**Atividade 1 – Clusters**

**Análise do Arquivo “Notas.csv” através dos algoritmos: Simple Kmeans e EM.**

Inicialmente executamos o algoritmo EM, e obtivemos a saída abaixo:

=== Run information ===

Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Relation: Notas

Instances: 423

Attributes: 173

[list of attributes omitted]

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 3

Number of iterations performed: 1

Cluster

Attribute 0 1 2

(0.39) (0.34) (0.28)

======================================================================================

GENERO

M 148.1599 129.6302 101.2099

F 17.0007 14.9993 18

[total] 165.1606 144.6295 119.2099

ADMINISTRACAO

mean 8.15 8.15 8.15

std. dev. 0.0946 0.1627 0.0946

ALGORITMOS\_DE\_PROGRAMACAO

mean 3.9558 3.9558 3.9558

std. dev. 0.0018 2.7826 1.6177

ALGORITMOS\_DE\_PROGRAMACAO\_I

mean 0 0 0

std. dev. 0 0 0

ALGORITMOS\_DE\_PROGRAMACAO\_II

mean 0 0 0

std. dev. 0 0 0

ALGORITMOS\_E\_PROGRAMACAO\_DE\_COMPUTADORES

mean 4.8751 4.8749 4.875

std. dev. 0.0161 0.251 0.1463

ANALISE\_E\_PROJETO\_DE\_SISTEMAS

mean 6.5066 5.6847 6.1232

std. dev. 0.8905 1.8699 1.2689

ANALISE\_E\_PROJETO\_DE\_SISTEMAS\_I

mean 5.5 5.5 5.5

std. dev. 0 0 0

ANATOMIA\_GERAL\_E\_PRIMEIROS\_SOCORROS

mean 0 0 0

std. dev. 0 0 0

ANATOMIA\_VETERINARIA\_

mean 0 0 0

std. dev. 0 0 0

ANTROPOLOGIA\_E\_SOCIOLOGIA

mean 4.375 4.375 4.375

std. dev. 0 0 0

ANTROPOLOGIA\_SOCIOLOGIA\_E\_SAUDE\_DO\_AMBIENTE\_

mean 0 0 0

std. dev. 0 0 0

APLICATIVOS\_MOVEIS

mean 6.5473 3.326 0.0153

std. dev. 2.2116 1.7825 0.1031

ARQUITETURA\_DE\_COMPUTADORES

mean 5.8882 3.4521 0.0288

std. dev. 1.7905 1.5266 0.1443

ARQUITETURA\_DE\_REFERENCIA\_PARA\_OS\_NEGOCIOS

mean 7.889 7.6954 7.7987

std. dev. 0.4529 1.2158 0.7653

ARQUITETURA\_E\_ORGANIZACAO\_DE\_COMPUTADORES

mean 5.019 5.019 5.019

std. dev. 0 1.7851 1.0378

ATIVIDADE\_FISICA\_E\_APTIDAO\_FISICA

mean 0 0 0

std. dev. 0 0 0

ATIVIDADES\_AQUATICAS\_I

mean 0 0 0

std. dev. 0 0 0

ATIVIDADES\_GINASTICAS\_-\_ARTISTICA\_E\_RITMICA

mean 0 0 0

std. dev. 0 0 0

ATIVIDADES\_RITMICAS\_E\_DANCA

mean 0 0 0

std. dev. 0 0 0

BANCO\_DE\_DADOS

mean 7.0952 6.3128 6.7303

std. dev. 0.8601 1.9978 1.3211

BANCO\_DE\_DADOS\_I

mean 7.7417 7.7417 7.7417

std. dev. 0.2031 0.3493 0.2031

BEM\_ESTAR\_ANIMAL

mean 0 0 0

std. dev. 0 0 0

BIOLOGIA\_CELULAR\_HISTOLOGIA\_E\_EMBRIOLOGIA\_VETERINARIA

mean 0 0 0

std. dev. 0 0 0

BIOQUIMICA\_VETERINARIA\_

mean 0 0 0

std. dev. 0 0 0

BIOTERISMO

mean 0 0 0

std. dev. 0 0 0

BUSINESS\_ANALYTICS

mean 8.631 8.6601 8.6446

std. dev. 0.2379 0.4258 0.2887

CALCULO\_DIFERENCIAL\_E\_INTEGRAL

mean 2.4405 2.4374 2.4442

std. dev. 0.0037 0.0866 0.0396

CALCULO\_DIFERENCIAL\_E\_INTEGRAL\_I

mean 2.5299 2.5664 2.5169

std. dev. 0.2168 0.1755 0.2335

CALCULO\_DIFERENCIAL\_E\_INTEGRAL\_II

mean 5.0667 5.0667 5.0667

std. dev. 0 0 0

CIENCIA\_DOS\_MATERIAIS

mean 0 0 0

std. dev. 0 0 0

COMPORTAMENTO\_HUMANO\_NAS\_ORGANIZAC†ES

mean 0 0 0

std. dev. 0 0 0

COMPUTACAO\_GRAFICA

mean 0 0 0

std. dev. 0 0 0

COMUNICACAO\_E\_CULTURA\_DIGITAL

mean 8.5 8.5 8.5

std. dev. 0 0 0

COMUNICACAO\_E\_EXPRESSAO

mean 5.4624 4.2705 4.8846

std. dev. 1.4604 2.6961 0.1702

COMUNICACAO\_EMPRESARIAL

mean 0 0 0

std. dev. 0 0 0

COMUNICACAO\_EMPRESARIAL\_E\_PROFISSIONAL

mean 5.075 5.075 5.075

std. dev. 0 0 0

COMUNICACAO\_INSTITUCIONAL

mean 0 0 0

std. dev. 0 0 0

COMUNICACAO\_PROFISSIONAL

mean 1.85 1.85 1.85

std. dev. 0.2206 0.3794 0.2206

CONTABILIDADE\_BASICA\_E\_GESTAO\_DE\_CUSTOS

mean 0.7167 0.7167 0.7167

std. dev. 0.0855 0.147 0.0855

CONTABILIDADE\_E\_ANALISE\_DE\_CUSTOS

mean 3.7 3.7 3.7

std. dev. 0 0 0

CRIATIVIDADE\_E\_EXPRESSAO

mean 0 0 0

std. dev. 0 0 0

CULTURA\_STARTUP

mean 7.8102 7.8181 7.8139

std. dev. 0.1339 0.2068 0.1463

DEONTOLOGIA\_E\_ETICA\_PROFISSIONAL

mean 0 0 0

std. dev. 0 0 0

DESENHO\_TECNICO

mean 5.6389 5.6554 5.6187

std. dev. 0.1625 0.197 0.2172

DESENVOLVIMENTO\_DE\_SISTEMAS\_COM\_PHP

mean 0 0 0

std. dev. 0 0 0

DESENVOLVIMENTO\_DE\_SOFTWARE

mean 6.7109 4.014 0.0244

std. dev. 2.1357 1.8965 0.1969

DESENVOLVIMENTO\_E\_APRENDIZAGEM\_MOTORA

mean 0 0 0

std. dev. 0 0 0

DESENVOLVIMENTO\_GERENCIAL

mean 0 0 0

std. dev. 0 0 0

DESENVOLVIMENTO\_PESSOAL

mean 0 0 0

std. dev. 0 0 0

DESIGN\_DE\_INTERACAO

mean 7.6079 4.8771 0.1257

std. dev. 1.1915 1.6443 0.5518

DIREITO\_E\_LINGUAGEM

mean 2.9 2.9 2.9

std. dev. 0 0 0

DIREITO\_E\_PSICOLOGIA

mean 4 4 4

std. dev. 0 0 0

ECONOMIA\_E\_NEGOCIOS\_INTERNACIONAIS

mean 2.35 2.35 2.35

std. dev. 0.2319 0.3989 0.2319

ELETRICIDADE

mean 2.0563 2.081 2.0435

std. dev. 0.1898 0.3686 0.1896

ELETRICIDADE\_E\_ELETROMAGNETISMO

mean 0 0 0

std. dev. 0 0 0

ELETRONICA\_ANALOGICA

mean 3.72 3.72 3.72

std. dev. 0 0 0

ELETRONICA\_DIGITAL

mean 0 0 0

std. dev. 0 0 0

EMPREENDEDORISMO\_E\_PLANO\_DE\_NEGOCIOS

mean 5.4 5.4 5.4

std. dev. 0 0 0

EMPREENDEDORISMO\_I

mean 1.2667 1.2667 1.2667

std. dev. 0.151 0.2598 0.151

ENGENHARIA\_DE\_SOFTWARE

mean 4.4974 4.617 4.5532

std. dev. 0.501 1.6989 1.0369

EPISTEMOLOGIA\_FILOSOFIA\_POLITICA\_E\_TEORIAS\_DA\_JUSTICA

mean 5.175 5.175 5.175

std. dev. 0 0 0

ESPORTE\_COLETIVO\_-\_BASQUETE

mean 0 0 0

std. dev. 0 0 0

ESPORTE\_COLETIVO\_-\_HANDEBOL

mean 0 0 0

std. dev. 0 0 0

ESPORTE\_COLETIVO\_-\_VOLEIBOL

mean 0 0 0

std. dev. 0 0 0

ESTATISTICA

mean 6.4929 6.4929 6.4929

std. dev. 0.2825 0.4859 0.2825

ESTRATEGIAS\_DE\_PRODUCAO\_INDUSTRIAL

mean 0 0 0

std. dev. 0 0 0

ESTRUTURA\_DE\_DADOS\_I

mean 7.67 7.67 7.67

std. dev. 0.0873 0.1501 0.0873

ESTRUTURA\_DE\_DADOS\_II

mean 5.76 5.76 5.76

std. dev. 0.2437 0.4191 0.2437

FILOSOFIA\_E\_ETICA

mean 5.0714 5.1457 5.1061

std. dev. 1.1333 1.3793 1.068

FISICA\_A

mean 1.7833 1.7857 1.7805

std. dev. 0.0005 0.0229 0.0307

FISICA\_I

mean 1.5169 1.544 1.5116

std. dev. 0.121 0.2783 0.1402

FISICA\_II

mean 3.9 3.9 3.9

std. dev. 0 0 0

FISICA\_PARA\_JOGOS\_DIGITAIS

mean 0 0 0

std. dev. 0 0 0

FISIOLOGIA\_HUMANA\_E\_BIOQUIMICA

mean 0 0 0

std. dev. 0 0 0

FOTOJORNALISMO

mean 8.325 8.325 8.325

std. dev. 0 0 0

FUNDAMENTOS\_DA\_ADMINISTRACAO

mean 0 0 0

std. dev. 0 0 0

FUNDAMENTOS\_DA\_ENGENHARIA\_DE\_PRODUCAO

mean 8.5 8.5 8.5

std. dev. 0 0 0

FUNDAMENTOS\_DA\_MATEMATICA

mean 1.7433 1.7897 1.7649

std. dev. 0.1942 0.2937 0.2101

FUNDAMENTOS\_DE\_ADMINISTRACAO

mean 6.3828 6.4445 6.4115

std. dev. 0.8298 0.8376 0.7099

FUNDAMENTOS\_DE\_SISTEMAS\_DE\_INFORMACAO

mean 6.0909 6.1714 6.1284

std. dev. 0.4783 1.5511 0.9502

GENETICA\_APLICADA\_ú\_MEDICINA\_VETERINARIA\_

mean 0 0 0

std. dev. 0 0 0

GEOMETRIA\_ANALITICA\_E\_ALGEBRA\_LINEAR

mean 3.7375 3.7375 3.7375

std. dev. 0 0 0

GESTAO\_COMERCIAL

mean 2.9187 2.9187 2.9187

std. dev. 0.2859 0.4918 0.2859

GESTAO\_DE\_PROJETOS

mean 7.6254 7.503 7.5683

std. dev. 0.3807 0.8608 0.556

GESTAO\_DE\_SISTEMAS\_DE\_INFORMACAO

mean 8.5 8.5 8.5

std. dev. 0.0989 0.1701 0.0989

GESTAO\_ESTRATEGICA\_DE\_EMPREENDIMENTOS

mean 0 0 0

std. dev. 0 0 0

GOVERNANCA\_DE\_TI

mean 7.6521 7.6475 7.65

std. dev. 0.0408 0.302 0.1774

HISTORIA\_CONTEMPORANEA\_DO\_BRASIL

mean 8.9 8.9 8.9

std. dev. 0 0 0

HISTORIA\_DA\_ARTE\_E\_DA\_CULTURA

mean 0 0 0

std. dev. 0 0 0

HISTORIA\_DA\_COMUNICACAO

mean 8.575 8.575 8.575

std. dev. 0 0 0

HUMANIDADES

mean 6.9012 6.9076 6.9042

std. dev. 0.1639 0.0339 0.1038

INFRAESTRUTURA\_TECNOLOGICA

mean 5.8248 4.4265 5.1726

std. dev. 1.3571 1.7975 1.4686

INTRODUCAO\_ú\_ENGENHARIA

mean 8.375 8.375 8.375

std. dev. 0 0 0

INTRODUCAO\_ú\_ENGENHARIA\_

mean 4.5528 4.5787 4.5112

std. dev. 0.0351 0.3066 0.4185

INTRODUCAO\_ú\_ENGENHARIA\_DA\_COMPUTACAO

mean 1.852 1.9013 1.875

std. dev. 0.2063 0.4958 0.3162

INTRODUCAO\_ú\_MEDICINA\_VETERINARIA\_E\_ZOOTECNIA

mean 0 0 0

std. dev. 0 0 0

INTRODUCAO\_ú\_PROGRAMACAO

mean 4.3747 4.3747 4.3747

std. dev. 0.0024 2.9309 1.7039

INTRODUCAO\_AO\_JORNALISMO\_E\_LEGISLACAO

mean 8.45 8.45 8.45

std. dev. 0 0 0

JORNALISMO\_DIGITAL\_I

mean 0 0 0

std. dev. 0 0 0

JORNALISMO\_GRAFICO\_I

mean 0 0 0

std. dev. 0 0 0

LIBRAS

mean 6.8169 6.7858 6.8024

std. dev. 0.5846 0.425 0.4397

LINGUA\_PORTUGUESA\_I

mean 8.45 8.45 8.45

std. dev. 0 0 0

LINGUA\_PORTUGUESA\_II

mean 0 0 0

std. dev. 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_EM\_C

mean 0 0 0

std. dev. 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_I

mean 0 0 0

std. dev. 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_II

mean 0 0 0

std. dev. 0 0 0

LOCALIZACAO

mean 0 0 0

std. dev. 0 0 0

LOCALIZACAO\_AVANCADA

mean 0 0 0

std. dev. 0 0 0

LOGICA\_MATEMATICA

mean 3.8602 3.8597 3.86

std. dev. 0.0295 0.6045 0.3519

LOGICA\_MATEMATICA\_I

mean 4.3006 4.3006 4.3006

std. dev. 0.0018 2.6725 1.5537

LOGICA\_MATEMATICA\_II

mean 4.5306 4.5306 4.5306

std. dev. 0 1.89 1.0988

MARKETING\_E\_GESTAO\_DE\_QUALIDADE\_EM\_HOSPITALIDADE

mean 0 0 0

std. dev. 0 0 0

MATEMATICA

mean 6.1425 3.649 0.0671

std. dev. 1.6057 1.4074 0.2678

MATEMATICA\_APLICADA

mean 0 0 0

std. dev. 0 0 0

MATEMATICA\_APLICADA\_AOS\_NEGOCIOS

mean 3.1 3.1 3.1

std. dev. 0 0 0

MATEMATICA\_DISCRETA

mean 3.4057 3.4057 3.4057

std. dev. 0.0006 2.1751 1.2645

MEDIC†ES\_E\_CALIBRACAO

mean 0 0 0

std. dev. 0 0 0

METODOLOGIA\_CIENTIFICA

mean 1653.0779 331.7952 1024.6883

std. dev. 2455.4819 939.2841 94.3304

METODOLOGIA\_DE\_PROJETOS

mean 1.7667 1.7667 1.7667

std. dev. 0.2107 0.3623 0.2107

METODOLOGIA\_DO\_TRABALHO\_ACADEMICO

mean 0 0 0

std. dev. 0 0 0

METODOS\_QUANTITATIVOS\_I

mean 1.0333 1.0333 1.0333

std. dev. 0.1232 0.2119 0.1232

NOVO\_AMBIENTE\_ORGANIZACIONAL

mean 6.9605 6.3923 6.6954

std. dev. 0.8365 1.8543 1.221

NUTRICAO\_E\_DIETETICA

mean 0 0 0

std. dev. 0 0 0

OFICINA\_DE\_BANCO\_DE\_DADOS

mean 9 9 9

std. dev. 0.1377 0.2368 0.1377

OFICINA\_DE\_MATEMATICA

mean 0 0 0

std. dev. 0 0 0

OFICINA\_DE\_PROGRAMACAO\_I

mean 4.761 4.761 4.761

std. dev. 0 1.8126 1.0538

OFICINA\_DE\_PROGRAMACAO\_II

mean 6.6 6.6 6.6

std. dev. 0.2539 0.4367 0.2539

ORIENTACAO\_A\_OBJETOS

mean 4.4318 3.3501 3.9273

std. dev. 1.9845 1.8397 1.6964

PESSOAS\_BENS\_E\_NEGOCIO\_JURIDICO

mean 3 3 3

std. dev. 0 0 0

PLANEJAMENTO\_ARQUITETONICO\_E\_MANUTENCAO

mean 0 0 0

std. dev. 0 0 0

PLANEJAMENTO\_E\_ORGANIZACAO\_DE\_EVENTOS

mean 0 0 0

std. dev. 0 0 0

PLANEJAMENTO\_GRAFICO

mean 8.05 8.05 8.05

std. dev. 0 0 0

POLITICA\_ESTADO\_E\_CONSTITUICAO

mean 6.25 6.25 6.25

std. dev. 0 0 0

PRATICAS\_DE\_COZINHA\_E\_PRODUCAO\_DE\_ALIMENTOS\_E\_BEBIDAS

mean 0 0 0

std. dev. 0 0 0

PRE-CALCULO

mean 0.6 0.6042 0.5949

std. dev. 0.0015 0.0372 0.0552

PRODUCAO\_MULTIMIDIA

mean 0 0 0

std. dev. 0 0 0

PROGRAMACAO\_DE\_COMPUTADORES

mean 0 0 0

std. dev. 0 0 0

PROGRAMACAO\_E\_CALCULO\_NUMERICO

mean 4.775 4.775 4.775

std. dev. 0 0 0

PROGRAMACAO\_I

mean 2.9146 2.9662 2.9041

std. dev. 0.2332 1.9401 0.2694

PROGRAMACAO\_II

mean 7.1083 7.1083 7.1083

std. dev. 0.2256 0.388 0.2256

QUIMICA\_TECNOLOGICA

mean 3.9703 4.078 4.0165

std. dev. 0.4429 0.3758 0.0317

RACIOCINIO\_LOGICO

mean 6.6645 3.8298 0.103

std. dev. 1.4437 1.4724 0.3737

RADIOJORNALISMO\_I

mean 0 0 0

std. dev. 0 0 0

REDACAO\_JORNALISTICA

mean 8.125 8.125 8.125

std. dev. 0 0 0

REDE\_DE\_COMPUTADORES\_I

mean 5.6 5.6 5.6

std. dev. 0 0 0

REDES\_DE\_COMPUTADORES

mean 0 0 0

std. dev. 0 0 0

REPRESENTACAO\_E\_MODELAMENTO\_GEOMETRICO

mean 0 0 0

std. dev. 0 0 0

SAUDE\_PUBLICA\_APLICADA\_ú\_MEDICINA\_VETERINARIA

mean 0 0 0

std. dev. 0 0 0

SEGURANCA\_E\_HIGIENE\_DE\_ALIMENTOS

mean 0 0 0

std. dev. 0 0 0

SISTEMAS\_CONVERGENTES

mean 7.4471 7.1141 7.2918

std. dev. 0.5516 1.0276 0.7032

SISTEMAS\_CORPORATIVOS

mean 8.3345 8.2816 8.3098

std. dev. 0.4277 0.8593 0.5664

SISTEMAS\_DE\_INFORMACAO

mean 8.2843 5.5804 0.1392

std. dev. 0.8753 1.6189 0.8664

SISTEMAS\_EMBARCADOS

mean 6.8182 6.4022 6.6242

std. dev. 0.9 1.3382 0.9745

SISTEMAS\_EMBARCADOS\_I\_

mean 2.7 2.7 2.7

std. dev. 0 0 0

SISTEMAS\_FINANCEIROS

mean 7.6 7.6 7.6

std. dev. 0 0 0

SISTEMAS\_OPERACIONAIS

mean 6.9063 6.4704 6.703

std. dev. 0.6238 1.1947 0.8168

SISTEMAS\_OPERACIONAIS\_I

mean 7.4 7.4 7.4

std. dev. 0.0275 0.0474 0.0275

SISTEMAS\_OPERACIONAIS\_II

mean 9.6 9.6 9.6

std. dev. 0 0 0

SOCIOLOGIA\_DA\_COMUNICACAO

mean 0 0 0

std. dev. 0 0 0

TECNICA\_HOSPITALAR\_VETERINARIA

mean 0 0 0

std. dev. 0 0 0

TECNICAS\_DE\_APURACAO

mean 9.025 9.025 9.025

std. dev. 0 0 0

TECNICAS\_DE\_SERVICO\_EM\_HOSPITALIDADE

mean 0 0 0

std. dev. 0 0 0

TELEJORNALISMO\_I

mean 0 0 0

std. dev. 0 0 0

TEORIA\_E\_HISTORIA\_DO\_DIREITO

mean 7 7 7

std. dev. 0 0 0

TEORIA\_GERAL\_DA\_ADMINISTRACAO

mean 3.944 3.944 3.944

std. dev. 0.0007 2.3957 1.3928

TEORIAS\_DA\_COMUNICACAO

mean 0 0 0

std. dev. 0 0 0

TOPICOS\_ESPECIAIS\_EM\_SOFTWARE

mean 7.5821 7.5764 7.5794

std. dev. 0.2883 0.1403 0.197

TOPOGRAFIA

mean 2.2375 2.2375 2.2375

std. dev. 0 0 0

TRABALHO\_DE\_CONCLUSAO\_DE\_CURSO

mean 7.2838 7.2917 7.2875

std. dev. 0.3307 0.8716 0.5469

TRABALHO\_DE\_ENGENHARIA\_DA\_COMPUTACAO

mean 7.8375 7.8374 7.8375

std. dev. 0.006 0.093 0.0542

TRABALHO\_INTERDISCIPLINAR\_I

mean 8.9 8.9 8.9

std. dev. 0.0871 0.1498 0.0871

Time taken to build model (full training data) : 5.61 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 204 ( 48%)

1 109 ( 26%)

2 110 ( 26%)

Log likelihood: 1290.53808

Class attribute: Situacao

Classes to Clusters:

0 1 2 <-- assigned to cluster

43 48 107 | Evadido

161 61 3 | Ativo

Cluster 0 <-- Ativo

Cluster 1 <-- No class

Cluster 2 <-- Evadido

Incorrectly clustered instances : 155.0 36.643 %

Depois executamos o algoritmo SImple Kmeans, com a saída abaixo:

=== Run information ===

Scheme: weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: Notas

Instances: 423

Attributes: 173

[list of attributes omitted]

Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

kMeans

======

Number of iterations: 11

Within cluster sum of squared errors: 358.48512320261045

Initial starting points (random):

Cluster 0: M,8.15,3.955814,0,0,4.875,6.1232,5.5,0,0,4.375,0,7.25,5.125,7.798686,5.019,0,0,0,0,6.730294,7.741667,0,0,0,0,8.644608,2.440476,2.538607,5.066667,0,0,0,8.5,4.900418,0,5.075,0,1.85,0.716667,3.7,0,7.813889,0,5.638889,0,6.575,0,0,0,7.625,2.9,4,2.35,2.061071,0,3.72,0,5.4,1.266667,4.553191,5.175,0,0,0,6.492857,0,7.67,5.76,5.106071,1.783333,1.524567,3.9,0,0,8.325,0,8.5,1.764931,6.411538,6.128431,0,3.7375,2.91875,7.568333,8.5,0,7.65,8.9,0,8.575,6.904167,5.1726,8.375,4.55,1.875,0,4.374706,8.45,0,0,6.802381,8.45,0,0,0,0,0,0,3.86,4.300581,4.530612,0,6,0,3.1,3.405682,0,1033.438732,1.766667,0,1.033333,6.695433,0,9,0,4.761,6.6,3.927254,3,0,0,8.05,6.25,0,0.6,0,0,4.775,2.929091,7.108333,4.019444,7.375,0,8.125,5.6,0,0,0,0,7.29181,8.309828,8.75,6.624167,2.7,7.6,6.702966,7.4,9.6,0,0,9.025,0,0,7,3.944048,0,7.579412,2.2375,7.2875,7.8375,8.9

Cluster 1: M,8.15,3.955814,0,0,4.875,7.875,5.5,0,0,4.375,0,7.925,7.8,7.798686,5.019,0,0,0,0,7.125,7.741667,0,0,0,0,8.644608,2.440476,2.538607,5.066667,0,0,0,8.5,3.7,0,5.075,0,1.85,0.716667,3.7,0,7.813889,0,5.638889,0,4.05,0,0,0,7.25,2.9,4,2.35,2.061071,0,3.72,0,5.4,1.266667,4.553191,5.175,0,0,0,6.492857,0,7.67,5.76,5.106071,1.783333,1.524567,3.9,0,0,8.325,0,8.5,1.764931,6.411538,6.128431,0,3.7375,2.91875,7.568333,8.5,0,7.65,8.9,0,8.575,6.904167,7.5,8.375,4.55,1.875,0,4.374706,8.45,0,0,6.802381,8.45,0,0,0,0,0,0,3.86,4.300581,4.530612,0,7,0,3.1,3.405682,0,5.3,1.766667,0,1.033333,8.35,0,9,0,4.761,6.6,7.5,3,0,0,8.05,6.25,0,0.6,0,0,4.775,2.929091,7.108333,4.019444,6,0,8.125,5.6,0,0,0,0,7.29181,8.309828,9,6.624167,2.7,7.6,6.702966,7.4,9.6,0,0,9.025,0,0,7,3.944048,0,7.579412,2.2375,7.2875,7.8375,8.9

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#

Attribute Full Data 0 1

(423.0) (168.0) (255.0)

=====================================================================================

GENERO M M M

ADMINISTRACAO 8.15 8.15 8.15

ALGORITMOS\_DE\_PROGRAMACAO 3.9558 3.2539 4.4182

ALGORITMOS\_DE\_PROGRAMACAO\_I 0 0 0

ALGORITMOS\_DE\_PROGRAMACAO\_II 0 0 0

ALGORITMOS\_E\_PROGRAMACAO\_DE\_COMPUTADORES 4.875 4.875 4.875

ANALISE\_E\_PROJETO\_DE\_SISTEMAS 6.1232 6.0411 6.1773

ANALISE\_E\_PROJETO\_DE\_SISTEMAS\_I 5.5 5.5 5.5

ANATOMIA\_GERAL\_E\_PRIMEIROS\_SOCORROS 0 0 0

ANATOMIA\_VETERINARIA\_ 0 0 0

ANTROPOLOGIA\_E\_SOCIOLOGIA 4.375 4.375 4.375

ANTROPOLOGIA\_SOCIOLOGIA\_E\_SAUDE\_DO\_AMBIENTE\_ 0 0 0

APLICATIVOS\_MOVEIS 3.6512 0.806 5.5256

ARQUITETURA\_DE\_COMPUTADORES 3.4432 0.8567 5.1473

ARQUITETURA\_DE\_REFERENCIA\_PARA\_OS\_NEGOCIOS 7.7987 7.7969 7.7999

ARQUITETURA\_E\_ORGANIZACAO\_DE\_COMPUTADORES 5.019 4.9305 5.0773

ATIVIDADE\_FISICA\_E\_APTIDAO\_FISICA 0 0 0

ATIVIDADES\_AQUATICAS\_I 0 0 0

ATIVIDADES\_GINASTICAS\_-\_ARTISTICA\_E\_RITMICA 0 0 0

ATIVIDADES\_RITMICAS\_E\_DANCA 0 0 0

BANCO\_DE\_DADOS 6.7303 6.701 6.7496

BANCO\_DE\_DADOS\_I 7.7417 7.7417 7.7417

BEM\_ESTAR\_ANIMAL 0 0 0

BIOLOGIA\_CELULAR\_HISTOLOGIA\_E\_EMBRIOLOGIA\_VETERINARIA 0 0 0

BIOQUIMICA\_VETERINARIA\_ 0 0 0

BIOTERISMO 0 0 0

BUSINESS\_ANALYTICS 8.6446 8.6446 8.6446

CALCULO\_DIFERENCIAL\_E\_INTEGRAL 2.4405 2.443 2.4388

CALCULO\_DIFERENCIAL\_E\_INTEGRAL\_I 2.5386 2.5294 2.5447

CALCULO\_DIFERENCIAL\_E\_INTEGRAL\_II 5.0667 5.0667 5.0667

CIENCIA\_DOS\_MATERIAIS 0 0 0

COMPORTAMENTO\_HUMANO\_NAS\_ORGANIZAC†ES 0 0 0

COMPUTACAO\_GRAFICA 0 0 0

COMUNICACAO\_E\_CULTURA\_DIGITAL 8.5 8.5 8.5

COMUNICACAO\_E\_EXPRESSAO 4.9004 3.9866 5.5024

COMUNICACAO\_EMPRESARIAL 0 0 0

COMUNICACAO\_EMPRESARIAL\_E\_PROFISSIONAL 5.075 5.075 5.075

COMUNICACAO\_INSTITUCIONAL 0 0 0

COMUNICACAO\_PROFISSIONAL 1.85 1.839 1.8573

CONTABILIDADE\_BASICA\_E\_GESTAO\_DE\_CUSTOS 0.7167 0.7124 0.7195

CONTABILIDADE\_E\_ANALISE\_DE\_CUSTOS 3.7 3.7 3.7

CRIATIVIDADE\_E\_EXPRESSAO 0 0 0

CULTURA\_STARTUP 7.8139 7.8139 7.8139

DEONTOLOGIA\_E\_ETICA\_PROFISSIONAL 0 0 0

DESENHO\_TECNICO 5.6389 5.6248 5.6481

DESENVOLVIMENTO\_DE\_SISTEMAS\_COM\_PHP 0 0 0

DESENVOLVIMENTO\_DE\_SOFTWARE 3.9488 0.9411 5.9302

DESENVOLVIMENTO\_E\_APRENDIZAGEM\_MOTORA 0 0 0

DESENVOLVIMENTO\_GERENCIAL 0 0 0

DESENVOLVIMENTO\_PESSOAL 0 0 0

DESIGN\_DE\_INTERACAO 4.6139 1.3225 6.7823

DIREITO\_E\_LINGUAGEM 2.9 2.9 2.9

DIREITO\_E\_PSICOLOGIA 4 4 4

ECONOMIA\_E\_NEGOCIOS\_INTERNACIONAIS 2.35 2.336 2.3592

ELETRICIDADE 2.0611 2.0567 2.064

ELETRICIDADE\_E\_ELETROMAGNETISMO 0 0 0

ELETRONICA\_ANALOGICA 3.72 3.72 3.72

ELETRONICA\_DIGITAL 0 0 0

EMPREENDEDORISMO\_E\_PLANO\_DE\_NEGOCIOS 5.4 5.4 5.4

EMPREENDEDORISMO\_I 1.2667 1.2591 1.2716

ENGENHARIA\_DE\_SOFTWARE 4.5532 4.4754 4.6044

EPISTEMOLOGIA\_FILOSOFIA\_POLITICA\_E\_TEORIAS\_DA\_JUSTICA 5.175 5.175 5.175

ESPORTE\_COLETIVO\_-\_BASQUETE 0 0 0

ESPORTE\_COLETIVO\_-\_HANDEBOL 0 0 0

ESPORTE\_COLETIVO\_-\_VOLEIBOL 0 0 0

ESTATISTICA 6.4929 6.4929 6.4929

ESTRATEGIAS\_DE\_PRODUCAO\_INDUSTRIAL 0 0 0

ESTRUTURA\_DE\_DADOS\_I 7.67 7.67 7.67

ESTRUTURA\_DE\_DADOS\_II 5.76 5.76 5.76

FILOSOFIA\_E\_ETICA 5.1061 5.0855 5.1196

FISICA\_A 1.7833 1.7813 1.7846

FISICA\_I 1.5246 1.5134 1.5319

FISICA\_II 3.9 3.9 3.9

FISICA\_PARA\_JOGOS\_DIGITAIS 0 0 0

FISIOLOGIA\_HUMANA\_E\_BIOQUIMICA 0 0 0

FOTOJORNALISMO 8.325 8.325 8.325

FUNDAMENTOS\_DA\_ADMINISTRACAO 0 0 0

FUNDAMENTOS\_DA\_ENGENHARIA\_DE\_PRODUCAO 8.5 8.5 8.5

FUNDAMENTOS\_DA\_MATEMATICA 1.7649 1.7842 1.7522

FUNDAMENTOS\_DE\_ADMINISTRACAO 6.4115 6.4019 6.4179

FUNDAMENTOS\_DE\_SISTEMAS\_DE\_INFORMACAO 6.1284 6.0725 6.1653

GENETICA\_APLICADA\_ú\_MEDICINA\_VETERINARIA\_ 0 0 0

GEOMETRIA\_ANALITICA\_E\_ALGEBRA\_LINEAR 3.7375 3.7375 3.7375

GESTAO\_COMERCIAL 2.9188 2.9014 2.9302

GESTAO\_DE\_PROJETOS 7.5683 7.562 7.5725

GESTAO\_DE\_SISTEMAS\_DE\_INFORMACAO 8.5 8.5 8.5

GESTAO\_ESTRATEGICA\_DE\_EMPREENDIMENTOS 0 0 0

GOVERNANCA\_DE\_TI 7.65 7.65 7.65

HISTORIA\_CONTEMPORANEA\_DO\_BRASIL 8.9 8.9 8.9

HISTORIA\_DA\_ARTE\_E\_DA\_CULTURA 0 0 0

HISTORIA\_DA\_COMUNICACAO 8.575 8.575 8.575

HUMANIDADES 6.9042 6.9042 6.9042

INFRAESTRUTURA\_TECNOLOGICA 5.1726 5.0712 5.2394

INTRODUCAO\_ú\_ENGENHARIA 8.375 8.375 8.375

INTRODUCAO\_ú\_ENGENHARIA\_ 4.55 4.5256 4.5661

INTRODUCAO\_ú\_ENGENHARIA\_DA\_COMPUTACAO 1.875 1.875 1.875

INTRODUCAO\_ú\_MEDICINA\_VETERINARIA\_E\_ZOOTECNIA 0 0 0

INTRODUCAO\_ú\_PROGRAMACAO 4.3747 3.6324 4.8638

INTRODUCAO\_AO\_JORNALISMO\_E\_LEGISLACAO 8.45 8.45 8.45

JORNALISMO\_DIGITAL\_I 0 0 0

JORNALISMO\_GRAFICO\_I 0 0 0

LIBRAS 6.8024 6.8024 6.8024

LINGUA\_PORTUGUESA\_I 8.45 8.45 8.45

LINGUA\_PORTUGUESA\_II 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_EM\_C 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_I 0 0 0

LINGUAGEM\_DE\_PROGRAMACAO\_II 0 0 0

LOCALIZACAO 0 0 0

LOCALIZACAO\_AVANCADA 0 0 0

LOGICA\_MATEMATICA 3.86 3.86 3.86

LOGICA\_MATEMATICA\_I 4.3006 3.6094 4.756

LOGICA\_MATEMATICA\_II 4.5306 4.4075 4.6117

MARKETING\_E\_GESTAO\_DE\_QUALIDADE\_EM\_HOSPITALIDADE 0 0 0

MATEMATICA 3.6183 0.9144 5.3997

MATEMATICA\_APLICADA 0 0 0

MATEMATICA\_APLICADA\_AOS\_NEGOCIOS 3.1 3.1 3.1

MATEMATICA\_DISCRETA 3.4057 2.8137 3.7957

MEDIC†ES\_E\_CALIBRACAO 0 0 0

METODOLOGIA\_CIENTIFICA 1033.4387 793.7788 1191.3323

METODOLOGIA\_DE\_PROJETOS 1.7667 1.7562 1.7736

METODOLOGIA\_DO\_TRABALHO\_ACADEMICO 0 0 0

METODOS\_QUANTITATIVOS\_I 1.0333 1.0272 1.0374

NOVO\_AMBIENTE\_ORGANIZACIONAL 6.6954 6.6783 6.7067

NUTRICAO\_E\_DIETETICA 0 0 0

OFICINA\_DE\_BANCO\_DE\_DADOS 9 9 9

OFICINA\_DE\_MATEMATICA 0 0 0

OFICINA\_DE\_PROGRAMACAO\_I 4.761 4.6133 4.8583

OFICINA\_DE\_PROGRAMACAO\_II 6.6 6.6 6.6

ORIENTACAO\_A\_OBJETOS 3.9273 3.8602 3.9714

PESSOAS\_BENS\_E\_NEGOCIO\_JURIDICO 3 3 3

PLANEJAMENTO\_ARQUITETONICO\_E\_MANUTENCAO 0 0 0

PLANEJAMENTO\_E\_ORGANIZACAO\_DE\_EVENTOS 0 0 0

PLANEJAMENTO\_GRAFICO 8.05 8.05 8.05

POLITICA\_ESTADO\_E\_CONSTITUICAO 6.25 6.25 6.25

PRATICAS\_DE\_COZINHA\_E\_PRODUCAO\_DE\_ALIMENTOS\_E\_BEBIDAS 0 0 0

PRE-CALCULO 0.6 0.5964 0.6024

PRODUCAO\_MULTIMIDIA 0 0 0

PROGRAMACAO\_DE\_COMPUTADORES 0 0 0

PROGRAMACAO\_E\_CALCULO\_NUMERICO 4.775 4.775 4.775

PROGRAMACAO\_I 2.9291 2.8317 2.9932

PROGRAMACAO\_II 7.1083 7.1083 7.1083

QUIMICA\_TECNOLOGICA 4.0194 4.0262 4.015

RACIOCINIO\_LOGICO 3.8905 0.984 5.8054

RADIOJORNALISMO\_I 0 0 0

REDACAO\_JORNALISTICA 8.125 8.125 8.125

REDE\_DE\_COMPUTADORES\_I 5.6 5.6 5.6

REDES\_DE\_COMPUTADORES 0 0 0

REPRESENTACAO\_E\_MODELAMENTO\_GEOMETRICO 0 0 0

SAUDE\_PUBLICA\_APLICADA\_ú\_MEDICINA\_VETERINARIA 0 0 0

SEGURANCA\_E\_HIGIENE\_DE\_ALIMENTOS 0 0 0

SISTEMAS\_CONVERGENTES 7.2918 7.2901 7.293

SISTEMAS\_CORPORATIVOS 8.3098 8.3072 8.3115

SISTEMAS\_DE\_INFORMACAO 5.1157 1.6293 7.4126

SISTEMAS\_EMBARCADOS 6.6242 6.6007 6.6396

SISTEMAS\_EMBARCADOS\_I\_ 2.7 2.7 2.7

SISTEMAS\_FINANCEIROS 7.6 7.6 7.6

SISTEMAS\_OPERACIONAIS 6.703 6.6866 6.7138

SISTEMAS\_OPERACIONAIS\_I 7.4 7.4 7.4

SISTEMAS\_OPERACIONAIS\_II 9.6 9.6 9.6

SOCIOLOGIA\_DA\_COMUNICACAO 0 0 0

TECNICA\_HOSPITALAR\_VETERINARIA 0 0 0

TECNICAS\_DE\_APURACAO 9.025 9.025 9.025

TECNICAS\_DE\_SERVICO\_EM\_HOSPITALIDADE 0 0 0

TELEJORNALISMO\_I 0 0 0

TEORIA\_E\_HISTORIA\_DO\_DIREITO 7 7 7

TEORIA\_GERAL\_DA\_ADMINISTRACAO 3.944 3.315 4.3585

TEORIAS\_DA\_COMUNICACAO 0 0 0

TOPICOS\_ESPECIAIS\_EM\_SOFTWARE 7.5794 7.5794 7.5794

TOPOGRAFIA 2.2375 2.2375 2.2375

TRABALHO\_DE\_CONCLUSAO\_DE\_CURSO 7.2875 7.2875 7.2875

TRABALHO\_DE\_ENGENHARIA\_DA\_COMPUTACAO 7.8375 7.8375 7.8375

TRABALHO\_INTERDISCIPLINAR\_I 8.9 8.9 8.9

Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 168 ( 40%)

1 255 ( 60%)

Class attribute: Situacao

Classes to Clusters:

0 1 <-- assigned to cluster

151 47 | Evadido

17 208 | Ativo

Cluster 0 <-- Evadido

Cluster 1 <-- Ativo

Incorrectly clustered instances : 64.0 15.13 %

**Análise dos Resultados:**

Através do arquivo de notas, pudemos observar que o maior número de evasões está ligado a disciplinas da área de exatas, principalmente matemática e matérias importantes de ciências de computação. Em outras áreas, a evasão é muito pequena as vezes inexistente.

Tal comportamento nos leva a crer que matérias importantes com baixo aproveitamento leva os alunos a evadir os cursos.

Quanto aos resultados apresentados pelos algoritmos, vemos uma melhor acurácia do Kmeans. Ela era esperada, pois, o algoritmo foca nos atributos da classe informada na construção dos centroides e observação aos redores dele. Desta forma ele criou dois clusters (Ativo e Evadido), com um bom agrupamento muito próximo da realidade. Tal situação pode ser observada na matriz de confusão, que possui poucos falsos positivos e negativos, além de um número menor de instâncias classificadas incorretamente, 64, contra 155, mais que o dobro, apresentado pelo EM

Não que o EM seja ineficiente, mas nessa situação, como ele ignora os atributos da classe e tenta criar os centroides por observação dos agrupamentos, ele acabou criando um terceiro grupo sem classificação, que aparentemente roubou metade dos registros de cada classe.

Portanto, nesta aplicação o algoritmo Simple Kmeans teve uma eficácia maior sobre o EM.