

Machine Learning

Mücahit Söylemez



İÇİNDEKİLER

1. GİRİŞ

1.1. Problem Tanımı

1.2. Literatür Taraması

2. METODOLOJİ

3. BULGULAR

5. SONUÇLAR VE ÖNERİLER

- 1.1 Problem Tanımı

Son zamanlarda cinsiyet tanıma önemli bir araştırma haline geldi. Cinsiyet tanıma, güvenlik amaçlı, konuşmacı Tanıma gibi çeşitli alanlarda kullanılabilir.

- 1.2 Literatür Taraması

Çalışmalar	Yazarlar	Açıklama
This paper deals with identifying the gender using Acoustic properties of voice using Machine learning and how the accuracy vary when dataset goes through different transformation. It shows that we achieve maximum accuracy when uniform transformation is applied on dataset in case of KNN and SVM both.	Rohit Bhatia & Nagendra Pratap Singh	Cinsiyet sınıflandırma için KNN ve SVM sınıflandırma yöntemleri kullanarak makine öğrenmesi gerçekleştirmişler.

1.3. Anahtar Kelimelerle Projeye Bakış

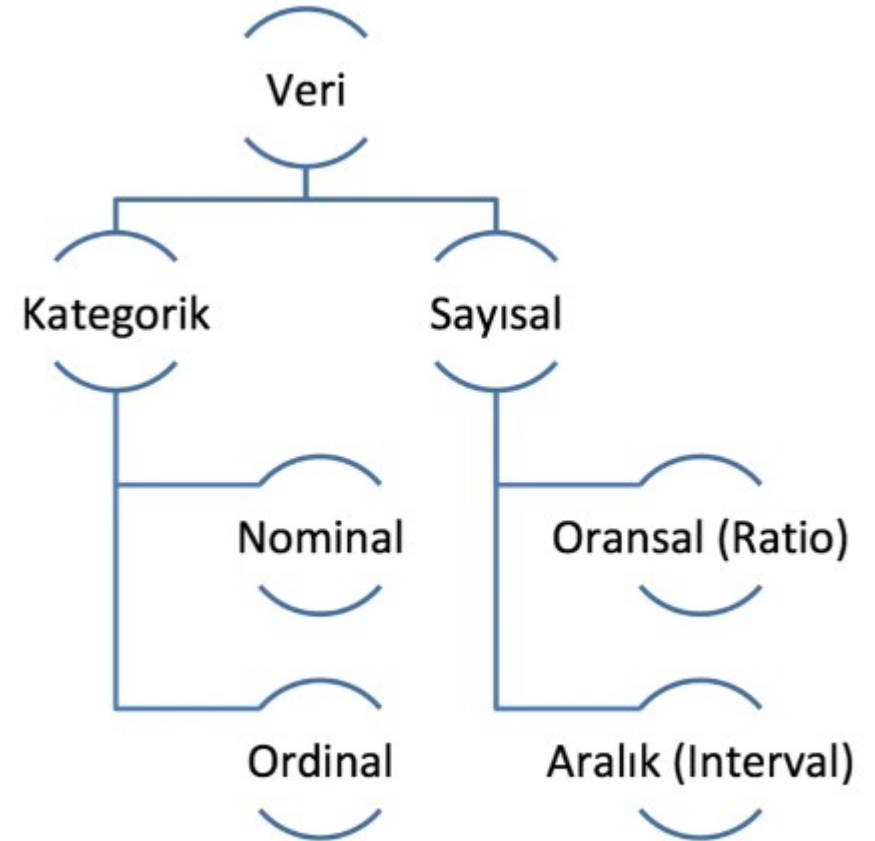
Makine Öğrenmesi
nedir ?

Sınıflandırma
(Classification)
Yöntemleri Nelerdir?

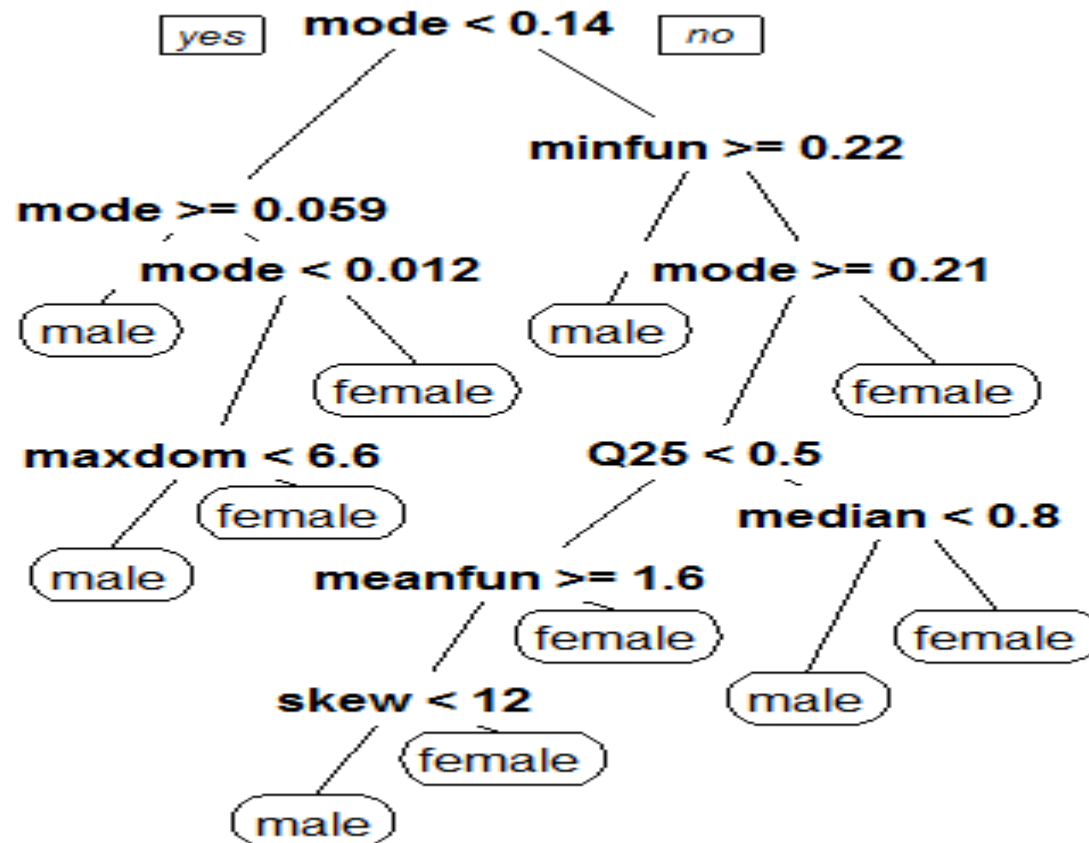
1.3. Anahtar Kelimelerle Projeye Bakış

Makine Öğrenmesi
Nedir ?

Sınıflandırma
(Classification)
Yöntemleri Nelerdir ?



1.3. Anahtar Kelimelerle Projeye Bakış



METODOLOJİ

Veri Seti

20 tane bağımsız ve 1 tane bağımlı özellikten oluşan verisetinde toplam 3169 adet satır bulunmaktadır.

KNN

Random
Forest
Classifier

SVC

METODOLOJİ

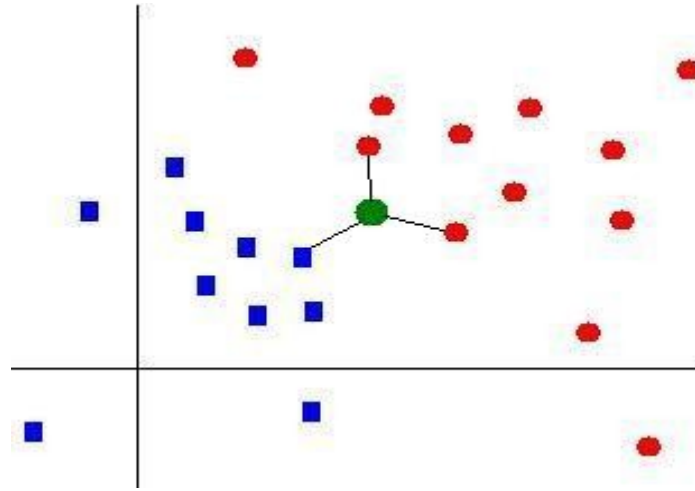
Veri Seti

KNN

Random
Forest
Classifier

SVC

KNN (K-Nearest Neighbors) algoritması sınıflandırma sırasında çıkarılan özelliklerden (feature extraction), sınıflandırılmak istenen yeni bireyin daha önceki bireylerden k tanesine yakınlığına bakılmasıdır.



METODOLOJİ

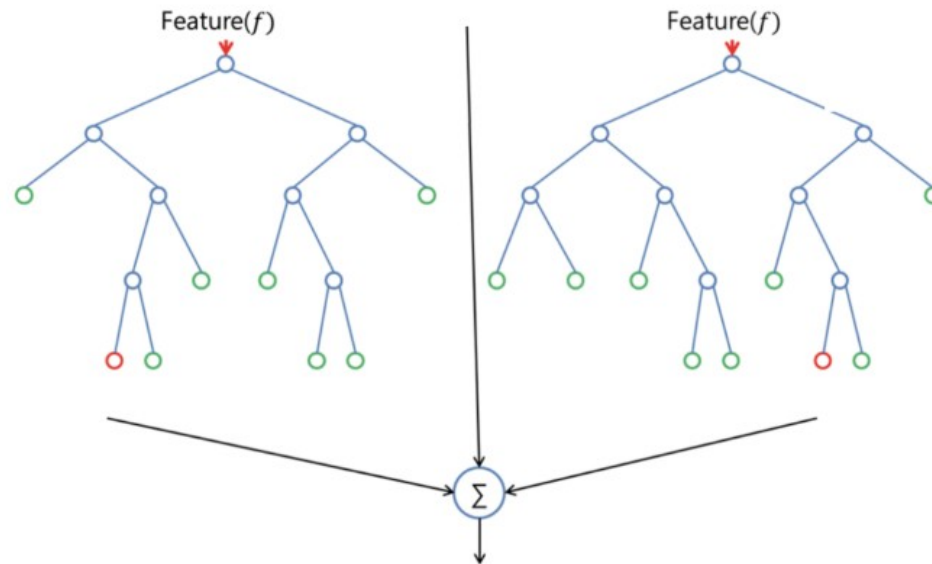
Veri Seti

KNN

Random
Forest
Classifier

SVC

Çok sayıda karar ağacı oluşturarak toplu öğrenme gerçekleştiren bir algoritmadır.



METODOLOJİ

Veri Seti

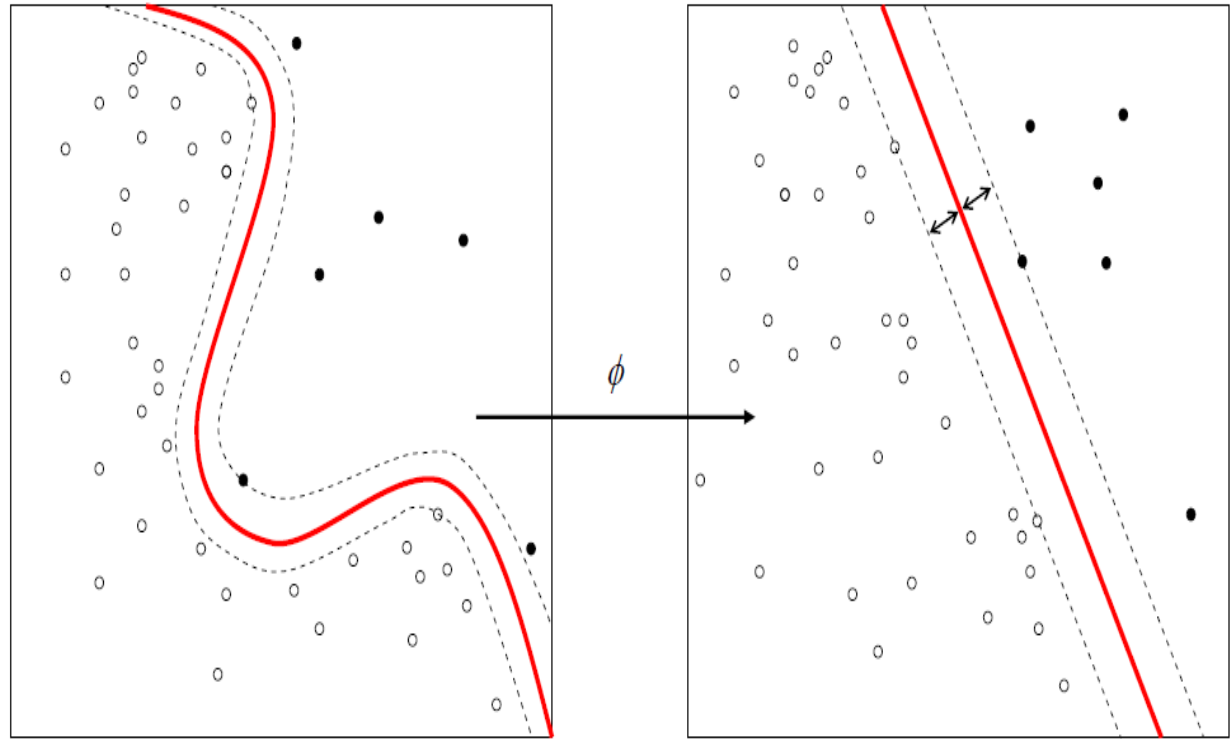
KNN

Random
Forest
Classifier

SVC

İki sınıfı bir doğru veya düzlem ile birbirinden ayırmaya çalışır. Bu ayırmayı da sınırdaki elemanlara göre yapar.

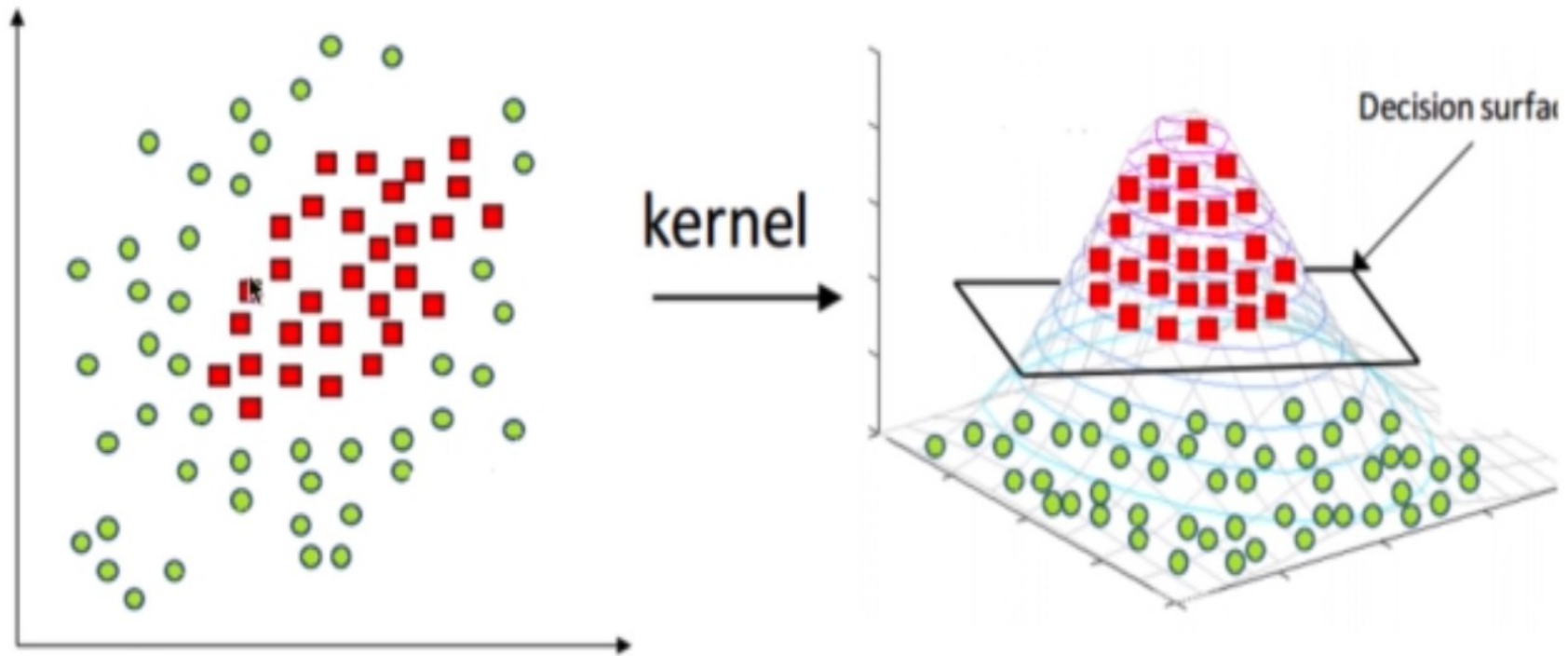
- Destek vektör algoritmasının en önemli özelliği, marjinal verilere karşı hassas olmasıdır.
- Yüksek boyutlu uzayda etkilidir.



METODOLOJİ

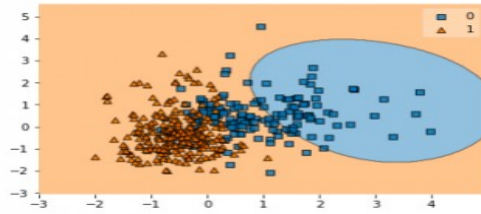
SVC'nin önemli parametreleri;

- Kernel
- C
- Gama

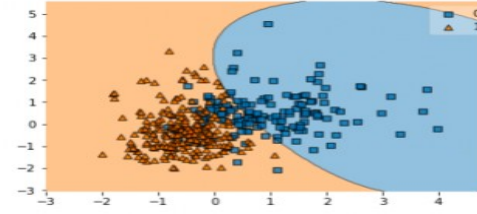


METODOLOJİ

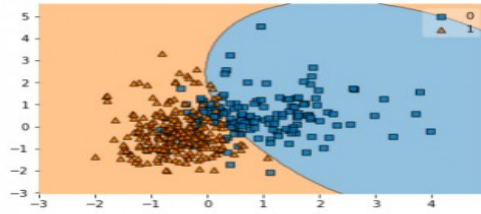
C parametresi yanlış sınıflandırma toleransını belirler. Eğer C parametresi küçük olursa hata toleransı yüksektir. Ama C değeri yüksek olursa hata toleransı düşüktür.



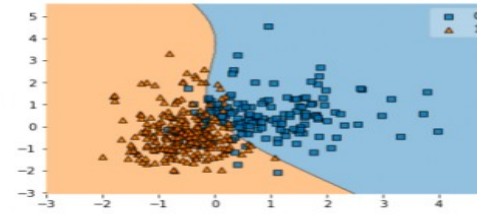
C = 0.02
Accuracy: 81.3%



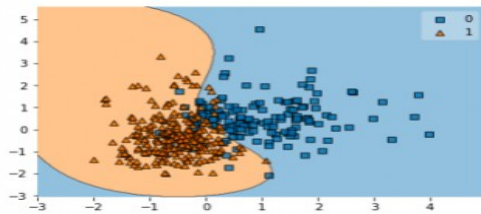
C = 0.03
Accuracy: 88.3%



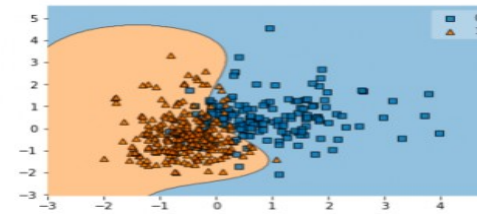
C = 0.08
Accuracy: 90.6%



C = 1.0
Accuracy: 90.6%



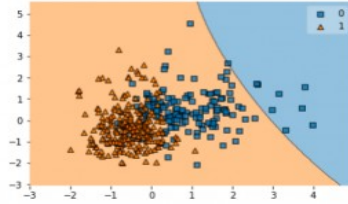
C = 10.0
Accuracy: 90.1%



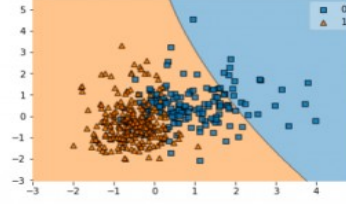
C = 100.0
Accuracy: 90.1%

METODOLOJİ

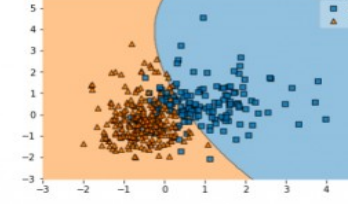
Gama, RBF çekirdeğinin bir parametresidir ve çekirdiğin yayılımı olarak düşünülebilir ve dolayısıyla karar bölgesidir. Gama düşük olduğunda karar sınırının “eğrisi” çok düşüktür. Büyük olursa karar sınırının “eğrisi” büyüktür.



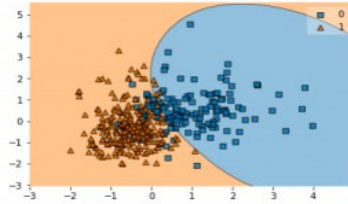
Gamma = 0.008
Accuracy: 63.7%



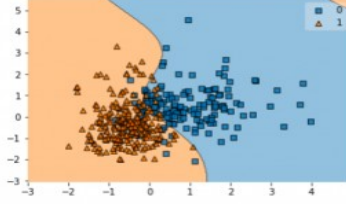
Gamma = 0.01
Accuracy: 68.4%



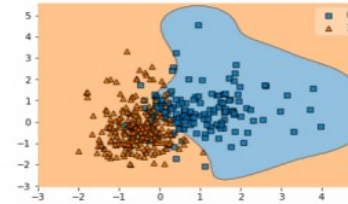
Gamma = 0.05
Accuracy: 88.9%



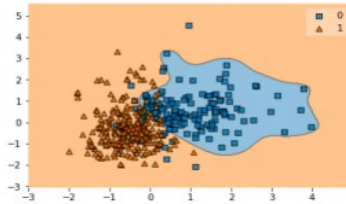
Gamma = 0.1
Accuracy: 90.1%



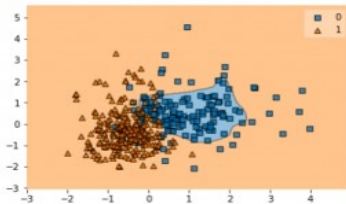
Gamma = 0.5
Accuracy: 91.6%



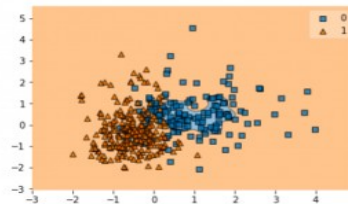
Gamma = 1.0
Accuracy: 90.1%



Gamma = 3.0
Accuracy: 88.9%



Gamma = 7.0
Accuracy: 84.8%



Gamma = 11.0
Accuracy: 74.9%

BULGULAR

K-NN Sınıflandırma Raporu

	Precision	Recall	f1-score	support
Female	0.98	0.97	0.98	496
Male	0.97	0.99	0.98	550
Accuracy			0.98	1046
Macro avg	0.98	0.98	0.98	1046
Weighted avg	0.98	0.98	0.98	1046

K-nn temel model doğruluk değeri = 0.9780

Confusion Matrix

481	15
8	542

Toplam 1046 test verisinden 23 tanesini yanlış sınıflandırma yapmış.

Cross Validation Score = 0.9726
Standart Sapma = 0.003

BULGULAR

Random Forest Sınıflandırma Raporu

	Precision	Recall	f1-score	support
Female	0.98	0.97	0.97	496
Male	0.97	0.98	0.98	550
Accuracy			0.98	1046
Macro avg	0.98	0.98	0.98	1046
Weighted avg	0.98	0.98	0.98	1046

RFC temel model doğruluk değeri = 0.9760

Confusion Matrix

482	14
11	539

Toplam 1046 test verisinden 25 tanesini yanlış sınıflandırma yapmış.

Cross Validation Score = 0.9764
Standart Sapma = 0.006

BULGULAR

Support Vector Classifier (SVC) Sınıflandırma Raporu

	Precision	Recall	f1-score	support
Female	0.99	0.98	0.98	496
Male	0.98	0.99	0.98	550
Accuracy			0.98	1046
Macro avg	0.98	0.98	0.98	1046
Weighted avg	0.98	0.98	0.98	1046

SVC temel model doğruluk değeri = 0.9827

Confusion Matrix

485	11
7	543

Toplam 1046 test verisinden 18 tanesini yanlış sınıflandırma yapmış.

Cross Validation Score = 0.9849
Standart Sapma = 0.003

{'C': 5, 'gamma': 0.1, 'kernel': 'rbf'}

SONUÇ

Sınıflandırma	Accuracy (Doğruluk Oranı)	Cross Validation Ortalaması
Logistic Classifier	0.9770	0.9698
KNN	0.9780	0.9726
Naive Bayes	0.9072	0.8911
Decision Tree	0.9062	0.9590
Random Forest	0.9760	0.9764
SVC	0.9827	0.9849