1 - Veri Seti

[**Veri Seti**](https://scikit-learn.org/stable/datasets/toy_dataset.html#wine-dataset)

Şarap veri seti, klasik ve çok kolay, çok sınıflı bir sınıflandırma veri setidir.

|  |  |
| --- | --- |
| Classes | 3 |
| Samples per class | [59,71,48] |
| Samples total | 178 |
| Dimensionality | 13 |
| Features | real, positive |

Bu veriler, İtalya'da aynı bölgede yetiştirilen, ancak üç farklı çeşitten elde edilen şarapların kimyasal analizinin sonuçlarıdır. Analiz, üç şarap türünün her birinde bulunan 13 bileşenin miktarını belirledi.

1) Alkol

2) Malik asit

3) Kül

4) Külün alkalinitesi

5) Magnezyum

6) Toplam fenoller

7) Flavanoidler

8) Flavanoid olmayan fenoller

9) Proantosiyaninler

10) Renk yoğunluğu

11) Ton

12) OD280 / OD315 seyreltilmiş şaraplar

13) Prolin

1) Yukarıdaki veri seti üzerinde Python’da yer alan aşağıdaki sınıflama algoritmalarını kodlayacaksınız. Verinin %70’i eğitim %30’u test için kullanılacaktır. Her model için default değerlerle çalıştırdığınız kodlarınızı model başlığı altına kopyalayınız. Her model için farklı parametrelerle aldığınız sonuçların confusion matrix(karmaşıklık matrisi)’ini yine model başlığı altına eklenecektir. Ayrıca her modelin confusion matrix’inden elde edilen metrik değerleri hesaplanmalıdır.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Modeller | Accuracy | Precision | Recall | F1 Score | Support |
| SVC-default | 0,67 | 0,51 | 0,67 | 0,57 | 54 |
| SVC-linear | 0,94 | 0,95 | 0,94 | 0,94 | 54 |
| SVC-sigmoid | 0,20 | 0,13 | 0,20 | 0,16 | 54 |
| SVC-poly | 0,67 | 0,54 | 0,67 | 0,58 | 54 |
| SVC-linear C=10 | 0,96 | 0,97 | 0,96 | 0,96 | 54 |
| SVC-linear C=0,1 | 0,98 | 0,98 | 0,98 | 0,98 | 54 |
| SVC-linear C=0,1 Gamma = ‘auto’ | 0,98 | 0,98 | 0,98 | 0,98 | 54 |
| DecisionTreeClassifier-defauld | 0,96 | 0,97 | 0,96 | 0,96 | 54 |
| DecisionTreeClassifier criterion=’entropy’ | 0,91 | 0,91 | 0,91 | 0,91 | 54 |
| DecisionTreeClassifier criterion=’ log\_loss’ | 0,93 | 0,93 | 0,93 | 0,93 | 54 |
| DecisionTreeClassifier criterion= ‘gini’ random\_state = 42 min\_samples\_split = 4 | 0,96 | 0,97 | 0,96 | 0,96 | 54 |
| KNeighborsClassifier-default | 0,72 | 0,74 | 0,72 | 0,73 | 54 |
| KNeighborsClassifier- n\_neighbors=8 | 0,72 | 0,77 | 0,72 | 0,73 | 54 |
| KNeighborsClassifier- n\_neighbors=8 weights='distance' | 0,72 | 0,75 | 0,72 | 0,73 | 54 |
| GaussianNB-default | 1,00 | 1,00 | 1,00 | 1,00 | 54 |
| GaussianNB- priors=None | 1,00 | 1,00 | 1,00 | 1,00 | 54 |
| GaussianNB- priors=None var\_smoothing=1e-6 | 1,00 | 1,00 | 1,00 | 1,00 | 54 |
| MLPClassifier-default | 0,52 | 0,49 | 0,52 | 0,44 | 54 |
| MLPClassifier(hidden\_layer\_sizes=(45,30,15),solver='sgd',learning\_rate\_init=0.01,max\_iter=500) | 0,28 | 0,08 | 0,28 | 0,12 | 54 |
| MLPClassifier(hidden\_layer\_sizes=(62,28,40),solver='adam',learning\_rate\_init=0.01,max\_iter=500) | 0,59 | 0,66 | 0,59 | 0,57 | 54 |
| MLPClassifier(hidden\_layer\_sizes=(62,28,40),solver='sgd',learning\_rate\_init=1000,max\_iter=500) | 0,33 | 0,11 | 0,33 | 0,17 | 54 |
| RandomForestClassifier-default | 1,00 | 1,00 | 1,00 | 1,00 | 54 |
| AdaBoostClassifier-default | 0,87 | 0,88 | 0,87 | 0,87 | 54 |
| svc=SVC(probability=True, kernel='linear')  AdaBoostClassifier | 0,87 | 0,88 | 0,87 | 0,87 | 54 |
| BaggingClassifier-default | 0,98 | 0,98 | 0,98 | 0,98 | 54 |
| svc=SVC(probability=True, kernel='linear')  BaggingClassifier | 0,96 | 0,96 | 0,96 | 0,96 | 54 |
| GradientBoostingClassifier-default | 0,96 | 0,97 | 0,96 | 0,96 | 54 |
| estimators = [('rf', RandomForestClassifier()), ('gbc', GradientBoostingClassifier())]  VotingClassifier(estimators=estimators, voting='hard', weights=None, n\_jobs=None, flatten\_transform=True, verbose=False) | 1,00 | 1,00 | 1,00 | 1,00 | 54 |