

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
FEN FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

ÜRETİM HATALARININ ANALİZİ VE MAKİNE ÖĞRENMESİ İLE
TAHMİNLEMESİ

Staj Raporu

ALPER ENGİN

08,2022

Özet

Bu projemizde, üretilen kumaş toplarında oluşan hatalara ilişkin analizleri ve kumaşta oluşan hatalara ve üretim miktarına bakarak kumaş kalitesinin makine öğrenmesi ile tahminlemesi yapılmıştır.

Anahtar kelimeler: hata, makine öğrenmesi

İçindekiler

Giriş	3
Bölüm 1: Tanımlayıcı İstatistikler	5
1.0.1 Ek Hatası	7
1.0.2 Leke Hatası	8
1.0.3 Delik Hatası	9
1.0.4 Baskı Hatası	10
Bölüm 2: Kalite Düzeyinin Tahminlenmesi	11
2.0.1 Veriyi Ön İşleme	11
2.0.2 Random Forest Modeli	12
2.0.3 Knn Modeli	13
2.0.4 XGboost Modeli	14
Bölüm 3: Hatanın Konumunun Tahminlenmesi	15
3.0.1 Veriyi Hazırlama	15
3.0.2 Random Forest Modeli	16
3.0.3 Knn Modeli	17
3.0.4 Xgboost Modeli	18
Sonuç	19
Kaynaklar	21

Şekil Listesi

1.2	Veri Seti	5
1.3	Verinin tanımlayıcı istatistikleri	6
1.4	Ek Hatasının histogramı	7
1.5	Ek Hatasının Kalite Düzeyleri Grafiği	7
1.6	Leke Hatası Histogramı	8
1.7	Leke Hatası Kalite Düzeyleri Grafiği	8
1.8	Delik hatası histogramı	9
1.9	Delik hatası Kalite Düzeyleri Grafiği	9
1.10	Baskı hatası histogramı	10
1.11	Baskı Hatası Kalite Düzeyleri Grafiği	10
2.1	Yeninden Hazırlanmış Veri Seti	11
2.2	Üç Sınıflı Random Forest Modeli	12
2.3	Üç Sınıflı Random Forest Modeli Matrisi	12
2.4	Üç Sınıflı Random Forest Modeli Roc Curve	12
2.5	Üç Sınıflı Knn Modeli	13
2.6	Üç Sınıflı Knn Modeli Matrisi	13
2.7	Üç Sınıflı Knn Roc Curve	13
2.8	Üç Sınıflı XGboost Modeli	14
2.9	Üç Sınıflı XGboost Modeli Matirisi	14
2.10	Üç Sınıflı Knn Roc Curve	14
3.1	Tekrar Oluşturulan Veri Seti	15
3.2	İki sınıflı Random Forest Modeli	16
3.3	İki sınıflı Random Forest Modeli Matrisi	16
3.4	Random Forest Modeli Roc Curve	16
3.5	İki sınıflı Knn Modeli	17
3.6	İki sınıflı Knn Modeli Matrisi	17
3.7	İki sınıflı Knn Roc Curve	17
3.8	İki sınıflı Xgboost Modeli	18
3.9	İki sınıflı Xgboost Modeli Matrisi	18
3.10	İki sınıflı Xgboost Modeli Roc Curve	18

Giriş

Stajımı Beztas  firmasında ham kumaş üretim bölümünde kalite departmanında yaptım. Stajdaki görevlerim üretilen kumaşlarda hata analizi, işletmenin üretim raporlarını hazırlama ve üretim planları oluşturmadır. Bu çalışmada işletmenin üretimi sırasında oluşan hataların analizi ve makine öğrenmesi yöntemlerini kullanarak kumaşın kalitesinin tahminlemesi yapılmıştır.

Bölüm 1

Tanımlayıcı İstatistikler

Projemde kullandığım veri 3 adet kalite ve 4 hata çeşidinden oluşmaktadır. Kalite çeşitleri iyiden kötüye FASON , JAVANEZ ve RAYON GEORGETTE olarak sıralanmaktadır. Hata çeşitleri ise Ek, Baskı hatası, delik ve lekedir. Bu veriyi kullandandan önce hatanın toplam metre fazla girildiği yanlış veriler temizlendi.

FASON	1496
15302 RAYON GEORGETTE	314
15288 JAVANEZ	74

(a) Kalite çeşitleri

LEKE	907
EK	333
DOKUMA HATASI	330
DELİK	314

(b) Hata çeşitleri

	KALİTE	TOP NO	TOP MT	HATA ADI	HATA METRESİ
0	15288 JAVANEZ	S22001221M08	100.0	DELİK	77.69
1	15288 JAVANEZ	S22001221M10	100.0	DELİK	68.32
2	15288 JAVANEZ	S22001221M15	91.2	DELİK	21.16
3	15288 JAVANEZ	S22001221M23	100.0	DELİK	19.02
4	15288 JAVANEZ	S22001221M32	103.0	DELİK	50.32
...
1879	FASON	F22000042M05	90.0	EK	6.81
1880	FASON	F22000042M05	90.0	EK	83.53
1881	FASON	F22000042M06	211.9	EK	56.39
1882	FASON	F22000042M07	142.8	EK	23.17
1883	FASON	F22000042M07	142.8	EK	72.21

Şekil 1.2: Veri Seti

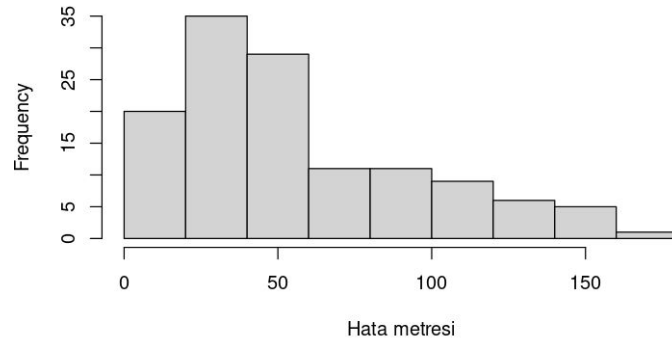
Aşağıda veri setinin tanımlayıcı istatistikleri yer almaktadır. En yüksek üretimden elde edilen top metresi 294m ve hatanın görüldüğü en yüksek metre de 241m olarak görülmüştür.

	vars	n	mean	sd	median	trimmed	mad	min	max
X	1	1884	18868.29	9519.00	23396.50	20476.24	1153.46	0	24846.00
KALİTE*	2	1884	2.75	0.51	3.00	2.87	0.00	1	3.00
TOP.NO*	3	1884	276.22	162.67	252.00	266.61	176.43	1	623.00
TOP.MT	4	1884	122.10	49.88	121.70	121.51	50.85	12	294.00
HATA.ADI*	5	1884	2.97	1.15	3.00	3.09	1.48	1	4.00
HATA.METRESI	6	1884	57.17	42.70	47.88	52.55	41.85	0	241.64
	range	skew	kurtosis	se					
X	24846.00	-1.44	0.09	219.31					
KALİTE*	2.00	-2.01	3.17	0.01					
TOP.NO*	622.00	0.42	-0.85	3.75					
TOP.MT	282.00	0.12	-0.51	1.15					
HATA.ADI*	3.00	-0.60	-1.16	0.03					
HATA.METRESI	241.64	0.92	0.42	0.98					

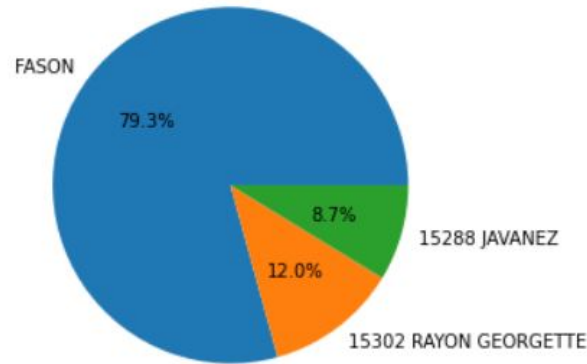
Şekil 1.3: Verinin tanımlayıcı istatistikleri

1.0.1 Ek Hatası

Üretim sırasında oluşan ek hatasının ilk 50 m de daha sık karşılaşıldığı görülmüştür. En çok delik hatasına rastlanan kumaş kalite düzeyi ise Fason olarak görülmektedir. 0.95 güven ile ek hatası 36.8330134801017 ile 41.60452405743584 metre arasında beklenmektedir.



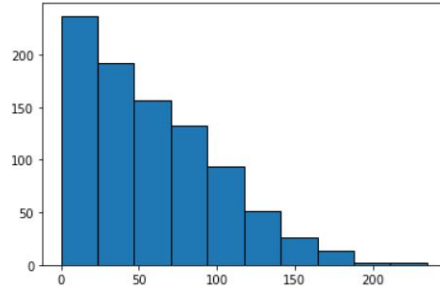
Şekil 1.4: Ek Hatasının histogramı



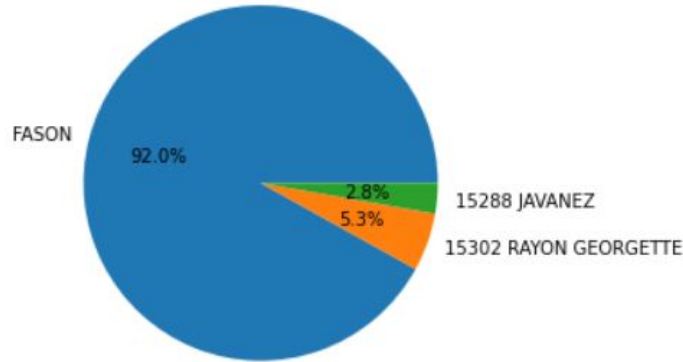
Şekil 1.5: Ek Hatasının Kalite Düzeyleri Grafiği

1.0.2 Leke Hatası

Üretim sırasında oluşan leke hatasının ilk 50 metrede çok sık , son metrelerine doğru ise azaldığı görülmektedir. En çok hatayı da Fason tipi kumaşta görülmektedir. 0.95 güven ile ek hatası 57.0198686168961 ile 66.87216959966443 metre arasında beklenmektedir.



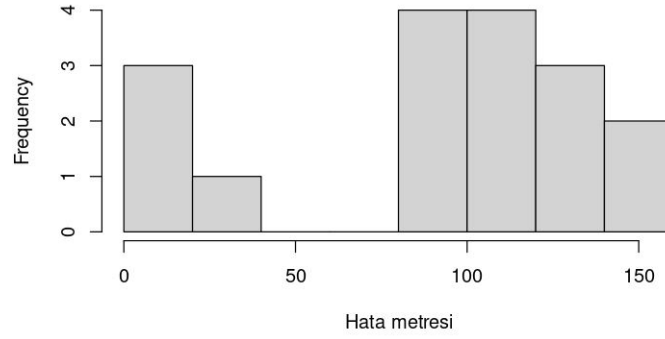
Şekil 1.6: Leke Hatası Histogramı



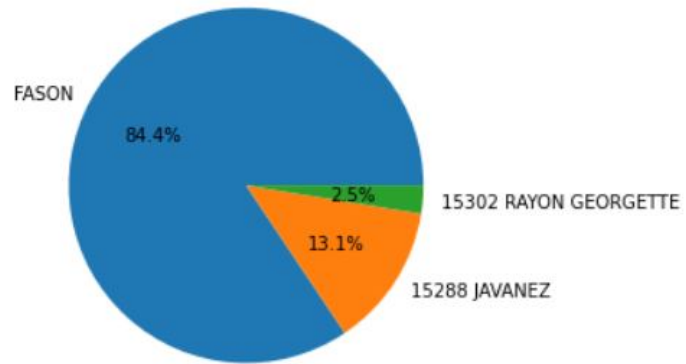
Şekil 1.7: Leke Hatası Kalite Düzeyleri Grafiği

1.0.3 Delik Hatası

Üretim sırasında oluşan leke hatası en sık 75 ile 100m arasında FASON kumaşta görülmüştür. 0.95 güven ile ek hatası 55.545003780915465 ile 61.163882657893836 metre arasında beklenmektedir.



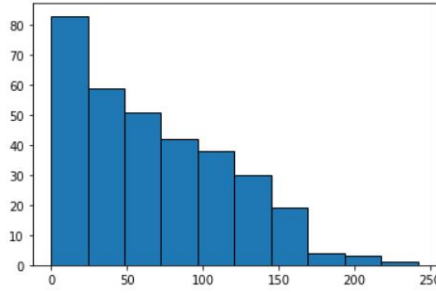
Şekil 1.8: Delik hatası histogramı



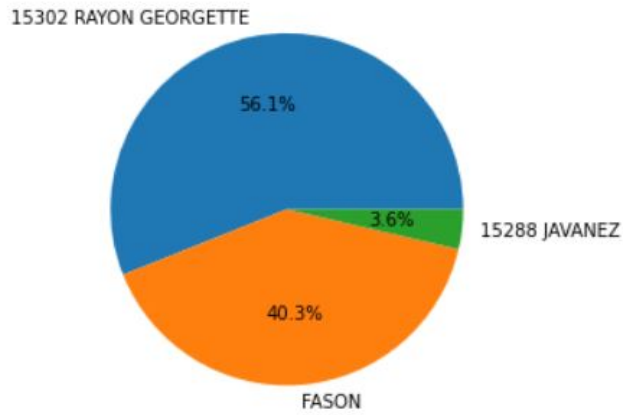
Şekil 1.9: Delik hatası Kalite Düzeyleri Grafiği

1.0.4 Baskı Hatası

Üretim sırasında oluşan baskı hatasının ilk 50m de ve RAYON GEORGETTE kumaş kalitesinde çok sık görülmüştür. 0.95 güven ile ek hatası 62.07816001546602 ile 72.88917331786729 metre arasında beklenmektedir.



Şekil 1.10: Baskı hatası histogramı



Şekil 1.11: Baskı Hatası Kalite Düzeyleri Grafiği

Bölüm 2

Kalite Düzeyinin Tahminlenmesi

2.0.1 Veriyi Ön İşleme

Veri setimizde bulunan kalite düzeyleri, birbirlerine üstünlükleri nedeniyle ordinal factor olarak dönüştürülmüştür. Hata adları ise dummy oluşturularak veri setine dahil edilmiştir. Eğitim ve test verisi ayrımı içinde 0.3 test olarak belirlenmiştir.

- Kalite düzeyleri;
 - 0 : FASON
 - 1 : JAVANEZ
 - 2 : RAYON GEORGETTE

	KALİTE	TOP MT	HATA ADI	HATA METRESİ	DELİK	DOKUMA HATASI	EK	LEKE
0	1	100.0	DELİK	77.69	1	0	0	0
1	1	100.0	DELİK	68.32	1	0	0	0
2	1	91.2	DELİK	21.16	1	0	0	0
3	1	100.0	DELİK	19.02	1	0	0	0
4	1	103.0	DELİK	50.32	1	0	0	0
...
1879	0	90.0	EK	6.81	0	0	1	0
1880	0	90.0	EK	83.53	0	0	1	0
1881	0	211.9	EK	56.39	0	0	1	0
1882	0	142.8	EK	23.17	0	0	1	0
1883	0	142.8	EK	72.21	0	0	1	0

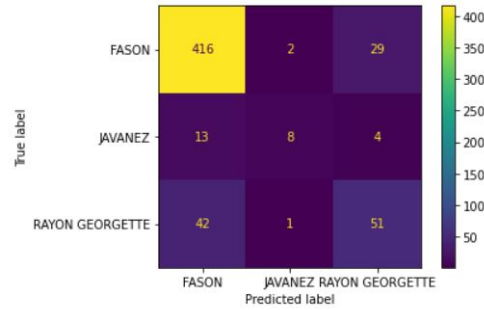
Şekil 2.1: Yeninden Hazırlanmış Veri Seti

2.0.2 Random Forest Modeli

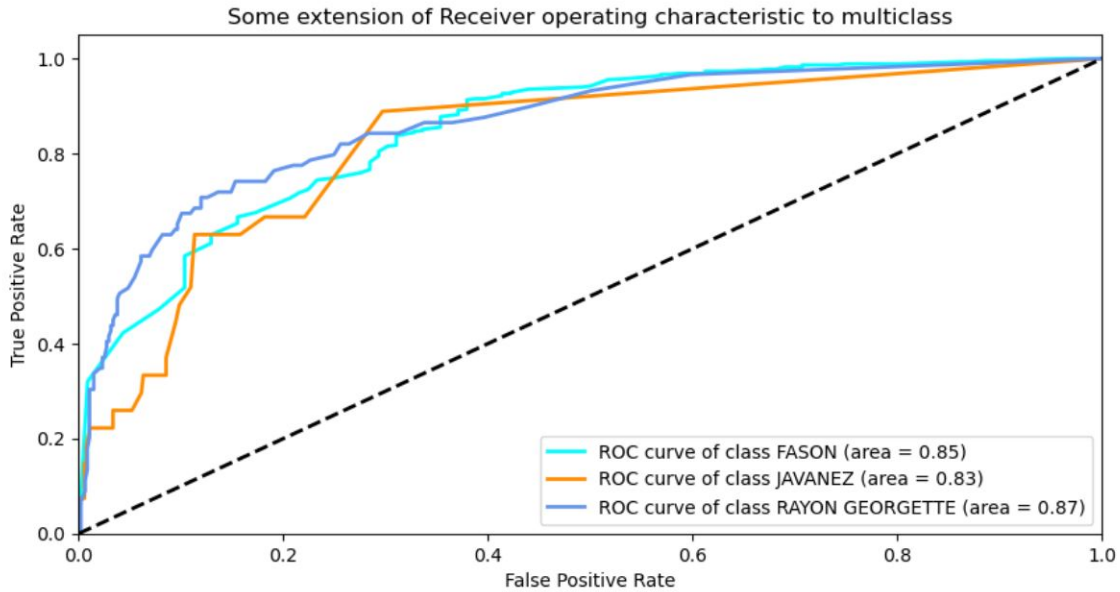
Model Fason(0) kumaş türünü tahminlemede iyi iş çıkarmış fakat javanez(1) için kötü tahminlemeler yapmıştır. Bunlara rağmen 0.82 gibi yüksek bir açıklayıcılık oranına ulaşmaktayız. TIBCO (2021)

	precision	recall	f1-score	support
0	0.86	0.94	0.90	437
1	0.86	0.18	0.30	33
2	0.61	0.51	0.56	96
accuracy			0.82	566
macro avg	0.78	0.54	0.58	566
weighted avg	0.81	0.82	0.80	566

Şekil 2.2: Üç Sınıflı Random Forest Modeli



Şekil 2.3: Üç Sınıflı Random Forest Modeli Matrisi



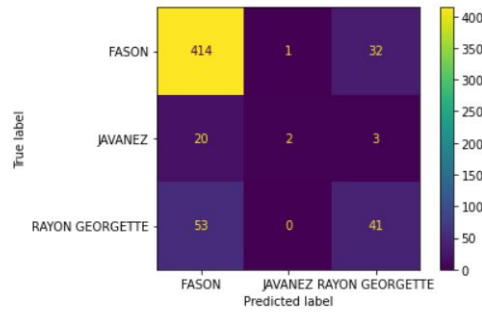
Şekil 2.4: Üç Sınıflı Random Forest Modeli Roc Curve

2.0.3 Knn Modeli

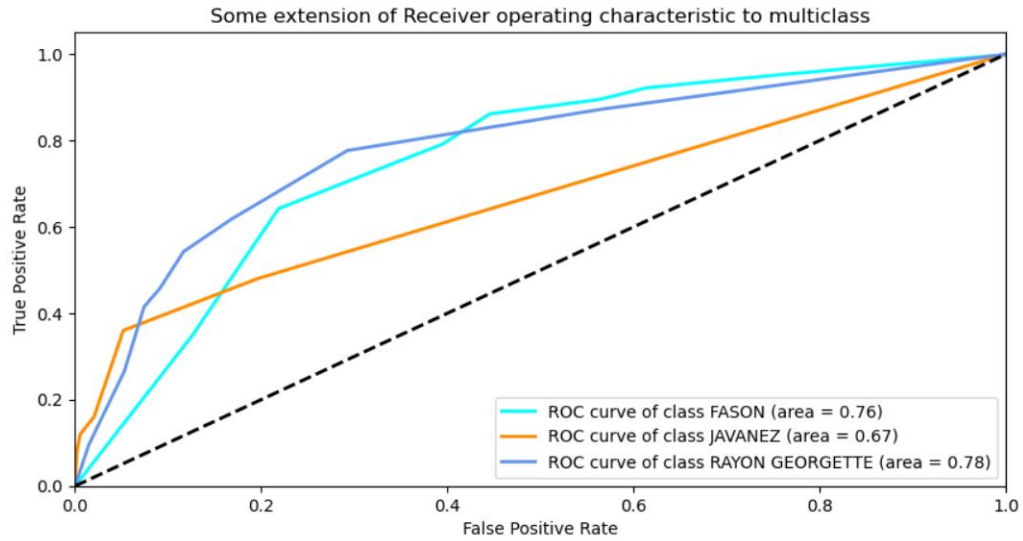
Knn modeli yüksek açıklayıcılığa ulaşmasına rağmen, 1. seviye ve 2. seviye kumaş kalite seviyelerini tahminleme de iyi bir sonuç verememiştir. Cover & Hart (1967)

	precision	recall	f1-score	support
0	0.85	0.93	0.89	447
1	0.67	0.08	0.14	25
2	0.54	0.44	0.48	94
accuracy			0.81	566
macro avg	0.69	0.48	0.50	566
weighted avg	0.79	0.81	0.79	566

Şekil 2.5: Üç Sınıflı Knn Modeli



Şekil 2.6: Üç Sınıflı Knn Modeli Matrisi



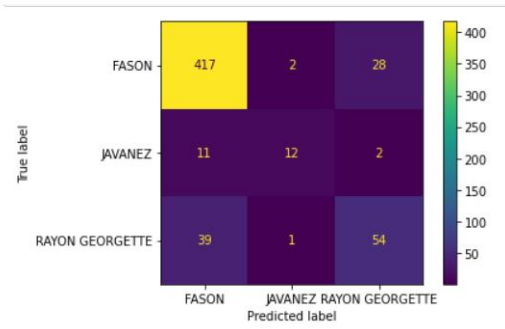
Şekil 2.7: Üç Sınıflı Knn Roc Curve

2.0.4 XGboost Modeli

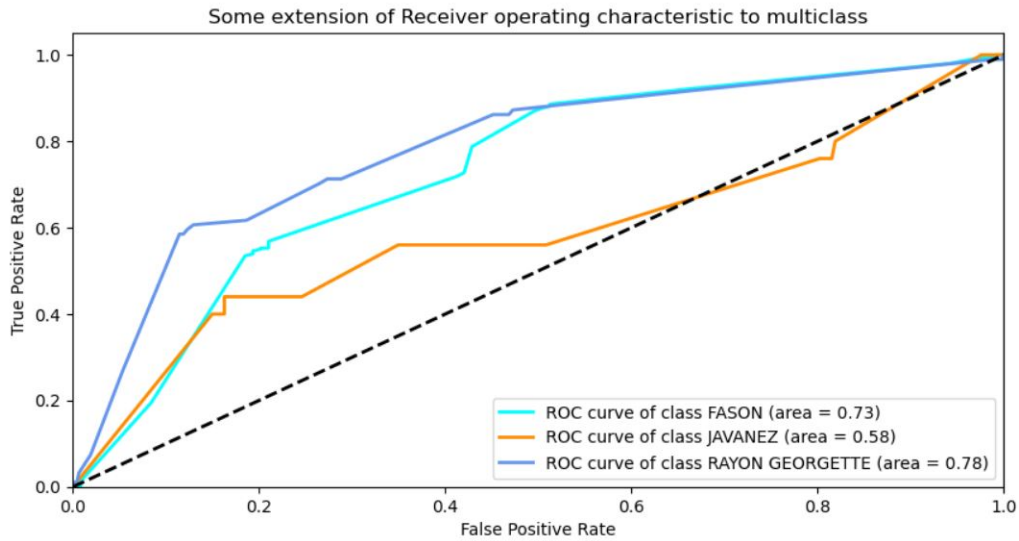
Xgboost modeli knn modeline göre bize daha iyi sonuçlar vermektedir. Roc curve de ise 1. seviye kalite düzeyi 0.5'in altında kalmıştır. Friedman (2001)

	precision	recall	f1-score	support
0	0.89	0.93	0.91	447
1	0.80	0.48	0.60	25
2	0.64	0.57	0.61	94
accuracy			0.85	566
macro avg	0.78	0.66	0.71	566
weighted avg	0.85	0.85	0.85	566

Şekil 2.8: Üç Sınıflı XGboost Modeli



Şekil 2.9: Üç Sınıflı XGboost Modeli Matirisi



Şekil 2.10: Üç Sınıflı Knn Roc Curve

Bölüm 3

Hatanın Konumunun Tahminlenmesi

3.0.1 Veriyi Hazırlama

Test-train veri setimizi bu uygulama için tekrar oluşturmaktayız. Çünkü bağımlı değişkenimiz hatanın 58mden önce ve sonra olmasına göre yeniden oluşturulmuştur.

- 0 : 58m den önce
- 1 : 58m den sonra

	KALİTE	TOP MT	HATA METRESİ	DELİK	DOKUMA HATASI	EK	LEKE
0	1	100.0	1.0	1	0	0	0
1	1	100.0	1.0	1	0	0	0
2	1	91.2	0.0	1	0	0	0
3	1	100.0	0.0	1	0	0	0
4	1	103.0	0.0	1	0	0	0
...
1879	0	90.0	0.0	0	0	1	0
1880	0	90.0	1.0	0	0	1	0
1881	0	211.9	0.0	0	0	1	0
1882	0	142.8	0.0	0	0	1	0
1883	0	142.8	1.0	0	0	1	0

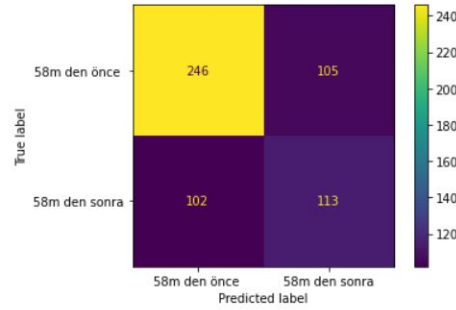
Şekil 3.1: Tekrar Oluşturulan Veri Seti

3.0.2 Random Forest Modeli

Bu tahmin de hatanın konumu, hatanın çeşidi ve kalite seviyesi gibi birden çok değişken etkili olduğu için 0.66 gibi orta düzey bir açıklayıcılık oranına ulaşmaktayız. TIBCO (2021)

	precision	recall	f1-score	support
0.0	0.69	0.75	0.72	323
1.0	0.62	0.54	0.58	243
accuracy			0.66	566
macro avg	0.65	0.65	0.65	566
weighted avg	0.66	0.66	0.66	566

Şekil 3.2: İki sınıflı Random Forest Modeli



Şekil 3.3: İki sınıflı Random Forest Modeli Matrisi



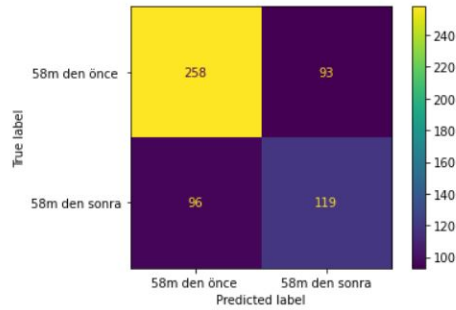
Şekil 3.4: Random Forest Modeli Roc Curve

3.0.3 Knn Modeli

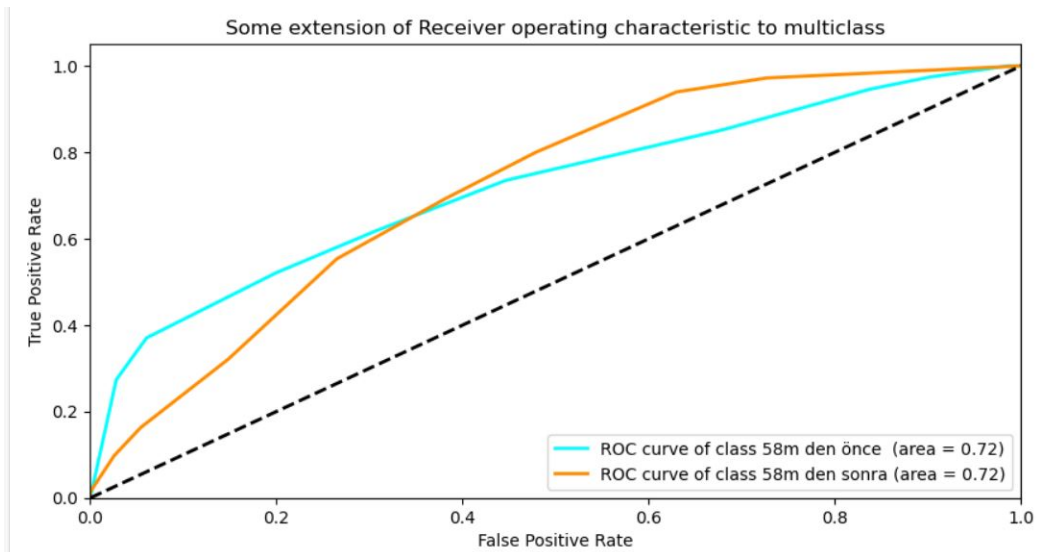
Knn modeli, 2. sınıfı tahminlemede 1. sınıfı göre daha kötü sonuçlar vermiştir. Fakat Random Forest modeline göre daha iyi iş çıkarmıştır. Cover & Hart (1967)

	precision	recall	f1-score	support
0.0	0.73	0.74	0.73	351
1.0	0.56	0.55	0.56	215
accuracy			0.67	566
macro avg	0.65	0.64	0.64	566
weighted avg	0.67	0.67	0.67	566

Şekil 3.5: İki sınıflı Knn Modeli



Şekil 3.6: İki sınıflı Knn Modeli Matrisi



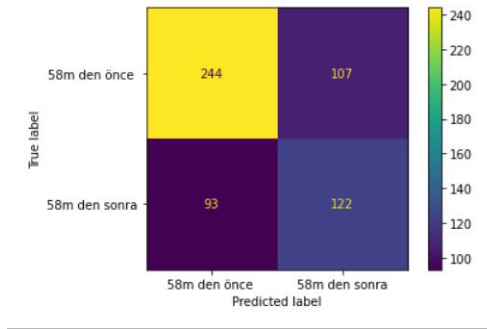
Şekil 3.7: İki sınıflı Knn Roc Curve

3.0.4 Xgboost Modeli

Xgboost modeli, tüm modellerde olduğu gibi 2. sınıfı tahminlemede kötü sonuçlar vermiş ve Random Forest modelinden daha kötü açıklayıcılığa sahip olmuştur. Friedman (2001)

	precision	recall	f1-score	support
0.0	0.72	0.70	0.71	351
1.0	0.53	0.57	0.55	215
accuracy			0.65	566
macro avg	0.63	0.63	0.63	566
weighted avg	0.65	0.65	0.65	566

Şekil 3.8: İki sınıflı Xgboost Modeli



Şekil 3.9: İki sınıflı Xgboost Modeli Matrisi



Şekil 3.10: İki sınıflı Xgboost Modeli Roc Curve

Sonuç

Çalışmamızda üretim sırasında oluşan hataları kullanarak tanımlayıcı istatistikler, oluşan hataya göre kumaş kalitesinin tahminini ve hataların konum tahminleri yapılmıştır. Genel olarak bakıldığında Fason kalite düzeyi dört hatanın üçünde en yüksek değerlere ulaşmıştır. Bu da en kalitesiz kumaş olduğunu destekleyen nitelikte bir sonuçtur. Ayrıca modellerimizin kumaş kalite tahminleri yüksek doğruluk oranına sahip olmuş ancak hataların konumunu tahminlemede yetersiz olmuştur. Daha büyük bir veri seti ve daha fazla kumaşla ilgili açıklayıcı ve tanımlayıcı değişkenler ile yüksek doğruluk oranına sahip modeller kurup, kumaştan beklenen hataları önceden tahmileyip önlem alabiliriz.

Kaynaklar

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5).

TIBCO (2021). What is a random forest?

URL <https://www.tibco.com/reference-center/what-is-a-random-forest>