

NIFTY 50 ENDEKSİNDEKİ VERİLERİN ZAMAN SERİSİ MODELİ PROPHET İLE ANALİZİ

ANALYSIS OF NIFTY 50 INDEX DATA WITH TIME SERIES MODEL PROPHET

Yasir SAYAR

Bilgisayar Mühendisliği
Amasya Üniversitesi
yasirrrsayarr@hotmail.com

Özet

Finansal zaman serisi analizleri, son dönemde büyük önem kazanmış olan alanlardan birisidir. Bu çalışma kapsamında, Hindistan'da bulunan borsa piyasasında önemli endekslerden birisi olan NIFTY50'de yer alan şirketlerin finansal verileri üzerinden zaman serisi analizi ile tahmin modelleri geliştirilmesi hedeflenmiştir. Zaman serisi analizi, finansal piyasalarda trend, mevsimsellik ve tahmin çalışmalarında çokça kullanılan bir tekniktir. Bu analiz belirli bir süre boyunca toplanan verilerin incelenip geleceğe dair bir tahmin yapmamızı sağlar. Zaman serisi analizinde sıklıkla ARIMA, Prophet, LSTM ve SARIMA gibi yöntemler özellikleri yakalamakta kullanılmaktadır. Bu çalışma kapsamında, Prophet zaman serisi analiz yöntemi ile NIFTY50 endeksine ait bir şirketin kapanış fiyatları üzerinde tahmin çalışması gerçekleştirilecektir. Prophet modeli, mevsimsellik ve trend özelliklerini esnek bir şekilde modelleyebilmesi ile öne çıkmaktadır.

Anahtar Kelimeler – Finans, zaman serisi, Prophet, Trend, Mevsimsellik, Borsa

1. Giriş

Finansal zaman serisi analizleri, son zamanlarda hem akademi hem de finans alanında araştırmacılar için üst tercihlerden birisidir. [1]. Bu çalışma kapsamında, Hindistan'da bulunan borsa piyasasının endekslerinden birisi olan NIFTY50'de yer alan şirketlerin finansal verileri üzerinden zaman serisi analizi ile tahmin modelleri geliştirilmesi hedeflenmiştir. Zaman serisi analizi, finansal verilerde trend, mevsimsellik ve tahmin çalışmalarında çokça kullanılan bir teknik olmakla birlikte aralarında en iyisidir.[2]. Bu analiz belirli bir süre boyunca toplanan verilerin incelenip geleceğe dair bir tahmin yapmamızı sağlar. Zaman serisi analizinde sıklıkla ARIMA, Prophet, LSTM ve SARIMA gibi yöntemler kullanılmaktadır[3][4]. Bu çalışma kapsamında, Prophet zaman serisi analiz yöntemleri ile NIFTY50 endeksine ait bir şirketin kapanış fiyatları üzerinde tahmin çalışması gerçekleştirilecektir. Prophet modeli, mevsimsellik ve trend özelliklerini esnek bir şekilde modelleyebilmesi ile öne çıkmaktadır.[5]Trend, zaman serisinin genel yönelimini/eğilimini ifade eder. Mevsimsellik ise düzenli aralıklarla veya periyotlarla tekrar eden desenlerdir.[6] ARIMA

modeli ise zaman serisinin özüne uygun yapısal modeller geliştirilerek tahminlerde bulunur[7].

2. Yöntemler

2.1. Zaman Serisi Analizi

Zaman serisi analizi, bir dizi verinin zamana göre incelenerek gelecekteki trend ve olasılıkların durumunun tahmin edilmesini sağlar. Finansal piyasalarda, zaman serisi analizi tek değişken üzerinden mevsimselliği belirlemek, trend analizi yapmak ve uzun vadeli tahminler yapmak için kullanılır. Özellikle ARIMA ve Prophet modelleri, finansal verilerin karmaşık yapısını çözümü konusunda daha başarılıdır.[5][8]

2.2. Prophet Modeli

Prophet, Facebook tarafından geliştirilen açık kaynak kodlu bir model. Facebook'un veri bilimi takımı tarafından hazırlandığı için tahmin edeceğimiz üzere mevsimsellik ve trend analizlerini diğer modellerden daha iyi yapmaktadır.[9][10] Bu model, doğrusal veya lojistik bir trend yaklaşımıyla çalışabilir ve kullanıcı tarafından tanımlanan tatil etkilerini de hesaba katabilir. Finansal piyasaların dalgalanmasına ve kolay etkilenmesine rağmen, Prophet modelinin çok yönlülüğü, onu özellikle kısa ve orta vadeli tahminlerde daha yakın bir seçenek haline getirmiştir. Bu model, trend değişiklik noktalarını otomatik olarak belirleyebilir ve zaman serisinde belirgin olan haftalık veya yıllık mevsimselliği modelleyebilir.[8][9][10][11]

3. Veri seti

NIFTY50 Veri seti kaggle sitesinden alınmıştır.

Veri, 2007'den 2021'e kadar 50 firmanın günlük (date(tarih),symbol(sembol),series(güvenlik tipi),prev close(önceki kapanış), open, high, low, last, close, VWAP (Hacim ağırlıklı ortalama)) değerlerini içermektedir. 15 sütundan, 3201 satırdan oluşmaktadır. 50 Firma için de aynı şekilde veriler mevcuttur.

4. Veri ön işleme

Veri ön işleme, her modelde olduğu gibi zaman serisi analizi sürecinde de kritik bir adımdır ve model performansının doğruluğu üzerine etkisi olabilir.[12] Bu çalışmada, NIFTY50 endeksine ait kapanış fiyatları verisi üzerinde aşağıdaki veri ön işleme adımları uygulanmıştır:

- Eksik veri kontrolü ve işlenmesi: Kaggle'dan aldığımız NIFTY 50 veri setinde eksik veri kontrolü sonucunda eksik veri olmadığı görülmüştür.
- Zaman Sütunu değişimi: Verimizde 'Date' sütunu object türünde olduğu için model eğitiminde kullanılamıyordu. Bunu çözmek için DateTime'a çevrilmiştir..
df['ds'] = pd.to_datetime(df['ds'])
- Trend ve Mevsimsellik : Prophet modeli varsayılan olarak yıllık ve haftalık mevsimsellik bileşenlerini analiz eder. İlgili grafikleri ve örneklerini ilerde değinilecektir.
- Veri setinin hazırlanması: Veri, eğitim için ilk %80, test için ise ilk %20'lik kısım olacak şekilde bölünmüştür.

Veri ön işleme adımları tamamlandıktan sonra , Prophet modelinin uygulanmasına geçilmiştir. Bu adımlar model performansını optimize etmek ve daha doğru tahminler için oldukça önemlidir.[13]

Tablo 1: Bir firmanın örnek verileri

Date	Symbol	Series	'rev Close	Open	High	Low	Last	Close
2008-05-26	BAJAJFINSV	EQ	2101.05	600	619	501	505.1	509.1
2008-05-27	BAJAJFINSV	EQ	509.1	505	610.95	491.1	564	554.65
2008-05-28	BAJAJFINSV	EQ	554.65	564	665.6	564	643	640.95
2008-05-29	BAJAJFINSV	EQ	640.95	656.65	703	608	634.5	632.4
2008-05-30	BAJAJFINSV	EQ	632.4	642.4	668	588.3	647	644
2008-06-02	BAJAJFINSV	EQ	644	650	699	622	687	686.95
2008-06-03	BAJAJFINSV	EQ	686.95	672	689.8	632.55	670.55	672.05
2008-06-04	BAJAJFINSV	EQ	672.05	674	674	566.6	595	598.95
2008-06-05	BAJAJFINSV	EQ	598.95	603	639.8	572.3	631.5	631.85

Bu haliyle 'Date' sütununun veri tipi object olduğu için Modelimize uygun bir şekilde tarihleri geçiremiyoruz. Çözüm olarak aşağıdaki satırlar ile çözüm ürettik.

```
df.columns = ['ds', 'y'] # Prophet formatına uygun hale getirme
df['ds'] = pd.to_datetime(df['ds'])
```

Tablo 2: Modele uygun tarih ve Close sütunu

Index	ds	y
0	2008-05-26 00:00:00	509.1
1	2008-05-27 00:00:00	554.65
2	2008-05-28 00:00:00	640.95
3	2008-05-29 00:00:00	632.4
4	2008-05-30 00:00:00	644
5	2008-06-02 00:00:00	686.95

Bu çalışmada kapanış değerleri üzerinden tahmin ve işlem yapacağımız için verinin son hali budur[4].

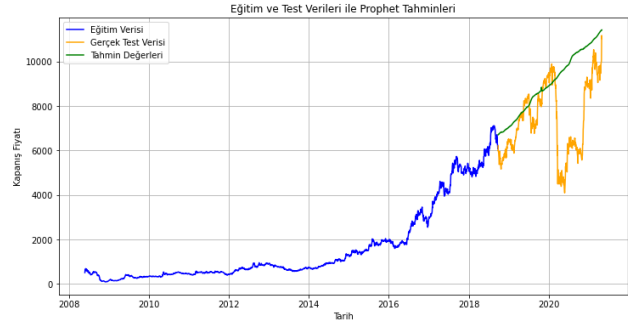
5. DeneySEL çalışma

Firmalardan bazıları Prophet modeliyle %80'i eğitim %20'si test şeklinde ayırdığımız veriyi girdi olarak verip eğitimi tamamladıktan sonra test verisindeki tarihlere karşılık gelen kapanış değerlerini modelimize tahmin ettirmeye çalışılmıştır. Bazı firmalar için model test edilmiştir. Firmaların iş alanları:

- BAJAJFINSV: Bankacılık dışı finansal hizmetler
- ADANI PORTS: Lojistik ve Liman operatörü

- MARUTI: Otomotiv
- RELIANCE: Enerji, petrokimya, tekstil, doğalgaz, eğlence, perakende, medya, eğlence vb.

Şekil 1: BAJAJFINSV firmasının tahmin ve gerçek değerleri



Tablolarda ; eğitim verisi mavi, gerçek değer turuncu, tahmin değerleri ise yeşil renkte gösterilecektir. Şekil 1 için, tahmin gerçek değerlerin genelde daha üstündedir.

Şekil 1'deki model tahmininin sonuçlarının performans ölçümleri şöyledir.

Accuracy: 70.74% (1-MAPE)

Mean Absolute Error (MAE): 1794.76

Mean Squared Error (MSE): 5849742.95

Mean Absolute Percentage Error (MAPE): 29.26%

5.1 Mean absolute error (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (1)$$

Burada N örnek sayısıdır, y_i gerçek değerdir, \hat{y}_i tahmin edilen değerdir, $e = y_i - \hat{y}_i$ tahmin hatasıdır, üstü çizgili y gerçek değerlerin ortalaması, üstü çizgili e hataların ortalamasıdır. Denklem (1) ortalama mutlak hatayı gösterir. Tahmin hatalarının mutlak değerlerinin ortalaması olarak tanımlanır.[14]

5.2 Mean square error (MSE)

$$MSE = \frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2 \quad (2)$$

(best value = 0; worst value = $+\infty$)

Tespit edilmesi gereken aykırı değerler için kullanılır. Formülde kare olduğu için aykırı değerler hatayı daha da büyütür.[15]

5.3 Mean absolute percentage Error (MAPE)

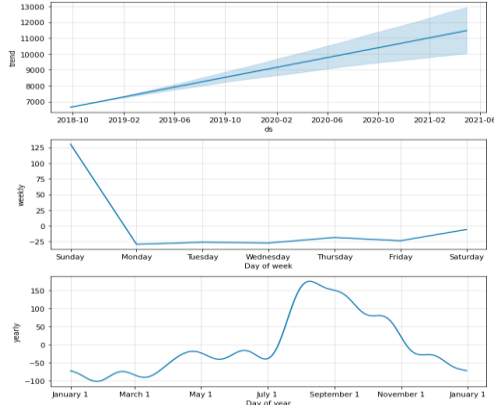
$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{Y_i - X_i}{Y_i} \right| \quad (3)$$

(best value = 0; worst value = $+\infty$)

MAPE göreceli hata açısından çok sezgisel yorum yapan bir modeldir. Mutlak varyasyonlardan ziyade göreceli varyasyonlara duyarlı olmanın daha önemli olduğu görevlerde kullanılması önerilir.[15][16][17]

Arkaplandaki analizin çıktısı için `components_fig = model.plot_components(forecast)` değişkeninin görseli aşağıdadır.

Şekil 2: BAJAJFINSV trend, haftalık ve yıllık mevsimsellik değerleri

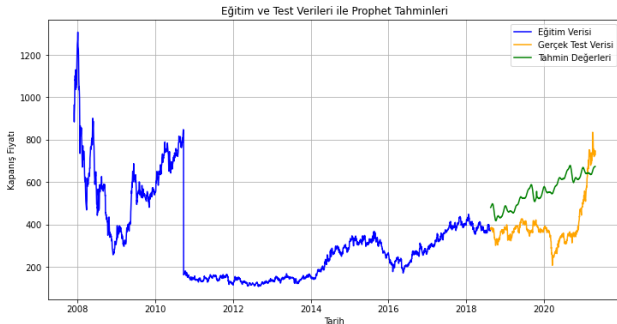


Görsel incelendiğinde, En üstteki grafik Trend grafiğidir. Zaman serisinin genel eğilimini gösterir. Grafik, zamanla kapanış fiyatlarında yükseliş trendi olduğunu göstermektedir. Mavi alan ise belirsizliği göstermektedir, zaman ilerledikçe belirsizlik artmaktadır.

Ortakdaki grafik, haftanın günlerine bağlı fiyat değişikliklerini göstermektedir. Pazartesi günleri fiyatlar daha düşükken cumartesi ve pazar gününe doğru artış olduğu görülmektedir.

En alttaki grafik, yıldaki aylara göre fiyat değişikliklerini göstermektedir. Temmuz ve eylül aylarında artarken, kasım ve aralık aylarında düştüğü görülmektedir. [19] Daha derin analizlerle bu farklılıkların neden kaynaklandığı araştırılıp daha doğru sonuçlar alınabilir. Örneğin firmanın alanına göre kışın daha fazla satış yapıp yazın daha az satış yapma durumlarına göre etkiler bulunabilir.

Şekil 3: ADANIPTS firmasının gerçek ve tahmini değerleri



Accuracy: 51.84%

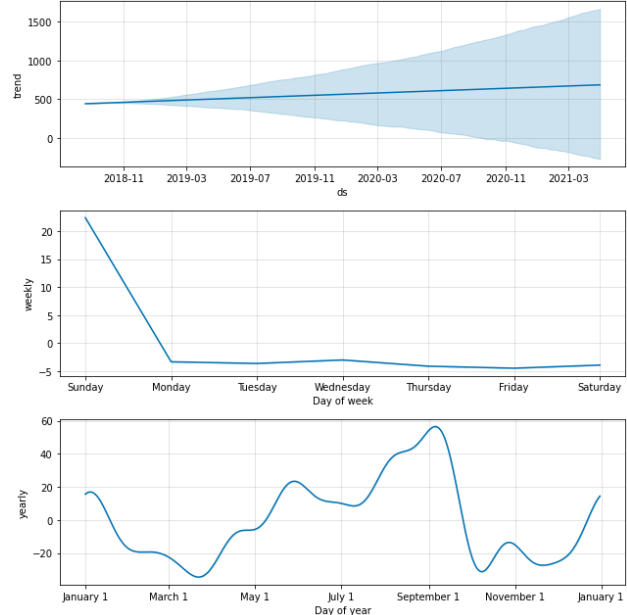
Mean Absolute Error (MAE): 171.33

Mean Squared Error (MSE): 36200.63

Mean Absolute Percentage Error (MAPE): 48.16%

Şekil 3 ve hata değerleri incelendiğinde önceki firmadan daha kötü bir tahmin yaptığı görülmektedir. Bu tahminlerde de gerçek değerden daha yüksek tahminler yapıldığı görülmektedir.

Şekil 4: ADANIPTS firmasının trend, haftalık ve yıllık mevsimsellik değerleri

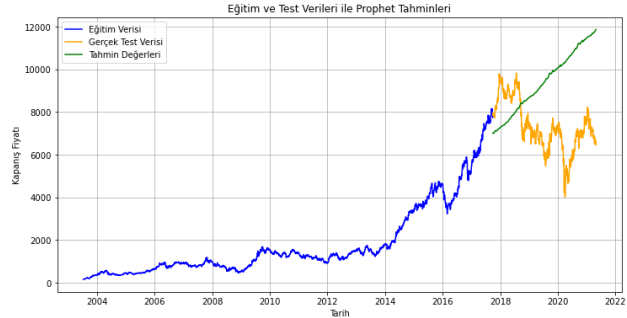


Trend grafiği, zaman geçtikçe artış çok az da olsa yatay trend göstermektedir ve belirsizlik önceki firmaya göre daha da fazla. Model uzun vadede fiyatların sabit kalacağını öngördü.

Haftalık mevsimsellik grafiği, Pazar günleri en yüksek ama pazartesi ve salı düşüş eğiliminde.

Yıllık mevsimsellik grafiği, Ekim-Kasım aylarında yükseliş görülmekte. Yani ekonomik hareketlilik veya yatırımcı durumunu yansıtabilir. Mayıs-Temmuz aylarındaki düşüş de mevsimsel bir faktöre bağlı olabilir.

Şekil 5: MARUTI firmasının gerçek ve tahmini değerleri



Accuracy: 58.24%

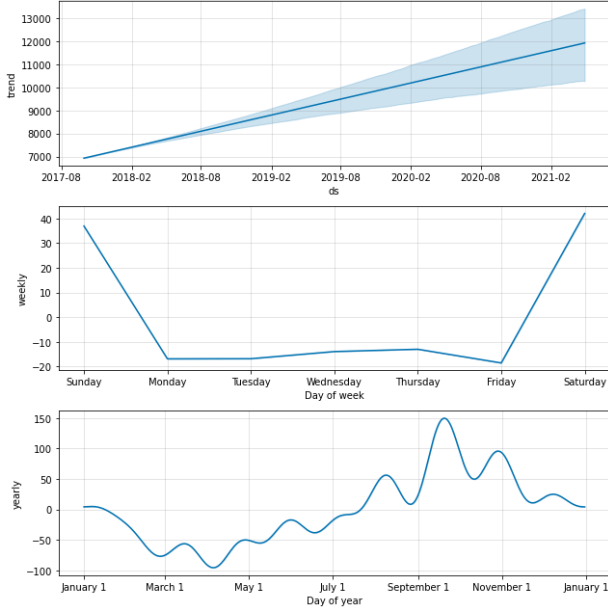
Mean Absolute Error (MAE): 2792.45

Mean Squared Error (MSE): 9964805.95

Mean Absolute Percentage Error (MAPE): 41.76%

Şekil 5 incelendiğinde, tahmin değerleri ilk safhada gerçek değerinin altında olurken 2019'dan sonra gerçek değerinin üstünde tahmin yaptığı görülmüştür.

Şekil 6: MARUTI firmasının trend ve mevsimsellik değerleri



Şekil 6'daki trend grafiği incelendiğinde, zaman geçtikçe artış olduğu görülmektedir. Belirsizlik bandı da zaman geçtikçe belirsizlik artıyor.

Haftalık mevsimsellik, Pazar ve cumartesi günleri fiyatlar daha yüksekken hafta içi fiyatlar daha düşük.

Yıllık mevsimsellik, eylül-kasım aralığında fiyatlar zirvede. Yıl sonuna doğru tekrar düşüşe geçtiği, yazın tekrar yükseldiği görülmektedir.

Şekil 7: RELIANCE firmasının gerçek ve tahmini değerleri



Accuracy: 80.20%

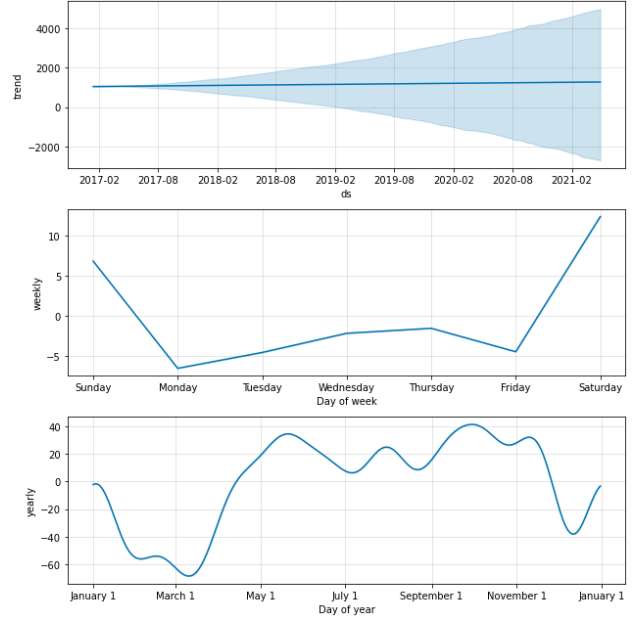
Mean Absolute Error (MAE): 309.55

Mean Squared Error (MSE): 170916.48

Mean Absolute Percentage Error (MAPE): 19.80%

Şekil 7 incelendiğinde, gerçek değerlerle içli dışlı ve yüksek doğrulukta bir tahmin yaptığı görülmektedir. Şu ana kadarki en yüksek doğrulukta tahmin bu firmaya aittir.

Şekil 8: RELIANCE firmasının trend ve mevsimsellik değerleri



Şekil 8'deki trend grafiği incelendiğinde, az artışa sahip yatay bir trend görülmektedir. Belirsizlik de zaman ilerledikçe çok artış göstermekte.

Haftalık mevsimsellik incelendiğinde, en yüksek değer cumartesi günü. En düşük değer ise pazartesi ve Salı günü en düşük değerdedir. Yıllık mevsimsellik, şubat ve mart aylarında en düşük değerdedir. Mayıs ile kasım arasında dalgalı artış ve iniş olsa da genel değer ortalamasının yüksek olduğu görülmektedir.

6. Sonuçlar

Bu çalışmada, NIFTY50 endeksinde yer alan firmaların kapanış fiyatları PROPHET modeli kullanılarak tahmin edilmiştir. PROPHET modeli, trend ve mevsimsellikteki başarısından tercih edilmiştir. Veri setinde 4 firmanın verilerinden, kullanılmayan sütunları ve boş verileri doldurduktan sonra 80 eğitim 20 test olacak şekilde model eğitilmiştir. Tahmin edilen 20lik kısmında MAE, MAPE, MSE metrikleriyle modelin başarısı incelenmiştir. Firmalara göre tahmin başarı yüzdeleri incelendiğinde en iyi tahmin yaptığı firmanın RELIANCE olduğu görülmüştür. En kötü tahmin ettiği firma ise ADANI PORTS çıkmıştır. Her bir firma için trend, haftalık ve yıllık mevsimsellik grafikleri çıkartılıp yorumlanmıştır. Daha ileri analizlerde verideki diğer sütunlar için modeller üretilip hangisi için daha doğru tahminler yapılabileceği incelenebilir. Daha fazla firma için bu model denenebilir. Hindistan'daki borsa piyasasından eğittiğimiz model ile başka ülkelerdeki borsa piyasalarında tahmin yapılabilir.

7. Kaynaklar

- [1] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing*, 90, 106181.
- [2] Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process*, 3(1), 65.
- [3] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37–45.
- [4] Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of

coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 179, 524-532.

- [5] Yenidoğan, I., Çayır, A., Kozan, O., Dağ, T., & Arslan, Ç. (2018, September). Bitcoin forecasting using ARIMA and PROPHET. In *2018 3rd international conference on computer science and engineering (UBMK)* (pp. 621-624). IEEE.
- [6] Kitagawa, G., & Gersch, W. (1984). A smoothness priors–state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, 79(386), 378-389.
- [7] Hyndman, R. J. (2018). *Forecasting: principles and practice*. OTexts.
- [8] Garlapati, A., Krishna, D. R., Garlapati, K., Rahul, U., & Narayanan, G. (2021, April). Stock price prediction using Facebook Prophet and Arima models. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-7). IEEE.
- [9] Jha, B. K., & Pande, S. (2021, April). Time series forecasting model for supermarket sales using FB-prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 547-554). IEEE.
- [10] Khayyat, M., Laabidi, K., Almalki, N., & Al-Zahrani, M. (2021). Time Series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction. *Computers, Materials & Continua*, 67(3).
- [11] Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences*, 164, 105126.
- [12] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big data analytics*, 1, 1-22.
- [13] Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. *European journal of operational research*, 156(2), 483-494.
- [14] González-Sopeña, J. M., Pakrashi, V., & Ghosh, B. (2021). An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renewable and Sustainable Energy Reviews*, 138, 110515.
- [15] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*, 7, e623.
- [16] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192, 38-48.
- [17] Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1), 69-80.
- [18] Gupta, G. K., & Ghosh, J. (2001, April). Detecting seasonal trends and cluster motion visualization for very high dimensional transactional data. In *Proceedings of the 2001 SIAM International Conference on Data Mining* (pp. 1-17). Society for Industrial and Applied Mathematics.
- [19]