

1906003132015

Doğal Dil İşleme

BAİBÜ Bilgisayar Müh.

Dr. Öğr. Üyesi İsmail Hakkı Parlak

ismail.parlak@ibu.edu.tr

Oda: 335

Dengesiz Veri Dağılımları

- Veri seti = {900 Türkçe, 100 İng.}
 - Eğitim kümesi (X_train) = {810 Türkçe, 90 İng.}
 - Test kümesi (X_test) = {90 Türkçe, 10 İng.}

```
model.train(X_train)
tahminler = model.predict(X_test)
np.mean(tahminler == y_test) # 0.9
```

- Model başarılı mı?

Dengesiz Veri Dağılımları

Class Imbalance

- Veri seti = {90 sağlıklı (-), 10 hasta (+)}
- positive (+): hastalık var.
- negative (-): hastalık yok.

		Gerçekler	
		(+)	(-)
Tahminler	(+)	True positive	False positive
	(-)	False negative	True negative

Dengesiz Veri Dağılımları

		Gerçekler	
		(+)	(-)
Tahminler	(+)	True positive	False positive
	(-)	False negative	True negative

- True positive (TP): Tahmin = hasta ve doğru
- True negative (TN): Tahmin = hasta değil ve doğru
- False positive (FP): Tahmin = hasta ve yanlış
- False negative (FN): Tahmin = hasta değil ve yanlış

Dengesiz Veri Dağılımları

- True positive (TP): Tahmin = hasta ✓
- True negative (TN): Tahmin = hasta değil ✓
- False positive (FP): Tahmin = hasta ✗
- False negative (FN): Tahmin = hasta değil ✗
- Precision = $\frac{TP}{TP+FP}$
- Recall = $\frac{TP}{TP+FN}$
- F1 Score = $\frac{2 \times Precision \times Recall}{Precision + Recall}$
- F1 Score = Precision ve Recall'un harmonik ortalaması
- Precision, recall ve F1 skorunun önemi çalışılan alana göre belirlenir.

Dengesiz Veri Dağılımları

```
from sklearn.metrics import f1_score
```

```
model.fit(X_train, y_train)
```

```
predictions = model.predict(X_test)
```

```
f1_score(y_test, predictions)
```