

Data Mining aplicado ao contexto de transtornos depressivos: analisando perfis dos indivíduos

Alunos: Gabriel Luciano
Geovane Fonseca
Luigi Domenico
Paulo Junio

13 de Junho de 2019



Sumário

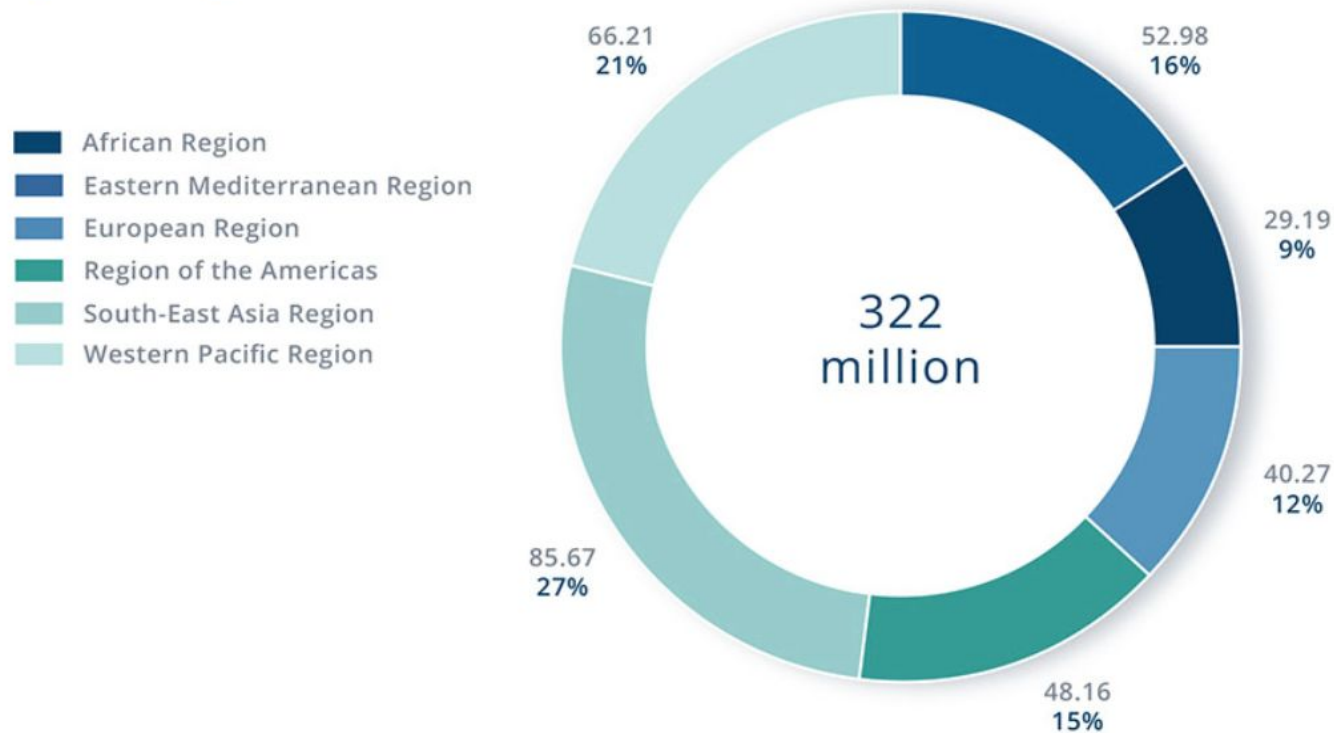
- Problema
- Metodologia
- Resultados
- Conclusão



Problema

- Depressão afeta várias pessoas ao redor do mundo
- **Objetivo:** descobrir perfis de pessoas diagnosticadas com o transtorno
- **Método:** algoritmo de agrupamento (K-modes)

Cases of depressive disorder (millions), by WHO Region





Metodologia

Base de dados: Foi utilizada uma base de dados disponibilizada pelo IBGE de uma pesquisa feita em 2013

- Dados sobre a moradia (Não foi utilizada para o problema)
- Dados sobre o indivíduo

205.546 instâncias x 943 atributos



Metodologia

Classes	Quantidade de Registros
Com Depressão	4.235
Sem Depressão	55.967
Sem Resposta	145.344

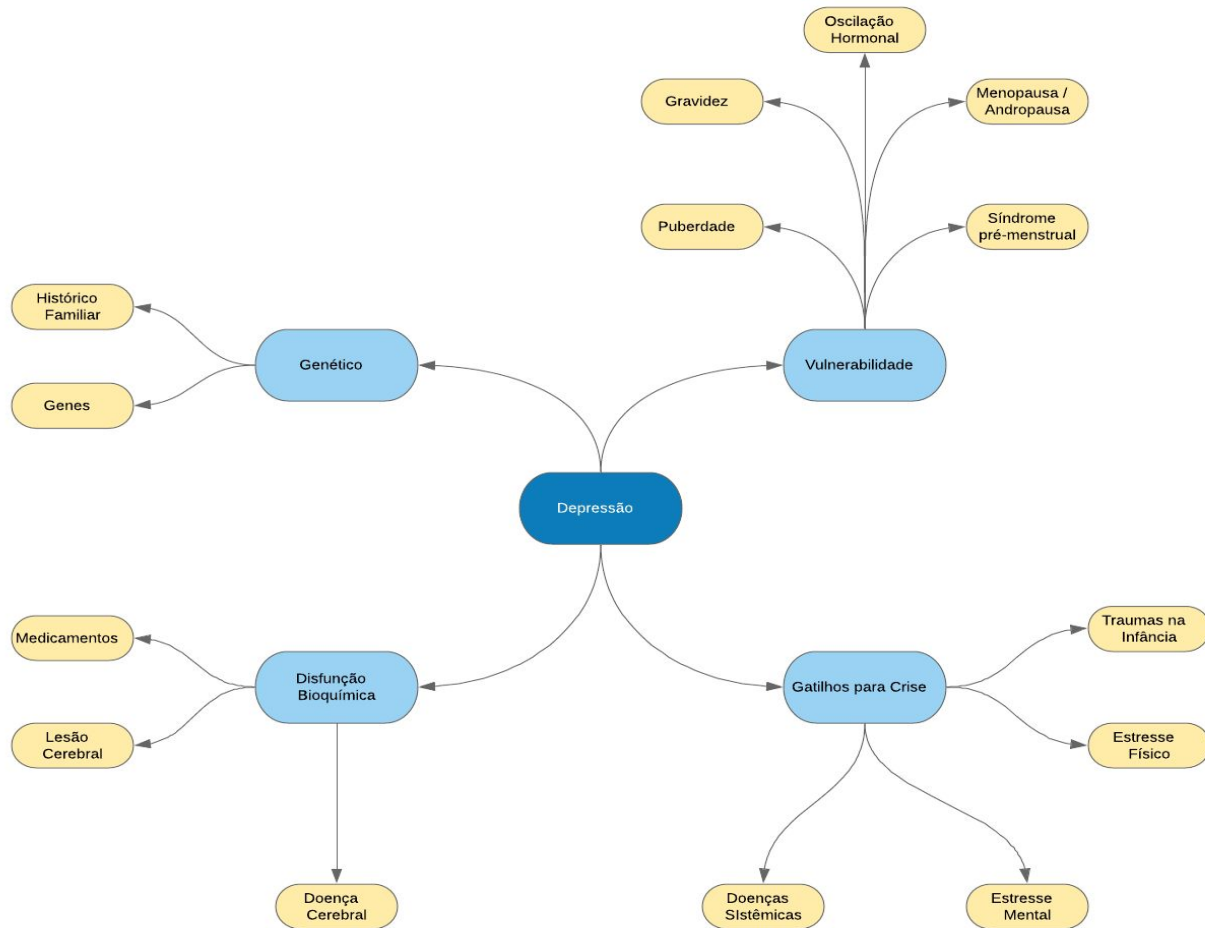


Metodologia

Domínio do Problema: Foi criado um mapa conceitual com as possíveis causas da Depressão

- Genética
- Disfunções Bioquímicas
- Gatilhos
- Vulnerabilidades

60.202 instâncias x 68 atributos





Metodologia

Pré-processamento:

- Seleção Manual dos Atributos
- Eliminação de *Outliers*
- Eliminação de Atributos com dados ausentes
- Outras transformações (i.e. juntar/balancear atributos)

8470 instâncias x 24 atributos



Metodologia

Classes	Quantidade de Registros
Com Depressão	4.235
Sem Depressão	4.235



Metodologia

Pré-processamento:

- Discretização de dados numéricos
 - Idade do Indivíduo
 - Quantidade de dias em que o indivíduo usa remédios para dormir

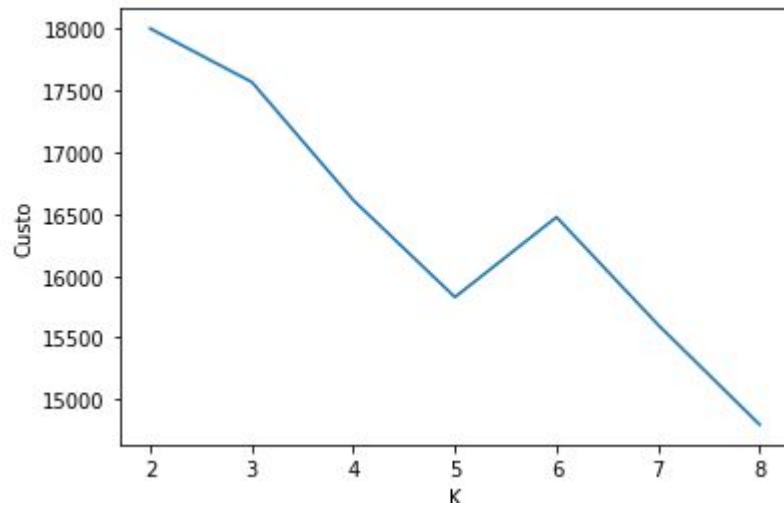


Metodologia

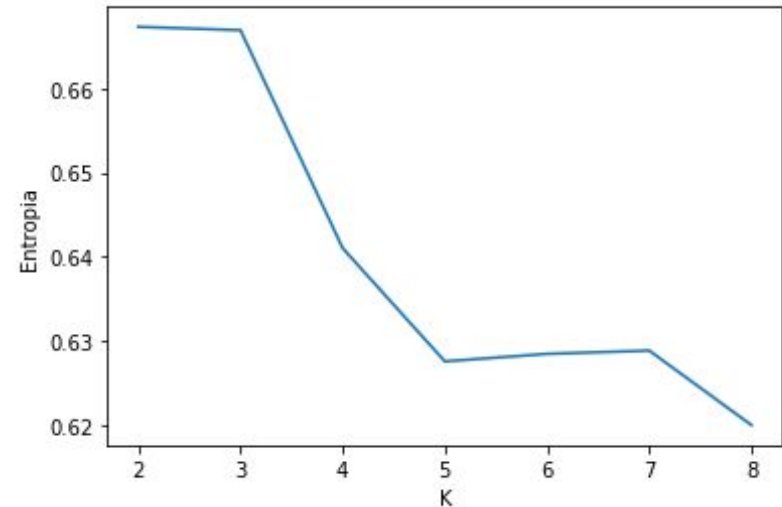
Data Mining:

- Aplicou-se o algoritmo de agrupamento K-modes
- Foram testados diversos valores para k ($2 \leq k \leq 8$)
- Avaliou a qualidade da clusterização utilizando:
 - Pureza
 - Entropia
 - *Silhouette Index*

Resultados

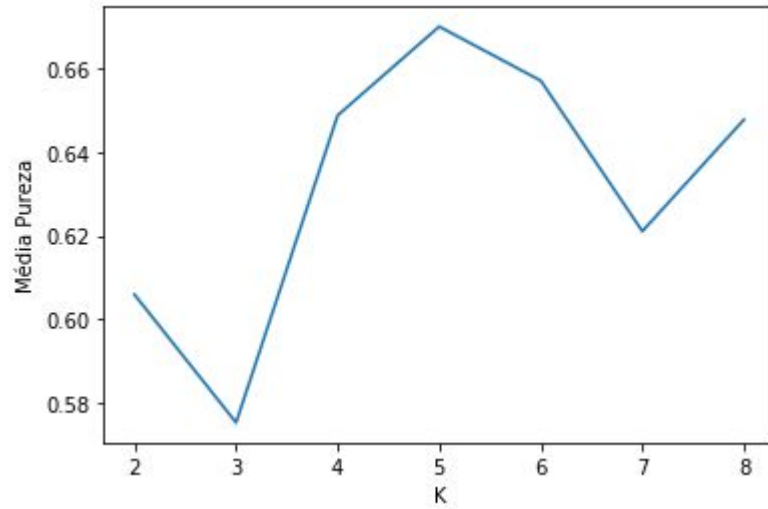


Custo de criação por Cluster

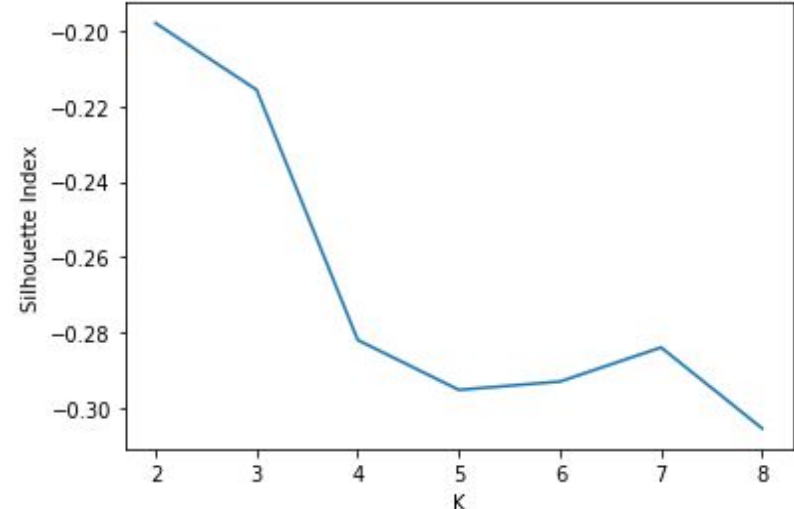


Entropia por cluster

Resultados



Média de pureza para cada cluster



Silhouette index para cada cluster



Resultados

Cluster 3:

- Quantidade de dias em que usa remédio para dormir
- Ocorrência de menstruação
- Presença de doenças físicas ou mentais
- Diagnóstico de doenças pulmonares
- Realização de atividades físicas semanais



Resultados

Cluster 4:

- Esquizofrenia, bipolaridade, psicose ou TOC
- Quantidade de dias em que usa remédio para dormir
- Cansaço em relação a subidas ou dores no peito
- Presença de doenças físicas ou mentais
- Diagnóstico de doenças pulmonares



Conclusão

- Resultados não satisfatórios
- Identificação de Clusters
 - Entropia baixa
 - Pureza alta
 - Silhouette index negativo (**Problema**)



Dúvidas?

