**Relatório da disciplina de datawarehouse e datamining**

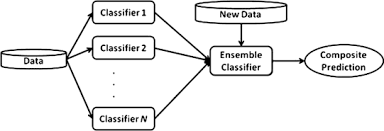
**Trabalho sobre comitê de classificadores.**

Rodrigo Rafael da Maceno de Souza

**Comitê de classificadores**

Um comitê de classificadores é um conjunto de classificadores que são utilizados para realizar uma predição conjunta (Imagem 1).

Imagem 1 – Exemplo de um comitê de classificador



Fonte: Felipe Augusto (2020)

Alguns motivos para se usar comitês de classificadores:

* Diferentes classificadores com diferentes inputs: classificação por voz, características, imagens entre outros;
* Diferentes características, podem existir casos onde em um conjunto de dados falte algumas informações que em outros não e vice versa;
* A união faz a força, unir vários classificadores em uma mesma tarefa pode melhorar o desempenho da classificação.

Um comitê de classificadores tende a seguir uma configuração padrão

* Vários classificadores que avaliam o conjunto de dados separadamente, usando métricas, valores, inputs distintos ou iguais;
* Um combinador que ira agrupar esses valores, seguindo alguma técnica, para o exemplo foi se utilizado o somatório dos classificadores e o produto dos mesmos.

**Implementação do algoritmo**

Primeiramente foi necessário rodar o dataset Iris nos classificadores do weka, os classificadores escolhidos foram: J48 (Imagem 2), BayesNet (Imagem 3) e RandomForest (Imagem 4), com as configurações que o professor passou, porém foi percebido que no item 1.3.5 da especificação do trabalho dizia para habilitar o Output predictions, porém para essa versão do weka, essa opção não existia, ou esta com o nome trocado, utilizei a OutputDistribution (Imagem 5), creio que seja apenas um erro na descrição, ou então o fato de estar usando a versão de MacOs.

Imagem 2 – J48

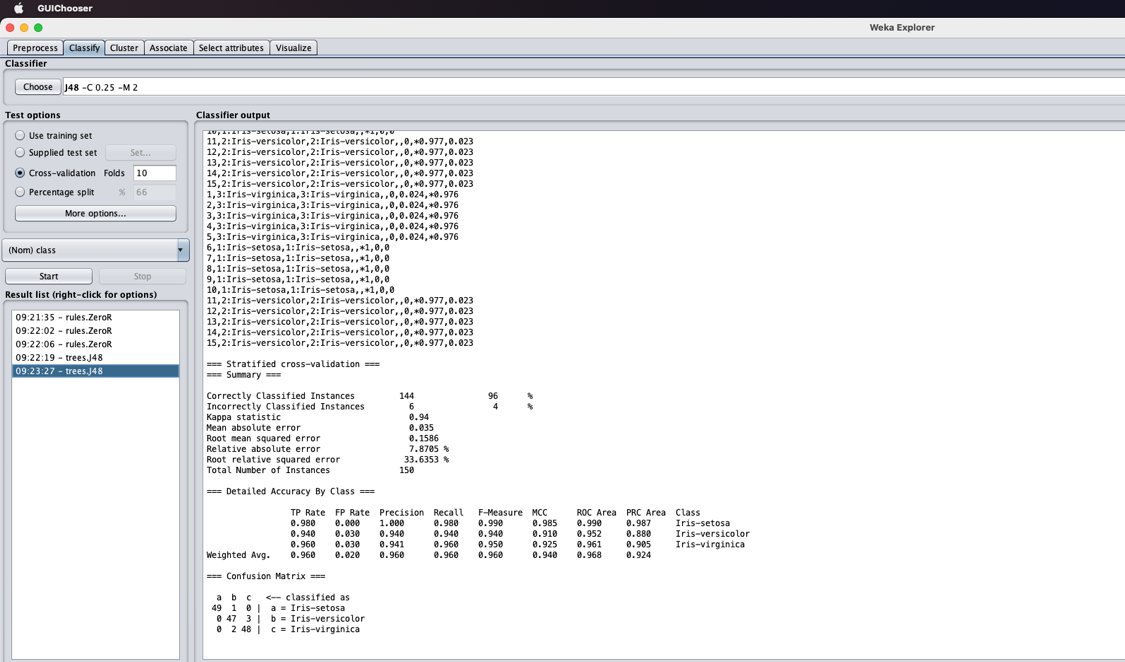


Imagem 3 - BayesNet

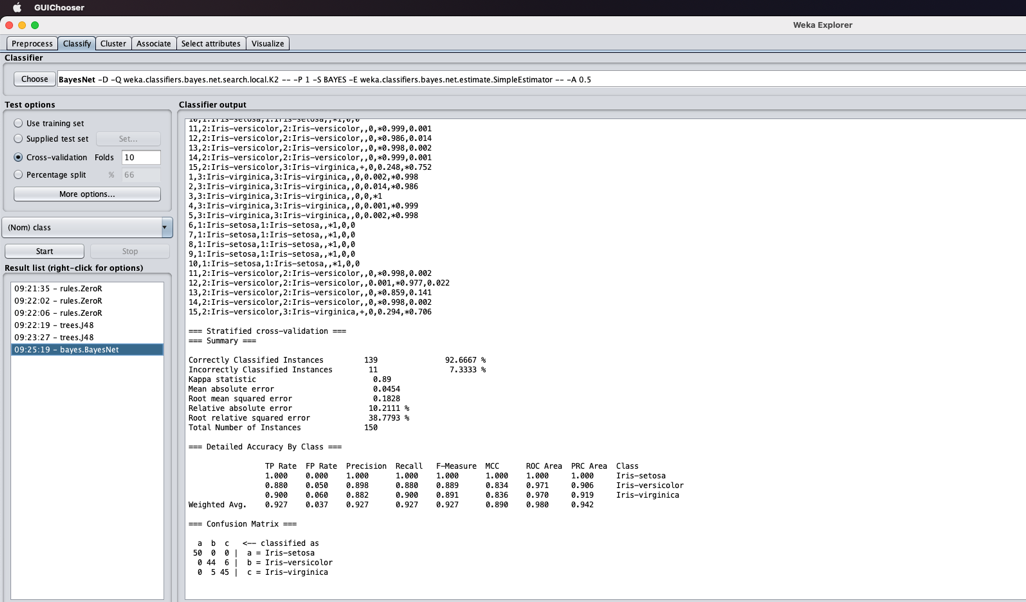


Imagem 4 - RandomForest

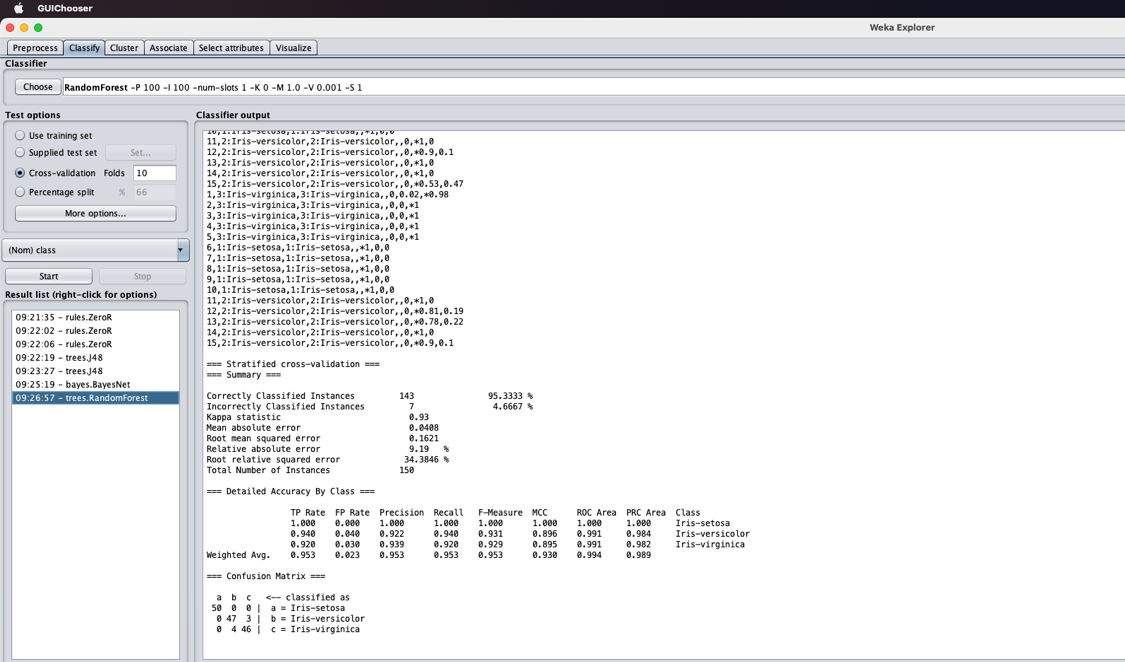
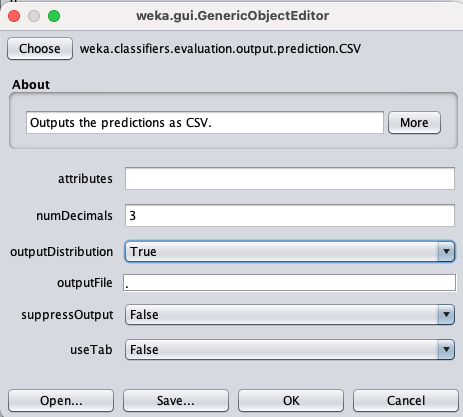


Imagem 5 - outputDistribution



Após rodados os classificadores e exportados os mesmos para CSV foi iniciado a implementação do código, inicialmente tentou-se se utilizar da biblioteca pandas para manipular os arquivos, porém por conta da falta de experiência com a mesma se optou em se utilizar apenas a csv, padrão do python.

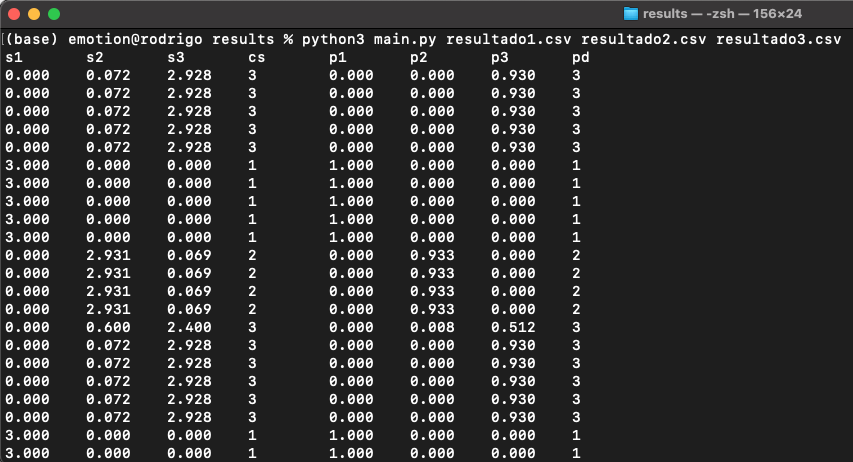
O código em si ficou bem simples, basicamente ele inicia carregando os arquivos, com base nos nomes passados via arguments: **python3 main.py resultado1.csv resultado2.csv resultado3.csv**

Com isso o algoritmo irá carregar os três arquivos, armazenar seus registros em listas, após isso o mesmo irá percorrer as listas realizando os cálculos de s1, s2, s3, cs, p1, p2 e pd. Como os datasets são iguais, o numero de registros em cada lista será o mesmo, logo foi tranquilo de implementar os cálculos. Ele realiza um somatório com os elementos das 3 listas para as variáveis correspondentes e exibe na tela.

Também foi criado um método chamado formatNumber, afim de formatar melhor as saídas.

A saída do algoritmo pode ser verificada na Imagem 6, e o código do mesmo se encontra no github[[1]](#footnote-1).

Imagem 6 – Saída do algoritmo



**Referencias**

Felipe Augusto, **Comitê de classificadores, Classificação multiclasse e multirrótulo. 2020** Disponível em: <https://medium.com/@f2acode/comit%C3%AA-de-classificadores-classifica%C3%A7%C3%A3o-multiclasse-e-multirr%C3%B3tulo-7a01d87ee9b8>. Acesso em: 14/12/2021.

1. https://github.com/rodrigorafaeldamaceno/classifiers\_committee [↑](#footnote-ref-1)