**TCC - Felipe Assad e Jorge Chagas - (09/05/2019)**

**Objetivo da atividade:**

Analisar a eficiência de alguns métodos de aprendizagem para a base com todos os dados.

**Base de entrada:**

todos\_os\_dados

**Treinamento:**

NUM\_BO fora o intervalo de 200 a 500 (21073 instâncias)

1. Dolosos: 651
2. Não dolosos: 20422

**Teste:**

NUM\_BO entre 200 e 500 (5361 instâncias)

1. Dolosos: 250
2. Não dolosos: 5111

**Target:**

Doloso

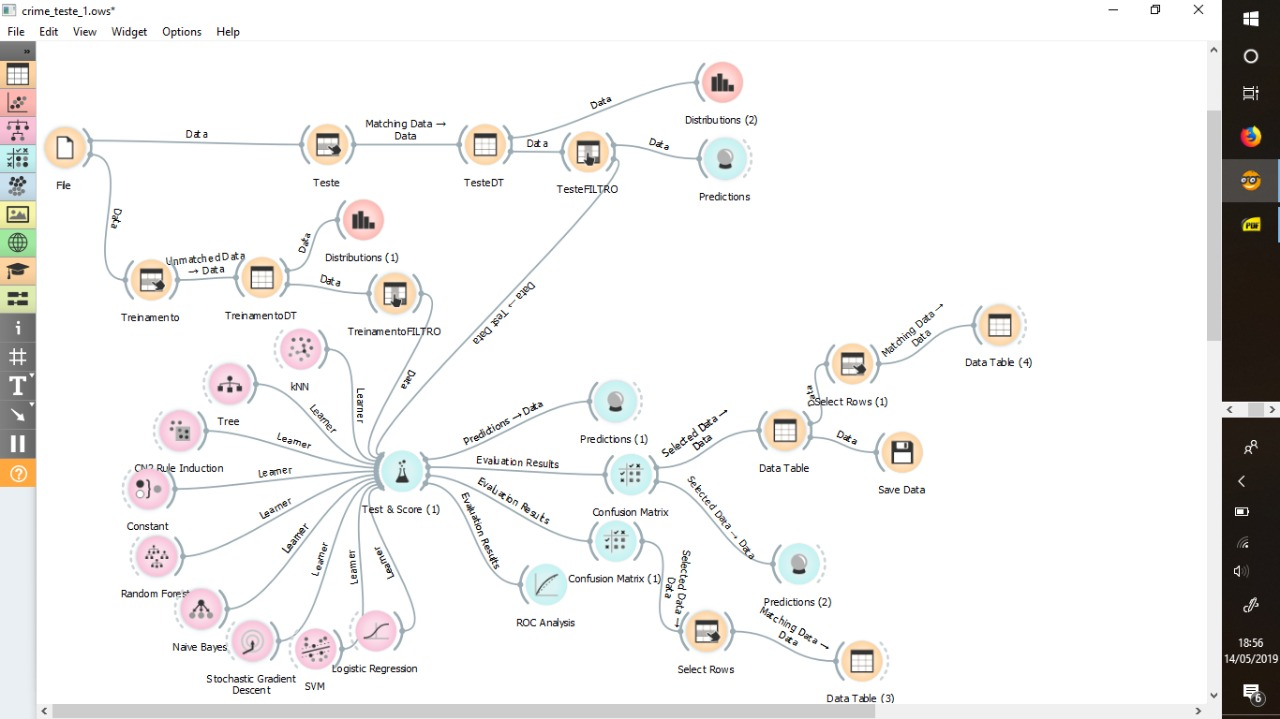
**Features:**

1. PERIODOOCORRENCIA
2. DESCRICAOLOCAL
3. IDADE

**Descrição:**

Dados de treinamento e teste misturados entre doloso e não doloso.

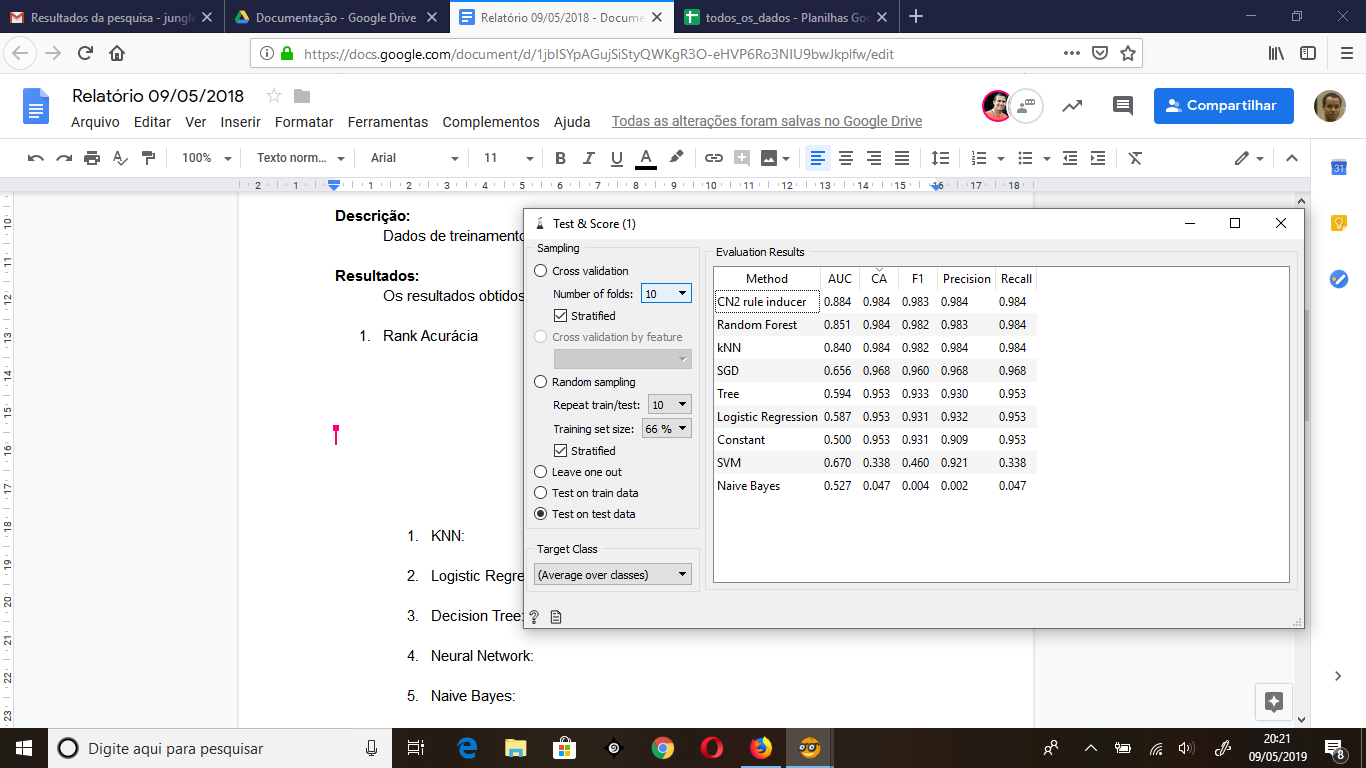
**Workflow:**



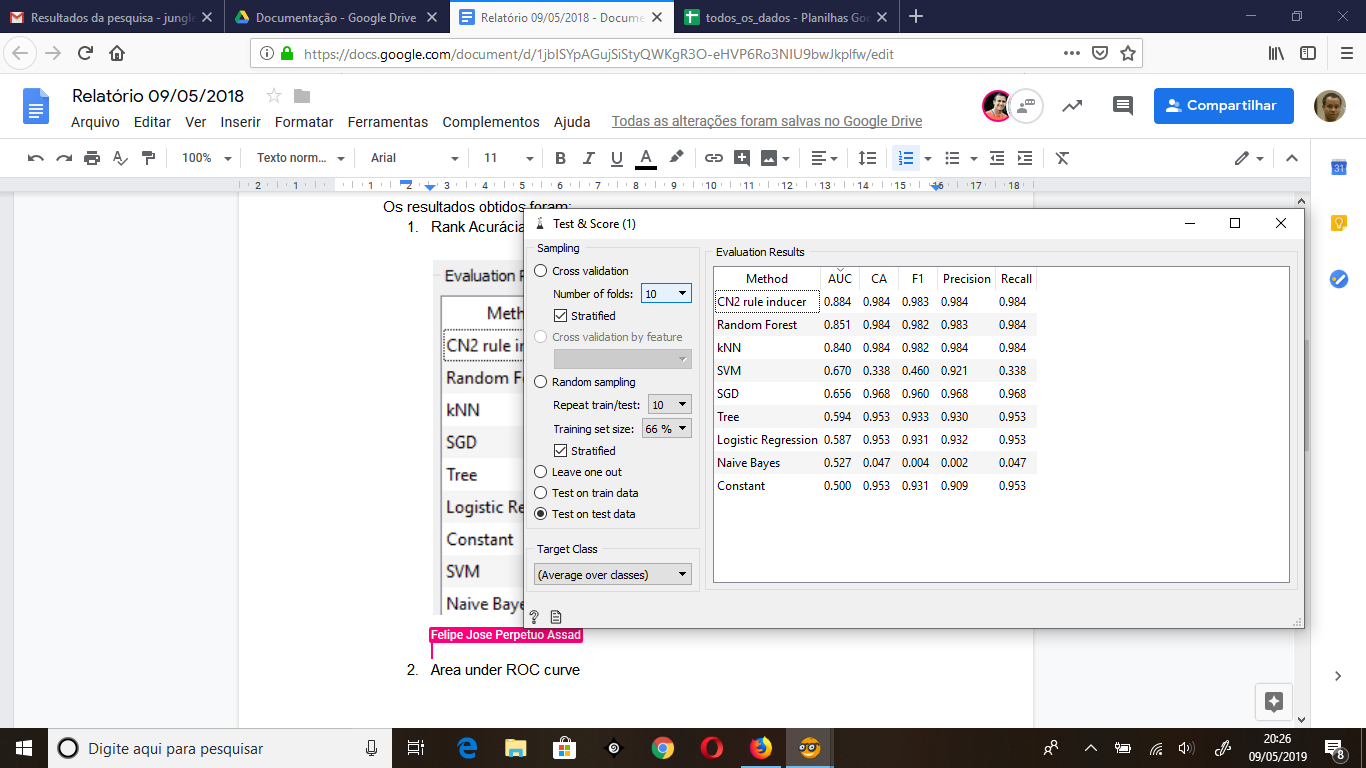
**Resultados:**

Os resultados obtidos foram:

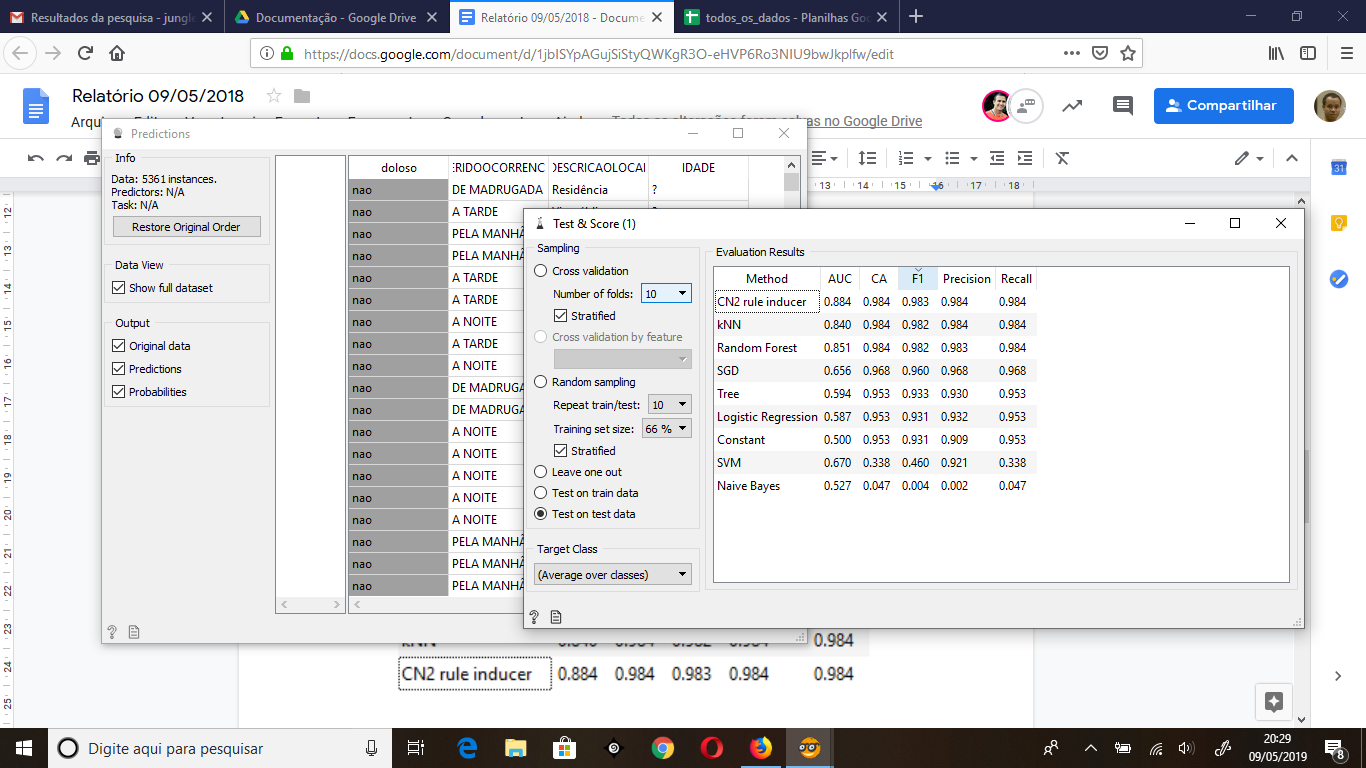
1. Rank Acurácia



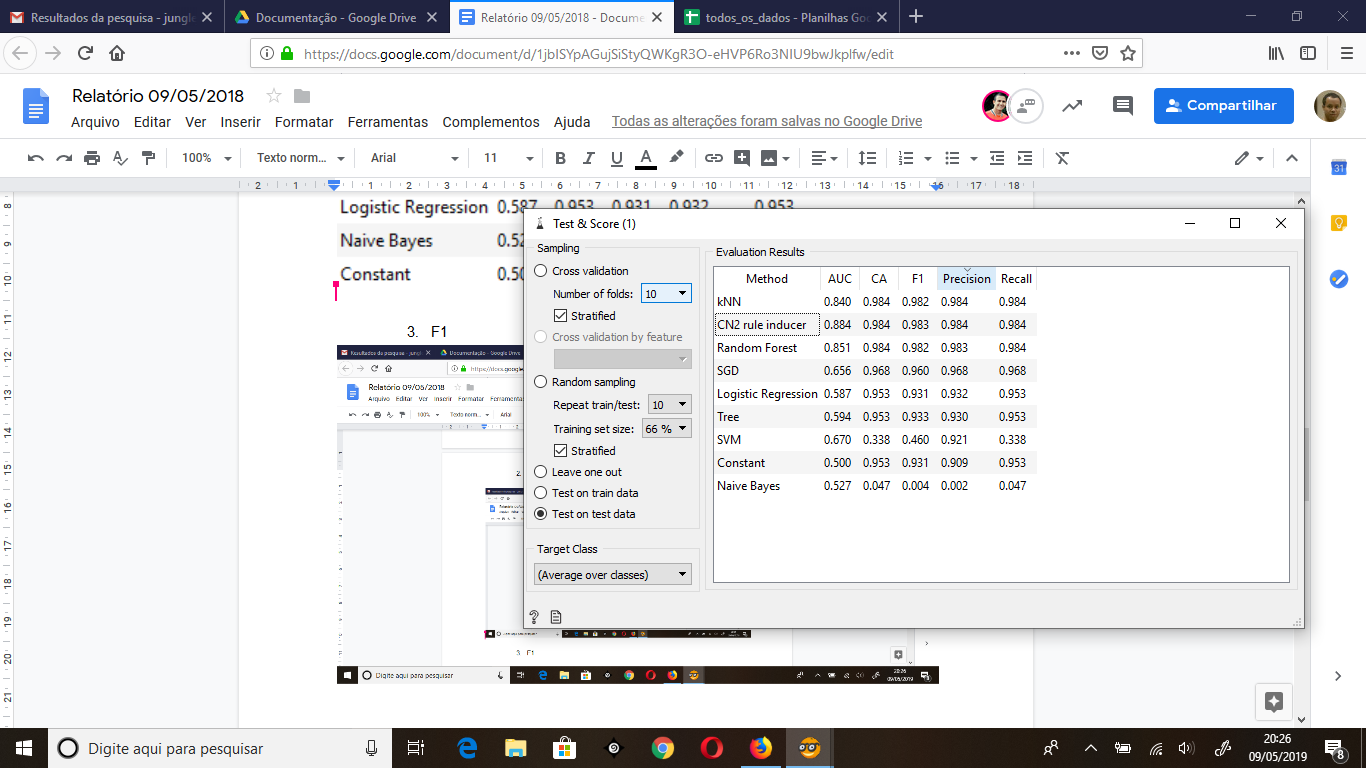
1. Area under ROC curve



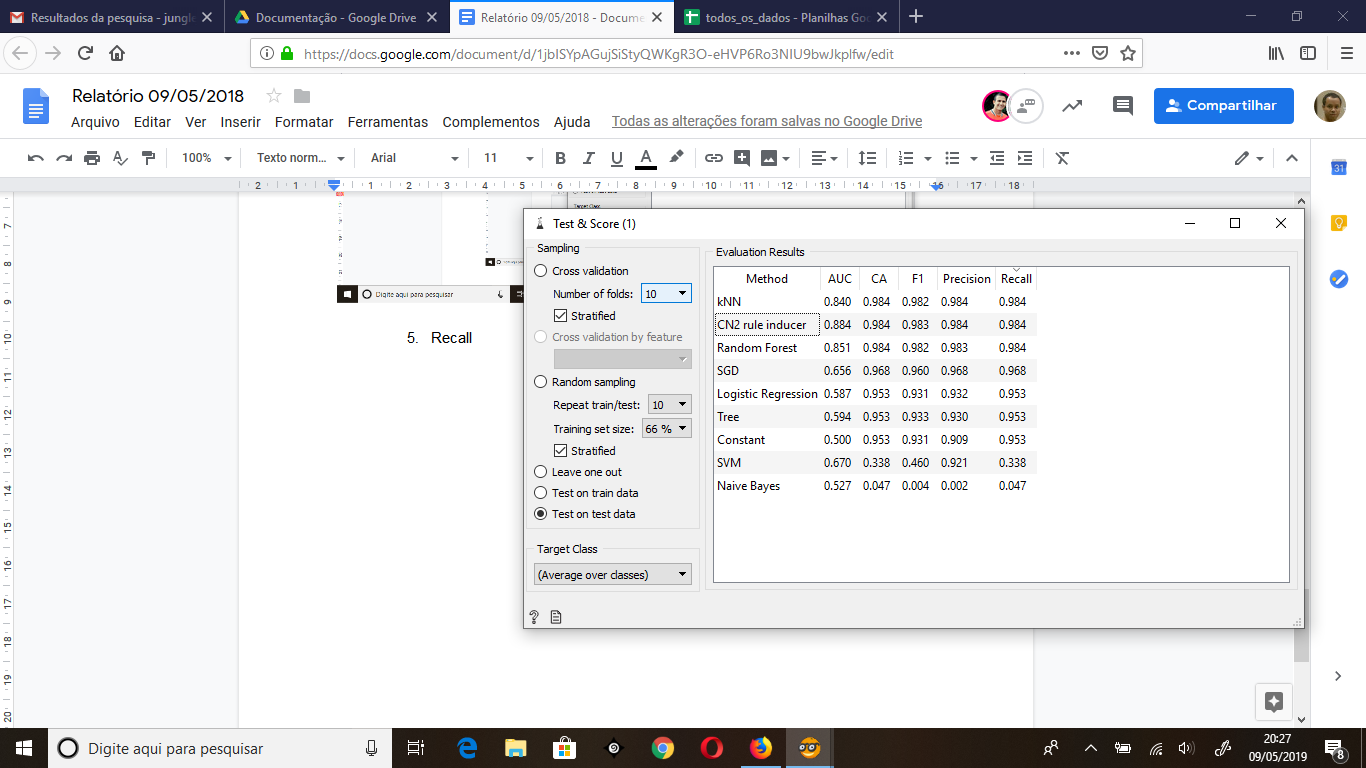
1. F1



1. Precision



1. Recall



**Conclusão:**

Estamos usando poucos atributos e isso não nos permite chegar numa conclusão com sentido, pois a ausência ou a presença da idade não é determinístico no processo de acerto/erro. Precisamos incluir novos atributos ao processo.

