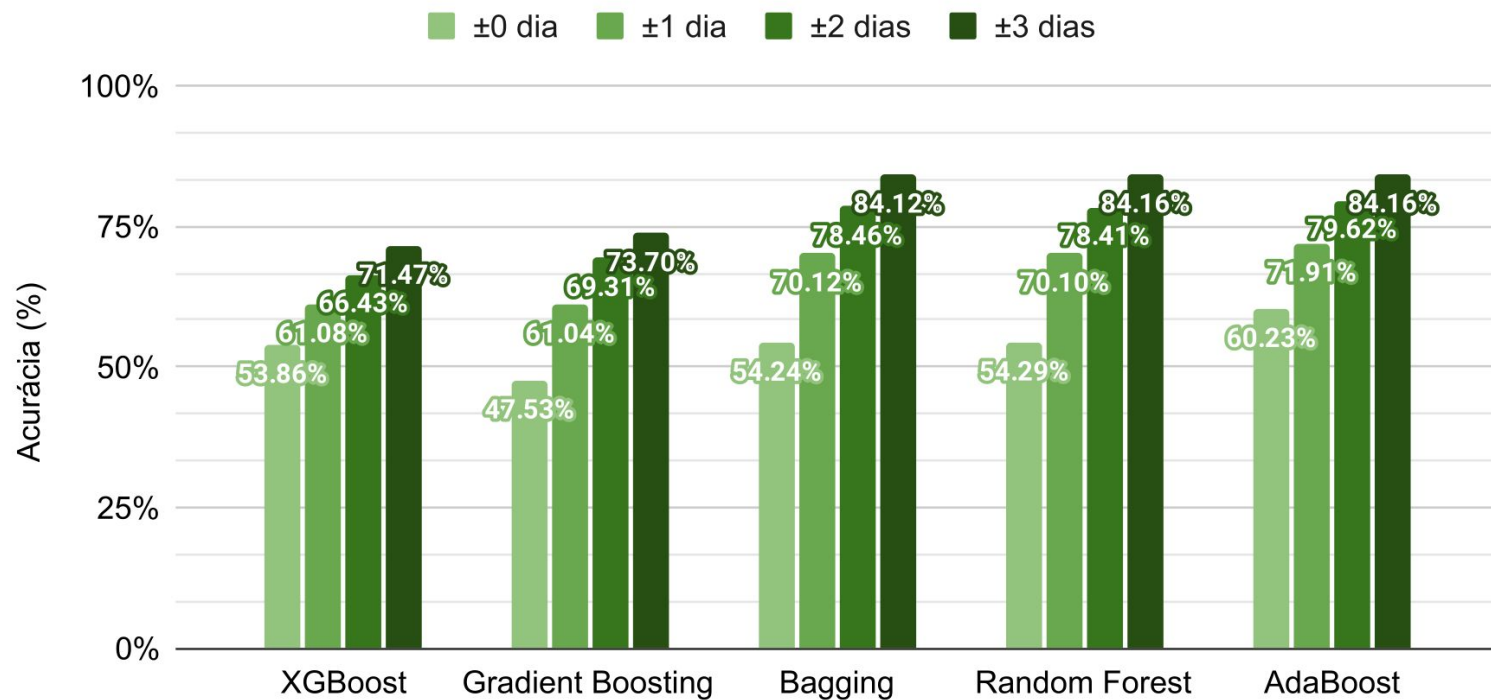
Resultados (Etapa 2)

Melhor resultado alcançado com Balanced Random Forest:

- **88,96%** dos dados foram reconhecidos corretamente;
- **77,76%** das faturas atrasadas além do mês de vencimento foram reconhecidas.



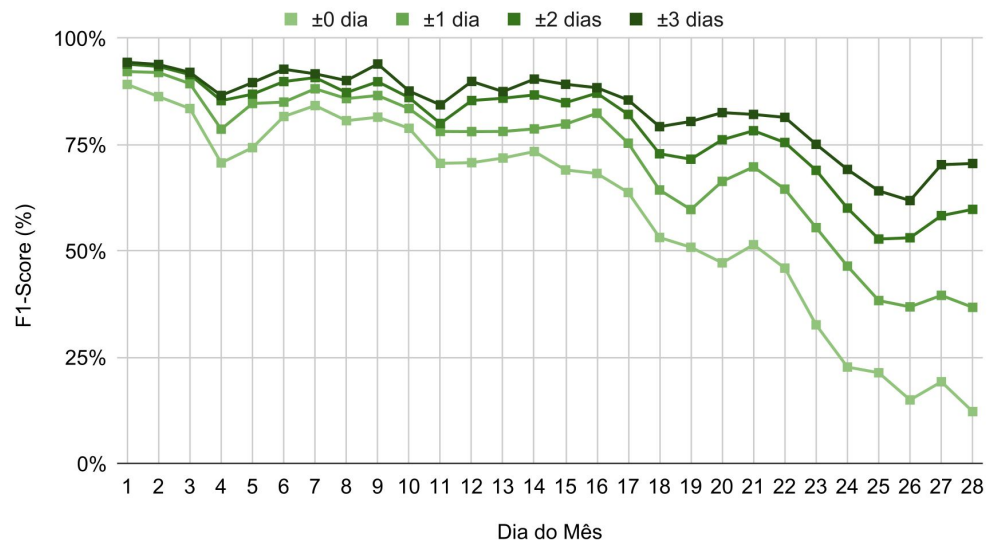
Resultados (Etapa 3)



Resultados (Etapa 3)

Melhor resultado alcançado com AdaBoost:

- 60,23% de acurácia sem variação de dia (predição correta);
- 84,16% de acurácia com variação de ± 3 dias.



Conclusão

- O trabalho proposto possibilita o auxílio no processo de coleta de contas a receber, oferecendo até 3 etapas de detalhamento para a tomada de decisão;
- Foi obtido uma precisão média de 80% no decorrer de todo o processo, através dos resultados obtidos dos melhores modelos;
- Como trabalhos futuros:
 - Modelar a etapa 3 para o contexto de intervalos de dias (buckets);
 - Ajustar o intervalo dos atributos históricos (atualmente em 7-14 dias);
 - Utilizar outras técnicas de aprendizagem de máquina.



Obrigado!