

## Veri Bilimine Giriş Dersi Proje Ödevi

### Regresyon Yapılacak Projeler İçin;

- Derste anlatılan regresyon modellerinden seçeceğiniz 4 tanesi ile en iyi parametrelere sahip (tune edilmiş) tahmin modellerinizi oluşturup test ediniz. Çalışma sonucunda 4 modeli test rmse hataları ve  $r^2$  değerleri bakımından karşılaştırınız. Karşılaştırma işleminin daha iyi anlaşılabilmesi için aşağıdaki örnekte verildiği gibi bir tablodan faydalanınız. Tablo görünümü için isterseniz pyhton ile jupyter lab ortamında dataframe veya pivot tablo vb. yöntemlerden, isterseniz harici bir Word veya excel dosyasından faydalanabilirsiniz. Ayrıca kullandığınız modellerinizin karşılaştırılması işlemini derste sınıflandırma modelleri konusunda anlatılan **Tüm Modellerin Karşılaştırılması** başlığı altında gösterildiği gibi rmse bakımından barplot ile görsel olarak da gerçekleştiriniz.

Veri Seti Hitters	Sınıflandırıcı Algoritması	rmse	$r^2$
	KNN	413.71	0.21
	Doğrusal SVR	367.54	0.38
	Doğrusal Olmayan SVR	428.0	0.15
	Çok Katmanlı Algılayıcı	377.03	0.34
	CART	376.39	0.35
	<b>Bagged Trees</b>	<b>346.45</b>	<b>0.45</b>
	Random Forest	346.79	0.44

### Sınıflandırma Yapılacak Projeler İçin;

- Derste anlatılan sınıflandırma modellerinden seçeceğiniz 4 tanesi ile en iyi parametrelere sahip (tune edilmiş) tahmin modellerinizi oluşturup test ediniz. Çalışma sonucunda 4 modeli test accuracy(doğruluk), precision(kesinlik), recall(duyarlılık), f1-score (f1 –skoru) ölçüt değerleri bakımından karşılaştırınız. Karşılaştırma işleminin daha iyi anlaşılabilmesi için aşağıdaki örnekte verildiği gibi bir tablodan faydalanınız. Tablo görünümü için isterseniz pyhton ile jupyter lab ortamında dataframe veya pivot tablo vb. yöntemlerden, isterseniz harici bir Word veya excel dosyasından faydalanabilirsiniz. Ayrıca kullandığınız modellerinizin karşılaştırılması işlemini derste anlatılan **Tüm Modellerin Karşılaştırılması** başlığı altında gösterildiği gibi accuracy (doğruluk) bakımından barplot ile görsel olarak da gerçekleştiriniz.

Veri Seti Diabetis	Sınıflandırıcı Algoritması	Doğruluk	Ortalama_Kesinlik	Duyarlılık		Ortalama _f1-Score
				Sınıf	Değer	
	Lojistik Regresyon	0.75	0.73	0.	0.83	0.72
				1.	0.60	
	Naive Bayes	0.74	0.72	0.	0.79	0.72
				1.	0.66	
	KNN	0.73	0.70	0.	0.81	0.70
				1.	0.57	
	Doğrusal SVM	0.75	0.72	0.	0.81	0.74
				1.	0.62	
	Doğrusal Olmayan SVM	0.74	0.71	0.	0.79	0.71
				1.	0.62	

	YSA(ANN)	0.75	0.72	0.	0.83	0.71
				1.	0.60	
	CART	0.75	0.73	0.	0.85	0.71
				1.	0.56	
	Random Forest	0.76	0.74	0.	0.80	0.74
				1.	0.69	
	GBM	0.74	0.71	0.	0.76	0.72
				1.	0.69	
	XGBoost	0.77	0.74	0.	0.82	0.74
				1.	0.66	
	Light GBM	0.75	0.72	0.	0.80	0.72
				1.	0.65	