



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

# **Caracterização e predição da (in)satisfação de clientes de uma operadora de telefonia móvel**

Por: Luiza Batista Laquini

Orientador: Prof. Vinícius Fernandes Soares Mota

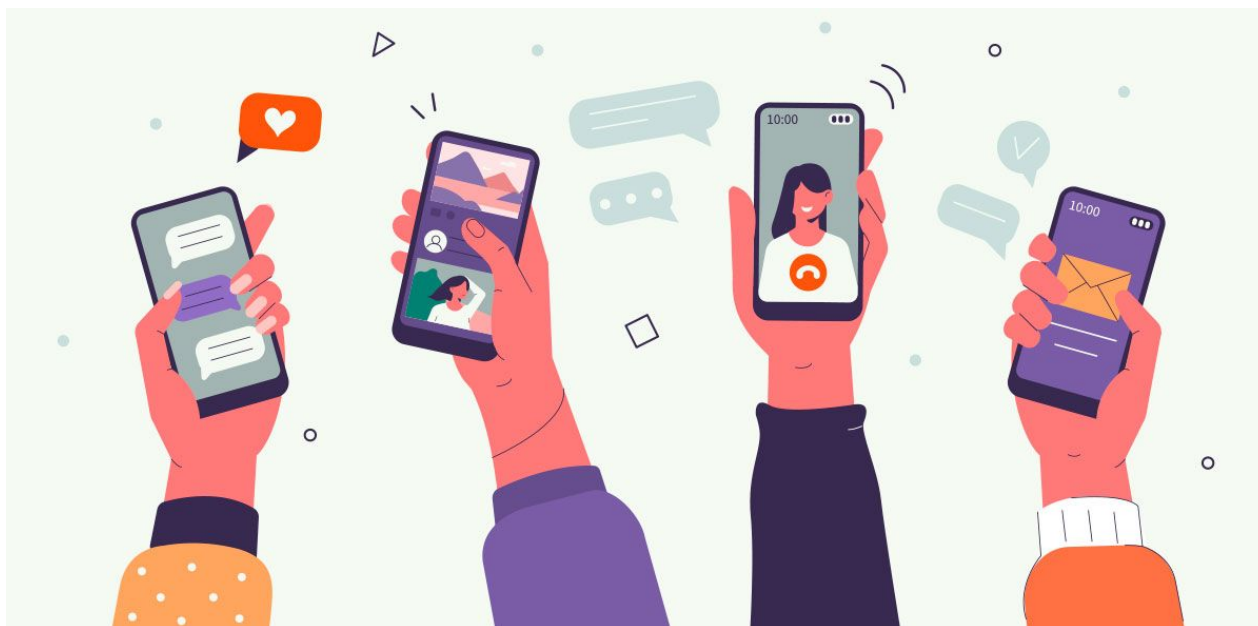
# Sumário

1. Introdução
2. Trabalhos relacionados
3. Análise inicial
4. Caracterização da (in)satisfação
5. Predição de detratores
6. Considerações finais

A solid blue vertical bar is positioned on the left side of the slide.

# Introdução

# Motivação



# Motivação

Clientes insatisfeitos possuem maior probabilidade de romper com o serviço da telefonia, comumente chamamos isso de *churn*.

Portanto, é de interesse da TIM saber de antemão quais clientes ficarão insatisfeitos para direcionar com mais eficiência ações de retenção em cima dos mesmos.

# Cenário

## Dicionário de Dados:

- Disponibilizado pela operadora de telefonia TIM
- Coleta feita de julho à dezembro de 2022 (6 meses)
- 74 colunas x 46.640 linhas (CSV)
- Cada linha é um cliente de telefonia móvel
- Cada coluna é uma *feature* (característica)

## Objetivos:

- Caracterizar a (in)satisfação » NPS
- Predizer detratores » Aprendizado de máquina

# Cenário

## Dicionário de Dados:

- Disponibilizado pela operadora de telefonia TIM
- Coleta feita de julho à dezembro de 2022 (6 meses)
- 74 colunas x 46.640 linhas (CSV )
- Cada linha é um cliente de telefonia móvel
- Cada coluna é uma *feature* (característica)

## Objetivos:

- Caracterizar a (in)satisfação » NPS
- Predizer detratores » Aprendizado de máquina

# Dicionário de Dados

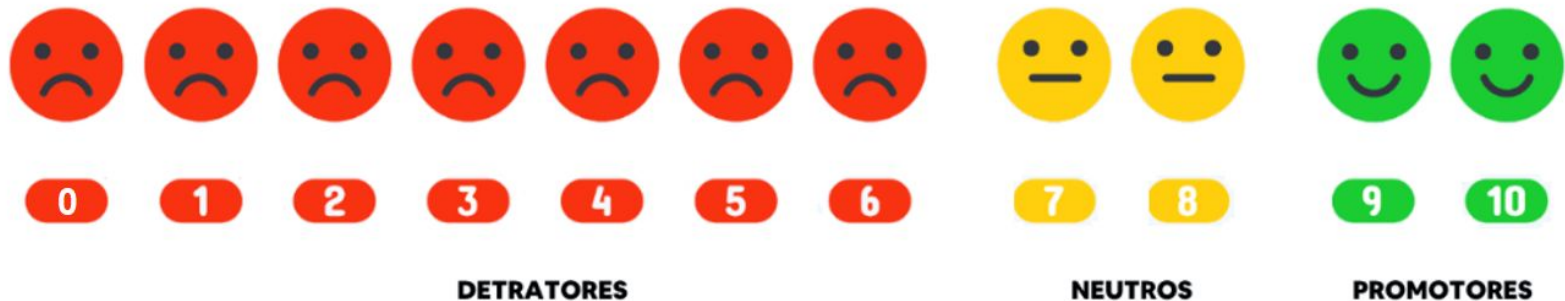
Descrição completa das 74 colunas: apêndice A da monografia. Destaques:

CAMPO	DESCRIÇÃO	TIPO
NPS_Class	Classe NPS (detrator, neutro ou promotor)	Categórico
USERVALUEQ2_VALUE	Atribuição à nota dada (sinal, estabilidade, etc)	Categórico
ANF	DDD do cliente	Numérico
term_fabr	Marca fabricante do dispositivo móvel	Categórico
tot_chamada	Total de chamadas	Numérico
originada_perc	Percentual de chamadas originadas	Numérico
qtde_celula	Quantidade de antenas utilizadas pelo usuário	Numérico
CUSTOMER_SEGMENT	Categoria do Plano (pré, controle ou pós)	Categórico
Qtd_CRM	Interações com CRM (duvidas, reclamações, etc)	Numérico
vol_Rede_Social	Consumo de dados em redes sociais	Numérico
vol_Navegacao	Consumo de dados em navegadores	Numérico
vol_MarketPlace	Consumo de dados na loja de aplicativos	Numérico
vol_Musica	Consumo de dados em música	Numérico



# ***Net Promoter Score (NPS)***

"Em uma escala de 0 a 10, o quanto você recomendaria nossa empresa/produto/serviço para um amigo ou colega?"



$$\text{NPS} = \% \text{😊} - \% \text{😞}$$

# Pesquisa realizada

## Pergunta 1 (quantitativa):

"Em uma escala de 0 a 10, o quanto você recomendaria nossa empresa/produto/serviço para um amigo ou colega?"

## Pergunta 2 (qualitativa):

"Com base em que cenário a nota foi escolhida?" Opções prontas (Justificativa para a nota):

- Sinal/Cobertura
- Velocidade
- Duração pacote dados
- Plano/Bônus
- Estabilidade
- Qualidade da internet
- Planos e ofertas
- Qualidade das ligações
- Atendimento
- Recarga/Conta
- Outros

A solid blue vertical bar is positioned on the left side of the slide, extending from the top to the bottom.

# **Trabalhos relacionados**

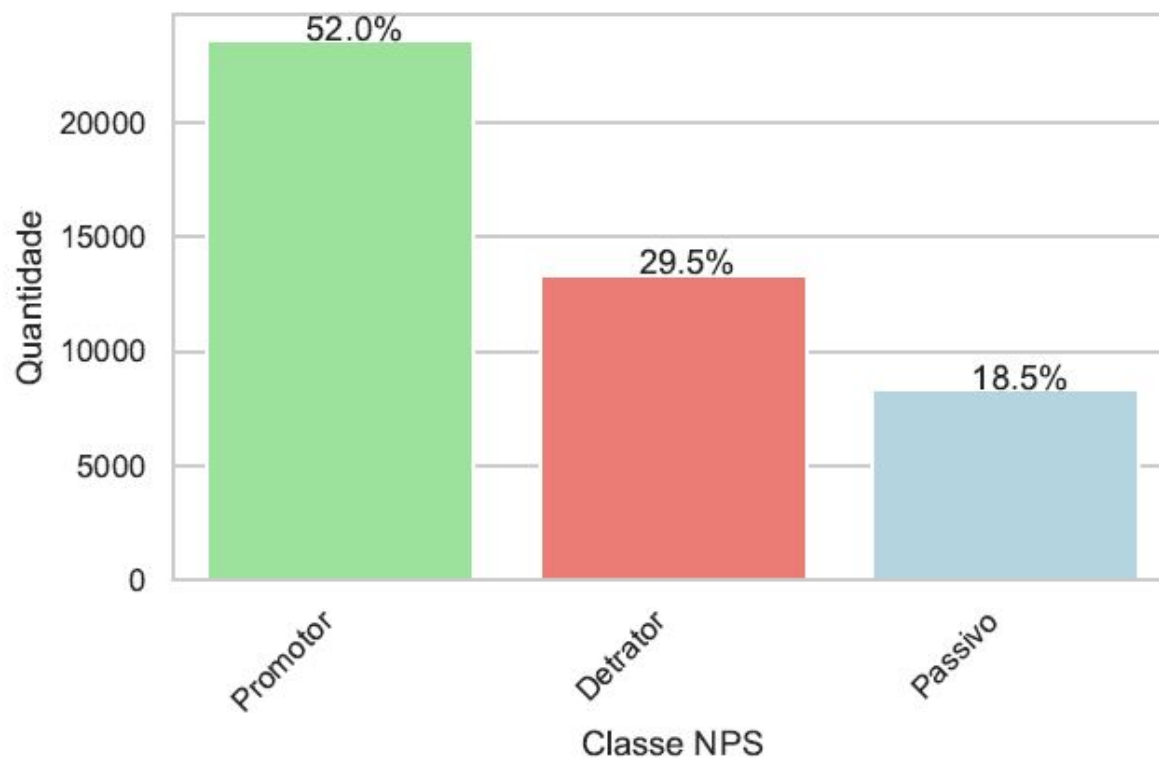
# Trabalhos relacionados

Autor	Base de dados	Metodologia	Melhor Algoritmo	Métrica Avaliadora
(HASSOUNA et al., 2015)	Operadora do Reino Unido	Predição de <i>churn</i>	Árvore de Decisão C5	AUC = 76.3% e Acurácia = 70.25%
(GAUR; DUBEY, 2018)	Real; Operadora não informada	Predição de <i>churn</i>	<i>Gradient Boosting</i>	AUC = 84.57%
(ULLAH et al., 2019)	GSM (sul da Ásia)	Predição de <i>churn</i>	<i>Random Forest</i>	Acurácia = 88.63%
(AHMAD; JAFAR; ALJOUAA, 2019)	SyriaTel (Síria)	Predição de <i>churn</i>	<i>Extreme Gradient Boosting</i>	AUC = 93.30%
(COPACEANU, 2021)	Real e pública; Operadora não informada	Predição de <i>churn</i>	Árvore de Decisão	<i>Recall</i> = 90.7%
(MUSTAFA; LING; RAZAK, 2021)	Operadora da Malásia	Predição de <i>churn</i> Análise do NPS	Classification and Regression Trees (CART)	Acurácia = 98%
(MARKOULIDAKIS et al., 2020)	Pequena amostra real + grande parte gerada	Predição da classe NPS	<i>Random Forest</i>	<i>F1-Score</i> = 79%
Presente Projeto	TIM (Brasil)	Predição de detratores Análise do NPS	<i>Random Forest</i>	<i>F1-Score</i> = 61%

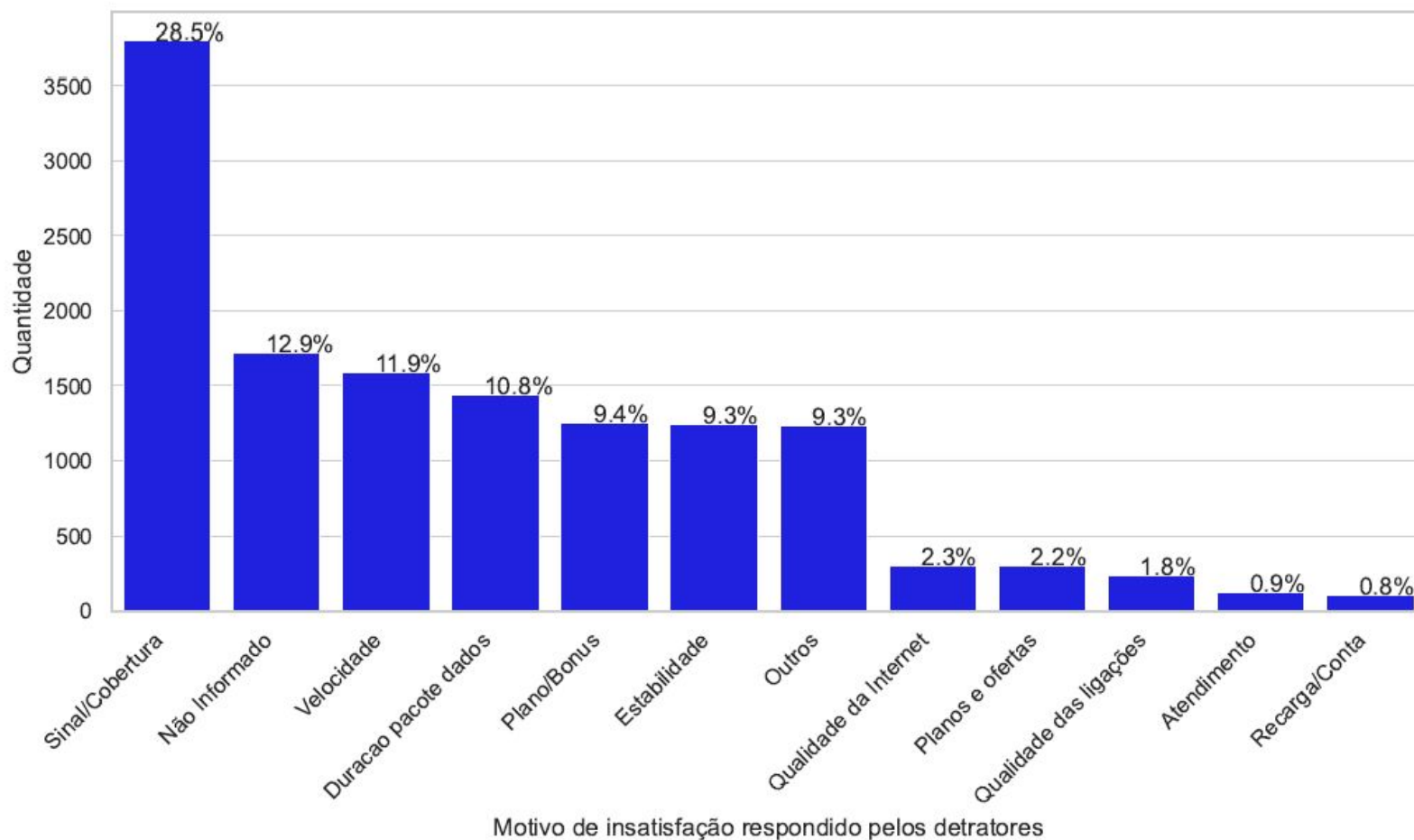
A solid blue vertical bar is positioned on the left side of the slide, extending from the top to the bottom.

# **Análise inicial**

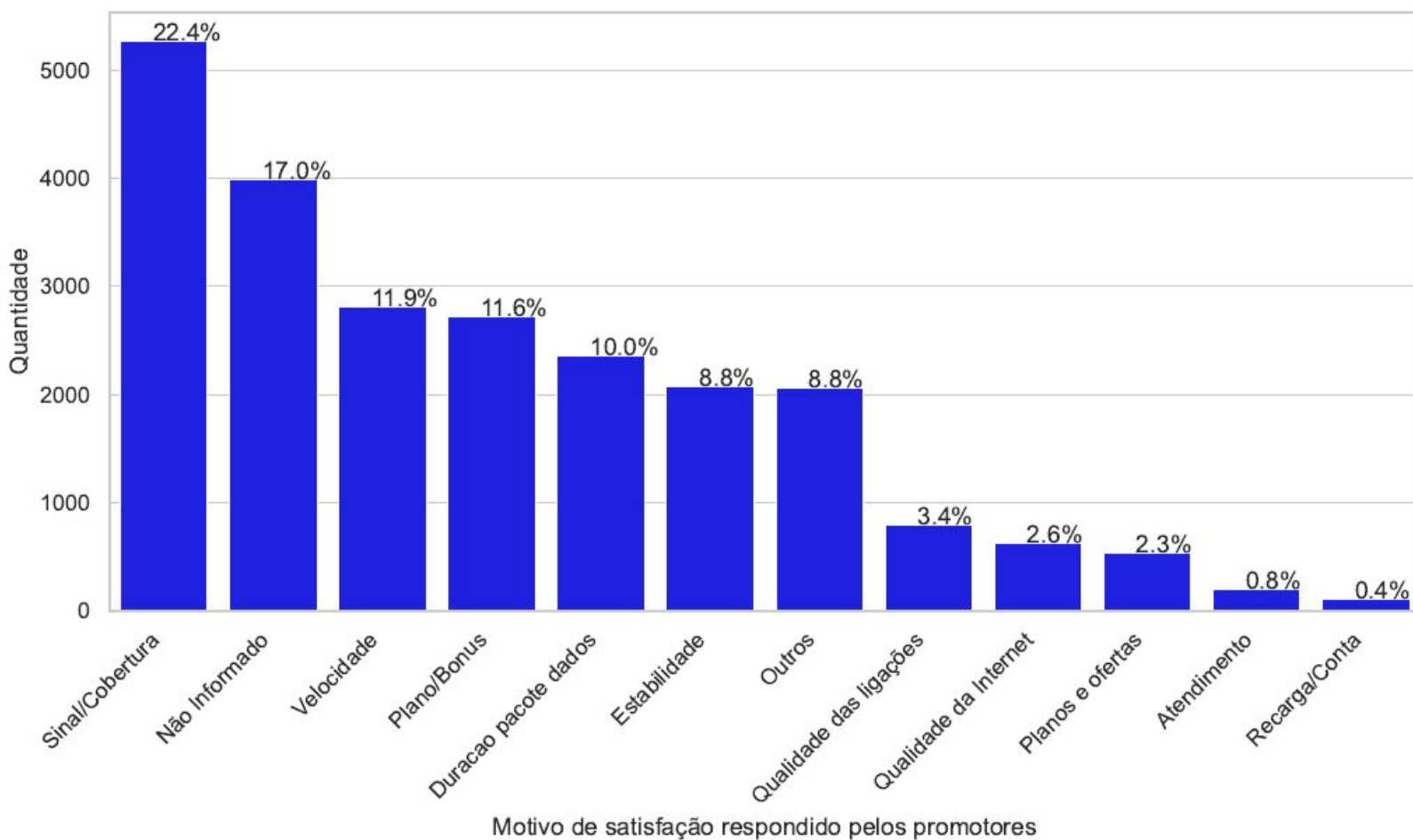
# Análise Inicial das respostas à pesquisa



# Análise Inicial das respostas à pesquisa



# Análise Inicial das respostas à pesquisa





# **Caracterização da (in)satisfação**

# Caracterização com o NPS - Social e demográfico

NPS e IDH:

ESTADO			REGIÃO		
Sigla	NPS	IDH	Nome	NPS médio	IDH médio
AC	0.146	0,710	Norte	0.33	0.703
AM	0.421	0,700			
AP	0.329	0,688			
PA	0.333	0,690			
RO	0.131	0,700			
RR	0.286	0,699			
TO	0.041	0,731			
AL	0.239	0,684	Nordeste	0.294	0.702
BA	0.266	0,691			
CE	0.330	0,734			
MA	0.341	0,676			
PB	0.337	0,698			
PE	0.279	0,719			
PI	0.320	0,690			
RN	0.327	0,728			
SE	0.268	0,702			

# Caracterização com o NPS - Social e demográfico

NPS e IDH:

ESTADO			REGIÃO		
Sigla	NPS	IDH	Nome	NPS médio	IDH médio
DF	0.135	0.814	Centro-Oeste	0.169	0.757
GO	0.111	0.737			
MS	0.351	0.742			
MT	0.247	0,736			
PR	0.191	0,769	Sul	0.174	0.777
RS	0.151	0,771			
SC	0.156	0,792			
ES	0.209	0,771	Sudeste	0.215	0.778
MG	0.150	0,774			
RJ	0.265	0,762			
SP	0.209	0,806			

# Caracterização com o NPS - Social e demográfico

Hipótese levantada:

Fatores sociais e demográficos como maior nível de escolaridade e maior poder econômico podem estar intimamente relacionados com expectativas mais elevadas, maiores sensibilidades a problemas ou falhas, maior consciência das alternativas de mercado e maior exigência de modo geral.

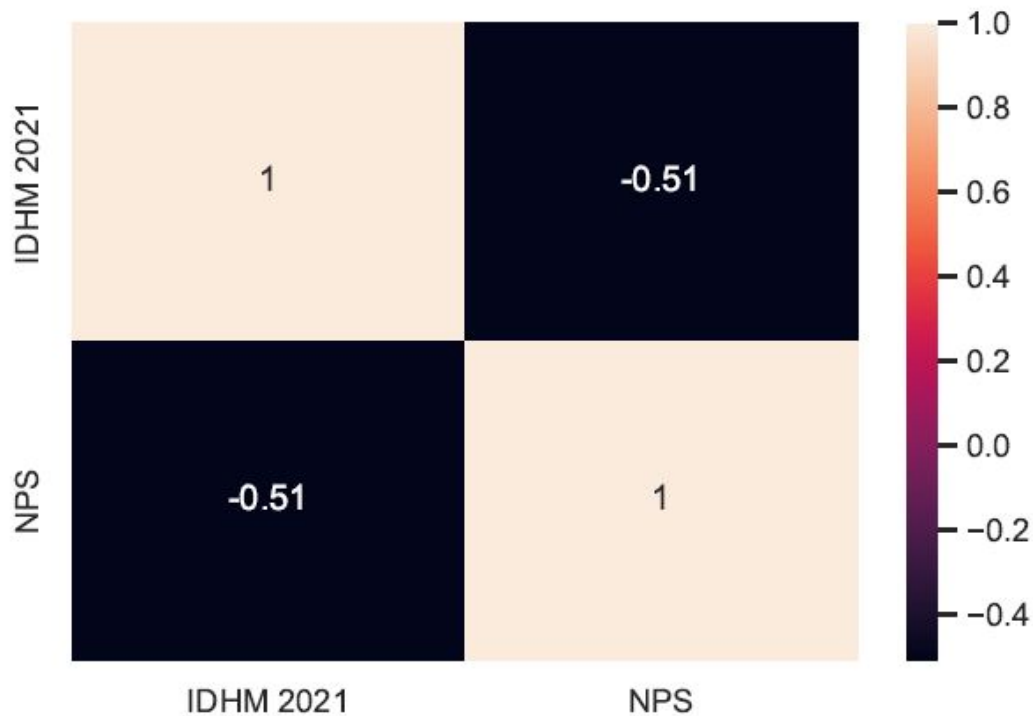
# Caracterização com o NPS - Social e demográfico

Marca fabricante do dispositivo móvel:

Marca Fabricante	Detrator	Passivo	Promotor	NPS	Percentual
lg	25.31	17.18	57.51	32.2	7.59
outros	24.57	20.42	55.01	30.44	2.45
não informado	27.97	17.85	54.18	26.21	4.9
motorola	29.28	18.54	52.18	22.91	28.04
samsung	29.68	18.58	51.74	22.07	47.04
asus	32.93	19.51	47.56	14.63	1.27
xiaomi	34.18	17.24	48.58	14.39	4.35
apple	35.09	19.93	44.98	9.89	4.36

# Caracterização com o NPS - Social e demográfico

Correlação linear entre NPS e IDH:



# Caracterização com o NPS - Consumo de dados

## Redes Sociais:

Consumo de Dados	Detrator	Passivo	Promotor	NPS	Percentual
Baixo Consumo	26.9	19.09	54.01	27.11	33.83
Médio Consumo	29.62	18.41	51.96	22.34	35.74
Alto Consumo	31.18	18.52	50.3	19.13	13.86
Extremo Consumo	33.13	17.37	49.51	16.38	16.57

## Navegação na internet:

Consumo de Dados	Detrator	Passivo	Promotor	NPS	Percentual
Baixo Consumo	27.79	18.12	54.09	26.3	78.84
Médio Consumo	34.39	19.94	45.67	11.27	8.91
Alto Consumo	35.02	20.66	44.32	9.3	4.85
Extremo Consumo	38.18	19.16	42.66	4.48	7.4

# Caracterização com o NPS - Consumo de dados

Música:

Consumo de Dados	Detrator	Passivo	Promotor	NPS	Percentual
Baixo Consumo	28.2	17.98	53.82	25.62	78.84
Alto Consumo	32.11	21.1	46.79	14.68	2.65
Médio Consumo	32.69	20.21	47.09	14.4	3.35
Extremo Consumo	35.1	20.25	44.65	9.54	15.16

*Marketplace:*

Consumo de Dados	Detrator	Passivo	Promotor	NPS	Percentual
Baixo Consumo	28.38	18.49	53.13	24.74	86.84
Médio Consumo	34.68	18.95	46.37	11.69	6.45
Alto Consumo	37.73	18.32	43.95	6.23	3.05
Extremo Consumo	39.94	17.7	42.36	2.42	3.65



# Caracterização com o NPS - Consumo de dados

Hipótese levantada:

Quanto maior o consumo, maiores as chances de enfrentar problemas de sinal, estabilidade, entre outras questões que podem levar a uma experiência negativa com o serviço. Dessa forma, quanto maior o consumo, maiores as chances de o usuário se tornar um detrator.

# Caracterização com o NPS - Técnico/interativo

Consumo de chamadas telefônicas:

Consumo de Chamadas	Detrator	Passivo	Promotor	NPS	Percentual
Baixo Consumo	26.78	17.82	55.4	28.61	25.62
Médio Consumo	29.92	19.29	50.79	20.87	43.56
Alto Consumo	30.77	18.48	50.76	19.99	16.69
Extremo Consumo	31.61	17.22	51.17	19.56	14.13

Situação das chamadas:

Situação Chamadas	Detrator	Passivo	Promotor	NPS	Percentual
Mais recebidas	28.28	17.28	54.44	26.16	29.46
Mais originadas	29.85	18.99	51.16	21.31	57.99
Equilibrado	30.74	18.95	50.32	19.58	12.55

# Caracterização com o NPS - Técnico/interativo

Mobilidade:

Mobilidade	Detrator	Passivo	Promotor	NPS	Percentual
Baixa Mobilidade	25.21	16.78	58.01	32.8	28.3
Média Mobilidade	28.56	18.53	52.92	24.36	39.76
Alta Mobilidade	32.77	20.32	46.91	14.14	21.84
Extrema Mobilidade	38.16	19.12	42.72	4.56	10.1

Plano/Segmento de contrato:

Plano	Detrator	Passivo	Promotor	NPS	Percentual
Pré-pago	25.77	15.55	58.68	32.92	34.84
Pós-pago	30.24	20.8	48.96	18.72	32.88
Controle	32.77	19.3	47.93	15.17	32.28

# Caracterização com o NPS - Técnico/interativo

Situação chamados:

Situação Chamados	Detrator	Passivo	Promotor	NPS	Percentual
Nunca abriu chamado	28.68	18.57	52.75	24.07	93.69
Já abriu pelo menos 1 chamado	41.65	17.22	41.13	-0.53	6.31

# Caracterização com o NPS - Técnico/interativo

Hipóteses levantadas:

Quando o usuário utiliza muito do recurso de ligações ou quando ele se desloca mais, maiores as chances de se deparar com problemas de rede. Portanto, maiores as probabilidades de se tornarem detratores

Clientes com maior controle financeiro sobre seu plano (pré-pago) tendem a uma maior satisfação.

Usuários que abrem reclamações possivelmente tiveram experiências negativas que já afetaram negativamente a percepção deles em relação ao serviço prestado pela operadora.

# Perfil do cliente

Características em que sua ocorrência pode indicar uma maior probabilidade do usuário se tornar um **detrator**:

- Possuir um iPhone;
- Consumir muitos dados seja em redes sociais, navegando na internet, ouvindo música, navegando na loja de aplicativos ou outros cenários;
- Realizar muitas ligações telefônicas;
- Possuir alta mobilidade;
- Possuir plano controle ou pós-pago;
- Abrir chamados ou reclamações.

# Perfil do cliente

Características em que sua ocorrência pode indicar uma maior probabilidade do usuário se tornar um **promotor**:

- Possuir um aparelho celular Android;
- Consumir poucos dados;
- Realizar poucas ligações telefônicas;
- Possuir baixa mobilidade;
- Possuir plano pré-pago;
- Não abrir chamados ou reclamações.

A solid blue vertical bar is positioned on the left side of the slide.

# **Predição de detratores**



# Predição de detratores

A predição de detratores nos dados de telefonia se trata de um **problema de classificação**, pois temos uma classe (Detratores) e queremos prever se os clientes pertencem ou não a ela.

Problemas de classificação são problemas de **aprendizado supervisionado**, já que temos todos os dados rotulados.

As métricas utilizadas para avaliação do desempenho dos modelos foram a acurácia, a precisão, o recall e a f1-score. A métrica principal, que buscou ser maximizada nos modelos, foi a métrica **f1-score**, por ser considerada um equilíbrio entre precisão e recall.

# Modelos escolhidos

Foram testados cinco modelos inicialmente: KNN, Regressão Logística, *Random Forest*, XGBoost e Árvore de Decisão. Entretanto, foi decidido prosseguir para a otimização de hiperparâmetros somente com os modelos Random Forest e Extreme Gradient Boosting, por terem apresentado os melhores resultados iniciais.

Todavia, os resultados aqui trazidos são do modelo ***Random Forest***, que se destacou em todas as abordagens.

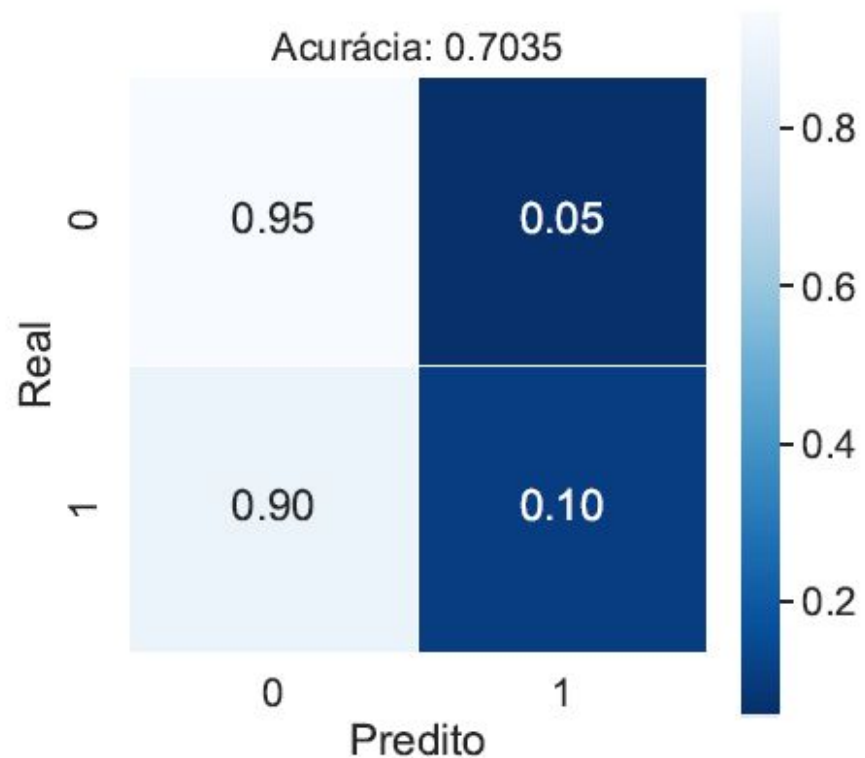
# Primeira abordagem: conjunto completo

Acurácia: 70%

Precisão: 65%

Recall: 70%

F1-score: 60%



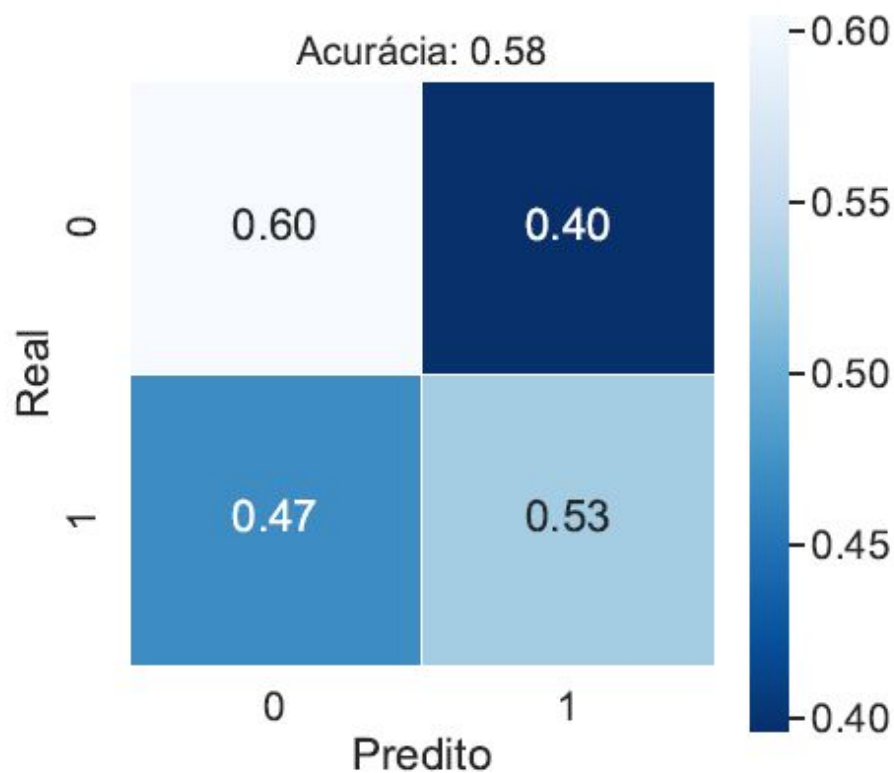
## Segunda abordagem: extremos

Acurácia: 58%

Precisão: 64%

Recall: 58%

F1-score: 60%



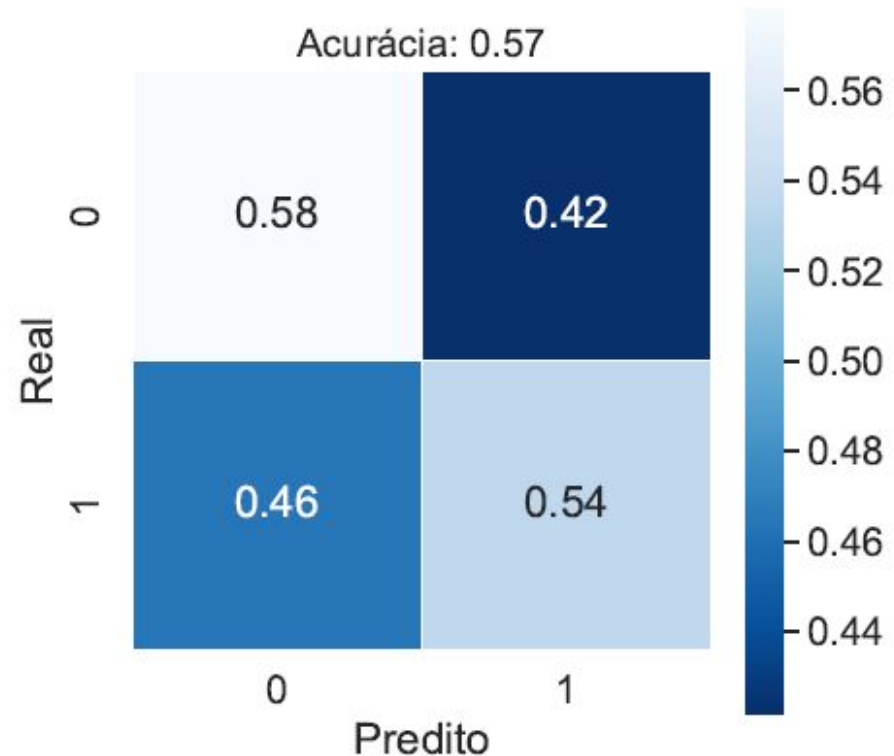
## Terceira abordagem: extremos estendidos

Acurácia: 57%

Precisão: 63%

Recall: 57%

F1-score: 59%



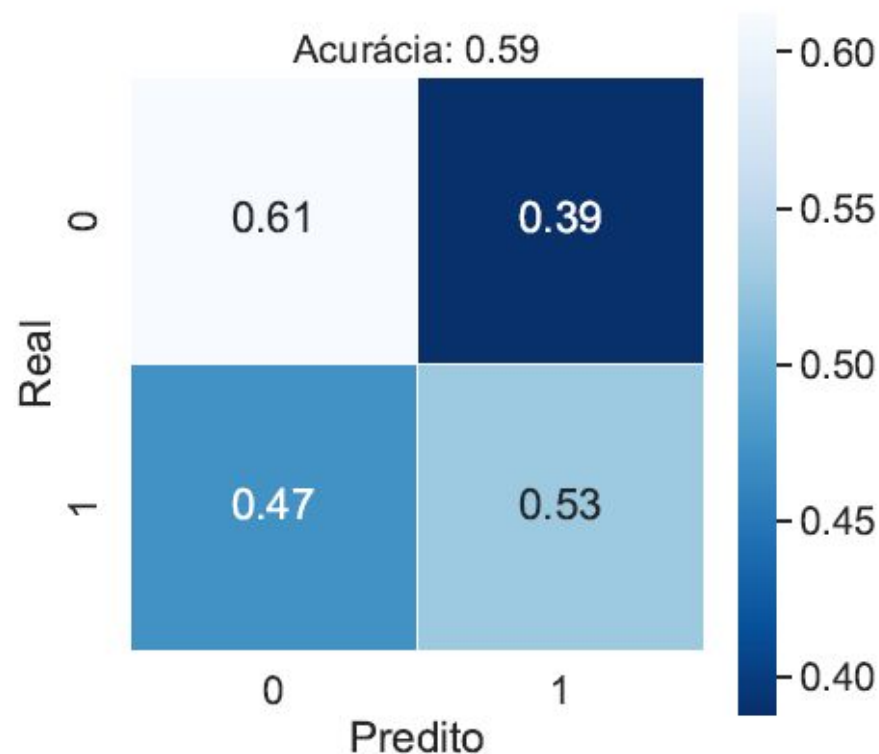
## Quarta abordagem: amostras equilibradas

Acurácia: 59%

Precisão: 64%

Recall: 59%

F1-score: 61%



## Quinta abordagem: três classes

- Classe 0: clientes certamente insatisfeitos - que deram notas 0, 1, 2 ou 3.
- Classe 1: clientes confusos para o modelo - que deram notas 4, 5, 6, 7 e 8.
- Classe 2: clientes certamente satisfeitos - que deram notas 9 e 10.

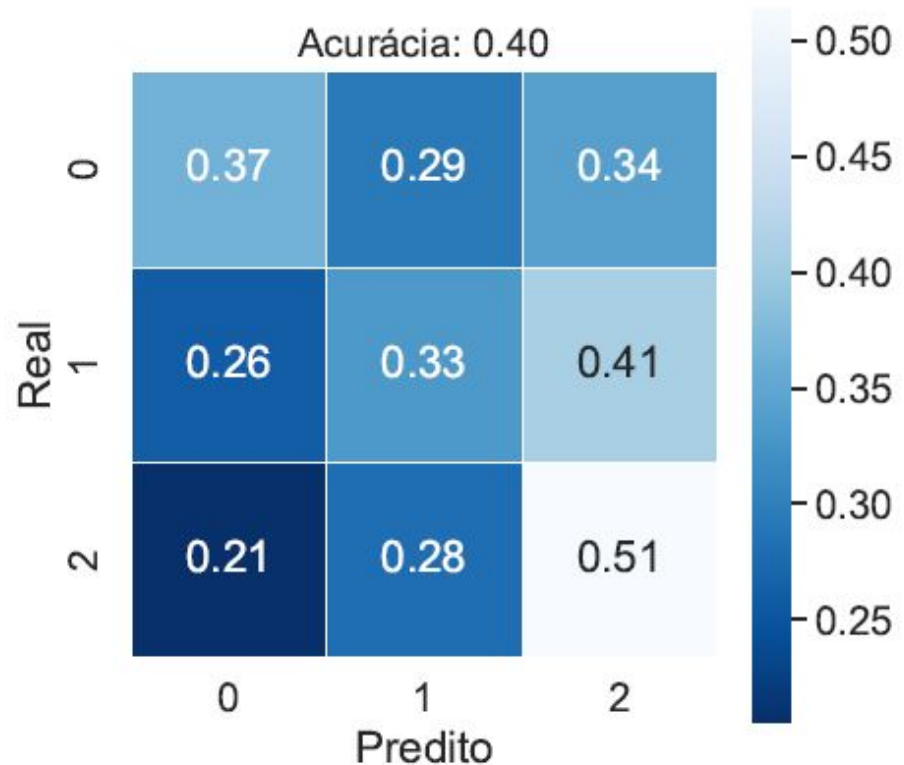
## Quinta abordagem: três classes

Acurácia: 40%

Precisão: 41%

Recall: 40%

F1-score: 40%

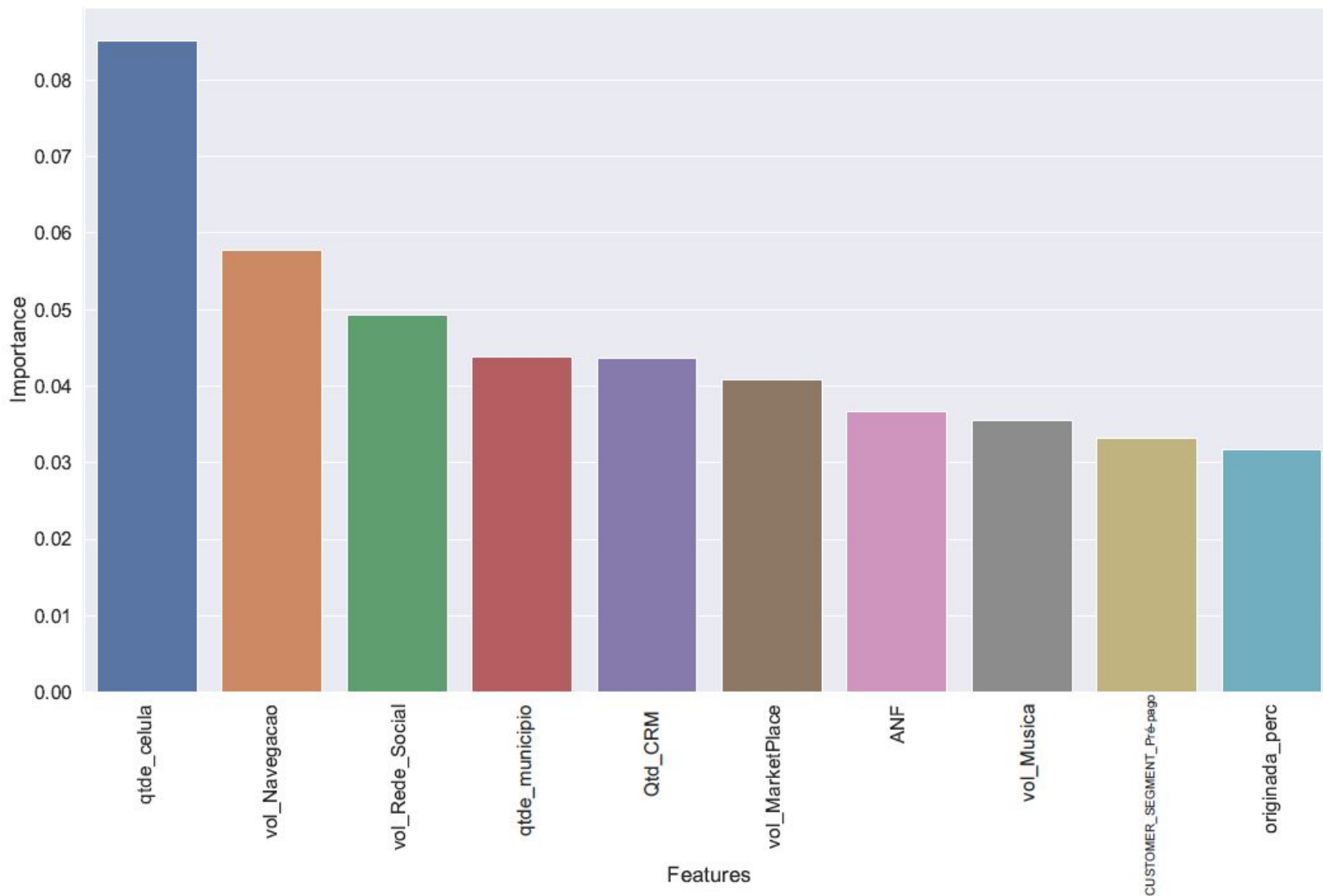




# Comparativo das abordagens

Abordagem	Acurácia(%)	Precisão(%)	<i>Recall</i> (%)	<i>F1-Score</i> (%)
Conjunto completo	70	65	70	60
Extremos	58	64	58	60
Extremos estendidos	57	63	57	59
Amostras equilibradas	59	64	59	61
Três classes	40	41	40	40

# *Feature Importance*



A solid blue vertical bar is positioned on the left side of the slide.

# **Considerações finais**

# Limitações e desafios

- Desbalanceamento das classes que impediu a utilização da base completa no treino dos modelos;
- Veracidade das respostas à pesquisa;
- Não conhecimento das unidades de medida das variáveis numéricas e ausência da documentação completa.

# Conclusões

Foi analisado o NPS para cada variável presente no dicionário de dados. Foram selecionadas e trazidas aqui as que se destacaram.

Foram treinados diferentes algoritmos de aprendizado de máquina e o que se destacou foi o *Random Forest*. Esse obteve 61% de *f1-score* para a abordagem com amostras equilibradas

Espera-se que a TIM possa utilizar desses resultados para o sucesso do seu negócio. Entretanto, apesar do bom resultado já alcançado, melhorias ainda podem ser implementadas.

# Trabalhos futuros

Sugestões para trabalhos futuros:

- Explorar outras métricas de satisfação como a CSAT e a *Retention Rate*;
- Calcular custos dos erros para priorizar *precision* ou *recall*;
- Obter mais dados dos detratores para um melhor balanceamento dos dados e também para que o modelo possa aprender mais sobre o alvo;
- Testar a atribuição de pesos, possível para os modelos de árvore;
- Implementar busca por *outliers* multivariados.



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Obrigada!