

Discente: Wilder Siqueira Medeiros

Docente: Arnaldo Candido Junior

Disciplina: Data Warehouse E Data Mining

Data: 09 de Dezembro de 2021

Relatório Do Comitê De Classificadores

Primeiramente, os dados no formato .csv são lidos e guardados em dataframes. Em seguida, são formatados para definir as colunas que serão trabalhadas e remover os caracteres especiais "*" da tabela de dados. Cada arquivo .csv estará armazenado em um dataframe, com isso, os data frames serão somados e multiplicados com o correspondente índice da LinhaXColuna de cada data frame e resultará em um novo data frame. Por exemplo, data frame 1 (df1) será somado com o data frame 2 (df2) e o resultado disso será um novo data frame que será somado com o terceiro data frame (df3) resultando na soma final dos índices correspondentes das instâncias da tabela de dados. O mesmo será realizado para o produto das probabilidades. Com isso os dados serão formatados novamente para que haja apenas 2 decimais após a vírgula, facilitando a leitura. Para a escolha, foi utilizado o método "idxmax" com parâmetro "axis=1" definido para selecionar o maior valor de cada linha entre as colunas, ou seja, dentre cada instância foi obtido o maior valor conforme as probabilidades de cada coluna. Com esse processo realizado para o data frame resultante da soma e produto, será elaborada a concatenação final dos dataframes, gerando um data frame final com os valores das somas e produtos das probabilidades de cada instância conforme cada categoria, isso pode ser observado exemplificadamente na Figura 2.

As ferramentas utilizadas foram:

- Python
 - Pandas
- Visual Studio Code
- Excel
- Weka
- Windows 10

Observação

Os arquivos .csv são lidos conforme um caminho específico do sistema operacional Windows. Para ler outros arquivos, cabe a alteração do caminho, assim como o nome dos arquivos.

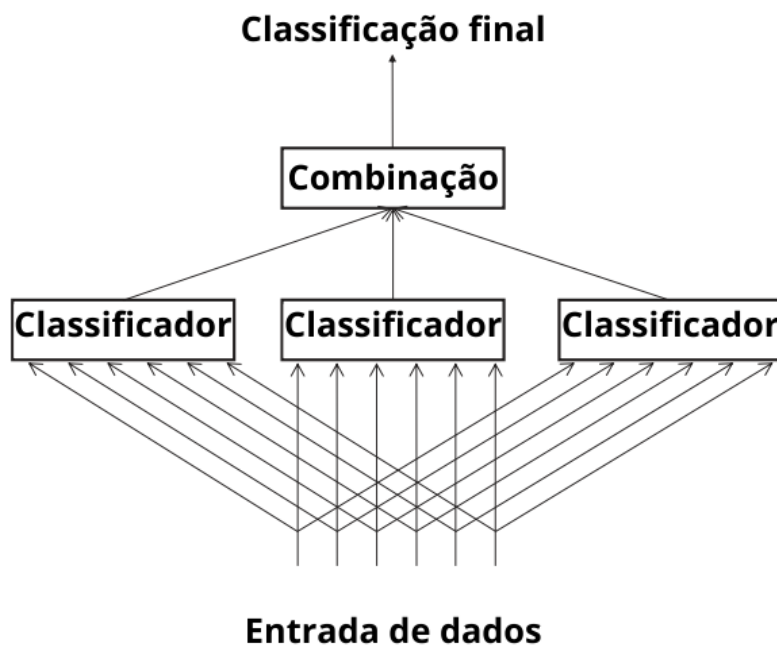
Conforme os nomes dos classificadores no aplicativo Weka, para os resultados, o arquivo "resultado.csv" é a classificação do classificador "NaiveBayes", o arquivo "resultado1.csv" é a classificação do classificador "MultilayerPerceptron" e o arquivo "resultado2.csv" é a classificação do classificador "IBk". Os três classificadores classificaram a base de dados "iris.arff" conforme a requisição do trabalho.

Comitê De Classificadores

O comitê de classificadores é um sistema que possui um conjunto de classificadores que classificam uma determinada instância e depois os resultados dessas classificações é combinado por meio de um critério em uma única classificação final dessa determinada instância KUNCHEVA (2014). Em outras palavras, um conjunto de respostas para um determinado problema é analisado e conforme os critérios de um método de combinação é realizado a escolha de uma resposta final, a Figura 1 representa esse comportamento. O objetivo desse sistema é melhorar o resultado de classificação com a

combinação das classificações individuais, obtendo uma classificação final mais robusta KUNCHEVA (2014).

Figura 1 - Comitê de Classificadores onde cada classificador classifica uma determinada instância e posteriormente essa escolha é combinada com outras classificações para obter uma classificação final



Fonte: Adaptado de Kuncheva (2014, p. 95)

Resultados

A Figura 2 exemplifica algumas saídas do data frame final com as somas e produtos das probabilidades de cada instância, assim como a escolha da categoria pertencente de cada uma. Sendo, 1 para Setosa, 2 para Versicolor e 3 para Virginica. S1, s2 e s3 são as somas dos três classificadores para as determinadas classes e cs é a classe escolhida. P1, p2, p3 e cp, são análogos, mas para o produto das probabilidades. Todas as instâncias podem ser visualizadas no arquivo "saida.csv" no repositório que acompanha o relatório.

Figura 1 - Exemplo do data frame final com as somas e produtos das probabilidades e a escolha da categoria conforme cada instância.

	s1	s2	s3	cs	p1	p2	p3	cp
0	0.00	0.01	2.99	3	0.0	0.00	0.99	3
1	0.00	0.26	2.74	3	0.0	0.00	0.76	3
2	0.00	0.09	2.91	3	0.0	0.00	0.91	3
3	0.00	0.04	2.95	3	0.0	0.00	0.95	3
4	0.00	0.01	2.99	3	0.0	0.00	0.99	3
..
145	0.01	2.97	0.02	2	0.0	0.97	0.00	2
146	0.01	2.70	0.29	2	0.0	0.72	0.00	2
147	0.01	2.45	0.54	2	0.0	0.49	0.00	2
148	0.01	2.98	0.01	2	0.0	0.98	0.00	2
149	0.01	2.82	0.16	2	0.0	0.83	0.00	2

[150 rows x 8 columns]

Fonte: Autoria Própria (2021)

Trabalhos Citados

KUNCHEVA, L.I. **Combining pattern classifiers: methods and algorithms**. John Wiley & Sons, 2014.