**Qual principal questão deve ser respondida?**

• prever perfil de pacientes e desdobramentos clínicos.

**Ideia Geral:**

A complexidade de opções diagnósticas e terapêuticas disponíveis para área de cirurgia neurológica é evidente no início do século XXI. Da mesma forma, também é variado o espectro de resultados possíveis porque existem muitas dimensões de interpretação. Ainda influenciados pelo contexto social, pacientes e médicos estão sobrecarregados com informações da era digital. A tomada de decisão hoje é cardinal e crítica. O recurso contemporâneo para essa demanda desafiante é a aplicação das tecnologias de sistemas de gestão da informação como a inteligência artificial.

Simplificar a complexidade não é uma solução. Ciência aplicada aos dados, compreendida como conhecimento humano agregado as tecnologias digitais é a melhor alternativa disponível para tomada de decisão na área de cirurgia neurológica.

Discutir opções baseadas em dados é uma decisão mais importante que uma incisão.

**Variáveis binárias importantes:**

O "diagnóstico" foi divido em dois para facilitar o preenchimento do protocolo.  
1 - Dor Lombar (Ou Lombalgia)  
2 - Dor na Perna (Ou Ciática)  
Também pode existir um terceiro que poderia ser Dor Lombar + Dor na Perna.  
  
Existe também dois tipos de "tratamento" que foram usados no preenchimento do protocolo.  
1 - Não cirúrgico   
2 - Cirúrgico  
  
Também duas condições de "trabalho" que foram preenchidas que são importantes:  
1 - Trabalhando  
2 - Não trabalhando

Por exemplo, correlacionar tudo que estiver associado com:  
  
$&:$-)):$/)/&:)/$.... dados = Dor Lombar + Tratamento Cirúrgico + Não trabalhando   
  
$&:$-)):$/)/&:)/$.... dados = Dor na Perna + Tratamento Não Cirúrgico + Trabalhando  
  
Ou seja, saber quais as três ou quatro variáveis (de todo banco de dados enviados) que mais estão correlacionadas com Diagnóstico/Tratamento/Trabalho seria muito importante e valioso  
  
Será que dá?

**SEGUNDO E-MAIL:**

Oi Natalia, tudo bem?

Excelente!

Os dados e mesmo o texto são muito ricos. Acho que temos bastante chances de competir pelo prêmio se seguir assim.

Por exemplo: "77% das pessoas que responderam apresentar dor nas pernas responderam que a dor que eles mais gostariam de se livrar é a dor nas costas. O que indica que para essas pessoas, a dor nas costas apresenta um incômodo muito maior do que a dor nas pernas." **Essa informação é importantíssima!**

Será que conseguimos colocar essas informações nos formatos exemplificados "em anexo"?

Será que eles conseguiriam cruzar essas informações com aquelas do questionário OSWESTRY? E os demais se for possível?

Enquanto isso vou trabalhando no texto do resumo para enviarmos.

Também precisaria de uma sugestão do título técnico sobre qual modalidade de análise (dentro de inteligência artificial que eles estão utilizando).

E o nome completo e instituição dos autores que vão participar do seu lado.

Muito obrigado,

 Vamos falando

Abraço

Vinícius

Resumo das técnicas e ferramentas aplicadas:

1. Pré-processamento dos dados:
   1. *One Hot Encoding*
   2. *Label Encoding*
   3. Tabela de contingência (tabela cruzada) para obter associação entre variáveis categóricas binárias
   4. Correlação de ponto bisserial para associação entre variável categórica binária e variáveis contínuas ou categóricas ordinais
   5. Construção de correlogramas
   6. Matrix de proximidade para tratamento de valores nulos
2. Modelos de aprendizado de máquina não supervisionados para agrupamento dos dados em grupos (*clustering*)

- Algoritmos que podem ser testados e avaliados quanto a performance:

a. KMeans

b. DBScan - *Density Based Spatial Clustering for Applications with Noise*

c. Agrupamento Hierárquico

1. Modelos de aprendizado de máquina para predição de variáveis resposta desejadas:
2. CART *Classification and Regression Trees* – (Árvores de decisão)
3. Modelo de *ensemble Random Forest* (utilização de n árvores de decisão para classificação ou regressão)
4. *Support Vector Machines (SVM)*
5. Redes neurais
6. Linguagens de programação utilizadas:
   * Python
   * R

**TEXTO VITOR E JOÃO:**

Em questões de múltipla escolha que não apresentam na resposta uma relação de ordinalidade, aplica-se a técnica de *One Hot Encoding*

A correlação entre as variáveis binárias pode ser realizada através da utilização de tabelas de contingência, também conhecidas como tabelas cruzadas.

Exemplo:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Livrar-se da dor nas costas | |
|  |  | 0 (não assinalou) | 1 (assinalou a opção) |
| Dor nas pernas | 0 (não possui) |  |  |
|  | 1 (possui) |  |  |

Processo de análise de dados:

• Quais variáveis, vetores colunas, são relevantes para o diagnóstico de doença crônica?

• Aquelas que não forem relevantes serão retiradas do banco de dados.

• Quantificar quantos valores faltantes há em cada coluna:

• Nas perguntas Se\_Sim, criar colunas com as respostas

Criar colunas para cada medicamento

Os medicamentos podem ser classificados em analgésicos e outros tipos, logo faz sentido utilizar o próprio nome do medicamento?

• O mesmo vale para a pergunte Se\_sim quais as doenças

•Da linha 141 para frente há informação da dose em mg dos medicamentos e antes da linha 141 não há esta informação ao lado do nome dos medicamentos. Por que?

Passaram a solicitar apenas a partir da linha 141 essa informação?

**25/08/2020**

• Quais dores em uma parte do corpo estão mais correlacionadas com dores em outra parte do corpo? (Correlação das dores, matriz de correlação?; correlação de matriz binária?, tabela cruzada?)

* Exemplo de tabela cruzada, a porcentagem simples, ou seja: 52 % dos pacientes com dor na lombar, tem dor nas costas a mais de 2 anos.

Uma tabela desse tipo é útil para construção de por exemplo árvore de decisão, a qual é facilmente interpretável e possível de ser visualizada.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **DOR A** | **Dor\_costas** | **Dor\_costas\_VAS** | **Dor\_pernas** | **Dor\_pernas\_VAS** |
| 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 7 | 1 | 6 | 0 | 0 |
| 3 | 22 | 1 | 10 | 1 | 10 |
| 4 | 5 | 1 | 4 | 0 | 0 |
| 5 | 8 | 1 | 7 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 |
| 7 | 3 | 1 | 2 | 0 | 0 |
| 8 | 6 | 1 | 5 | 0 | 0 |
| 9 | 4 | 1 | 3 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 |