

Zaman Serisi Analizi (Time Series Analysis)

Sadi Evren SEKER

Istanbul Medeniyet University, Department of Business

Özet

Bu yazıda zaman serisi analizinin ne olduğu, hangi durumlarda kullanıldığı, diğer veri madenciliği yöntemleri ile arasındaki farkların ne olduğu ve zaman serisi analizinin nasıl yapıldığı ile ilgili temel bir giriş yapılmıştır. Bu yazıda anlatılan zaman serisi analizi yöntemleri zamanı doğrusal kabul eden yöntemlerdir. Doğrusal olmayan zaman serisi analizleri de bulunmaktadır ancak yazının amacı konuya genel bir giriş yapmak olduğu için kavramsal olarak doğrusal zaman serilerine giriş yapılmıştır. Yazının diğer bir amacı ise doğrusal zaman serileri üzerinden özellik çıkarımı yaparak bu çıkarılan özelliklerin veri madenciliği adımlarına doğru şekilde verilmesidir.

Anahtar Kelimeler: Zaman Serisi Analizi, Hareketli Ortalama, Veri Madenciliği, Hata Düzeltme

Summary

This paper makes a brief introduction to time series concepts by answering the questions, “when to use time series?”, “what are the differences between time series analysis and other data mining techniques?”, “how to apply time series analysis?”. All the time series analysis methods in this paper are considered as linear methods. Although there are some non-linear time series analysis, this paper aims to make a conceptual introduction. Another purpose of the paper is making an introduction to feature extraction through time series analysis and using the features from this step on the data mining steps.

Keywords: Time Series Analysis, Moving Average, Data Mining, Data Scrubbing

Teşekkür: Bu yazının hazırlanmasında emeklerini eksik etmeyen Havva Yüksel ve Gülsüm Yiğit’e teşekkürü bir borç bilirim.

1. Giriş ve Tanım

Zaman serisi analizinin nerelerde kullanıldığını anlamak için zaman serisinin ne olduğu iyi bilinmelidir. Her seri zaman serisi değildir, bir serinin zaman serisi olabilmesi için zamana bağlı bir durum olmalıdır. Örneğin borsa değeri bir zaman serisidir, borsa değeri hesaplanırken bir önceki günün kapanış değeri bir sonraki günün değerini etkilemektedir. (Seker, Cankir, & Arslan, Information and Communication Technology Reputation for XU030 Quote Companies, 2014) Ancak borsada her bir hisse değeri hesaplanırken bir önceki günün değerlerinden etkilenmeseydi ve buna bağlı olarak her gün ölçülen hisse değerleri bir önceki günden bağımsız olarak hesaplanırdı, bu durumda değeri bir zaman serisi olmazdı. Dolayısıyla bir serinin zaman serisi olabilmesi için

eldeki verilerin en az bir tanesinin zamana bağlı olması gerekmektedir. (Seker, Mert, Al-Naami, OZALP, & AYAN, 2014)

Zaman ile ilgili olan her şey bir olayı ifade eder ve olayların hepsi bir mekânda gerçekleşir. Örneğin gün, ay ve yılın oluşması coğrafi bir olaydır ve belli bir zaman içerisinde gerçekleşir.

Kurulan bir sistem modelinde zamanın yer alıp almaması o sistemin zaman serisi olup olmamasını etkiler. Örneğin bir taşın yüksek bir yerden atılıp çarpma hızının hesaplandığı bir sistemde sonuç zamandan bağımsızdır, yani taş hangi zamanda atılırsa atılsın çarpma hızı bu zamandan etkilemez. Ancak taşın gün içinde kaç defa yüksekten atıldığına bakan bir sistemde, taşın kaç defa atıldığı gün içinde belli saatlerde farklı sonuçlar veriyorsa bu sistem bir zaman serisi olabilir.

Zaman serisi analizi yapılırken verilerin zamana bağlı değişimleri incelenir. Örneğin bir çağrı merkezinde gün içinde gelen çağrı sayısı günün belli saatlerinde fazla olurken belli saatlerinde de az olmaktadır, yani zamana bağlı olarak değişmektedir.

2. Zaman Serisi Analizinin Kullanım Amaçları

2.1. Aykırı(Outlier)Verileri Yakalama

Aykırı veri; eldeki veriler zamana göre bir sıraya yerleştirildiğinde serinin uzağında bulunan verilerdir. Verilen bir seride aykırı değeri yakalamak, o serinin hareketini yakalamayı gerektirir (Barnett, 1994). Yani aykırı değeri yakalamak; eldeki veriler üzerinde zaman serisi analizi yaparak, verilerin zaman üzerindeki hareketlerinde olan farklılığı yakalamaktır. Örneğin, günlük satışların dönemsel değiştiği veriler bir zaman aralığına yerleştirildiğinde serinin üzerinde zamana bağlı bir hareket yakalanır. Senenin belli dönemlerinde satışlardaki artış veya azalış bu hareketin belirlenmesinde rol oynar. Belirlenen zaman modeline göre veriler incelendiğinde bazı durumlarda verilerin çok küçük bir kısmının bu modelin dışında olduğu görülür ve bu veriler aykırı verilerdir. Sonuç olarak bu aykırı verilerin yakalanması için öncelikle verilerin zamana göre hareketlerinin yakalanması ve sonrasında bu hareketin dışında kalan verilerin belirlenmesi gerekmektedir.

Aykırı veriler; değerleri yanlış girilmiş veriler, sistemdeki hata sonucu oluşmuş veriler veya belli bir olaya bağlı olarak o olayın sonucunda elde edilen veriler olabilir. Dolayısıyla bu verileri yakalamak ve aykırı veri olmalarına sebep olan olayı anlamak, o olaya müdahale edebilmek için önemlidir. (Seker, Mert, Al-Naami, Ozalp, & Ayan, Correlation between the Economy News and Stock Market in Turkey, 2013)

2.2. Tahmin (Prediction)

Zaman Serisi Analizi yapılarak gerçekleşmesi muhtemel olaylar belirlenebilir. Örneğin dondurma satışlarının yazın artıp kışın azalmasına bağlı olarak veriler üzerinde zamana bağlı bir hareket bulunabilir ve dönemsel olarak bu artış ve azalışın sebebine bağlı olarak bir sonraki yılın satışları tahmin edilebilir. Başka bir örnek ise borsa teknik analizinde ve birçok alanda kullanılan hareketli ortalama yöntemidir. Hareketli ortalama grafiği ile borsanın nasıl değişeceği üzerine tahminler yapılabilmektedir. Sonuç olarak zaman serisi doğru bir şekilde yapıldığı takdirde ileriye dönük tahminler de yapılabilmektedir. Bu tahminler özellikle borsa analizi gibi finansal sistemlerde ve stratejik yönetim aşamalarında oldukça önemlidir (Seker, Cankir, & Okur, Strategic Competition of Internet Interfaces for XU30 Quoted Companies, 2014).

2.3. Eksik Verileri Tamamlama(Imputation)

Eksik olan verilerin yerine hangi değerlerin kullanılacağı zaman serisi analizi uygulanarak bulunabilir (Honaker, 2010). Örneğin veri madenciliği yapılacak olan bir satış verisinde eksik veriler olsun (Seker, Sosyal Ağlarda Akan Veri Madenciliği, 2015). Bu durumda bu veri üzerinde veri madenciliği algoritmasının uygulanabilmesi için eksik verilerin tamamlanması ya da silinmesi gerekir. Eksik verilerin tamamlanabilmesi için, eksik olan yere gelecek olan değerler zaman serisi analizi kullanılarak tahmin edilebilir. Yani eldeki satış verileri üzerinde zaman serisi analizi uygulandığında o verilerin zamana bağlı olarak hareketi yakalanır ve bu hareket doğrultusunda eksik verilerin değerleri tahmin edilebilir.

2.4. Hata Düzeltme(Data Scrubbing)

Verilerde bulunan aykırı değerlerin diğer değerlere belli yöntemlerle yaklaştırılması işleminde zaman serisi analizi kullanılır (Crosswhite, 2003).

3. Zaman Serisi Analiz Yöntemleri

3.1. Hareketli Ortalama

Hareketli ortalama yöntemi ile anlık, beklenmedik, hatalı ve aykırı verilerin genel verilerin zamana bağlı hareketi üzerindeki etkilerini azaltarak daha düz bir çizgide olmalarını sağlamaktadır. Farklı hareketli ortalama algoritmaları bulunmaktadır.

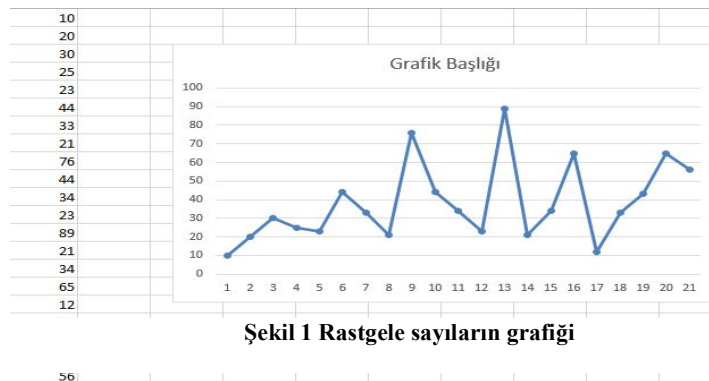
3.1.1. Basit Hareketli Ortalama(Simple Moving Average, SMA)

$$HO(SMA) = \frac{p_m + p_{m-1} + \dots + p_{m-(n-1)}}{n}$$

Formül 1 Hareketli ortalama

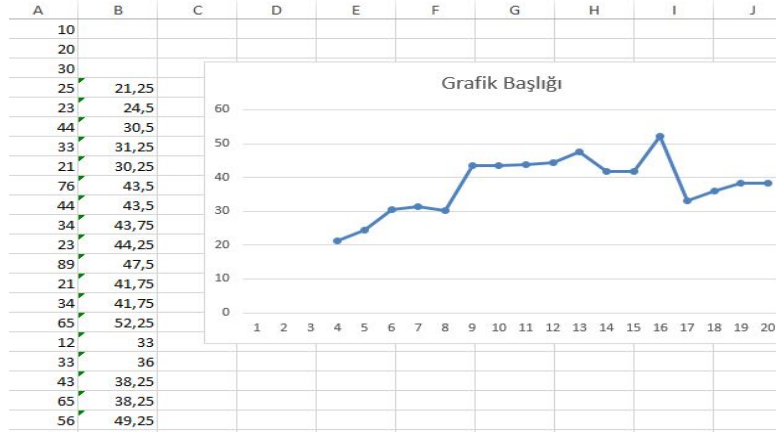
Basit Hareketli Algoritma(Simple Moving Average, SMA) yöntemi ile verilerin belirtilen bir verilen bir n değerine göre ortalamaları alınır (Wei, 1994). Örneğin n değeri 4 olarak verilirse ilk verinin hareketli ortalaması kendisi ve sonraki 3 verinin ortalaması alınarak bulunur. İkinci verinin hareketli ortalaması kendisi ve sonraki 3 verinin ortalaması alınarak bulunur. Bu şekilde bütün verilerin hareketli ortalama değerleri bulunur ve bu değerler üzerinden grafik çizdirilir.

Şekil 1'de rastgele sayılardan oluşturulmuş verilerin Excel ile çizilmiş grafiği verilmiştir. Grafik incelendiğinde verilerin dağınık bir şekilde dağıldığı görülmektedir.



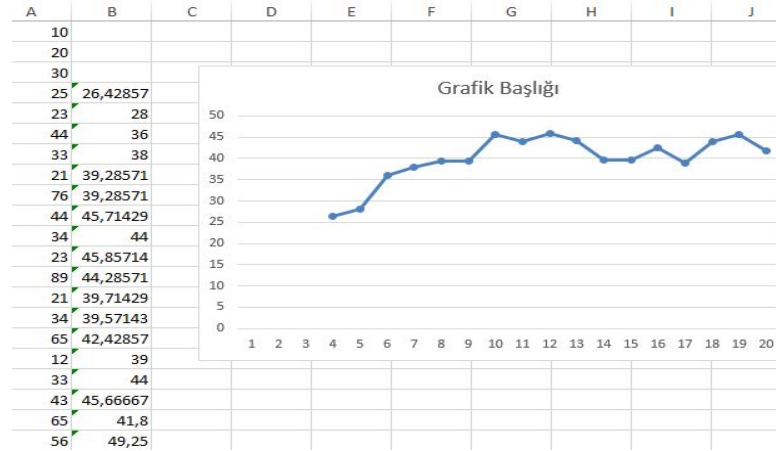
Şekil 1 Rastgele sayıların grafiği

Şekil 1'deki verilerin hareketli ortalamasını n=4 değerine göre almak için Excel'de ORTALAMA fonksiyonunu ilk 4 veri üzerinde kullandıktan sonra aynı işlemi diğer veriler üzerinde de uygulayarak bütün verilerin hareketli ortalaması bulunur. Bulunan hareketli ortalamaların grafiği Excel'de çizdirildiğinde Şekil 2'deki grafik oluşmaktadır. Grafik incelendiğinde verilerin biraz daha düzgün bir dağılım gösterdiği görülmektedir.



Şekil 2 Verilerin SMA(4)'e göre hareketli ortalamaları ve grafiği

Verilerin hareketli ortalaması n=7 değerine göre alınıp bu değerlerin grafiği çizdirildiğinde ise Şekil 3'teki grafik oluşmaktadır. Grafik incelendiğinde verilerin daha düzgün bir çizgide toplandığı görülmektedir. Sonuç olarak hareketli ortalama ile(SMA) dağınık, aykırı olan verilerin daha düz bir çizgide grafiği çıkartılmaktadır.



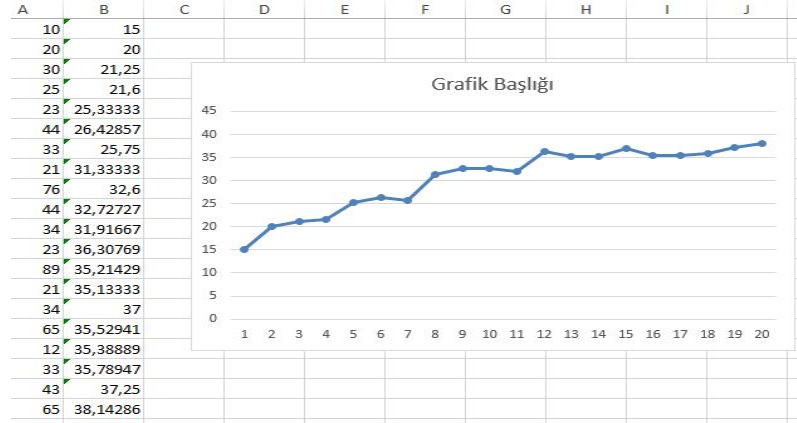
Şekil 3 Verilerin SMA(7) değerine göre hareketli ortalamaları ve grafiği

3.1.2. Kümülatif Hareketli Ortalama(Cumulative Moving Average, CMA)

$$KHO(CMA) = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Formül 2 Kümülatif Hareketli Ortalama

Hareketli ortalama algoritmalarından biri de Kümülatif Hareketli Ortalama(Cumulative Moving Average, CMA)'dır. Bu algoritmada SMA algoritmasındaki gibi bir n değeri verilmez. Her bir verinin hareketli ortalaması bulunurken verinin kendisi ve kendisinden önceki bütün verilerin ortalaması alınır. Şekil 4'te verilerin kümülatif hareketli ortalaması bulunmuş ve grafiği çizdirilmiştir.



Şekil 4 Verilerin Kümülatif ortalaması ve grafiği

3.1.3. Ağırlıklı Hareketli Ortalama(Weighted Moving Average, WMA)

$$AHO(WMA) = \frac{nP_m + (n-1)P_{m-1} + \dots + 2P_{m-n+2} + P_{m-n+1}}{n + (n-1) + \dots + 2 + 1}$$

Formül 3 Ağırlıklı Hareketli Ortalama

Ağırlıklı hareketli ortalamaya göre verilerin bir sonraki veriye olan ağırlıkları farklıdır (Lucas, 1990). Örneğin ilk 4 günün verilerinin ağırlıklı ortalamasına bakıldığında 3. Günün 4. Güne etkisi 1. Günden daha fazladır. Diğer hareketli ortalamalarda kullanılan rastgele sayılar üzerinde n=4 alınıp Excel'de formül uygulanırsa $(4 \cdot A_4 + 3 \cdot A_3 + 2 \cdot A_2 + 1 \cdot A_1) / (4 + 3 + 2 + 1)$ Şekil 8'deki grafik oluşmaktadır.



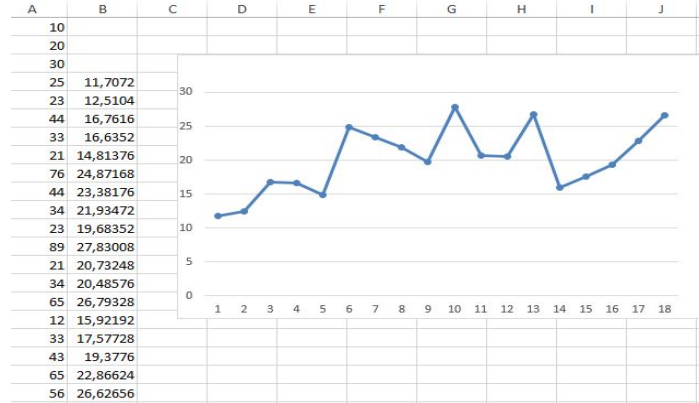
Şekil 5 Verilerin WMA formülüne göre hareketli ortalaması ve grafiği

3.1.4. Üstel Hareketli Ortalama(Exponential Moving Average)

$$EMA_{Bugün(today)} = \alpha * (p_1 + (1 - \alpha)p_2 + (1 - \alpha)^2 p_3 + (1 - \alpha)^3 p_4 + \dots$$

Formül 4 Üstel Hareketli Ortalama

Üstel hareketli ortalama (Lawrance, 1977), son günün sisteme etkisi olan α değeri ile hesaplanır ve bu α değeri 0 ile 1 arasındadır. Geçmiş günlere verilen ağırlık giderek azalır ve bu azalma $1 - \alpha$ 'nın kuvvetleri kadardır. Daha önce örnek olarak kullanılan veriler üzerinde $n=4$ ve $\alpha = 0,2$ alınarak verilerin üstel hareketli ortalamaları hesaplanmış ve Şekil 6'daki grafik elde edilmiştir.

**Şekil 6 Verilerin EMA formülüne göre hareketli ortalaması ve grafiği**

3.1.5. Basit Hareketli Orta Değer(Simple Moving Median, SMM)

Basit hareketli orta değer ile verilerin orta değeri alınır. Örneğin Şekil 6'daki verilerin ilk 4 değerine bakıldığında(10, 20, 30, 25), bu değerlerin orta değeri 25'tir. İkinci veri ile beraber 4 verinin(20, 30, 25, 23) orta değeri 25'tir. Bu şekilde devam ederek sürekli olarak orta değerler bulunur ve bu değerlerin grafiği çizdirilerek veriler daha düzgün bir hareket etrafında toplanmaktadır.

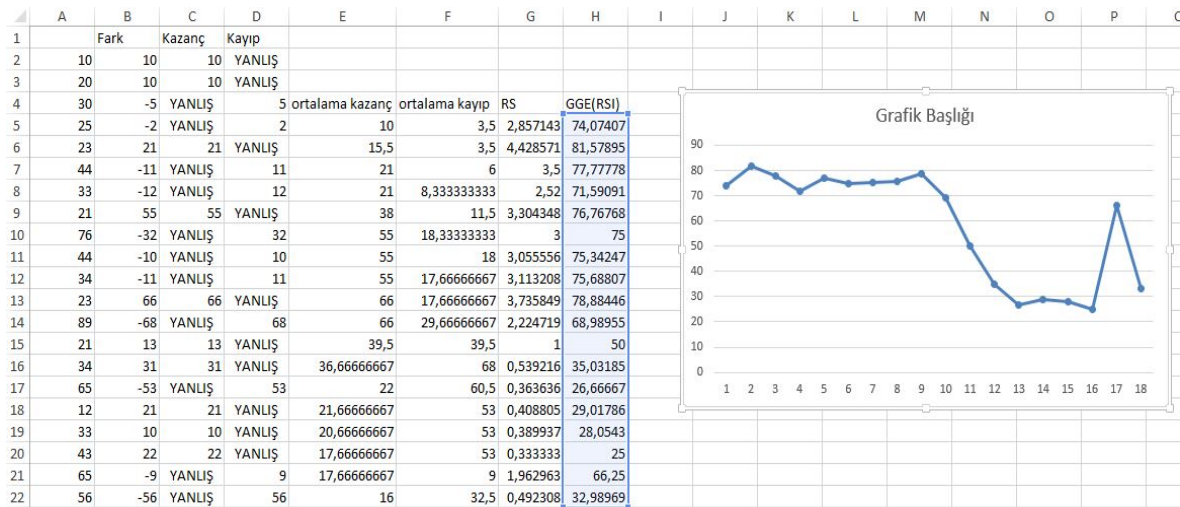
3.2. Göreceli Güç Endeksi(Relative Strength Index,RSI)

$$GGE(RSI) = 100 - \frac{100}{1 + RS}$$

$$RS = \frac{EMA_{U,n}}{EMA_{D,n}}$$

Formül 5 Göreceli Güç Endeksi

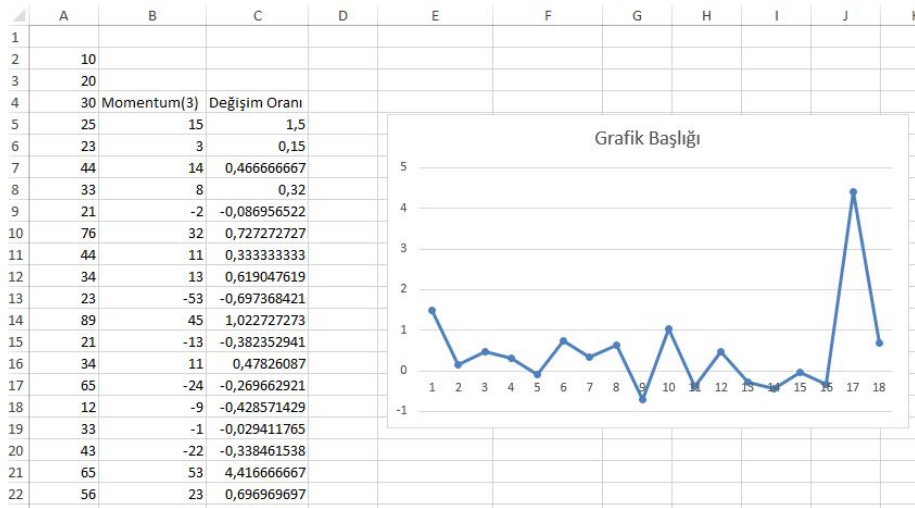
Göreceli Güç Endeksi ile verilerin güç endeksleri(RS) bulunur ve bulunan bu değer formülde yerine yazılır. Literatürde güç endeksleri Formül 5'teki gibi bulunmaktadır (Kim, 2003). Ancak kullanımı daha kolay olan RS=ortalama kazanç/ortalama kayıp formülü de kullanılmaktadır. Ortalama kazanç ve ortalama kayıp değerleri bir sonraki verinin değerinden bir önceki verinin değeri çıkartılarak bulunmaktadır. Çıkarma işleminin sonucu pozitif bir değer ise; ortalama kazanç, negatif bir değer ise ortalama kayıp değeri olmaktadır. Şekil 7'de daha önce örnek olarak kullanılan rastgele sayılardan oluşmuş veri kümesi üzerinde göreceli güç endeksi uygulanmış ve grafiği gösterilmiştir. Excel'de kazanç ve kayıp değerlerini bulmak için C2 ve D2 sütunlarına sırasıyla =EĞER(B2>0;+B2) ve =EĞER(B2<0;-1*B2) formülleri yazılmıştır ve diğer satırlara da uygulanmıştır. Böylece kazanç sütununda 0'dan büyük olan sayılar alınırken 0'dan küçük olan sayılar yanlış(false) değer olarak alınmıştır. Kayıp Sütununda ise 0'dan küçük olan sayıların değerleri pozitif değerlere dönüştürülerek alınmış ve 0'dan büyük olan sayılar yanlış(false) değer olarak alınmıştır. Daha sonra RS değerini bulmak için bu kazanç ve kayıpların n=4 için ortalamaları bulunup RS=ortalama kazanç/ortalama kayıp formülü uygulanmıştır. Son olarak Göreceli Güç Endeksi formülünde bulunan RS değeri kullanılmıştır.



Şekil 7 Göreceli Güç Endeksi (RSI)

3.3. Momentum ve Değişim Oranı(Rate of Change)

Verilerin momentum değerini bulmak için bir n değeri verilir. Örneğin n=3 alındığında 4. Veri ile 1. Veri birbirinden çıkartılarak momentum değeri bulunur ve aynı işlem diğer verilere de uygulanır. Değişim oranı ise momentumdan hesaplanan bir değerdir. Momentum değerinin n değerine göre verilere bölünmesi ile elde edilir. Örneğin n=3 alınarak bir momentum hesaplandysa 4. Sıradaki verinin değişim oranını(rate of change) bulmak için momentum değeri 1. verinin değerine bölünür. Aynı şekilde 5. Sıradaki verinin değişim oranını bulmak için momentum değeri 2. verinin değerine bölünür. Şekil 8'de örnek veri kümesinin momentum ve değişim oranları bulunmuştur.



Şekil 8 Verilerin Momentum ve Değişim Oranlarının grafiği

Kullanılan bu zaman serisi analizi yöntemleri, veri madenciliğinde veri kümeleri üzerinden özellik çıkarmayı sağlamaktadır. Yani verilerin zamana bağlı hareketlerinin bulunmasını sağlamaktadır. Böylece yapılan tahminler daha tutarlı ve doğru olmaktadır. Veri kümeleri üzerinde zaman serisi analizleri yapılarak o veri kümesindeki veriler arasında bir ilişki kurulmaktadır. Zaman serisi analizi yapılmayan veriler üzerinde veri madenciliği yapıldığında, her bir veri birbirinden bağımsız olarak ele alınmaktadır.

Referanslar

- Barnett, V. a. (1994). *Outliers in statistical data*.
- Bonifati, A. (2001). *Designing data marts for data warehouses*.
- Crosswhite, C. E. (2003). Method for determining optimal time series forecasting parameters.
- Hoffer, J. A., Prescott, M., & McFadden, F. R. (2001). *Modern Database Management*. Prentice Hall.
- Honaker, J. a. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science* , 561-581.
- Jarke, M. (1999). *Architecture and quality in data warehouses: An extended repository approach*.
- Kim, K.-j. (2003). Financial time series forecasting using support vector machines. 307-319.
- Kimball, R. a. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*.
- Lawrance, A. J. (1977). An exponential moving-average sequence and point process (EMA1). *Journal of Applied Probability* , 98-113.
- Lucas, J. M. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics* , 1-12.
- Seker, S. E. (2015). Büyük Veri ve Büyük Veri Yaşam Döngüleri. *YBS Ansiklopedi* , 2 (3), 10-17.
- Seker, S. E. (2015). Sosyal Ağlarda Akan Veri Madenciliği. *YBS Ansiklopedi* , 2 (2), 30-39.
- Seker, S. E., Cankir, B., & Arslan, M. L. (2014). Information and Communication Technology Reputation for XU030 Quote Companies. *International Journal of Innovation, Management and Technology* , 5 (3), 221-225.
- Seker, S. E., Cankir, B., & Okur, M. E. (2014). Strategic Competition of Internet Interfaces for XU30 Quoted Companies. *International Journal of Computer and Communication Engineering* , 3 (6), 464-468.
- Seker, S. E., Mert, C., Al-Naami, K., Ozalp, N., & Ayan, U. (2013). Correlation between the Economy News and Stock Market in Turkey. *International Journal of Business Intelligence and Review (IJBIR)* , 4 (4), 1-21.
- Seker, S. E., Mert, C., Al-Naami, K., OZALP, N., & AYAN, U. (2014). Time Series Analysis on Stock Market for Text Mining Correlation of Economy News. *International Journal of Social Sciences and Humanity Studies* , 6 (1),

69-91.

Turban, E. (1990). *Decision support and expert systems: management support systems*.

Wei, W. W.-S. (1994). *Time series analysis*.