

Herval Deep Mailing para Pagamentos

Objetivo

Desenvolver um metodo para que se possa apartir das caracteristicas do devedor, determinar a probabilidade de pagamento do acordo desse devedor.

Desenvolvimento e metodo

1. Foi gerado um excel com varias caracteristicas de pagantes e nao pagantes
2. Esse excel foi carregado e tratado usando pandas, gerando um dataframe em torno de 9000 casos
3. Nesse primeiro momento foi utilizado casos com ate 1000 dias de atraso e com valor de divida de ate 1000 reais
4. Esse dataframe foi normalizado (Valores sendo projetados de forma proporcional para valores entre 0.0 e 1.0), e variaveis discretas como Escolaridade e Classe social foram normalizadas para 0 ou 1 em cada categoria.
5. Apartir desse dataframe que agora tem em torno de 180 colunas com valores entre 0.0 e 1.0, iniciamos a execucao do algoritmo de reconhecimento de padroes
6. O Algoritmo escolhido foi o Gradient Boosting atraves da biblioteca XGBoost rodando num notebok jupyter com python3
7. Testamos varios hiperparametros e os resultados estao abaixo:

Resultados

Dados relevantes para o pagamento:

Criamos diversas arvores de decisao, mas, basicamente as variaveis relevantes sao as seguintes

```
[(6431, 'NORM_CLIENTE_VALOR_DIVIDA'),  
(5679, 'NORM_CONTRATO_ATRASO'),  
(4872, 'NORM_RENDA_PRESUMIDA'),  
(567, 'NORM_CLASSE_SOCIAL_C1'),  
(560, 'NORM_ESCOLARIDADE_ENSINO_SUPERIOR'),  
(494, 'NORM_ESCOLARIDADE_ENSINO_MEDIO/TECNICO'),  
(371, 'NORM_CLASSE_SOCIAL_D'),  
(230, 'NORM_CLASSE_SOCIAL_C2'),  
(189, 'NORM_CLASSE_SOCIAL_B2'),  
(126, 'NORM_ESCOLARIDADE_ENSINO_MEDIO'),  
(106, 'NORM_ESCOLARIDADE_ENSINO_FUNDAMENTAL'),  
(75, 'NORM_CLASSE_SOCIAL_B1'),  
(40, 'NORM_CLASSE_SOCIAL_A2'),  
(21, 'NORM_CLASSE_SOCIAL_A1')]
```

No modelo acima, foi forçado o overfitting exatamente para que se possa investigar as variaveis relevantes, os hiperparametros usados foram:

```
Number of Trees:20,  
eta:0.3,  
depth:20  
Precisao: 99.10741029173906%
```

Atualmente estamos testando o modelo e conseguimos o seguinte resultado:

True Positive:208
True Negative:649
False Positive:839
False Negative:43

Mas, o modelo ainda tem muitos valores a serem tratados e necessita de ajustes antes de entrar em producao