**TRABALHO DE FORMATURA – T.2020154**

**BMA e BMAC**

**Relatório dos meses de Maio e Junho**

1. NOME DO ALUNO: **Fabio Carvalho de Souza**

2. NÚMERO USP: **9425125**

3. CURSO: [ ] BMA **[x] BMAC**

4. HABILITAÇÃO:

BMA: [ ]101 [ ] 501 [ ] 611 [ ] 801

BMAC: [ ] 104 **[x] 204** [ ] 304 [ ] 404 [ ] 504 [ ] 604 [ ] 704 [ ] 804 [ ] 904 [ ] 1004

5. NOME DO ORIENTADOR: **Prof. Dr.** [**Helder Takashi Imoto Nakaya**](http://csbiology.com/)

UNIDADE DO ORIENTADOR: **Faculdade de Ciências Farmacêuticas (FCF)**

6. NOME DO CO-ORIENTADOR (se houver): À Definir

UNIDADE DO CO-ORIENTADOR: À Definir

7. TÍTULO DO PROJETO: **APLICAÇÃO DE MACHINE LEARNING PARA CLUSTERIZAÇÃO DE PACIENTES COM ARBOVIROSES**

8. RESUMO DAS ATIVIDADES DO BIMESTRE

As atividades referentes ao bimestre consistiram em melhorar os parâmetros de modelagem com o objetivo de resultar em métricas mais consistentes e que expressassem a melhor classificação do modelo quanto à resposta desejada. Mas se verificou que a separação das informações de severidade não está seguindo o esperado para o percentual marcado previamente, que se sabe o grau da infecção, pelo seguinte motivo que sugiro ser a motivação:

- Podemos ter 2 grupos severos (ou mais), com características diferentes, então como solução e passos seguintes, poderemos fazer uma análise descritiva das variáveis, ou seja, será que as variáveis fortes do grupo que está em destaque na base, não se relacionam totalmente com o grupo que ainda não foi definido e que está apresentando probabilidades baixas.

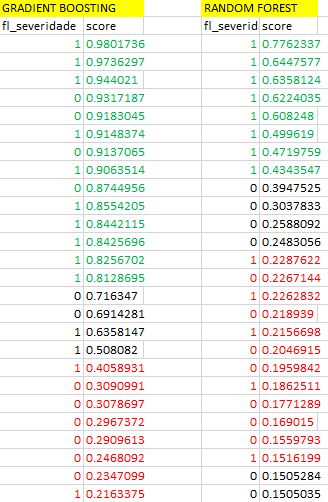


FIGURA 1:Exemplificação do caso possivel de 2 grupos, devido ao score baixo para casos severos confirmados (marcação 1)

Também considero testar como opção alternativa a técnica do K-means (já sendo aplicada mas ainda sem resultados), casos não tenhamos resultados significativos dos modelos iniciais, pois dessa forma teremos um meio de reunir informações com certa semelhança, e com base nisso verificar como classificar os pacientes tanto em grupo de características quanto em definir a severidade.

Vale ressaltar também que, a severidade em si pode não estar relacionada completamente com as variáveis que estão disponíveis, sendo algum fator não observado ou presente nos dados que pode ter influência forte na severidade, como por exemplo a variável com o número baixo de plaquetas que neste caso temos que descartar, por causa da variação da métrica entre os locais da coleta da informação e o desconhecimento das correspondentes para realizar a transformação. Outro ponto a mencionar dadas estas observações, é o fato de podermos nos deparar com a chance de ser inviavel separar os grupos de forma corrreta por meio das características.

|  |  |
| --- | --- |
| Variavel | Importance |
| laboratorio\_hemograma\_hemoglobina | 0.3177231770688286 |
| laboratorio\_hemograma\_hematocrito | 0.31269219910689766 |
| esteve\_hospitalizado\_7D | 0.060881000176705995 |
| Idade | 0.03917347288203914 |
| tot\_sinais\_de\_alarme | 0.03737414354032871 |
| exantema\_onde | 0.028013232027646096 |
| ja\_teve\_dengue | 0.01989123227532029 |
| SAQ\_Dor intensa e contÃ­nua no abdome (barriga) | 0.017425117819495076 |
| dor\_articulacao\_rigidez | 0.016323854586030445 |
| manchas\_pele\_local | 0.016111305739175852 |
| fl\_sexo | 0.013410253780047007 |
| outros\_sinais\_alarme | 0.013398320371714673 |
| OSAQ\_Dor Ã  palpaÃ§Ã£o abdominal | 0.012789025044478623 |
| SF7D\_Dor de cabeÃ§a | 0.012189042055724662 |
| manchas\_pele\_cocam | 0.010320431856301531 |
| Exantema | 0.010292201439868084 |
| dor\_articulacao\_qtde | 0.010081944125078585 |
| dor\_articulacao\_lado | 0.009019700316396184 |
| dor\_articulacao\_local | 0.008380706964504254 |
| esteve\_hospitalizado\_agravamento\_dengue\_7D | 0.006873253189661222 |
| SA7D\_Dor intensa e contÃ­nua no abdome (barriga) | 0.006653951072059115 |
| Sangramento | 0.006549025042948457 |
| Raca | 0.005314329667122733 |
| manchas\_pele\_cor | 0.0043971382203358824 |
| sangramento\_como | 0.004258119978614839 |
| SA14D\_PressÃ£o baixa | 0.00028088124826799023 |
| esteve\_hospitalizado\_14D | 0.0001742108913441259 |

TABELA 1: Variáveis com valor de importância acima de 0 modelo Gradiente

|  |  |
| --- | --- |
| Variavel | importance |
| laboratorio\_hemograma\_hematocrito | 0.1903334697088079 |
| esteve\_hospitalizado\_quantos\_dias\_7D | 0.12838800682224003 |
| laboratorio\_hemograma\_hemoglobina | 0.11244668118270697 |
| esteve\_hospitalizado\_agravamento\_dengue\_14D | 0.053047205789191976 |
| Idade | 0.04462601159232275 |
| esteve\_hospitalizado\_7D | 0.04246756000886816 |
| esteve\_hospitalizado\_agravamento\_dengue\_7D | 0.0351902537242753 |
| esteve\_hospitalizado\_14D | 0.03018833273044926 |
| OSAQ\_Letargia/irritabilidade/rebaixamento de nÃ­vel de consciÃªncia | 0.02862755030682664 |
| exantema\_onde | 0.025923752125935243 |
| esteve\_hospitalizado\_quantos\_dias\_14D | 0.022812656982191284 |
| SA14D\_Dor intensa e contÃ­nua no abdome (barriga) | 0.017697525138681735 |
| Raca | 0.01677724081946573 |
| tot\_sinais\_de\_alarme | 0.01561947234664292 |
| OSAQ\_Dor Ã  palpaÃ§Ã£o abdominal | 0.013492106708964613 |
| SAQ\_PressÃ£o baixa | 0.01196047764333225 |
| SA7D\_Manchas vermelhas na pele | 0.01179936384072247 |
| outros\_sinais\_alarme | 0.011497563378572285 |
| fl\_sexo | 0.011327568266845141 |
| dor\_articulacao\_local | 0.011109212296677873 |
| SAQ\_Dor intensa e contÃ­nua no abdome (barriga) | 0.010156633174233046 |
| dor\_articulacao\_rigidez | 0.009848903373599796 |
| SA7D\_Dor intensa e contÃ­nua no abdome (barriga) | 0.009750736056885312 |
| sangramento\_como | 0.009038894719635997 |
| manchas\_pele\_local | 0.008625936900934007 |
| SA14D\_PressÃ£o baixa | 0.008352351232096825 |
| OSAQ\_! 'Sinais clÃ­nicos ou Radiografia de tÃ³rax com derrame pleural ' | 0.0077308499219121436 |
| SF7D\_Dor atrÃ¡s dos olhos | 0.007724004925337092 |
| ja\_teve\_dengue | 0.007624227544214237 |
| Exantema | 0.007515698013609885 |
| dor\_articulacao\_lado | 0.007068036502896023 |
| SF7D\_Dor de cabeÃ§a | 0.0068938239207798025 |
| Sangramento | 0.006706046067147949 |
| dor\_articulacao\_qtde | 0.005890432477288881 |
| manchas\_pele\_cocam | 0.0058578103121918485 |
| SA7D\_VÃ´mitos persistentes | 0.005832077811605245 |
| SAQ\_VÃ´mitos persistentes | 0.005762891735472476 |
| SF7D\_Dor muscular (costas | 0.00508439464619334 |
| SA14D\_! 'Muito sono | 0.0048568046232062055 |
| manchas\_pele\_cor | 0.0045223720889130525 |
| SAQ\_Sangramento de mucosas (nariz | 0.004413178932713076 |
| SA7D\_PressÃ£o baixa | 0.0026454381716042234 |
| tem\_alguma\_doenca\_das\_articulacoes | 0.0024852270216109058 |
| SA14D\_Manchas vermelhas na pele | 0.002183006775435002 |
| edema\_onde | 0.0018494990330673827 |
| sintoma\_febre\_7dias\_inchaco | 0.001654508536745447 |
| SAQ\_Queda abrupta de plaquetas | 0.0015764642738843753 |
| Edema | 0.001121684390263504 |
| SA7D\_! 'Muito sono | 0.0005325183646621076 |
| SA7D\_VÃ´mitos com sangue | 0.0004540854181996017 |
| OSAQ\_HematÃ³crito > 44% (mulheres) ou 50% (homens) | 0.0003814669529829283 |
| SF7D\_Dor nas articulaÃ§Ãµes (juntas) | 0.0003235044471584519 |

TABELA 2: Variáveis com importância acima de 0 modelo Random Forest

Verificando as tabelas 1 e 2 e figura 1 temos uma exemplificação do caso mencionado, com os valores de significância de cada variável por modelo desenvolvido, bem como a classificação dos primeiros casos pelo score (probabilidade) e vemos que existe uma diferença para cada modelo, o que deveria ser bem semelhante, dada que a resposta esperada deve ser a mesma, sendo o ponto de trabalho atual encontrar o corte de cada grupo (o que poderemos ver pelas variáveis ou K-means) e conseguir assim, obter informações concretas dessa análise.

Fabio C.S

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Assinatura do aluno Assinatura do orientador

30 de Junho de 2020