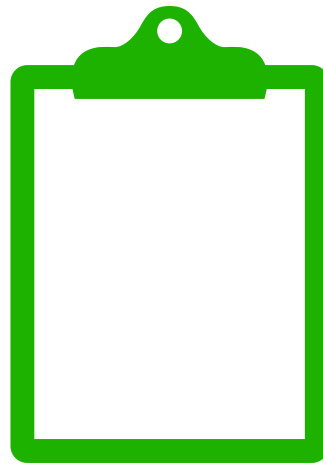


Bank Marketing Data Set

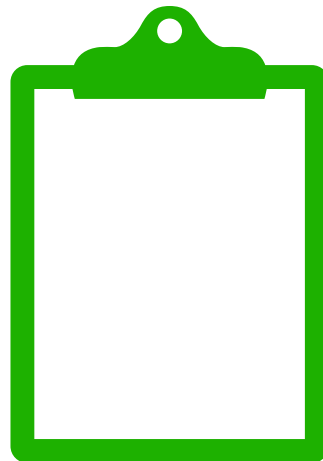
RIYADH SILVA | JEFFERSON PIRES



SOBRE O DATASET

BANK MARKETING DATA SET

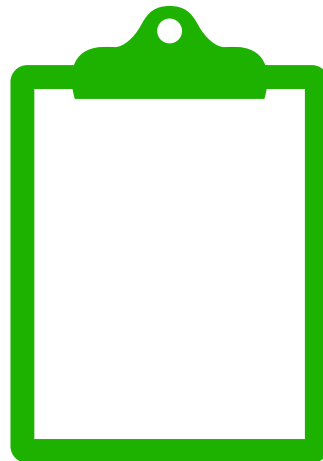
CAMPANHAS DE MARKETING DIRETA DE UMA INSTITUIÇÃO BANCÁRIA PORTUGUESA



SOBRE O DATASET

AS CAMPANHAS DE MARKETING FORAM BASEADAS EM CHAMADAS TELEFÔNICAS

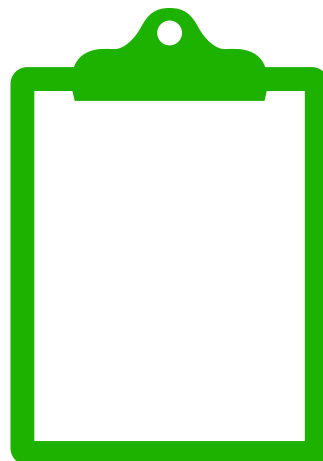
**MUITAS VEZES, ERA NECESSÁRIO CONTATAR MAIS DE UMA VEZ O MESMO CLIENTE, A FIM DE
ACESSAR SE O PRODUTO (DEPÓSITO A PRAZO BANCÁRIO) SERIA ASSINADO ('SIM') OU NÃO ('NÃO')**



SOBRE O DATASET

45211 INSTÂNCIAS (LINHAS)

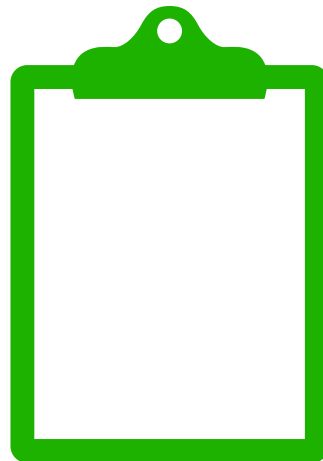
20 ATRIBUTOS (COLUNAS)



SOBRE O DATASET

COLUNAS DE 1 A 7 - REFEREM-SE AOS DADOS BANCÁRIOS DO CLIENTE

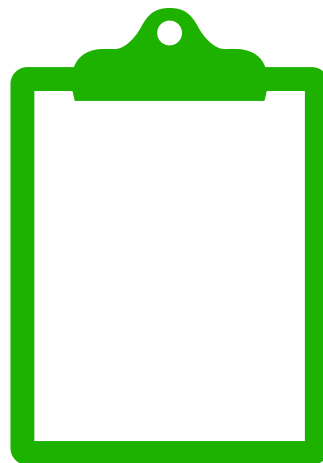
COLUNAS DE 8 A 11 - REFEREM-SE AO ÚLTIMO CONTATO DA CAMPANHA ATUAL



SOBRE O DATASET

COLUNAS DE 12 A 15 - REFEREM-SE AS CAMPANHAS

COLUNAS DE 16 A 20 - REFEREM-SE AOS ATRIBUTOS DE CONTEXTO SOCIAL E ECONÔMICO

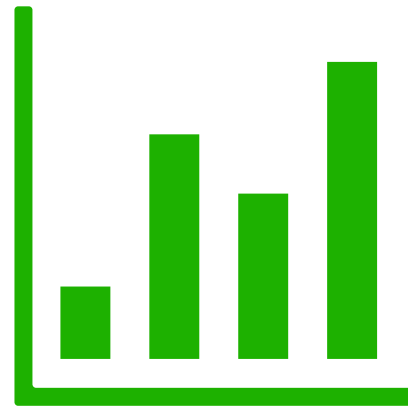


SOBRE O DATASET

COLUNAS 21 - REFERE-SE A VARIÁVEL DE SAÍDA (ALVO DESEJADO)



**O CLIENTE ASSINOU UM DEPÓSITO A PRAZO ?
(BINÁRIO: 'SIM', 'NÃO')**

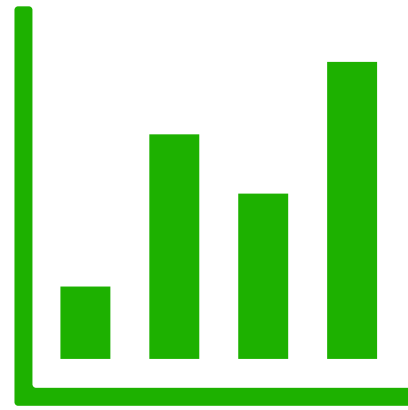


ANÁLISE DOS DADOS

5 ATRIBUTOS FLOAT64

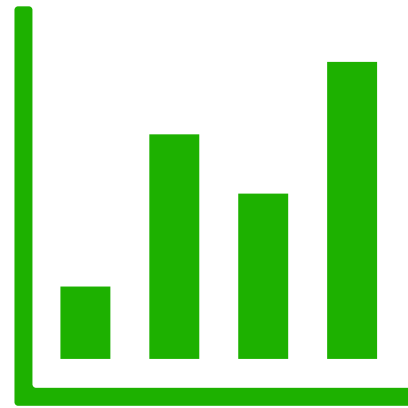
5 ATRIBUTOS INT64

11 ATRIBUTOS OBJECT



ANÁLISE DOS DADOS

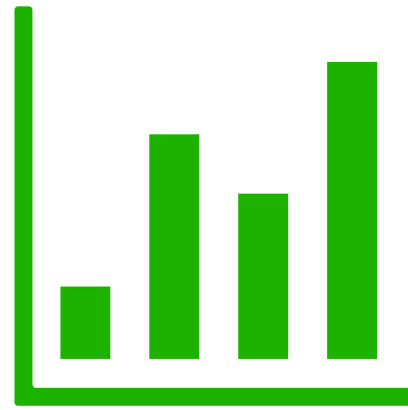
NÃO EXISTEM CAMPOS NULOS



ANÁLISE DOS DADOS

EXISTEM 10 COLUNAS COM VARIÁVEIS CATEGÓRICAS

**JOB | MARITAL | EDUCATION | DEFAULT | HOUSING
LOAN | CONTACT | MONTH | DAY_OF_WEEK | POUTCOME**



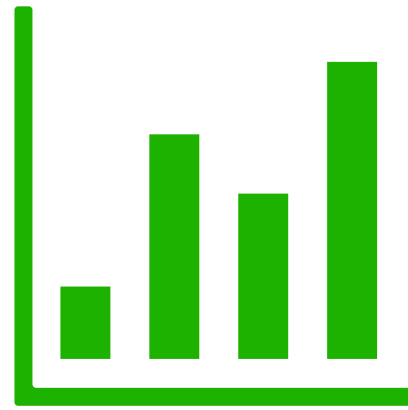
ANÁLISE DOS DADOS

NECESSÁRIO CRIAR NOVAS COLUNAS EM VIRTUDE DAS VARIÁVEIS CATEGÓRICAS

ANTES DESSE PROCESSO TEMOS 21 COLUNAS

43 NOVAS COLUNAS FORAM CRIADAS

APÓS O PROCESSO: 64 COLUNAS

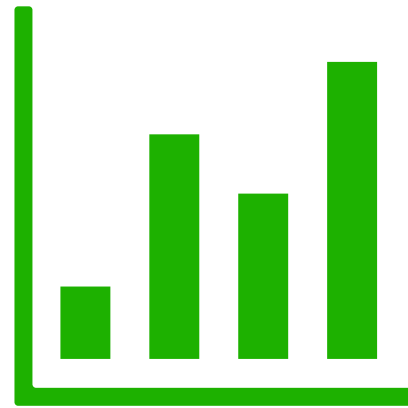


ANÁLISE DOS DADOS

CAMPOS DESCARTADOS COM JUSTIFICATIVA

EURIBOR3M, CONS.CONF.IDX, EMP.VAR.RATE

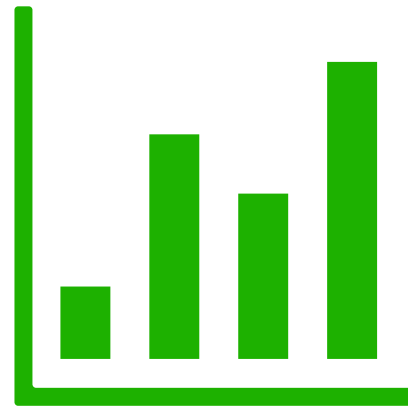
**O ALGORITMO MULTINOMIAL NÃO FUNCIONA COM VALORES NEGATIVOS BEM COMO AS
MATRIZES DA QUESTÃO 4**



ANÁLISE DOS DADOS

CONVERSÕES ENTRE TIPOS DE COLUNAS

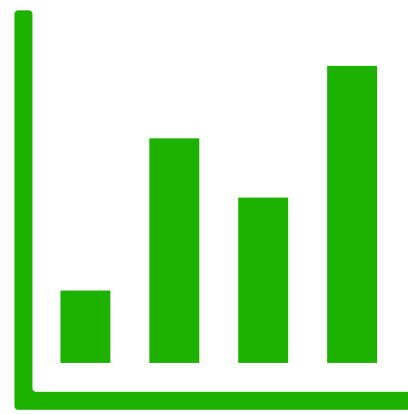
A ÚLTIMA COLUNA APRESENTAVA VALORES BINÁRIOS 'YES' E 'NO', FOI NECESSÁRIO A CONVERSÃO DESSES VALORES PARA '1' E '0' RESPECTIVAMENTE (PRESENTE NA LETRA D)



ANÁLISE DOS DADOS

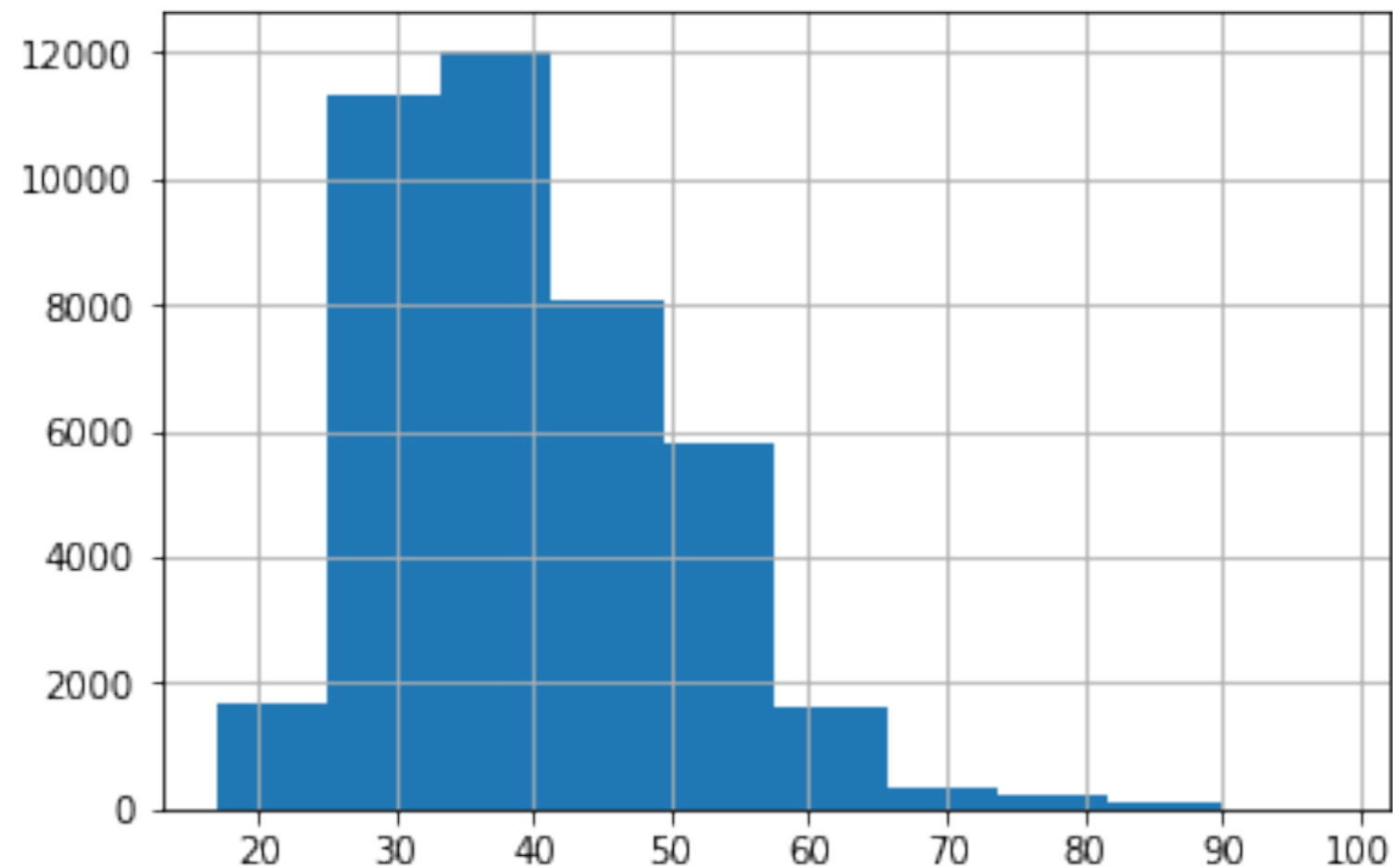
3 CARACTERÍSTICAS RELEVANTES DO DATASET

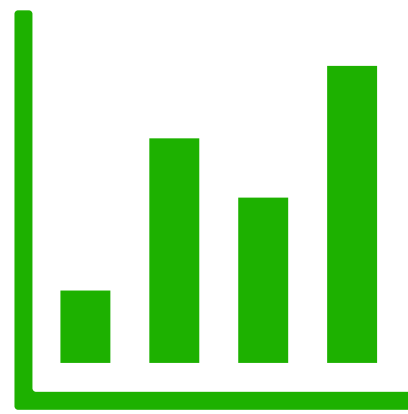
A IDADE DO PÚBLICO ALVO; DISTRIBUIÇÃO DAS ASSINATURAS; A CORRELAÇÃO ENTRE AS ASSINATURAS E AS DEMAIS COLUNAS



ANÁLISE DOS DADOS

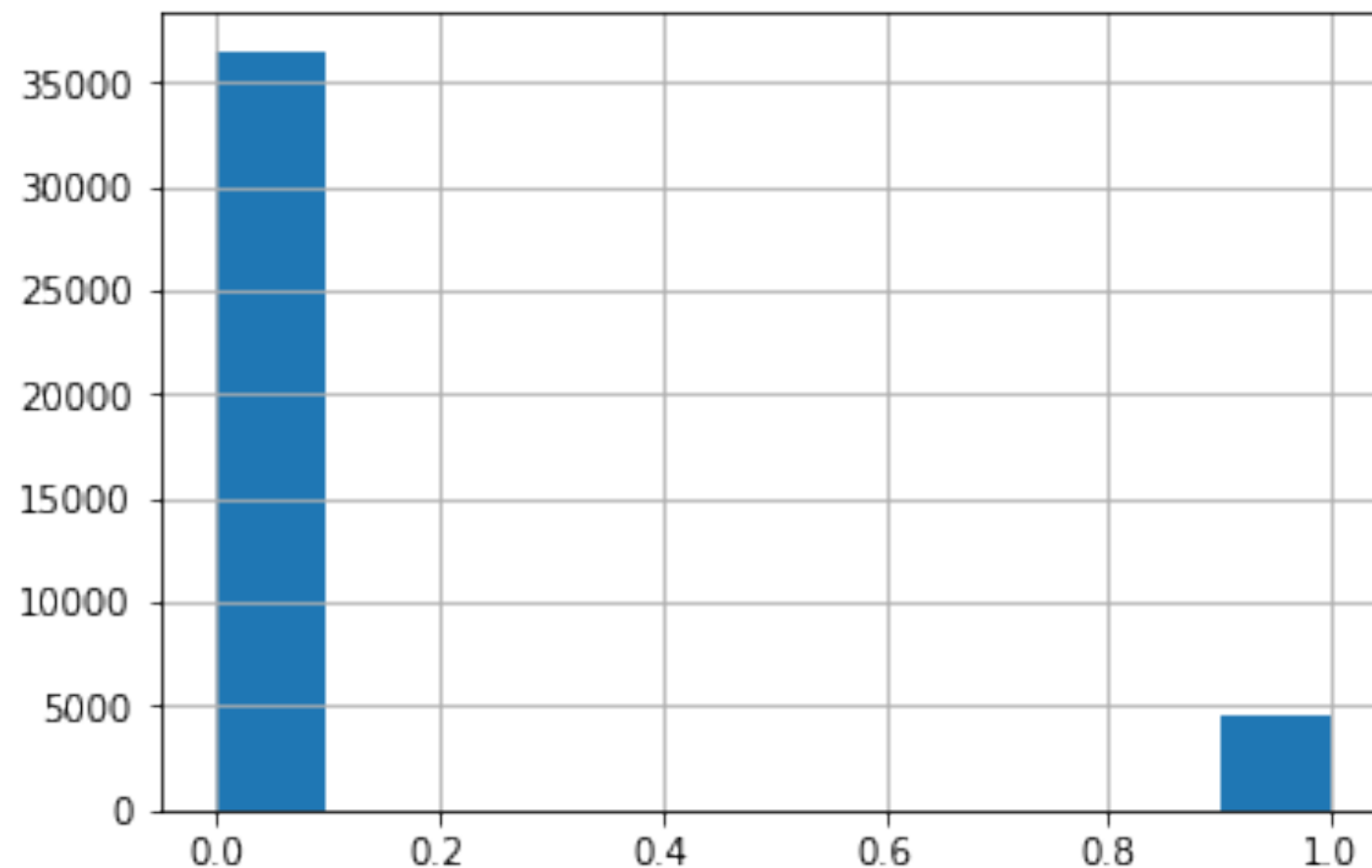
A IDADE DO PÚBLICO ALVO

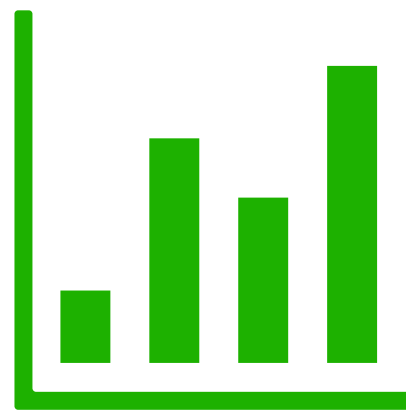




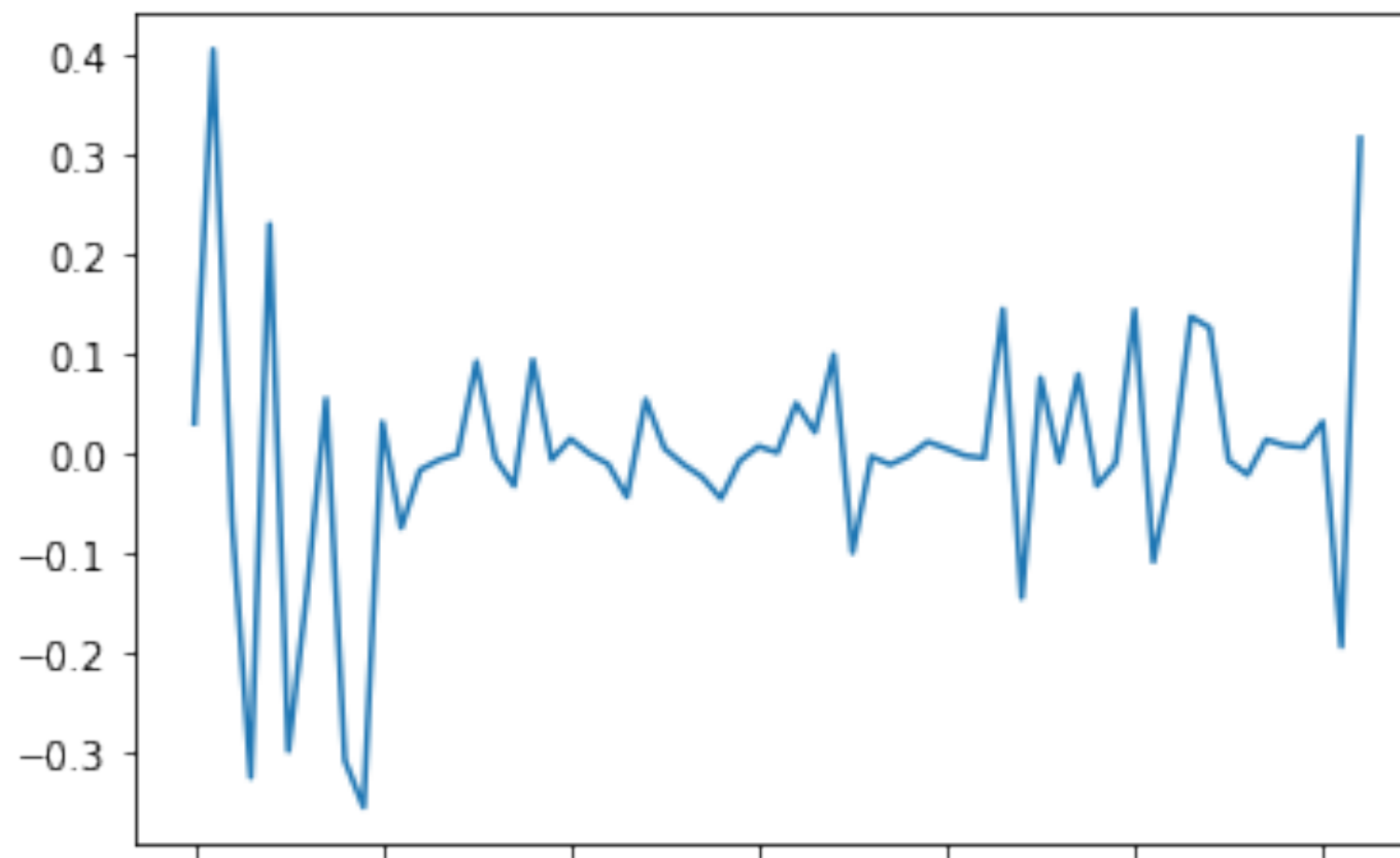
ANÁLISE DOS DADOS

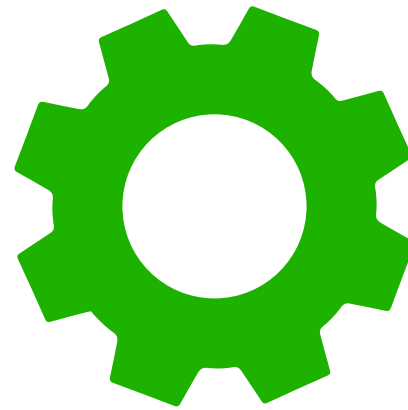
DISTRIBUIÇÃO DAS ASSINATURAS



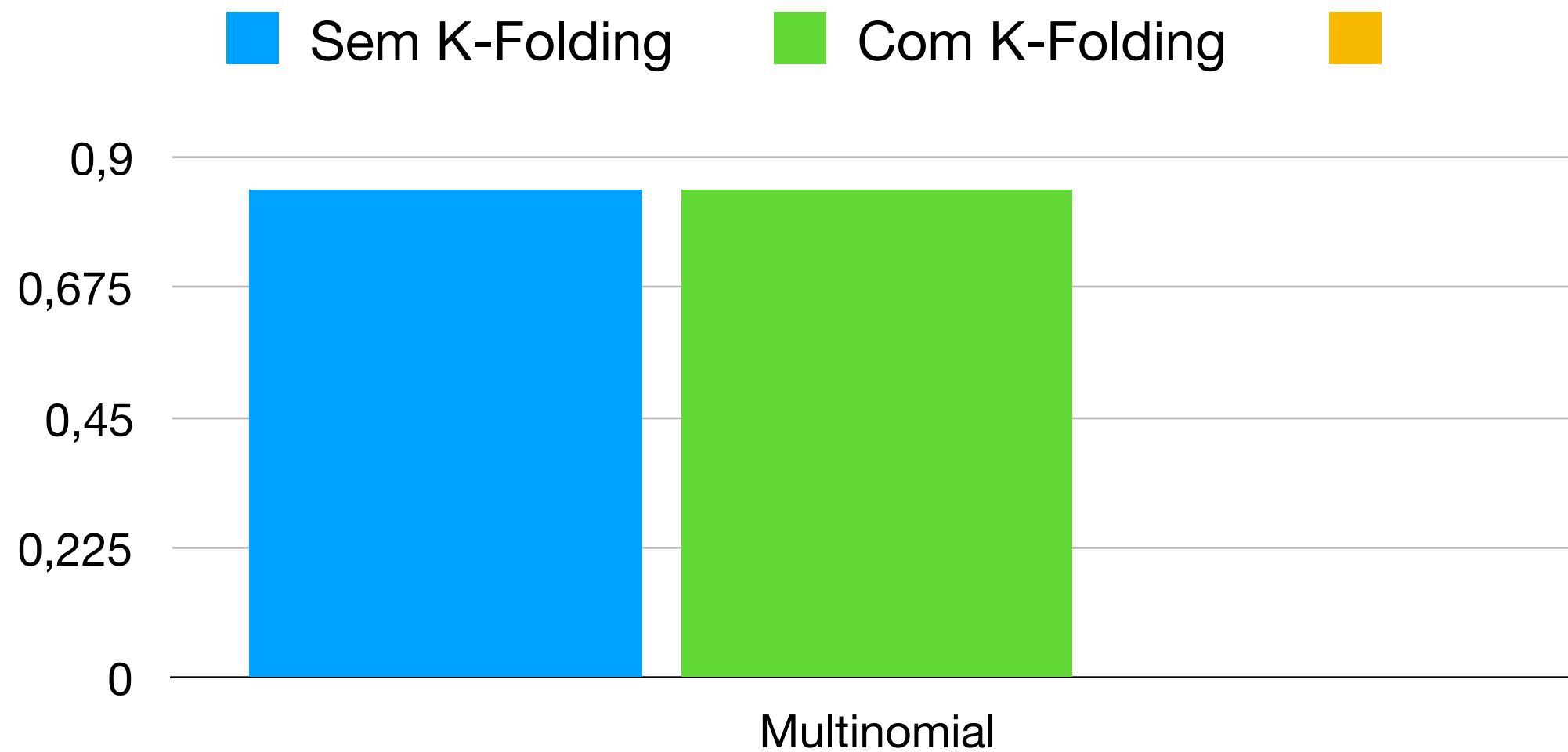


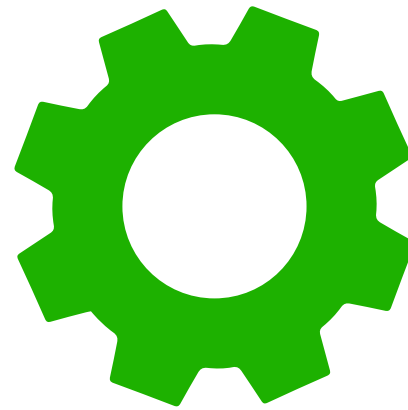
CORRELAÇÃO DAS ASSINATURAS COM AS DEMAIS COLUNAS



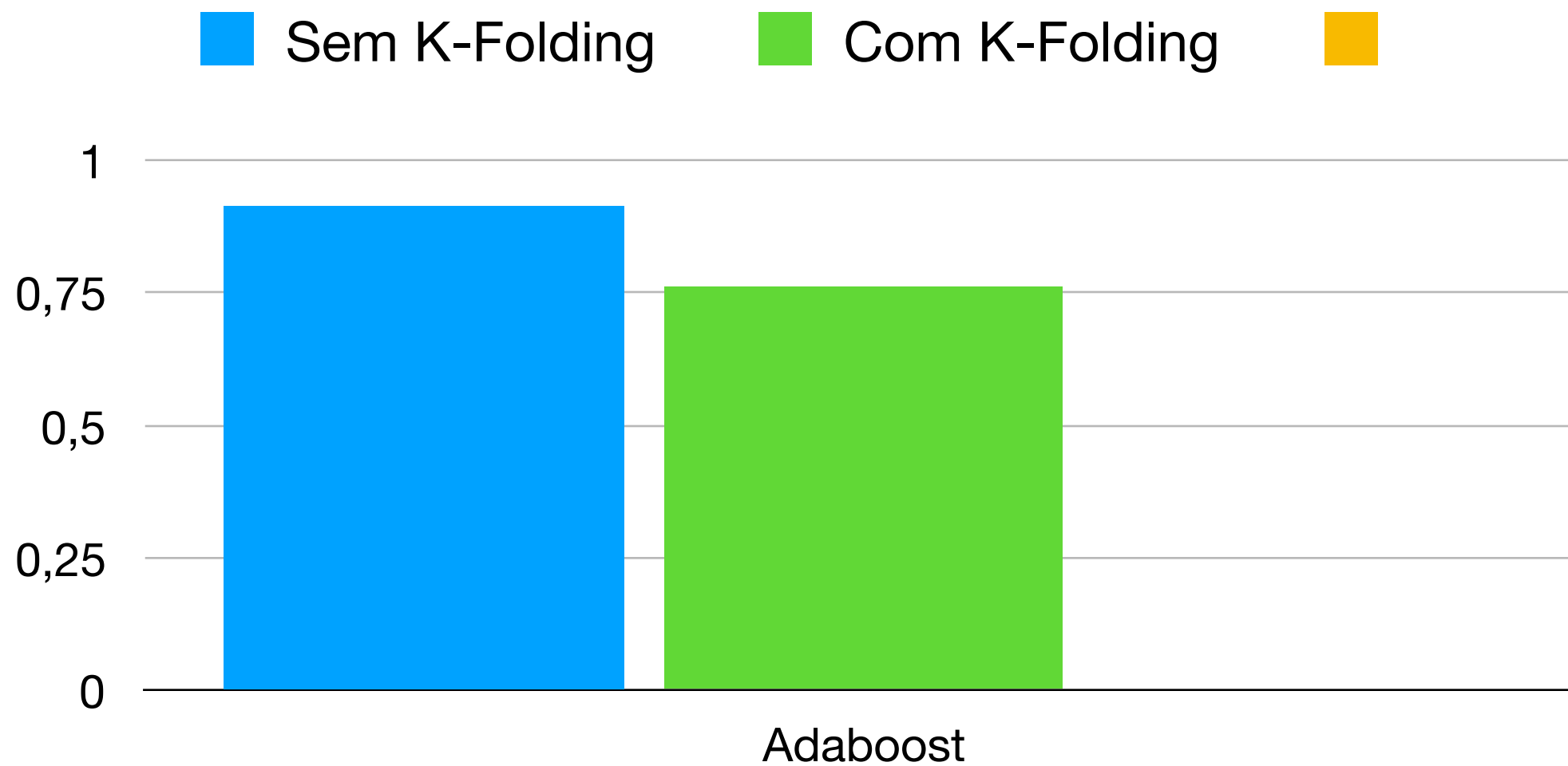


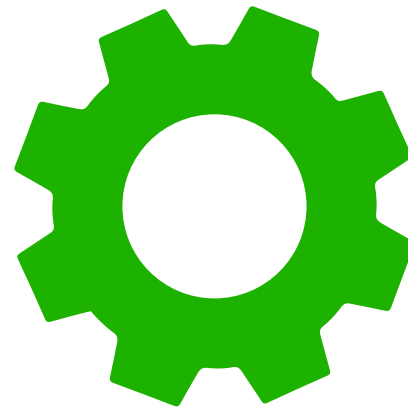
EXECUÇÃO DO ALGORITMO



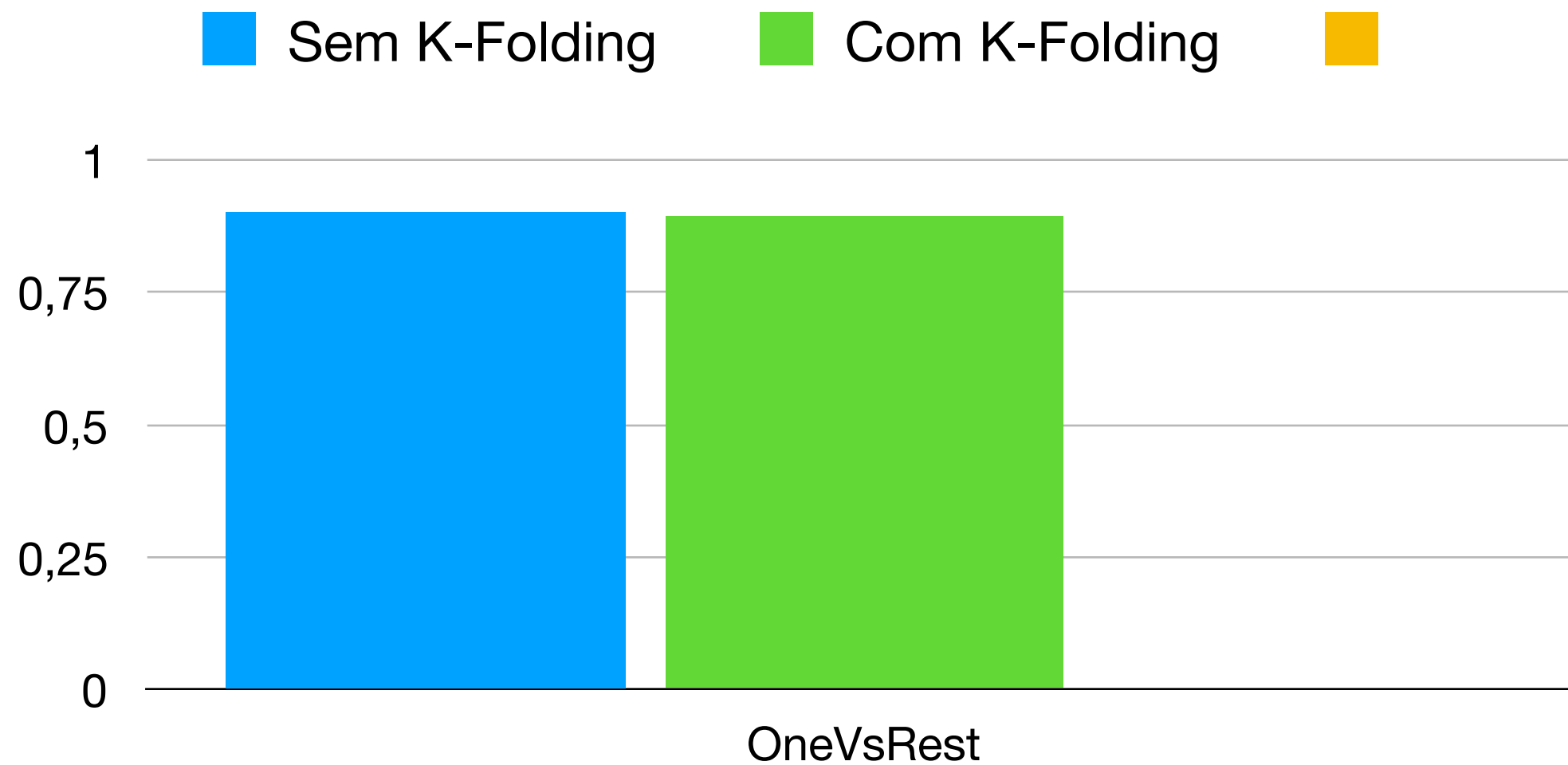


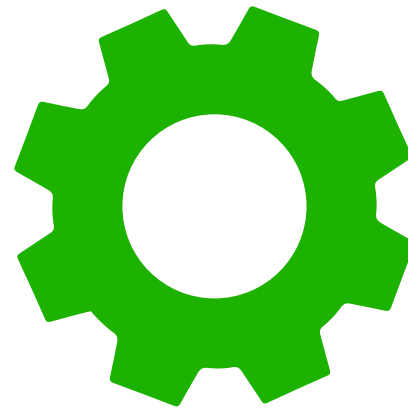
EXECUÇÃO DO ALGORITMO



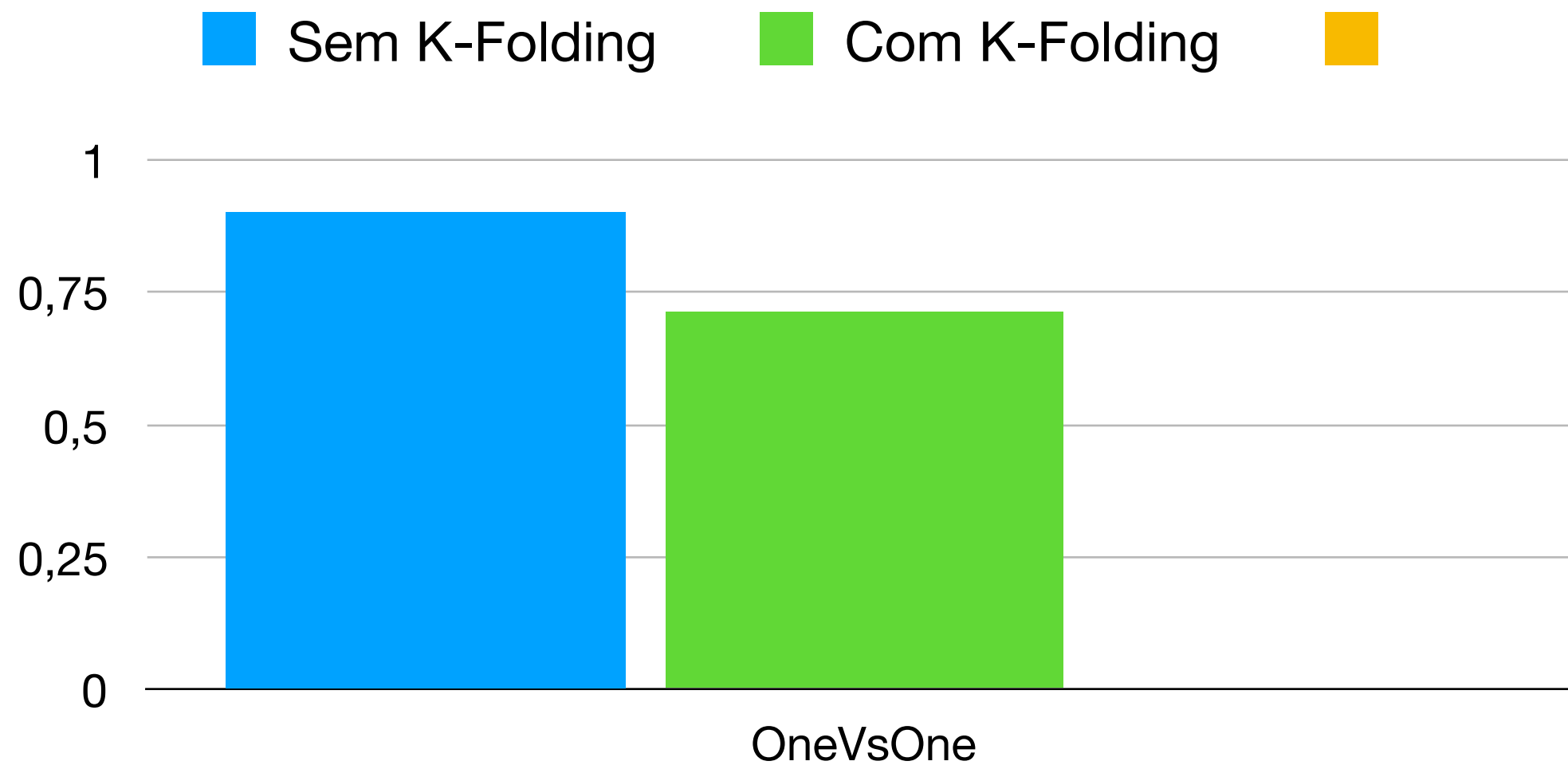


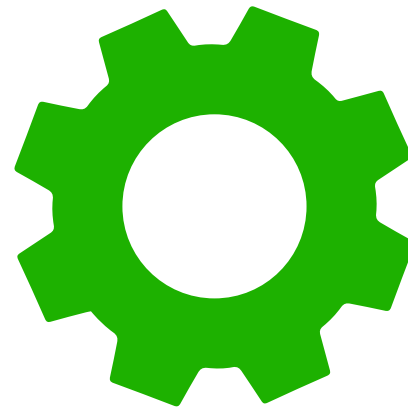
EXECUÇÃO DO ALGORITMO



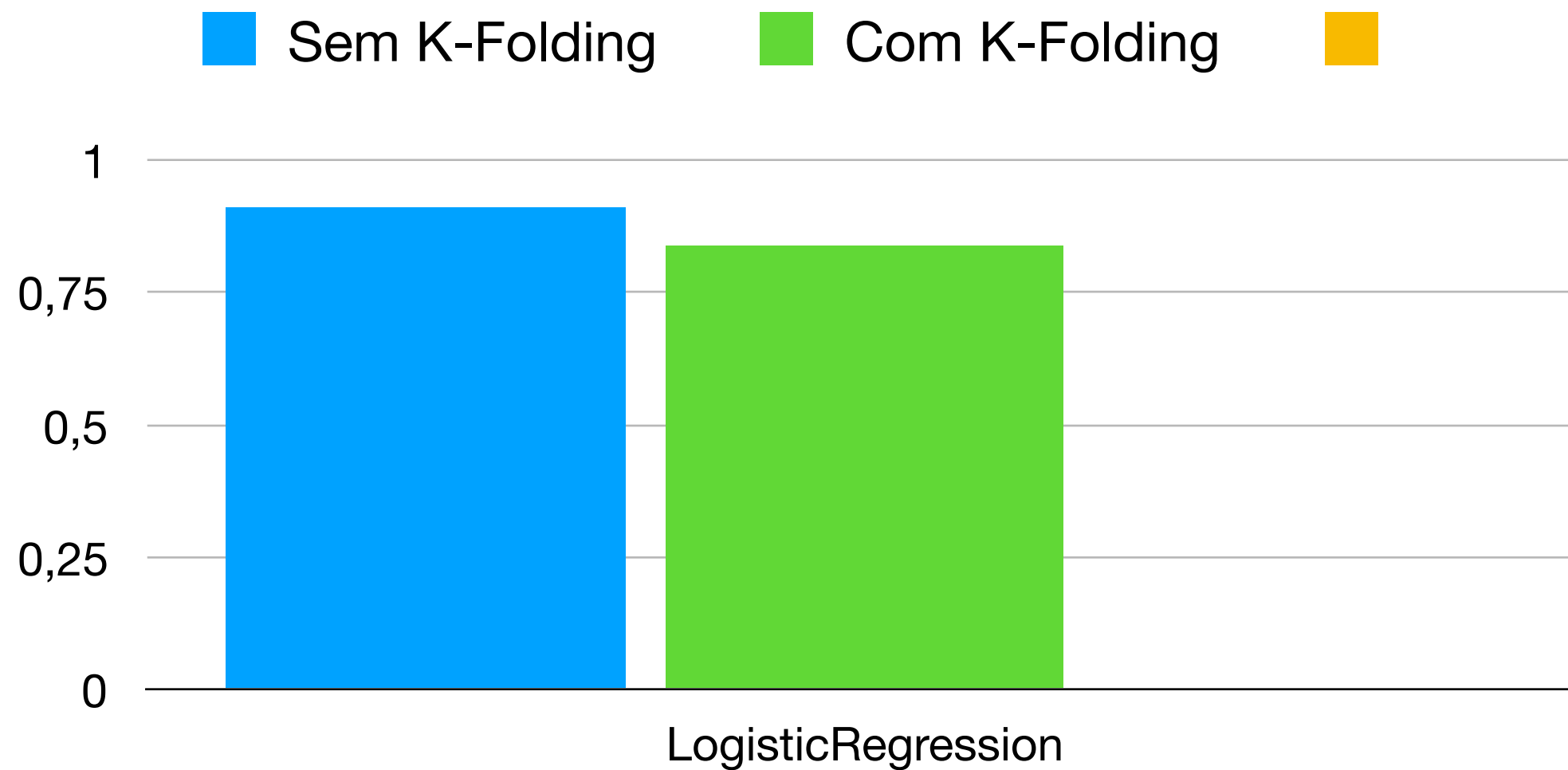


EXECUÇÃO DO ALGORITMO

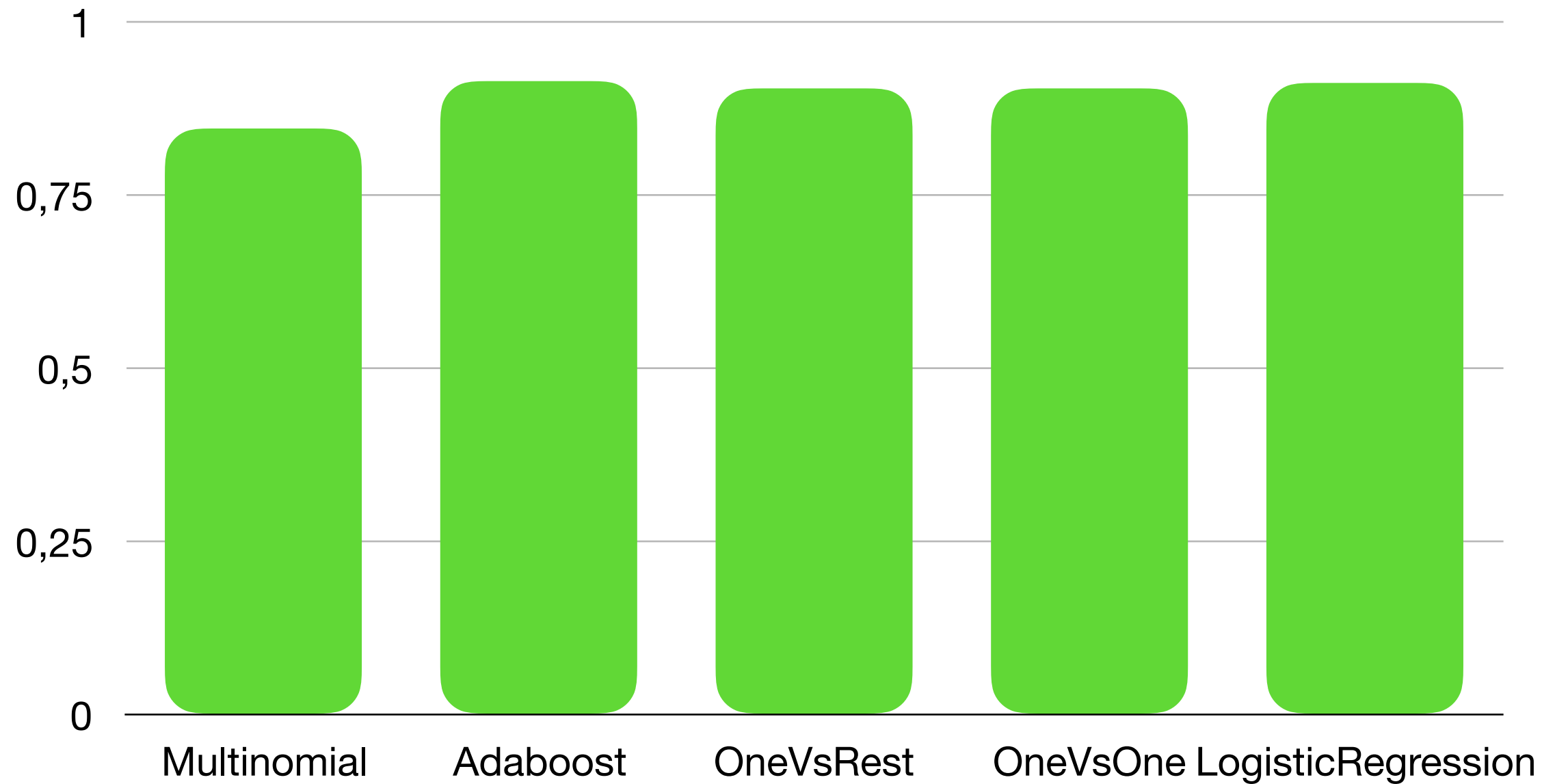




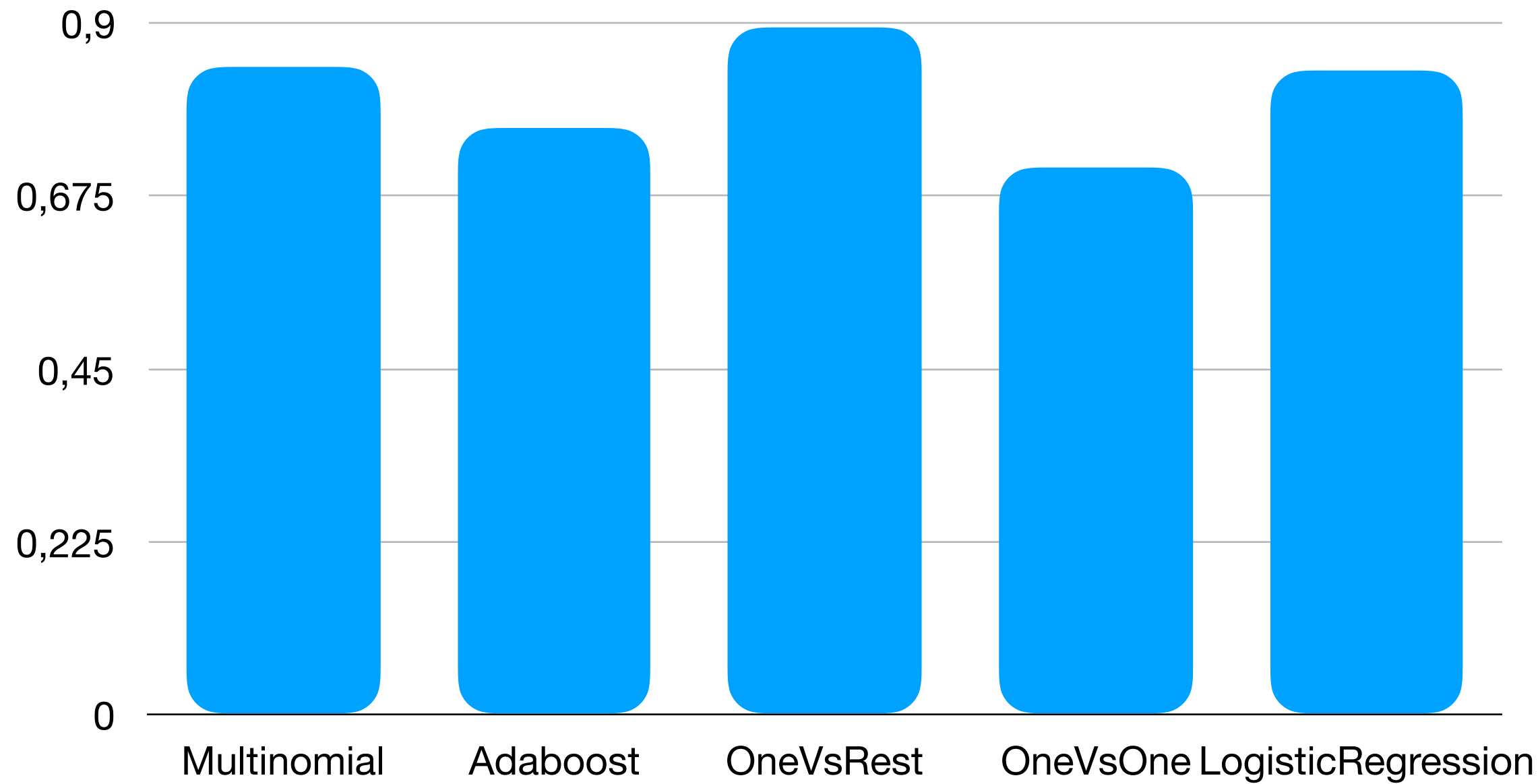
EXECUÇÃO DO ALGORITMO



Sem K-Folding



Com K-Folding





APRESENTAÇÃO DOS RESULTADOS

MATRIZ DE CONFUSÃO MULTINOMIALNB

3 150.6 504.2

143.2 320.8



APRESENTAÇÃO DOS RESULTADOS

MATRIZ DE CONFUSÃO ADABOOSTCLASSIFIER

3559.0 95.8

280.8 183.2



APRESENTAÇÃO DOS RESULTADOS

MATRIZ DE CONFUSÃO ONEVSRESTCLASSIFIER

2879.8 775.0

282.4 181.6



APRESENTAÇÃO DOS RESULTADOS

MATRIZ DE CONFUSÃO ONEVSONECLASSIFIER

2879.8 775.0

282.4 181.6

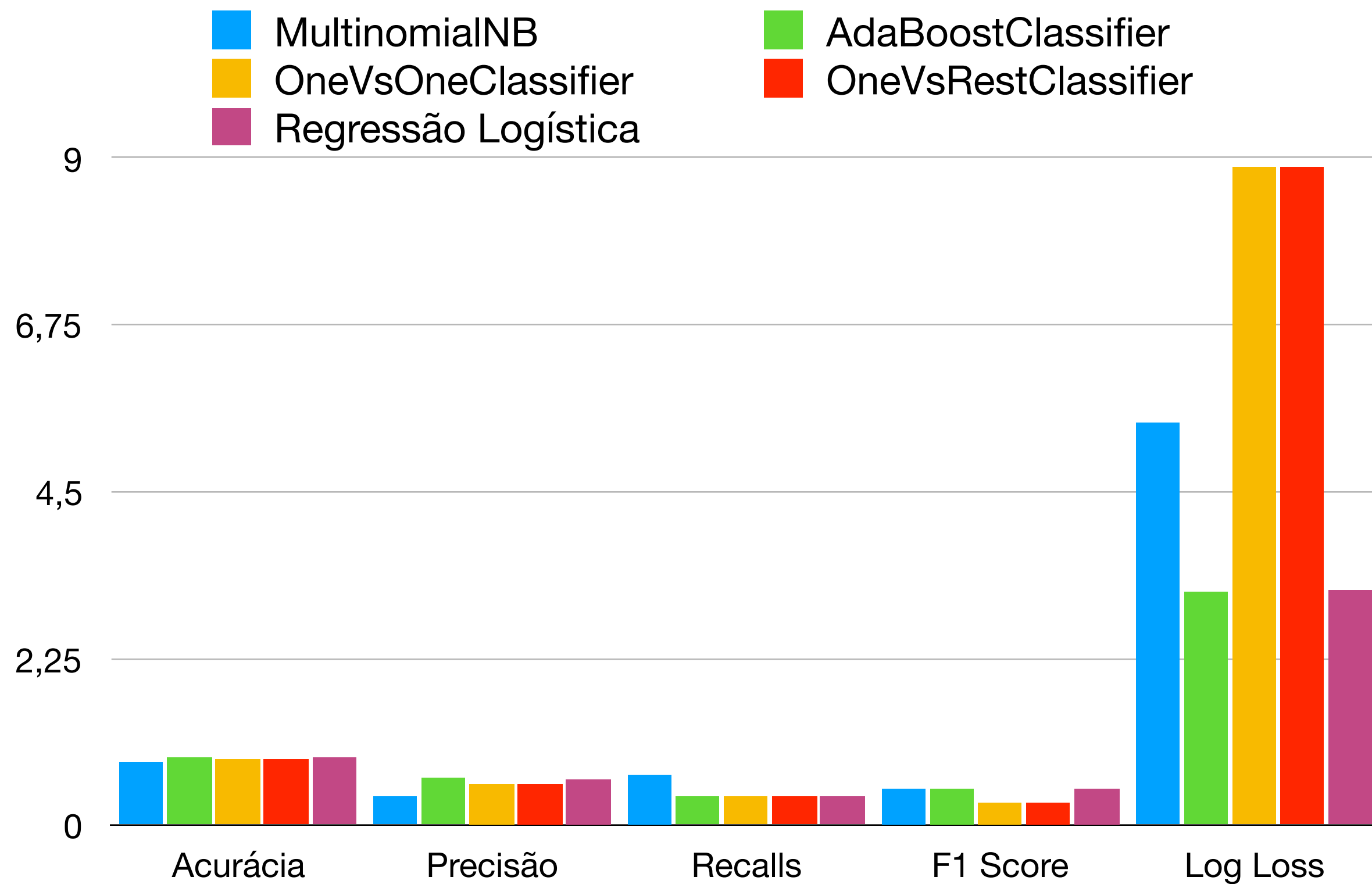


APRESENTAÇÃO DOS RESULTADOS

MATRIZ DE CONFUSÃO REGRESSÃO LOGÍSTICA

3560.4 94.4

284.8 179.2



Bank Marketing Data Set

RIYADH SILVA | JEFFERSON PIRES