

# Análise e Mineração de Dados sobre Síndrome Respiratória Aguda Grave (SRAG)

Uálaci dos Anjos Júnior

Orientadora: Luciana Lee

Universidade Federal do Espírito Santo

22 de setembro de 2024

laci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

#### Sumário

- 1. Introdução
- 2. Fundamentação Teórica
- 3. Pré-processamento de Dados
- 4. Análise Exploratória de Dados
- 5. Árvores de Decisão
- 6. Regras de Associação
- 7. Redes Neurais
- 8. Visualização de Dados
- 9. Conclusões



22 de setembro de 2024

#### Visão Geral da SRAG

- Surgimento
- A SRAG é uma condição respiratória grave com alta taxa de mortalidade, especialmente durante a pandemia de COVID-19.
- Afeta indivíduos de todas as idades, com maior incidência em grupos vulneráveis.

# Motivação para Mineração de Dados

- Pandemia de COVID-19.
- Iniciativa de dados abertos do governo brasileiro.



# Objetivos da Pesquisa

#### Gerais:

■ Apoiar a tomada de decisão em saúde pública e melhorar as intervenções clínicas.

#### Específicos:

- Analisar dados da SRAG para obter insights relevantes.
- Reduzir dimensionalidade e ajustar a base para melhor desempenho ao aplicar técnicas de Machine Learning.
- Identificar correlações entre parâmetros e evolução dos casos.
- Realizar predições sobre a evolução dos casos de SRAG.
- Otimizar modelos para melhor desempenho.
- Desenvolver uma plataforma interativa de visualização de dados.

álaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

## Conceitos e Técnicas de Mineração de Dados

- Mineração de dados é parte do processo de descoberta de conhecimento em bases de dados (KDD).
- Técnicas aplicadas: árvores de decisão, redes neurais, FP-Growth.

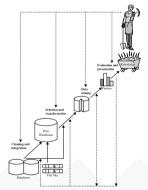


Figura 1 - Processo de Descoberta de Conhecimento em Bases de Dados (KDD). (Autor).

#### Fontes e Características dos Dados de SRAG

- Dados extraídos do OpenDataSUS entre 2017 e 2023.
- Parâmetros: idade, sexo, comorbidades, sintomas, desfecho, informações hospitalares, etc.
- Formato CSV
- Todo território Brasileiro, granularidade municipal, a nível de pacientes

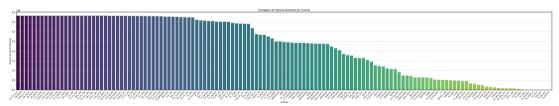


Figura 2 – Valores ausentes por parâmetro.(Autor).

ilaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

#### Trabalhos Relacionados

- Estudos prévios aplicaram árvores de decisão e análise de correlação em dados de saúde pública para prever desfechos clínicos.
- Este trabalho complementa essas pesquisas com redes neurais e FP-Growth.

Uálaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de s

## Pré-processamento inicial dos Dados

#### Limpeza e Integração dos Dados:

- Integração de dados de diferentes anos em um único dataset unificado.
- Exclusão de parâmetros irrelevantes e correção de inconsistências nos registros.

#### Tratamento de Parâmetros:

- Análise de valores ausentes por variável.
- Aplicação de métodos como imputação para valores críticos.
- Conversão de parâmetros

Feature 1	Feature 2	Corr.
SEM_NOT	NU_IDADE_N	0.9864
CO_REGIONA	CO_RG_RESI	0.9374
CO_REGIONA	CO_RG_INTE	0.9991
CO_MUN_NOT	CO_MUN_RES	0.9909
CO_MUN_NOT	CO_MU_INTE	0.9999
NU_IDADE_N	SEM_NOT	0.9864
CO_RG_RESI	CO_REGIONA	0.9374
CO_RG_RESI	CO_RG_INTE	0.9382
CO_MUN_RES	CO_MUN_NOT	0.9909
CO_MUN_RES	CO_MU_INTE	0.9905
HEMATOLOGI	HEPATICA	0.9210
HEPATICA	HEMATOLOGI	0.9210
M_AMAMENTA	CO_PS_VGM	0.9755
CO_RG_INTE	CO_REGIONA	0.9991
CO_RG_INTE	CO_RG_RESI	0.9382
RES_IGM	RES_IGG	0.9046
IDADE	QUINQUENIO	0.9994
QUINQUENIO	IDADE	0.9994

Tabela 1 – Correlação entre os parâmetros na base de dados. (Autor).

#### ■ Distribuição por faixa etária

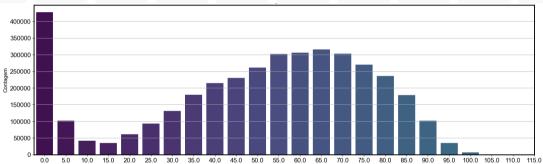


Figura 3 – Distribuição de casos por faixa etária. (Autor).

10 / 26

- Distribuição de evolução dos casos para crianças de até 5 anos
- Homogeneidade na distribuição de casos

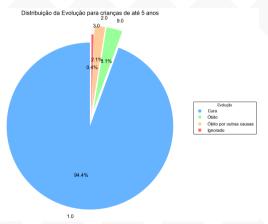


Figura 4 – Distribuição de evolução dos casos para crianças de até 5 anos. (Autor).

# Algoritmo de Árvores de Decisão

- Algoritmo baseado em divisões recursivas dos dados.
- Utilizado para prever desfechos de pacientes com SRAG.

Uálaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

# Características Importantes na Predição

- Principais variáveis preditivas: idade, necessidade de ventilação, diagnóstico inicial.
- O algoritmo destacou a idade como o principal fator de risco.



## Avaliação e Interpretação dos Resultados

- Acurácia do modelo de árvore de decisão: 75%.
- Análise da matriz de confusão para avaliação de erros do modelo.

# Algoritmo FP-Growth

- Algoritmo eficiente para descoberta de padrões frequentes em grandes bases de dados.
- Utilizado para encontrar associações entre sintomas e desfechos clínicos.

## Associações Encontradas

- Relação entre comorbidades hepáticas e risco aumentado de mortalidade.
- Associações frequentes entre uso de ventilação e desfecho desfavorável.

Uálaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

# Interpretação das Regras de Associação

- Regras de associação sugerem que pacientes com certas comorbidades têm maior risco de evolução negativa.
- Implicações para intervenções preventivas em pacientes de alto risco.



laci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

#### Arquitetura e Treinamento

- Rede neural com múltiplas camadas densas para predição de desfechos clínicos.
- Treinada com dados rotulados da base SRAG, com foco em precisão.

#### Previsão de Resultados

- A rede neural alcançou uma precisão de 77% na predição de desfechos clínicos.
- Fatores como idade e suporte ventilatório foram os mais relevantes.

22 de setembro de 2024

## Avaliação de Desempenho

- A avaliação foi feita utilizando métricas como acurácia, precisão e recall.
- A otimização de hiperparâmetros foi realizada com o uso de Optuna.



#### Painéis Interativos

- Desenvolvimento de dashboards interativos para explorar os dados.
- Painéis facilitam a análise de padrões epidemiológicos.

## Exploração de Padrões

- Visualização das correlações entre sintomas e desfechos em diferentes grupos etários.
- Permite identificar fatores de risco e tendências ao longo do tempo.

Uálaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024

#### Facilitação da Tomada de Decisão

- Painéis foram projetados para apoiar decisões rápidas em saúde pública.
- Visualizações facilitam a identificação de áreas críticas e grupos de risco.

Uálaci, A. Jr. (UFES) Mineração de Dados SRAG

#### Resumo dos Resultados

- A mineração de dados na base SRAG revelou fatores críticos como idade e suporte ventilatório.
- Algoritmos como Árvores de Decisão e FP-Growth foram eficazes na descoberta de padrões.

## Contribuições e Implicações

- Contribuição significativa para a análise preditiva em saúde pública.
- Implicações para a implementação de políticas preventivas e tratamento de grupos de risco.

## Limitações e Trabalhos Futuros

- Limitações: Dados restritos ao Brasil e inconsistências na qualidade dos registros.
- Futuras pesquisas podem expandir para dados internacionais e incorporar fatores socioeconômicos.



Uálaci, A. Jr. (UFES) Mineração de Dados SRAG 22 de setembro de 2024