

# Problema e Avaliação de Cenário:

A Fase de Acolhimento de Recurso sobre Processo Administrativo Previdenciário é essencial para celeridade dos julgamentos. Nessa perspectiva, foram metodificados e organizados vários passos para sua sistematização e automatização, sendo eles:

- coleta de dados, integração c/ bases de dados;
- geração de documentos e encaminhamento para o CRPS, CAJ ou JUNTA, dependendo do caso.

A automação permitiu otimizar a eficiência e acumular experiências; contudo, implicou no desmonte de setores e, mesmo obtendo sucesso em 91%, ainda represa 9% (resíduos para ação humana).

- **Proposta deste projeto:** desenvolver uma solução de ML para detectar e direcionar processos, evitando represamento de resíduos e garantindo inclusão dentro do prazo limite para criação de grupos de trabalho.

# Painel de Recursos

Tipo Benefício

Grupo Espécies

Espécies

Competência Encaminhamento

Lote

Origem Lote

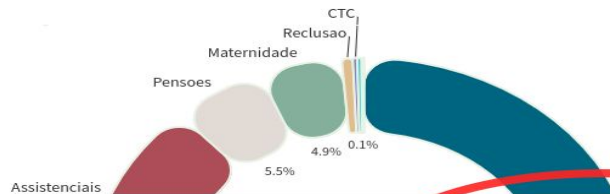
Total Processos Trabalhados

1.841.528 <sup>7.093.175</sup>  
Processamento

Total Processos Encaminhados

1.361.351

do por Grupo



Total Encaminhado por SR



Total Processos

125.718

*Total de Represados*

Processos Sem Automação

45.229

*Sem Tratamento*

Processos Sem Automação - Total Espécie

Processos Sem Automação - Total Superintendência Regional

Processos Trabalhados

1.841.528 <sup>7.093.175</sup>  
Execuções

Total com Ocorrências

5.441.726

Percentual de Ocorrências

76,72%

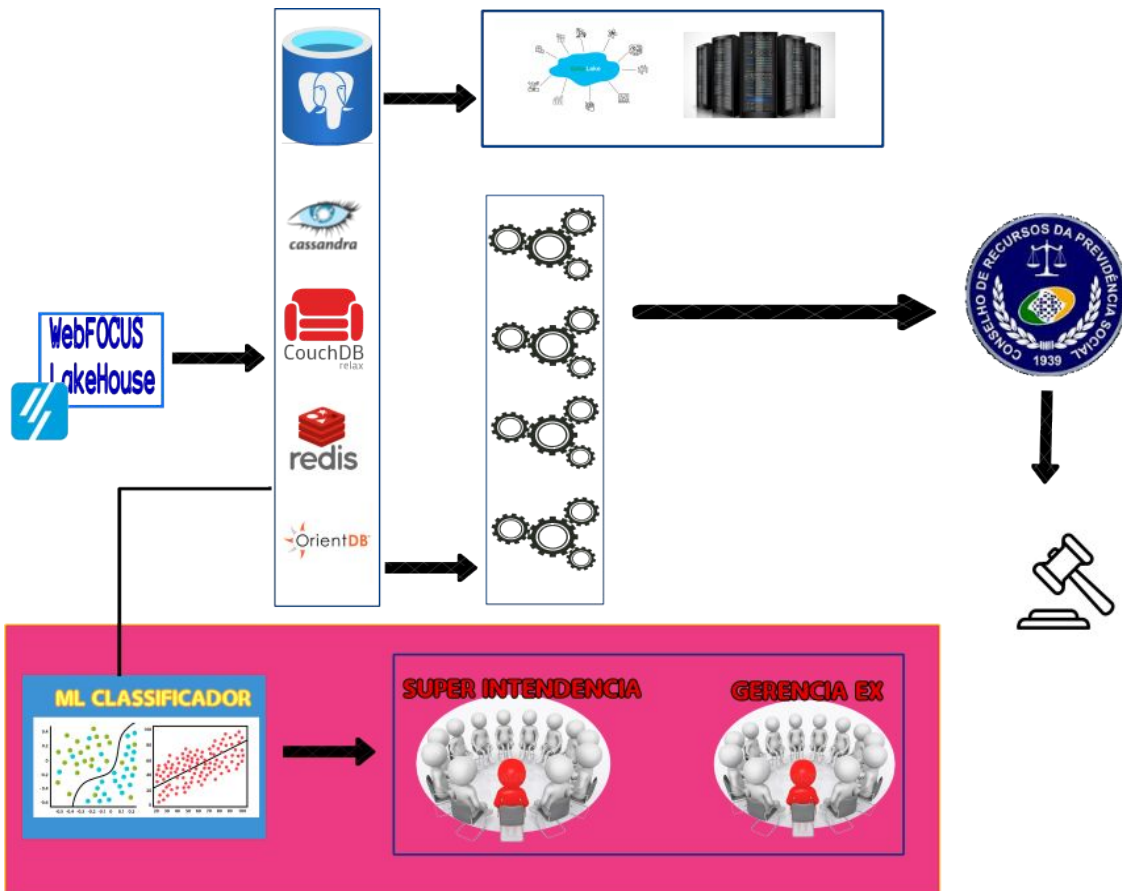
Ocorrências com Encaminhamentos - Últimos 5 Lotes

Legenda:  
■ Elegível sem Encamin...  
■ Fechado Forçosa...  
■ Falha Relatório  
■ Inelegível para Encami...  
■ Encaminhados

Percentual Ocorrências com Encaminhamentos



# Fluxo futuro com a Solução de Classificação em vermelho



## Notas de Implementação

A solução seria utilizar um mecanismo classificador em ML para interceptar o processamento do dado no pipeline e colocar uma label nas categorias :

**automatizável sem problemas**  
**necessária análise por servidor**  
**envio para definição política**  
**envio para grupo especializado**

Desta maneira, conseguiria enviar ao grupo correto os que sejam classificados para tal e continuar com o fluxo padrão.

O envio sendo feito em tempo hábil, mesmo que com pequenas falhas, permitiria tratá-los e acompanhá-los sem refazer a estrutura que já existe.

## Obtenção dos Dados e preparação da Massa com label's para os experimentos

Em parceria com a SR/SP, coletamos dados dos últimos dois anos. Inicialmente com 39 variáveis independentes para cada processo, Ex: *sub\_status*, *status\_final*, *eh\_elegivel*, *regra\_elegibilidade*, *processou*, *houve\_falha*, *inelegivel*, *eh\_primeira\_entrada*, *pesquisou\_tarefa\_principal*, *resposta\_pergunta*, *anexou\_arquivo*, *tem\_subtarefa\_instrucao*, *situacao\_encaminhamento*, *detalhe\_encaminhamento*, *tipo\_status\_envio*, *processou\_encaminhamento*, etc...

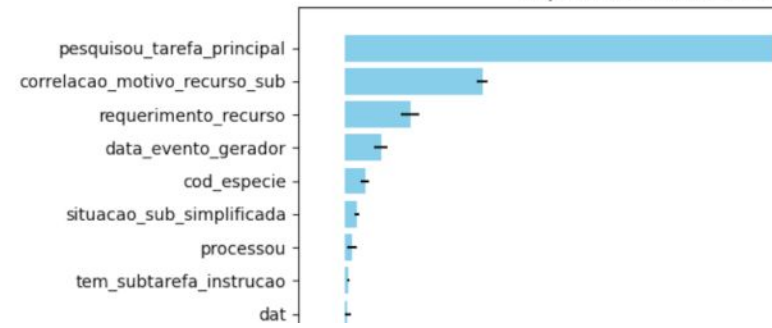
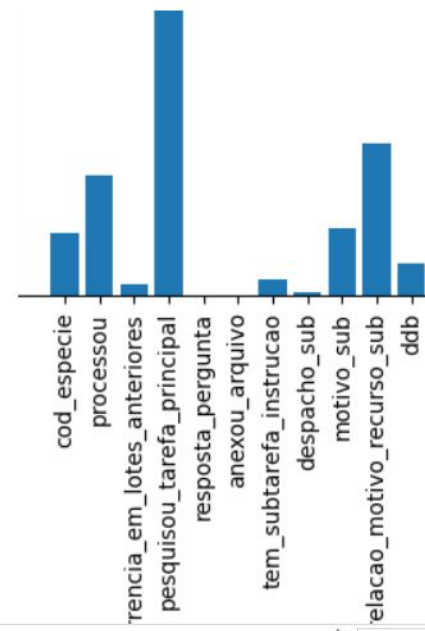
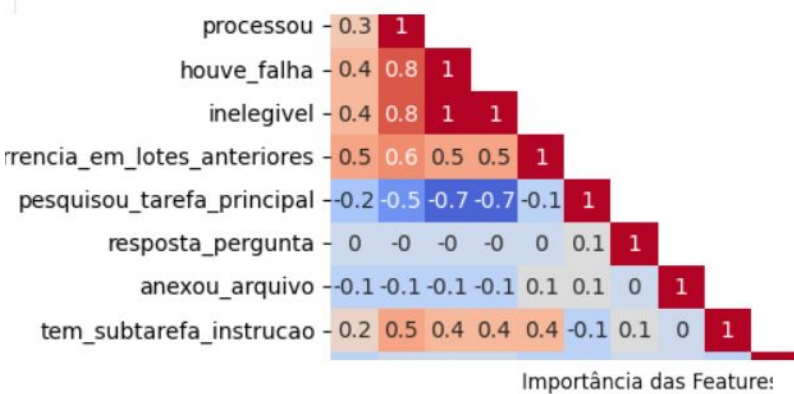
A partir dos experimentos (testes de importância do quocientes no cálculo das classes pelos modelos de Regressão linear e RandomForest, matrizes de correlação e experimentação empírica), as variáveis foram reduzidas/otimizadas Chegamos a 17.

Depois prospectou-se o tratamento e codificação para as variáveis não numéricas e o escalonamento para as variáveis numéricas.



	feature	importance
6	pesquisou_tarefa_principal	6.052116
3	houve_falha	1.050084
7	resposta_pergunta	0.149487
5	recorrencia_em_lotes_anteriores	0.104809
13	motivo_sub	0.005433
10	processou_encaminhamento	0.000000
11	erro_encaminhamento	0.000000
1	qtd_dias_inss	-0.001948
0	cod_especie	-0.006170
12	despacho_sub	-0.028662
8	anexou_arquivo	-0.207385
4	inelegivel	-1.570235
2	processou	-3.775700
9	tem_subtarefa_instrucao	-4.812946

feature	importance
processou	4.499878
tem_subtarefa_instrucao	3.336610
anexou_arquivo	1.099354
despacho_sub	0.049886
cod_especie	0.028485
motivo_sub	0.013499
erro_encaminhamento	0.000000
processou_encaminhamento	0.000000
qtd_dias_inss	-0.000288
recorrencia_em_lotes_anteriores	-0.090934
resposta_pergunta	-0.170469
inelegivel	-2.147676
houve_falha	-2.967028
pesquisou_tarefa_principal	-6.089483



## Engenharia de Features

Gráficos que demonstram o impacto Negativo das variáveis independentes processadas pelo meu modelo, que tinham forte correlação (por serem fruto de processamento e cruzamento nos meus pipelines), principalmente “processamento tarefa principal” e “houve\_falha” (ambas retiradas)

Ademais, foi inserida uma variável nova chamada **situacao\_comp** que é uma relação entre duas variáveis e que melhorou a predição do modelo

# Escolha dos Modelos e Treinamento:

Foram realizados cerca de 60 experimentos (tentativas de obtenção de dados com as mais diversas variações).

Após os treinamentos iniciais, removemos mais algumas variáveis para melhorar a generalização do modelo, aumentando o equilíbrio entre as variáveis preditoras.

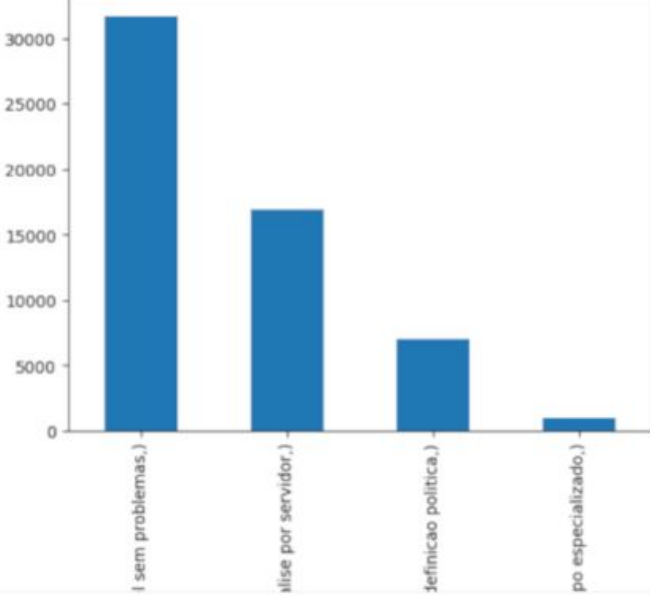
O modelo de regressão logística foi abandonado em detrimento do Random Forest devido à sua melhor performance e precisão, e o número de classes preditas foi reduzido para 3 e 2 classes para melhor ajuste do treinamento.

Verificou-se que com 3 classes, o modelo previa bem a classe "necessária análise por servidor", mas apresentava dificuldade ao tratar das 2 classes restantes.

Por isso, elas foram fundidas em uma única classe. Porém, essa abordagem, na prática, não contribui do ponto de vista do negócio, pois irá requerer intervenção humana no futuro para separá-las.

Desenvolvemos uma abordagem híbrida: um modelo binário para prever "automatizável sem problemas" com alto acerto e modelos de três classes para prever as demais categorias, melhorando a precisão e eficiência da predição.





scores desbalanceados por validacao cruzada  
 [0.97876806 0.9798817 0.97976416 0.98053852 0.98042546]  
 0.9798755804968902  
 f1 score desbalanceado  
 0.9830193685327673

relatorio de classificacao de classes desbalanceadas

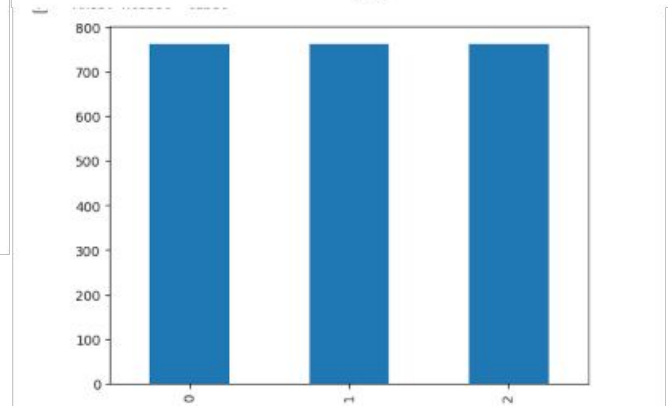
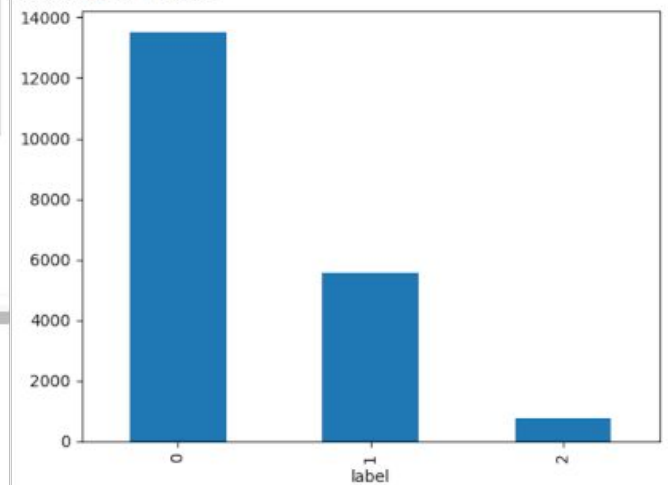
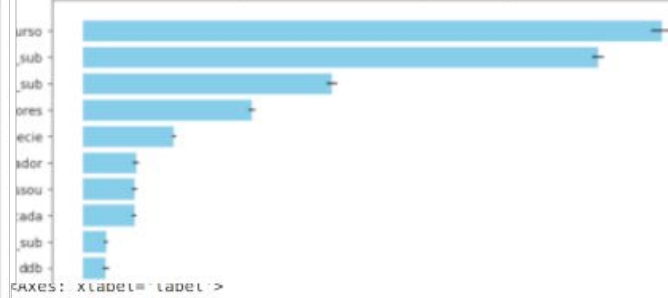
	precision	recall	f1-score	support
0	0.98	0.98	0.98	6340
1	0.98	0.98	0.98	4967

accuracy			0.98	11307
macro avg	0.98	0.98	0.98	11307
weighted avg	0.98	0.98	0.98	11307

valores de probabilidade por classe prevista

5.1.1 Graficos de importancia de *features*



## Resultado Eng. Features

Nos Gráficos aparecem a melhoria no equilíbrio das variáveis preditores e também o desbalanceamento das classes tanto em quatro como em três classes. Existe também um score relativo a implementação de um modelo binário. O desbalanceamento das classes foi tratado com sub amostragem das classes majoritárias.

```

print("binario")
print("SHAPE", X_train.shape)
print("F1 Score(macro):", f1_score(y_test_bin, y_pred_bin, average='macro')
print("F1 Score(weighted)", f1_score(y_test_bin, y_pred_bin, average='wei
print("F1 Score(micro)", f1_score(y_test_bin, y_pred_bin, average='micro')

```

```

print("Modelo Intermediário")
print("SHAPE", xx_train.shape)
print("F1 Score(macro):", f1_score(yy_test, y_pred_1, average='macro')
print("F1 Score(weighted)", f1_score(yy_test, y_pred_1, average='weighted')
print("F1 Score(micro)", f1_score(yy_test, y_pred_1, average='micro')

```

```

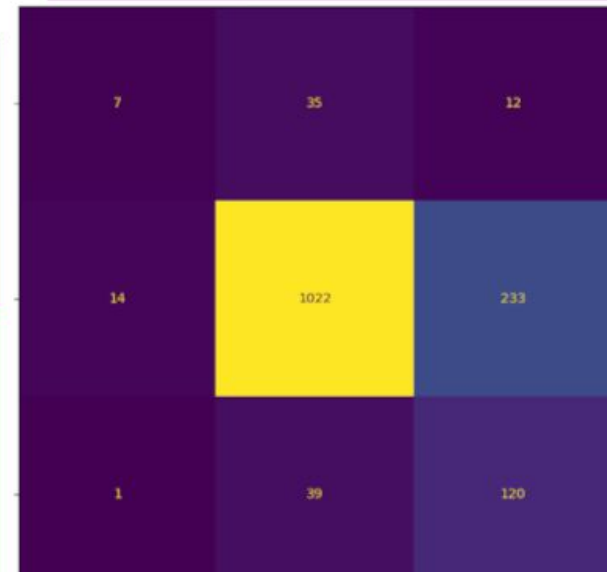
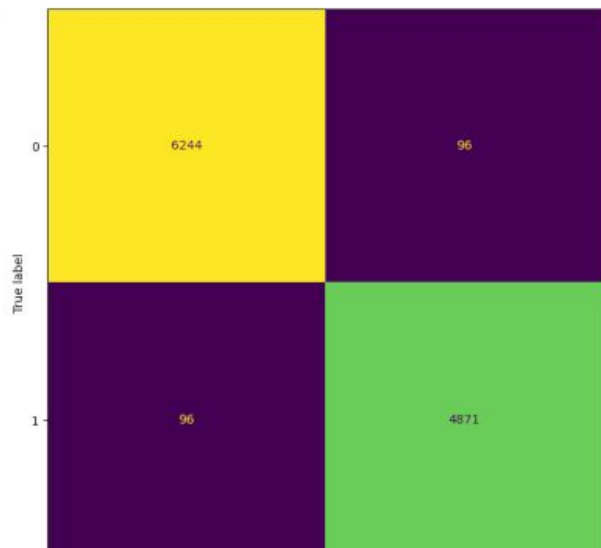
print("Modelo Final")
print("SHAPE", X_train_u.shape)
print("F1 Score(macro):", f1_score(yy_test[y_pred_1 != 0], y_pred_2, averag
print("F1 Score(weighted)", f1_score(yy_test[y_pred_1 != 0], y_pred_2, ave
print("F1 Score(micro)", f1_score(yy_test[y_pred_1 != 0], y_pred_2, averag

```

```

binario
SHAPE (45228, 14)
F1 Score(macro): 0.9827652411277206
F1 Score(weighted) 0.9838193685327673
F1 Score(micro) 0.9838193685327673
Modelo Intermediário
SHAPE (19868, 14)
F1 Score(macro): 0.7968584968184479
F1 Score(weighted) 0.9318081816904303
F1 Score(micro) 0.9341654922488424
Modelo Final
SHAPE (2289, 14)
F1 Score(macro): 0.5818746655221563
F1 Score(weighted) 0.7955838380726475
F1 Score(micro) 0.7747888496291382

```



## Resultado Classe à Classe

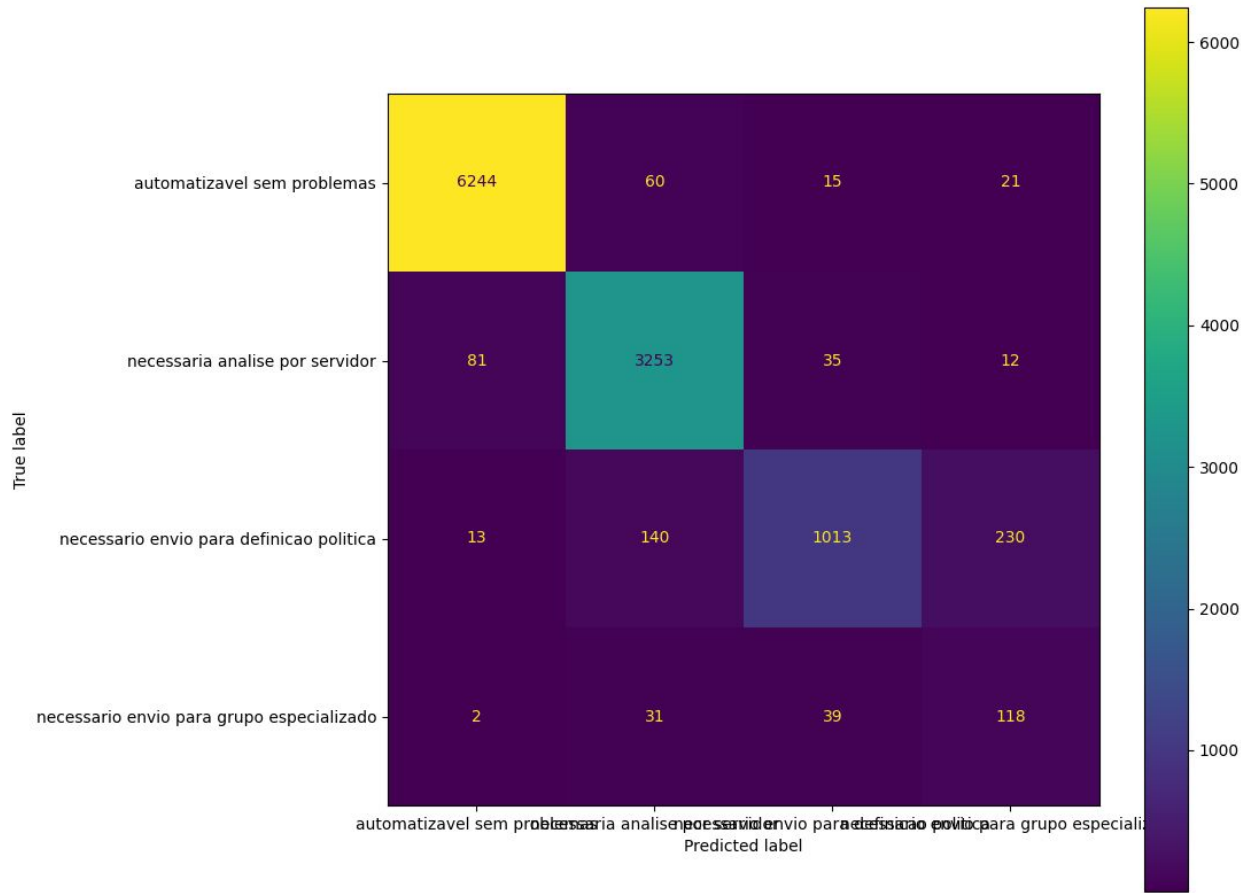
Demonstração das Matrizes de Confusão de um modelo binário e de dois modelos de três classes em que um deles foi treinado com classes desbalanceadas e outro com classes balanceadas . O modelo escolhido foi o XGBOOST

Com ajustes feitos pela melhor pontuação do Grid Searchcv para os seguintes hiperparâmetros:

`'min_child_weight', 'eta', 'gamma', 'max_depth', 'n_estimators'`

Os nomes dos modelos estão pensados como modelo de encadeamento sendo o binário a entrada da previsão o intermediário o modelo de 3 classes treinado sem balanceamento de classes e o final o de três classes com balanceamento

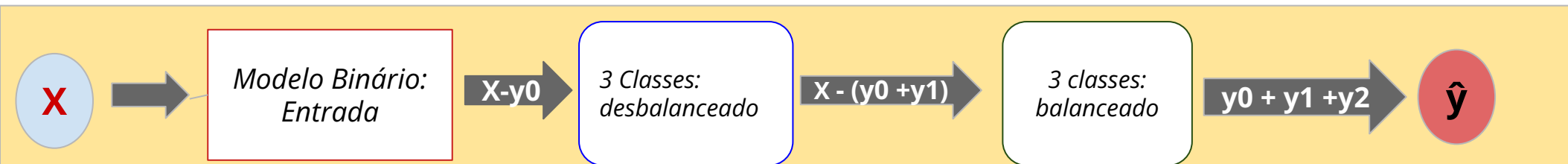




## Resultado Consolidação

A matriz mostra os resultados e f1 score. Os resultados poderiam ser melhores caso eu trate o vazamento entre os modelos ( usando o segundo ranqueado no predict\_proba) contudo isso ficará para outro momento.

↔ F1 Score(weighted): 0.942685613309813  
 F1 Score(macro): 0.7892306548864759  
 F1 Score(micro): 0.9399487043424427



# Considerações finais

Ao final do bootcamp a conclusão é que esse modelo pode ser usado para otimizar os processos internos de direcionamento de tarefas ( precisa ser testado com o tempo) e que a aplicação dele hoje, mesmo sendo inicial, pode dar ótimos frutos ( encaminhei alguns em que apliquei a previsão, que estavam fora da massa, e eles estão sendo tratados e não foram devolvidos (ainda não me responderam contudo se fosse erro crasso já teriam devolvido) ).

Também ao avaliar os pesos da previsão verifiquei algumas correlações interessantes ( caso da ddb e situacao\_sub com o despacho oque significa que durante um tempo determinado despacho feito levou a analises equivocadas da instrução de recurso além de várias correlações que estudarei)

Tive insights de como usar esse tipo de modelos no acolhimento da análise inicial podendo gerar uma melhor qualidade de análise através do direcionamento para especialistas quando necessário ( especialista em aposentadoria com nível de complexidade maior etc..) além do uso para analisar a qualidade de despachos e análises de processos iniciais.

Publicamos o caderno e o dataset anonimizado no github/gitlab

[https://github.com/claudiomarcosasilva/apresentacao\\_bootcamp\\_enap\\_2024](https://github.com/claudiomarcosasilva/apresentacao_bootcamp_enap_2024)

[https://gitlab.com/claudiomarcoz/apresentacao\\_bootcamp\\_enap\\_2024](https://gitlab.com/claudiomarcoz/apresentacao_bootcamp_enap_2024)

