

# Bir Transkripsiyon Faktörü için DNA Bağlanma Bölgelerinin Tahmini

İ. Topal<sup>1</sup>, A. Kabakçoğlu<sup>2</sup> ve L. Subaşı<sup>1</sup>

<sup>1</sup>İstanbul Teknik Üniversitesi

<sup>2</sup>Koç Üniversitesi

topali@itu.edu.tr



## Özet

Bu çalışmada, fare sinir hücresinin gelişiminde rol oynayan NeuroD2 proteininin hücre DNA'sı üzerindeki bağlanma bölgeleri yapay öğrenme teknikleriyle tahmin edilmiştir. Bu amaçla bağlanma bölgelerine dair deneysel veriler (Chip-Seq) ve bunlarla eşleştirdiğimiz 15 öznitelik kullanılmıştır. Bu ikili sınıflandırma probleminde makine öğrenmesi teknikleri ile başarımlar %65 olarak hesaplanmıştır. İleri aşamalarda başarımların farklı mimariler kullanılarak optimizasyonu ve optimal performans veren sinir ağının parametrelerinin incelenmesi ile bağlanmada kritik önemi olan biyolojik faktörlerin belirlenmesi planlanmıştır.

## Öznitelikler

ENCFF145FVU  
ENCFF145FVU  
ENCFF102IIL  
ENCFF676DBG  
ENCFF875CQU  
ENCFF152TUF

Histon bağlanma eğilimi

Adenine  
Thymine  
Guanine  
Cytosine

Nükleotit

TSSupstream  
CDS  
INTRON  
UTR5  
UTR3  
ND2

Transkripsiyon başlama bölgesine yakınlık

Kodlama bölgesi

Kodlamaya katılmayan kısım

Genin başı

Genin sonu

Hedef protein

## Veri

16 x 195471971 büyüklüğünde BitArray ND2 proteininin bağlandığı bölgeler tüm veri içinde %1'e denk geldiği için öncelikle dengeli bir set oluşturuldu.

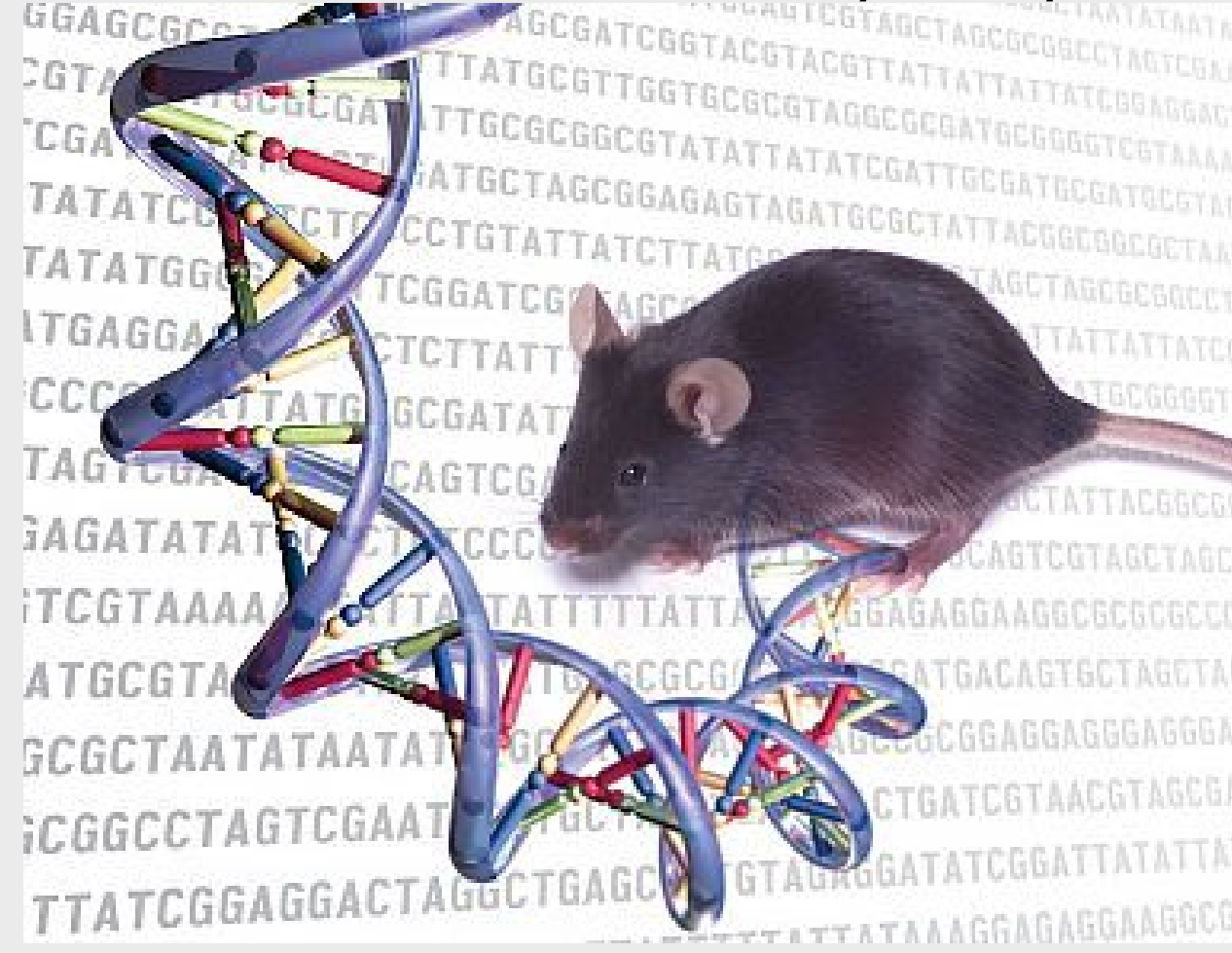
Verileri üreten Doç.Dr. Gülayşe İnce Dunn'a katkılarından dolayı teşekkür ederiz.

## Farklı modellerle başarımlar ve f1 skorlama

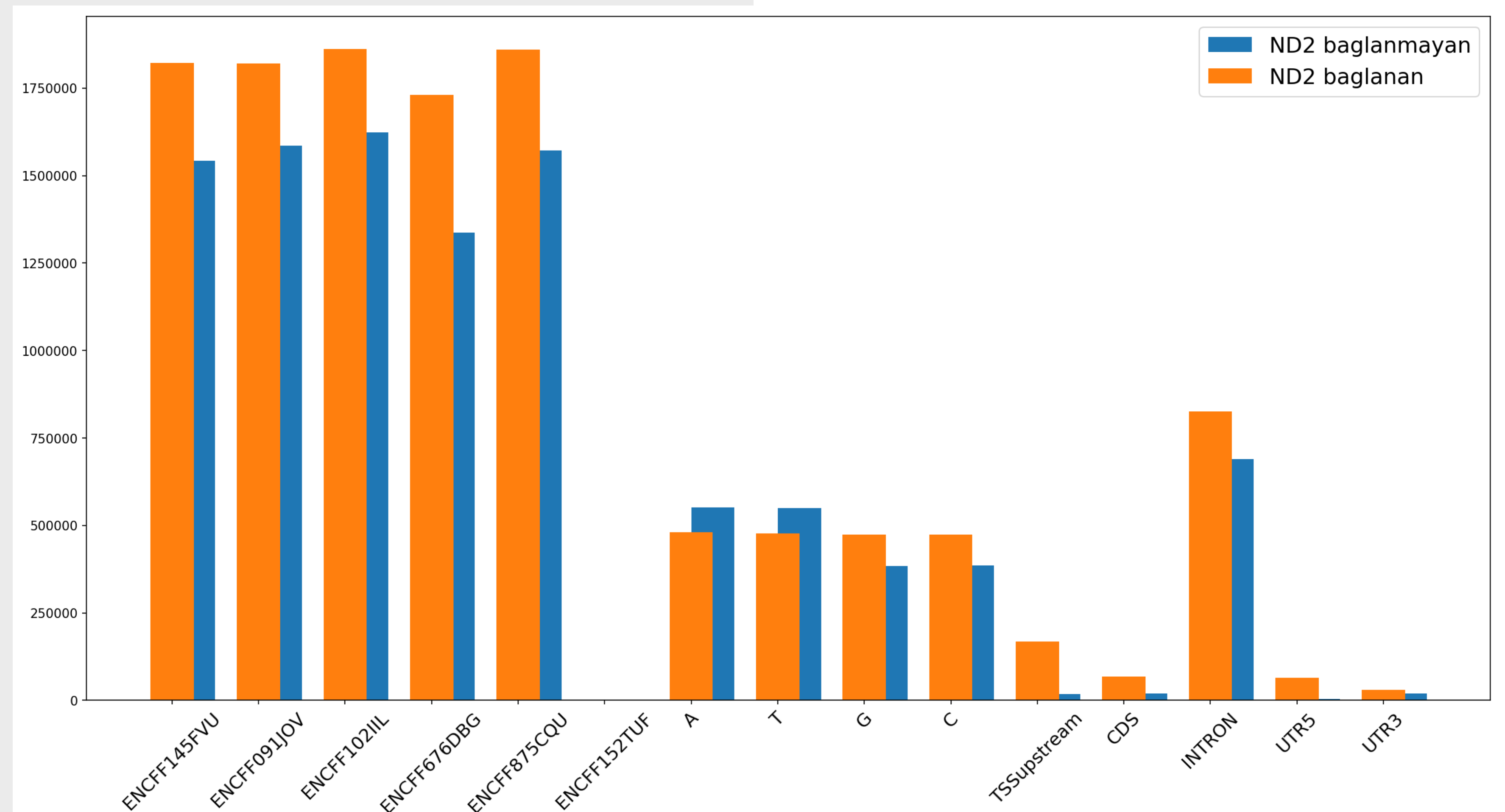
SVM	0.658 - 0.70
Logistic Regression	0.635 - 0.67
SGD	0.63 - 0.69
MLP(10)	0.634 - 0.65
MLP(1000)	0.63 - 0.65

## Öznitelik Analizi

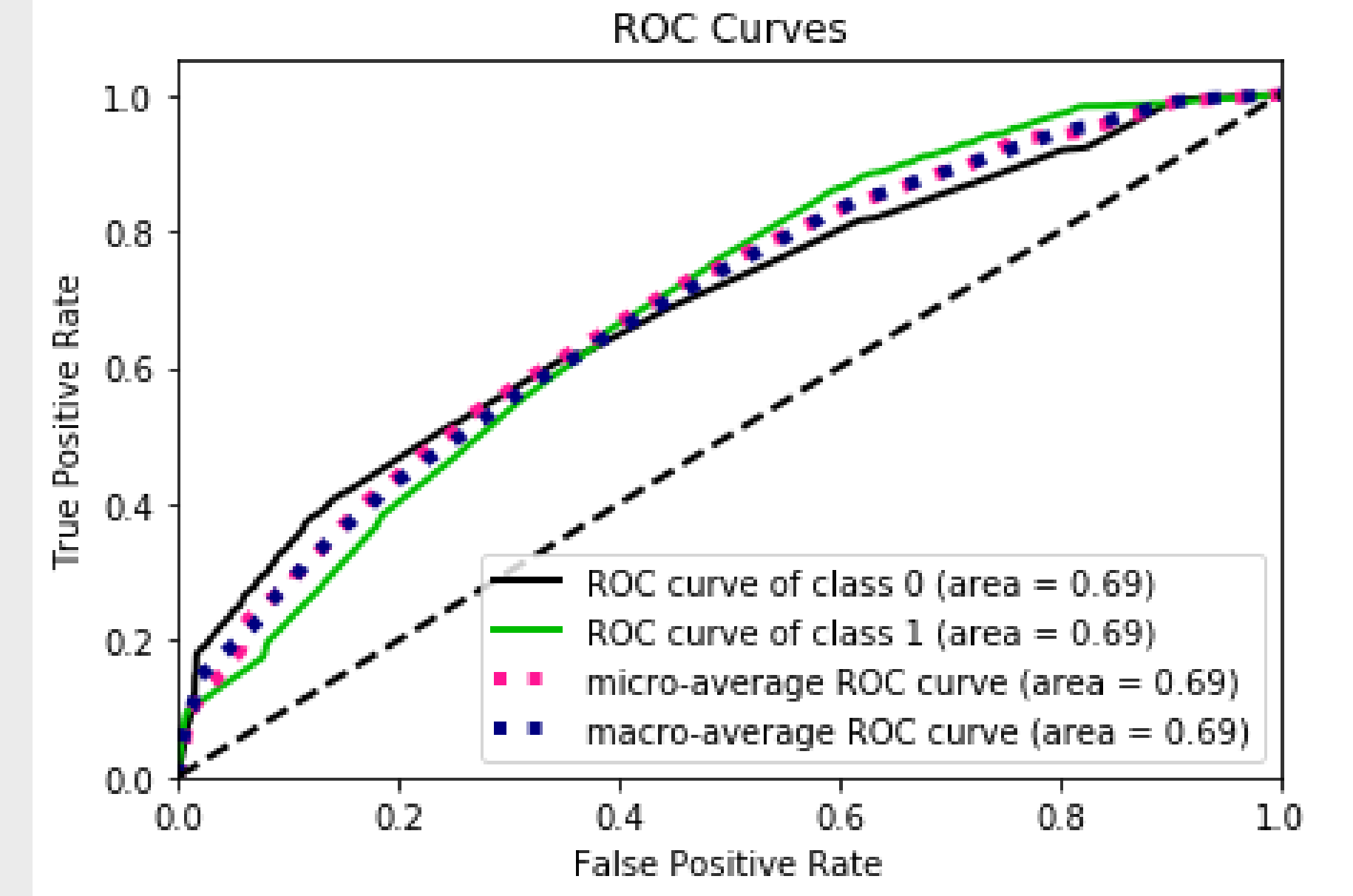
Histogramdan görüleceği gibi dengeli sette 6 farklı histon bağlanma eğilimi verisinin birinden hiç katkı gelmemektedir. TSSupstream özniteliği veriden çıkarılarak yapılan hesaplarda modeller bağlanma bölgelerini öğrenememiş, histogramdan da anlaşılacağı üzere bu öznitelik öne çıkmıştır.



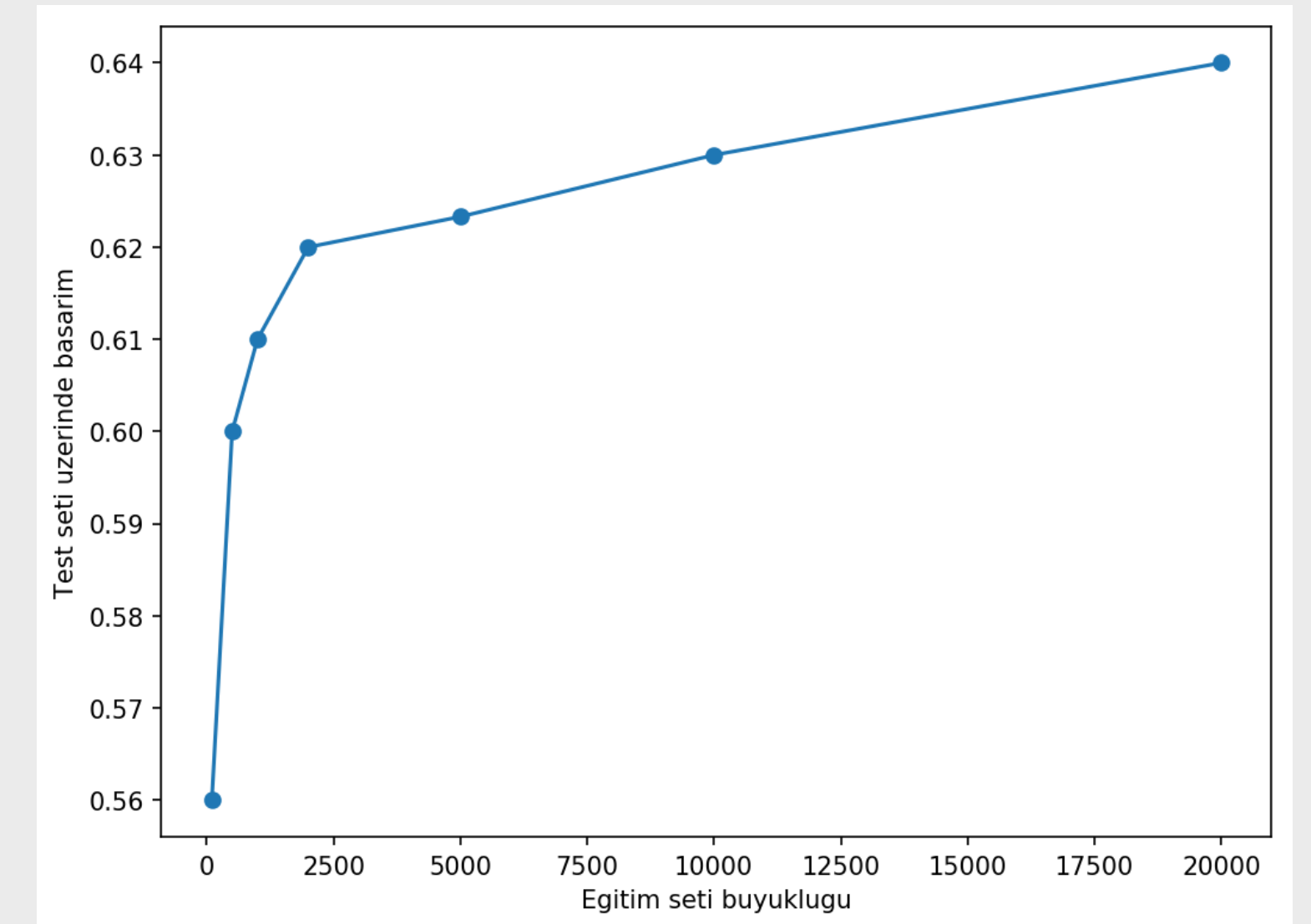
## Dengeli sette özniteliklerin dağılımı



## SVM için ROC Eğrisi



## SVM için Eğitim Seti Büyüklüğü



**Referans** B.Alipanahi, A. Delong, M.T. Weirauch and B. J. Frey, *Nat. Biotechnol.* **33**, 831-838 (2015). <https://www.nih.gov/>