

Modelo Preditivo de Nota de Satisfação Consumidor.gov



16/09/2022



MBA Analytics – Data Science



Nome do Aluno:

Andréia da Conceição Matias

Coordenadores:

Profª Drª Alessandra de Ávila Montini

Profª Dr. Adolpho Walter Pimazoni Canton

Agenda

- 1. Objetivo do Trabalho
- 2. Contextualização do Problema
- 3. Base de Dados
 - i. Filtros
 - ii. Principais variáveis
 - iii. Processo de redução de variáveis
- 4. Análise exploratória de dados
- 5. Modelagem estatística tradicional
 - i. Regressão logística
 - ii. Árvore de decisão
 - iii. Desempenho dos modelos
- 6. Inteligência Artificial
 - i. Random Forest
 - ii. Gradient boosting
 - iii. Extreme gradient boosting
 - iv. lightGBM

26/04/2014 14h35 - Atualizado em 26/04/2014 14h35

Pesquisa diz que consumidor prefere bom atendimento à preço baixo

Em Itapetininga (SP) e Avaré (SP) empresas investem em treinamento. Ser bem atendido é o mais importante para 61% dos consumidores.

Fonte: <https://g1.globo.com/sao-paulo/itapetininga-regiao/noticia/2014/04/pesquisa-diz-que-consumidor-prefere-bom-atendimento-preco-baixo.html>



Busca...

Agenda Liberal

Carreira

Dinheiro

Economia

ESG

Money Talks

NEGÓCIOS

Exame: gestão interna é a principal dor no atendimento

Da Redação | 28 de julho de 2022

<https://www.moneyreport.com.br/negocios/exame-gestao-interna-e-a-principal-dor-no-atendimento/>



Missão

Ajudar nossos clientes a melhorarem a satisfação de seus clientes com a utilização da análise de dados e da inteligência artificial.



Cliente feliz

Parceiros Chave

desenvolvedores

analistas de dados

cientistas de dados

Atividades Chave

desenvolvimento de solução para identificação de clientes propensos à insatisfação.

Recurso Chave

Estatística

programação

conhecimento de negócio

Proposta de Valor

Ajudar empresas a melhorar a satisfação de seus clientes com a utilização de inteligência artificial

Relação com o cliente

Consultoria para implantação e manutenção da ferramenta.

Canais

Integração com aplicativos de atendimento e de tratamento de ocorrências (SAC)

Segmentos de Mercado

Empresas que utilizem pesquisa de satisfação

Estrutura de Custos

pagamento de salários

aluguel

notebooks

Fontes de Renda

consultoria à distância e presencial

9. Matriz SWOT

Forças	Fraquezas
<ul style="list-style-type: none">• Possibilidade de identificar clientes cujo perfil seja mais propenso à insatisfação indicando como ponto de atenção na demanda, o que possibilitará ao analista envidar mais esforços para solução da demanda ou mesmo priorizá-la. É também possível criar ações de encantamento.	<ul style="list-style-type: none">• Há cliente que abrem diversas demandas, o que pode enviesar o resultado. Identificada a situação deverá ser averiguado a tratamento a ser dado.
Oportunidades	Ameaças
<ul style="list-style-type: none">• Priorizando a demanda do cliente mais propenso a avaliar o atendimento como insatisfatório, é possível dedicar mais tempo a ela para buscar alternativas e apresentar a solução de forma tempestiva, o que pode ajudar na melhora de sua satisfação, o que se refletiria na melhora do NPS.	<ul style="list-style-type: none">• Utilização de ferramenta semelhante por concorrentes.



- Prezamos pela privacidade de nossos cliente e protegemos os dados a nós confiados por meio da anonimização;
- Utilizamos apenas dados essenciais ao bom desempenho das atividades e pelo tempo necessário.



1. Objetivo do Trabalho

O objetivo do trabalho é identificar o perfil de clientes de um banco cuja nota de satisfação seja 1 ou 2.

Para tanto serão utilizadas as notas atribuídas pelos reclamantes no site Consumidor.gov entre os anos de 2015 e 2019 e serão utilizadas as técnicas de regressão logística ou árvore de decisão para identificação desses clientes.

Dessa forma, o banco poderá traçar estratégias para prever os clientes propensos a atribuir notas mais baixas e poderá envidar esforços para tentar solucionar a demanda de modo a reverter insatisfação.



2. Contextualização do Problema



O Consumidor.gov é uma plataforma de informação, interação e compartilhamento de dados que visa uma solução alternativa de conflitos de consumo. Os consumidores registram suas demandas pelo site e interagem com a empresa por ele. Findado o prazo de solução, podem avaliar se a demanda foi resolvida e atribuir uma nota.

Essas demandas são tratadas pelo Serviço de Atendimento ao Consumidor de um banco que tem por objetivo tirar dúvidas ou mesmo reverter a insatisfação de seus clientes e um dos indicadores acompanhados pela unidade é a nota de satisfação.

Para atender os clientes ou usuários que registram reclamação no site de forma mais proativa, seria interessante identificar aqueles mais propensos a ficar insatisfeitos para dar mais atenção à demanda ou mesmo de buscar ações que possam contribuir na melhoria do indicador.

3. Bases de Dados



<https://dados.gov.br/dataset/reclamacoes-do-consumidor-gov-br1>

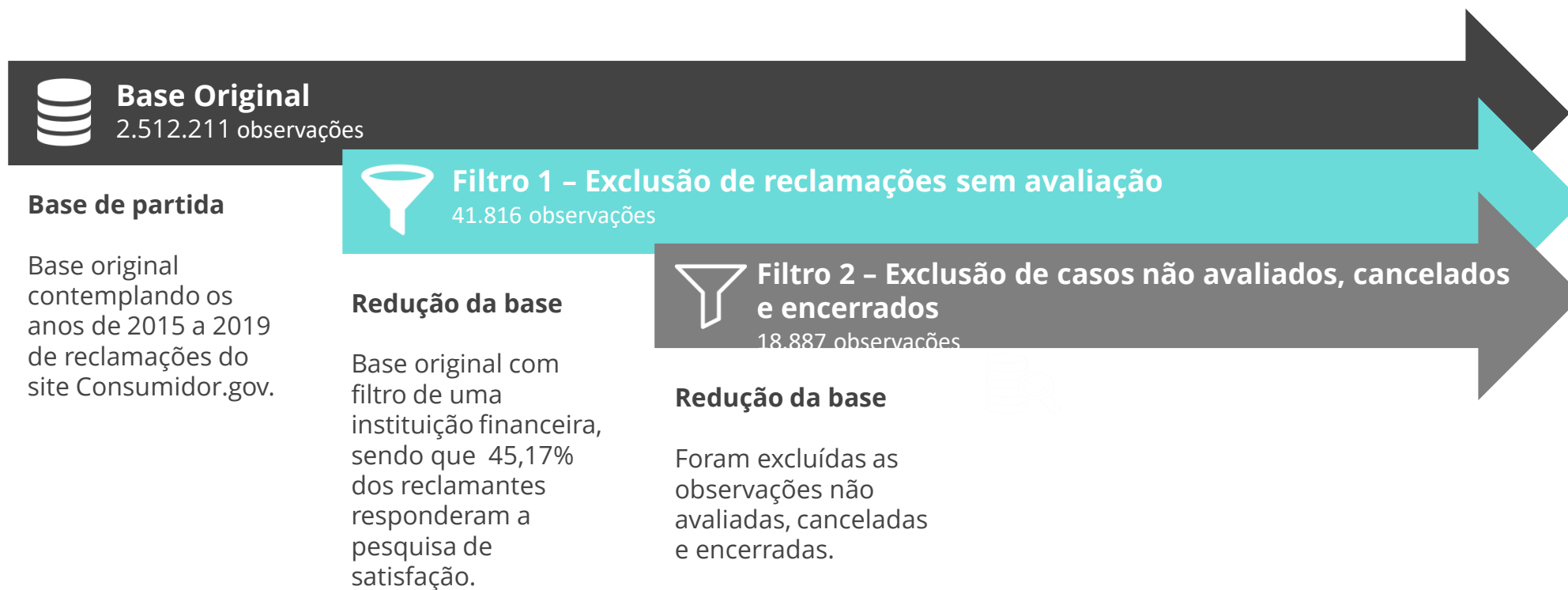


2015-2019



Base original: 2.512.211 (20 variáveis)
Base filtrada com casos da instituição financeira: 41.816 (17 variáveis)

3.i. Filtros



3.ii. Principais variáveis



Variáveis da reclamação

- assunto
- avaliação da reclamação
- como comprou/contratou
- data finalização
- grupo problema
- nota do consumidor
- Problema
- Procurou empresa
- Respondida
- Situação
- Tempo resposta



Variáveis demográficas

- cidade
- região
- faixa etária
- sexo
- UF



3.ii. Principais variáveis



Variável	Descrição	Exemplo de Dado
Assunto	Assunto objeto da reclamação	"Atendimento Bancário", "Telefonia Fixa", etc.
Avaliação Reclamação	Classificação atribuída pelo consumidor sobre o desfecho da reclamação	"Resolvida", "Não resolvida" e "Não Avaliada".
Cidade	Município do consumidor reclamante	"Brasília", "Manaus", "Recife", "São Paulo", etc.
Como Comprou Contratou	Descrição do meio utilizado para contratação/aquisição do produto ou serviço reclamado	"Loja física", "Telefone", "Internet", etc.
Faixa Etária	Faixa etária do consumidor	"até 20 anos", "entre 21 a 30 anos", etc.
Grupo Problema	Agrupamento do qual faz parte o problema classificado na reclamação	"Atendimento / SAC", "Contrato / Oferta", etc.
Nota do Consumidor	Número da nota de 1 a 5 atribuída pelo consumidor ao atendimento da empresa	"1", "2", "3", "4" e "5"
Problema	Descrição do problema objeto da reclamação	"Portabilidade não efetivada", etc.
Procurou Empresa	Sigla da resposta do consumidor à pergunta: "Procurou a empresa para solucionar o problema?"	"S" (sim) e "N" (não)

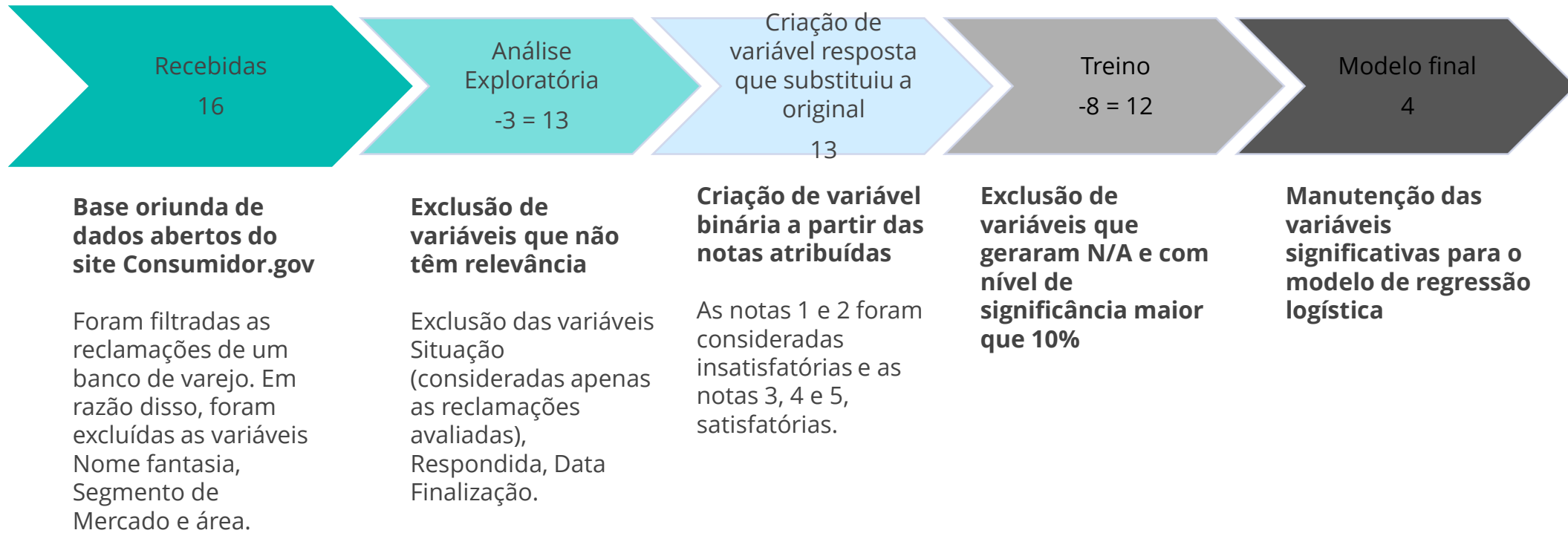


3.ii. Principais variáveis

Variável	Descrição	Exemplo de Dado
Região	Sigla da região geográfica do consumidor reclamante	"CO" (centro-oeste), "N" (norte), "SE" (sudeste), etc.
Sexo	Sigla do sexo do consumidor reclamante	"F" (feminino) e "M" (masculino)
Tempo Resposta	Número de dias para a resposta da reclamação, entre a Data de Resposta e a Data de Abertura, desconsiderado o tempo que a reclamação tenha ficado em análise pelo Gestor (se for o caso)	"0", "1", "2", "3", "4", "5", "6", "7", "8", "9" e "10".
UF	Sigla do estado do consumidor reclamante	"AM", "DF", "ES", "PE", "SP", etc.



3.iii. Processo de redução de variáveis



4. Análise Exploratória de Dados

Análise Univariada

17



Variáveis demográficas

- cidade
- região
- faixa etária
- sexo
- UF

Persona

- A maior parte dos consumidores que avaliou sua reclamação é do sexo masculino (62,46%);
- 65,32% têm entre 21 e 40 anos;
- 42,38% moram na região sudeste e 21,84% na região Nordeste;
- Os 5 estados com maior percentual de reclamações são São Paulo (24,78%), Paraná (9,59%), Minas Gerais (8,31%), Rio de Janeiro (6,77%), Bahia (6,29%);
- As reclamações estão mais pulverizadas pelas cidades, porém a cidade de São Paulo concentra o maior percentual (8,73%), seguida por Brasília (5,18%), Curitiba (4,09%), Rio de Janeiro (3,79%).

4. Análise Exploratória de Dados

Análise Univariada

18



Variáveis demográficas

- região
- faixa etária
- sexo

Sexo	Frequência
M	0,6246
F	0,3754

Região	Frequência
SE	0,4238
NE	0,2184
S	0,1626
CO	0,1439
N	0,0513

Faixa Etária	Frequência
entre 31 a 40 anos	0,3467
entre 21 a 30 anos	0,3065
entre 41 a 50 anos	0,1778
entre 51 a 60 anos	0,0967
entre 61 a 70 anos	0,0417
mais de 70 anos	0,0165
até 20 anos	0,0140



4. Análise Exploratória de Dados

Análise Univariada

19



Variáveis da reclamação

- assunto
- avaliação da reclamação
- como comprou/contratou
- data finalização
- grupo problema
- nota do consumidor
- Problema
- Procurou empresa
- Respondida
- Situação
- Tempo resposta

Grupo Problema	Frequência
Cobrança / Contestação	0,5439
Atendimento / SAC	0,1713
Vício de Qualidade	0,1414
Contrato / Oferta	0,1071
Informação	0,0299
Saúde e Segurança	0,0037
Entrega do Produto	0,0028

Reclamação

Mais de 50% das reclamações se referem ao grupo Cobrança/Contestação. O segundo grupo com mais reclamações é o de Atendimento/SAC, porém o percentual é de 17,13%.



4. Análise Exploratória de Dados

Análise Univariada

20



Variáveis da reclamação

- assunto
- avaliação da reclamação
- como comprou/contratou
- data finalização
- grupo problema
- nota do consumidor
- Problema
- Procurou empresa
- Respondida
- Situação
- Tempo resposta

Assunto	Frequência
Conta corrente / Salário / Poupança /Conta Aposentadoria	0,3065
Cartão de Crédito / Cartão de Débito / Cartão de Loja	0,1939
Atendimento Bancário	0,1348
Crédito Pessoal e Demais Empréstimos (exceto financiamento de imóveis e veículos)	0,1306
Crédito Consignado (Empréstimo descontado em folha de pagamento)	0,0673
Banco de Dados e Cadastros de Consumidores (SPC, Serasa, SCPC etc)	0,0537
Programas de Fidelidade / Benefícios (pontos, milhagem etc)	0,0287
Financiamento de Imóveis	0,0214
Serviço de pagamento online/ via celular/ maquininha	0,0128
Investimentos / Aplicações / Previdência Privada	0,0115
Financiamentos de Veículos / Leasing	0,0107
Crédito Consignado (para servidores públicos ou trabalhadores do setor privado)	0,0085
Seguro de Vida / Acidentes Pessoais / Desemprego	0,0075
Demais Seguros (exceto habitacional)	0,0038
Seguro Habitacional / Residencial	0,0021
Crédito Consignado / Cartão de Crédito Consignado / RMC (para beneficiários do INSS)	0,0017
Consórcios (exceto imóveis)	0,0017
Cartão de Programas Sociais	0,0013
Seguros de Veículos	0,0007
Consórcio de Imóveis	0,0005
Crédito Rural	0,0002
Seguro Garantia Estendida	0,0001
Seguro Viagem	0,0001

Reclamação

O assunto mais reclamado é “Conta corrente/Salário/Poupança/Conta Aposentadoria com 30,65% das reclamações; seguido por Cartão de Crédito, Cartão de Débito, Cartão de Loja com 19,39% das reclamações.

4. Análise Exploratória de Dados

Análise Univariada

21



Variáveis da reclamação

- assunto
- avaliação da reclamação
- como comprou/contratou
- data finalização
- grupo problema
- nota do consumidor
- Problema
- Procurou empresa
- Respondida
- Situação
- Tempo resposta

Como Comprou Contratou	Frequência
Loja física	0,4092
Não comprei / contratei	0,3401
Internet	0,1533
Telefone	0,0774
Domicílio	0,0101
Catálogo	0,0032
SMS / Mensagem de texto	0,0029
Stand, feiras e eventos	0,0022
Ganhei de presente	0,0015

Procurou Empresa	Frequência
S	0,812569492
N	0,187430508

Avaliação Reclamação	Frequência
Resolvida	0,5827
Não Resolvida	0,4173

Reclamação

- 40,92% dos reclamantes efetuaram a compra em loja física e 34,01% alegaram não ter comprado/contratado;
- Mais de 80% procuraram a empresa;
- 58,27% consideraram a reclamação resolvida.



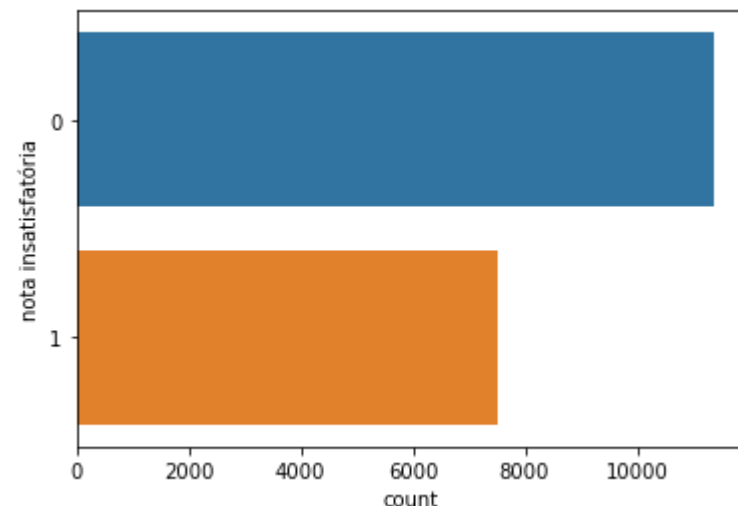
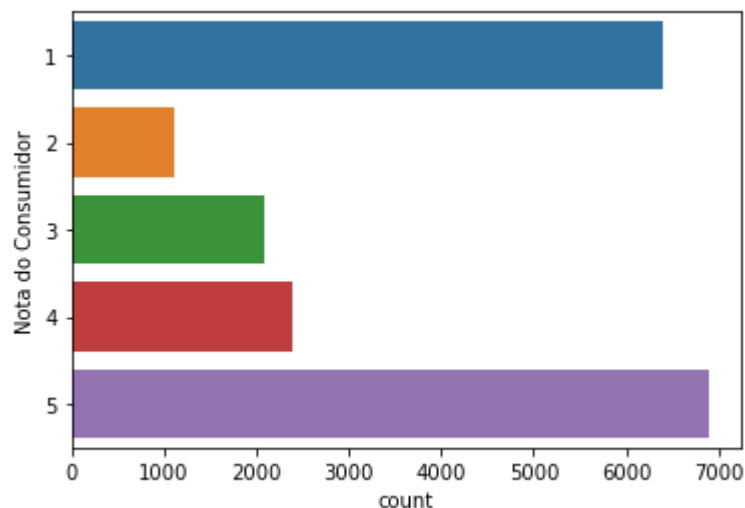
4. Análise Exploratória de Dados

Análise Univariada



Variáveis da reclamação

- assunto
- avaliação da reclamação
- como comprou/contratou
- data finalização
- grupo problema
- nota do consumidor
- Problema
- Procurou empresa
- Respondida
- Situação
- Tempo resposta



Reclamação

- O tempo de resposta médio foi de 6,94, com desvio-padrão de 2,50. A mediana era 7. O coeficiente de variação foi de 36,07%.
- A média de nota foi de 3,12, com desvio-padrão de 1,73 e a mediana era 3. O coeficiente de variação foi de 55,35%.
- Há maior quantidade de notas 1 e 5.
- Em razão do desbalanceamento, foi criada nova variável “nota insatisfatória, que considerou as notas 1 e 2 como insatisfatórias correspondendo a 39,70% da base.”

4. Análise Exploratória de Dados

Análise Bivariada

Região	Nota Satisfatória	Nota Insatisfatória	Total
CO	1679	1039	2718
	0,618	0,382	0,144
N	617	351	968
	0,637	0,363	0,051
NE	2598	1527	4125
	0,63	0,37	0,218
S	1878	1193	3071
	0,612	0,388	0,163
SE	4615	3390	8005
	0,577	0,423	0,424
Total	11387	7500	18887

Nota satisfatória/insatisfatória por região

- Após agrupamento das notas 1 e 2 como insatisfatórias e 3, 4 e 5 como satisfatórias, não nota-se diferença significativa entre as regiões.



4. Análise Exploratória de Dados

Análise Bivariada

Faixa etária	Nota satisfatória	Nota insatisfatória	Total
até 20 anos	204	61	265
	0,77	0,23	0,014
entre 21 a 30 anos	4073	1716	5789
	0,704	0,296	0,307
entre 31 a 40 anos	3919	2629	6548
	0,599	0,401	0,347
entre 41 a 50 anos	1749	1609	3358
	0,521	0,479	0,178
entre 51 a 60 anos	922	905	1827
	0,505	0,495	0,097
entre 61 a 70 anos	372	416	788
	0,472	0,528	0,042
mais de 70 anos	148	164	312
	0,474	0,526	0,017
Total	11387	7500	18887

Nota satisfatória/insatisfatória por Faixa Etária

- Após agrupamento das notas 1 e 2 como insatisfatórias e 3, 4 e 5 como satisfatórias, não nota-se diferença significativa entre as notas atribuídas conforme sexo, porém, parece haver maior propensão de atribuir nota insatisfatória conforme maior a faixa etária.



4. Análise Exploratória de Dados

Análise Bivariada

Tempo de resposta	Nota satisfatória	Nota insatisfatória	Total
0	72	18	90
	0,8	0,2	0,005
1	335	125	460
	0,728	0,272	0,024
2	599	223	822
	0,729	0,271	0,044
3	658	256	914
	0,72	0,28	0,048
4	763	363	1126
	0,678	0,322	0,06
5	946	520	1466
	0,645	0,355	0,078
6	1396	717	2113
	0,661	0,339	0,112
7	1714	1071	2785
	0,615	0,385	0,147
8	1758	1411	3169
	0,555	0,445	0,168
9	1432	1148	2580
	0,555	0,445	0,137
10	1689	1617	3306
	0,511	0,489	0,175
11	23	28	51
	0,451	0,549	0,003
12	0	3	3
	0	1	0
14	2	0	2
	1	0	0
Total	11387	7500	18887

Nota satisfatória/insatisfatória por Tempo de resposta

- Após agrupamento das notas 1 e 2 como insatisfatórias e 3, 4 e 5 como satisfatórias, parece haver maior propensão a atribuir nota insatisfatória quanto maior o tempo de resposta.



4. Análise Exploratória de Dados

Análise Bivariada

Grupo Problema	Nota satisfatória	Nota insatisfatória	Total
Atendimento / SAC	1962	1274	3236
	0,606	0,394	0,171
Cobrança / Contestação	6207	4065	10272
	0,604	0,396	0,544
Contrato / Oferta	1129	893	2022
	0,558	0,442	0,107
Entrega do Produto	26	27	53
	0,491	0,509	0,003
Informação	386	178	564
	0,684	0,316	0,03
Saúde e Segurança	33	36	69
	0,478	0,522	0,004
Vício de Qualidade	1644	1027	2671
	0,615	0,385	0,141
Total	11387	7500	18887

Nota satisfatória/insatisfatória por Grupo Problema

- Após agrupamento das notas 1 e 2 como insatisfatórias e 3, 4 e 5 como satisfatórias, não é possível notar predominância de atribuição de notas insatisfatórias por grupo/problema.
- As variáveis Assunto e Problema deixaram de ser analisadas.



4. Análise Exploratória de Dados

Análise Bivariada

Procurou Empresa	Nota satisfatória	Nota insatisfatória	Total
N	2537	1003	3540
	0,717	0,283	0,187
S	8850	6497	15347
	0,577	0,423	0,813
Total	11387	7500	18887

Nota satisfatória/insatisfatória por Procurou empresa

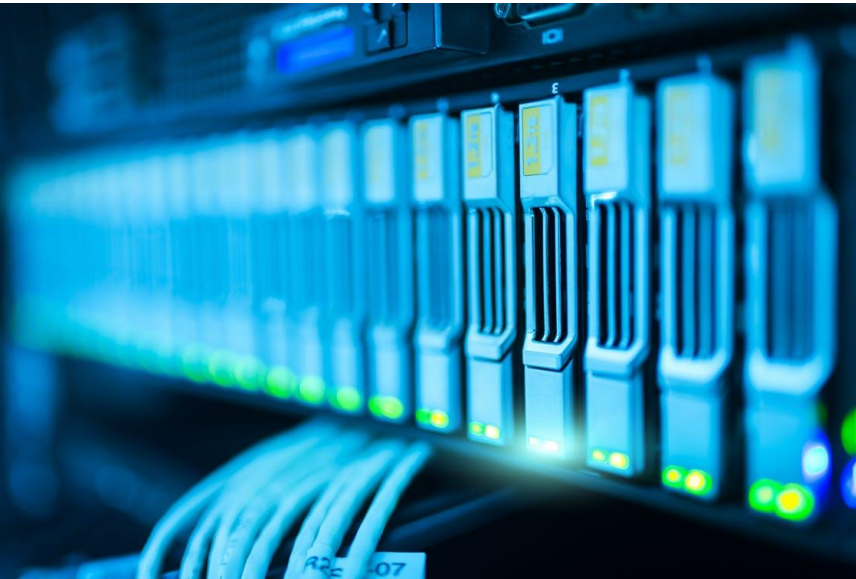
- Após agrupamento das notas 1 e 2 como insatisfatórias e 3, 4 e 5 como satisfatórias, é possível notar predominância de atribuição de notas insatisfatórias para aqueles que procuraram a empresa.



5. Modelagem Estatística Tradicional

BASES DE TREINO E TESTE

28



Tratamento das base de dados para modelagem

1. 75% aleatório para treino e 25% para teste
 - Treino: 14165 avaliações
 - Teste: 4722 avaliações
2. A variável nota insatisfatória conta com cerca de 40% das observações.

5.i Regressão Logística

MODELAGEM COM ESTATÍSTICA TRADICIONAL | INTERPRETAÇÃO DAS VARIÁVEIS

29

O modelo seleciona as variáveis mais relevantes e estima um peso para cada uma de suas categorias, atribuindo para cada consumidor a probabilidade de atribuir nota insatisfatória.

$$p = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4}}$$

Variável	Categoria	Coeficiente (β)	Interpretação em relação à nota
Faixa Etária	-	-2,59	Quanto maior a faixa etária, maior propensão a atribuir nota insatisfatória.
	De 21 a 30 anos	0,51	
	De 31 a 40 anos	0,90	
	De 41 a 50 anos	1,23	
	De 51 a 60 anos	1,29	
	De 61 a 70 anos	1,37	
Sexo	Mais de 70 anos	1,42	Homens são mais propensos a atribuir nota insatisfatória
	Homem	0,18	
Tempo de resposta	Prazo de resposta	0,10	Quanto maior o tempo de resposta, o consumidor fica mais propenso a atribuir nota insatisfatória
Procurou a empresa	Se consumidor procurou a empresa antes de registrar a reclamação no site	0,63	Se o consumidor já havia procurado a empresa antes de registrar a reclamação, ele é mais propenso a atribuir nota insatisfatória.
		0,51	



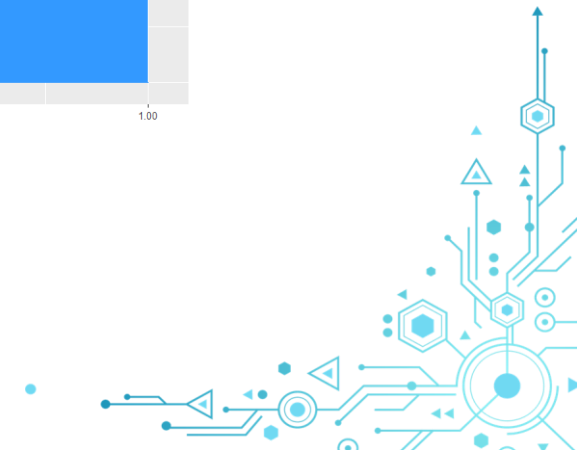
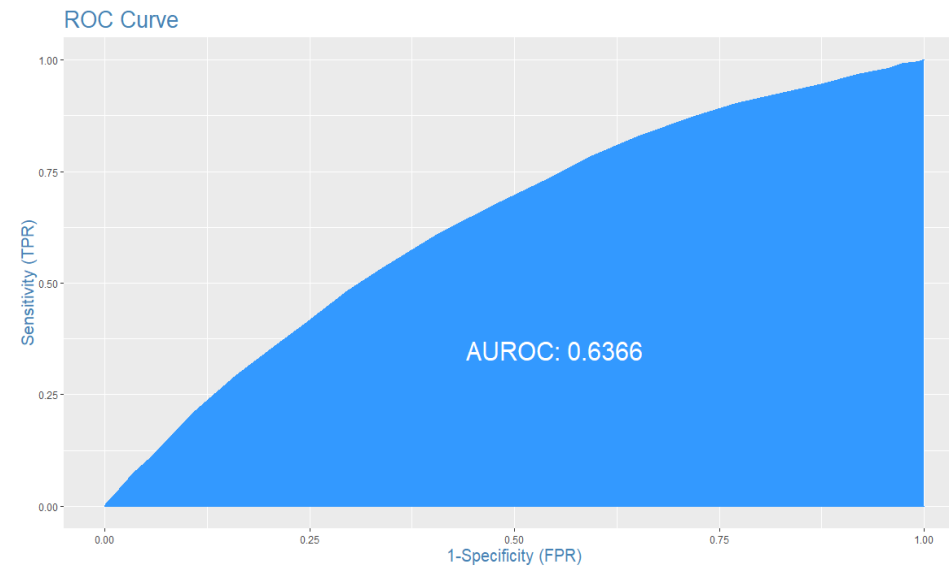
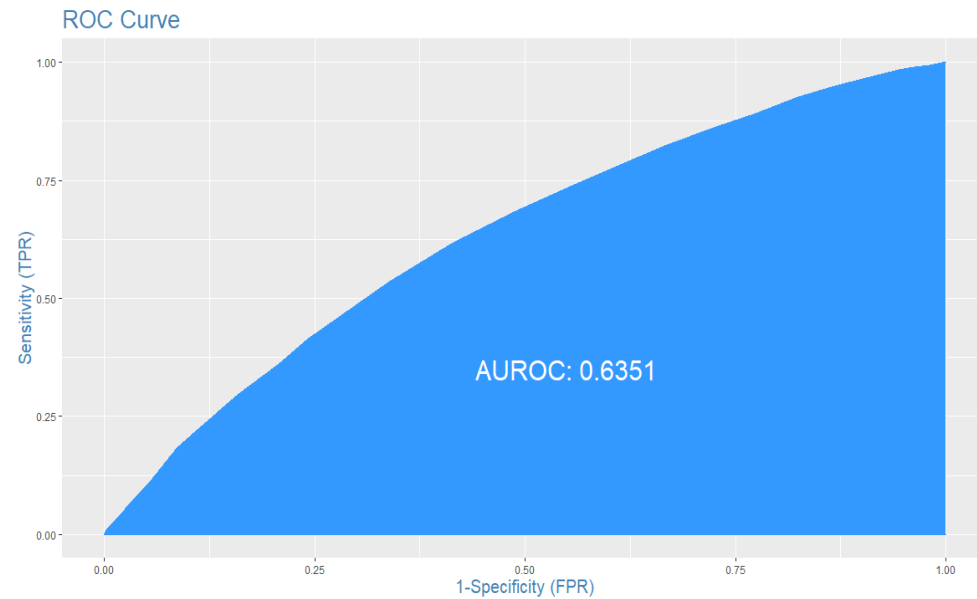
5.i Regressão Logística

MODELAGEM COM ESTATÍSTICA TRADICIONAL | INTERPRETAÇÃO DAS VARIÁVEIS

30

Treino
KS = 0,203

Teste
KS = 0,2

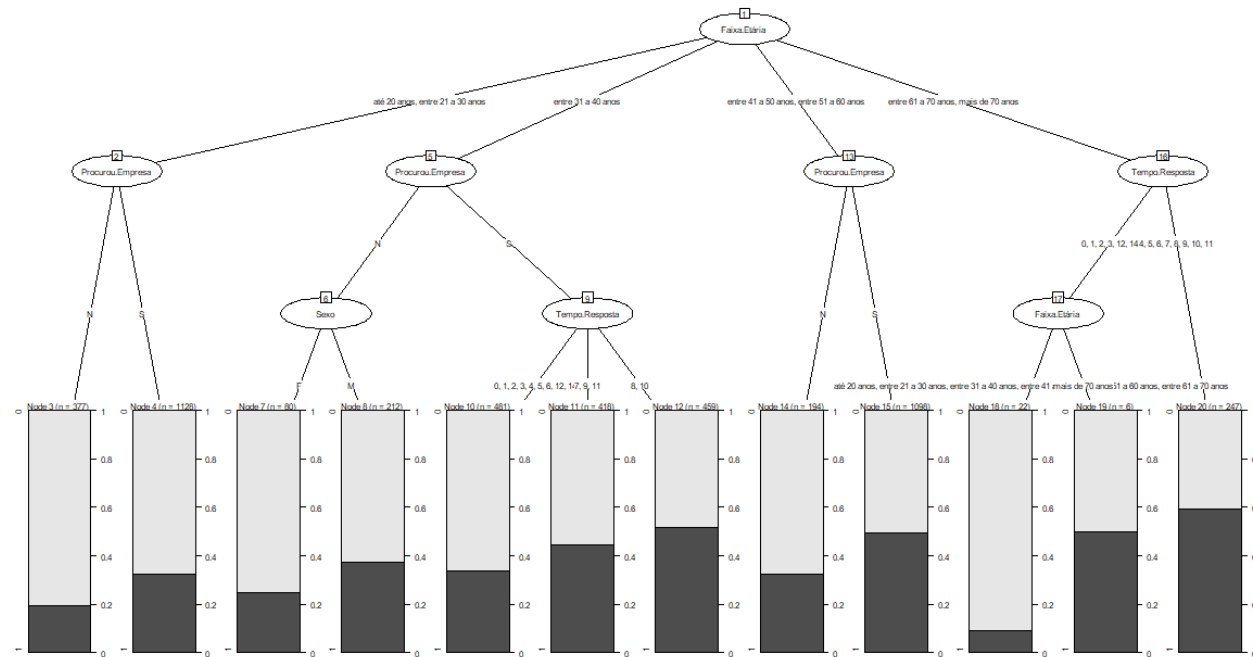


5.ii Árvore de decisão

MODELAGEM COM ESTATÍSTICA TRADICIONAL | INTERPRETAÇÃO DAS VARIÁVEIS

31

Classifica as observações pela combinação de características, por meio de uma árvore de classificação, predizendo se a nota é satisfatória ou não.



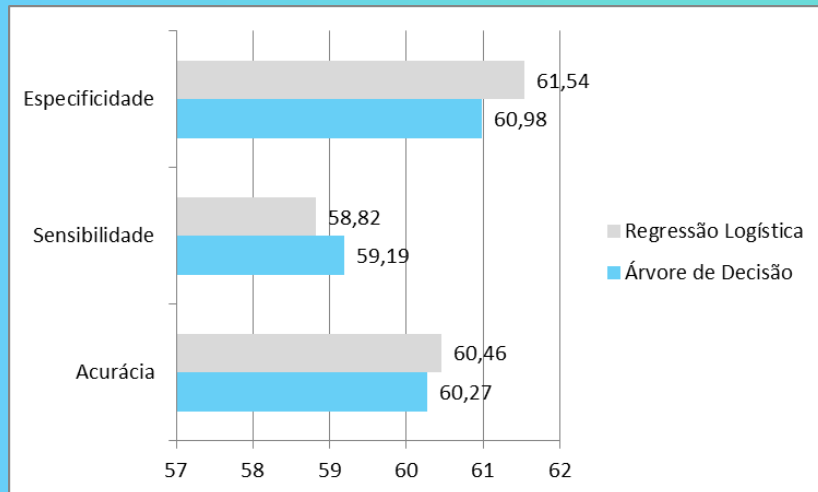
Intepretação:

- As variáveis mais importantes são respectivamente Faixa Etária, se procurou a empresa e o tempo de resposta.
- De acordo com o conjunto de teste, o perfil de quem atribui notas mais baixas é de pessoas a partir de 41 anos que tenham procurado a empresa.

5.iii Desempenho dos modelos

MODELAGEM COM ESTATÍSTICA TRADICIONAL | COMPARAÇÃO ENTRE TÉCNICAS

32



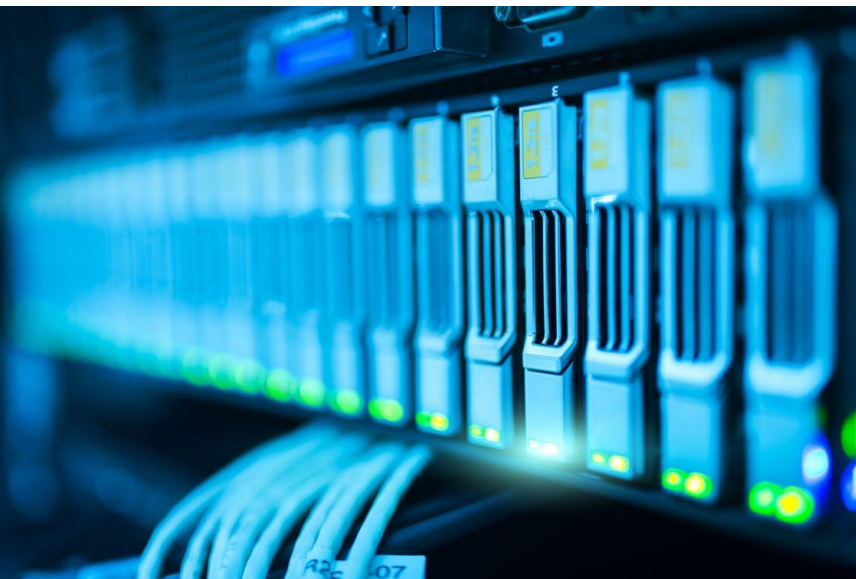
- Ambas as técnicas apresentaram **desempenho satisfatório**.
- A Árvore de Decisão apresentou **melhor desempenho** em relação a Regressão logística com relação à sensibilidade que é o acerto dos verdadeiros positivos. O ponto de corte da regressão logística foi otimizado para minimizar a diferença entre sensibilidade e especificidade.
- Ambos apresentaram desempenho semelhante nos conjuntos de treino e de teste.



6. Inteligência artificial

BASES DE TREINO E TESTE

33



Tratamento das base de dados para modelagem

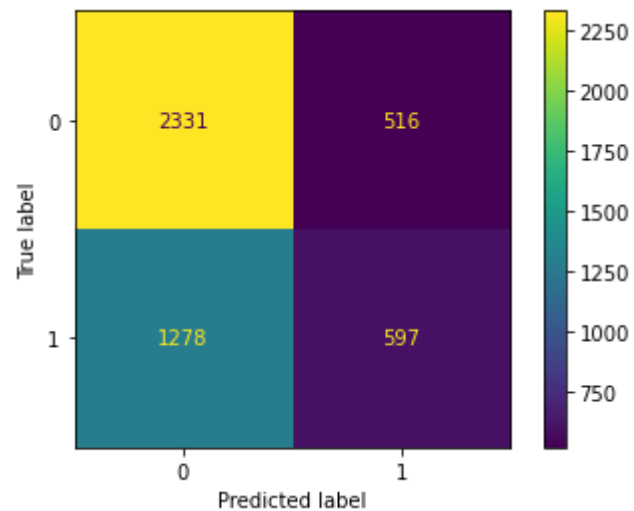
1. 75% aleatório para treino e 25% para teste
 - Treino: 14165 avaliações
 - Teste: 4722 avaliações
2. A variável nota insatisfatória conta com cerca de 40% das observações.

6.i Random Forest

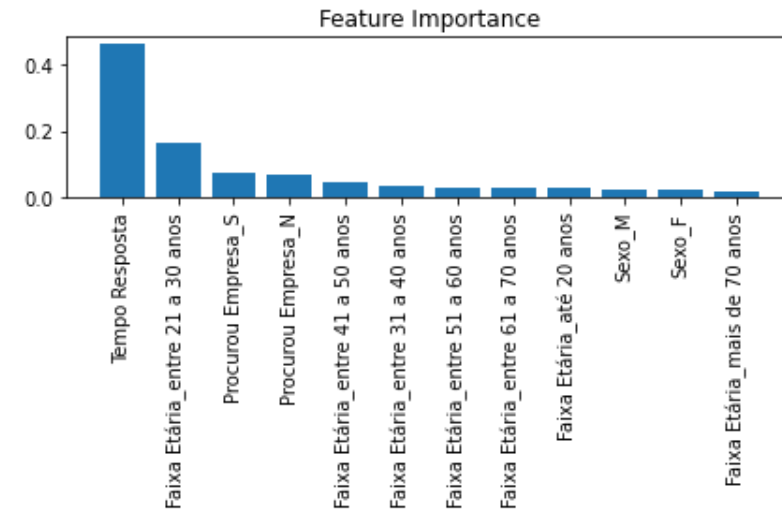
MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

34

Random forest é um algoritmo de aprendizagem supervisionada que cria uma combinação (*ensemble*) de árvores de decisão, na maioria dos casos treinados com o método de *bagging*.



Métrica	Base	valor
Acurácia	Treino	0,63
	Teste	0,62
Recall	Treino	0,34
	Teste	0,33
Precision	Treino	0,56
	Teste	0,54
ROC	Treino	0,65
	Teste	0,62



6.i Random Forest

MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

35

Random forest é um algoritmo de aprendizagem supervisionada que cria uma combinação (*ensemble*) de árvores de decisão, na maioria dos casos treinados com o método de *bagging*.

Variável	categoria	score	interpretação
Tempo Resposta		0,464	Quanto maior o tempo de resposta, maior a insatisfação.
	entre 21 a 30 anos	0,156	
	entre 41 a 50 anos	0,045	
Faixa Etária	entre 51 a 60 anos	0,041	Os clientes com idade entre 21 a 30 anos são os mais insatisfeitos.
	entre 31 a 40 anos	0,040	
	entre 61 a 70 anos	0,030	
	até 20 anos	0,029	
	mais de 70 anos	0,015	
Procurou Empresa	N	0,069	Não há diferença significativa entre aqueles que procuraram a empresa
	S	0,067	
Sexo	F	0,022	Não há diferença significativa entre o sexo dos clientes
	M	0,022	

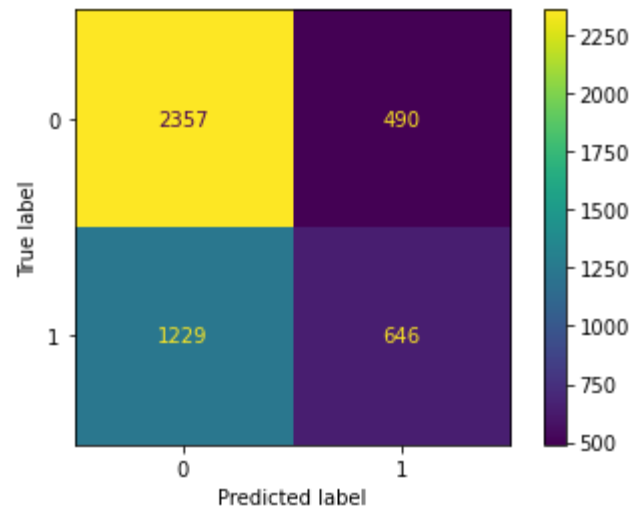


6.ii Gradient Boosting

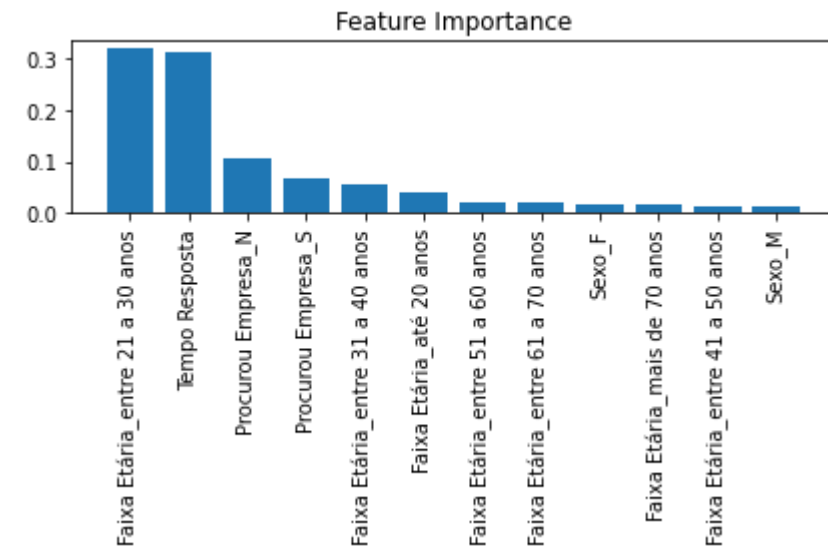
MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

36

Gradient boosting combina a descida de gradiente com a técnica de boosting. Tenta ajustar o novo preditor aos erros do preditor anterior.



Métrica	Base	valor
Acurácia	Treino	0,63
	Teste	0,63
Recall	Treino	0,32
	Teste	0,34
Precision	Treino	0,55
	Teste	0,56
ROC	Treino	0,64
	Teste	0,64



6.ii Gradient Boosting

MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

37

Gradient boosting combina a descida de gradiente com a técnica de boosting. Tenta ajustar o novo previsor aos erros do previsor anterior.

Variável	categoria	score	interpretação
Tempo Resposta		0,324	Quanto maior o tempo de resposta, maior a insatisfação.
	entre 21 a 30 anos	0,295	
	entre 31 a 40 anos	0,061	
Faixa Etária	até 20 anos	0,044	Os clientes com idade entre 21 a 30 anos são os mais insatisfeitos.
	entre 51 a 60 anos	0,038	
	entre 61 a 70 anos	0,021	
	entre 41 a 50 anos	0,018	
	mais de 70 anos	0,007	
Procurou Empresa	S	0,088	Não há diferença significativa entre aqueles que procuraram
	N	0,076	
Sexo	M	0,015	Não há diferença significativa entre o sexo dos clientes
	F	0,014	

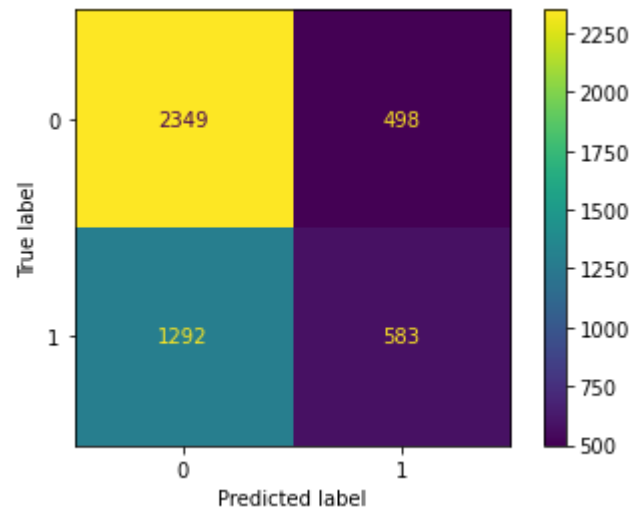


6.iii Extreme Gradient Boosting

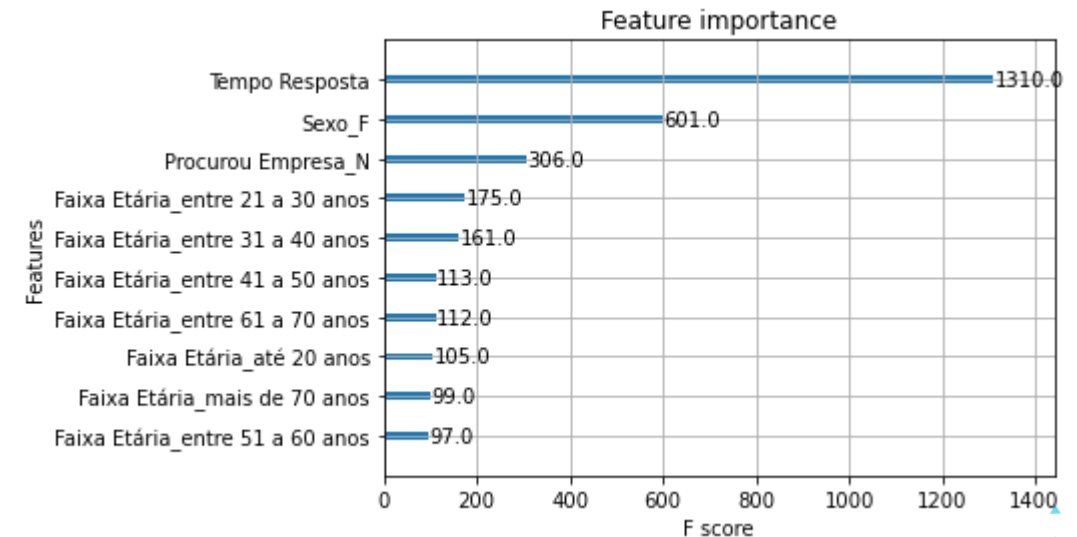
MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

38

Extreme Gradient boosting inclui uma variedade de regularizações que reduz o sobreajuste (overfitting) e melhora o desempenho geral.



Métrica	Base	valor
Acurácia	Treino	0,63
	Teste	0,62
Recall	Treino	0,32
	Teste	0,31
Precision	Treino	0,58
	Teste	0,54
ROC	Treino	0,65
	Teste	0,63

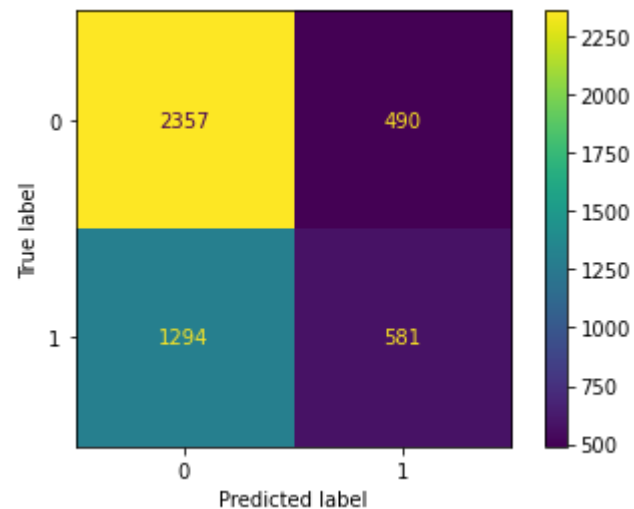


6.iv LightGBM

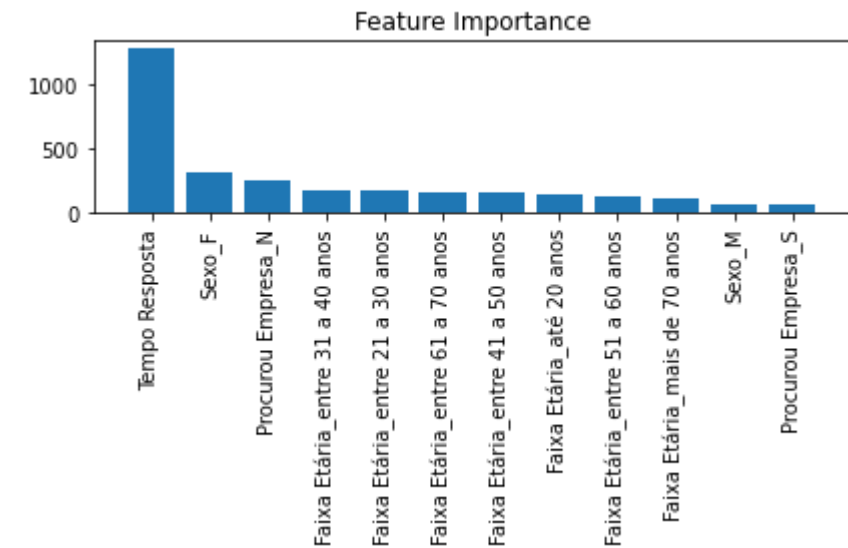
MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

39

Enquanto os outros algoritmos as árvores crescem horizontalmente, o algoritmo LGBM cresce verticalmente, ou seja, ele cresce em termos da folha da árvore.



Métrica	Base	valor
Acurácia	Treino	0,63
	Teste	0,62
Recall	Treino	0,32
	Teste	0,31
Precision	Treino	0,56
	Teste	0,54
ROC	Treino	0,65
	Teste	0,63



6.iv LightGBM

MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | IMPORTÂNCIA DAS VARIÁVEIS

40

Enquanto os outros algoritmos as árvores crescem horizontalmente, o algoritmo LGBM cresce verticalmente, ou seja, ele cresce em termos da folha da árvore.

Variável	categoria	score	interpretação
Tempo Resposta		1.393	Quanto maior o tempo de resposta, maior a insatisfação.
Sexo	F	295	Pessoas do sexo feminino tendem a ser mais insatisfeitas
	M	56	
	N	216	Pessoas que não
Procurou Empresa	S	58	procuraram a empresa
Faixa Etária	entre 31 a 40 anos	185	
	entre 21 a 30 anos	174	
	entre 41 a 50 anos	145	A faixa etária mais
	até 20 anos	136	insatisfeita está entre 31 a
	entre 61 a 70 anos	135	40 anos.
	entre 51 a 60 anos	119	
	mais de 70 anos	88	



6.v Desempenho dos modelos

MODELAGEM COM INTELIGÊNCIA ARTIFICIAL | COMPARAÇÃO ENTRE TÉCNICAS

41

- Mantidas as mesmas variáveis da análise estatística, todas as técnicas de inteligência artificial tiveram resultados muito próximos e não conseguiram traçar um perfil do consumidor. Com exceção de Gradient Boosting (idade entre 21 e 30 anos está muito próxima do tempo de resposta), em todos os casos a variável mais importante foi o tempo de resposta.

