AF-SM-04\_05.2020.G

|  |  |
| --- | --- |
| **კურსის სახელწოდება:** | ხელოვნური ინტელექტი და მანქანური სწავლების საწყისები |
| **ლექტორი:** | დავით დათუაშვილი |
| **სტუდენტი:** |  |

1. მოცემულია ფაილი Quiz\_2.csv, რომელშიც დამოკიდებული ცვლადი არის target , ხოლო ყველა სხვა ცვლადი არის დამოუკიდებელი ცვლადი, უპასუხეთ ქვევით მოცემულ კითხვებს(10 ქულა)

* დაყავით მონაცემები სატრენინგოდ და სატესტოდ და გამოიყენეთ GaussianNB ის ალგორითმი და გამოთვალეთ როგორც სატრენინგო , ასევე სატესტო Score (4 ქულა)

*from* sklearn.naive\_bayes *import* GaussianNB  
*import* pandas *as* pd  
*from* sklearn.model\_selection *import* train\_test\_split  
  
  
data = pd.read\_csv("Quiz\_2.csv")  
  
y = data['target'].values  
X = data.drop('target', axis=1).values  
  
  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=5)  
  
model = GaussianNB()  
model.fit(X\_train, y\_train)  
  
print(model.score(X\_test, y\_test))  
print(model.score(X\_train, y\_train))

სატესტო ქულა - 0.8933333333333333

სატრენინგო ქულა - 0.9033333333333333

* ფუნქცია value\_counts() ის მეშვეობით გაიგეთ თავდაპირველ target სვეტში როგორ არის კლასები განაწილებული პროპროციულად(normalize=True დამხმარე ფუნქციის მეშვეობით value\_counts() ში) და GaussianNB ალგორითმში priors პარამეტრის მეშვეობით გაუწერეთ სასურველი ალბათობები და შეამოწმეთ როგორ შეიცვლება შედეგები. აღწერეთ მიღებული შედეგი (6 ქულა)
* model = GaussianNB(priors=[0, 1])  
  model.fit(X\_train, y\_train)  
    
  print(model.score(X\_test, y\_test))  
  print(model.score(X\_train, y\_train))  
    
  print(data['target'].value\_counts(normalize=*True*))

priors=[0,1] შემთხვევაში :

0.15222222222222223

0.1619047619047619

0 0.841

1 0.159

Name: target, dtype: float64

Priors=[0.9, 0.1] შემთხვევაში :

0.8877777777777778

0.900952380952381

0 0.841

1 0.159

Name: target, dtype: float64

Priors=[0.5, 0.5] შემთხვევაში :

0.8322222222222222

0.8419047619047619

0 0.841

1 0.159

Name: target, dtype: float64

მთლიანი კოდი

*from* sklearn.naive\_bayes *import* GaussianNB  
*import* pandas *as* pd  
*from* sklearn.model\_selection *import* train\_test\_split  
  
  
data = pd.read\_csv("Quiz\_2.csv")  
  
y = data['target'].values  
X = data.drop('target', axis=1).values  
  
  
X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=5)  
  
model = GaussianNB(priors=[0.5, 0.5])  
model.fit(X\_train, y\_train)  
  
print(model.score(X\_test, y\_test))  
print(model.score(X\_train, y\_train))  
  
print(data['target'].value\_counts(normalize=*True*))