Adult dataset

Obtidos subdatasets com determinadas combinações de dados de forma a obter IR e GR desejados para avaliar o impacto de alteração destas métricas

Nesta experiencia um dos valores é mantido estático a 0.5 e o outro vai ser alterado para um dos seguintes valores [0.01,0.02,0.05,0.1,0.2, 0.3,0.4,0.5,0.6,0.7,0.8,0.9,0.95,0.98,0.99]

Para race:

|  |  |  |
| --- | --- | --- |
|  | <=50K | >50K |
| notWhite | 3560 | 669 |
| White | 19093 | 6839 |

GR = (3560 +669) / 32560 = 0.1299

IR = (669 + 6839) / 32560 = 0.2306

Fairness

Da análise das seguintes imagens para a variação dos valores de GR entre 0.2 e 0.8 com IR a 0.5 os valores de fairness tende a estar estáveis próximos de fairness perfeita para os diferentes classificadores utilizados ( como esperado datasets mais balanceados tendem a apresentar melhores valores de fairness)

Variar GR -> Para um GR elevado (mais exemplos da classe protegida do que da classe não protegida)

Uma imagem com texto, diagrama, Esquema, número

Descrição gerada automaticamenteUma imagem com texto, file, Paralelo, captura de ecrã

Descrição gerada automaticamente

Para a variação de GR as métricas de fairness aparentam todas ter comportamentos semelhantes sendo que quando possuímos um numero demasiado elevado ou reduzido da classe protegida a fairness tende a piorar.

Variar IR -> IR elevado (mais exemplos da classe positiva do que da classe negativa)

Uma imagem com texto, diagrama, Esquema, Desenho técnico

Descrição gerada automaticamenteUma imagem com texto, file, Paralelo, captura de ecrã

Descrição gerada automaticamente

Escolha das métricas com base no IR e GR dos dados do dataset:

Metricas de fairnes EQ e PPP nao aparentam lidar bem com um reduzido numero da classe positiva

Já as métricas PE e NPP lidam melhor com datasets com um reduzido numero de classe positiva mas aparentam maiores dificuldade em datasets com um elevado numero de classe positiva

Para as métricas AE e SP a variação da quantidade de casos da classe positiva não aparenta grande impacto no seu desempenho

O mesmo foi verificado para outros datasets como o caso de creditcard

Uma imagem com texto, file, Paralelo, Gráfico

Descrição gerada automaticamenteUma imagem com texto, file, Gráfico, Paralelo

Descrição gerada automaticamente

Ou bank

Uma imagem com texto, file, Paralelo, Gráfico

Descrição gerada automaticamenteUma imagem com texto, file, Gráfico, diagrama

Descrição gerada automaticamente

Métricas complexidade com adult race

Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

Uma imagem com texto, captura de ecrã, Paralelo, file

Descrição gerada automaticamente Uma imagem com texto, captura de ecrã, Paralelo, file

Descrição gerada automaticamente

R value variar IR borderline varir IR

Para a variação de GR não parece existir uma diferença tao notória. Algo de notar é que o intervalo de valores na complexidade R value e borderline é maior quando o IR varia, ou seja, o rácio classe positiva classe negativa aparenta impactar mais as métricas de complexidade do que o rácio de classe protegida, classe não protegida

Das métricas de complexidade analisadas as R value e borderline são as que mais se distinguem nos casos em que existe unfairness

O mesmo se verifica noutros datasets como creditCard  
Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

Métrica de performance f1

Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

Uma imagem com texto, captura de ecrã, número, Tipo de letra

Descrição gerada automaticamente

No caso da métrica de performance f1 quanto maior o valor de IR (maior numero de classe positiva do que negativa) maior tende a ser o valor da performance.

As alterações ao valor de GR não aparentao ter grande efeito na performance

Semelhante às métricas de complexidade as alterações no IR aparentam ter mais impacto na perfomance