**MÜCAHİT TAHA AZ**

**ENM593**

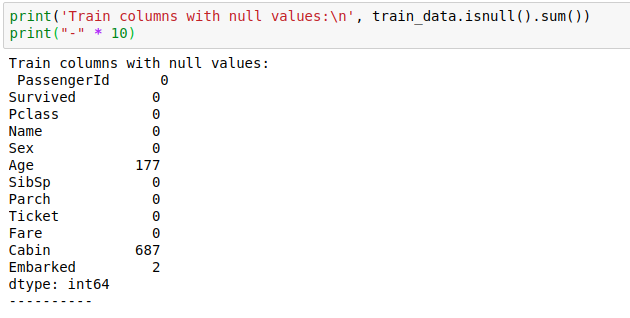
**Endüstriyel Veri Madenciliği Uygulamaları**

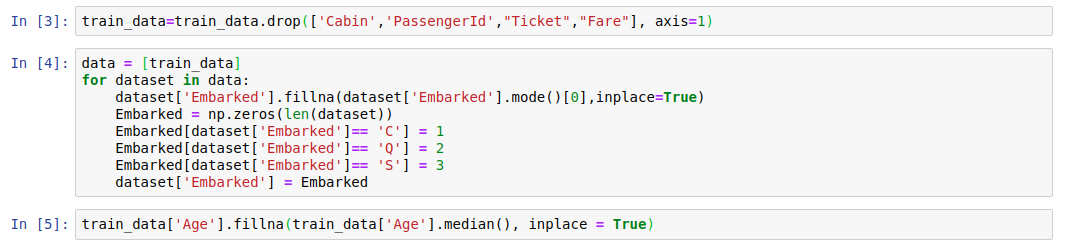
**ARA RAPOR - 2**

Gerekli importlar yapılır ve veri okutulur.

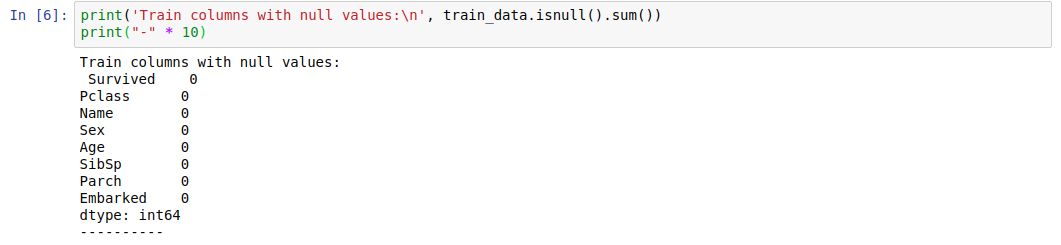


Eksik veri sayısı görüntülenir.



Cabin , Ticket , PassengerId , Fare featureları droplanır. Embarked featureı sayısal değerler alacak şekilde güncellenir. Age’de eksik olan veriler medyana göre tamamlanır.

Artık eksik veri kalmamıştır.



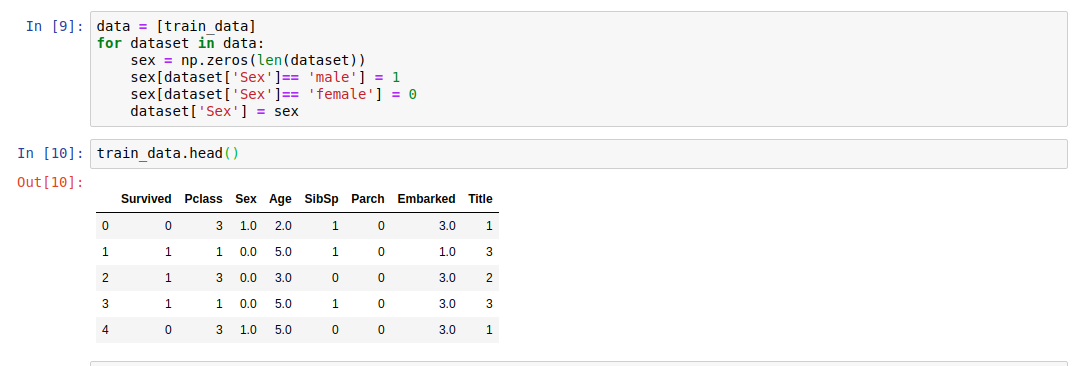
Age verileri kümelere ayrılır.



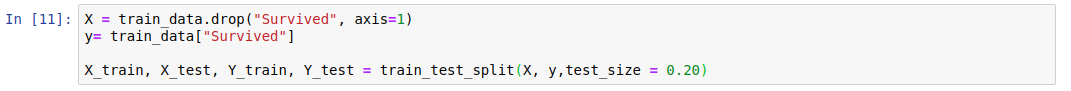
İsimlerin içinde yer alan lakaplar kümelenir ve sayısal değerler verilir.



Cinsiyet sayısal değerler ile belirlenecek hale getirilir.



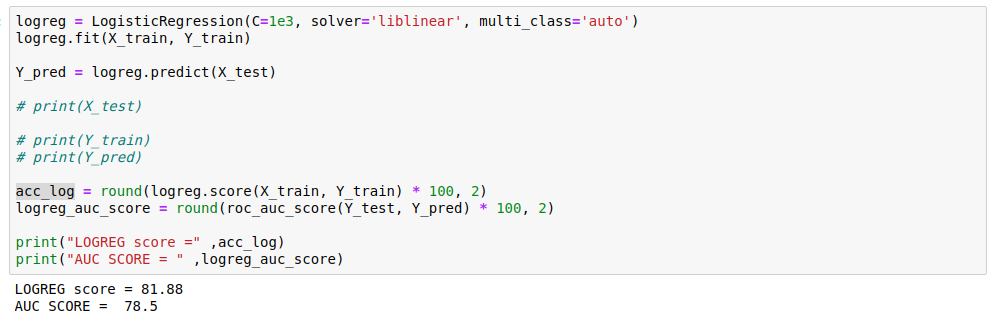
Test ve Train dataları ayrılır.



Normalizasyon yapılır.



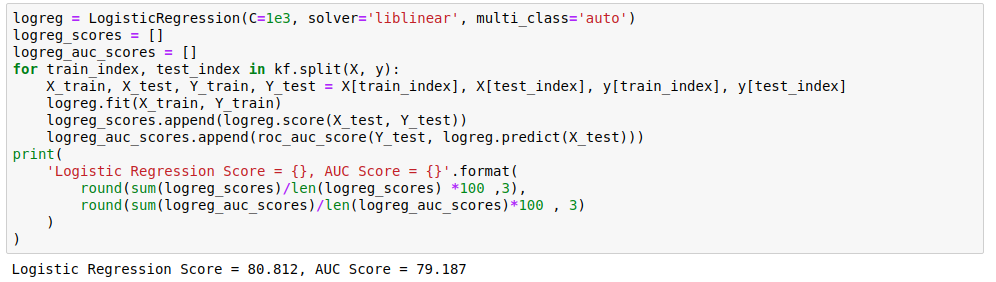
Logistic Regresyona göre çözüm yaptırılır.



Çapraz geçerleme için gerekli fonksiyon tanımlanır.

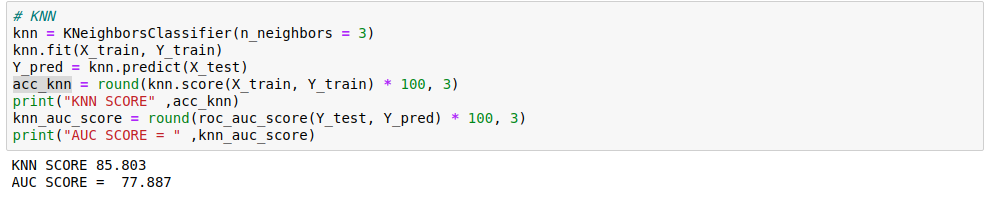


K Fold sonrası Logistic Regresyon yaptırılır.

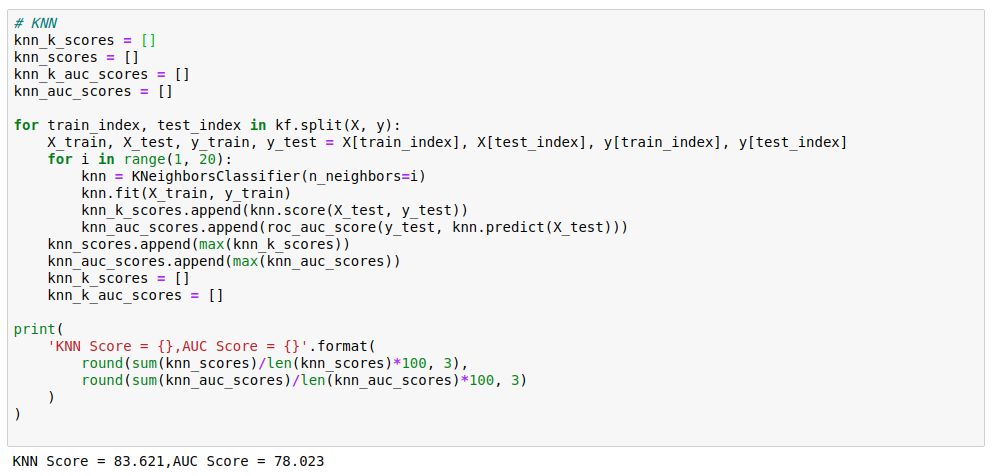


**K FOLD sonrası accuracy azalmıştır ama AUC skoru artmıştır.**

KNN ile çözüm yaptırılır.

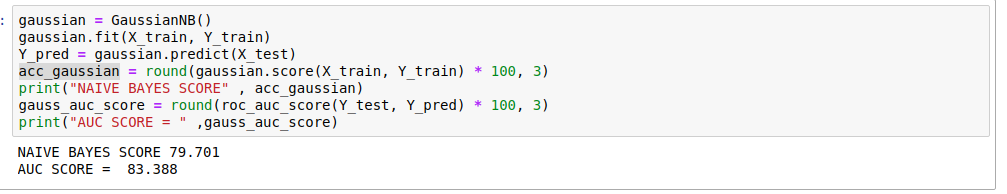


K Fold sonrası KNN çözümü yaptırılır.

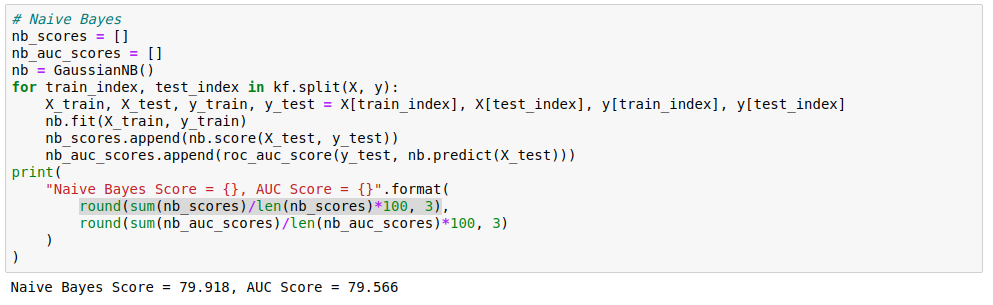


**K FOLD sonrası accuracy azalmıştır ama AUC skoru artmıştır.**

Naive Bayes ile çözüm yapılır.

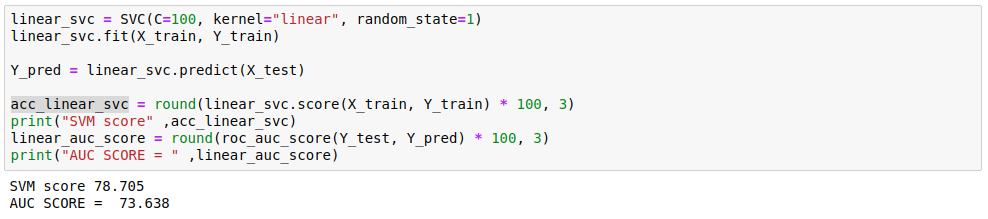


K Fold sonrası Naive Bayes çözümü yaptırılır.

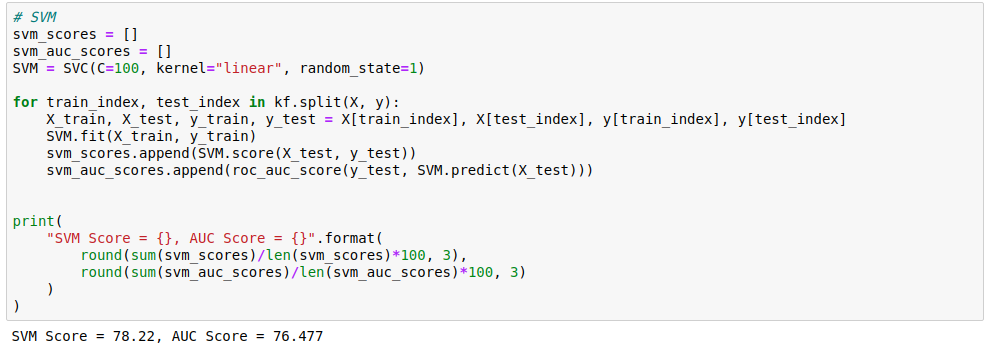


**K FOLD sonrası accuracy artmış ve AUC skoru azalmıştır.**

SVM ile çözüm yaptırılır.

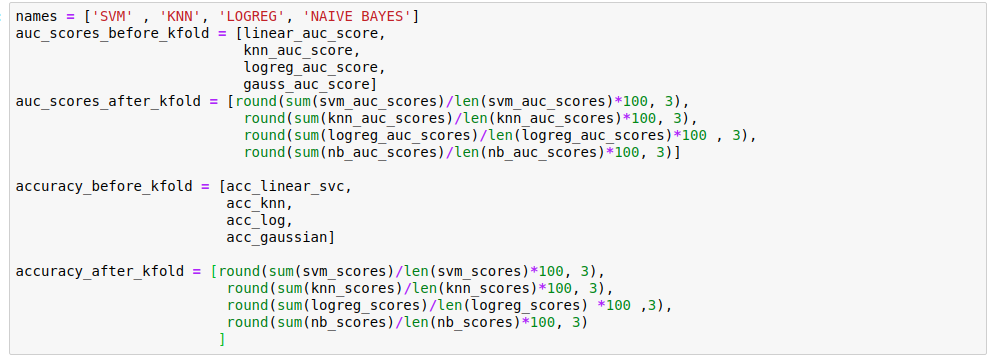


K Fold Sonrası SVM çözümü yaptırılır.

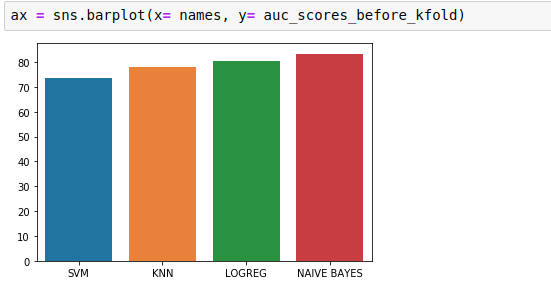


**K FOLD sonrası accuracy azalmıştır ama AUC skoru artmıştır.**

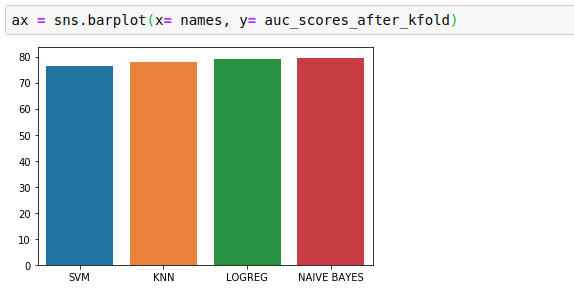
Grafikler için tanımlar yapılır.



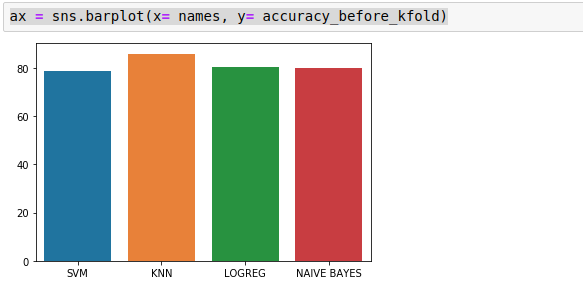
**K fold yapılmadan önce AUC skoru en yüksek olan Naive Bayes yöntemidir.**



**K fold sonrası Naive Bayes yine en yüksek AUC skoruna sahiptir ama diğer yöntemler ile arasında olan farkta azalma olmuştur.**



**K fold öncesinde KNN diğer yöntemlere göre daha yüksek accuracy değerine sahiptir.**



**K fold sonrasında KNN en iyi accuracy değerine sahip olmaya devam etmiştir ama diğer yöntemler ile arasındaki fark azalmıştır.**

