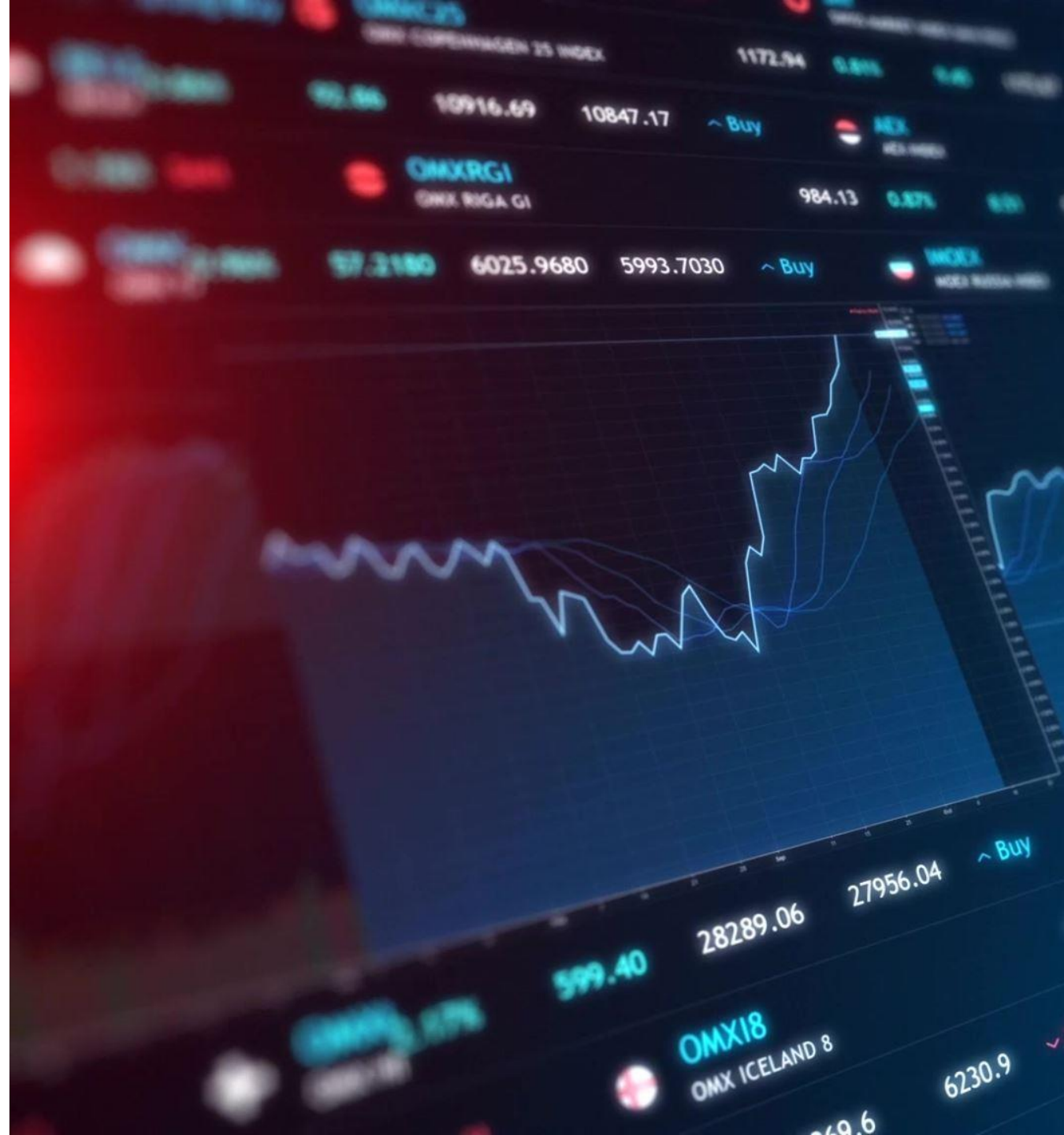




Bitcoin (BTC/USD) Fiyat Yönü Tahmini

NEDEN BTC TAHMİN?

- Günümüzde kripto piyasasında insanlar genellikle psikolojik faktörlerle hareket etmekte ve teknik analiz (indikatörler) ya da başkalarının etkisiyle işlem yapmaya yönelmektedir. Piyasanın en yüksek hacmine sahip dijital para birimi olan Bitcoin (BTC), yatırımcıların büyük ölçüde ona göre hareket etmesine neden olmaktadır. Bu nedenle, BTC'nin yarınki durumu hakkında bir öngöründe bulunabilirsek, piyasada daha bilinçli ve etkili kararlar alabiliriz.



Veriler

- Verinin ilk hali

	Price	Close	High	Low	Open	Volume
1	Ticker	BTC-USD	BTC-USD	BTC-USD	BTC-USD	BTC-USD
2	Date					
3	2020-01-01	7200.17431640625	7254.33056640625	7174.9443359375	7194.89208984375	18565664997
4	2020-01-02	6985.47021484375	7212.1552734375	6935.27001953125	7202.55126953125	20802083465
5	2020-01-03	7344.88427734375	7413.71533203125	6914.99609375	6984.4287109375	28111481032
6	2020-01-04	7410.65673828125	7427.3857421875	7309.51416015625	7345.37548828125	18444271275

- Verinin date kolonu silinmiş ve düzenlenmiş hali hali

Close	High	low	Open	Volume
457.334	468.174	452.422	465.864	21056800
424.440	456.860	413.104	456.860	34483200
394.796	427.835	384.532	424.103	37919700
408.904	423.296	389.883	394.673	36863600

- Verinin ekstra özellik eklenip NaN değerleri çıkarılınca oluşan hali

	Close	High	low	Open	Volume	RSI	MACD	Signal	SMA20	EMA20
35	386.475	392.646	380.834	382.421	14188900	47.33701	-0.823187737	-2.12075422	368.6641	383.7944
36	383.158	388.576	382.249	386.118	11641300	46.10027	-0.750194010	-1.84664218	369.0684	383.7338
37	358.417	385.048	356.447	382.962	26456900	38.10374	-1.202314313	-1.71777660	369.0136	381.3227
38	358.345	364.345	353.305	358.591	15585700	38.08304	-1.548790796	-1.68397944	370.4876	379.1343

- Son olarak gereksiz olduğunu düşündüğüm özellikleri çıkardığım ve target kolonu eklediğim hali

	Close	Open	Volume	RSI	MACD	Signal	SMA20	EMA20	Target
35	386.475	382.421	14188900	47.33701	-0.823187737	-2.12075422	368.6641	383.7944	0
36	383.158	386.118	11641300	46.10027	-0.750194010	-1.84664218	369.0684	383.7338	0
37	358.417	382.962	26456900	38.10374	-1.202314313	-1.71777660	369.0136	381.3227	0
38	358.345	358.591	15585700	38.08304	-1.548790796	-1.68397944	370.4876	379.1343	0
39	347.271	358.611	18127500	34.93873	-2.043087510	-1.75580105	371.8256	376.0997	1
40	354.704	347.487	11272500	38.60299	-2.253267409	-1.85529433	373.0569	374.0620	0
41	352.989	354.777	13033000	38.07023	-2.432461074	-1.97072768	373.8970	372.0551	1

Kullanılan Özellik Seçimleri

- **SelectFromModel:** Ortalama önem puanının üzerindeki alıyor.
- **RFE:** Özellikleri sırayla çıkararak en etkili 5 taneyi buluyor.
- **Feature Importance:** Rastgele ormandan gelen önem puanlarına göre en iyi 3 tanesini seçiyor.



Kullanılan Modeller

1. Logistic Regression (Lojistik Regresyon)

- **Avantajları:** Hızlıdır, yorumlanması kolaydır ve özellikle doğrusal ilişkilerde iyi performans gösterir.
- **Bizim kullanımımız:** "Fiyat yarın artacak mı?" sorusuna olasılıksal yanıt vermek için.

2. K-Nearest Neighbors (KNN) – En Yakın Komşular

- **Avantajları:** Parametre ayarı azdır ve modelin eğitimi gerekmez (hafıza tabanlıdır).
- **Dezavantajları:** Yavaş çalışabilir, özellikle büyük veri setlerinde.
- **Bizim kullanımımız:** Teknik göstergelere en çok benzeyen geçmiş verilerin sonucuna bakarak tahmin yapmak.

3. Random Forest (Rastgele Orman)

- **Avantajları:** Aşırı öğrenmeye (overfitting) karşı dayanıklıdır, önemli değişkenleri belirleme yeteneği vardır.
- **Bizim kullanımımız:**
 - Hem özellik seçimi için kullanıldı (feature importance),
 - Hem doğrudan model olarak test edildi.

4. Support Vector Machine (SVM)

- **Avantajları:** Özellikle küçük ve karmaşık veri setlerinde yüksek başarı gösterir.
- **Dezavantajları:** Büyük veri setlerinde yavaş olabilir, parametre ayarı önemlidir.
- **Bizim kullanımımız:** RFE yöntemiyle seçilen en iyi özellikler kullanılarak final model olarak kaydedildi.

Sonuç

- RFE yöntemiyle seçilen özelliklerle eğitilen SVM modeli en yüksek doğruluk oranını sağlamıştır (Doğruluk: 0.5272, Hassasiyet: 0.5315, Duyarlılık: 0.8954, F1 Skoru: 0.6671). Bazı modeller daha yüksek F1 skoru ve duyarlılık göstermiş olsa da, bir sonraki günü en doğru şekilde tahmin etmek benim için öncelikli hedef olduğundan, doğruluk oranı en yüksek olan bu modeli tercih ettim.