

UFRR – EMBRAPA

# CLASSIFICAÇÃO AUTOMÁTICA EM PASTAGEM USANDO MACHINE LEARNING

IBUKUN CHIFE DIDIER ADJITCHE

MAIO 2020

## PARTIE 2

IMPLEMENTAÇÃO

# ETAPAS

## DADOS

- Imagens Coletadas

## PROCESSAMENTO

- Recortes
- Tratamentos

## SEGMENTAÇÃO

- Remoção do Fundo

## RECURSOS

- Extração de Características

# ETAPAS

## CLASSIFICAÇÃO

- KNN
- SVM
- RDFC
- XGboost

## RESULTADOS

- Matriz de Confusão
- Accuracia

# DADOS

## COLETA DE DADOS

- Três de fonte de dados
- Lupa, Câmera foto, Celular
- lupa : 735 imagens
- Câmera : 322 imagens
- Celular : 434 imagens
- imagens de fundo magenta
- Uso da Caixa cubica para as imagens da Câmera e do celular
- Quatro iluminação LED 75 ml/W de 12 W

## PROCESSAMENTO

## RECORTE E TRATAMENTO

- Todas imagens foram recortados afim de focar mais nos insetos.
- Aplicar as propriedades nas imagens da Câmera e do celular : temperatura: 4550 , colorir: -2, exposição +0,40, vibração +25.
- Em todas as imagens foram aplicadas a Saturação -100.
- Photoshop

# SEGMENTAÇÃO

## REMOÇÃO DO FUNDO

- Converter a tom cinza a imagem
- Remover os ruídos
- Usar o algoritmo de erosão e de dilatação para uma suavização linear
- Abstrair o retrato do fundo .
- Subtrair o fundo da imagem original usando Bitwise\_and

## RECURSOS

## CARACTERÍSTICAS

- Foram extraídas as características de Cor, da textura e de forma
- As bibliotecas Mohatas e Open-Cv possibilitam o uso dos algoritmos: HU MOMENTS para as formas, HARALICK TEXTURE para as Texturas e COLOR HISTOGRAM para a Cor.



# CLASSIFICAÇÃO

## DATASET

Separar os dados de Treinamento de Teste e de Validação

## TRAIN

Treinar os algoritmos com os dados de treinamento.

## TEST

Usar o modelo preditivo e os dados de teste para avaliar a aprendizagem

## VALIDAÇÃO

Fazer uso da dados diferentes aos de Train e de Test na apreciação da performance do algoritmos

# DADOS DE VALIDAÇÃO

## VALIDAÇÃO

Foram considerados como dados de validação, imagem ulteriormente tiradas no inicio da nossas pesquisas pela lupa eletrônica tendo como cor de fundo: a cor branca.

As imagens dos 3 dados de Validação são as mesmas mas cadas apresentam umas particularidades :

# DADOS DE VALIDAÇÃO

## VALIDAÇÃO

A primeira : são imagens sem processamento obtida diretamente da Lupa

A segunda : são imagens da primeira segmentada afim de diminuir os custos de processamento e focar no alvo.

A terceira: são imagens da segunda com a retirada do parametro de saturação conforme usado nos dados de treinamento.



**KNN**

**RDFC**

**SVM**

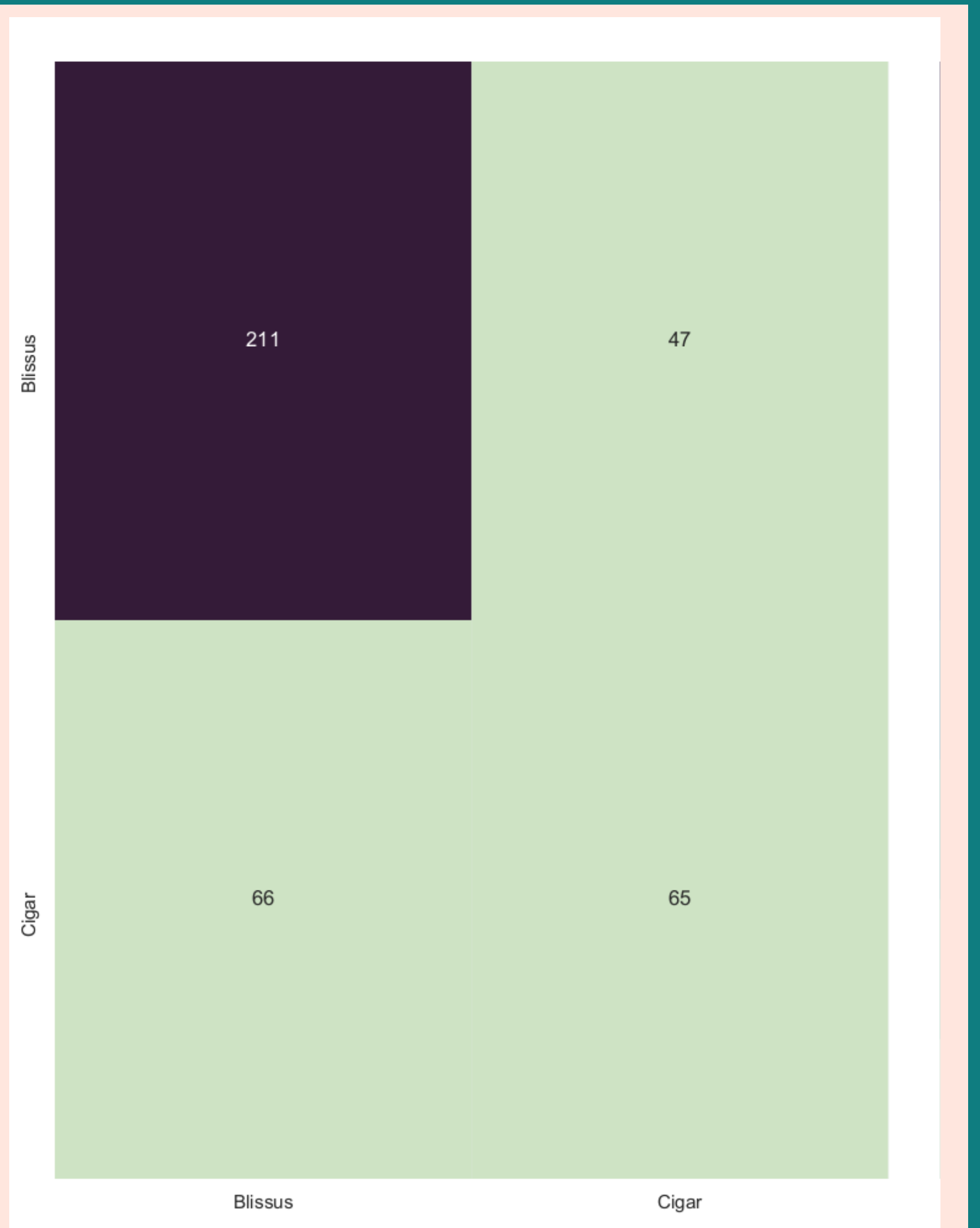
**XGB**

# KNN

## Dados de Teste

Acurácia : 70,95 %

	Precision	Recall	Num
B.pul	0.76	0.82	258
Ciga	0.58	0.50	131



# KNN – RESULTADO – VALIDAÇÃO

	SEM SEGMENTAÇÃO								COM SEGMENTAÇÃO						COM SEGMENTAÇÃO SATURAÇÃO -100										
	Acurácia : 76,06 %								Acurácia : 79,67 %						Acurácia : 60,83 %										
	Precision	Recall	Num					Precision	Recall	Num					Precision	Recall	Num								
	B.pul	0.94	0.79	265				B.pul	0.89	0.88	265				B.pul	0.94	0.61	200							
	Ciga	0.10	0.32	19				Ciga	0.24	0.25	40				Ciga	0.10	0.53	17							

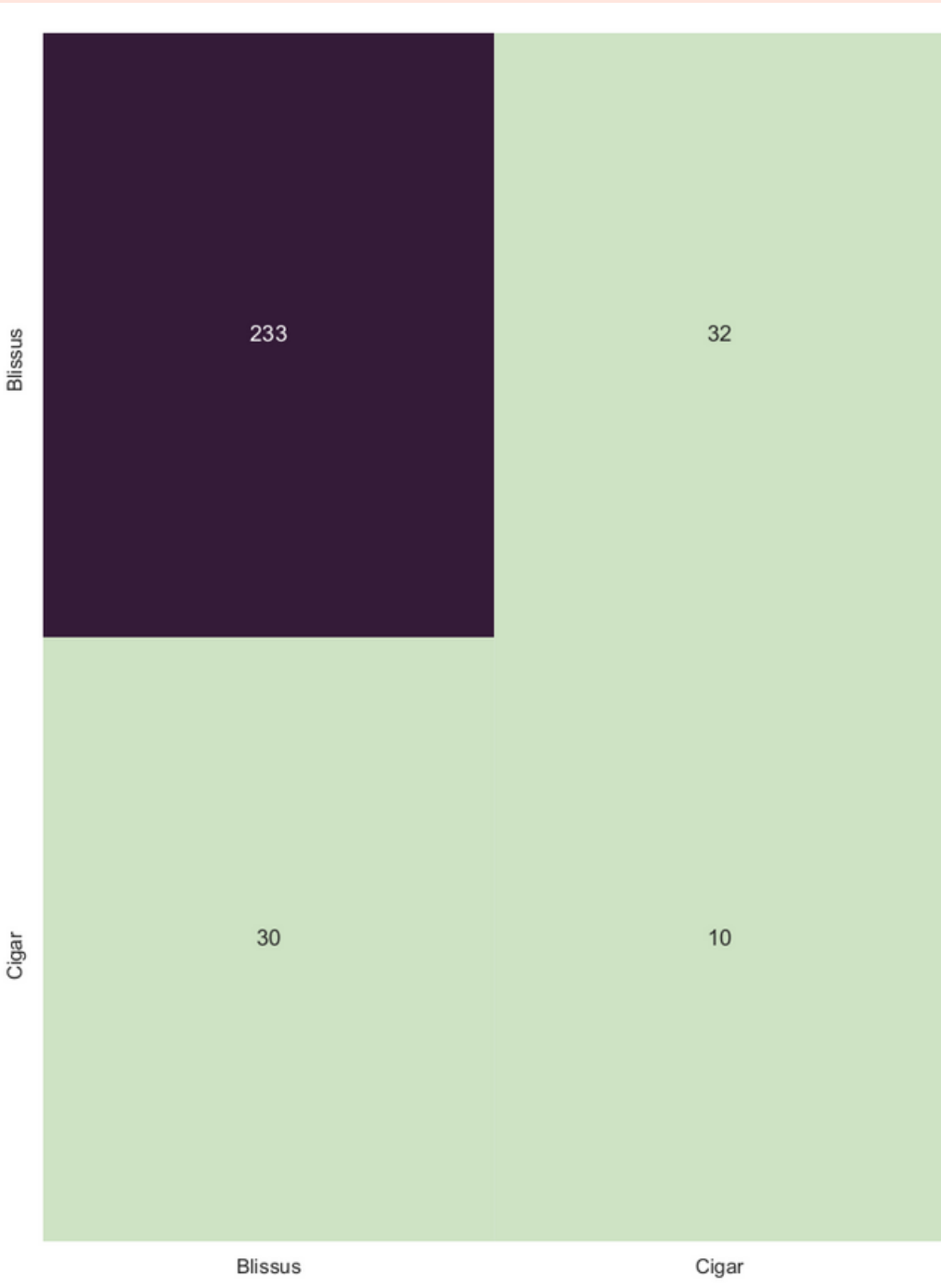
KNN – RESULTADO – VALIDAÇÃO

MATRIZ DE CONFUSÃO

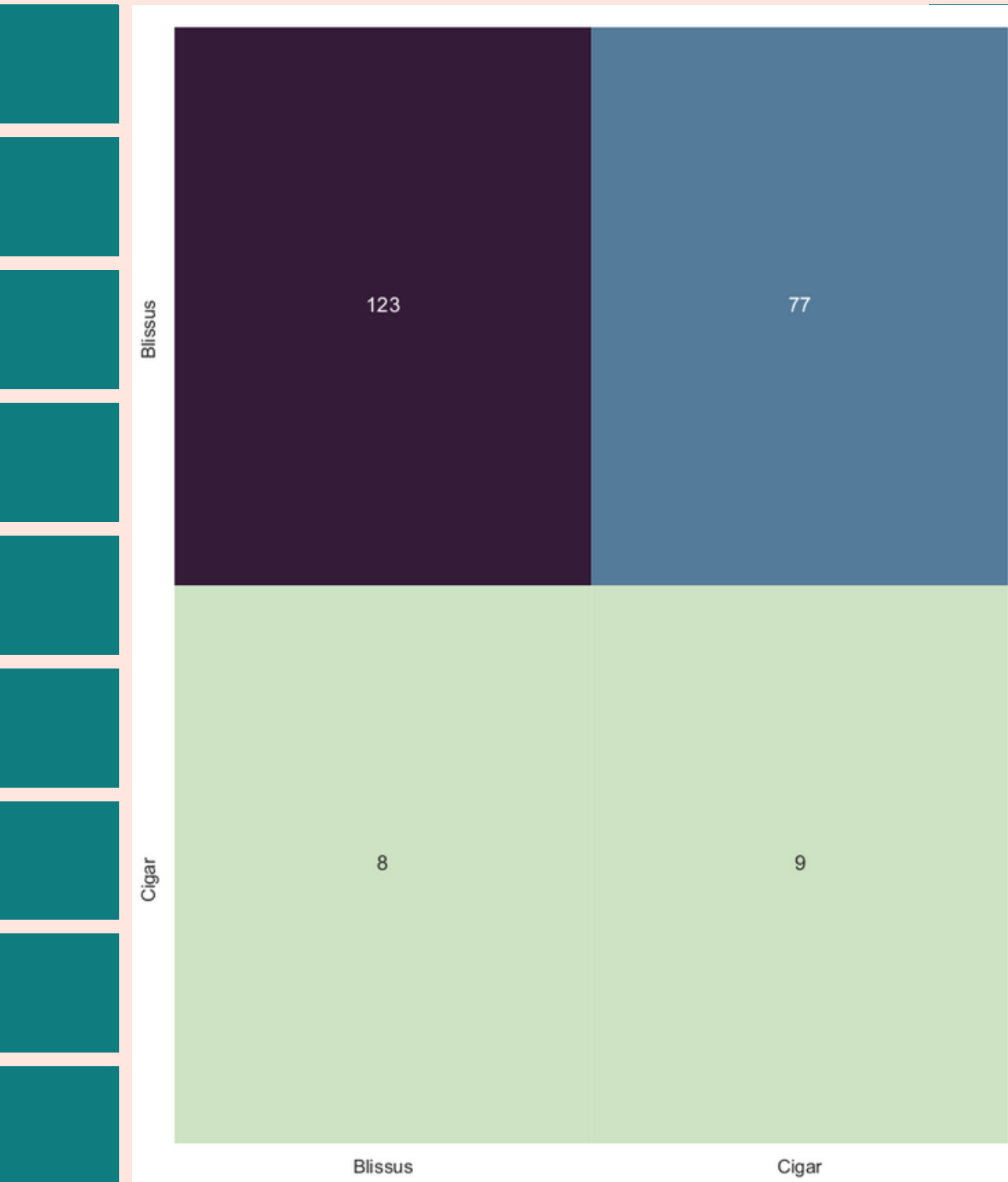
SEM  
SEGMENTAÇÃO



COM  
SEGMENTAÇÃO



COM  
SEGMENTAÇÃO  
SATURAÇÃO -100



# SVC

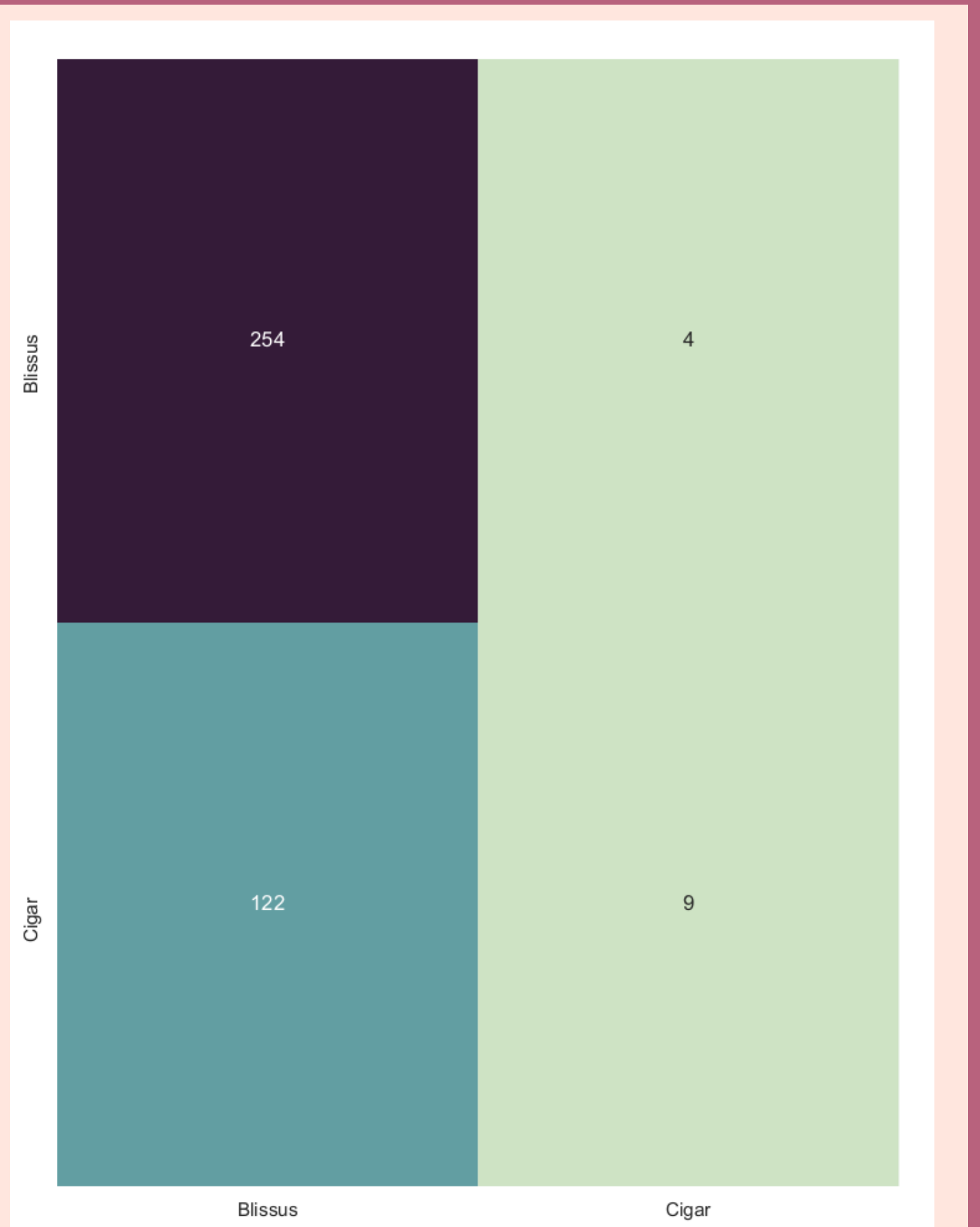
## Dados de Teste

Acurácia : 67,61 %

	Precision	Recall	Num
--	-----------	--------	-----

B.pul	0.68	0.98	258
-------	------	------	-----

Ciga	0.69	0.07	131
------	------	------	-----



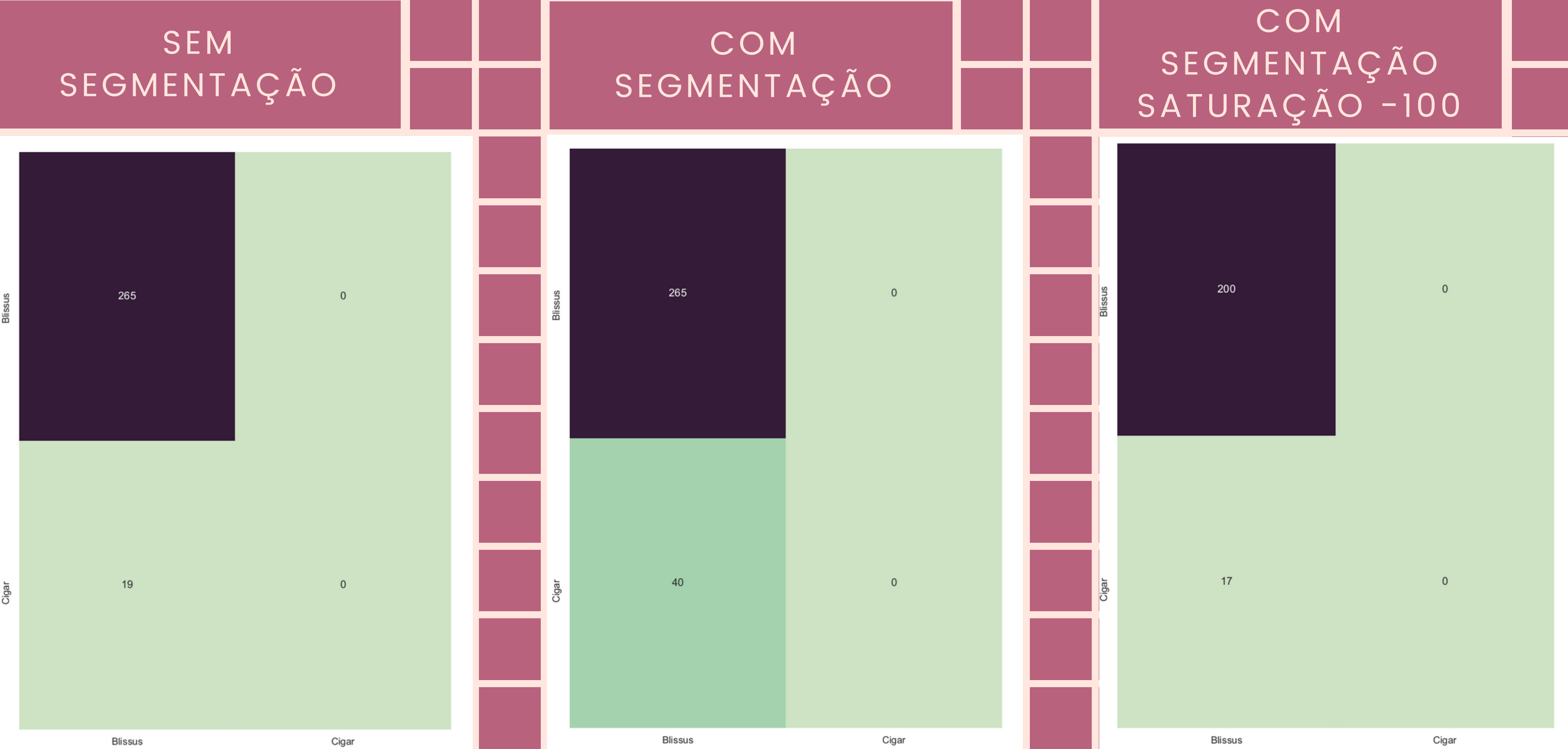


# SVC – RESULTADO – VALIDAÇÃO

	SEM SEGMENTAÇÃO								COM SEGMENTAÇÃO								COM SEGMENTAÇÃO SATURAÇÃO -100							
	Acurácia : 93.31 %								Acurácia : 86,89 %								Acurácia : 92,17 %							
	Precision	Recall	Num						Precision	Recall	Num						Precision	Recall	Num					
	B.pul	0.93	1.00	265					B.pul	0.87	1.00	265					B.pul	0.92	1.00	200				
	Ciga	0.00	0.00	19					Ciga	0.00	0.00	40					Ciga	0.00	0.00	17				

SVC - RESULTADO - VALIDAÇÃO

MATRIZ DE CONFUSÃO

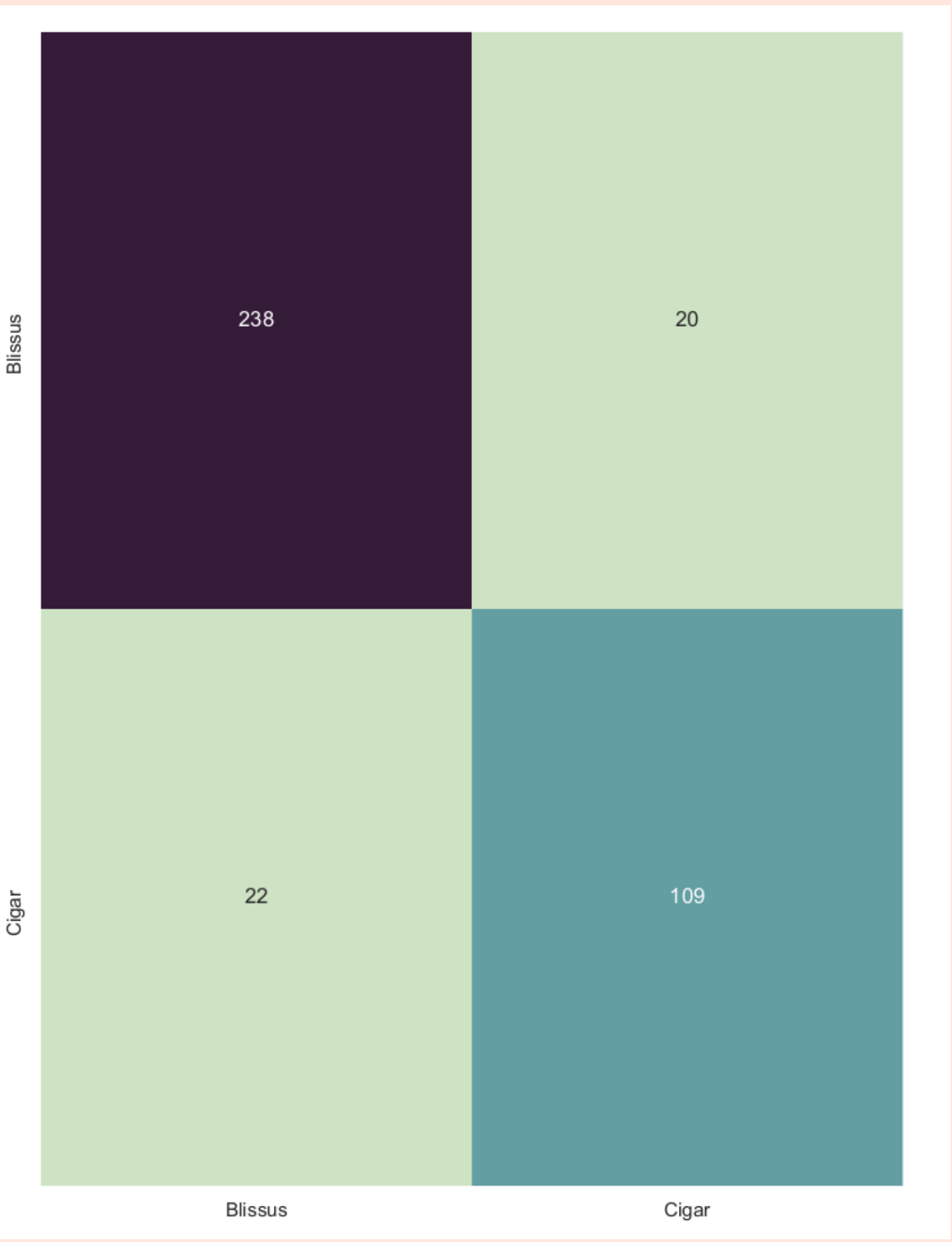


# RDFC

## Dados de Teste

Acurácia : 89,20 %

	Precision	Recall	Num
B.pul	0.92	0.92	258
Ciga	0.84	0.83	131



# RDFC – RESULTADO – VALIDAÇÃO

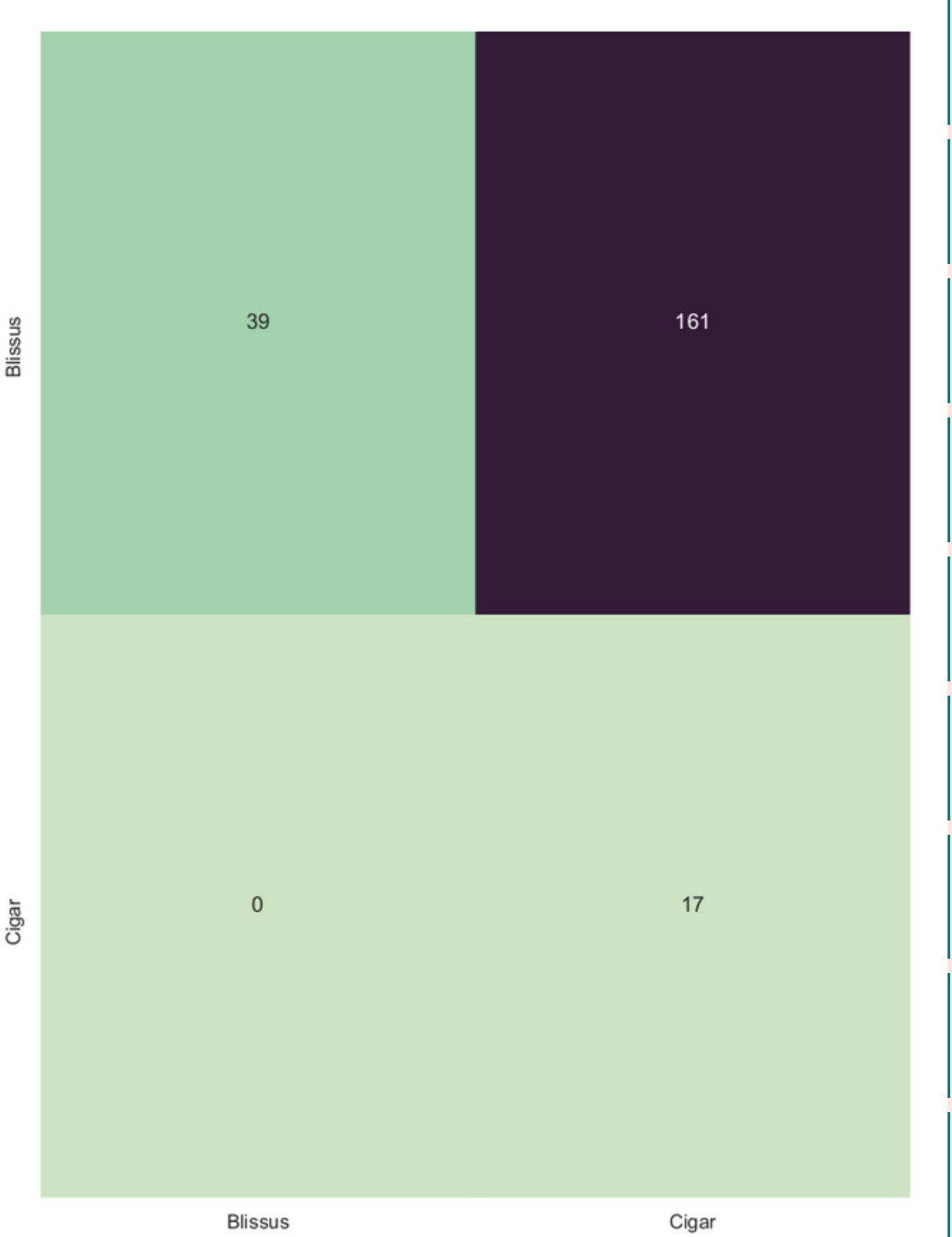
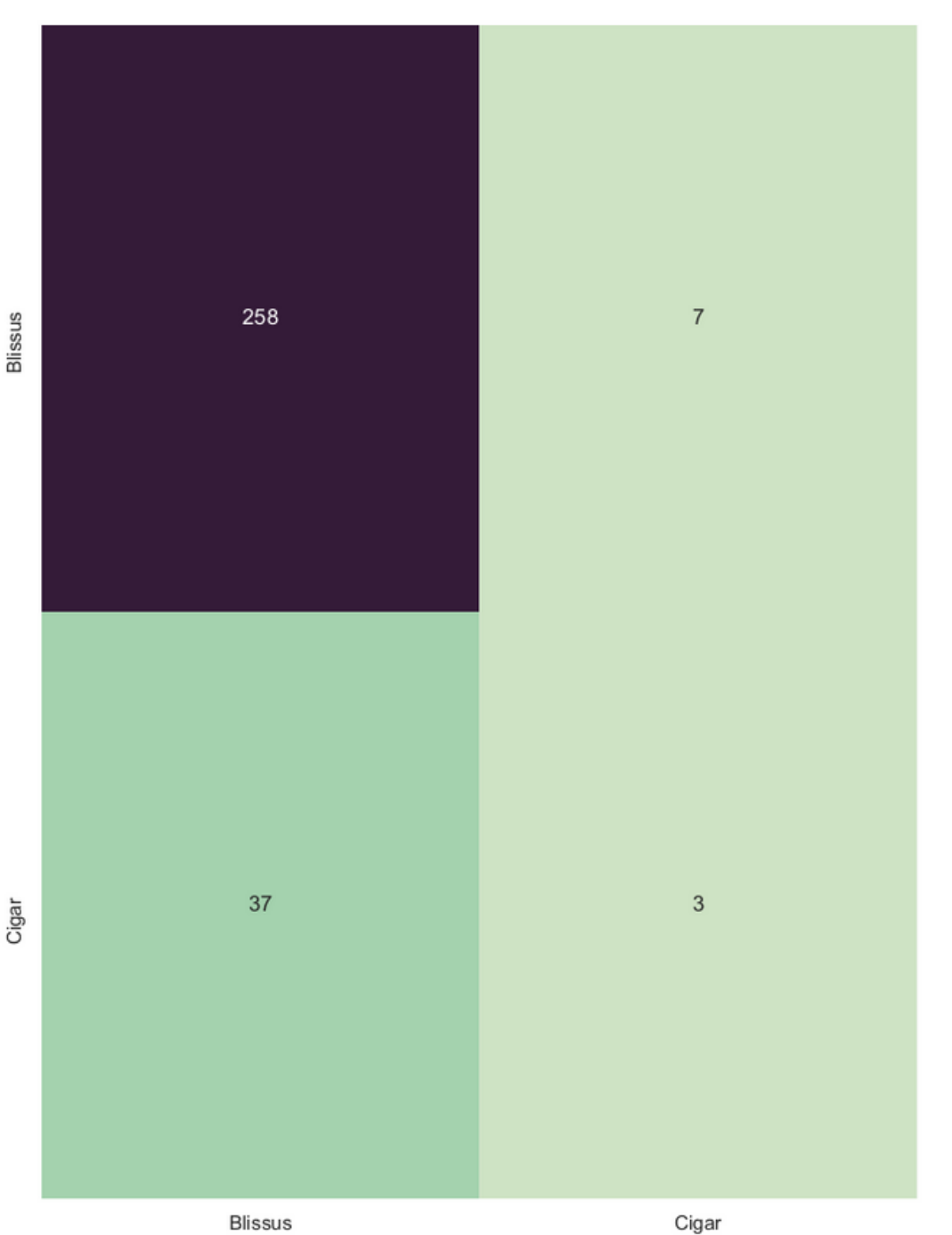
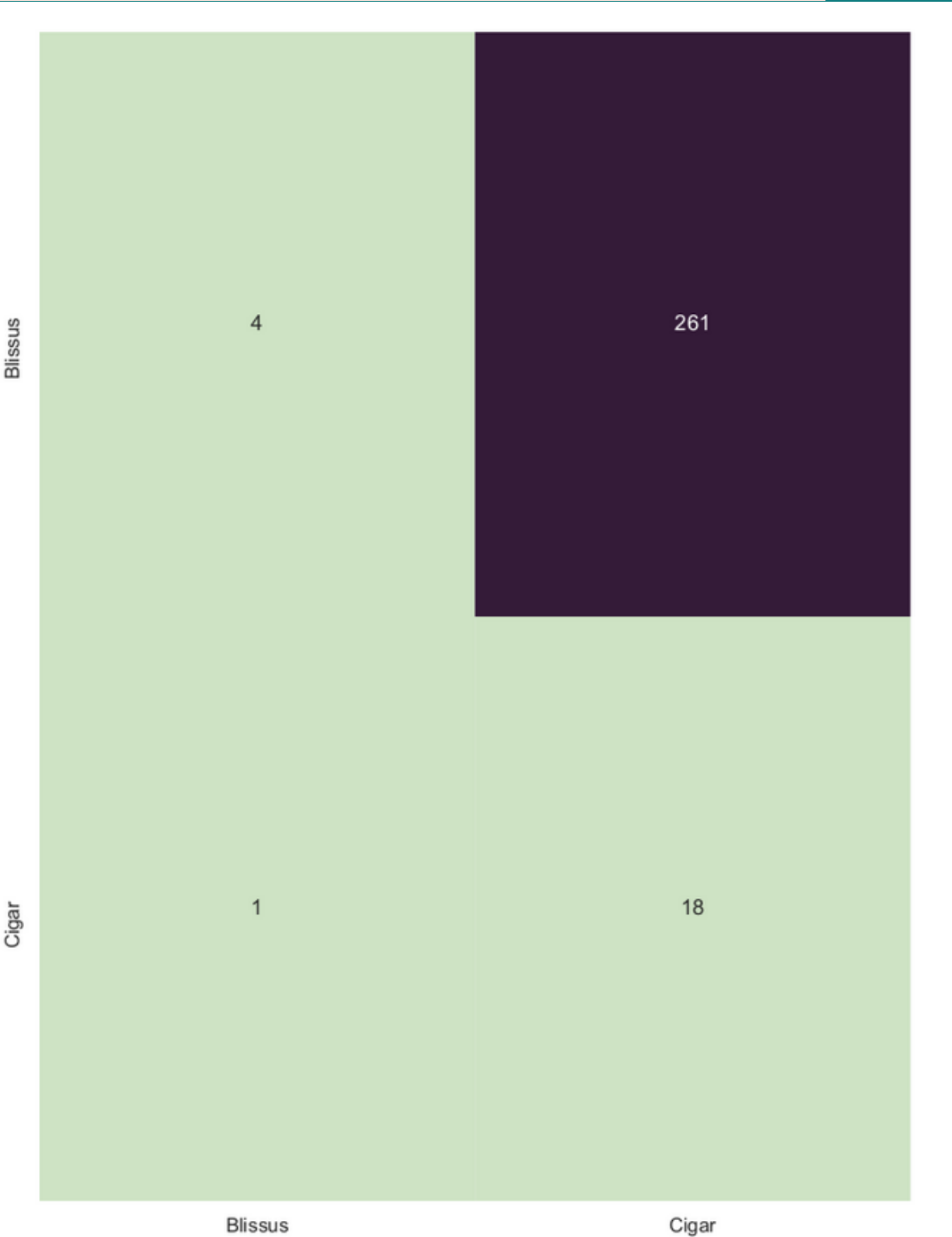
RDFC – RESULTADO – VALIDAÇÃO

MATRIZ DE CONFUSÃO

SEM  
SEGMENTAÇÃO

COM  
SEGMENTAÇÃO

COM  
SEGMENTAÇÃO  
SATURAÇÃO -100

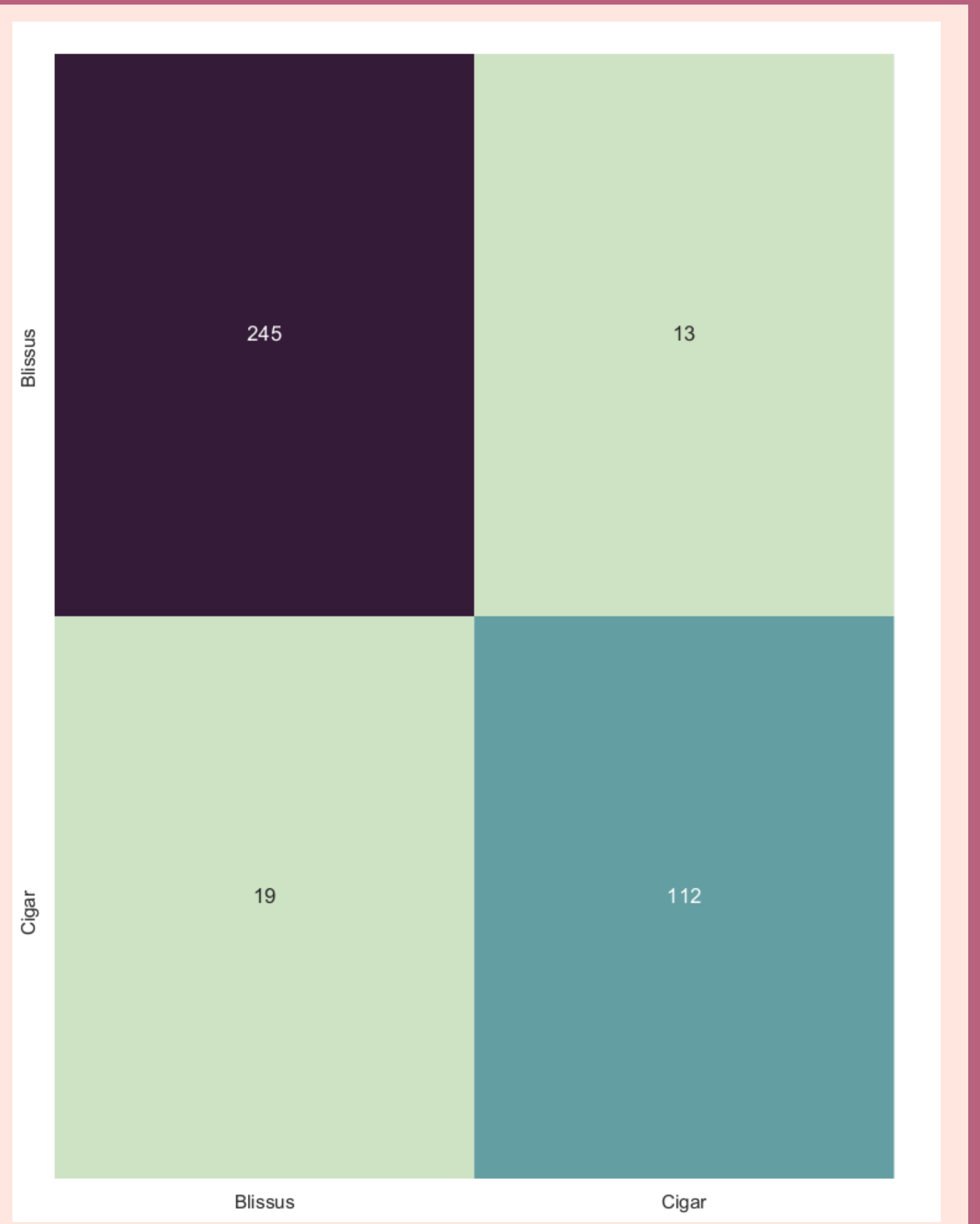


# XGBoost

Dados de Teste

Acurácia : 91,77 %

	Precision	Recall	Num
B.pul	0.93	0.95	258
Ciga	0.90	0.85	131



# XGBoost – RESULTADO – VALIDAÇÃO

	SEM SEGMENTAÇÃO								COM SEGMENTAÇÃO								COM SEGMENTAÇÃO SATURAÇÃO -100								
	Acurácia : 72.54 %								Acurácia : 86,89 %								Acurácia : 32,26 %								
	Precision	Recall	Num					Precision	Recall	Num					Precision	Recall	Num								
	B.pul	0.92	0.78	265					B.pul	0.87	1.00	265					B.pul	0.98	0.27	200					
	Ciga	0.03	0.00	19					Ciga	0.00	0.00	40					Ciga	0.10	0.94	17					

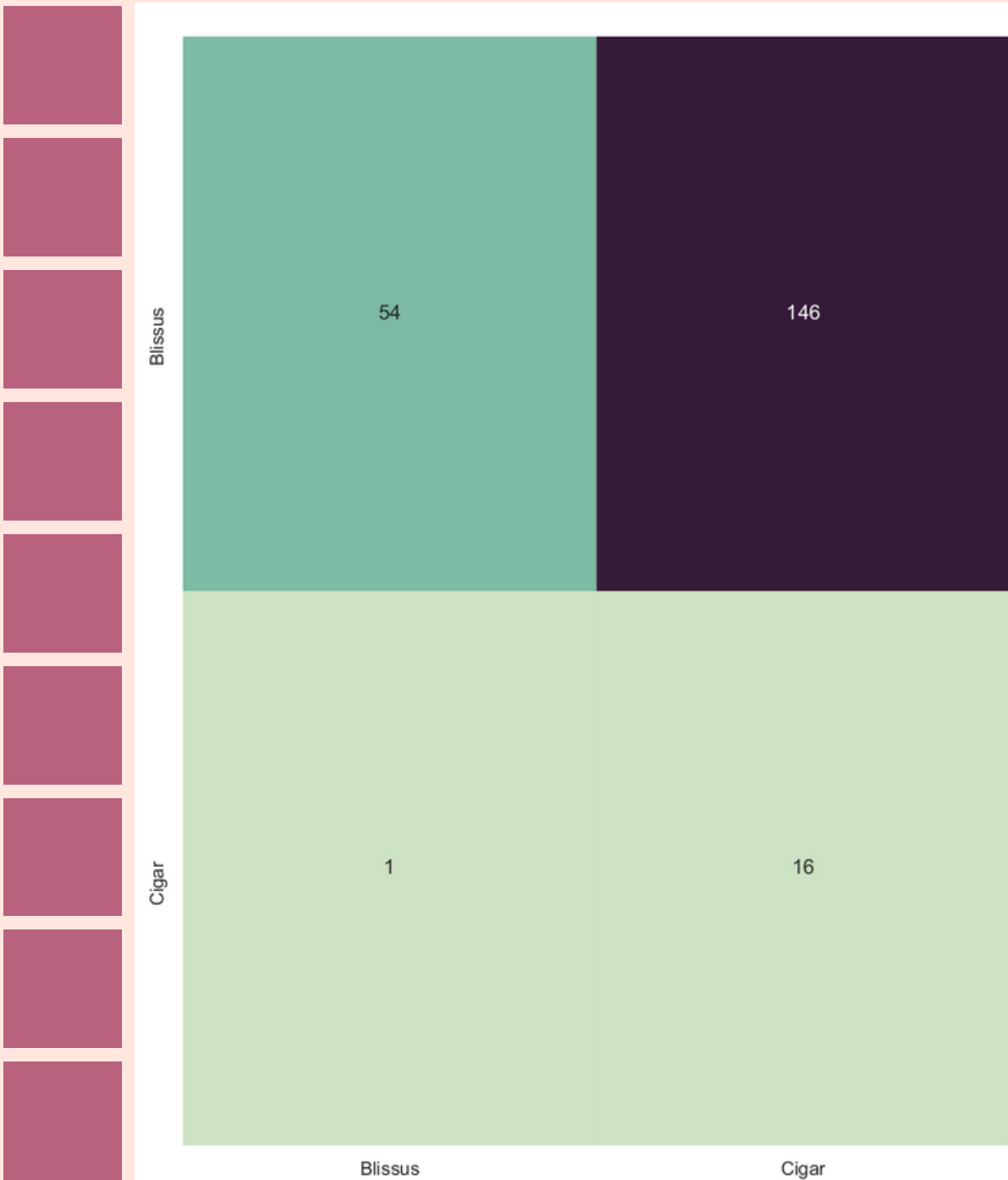
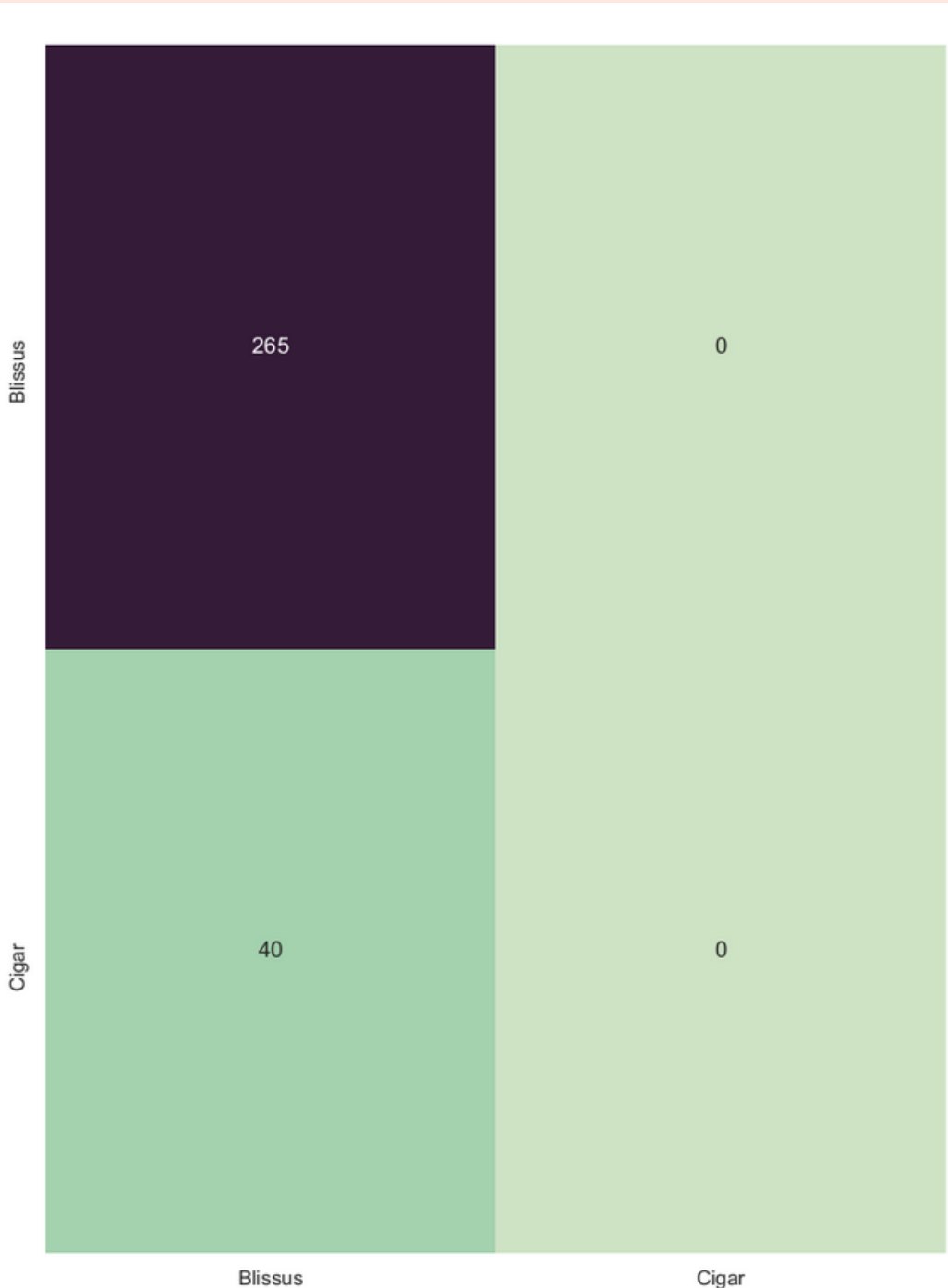
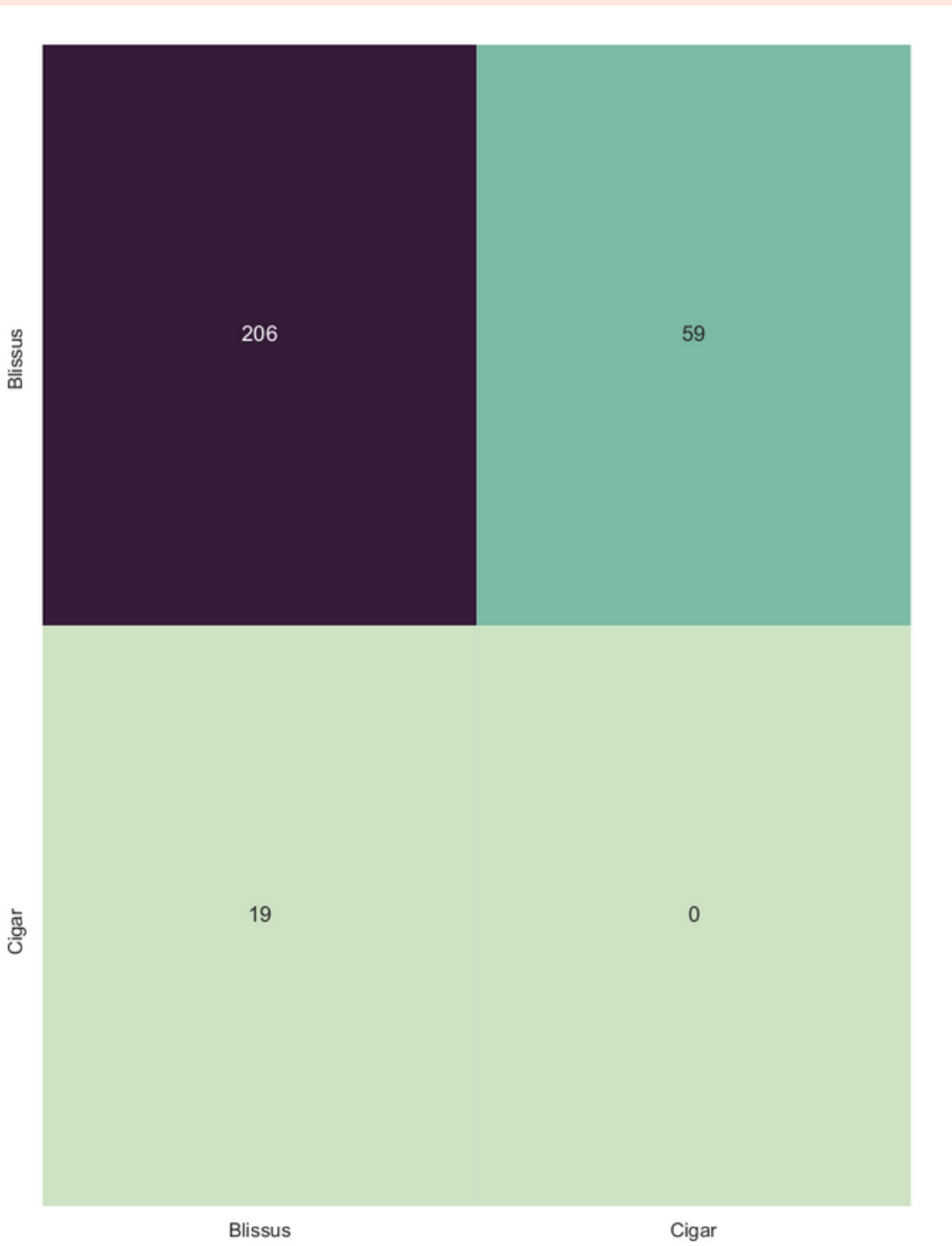
XGBoost – RESULTADO – VALIDAÇÃO

MATRIZ DE CONFUSÃO

SEM  
SEGMENTAÇÃO

COM  
SEGMENTAÇÃO

COM  
SEGMENTAÇÃO  
SATURAÇÃO -100





**CONCLUSÃO SOBRE OS  
DADOS DE TESTES**

**CONCLUSÃO SOBRE OS  
DADOS DE VALIDAÇÃO**

# Dados de Testes

- Sobre um total de 258 Elementos de Blissus e 131 Elementos da Cigarrinha
- O KNN (70,95%) deu 211 Blis e 65 Cig
- O SVC (67,61%) deu 254 Blis e 9 Cig
- O RDFC (89,20%) deu 238 Blis e 109 Cig
- O XGBoost (91,77%) deu 245 Blis e 112 Cig

# Dados de Testes

- Dos resultados acima podemos afirmar que os Algoritmos XGboost e RDFC apresentarão uma assertividade muito boa e acertável a respeito da nossa pesquisa

# Dados de Validação

SEM  
SEGMENTAÇÃO

- Sobre um total de 265 Elementos de Blissus e 19 Elementos da Cigarrinha
- O KNN (76,06%) deu 210 Blis e 6 Cig
- O SVC (93,31%) deu 265 Blis e 0 Cig
- O RDFC (7,75%) deu 4 Blis e 18 Cig
- O XGBoost (72,54%) deu 206 Blis e 0 Cig

# Dados de Validação

SEM  
SEGMENTAÇÃO

- Dos Resultados das imagens de validação sem pré-processamento, tivermos uma performance de acurácia bastante positiva no que diz respeito a identificação do Blissus, mas uma pouca identificação ou quase nenhuma da cigarrinha.

# Dados de Validação

COM  
SEGMENTAÇÃO

- Sobre um total de 265 Elementos de Blissus e 40 Elementos da Cigarrinha
- O KNN (79,67%) deu 233 Blis e 10 Cig
- O SVC (86,89%) deu 265 Blis e 0 Cig
- O RDFC (85,57%) deu 258 Blis e 3 Cig
- O XGBoost (86,89%) deu 265 Blis e 0 Cig

# Dados de Validação

COM  
SEGMENTAÇÃO

- Dos resultados das imagens de validação com segmentação, todos os algoritmos apresentaram uma boa acurácia, e uma identificação significativa do Blissus. No que diz respeito a cigarrinha poderá se considerar o KNN como, o algoritmo que apresentou um TradeOff aceitável na identificação.

# Dados de Validação

COM  
SEGMENTAÇÃO  
SATURAÇÃO -100

- Sobre um total de 200 Elementos de Blissus e 17 Elementos da Cigarrinha
- O KNN (60,83%) deu 123 Blis e 9 Cig
- O SVC (92,17%) deu 200 Blis e 0 Cig
- O RDFC (25,81%) deu 39 Blis e 17 Cig
- O XGBoost (32,26%) deu 54 Blis e 16 Cig



# Dados de Validação

COM  
SEGMENTAÇÃO  
SATURAÇÃO -100

- Dos resultados das imagens de validação com segmentação e saturação -100, os algoritmos de KNN e de SVC apresentaram uma acurácia maior do que 50%, mas percebermos que no SVC nenhuma Cigarrinha foi reconhecida . Isso leva a considerar o KNN como o algoritmo que poderá apresentar uma assertividade compartilhada na identificação dos insetos.

# CONCLUSÃO

- Apesar de apresentar uma acurácia muito boa, os algoritmos tiveram algumas dificuldades a realizar a classificação com os dados de validação
- Uma explicação destes resultados, poderá a forma pelo qual se apresenta as imagens de validação (pré-processado ou não)
- De modo geral, percebermos que as imagens de validação segmentadas apresentaram com os algoritmos resultados melhores que as imagens de validação bruta e inclusive as imagens segmentadas com efeito de saturação – 100 como os dados de treinamento e de Teste