

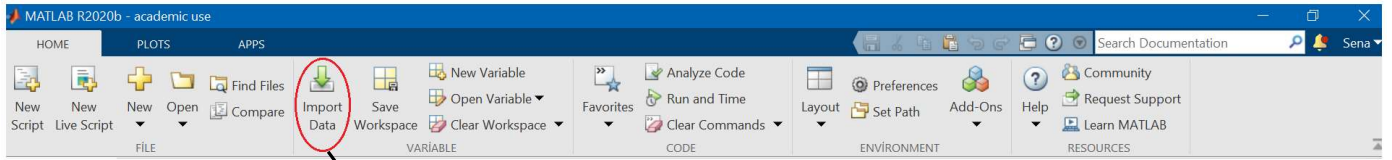
# YAPAY ZEKA FİNAL PROJESİ

Excel tablosunda veri setlerinin oluşturduğu diyabet hastalığının bazı ölçütlerine göre sınıflandırılan verilerden bazıları aşağıdaki gibidir;

768x9'luk  
bir veri  
topluluğu  
oluşturuldu.

	A	B	C	D	E	F	G	H	I
1	Glikoz	Tansiyon	Insulin	Kalori	BMI	Uyku süresi	Yas	Cilt Kalınlığı	Sonuc
2	148	72	0	2760	266	8	31	35	0
3	85	66	0	6480	233	8	32	29	1
4	183	64	0	3912	281	8	21	0	0
5	89	66	94	1970	431	9	33	23	1
6	137	40	168	2760	256	4	30	35	0
7	116	74	0	4140	31	6	26	0	1
8	78	50	88	2853	353	6	29	32	0
9	115	0	0	3980	305	7	53	0	1
10	197	70	543	4340	0	8	54	45	1
11	125	96	0	2507	376	5	30	0	0
12	110	92	0	1570	38	10	34	0	1
13	168	74	0	2853	271	6	57	0	0
14	139	80	0	4280	301	8	59	0	1
15	189	60	846	5180	258	7	51	23	1
16	166	72	175	9990	30	7	32	19	1
17	100	0	0	4100	458	9	31	0	1
18	118	84	230	6820	296	7	31	47	1
19	107	74	0	2680	433	8	33	0	0
20	103	30	83	7370	346	8	32	38	1
21	115	70	96	2733	393	6	27	30	0
22	126	88	235	3813	354	8	50	41	0
23	99	84	0	4100	398	7	41	0	1
24	196	90	0	3128	29	9	29	0	1
25	119	80	0	1890	366	11	51	35	1
26	143	94	146	7860	311	10	41	33	1
27	125	70	115	4270	394	7	43	26	1
757	128	88	110	3320	32	7	39	39	0
758	137	90	0	4140	363	9	52	41	1
759	123	72	0	2160	375	8	26	0	0
760	106	76	0	1970	355	6	66	0	1
761	190	92	0	2933	284	7	22	0	0
762	88	58	16	4340	44	9	43	26	1
763	170	74	0	2853	225	5	33	31	0
764	89	62	0	2860	329	6	63	0	0
765	101	76	180	3320	368	7	27	48	0
766	122	70	0	3453	262	4	30	27	0
767	121	72	112	8850	301	8	47	23	1
768	126	60	0	2620	304	8	23	0	0

Bu kısımla kaggle tarafından alıntılanan örnek veri setinden yola çıkılarak matlab programında import data ile veriler workspace ortamına aktarıldı.



Import data butonuna tıkladıktan sonra Excel üzerinden veri dosyasını seçip verilerin Matlab üzerine aktarımı sağlanır.

Range: A2:I768 Output Type: Table

Variable Names Row: 1

SELECTION IMPORTED DATA UNIMPORTABLE CELLS IMPORT

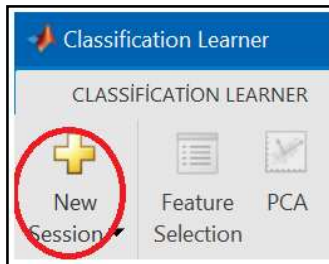
DIYABET.xlsx

Diyabet_Rahatsızlığı									
Glikoz	Tansiyon	Insulin	Kalori	BMI	UykuSuresi	Yas	CiltKalinligi	Sonuc	
Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
1	Glikoz	Tansiyon	Insulin	Kalori	BMI	Uyku süresi	Yas	Cilt Kalinligi	Sonuc
2	148	72	0	2760	266	8	31	35	0
3	85	66	0	6480	233	8	32	29	1
4	183	64	0	3912	281	8	21	0	0
5	89	66	94	1970	431	9	33	23	1
6	137	40	168	2760	256	4	30	35	0
7	116	74	0	4140	31	6	26	0	1
8	78	50	88	2853	353	6	29	32	0
9	115	0	0	3980	305	7	53	0	1
10	197	70	543	4340	0	8	54	45	1
11	125	96	0	2507	376	5	30	0	0
12	110	92	0	1570	38	10	34	0	1
13	168	74	0	2853	271	6	57	0	0
14	139	80	0	4280	301	8	59	0	1
15	189	60	846	5180	258	7	51	23	1
16	166	72	175	9990	30	7	32	19	1
17	100	0	0	4100	458	9	31	0	1
18	118	84	230	6820	296	7	31	47	1
19	107	74	0	2680	433	8	33	0	0
20	103	30	83	7370	346	8	32	38	1

Veriler seçildikten sonra workspace kısmına aktarılır.

Workspace		
Name	Value	Size
Diyabet_Rahatsızlığı	768x9 table	768x9

Veriler adlandırılıp workspace kısmında aktarıldıktan sonra Apps bölümünden 'classification learner' butonuna tıklanır.



APPS---> Classification Learner butonuna tıklandığında new session kısmından workspace de bulunan veri seçilir.

Veri seçildi

Otomatik olarak alınıyor.

New Session

Data set

Data Set Variable

Diyabet\_Rahatsızlığı 768x9 table

Response

From data set variable

From workspace

Sonuc

double

0 .. 1

Predictors

	Name	Type	Range
<input checked="" type="checkbox"/>	Glikoz	double	0 .. 199
<input checked="" type="checkbox"/>	Tansiyon	double	0 .. 122
<input checked="" type="checkbox"/>	Insulin	double	0 .. 846
<input checked="" type="checkbox"/>	Kalori	double	380 .. 9990
<input checked="" type="checkbox"/>	BMI	double	0 .. 671
<input checked="" type="checkbox"/>	UykuSuresi	double	4 .. 15
<input checked="" type="checkbox"/>	Yas	double	21 .. 81
<input checked="" type="checkbox"/>	CiltKalınlığı	double	0 .. 99
<input type="checkbox"/>	Sonuc	double	0 .. 1

Add All

Remove All

How to prepare data

Validation

Cross-Validation

5 katlamalı çapraz doğrulama

Holdout Validation

Recommended for large data sets.

No Validation

No protection against overfitting.

Start Session

Cancel

Response variable is numeric. Distinct values will be interpreted as class labels.

Start Session'a tıklandıktan sonra Classification Learner kısmında bulunan çeşitli algoritmaları ortaya çıktı.

Ağaç algoritma oranlarının %92-%94 aralığında başarı gösterdiği görüldü.

History

1.1	Tree	Accuracy: 92.6%
Last change: Fine Tree		
1.2	Tree	Accuracy: 92.6%
Last change: Medium Tree		
1.3	Tree	Accuracy: 94.1%
Last change: Coarse Tree		

Current Model

Model 1.1: Trained

Results

Accuracy	92.6%
Total misclassification cost	57
Prediction speed	~60000 obs/sec
Training time	0.59614 sec

Model Type

Preset: Fine Tree

Maximum number of splits: 100

Split criterion: Gini's diversity index

Surrogate decision splits: Off

Optimizer Options

Hyperparameter options disabled

Scatter Plot

Original data set: Diyabet\_Rahatsızlığı

Plot

Data

Model predictions

Predictors

X: Glikoz

Y: Tansiyon

Classes

0

1

Scatter Plot

How to investigate features

History

1.1	Tree	Accuracy: 92.6%
Last change: Fine Tree		
1.2	Tree	Accuracy: 92.6%
Last change: Medium Tree		
1.3	Tree	Accuracy: 94.1%
Last change: Coarse Tree		

Current Model

Model 1.3: Trained

Results

Accuracy	94.1%
Total misclassification cost	45
Prediction speed	~44000 obs/sec
Training time	0.83224 sec

Model Type

Preset: Coarse Tree

Maximum number of splits: 4

Split criterion: Gini's diversity index

Surrogate decision splits: Off

Optimizer Options

Model 1.3

Confusion Matrix

	0	1
0	99.6%	0.4%
1	16.1%	83.9%

TPR

FNR

Plot

Number of observations

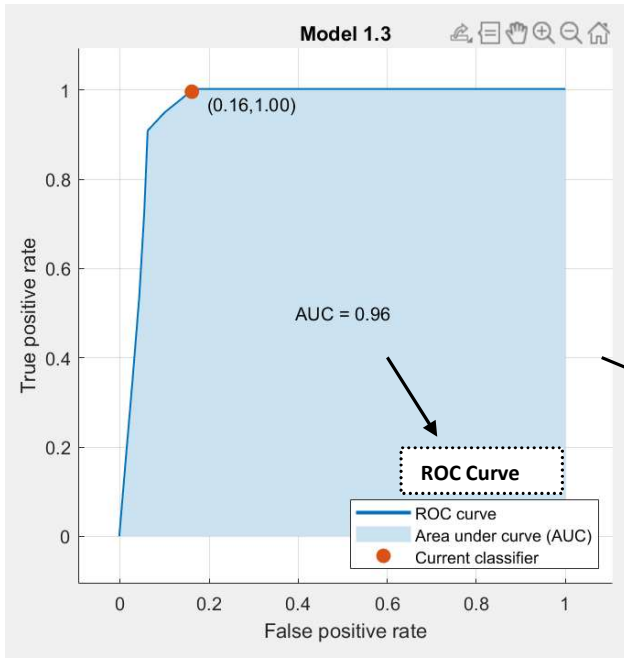
True Positive Rates (TPR)

False Negative Rates (FNR)

Positive Predictive Values (PPV)

False Discovery Rates (FDR)

What is the confusion matrix?

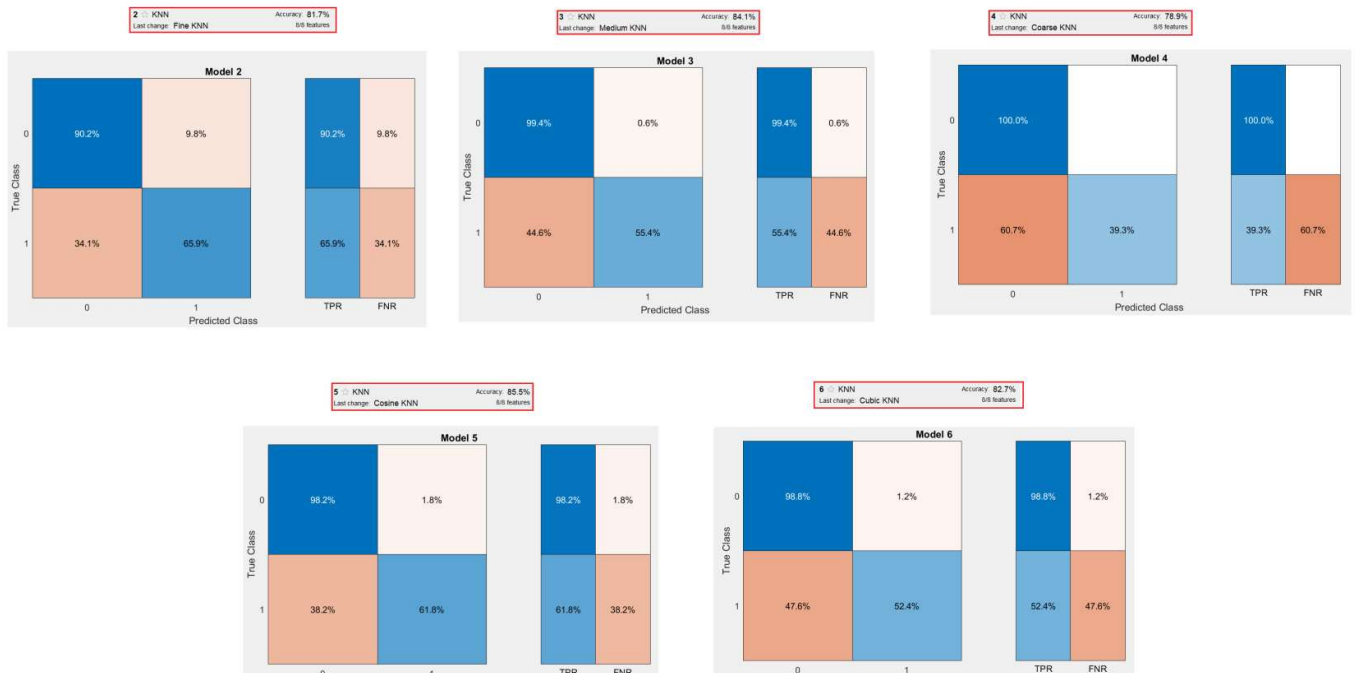


Decision Tree algoritmasının farklı model yapıları böylelikle gösterilmiş oldu.

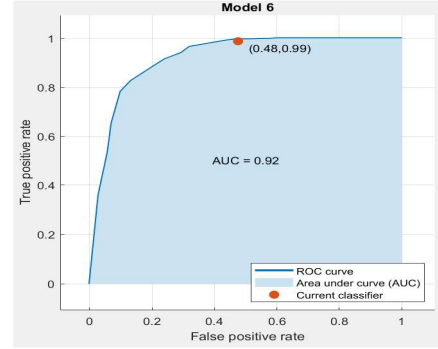
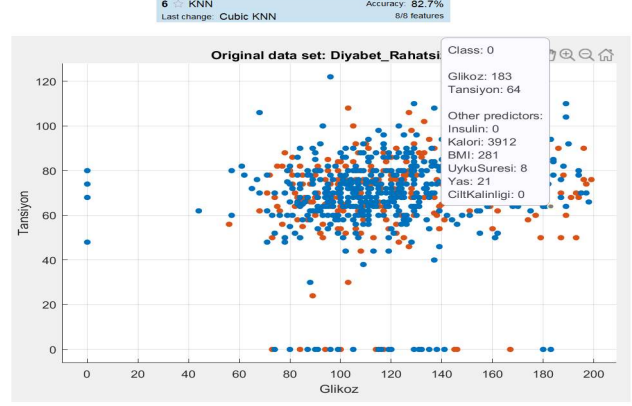
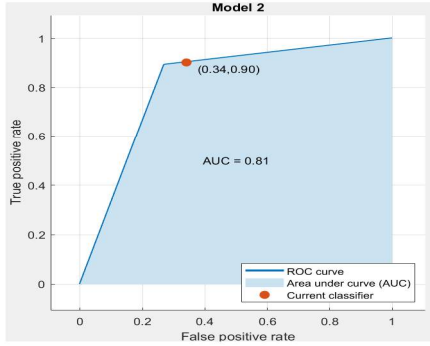
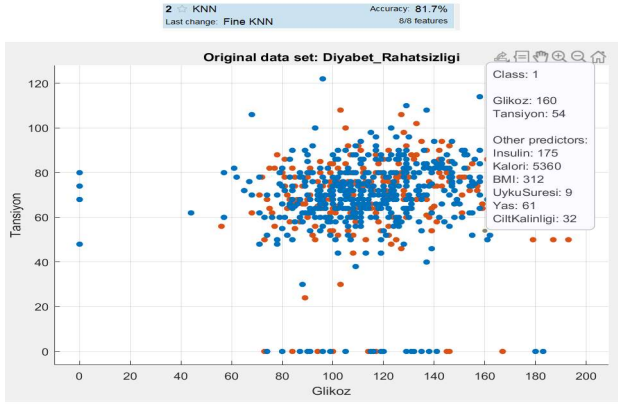
**İkinci algoritmamız ise K-En Yakın Komşu (K-Nearest Neighbors(KNN)) algoritmasıdır ve model karşılaştırılması aşağıdaki gibidir;**

2 ☆ KNN	Accuracy: 81.7%
Last change: Fine KNN	8/8 features
3 ☆ KNN	Accuracy: 84.1%
Last change: Medium KNN	8/8 features
4 ☆ KNN	Accuracy: 78.9%
Last change: Coarse KNN	8/8 features
5 ☆ KNN	Accuracy: 85.5%
Last change: Cosine KNN	8/8 features
6 ☆ KNN	Accuracy: 82.7%
Last change: Cubic KNN	8/8 features

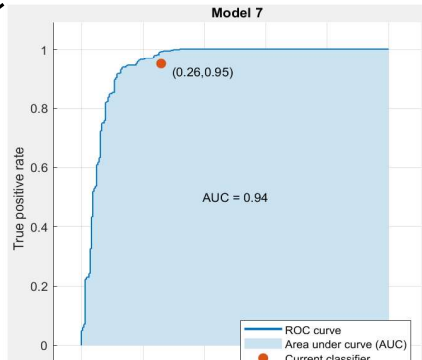
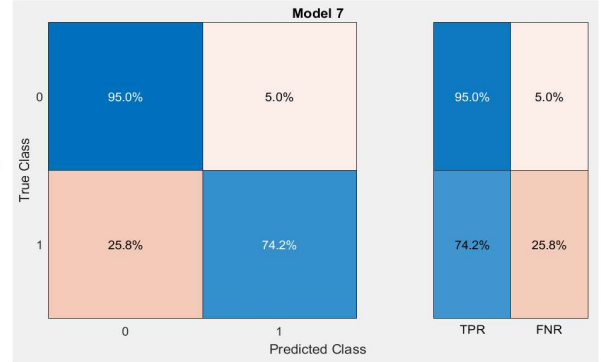
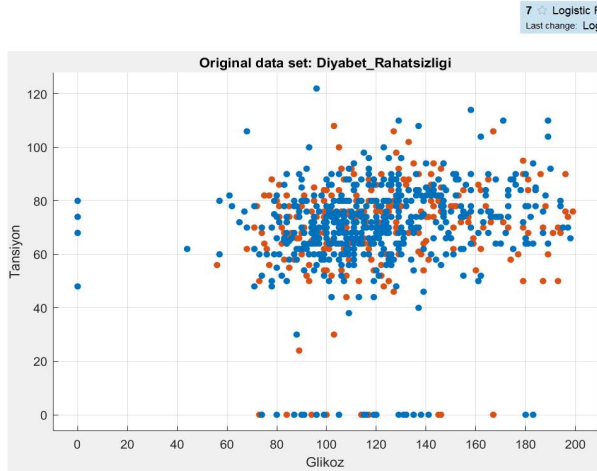
K-En Yakın Komşu sınıflandırma modellerinin başarı oranları %78 ve %85.5 arasında değişmektedir.



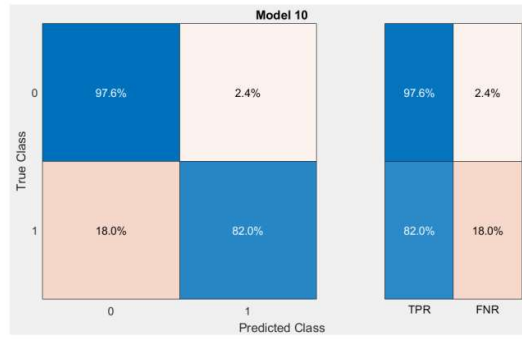
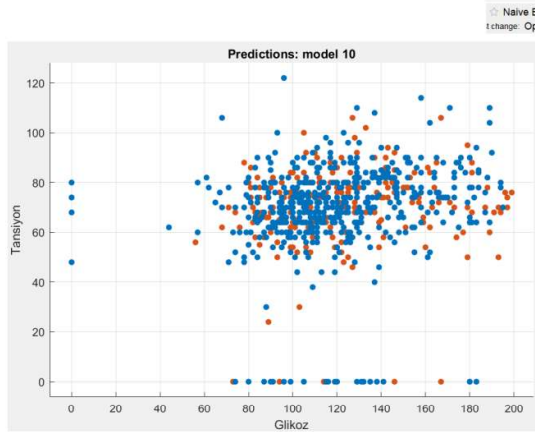
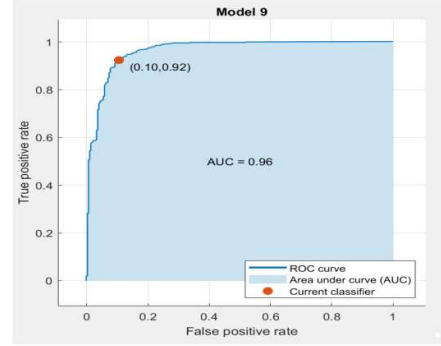
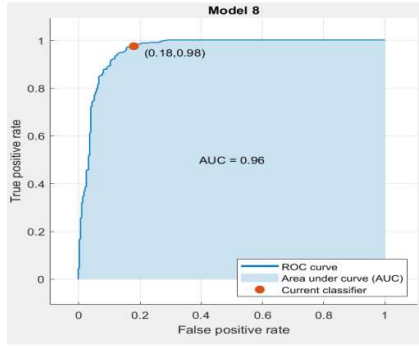
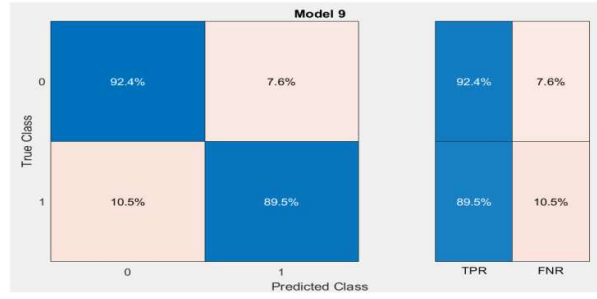
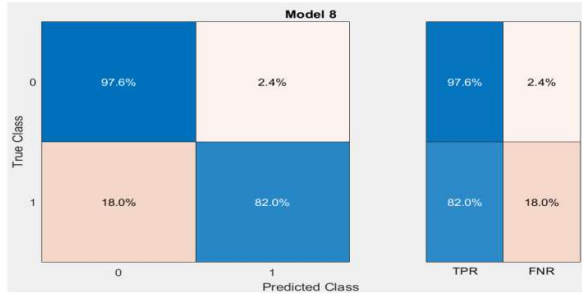
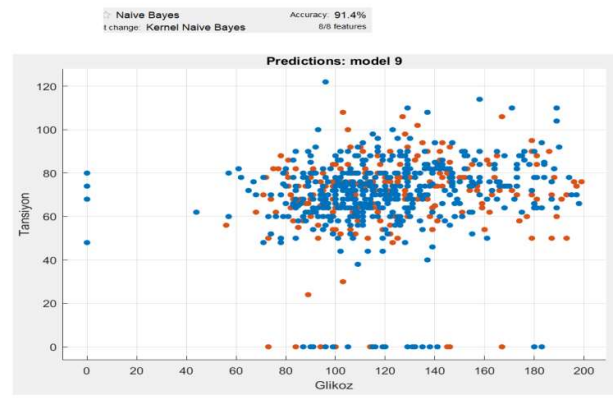
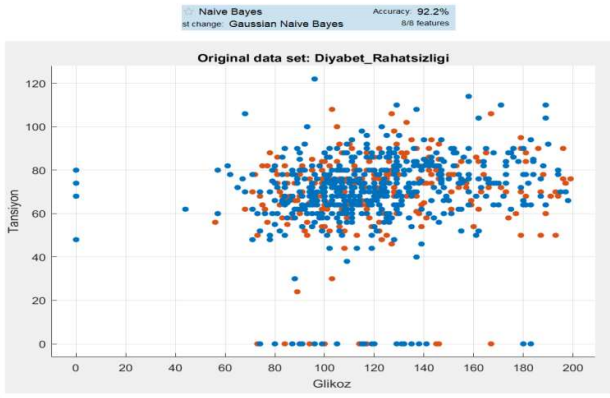




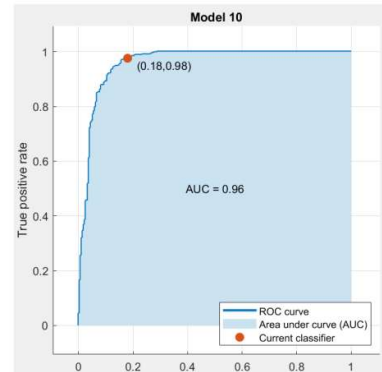
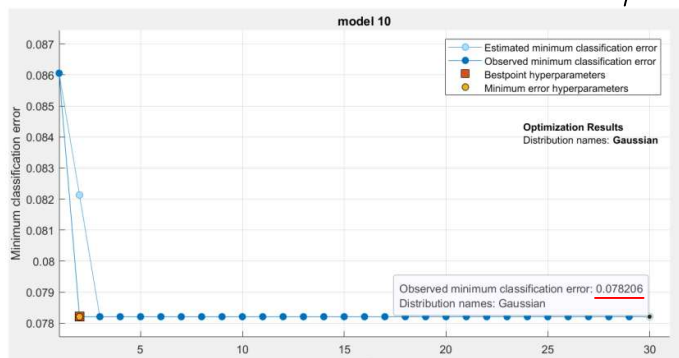
K-En Yakın Komşu sınıflandırma modellerinin başarı oranları görüldüğü üzere Scatter Plot, Confusion Matrix ve ROC Curve modellerine göre ayırım yapıldı.



Logistic Regression sınıflandırması  
Scatter Plot, Confusion Matrix,  
ROC Curve modeli olarak  
karşılaştırılması yapılarak %87.7  
oranında başarı sağlandı.

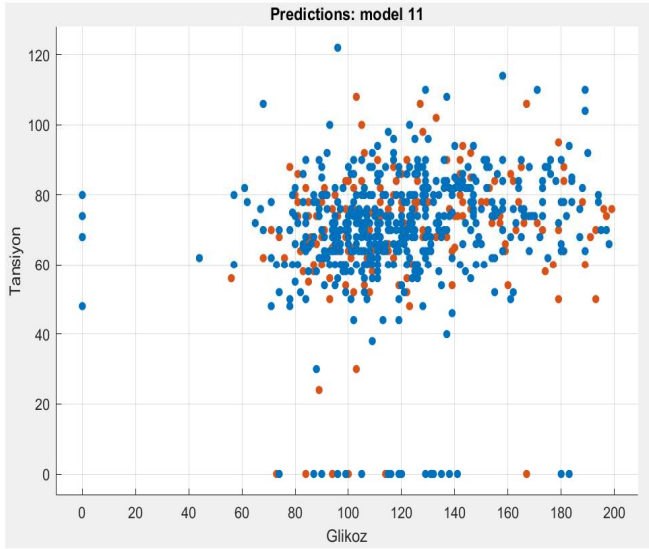


Naive Bayes sınıflandırma yönteminde üç farklı alt metodu kullanıldı.

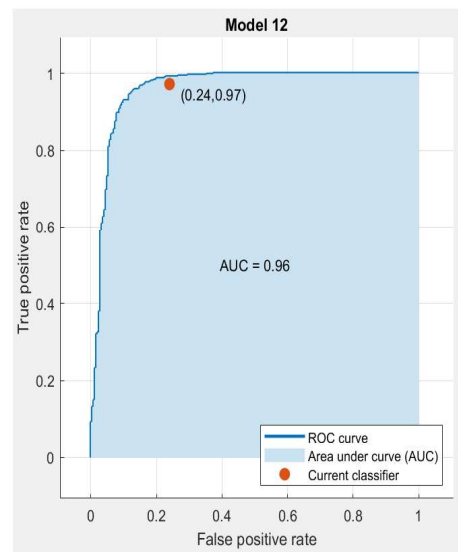
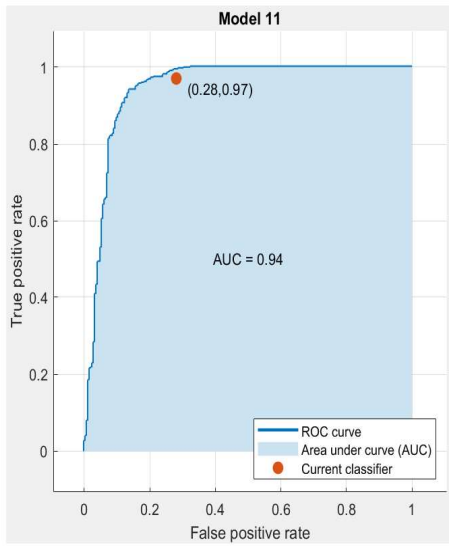
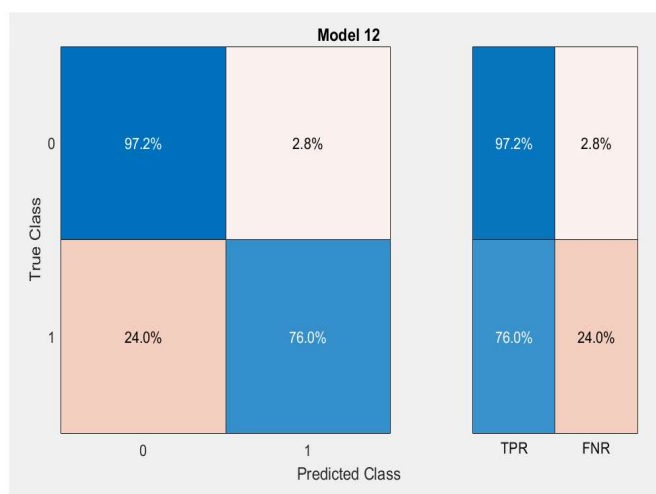
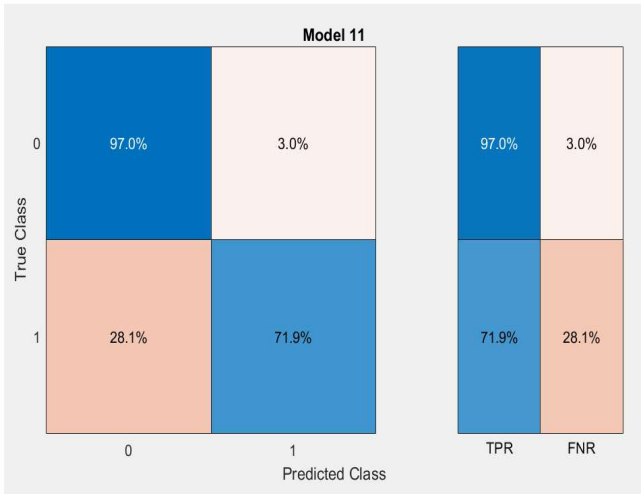
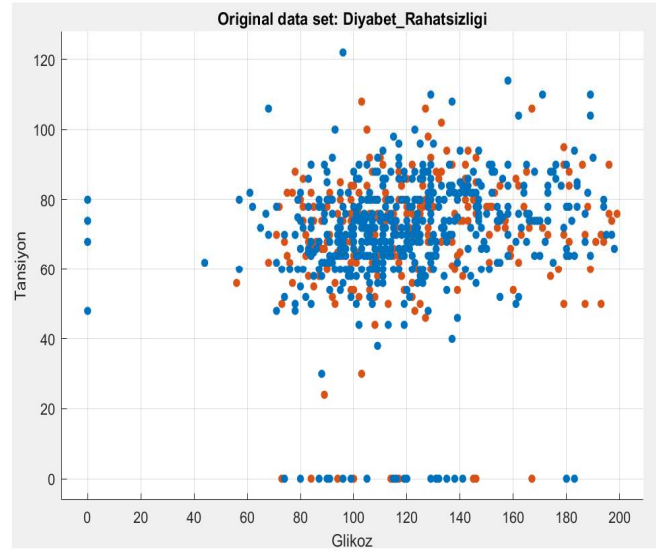


Minimum Classification Error Plot (Minimum Sınıflandırma Hata Grafiği)

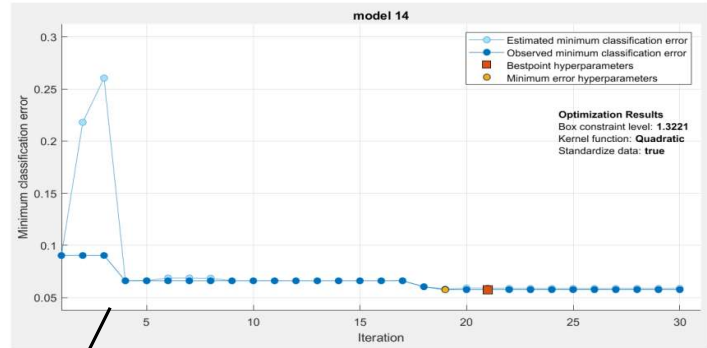
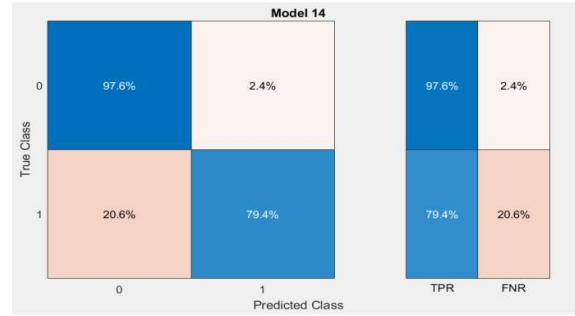
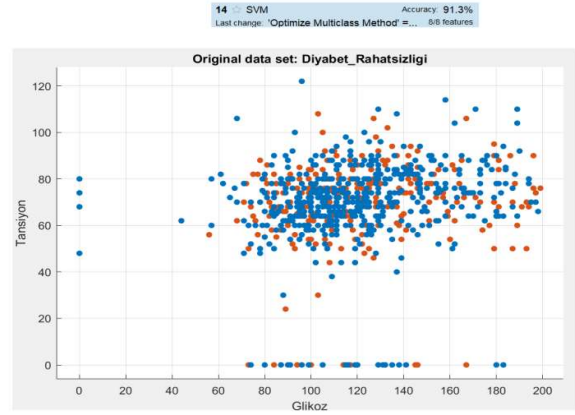
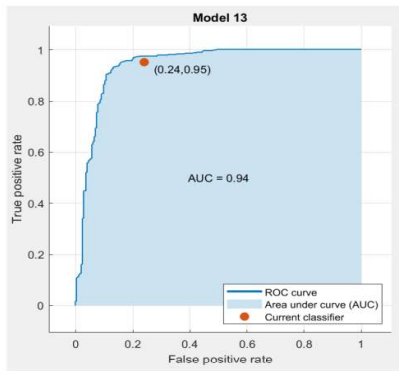
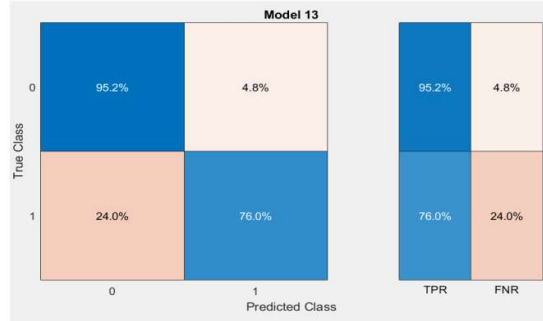
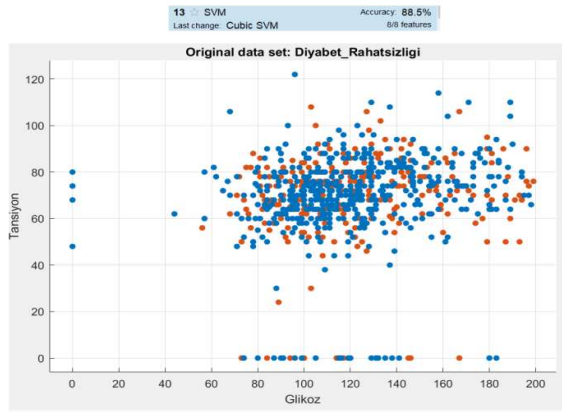
11 ☆ SVM Accuracy: 88.3%  
Last change: Linear SVM 8/8 features



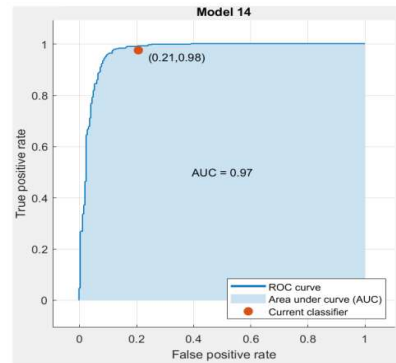
12 ☆ SVM Accuracy: 89.8%  
Last change: Quadratic SVM 8/8 features



DEVAMI



Minimum Sınıflandırma Hata (Minimum Classification Error Plot) grafiğinde çekirdek fonksiyonu quadratic olarak ve kutu kısıtlama seviyesi (box constraint level) sonucu 1.3221 olarak sonuca ulaşıldı.



Sonuç olarak sınıflandırma veri setinin çeşitli makine algoritmalarıyla eğitilerek birden farklı modellerin performanslarının farklılıkları karşılaştırılmış olup her algoritmanın farklı sonuçlar ortaya çıkardığı görülmüştür.

#### VERİ SET KAYNAKLARI:

kaggle datasets download -d ahmettezcantekin/beginner-datasets

kaggle datasets download -d dslearner0406/diabetes-dataset

kaggle datasets download -d abdallamahgoub/diabetes