



ENTERTECH AKADEMİ DATA SCIENCE FİNAL PROJESİ

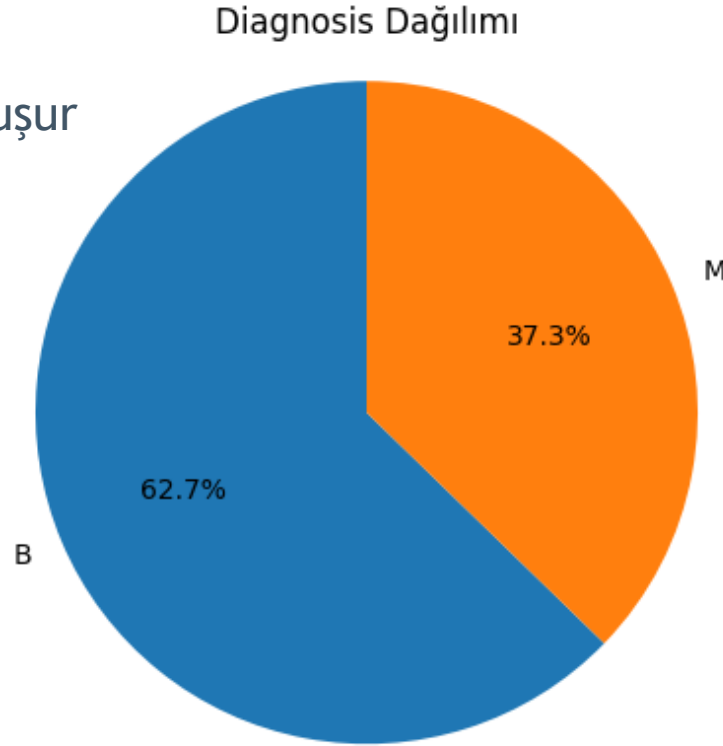
HÜSEYİN DOĞDU

İÇERİK

- 1) Wisconsin Breast Cancer Dataset Hakkında Bilgi
- 2) Bu Alanda Veri Bilimi'nin Kullanmanın Amacı
- 3) Sütunların B ve M bazlı Ortalamaları
- 4) Modeller
- 5) Modellerin Performansı ve Tıbbi Bağlamda Yorumlanması

1) Wisconsin Breast Cancer Dataset Hakkında Bilgi

- ★ Wisconsin Breast Cancer dataset, meme kanseri teşhisi konusundaki bir veri setidir.
- ★ Meme biyopsisi sonuçlarına dayanarak tümörlerin iyi huylu (B - Benign) veya kötü huylu (M - Malignant) olup olmadığını sınıflandırmak için kullanılır.
- ★ 569 Satır ve 32 sütundan(özellikten) oluşur



Data Set içindeki Özellikler



#	Column
0	id
1	diagnosis
2	radius_mean
3	texture_mean
4	perimeter_mean
5	area_mean
6	smoothness_mean
7	compactness_mean
8	concavity_mean
9	concave points_mean
10	symmetry_mean
11	fractal_dimension_mean
12	radius_se
13	texture_se
14	perimeter_se
15	area_se
16	smoothness_se
17	compactness_se
18	concavity_se
19	concave points_se
20	symmetry_se
21	fractal_dimension_se
22	radius_worst
23	texture_worst
24	perimeter_worst
25	area_worst
26	smoothness_worst
27	compactness_worst
28	concavity_worst
29	concave points_worst
30	symmetry_worst
31	fractal_dimension_worst

2) Bu Alanda Veri Bilimi'nin Kullanmanın Amacı

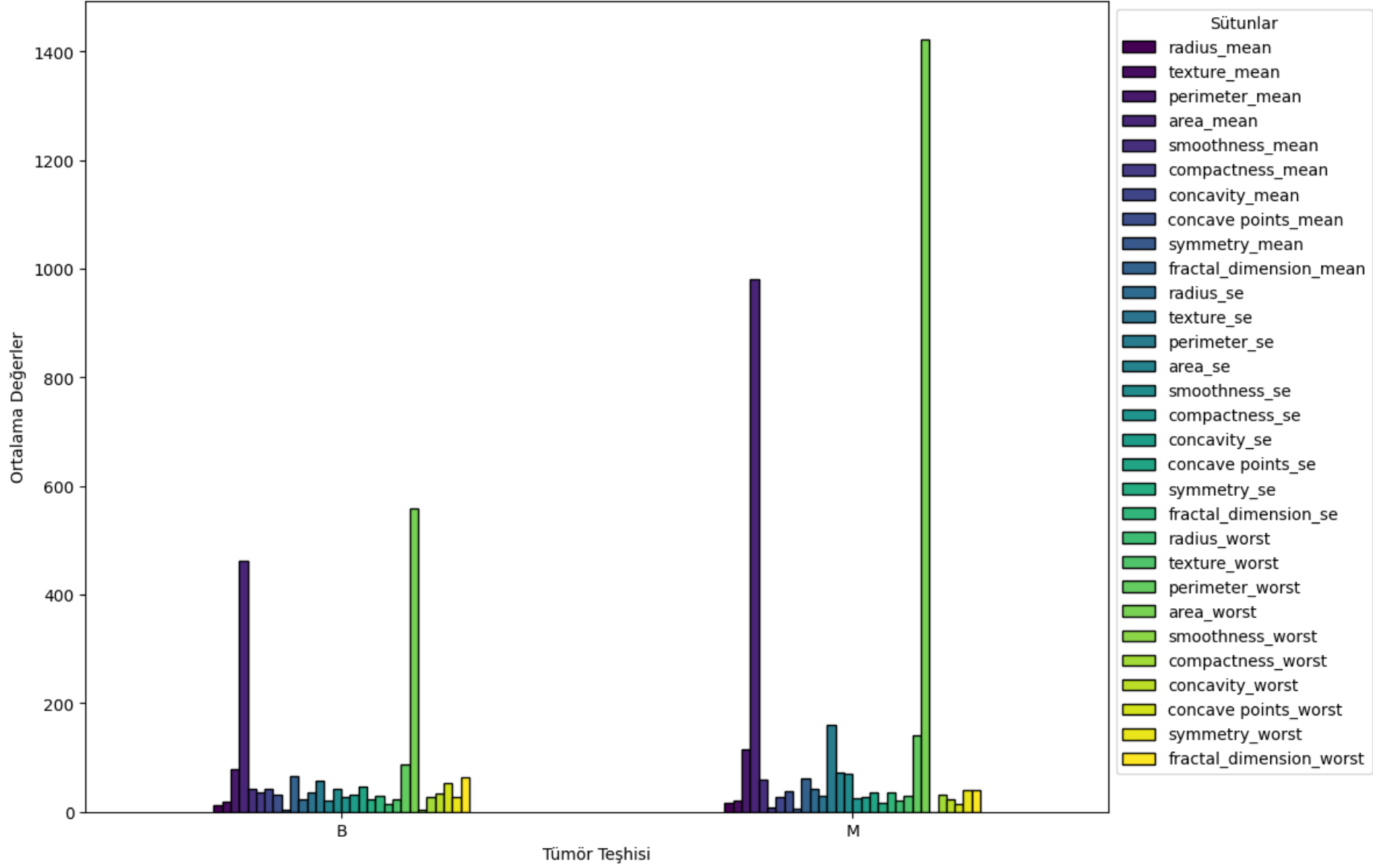
★ **Erken Teşhis** = Hastalığın erken aşamalarında teşhisin yapılmasına yardımcı olabilir. Erken teşhis, tedavi başarısını artırabilir.

★ **Veri Keşfi** = Veri setlerinin analizi, göğüs kanseri ile ilgili yeni bilgilerin keşfedilmesine yardımcı olabilir. Örneğin, verilerden yeni mutasyonlar keşfedildiğinde ona göre ilaçlar , tedaviler üretilir ve ölüm riski azaltılmış olur.

3) Sütunların B ve M bazlı Ortalamaları

diagnosis	B	M
radius_mean	12.146524	17.462830
texture_mean	17.914762	21.604906
perimeter_mean	78.075406	115.520048
area_mean	462.483989	979.957346
smoothness_mean	43.115004	59.402658
compactness_mean	35.588609	7.315709
concavity_mean	42.190003	26.737373
concave points_mean	32.133070	37.636424
symmetry_mean	4.092240	6.401542
fractal_dimension_mean	65.601529	62.104519
radius_se	22.958942	42.259741
texture_se	36.781057	28.651371
perimeter_se	57.624225	160.723976
area_se	21.135148	72.672406
smoothness_se	41.962549	70.563052
compactness_se	27.688648	25.017528
concavity_se	31.432529	27.468445
concave points_se	45.508017	35.655019
symmetry_se	22.573941	16.331892
fractal_dimension_se	28.505454	36.116479
radius_worst	13.379801	21.120521
texture_worst	23.515070	29.318208
perimeter_worst	87.005938	141.370330
area_worst	559.670506	1422.286321
smoothness_worst	3.900430	0.257506
compactness_worst	26.633386	30.514564
concavity_worst	34.684294	23.713164
concave points_worst	53.353696	13.507145
symmetry_worst	26.661637	39.095723
fractal_dimension_worst	63.259112	39.794865

Kötü ve İyi Huylu Tümörler İçin Sütun Ortalamaları



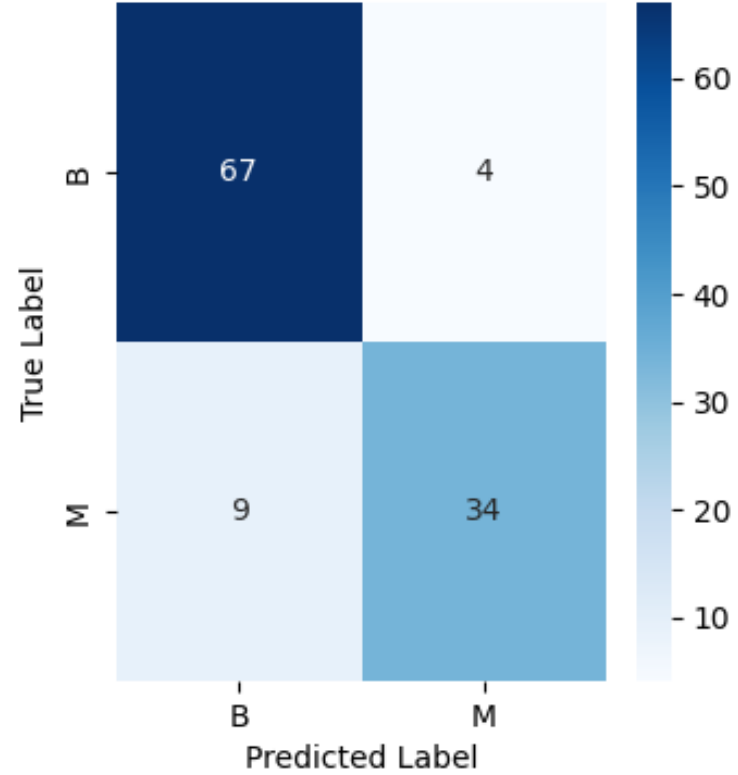
4) Modeller

- 1) K-Nearest Neighbors
- 2) Random Forest
- 3) Gradient Boosting
- 4) Logistic Regression
- 5) Support Vector Machine
- 6) Decision Tree
- 7) Naive Bayes

5) Modellerin Performansı ve Tıbbi Bağlamda Yorumlanması

K-Nearest Neighbors Modeli Test Seti Performansı:
F1 Score: 0.8395
Precision: 0.8947
Recall: 0.7907
Accuracy: 0.8860

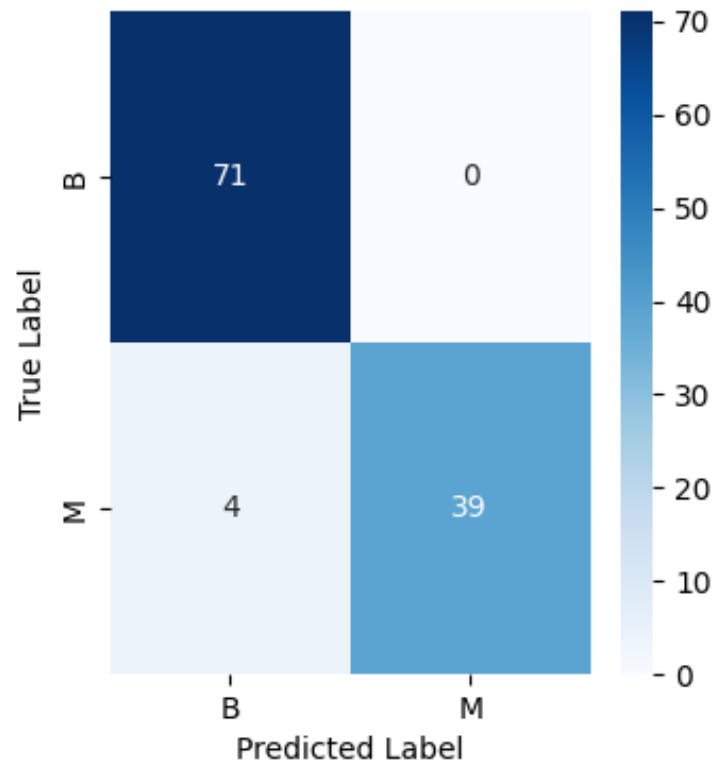
K-Nearest Neighbors - Confusion Matrix



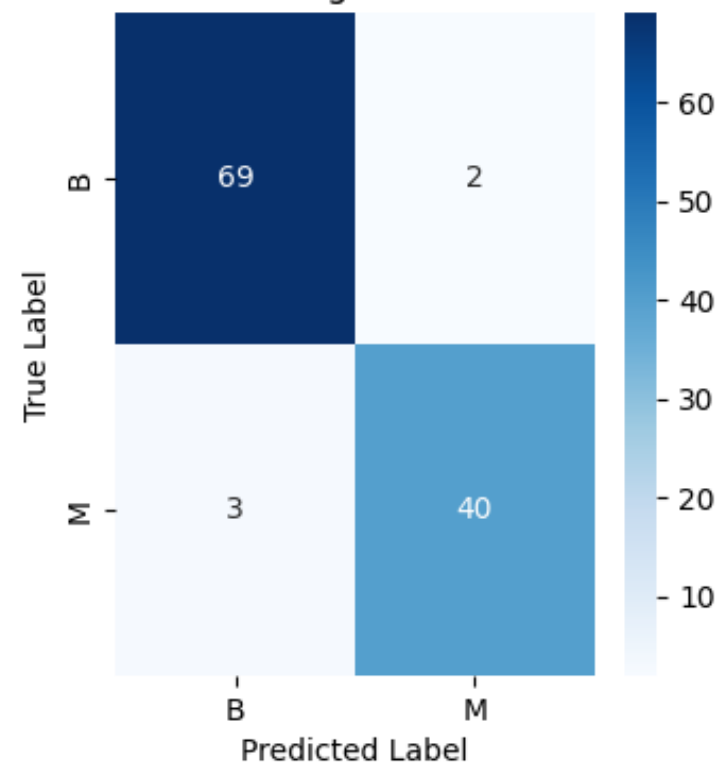
Random Forest Modeli Test Seti Performansı:
F1 Score: 0.9512
Precision: 1.0000
Recall: 0.9070
Accuracy: 0.9649

Gradient Boosting Modeli Test Seti Performansı:
F1 Score: 0.9412
Precision: 0.9524
Recall: 0.9302
Accuracy: 0.9561

Random Forest - Confusion Matrix

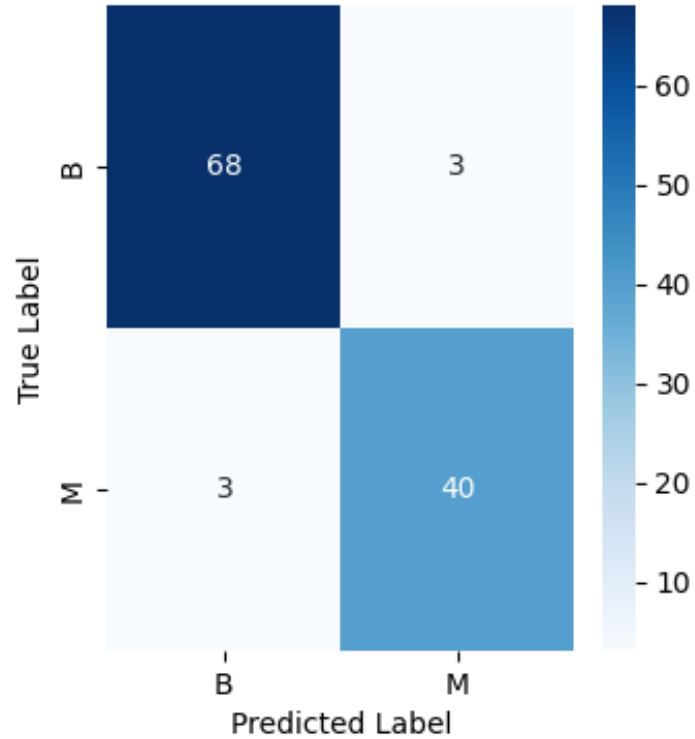


Gradient Boosting - Confusion Matrix



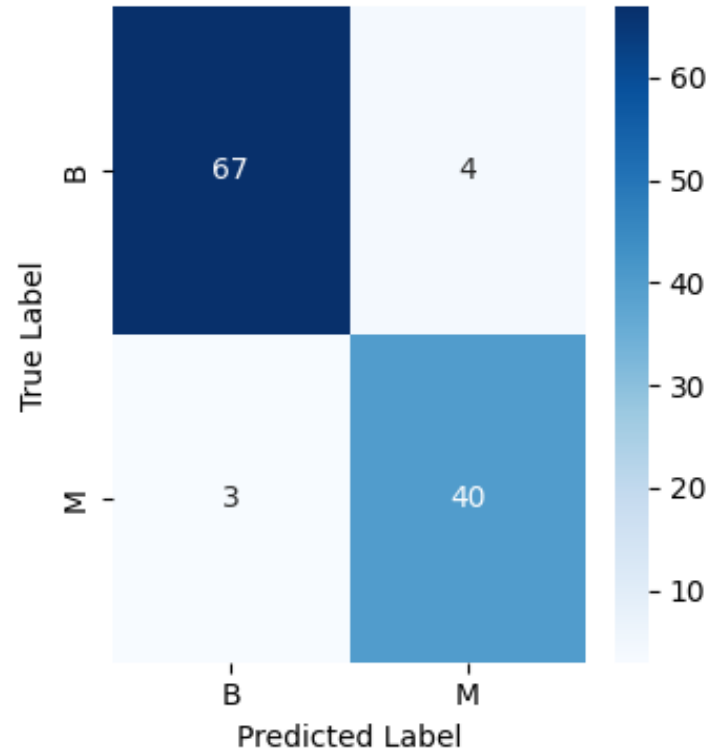
Logistic Regression Modeli Test Seti Performansı:
F1 Score: 0.9302
Precision: 0.9302
Recall: 0.9302
Accuracy: 0.9474

Logistic Regression - Confusion Matrix

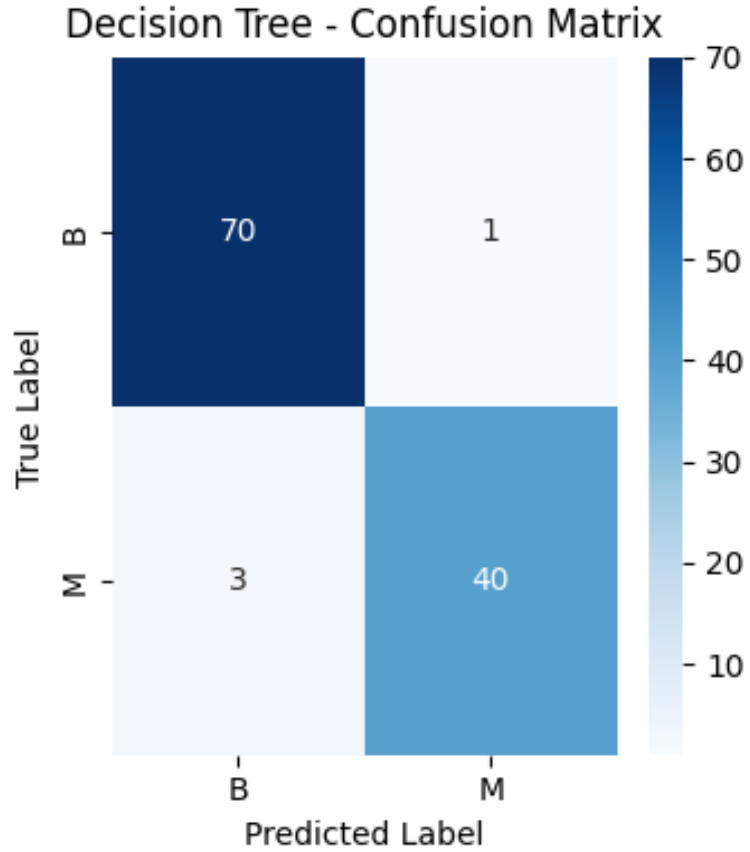


Support Vector Machine Modeli Test Seti Performansı:
F1 Score: 0.9195
Precision: 0.9091
Recall: 0.9302
Accuracy: 0.9386

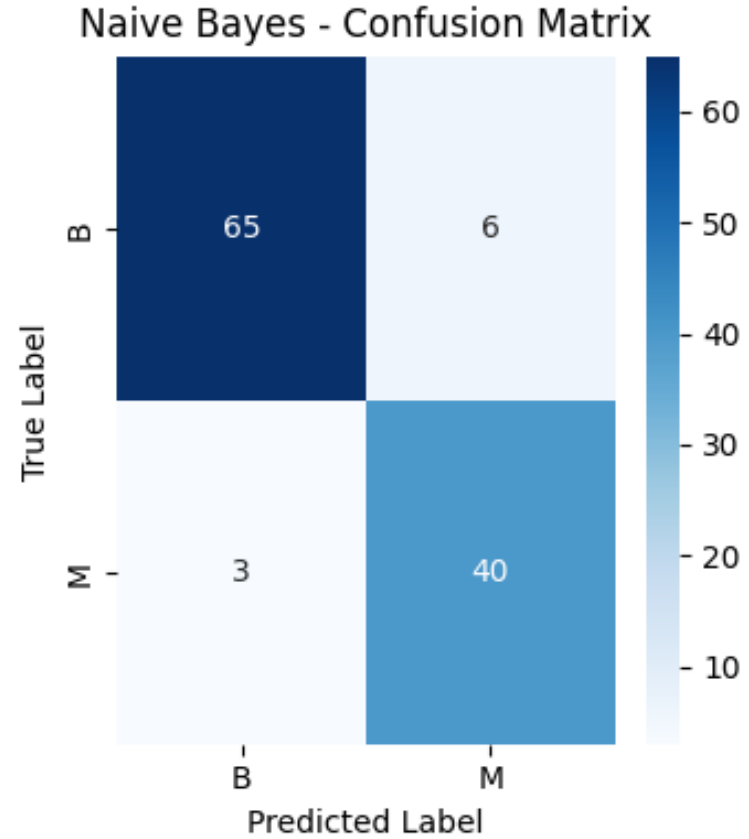
Support Vector Machine - Confusion Matrix



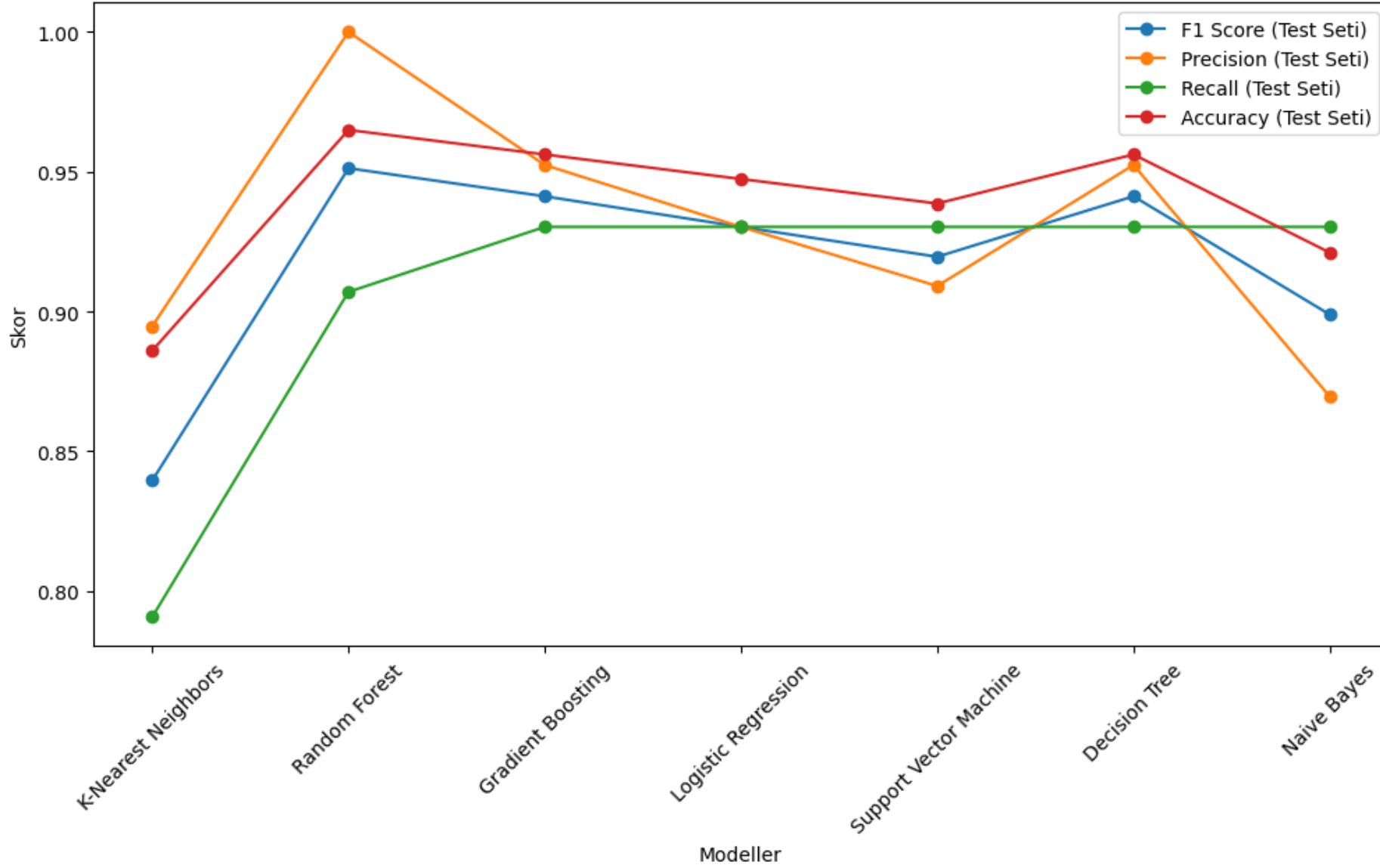
Decision Tree Modeli Test Seti Performansı:
F1 Score: 0.9524
Precision: 0.9756
Recall: 0.9302
Accuracy: 0.9649



Naive Bayes Modeli Test Seti Performansı:
F1 Score: 0.8989
Precision: 0.8696
Recall: 0.9302
Accuracy: 0.9211



Modellerin Test Seti Performans Karşılaştırması



TEŞEKKÜRLER

