



İçerik

1. Değer Tutum ve Yaşam Şekilleri (Value Attitude and Lifestyles)
2. Veri Ambarları (Data Warehouse)
3. Doğal Dil İşleme (Natural Language Processing)
4. Zaman Serisi Analizi (Time Series Analysis)



Yönetim Bilişim
Sistemleri
Ansiklopedisi
Cilt 2, Sayı 4
Aralık 2015

Başlarken

Bu derginin amacı, sureli olarak yönetim bilişim sistemleri (YBS) dünyasındaki akademik literatürü popüler hale getirerek okuyucularla buluşturmadır. Derginin yayın dili Türkçe olup, dergideki yayınlar güncel akademik YBS dergilerinden ve konferanslarından derlenmektedir. Ayrıca temel kavramlara da yer verilmektedir.

Derginin amacı, sürekli yapılan yayınlar ile ansiklopedik bir kaynak oluşturmanın yanında YBS alanında çalışan araştırmacılara akademik bir kaynak sunabilmektir.

Derginin web sitesinde duyurulduğu üzere, akademik ansiklopedi maddesi çağrımız ve mevcut maddelerin genişletilme çağrısı bütün araştırmacılara açıktır.

Şadi Evren ŞEKER

İçindekiler

1. Değer Tutum ve Yaşam Şekilleri (Value Attitude and Lifestyles)
syf. 1 – 5
2. Veri Ambarları (Data Warehouse)
syf. 6 – 13
3. Doğal Dil İşleme (Natural Language Processing)
syf 14 – 22
4. Zaman Serisi Analizi (Time Series Analysis)
syf. 23 – 31

Değer Tutum ve Yaşam Şekilleri (Value Attitude and Lifestyles)

Şadi Evren ŞEKER

^aIstanbul Medeniyet University, Department of Business

Özet

İngilizce Value Attitude and Lifestyle kelimelerinin baş harflerinden oluşan “VALS” kavramı pazarlama dünyasındaki farklı müşteri segmentlerini belirlemek için kullanılan ve müşterileri motivasyonları, satın alma gücü ve yaşam standartlarına göre 8 seviyede inceleyen yaklaşımın ismidir. Temel olarak psikografik pazar segmentasyonu oluşturmaktadır ve müşterilerin, kişilik, değerler, kanaat, tutum ve yaşam şekillerine göre gruplanması esasına dayanmaktadır. Pazarlamada başarı için doğru müşteri segmentine doğru mesajın verilmesi ve doğru ürünün hedeflenmesi prensibine dayanır. Günümüzde gelişen sosyal ağlar ve web 2.0 teknolojileri ile hedeflenmiş pazarlama açısından oldukça önemli hale gelen yaklaşım sayesinde hedefteki müşterinin ürüne ilgi duyması ve doğru ürünün pazarlanması sağlanmaktadır.

Anahtar Kelimeler: Yönetim Bilişim sistemleri, Bilgi Yönetimi, Pazarlama

Summary

VALS is the shortening of Value Attitude and Lifestyle words and is a tool for segmentation of customers in marketing studies. The segmentation is mostly rely on the income level, life standards and motivation of customers. Basically the lifestyles are divided into 8 groups. The technique classifies the customers depending on their values, opinions, attitudes and life styles. The technique is also useful for targeting correct customer for the correct product. By the increasing trend of social networks and web 2.0, the targeted marketing has an increasing importance now a day. The technique can be useful to determine the way of marketing for the correct customer and product also.

Keywords: Management Information Systems, Knowledge Management, Marketing

1. Giriş ve Tanım

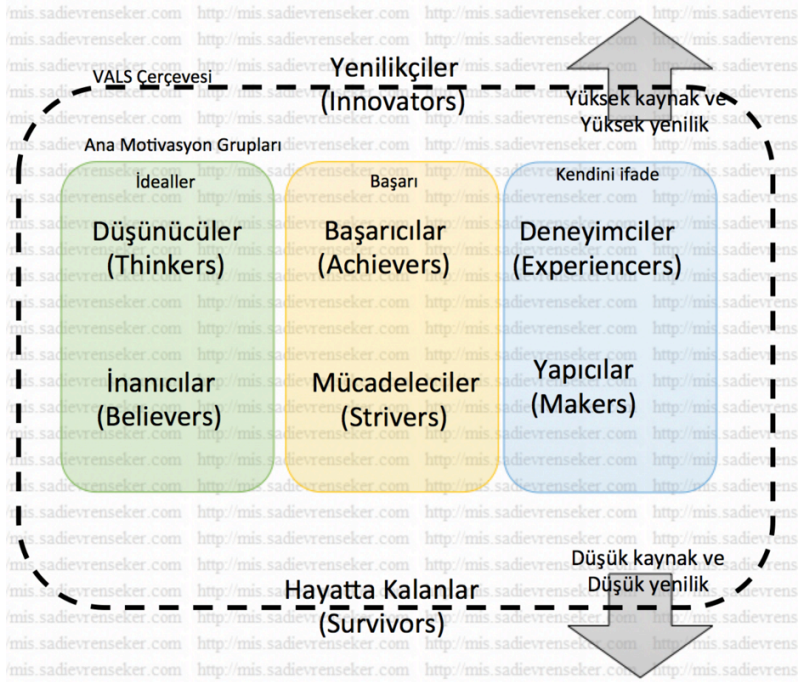
Toplumdaki her bireyin ait olduğu farklı bir segment ve bu segmentlerin özelliklerinin detaylı olarak çıkarıldığı çalışmadır. Temel olarak davranışsal analiz üzerine kuruludur ve her segmentteki müşteri gruplarının farklı davranışlarını araştırmayı hedefler. Davranış şekilleri, kişilik, değerler, kanaat, tutum ve yaşam şekli olarak beş ana unsur üzerine dayanan psikografik çalışmadır. Ayrıca psikografik çalışmaları kültür kavramı ile yakından ilişkili olup her segmentteki kişiliğin farklı bir kültürden olduğu da kolaylıkla iddia edilebilir. Örneğin Amerikan toplumu için yapılan çocuk patlaması nesli (baby boom generation) veya X nesli olarak geçen nesil (Ulrich, 2003) ve 2000 yılını ifade eden Millennials nesli veya Y nesli (William Strauss, 2000) arasında psikografik yöntemler kullanarak farkları analiz etmektedirler.

VALS kavramı ilk olarak 1978 yılında bir sosyal bilimci olan Arnold Mitchell tarafından SRI International firmasında geliştirilmiştir. Ardından kavrama Harvard sosyolistlerinden David Riesman ve psikolog Abraham Maslow tarafından sahip

çıkılmış ve geliştirilmiştir (Yankelovich & Meer, 2006).

Mitchell tarafından geliştirilen fikrin ilk haliyle Amerikan toplumundaki yetişkinlerin alışveriş alışkanlıkları analiz edilmeye çalışılmış ve SRI International firması tarafından 1980 yılında 1635 yetişkinden toplanan verilere göre ilk segmentasyon yapılmıştır (Beatty, Homer, & Kahle, 1988). Bu ilk çalışmadaki müşteri sınıfları ve dağılım oranları şu şekildedir: survivors (4%), sustainers (7%), belongs (35%), emulators (9%), achievers (22%), I-am-me (5%), experiential (7%), societally conscious (9%), and integrated (2%).

VALS yapısı itibarıyla iki boyutlu bir analiz olarak düşünülebilir. Buna göre düşey boyutta kaynaklar ve kaynaklara erişim ve kontrol imkanları bulunurken yatay boyutta ana motivasyon değerleri bulunmaktadır.



Şekil 1 VALS Çerçeve çizimi

Yukarıdan aşağıya doğru, yüksek kaynak erişimi ve kontrolünden daha düşüğe doğru giden bir yapı vardır. Buradaki kaynakla kast edilen bireyin yenilik (innovation) katılımını da etkileyen faktörlerdir (Seker, 2014). Örneğin gelir düzeyi, eğitim seviyesi, kendine güven, bilgi birikimi, liderlik vasıfları veya enerjisi gibi özellikleri bu grupta düşünülebilir.

Klasik motivasyon teorisine (SEKER, 2015) benzer şekilde dış motivasyon faktörlerinde iç motivasyon faktörlerine doğru (sağdan sola) toplumsal grupları yatay eksenle sınıflandırmıştır. Buna göre üç temel sınıflandırma aşağıdaki şekilde sıralanabilir:

1. **İdealler:** Genellikle idealler tarafından şekillenen bilgi (knowledge) ve prensiplerin algılanması ile ölçülebilir. Bu grupta düşüncü ve inanıcılar bulunmaktadır. Kişilik olarak idealist kişilikler bu grupta gösterilebilir. Genel olarak kaynaklara daha fazla erişen grupların fikir üretimi ve düşünce düzleminde tartışma imkanı da daha yüksektir. Daha düşük kaynaklara erişen gruplarda ise düşünce yerini inanca bırakmıştır ve ideallere ulaşmak için açıklamaya ihtiyaç duymaksızın inanma eğilimi görülmektedir.
2. **Başarı:** Genel olarak çevresine başarıyı gösteren veya çevresi tarafından başarıları takdir edilen kişilik grubudur ve başarıları ile mutlu olurlar. Bu grupta başarılılar ve mücadeleçiler vardır. Genel olarak başarıyı etkileyen en önemli faktörlerden birisi kaynaklara erişim olduğu için, kaynaklara erişim imkanı daha yüksek olan grupların başarıya da daha yakın olduğu söylenebilir.
3. **Kendini ifade:** Bu gruptaki kişilikler genelde sosyal veya fiziksel aktivitelerini, risk algılarını, kendi kişisel ifadeleri üzerine inşa ederler. Toplumdan kopuk, topluma muhalif veya eleştirel bakan ve kendi kişiliği üzerine seçimlerini ve kararlarını inşa eden kişilik gruplarıdır. Daha fazla kaynak erişimi olan grupların kendi kişilikleri üzerinden inşa ettikleri yeni deneyimlere açık olduğunu, bununla birlikte kaynaklara kısıtlı erişimleri olanların (yapıcıların) sadece hayatlarını idame ettirecek kanaatkarlıkta kendi kendine yeter bir hayat yaşadıkları görülmektedir.

2. VALS Segmentleri

Şekil 1’de gösterilen ve iki boyutlu uzayda motivasyon ve kaynaklara erişim boyutlarına göre yerleştirilen 8 farklı segment, aşağıdaki şekilde açıklanabilir :

1. Yenilikçiler (Innovators): Değişimin kenarında yaşayan ve değiştirme gücü olan kesimi ifade eder. En yüksek gelir düzeyine sahip kişilerdir. Kendi yönlerini kendileri tayin ederler ve genelde yöneldikleri alanlarda başarılı olurlar. Bağımsız, güçlü karakterli, ve ince zevkleri olan kişilerdir. Ayrıca genel olarak aşağıdaki özellikleri taşıdıkları söylenebilir.
 - a. Bilgiye sürekli açık olmaları ve yeni şeyler öğrenmeleri
 - b. Deneyimleyecek kadar özgüvenleri olması
 - c. En yüksek finansal hareketlilikleri yönetmeleri
 - d. Reklamlar konusunda şüpheci olmaları
 - e. Uluslararası yaşamları olması
 - f. Gelecekle ilgili yönelimleri olması
 - g. Kendi kararlarını almaları
 - h. Yeni fikir ve teknolojilere açık olmaları
 - i. Problem çözme konusunda meydan okumalardan zevk almaları
 - j. En geniş aktivite ve ilgi yelpazesine sahip grup olmaları
2. Düşünücüler (Thinkers): Bu gruptakiler yüksek gelir düzeyine sahip ve idealleri olan kişilerdir. Olgun, sorumluluk sahibi ve iyi eğitim görmüş profesyonellerdir. Genelde kişisel ve aile yaşantıları ile zevklerini birleştirmiş kişiler olmakla birlikte, dünyada nelerin döndüğü hakkında da bilgileri vardır ve yeni fikirler ve sosyal değişimlere açıktırlar. Yüksek gelir düzeyinde olmalarına karşılık uygulamada birer tüketici davranışı sergilemenin yanı sıra genelde kararlarını bir mantıklı şekilde alırlar.
 - a. Sosyal davranışları için “hep” ve “hiç” yargıları vardır
 - b. Analiz felçlerine (analysis paralysis) eğilimleri vardır. Kısaca aşırı analiz veya bir anti-örüntü ile felç olma durumudur.
 - c. Harekete geçmeden önce planlama ve araştırma eğilimleri vardır
 - d. Olayların hikayesini ve geçmişini incelemeyi severler
 - e. Finansal olarak gelişmişlerdir
 - f. Modadan kolay kolay etkilenmezler
 - g. Teknolojiyi fonksiyonel olarak kullanırlar
 - h. Geleneksel olarak entelektüel yolu tercih ederler
 - i. Kendini kanıtlamış ürünleri tercih ederler
3. İnanıcılar (Believers): İdealleri olan düşük gelir grubunu ifade eder. Bu gruptakiler muhafazakar ve tahmin edilebilir müşterilerdir. Genelde yerel ürünleri ve bilinen markaları tercih ederler. Hayatları aile, toplum ve ulus ekseninde şekillenir. Hatırı sayılır gelirleri bulunmaktadır:
 - a. İyi bir hayat için basit doğrular ve yanlışlara inanırlar
 - b. İlham kaynağı olarak ruhsal ve dinsel öğelere dayanırlar
 - c. Arkadaşlık ilişkileri önemlidir
 - d. Kaçış için televizyon seyretmek ve romantik kitaplar okumayı tercih ederler
 - e. Muğlaklığa toleransları yoktur ve her kavramın kesin ve net olarak bir zemine oturmasını isterler
 - f. Toplumun değişmesini istemezler
 - g. Reklamları meşru birer bilgi kaynağı olarak görürler
 - h. Değerlerin sabit ve sürekli kalmasını isterler (sadıktırlar)
 - i. Davranışsal olarak “ben de” eğilimi yüksek oranda gözlenir
4. Başarıcılar (Achievers): Başarıyla motive olan grubun yüksek gelir seviyesinde yer alan bireyleridir. İş hedefli kişilerdir ve iş ve aileleri ile mutlu olurlar. Politik olarak muhafazakar ve değişime kapalı, otoriteye saygılı kişilerdir. Bilinirliği yüksek (meşhur) markaları kullanarak çevrelerine kendilerini başarılı olarak gösterirler:
 - a. Tutum olarak “önce ben ve ailem” eğilimleri vardır
 - b. Paranın gücün kaynağı olduğuna inanırlar
 - c. Ailelerine ve işlerine adanmışlardır
 - d. Zamanlamaya önem verirler ve genelde yoğunlurlar
 - e. Hedef odaklıdırlar
 - f. Çok çalışırlar
 - g. Politik olarak merkezde yer alırlar
 - h. Status quo’nun birer abidesi gibi yaşarlar ve değişime karşıdırlar

- i. Kendi hakkında farkındalıkları vardır ve kendilerini tanımaya çalışırlar
- j. Özeldirler ve özel hissetmek isterler
- k. Profesyoneldirler
- l. Teknolojiyi çıktılarına göre yargırlar. Örneğin üretimi ne kadar arttırdığına göre değerlendirirler
- 5. Mücadeleciler (Strivers): Başarı hedefleri olan kişiliklerin düşük gelir düzeyinde yer alan bireylerdir. Başarıncılar ile aynı karakterlerde olmakla birlikte kaynaklara erişim açısından daha sıkıntılı gruptur ve daha kısıtlı ekonomik, sosyal veya psikolojik kaynağa erişebilirler. Bu bireyler için stil ve moda oldukça önemlidir çünkü bu sayede hayranı oldukları kişilere benzeyebilmektedirler.
 - a. Genelde işte kalma süreleri kısa ve sık sık işsiz kalma oranlarına sahiptirler
 - b. Filim izlemek, bilgisayar oyunları gibi kendilerini hayallerindeki konumda gösteren araçlara eğilimleri vardır.
 - c. Eğlenceyi severler
 - d. Taklit ürün kullanma ve kişilik taklidi yapma eğilimleri vardır
 - e. Genelde toplu taşıma kullanıcıları
 - f. Genelde düşük seviye ve sokak kültürüne sahiptirler
 - g. Hayatlarını değiştirmeyi tutku ile isterler ama genelde bu tutkularını gerçekleştirecek imkanları yoktur
 - h. Bütün gelirlerini görünüşlerine harcaabilirler
- 6. Deneyimciler (Experiencers): Kendi kişiliklerini merkeze alan grubun yüksek gelir seviyesini ifade eder. Diğer bütün gruplara göre en genç yaş grubunu oluştururlar ve ortalama 25 yaşındadırlar. Fiziksel ve sosyal zevkler ve deneyimler için çok fazla enerjileri vardır. Hırslı müşteri grubudur ve giyim, müzik, fastfood ve diğer bütün gençlik harcamalarına eğilimleri vardır. Yeni ürün ve servisleri deneyimleme heyecanı ile yaşarlar.
 - a. Herşeyi isterler
 - b. İlk giren ilk çıkar eğilimini benimserler ve bir ürün ilk çıktığında sahip olmak isterler
 - c. Ana akımın tersine gitmeye çalışan muhalif yapıları vardır
 - d. Son modayı takip ederler
 - e. Fiziksel aktivite peşinden koşarlar
 - f. Kendilerini sosyal olarak görürler
 - g. Arkadaşlığın önemine inanırlar.
 - h. Ani yaşar ani kararlar alırlar
 - i. Görsel olarak yüksek zevklere sahiptirler ve kolay etkilenirler
- 7. Yapıcılar (Makers): Bu gruptaki bireyler kişilik merkezli grubun düşük gelir seviyesinde yer alırlar. Kişinin kendine yeterli olmasını hedefleyen pratik kişilerdir. Genelde içlerine kapanık ve aile ve yakın tanıdıklarla sınırlı sosyal hayatları vardır ve dış dünyayı çok önemsemezler. Müşteri olarak genelde pratik ve kullanışlı ürünleri tercih ederler.
 - a. Devlet başta olmak üzere çoğu kuruma güvenmezler
 - b. Otomatik işleyen hemen her şeye ilgileri vardır (bir önceki maddedeki güvensizliğin de etkisiyle)
 - c. Dış mekan ilgileri yüksektir (avcılık, balık avlamak gibi)
 - d. Cinsiyetler arası algıları nettir ve kesin şekilde cinsiyetleri ve rollerini ayırt ederler
 - e. Kendilerinin olduğuna inandıkları şeyleri korumaya çalışırlar
 - f. Kendilerine göre oldukça düz mantığa sahiptirler ancak başkalarına cahil olarak görülebilirler
 - g. Kendi sahaları olmasını isterler (hem fiziksel bir daire, bina, arsa anlamında hem de sosyal alan anlamında)
- 8. Hayatta Kalanlar (Survivors): En düşük gelir seviyesindeki bireylerdir. Herhangi bir motivasyon belirtisi gösteremeyecek kadar düşük kaynakları vardır. Genelde bütün segmentlerin en yaşlı grubunu oluşturur ve ortalama 61 yaşındadırlar. Genelde değişime kapalı olduklarından markalara karşı en yüksek sadakatin olduğu gruptur.
 - a. Dikkatli ve riski sevmeyen gruptur
 - b. En yaşlı gruptur
 - c. Tutumludurlar
 - d. Geleneksel veya trendi takip eden görülüp görülmemeyi umursamazlar
 - e. Tekrarlardan zevk alır ve rahat hissederek, aynı kişiler ve aynı mekanları tercih ederler
 - f. En fazla televizyon izleyen gruptur
 - g. Genelde çoğu zamanlarını yalnız geçirirler
 - h. İnterneti en az kullanan gruptur

Kaynaklar

- Beatty, S. E., Homer, P. M., & Kahle, L. R. (1988). "Problems With Vals in International Marketing Research: an Example From an Application of the Empirical Mirror Technique".
- Seker, S. E. (2014). "Ağ Etkisi, Moore Yasası ve Dijitalleşmenin Yenilik ile İlişkisi (Relation between Innovation and Network Effect, Moore Law and Digitalization)". *YBS Ansiklopedi* , 1 (1), 8-12.
- SEKER, S. E. (2015). Motivasyon Teorisi (Motivation Theory). *YBS Ansiklopedi* , 2 (1), 22 - 26.
- Ulrich, J. (2003). "Introduction: A (Sub)cultural Genealogy". In *Andrea L. Harris. GenXegesis: essays on alternative youth*. p. 3. ISBN 9780879728625.
- William Strauss, N. H. (2000). *Millennials Rising: The Next Great Generation*. Cartoons by R.J. Matson. New York, NY: *Vintage Original*. p. 370. ISBN 0-375-70719-0.
- Yankelovich, D., & Meer, D. (2006). "Rediscovering Market Segmentation".

Veri Ambarı (Data Warehouse)

Sadi Evren SEKER

^aIstanbul Medeniyet University, Department of Business

Özet

Bu çalışmada genel olarak veri ambarı kavramına giriş yapılmıştır. Veri tabanı ile veri ambarı arasındaki ilişkiler, veri ambarı ile veri marketleri arasındaki ilişkilerden bahsedilmiştir. ETL süreçleri, veri ambarı mimarileri ve veri küplerinden bahsedilerek veri ambarı ile ilgili temel konular anlatılmıştır.

Anahtar Kelimeler: ETL süreçleri, Veri Ambarı Mimarileri, Yıldız şeması, Veri Küpleri

Summary

This paper makes a brief introduction to the data warehouse concept. Conceptual view of relationships among datawarehouse, database and data marts has been introduced as well as ETL processes, data warehouse architectures and datacubes.

Keywords: Misspellings, ETL Process, Data Warehouse, Data Architecture, Star Scheme, Data Cubes

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

1. Giriş ve Tanım

Veri ambarları(data warehouse) (Hoffer, Prescott, & McFadden, 2001), veri tabanlarının birer parçası olarak düşünülmektedir. Veri tabanlarını yormamak için oluşturulmuş, daha hızlı çalışan, özelleştirilmiş ve veri tabanlarına göre daha az veri saklayan yapıdadırlar. İlgili veriyi kolay, hızlı ve doğru biçimde analiz etmek için gerekli işlemleri yerine getirir.

1.1. Veri Ambarının Temel Özellikleri ve Veri Tabanı Arasındaki Farklar

- Veri tabanlarının amacı bütün verileri tutmaktır, veri ambarı ise işlenmiş, özelleştirilmiş verileri tutar.
- Veri tabanları ile veri ambarları arasındaki en önemli farklardan birisi farklı kaynaklardan besleniyor olmalarıdır. Örneğin bir alışveriş merkezinde yazar kasalarından gelen veriler, müşteri yönetimi ile ilgili olan veriler, stok ile ilgili veriler veri tabanlarında tablolar halinde farklı veri kaynakları olarak saklanır. Saklanan bu farklı veri kaynaklarının birleştirilerek, bir

amaca yönelik olarak raporlarının hazırlanması ve verinin daha verimli kullanılabilmesi için geliştirilen sistemlere veri ambarı denilmektedir.

- Veri ambarları konu odaklıdır; müşteri, ürün veya satış odaklı özelleştirilebilirler.
- Veri ambarları trend ve değişim değerlerini takip ederek zaman içinde kendini güncellemektedir. Veri tabanlarındaki güncellemelerden haberdar olabilmek için çeşitli algoritmalar kullanılmaktadır.
- Yapı olarak veri ambarları doğrudan güncellenemez, veri tabanlarından gerekli güncelleme işlemleri yapılır. Dolayısıyla veri tabanlarından veri ambarlarına tek yönlü bir veri akışı bulunmaktadır.
- Veri ambarları yönetim kademesinde veya karar verme süreçlerinde kullanılan sistemlerdir.

Veri ambarlarının daha küçültülmüş ve hedefe yönlendirilmiş alt kümelerine Veri Marketleri(Data Mart) (Bonifati, 2001) denilmektedir. Veri marketleri, veri ambarlarının alan olarak daha daraltılmış halidir; veri ambarları gibi bir problemin bütününe değil, belli bir kısma yönelik bir bakış sağlar.

Veri ambarları şirketlerin ihtiyaç duyduğu raporların alınmasını sağlar. Operasyonel ve enformasyonel olmak üzere ikiye ayrılır.

1.2. Operasyonel ve Enformasyonel Veri Ambarları Arasındaki Farklar

Özellik	Operasyonel	Enformasyonel
Amacı	İşletmenin süreçlerinin Devamı	Yönetime karar vermede destek
Veri Tipi	İş Süreçlerindeki Durumlar	Geçmişe ait rapor ve Gelecek tahminleri
Ana kullanıcılar	Operasyonel kişiler, satış, idari personel,	Yöneticiler, iş analistleri, müşteriler
Kullanım Genişliği	Dardır, Planlıdır, düşük güncelleme gerektirir	Geniştir, ad-hoc, komplekstir, hesaplama ve analiz gerektirir
Tasarım Amacı	Performans, Veriye erişim imkanı	Basit ve esnek erişim imkanı
Hacim	Çok sayıda ama kısıtlı veri kaynaklarına erişim	Hemen her veri kaynağına, seyrek erişim

Şekil 1 Operasyonel ve Enformasyonel Veri Ambarları Arasındaki Farklar

- Enformasyonel veri ambarlarında iş problemlerinin rapor ve bilgilerinin alınması hedeflenirken operasyonel veri ambarlarında bir işin veya sürecin hedef alınması söz konusudur. Örneğin satışı yapılan bir ürün için firma müşteri tarafından sorun bildirmek üzere arandığında, o ürünün hangi kasiyer tarafından satıldığı, hangi kargo şirketi ile gönderildiği, stokta nerede beklediği gibi detayların alınması operasyonel bir işlemdir ve hata bildirme sürecinin bir parçasıdır. Bir yöneticinin herhangi bir iş problemi ile ilgili bilgiler ve raporların hazırlanması enformasyonel veri ambarına örnek olarak verilebilir.
- Veri tipine bakıldığında operasyonel veri ambarlarında süreç ve durumlar ile ilgili bilgi ve raporlar alınırken, enformasyonel veri ambarlarında geçmişe ait raporlar ve geleceğe dair tahminler ele alınır.
- Operasyonel veri ambarlarının kullanıcıları operasyonel kişiler, idari personel gibi daha alt ve ya orta seviyede konumlanan yöneticiler ve bu bilgileri doğrudan işlerinde kullanacak olan kişilerdir. Enformasyonel veri ambarının kullanıcıları ise üst seviyedeki yöneticiler veya iş analistleri ve müşterilerdir.
- Operasyonel veri ambarlarının kullanım genişliği daha dardır, planlıdır ve düşük güncelleme gerektirir. Boyut olarak daha küçük olduğu için güncelleme yapılırken çekilen veri azdır, dolayısıyla güncelleme maliyeti de azdır. Enformasyonel veri

ambarının kullanım genişliği daha fazladır. Anlık olarak değişen(ad-hoc) ve kompleks yapıdadır. Raporların karmaşıklığından dolayı raporlama ve analiz gerektirir.

- Operasyonel veri ambarlarının tasarım amacı daha çok performans ve veriye eşim imkânı sağlarken, enformasyonel veri ambarlarında daha esnek ve basit erişim imkânları bulunmaktadır.
- Operasyonel veri ambarlarında bir konuya ait bütün verilerin işlenmesi söz konusudur ve işlenen veriler belirli bir hedefe ve amaca yöneliktir. Enformasyonel veri ambarlarında daha geniş açılımda veri gurubuna ihtiyaç varken, daha seyrek erişim söz konusudur. Örneğin yılda bir defa alınan plan raporu için bütün verilere erişim yapılır, ancak erişim sayısı azdır.

1.3. Veri Ambarı ve Veri Marketi Arasındaki Farklar

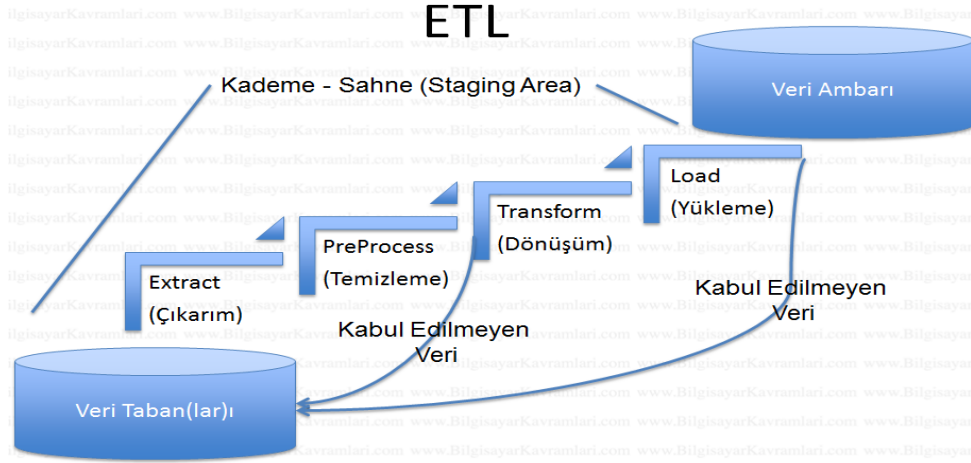
Özellik	Veri Ambarı	Veri Marketi
Ölçek	Uygulama Bağımsız, Merkezi ve Bütün işletmeyi ilgilendiren, Planlı	Özel DSS uygulamaları, Kullanıcıya özel, organik, planlı olmayan
Veri	Tarihsel, detaylı ve özet, hafifçe denormalize edilmiş	Kısmi olarak geçmişe ait, detaylı ve özetlenmiş, yüksek oranda denormalize
Konusu	Çok farklı konulardan	Kullanıcıyı ilgilendiren özel bir konuda
Kaynakları	Çok sayıda iç ve dış veri kaynağından	Kısıtlı iç veya dış kaynaklardan
Diğer	Esnek, Veri Yönelimli, Uzun ömürlü, Büyük ölçekte, Tek ve Karmaşık yapıda	Kısıtlı, Proje yönelimli, Kısa ömürlü, küçük başlayıp büyüyen, çok ve yarı karmaşık veya karmaşık yapıda

Şekil 2 Veri Ambarı ve Veri Marketi Arasındaki Farklar

- Veri ambarları uygulama bağımsız, merkezi ve bütün işletmeyi ilgilendiren, planlı ve büyük ölçeklidir. Veri marketleri ise karar verme süreçlerine(DSS) (Turban, 1990) daha yakın, kullanıcıya özel, organik, planlı olmayan veri ambarları olarak düşünülebilir.
- Veri ambarlarının verileri tarihsel, detaylı ve özet, hafifçe denormalize edilmiş yapıdadır. Veri marketlerinin verileri ise kısmi olarak geçmişe ait, detaylı ve özetlenmiş, yüksek oranda denormalize edilmiştir. Denormalize, normalizasyonun ters işlemidir. Veri tabanındaki tabloların en az yer kaplayacak şekilde ve buna karşılık hızından ödün vererek düzenlenmesine normalizasyon denilmektedir. Denormalizasyon ise hızı öncelik haline getiren ancak buna bağlı olarak veri tabanının daha fazla yer kaplamasına sebep olan sistemdir. Sonuç olarak veri marketlerinin yüksek oranda denormalize edilmesi hızı öncelik aldığını göstermektedir.
- Veri ambarı çok farklı konulardan toplanırken, veri marketinin konusu kullanıcıyı ilgilendiren özel bir konudur.
- Veri ambarları çok sayıda iç ve dış kaynaklardan beslenirken, veri marketi kısıtlı iç veya dış kaynaklardan toplanır.
- Veri ambarları esnek yapıda, veri yönelimli, uzun ömürlü, büyük ölçekte tek ve karmaşık yapıdadır; veri marketi ise kısıtlı, proje yönelimli, kısa ömürlü, küçük başlayıp büyüyen, çok ve yarı karmaşık veya karmaşık yapıda olabilmektedir.

2. ETL Süreçleri

Veri ambarının en önemli süreçlerinden biridir. ETL (Kimball, 2011); Extract(çıkartım), transform(dönüşüm) ve load(yükleme) kelimelerinin kısaltılmışıdır.



Şekil 3 ETL Süreçleri

Şekil 3'te ETL sürecinin adımları gösterilmiştir. Bu süreçler incelendiğinde silindir ile ifade edilen veri tabanlarından alınan veriler extract(çıkartım) kısmında işlenir. Hangi verilerin, ne kadarının, ne sıklıkta alınacağı gibi kararlar extract(çıkartım) kısmında verilir. Verinin ön işleminin yapıldığı kısım PreProcess(temizleme) kısmıdır. Temizleme kısmında veriler incelenir, tutarsız, hatalı girilen ve boş değerlere belli yöntemler uygulanarak veriler işlenebilecek hale getirilir. Transform(dönüşüm) kısmında verinin amaca yönelik olarak dönüştürülme işlemi yapılır. Örneğin eldeki bir veride adres bilgisi içinden sadece şehir bilgisinin alınması veya kişilerin mesleklerinin tutulduğu veriden sadece bilişim sektörüne ait olanların raporlanması istendiğinde, meslekler arasında bilişim ile ilgili olanların çeşitli yöntemlerle seçilmesi dönüşüm işlemidir. Son aşamada ise dönüştürülen işlenmiş veriler veri ambarına yüklenir(load) ve erişime hazır hale getirilir. Verinin erişime açık hale gelmesi o veri üzerindeki işlemlerin bittiğini ve raporlama işlemlerinin yapılabileceğini gösterir. Bazı durumlarda veri ambarındaki işlenmiş veri tekrar ETL süreçlerinden geçebilmektedir. Bunun dışında veri ambarları üzerinde veri madenciliği algoritmaları, karar verme mekanizmaları uygulanabilmektedir. Örneğin stokta biten bir ürünün ilgili yerlere stok bilgisi raporlanabilir. Sonuç olarak veri ambarı verinin hazırlanma süreci olarak düşünülebilir. Bu süreç sırasında kademelerden kabul edilmeyen veriler veri tabanlarından çıkarılabilir. Verilerin veri tabanından veri ambarına kadar olan süreçte işlenmesine Kademe-Sahne(Staging Area) denilmektedir.

3. Veri Uzlaşması(Data Reconciliation)

3.1. Extract(Çıkartım)

Extract aşaması statik veya incremental(artan) olabilmektedir. Statik algoritmalarda bloklar halinde alınan veriler ile raporlar oluşturulur ve bir sonraki ay tekrar rapor oluşturulmak istendiğinde bu bir aylık veri tekrar extract(çıkartım) edilir. Incremental(artan)'da ise sadece aradaki artışlar güncellenir. Aylık raporlar için statik extract(çıkartım), günlük raporlar için ise değişiklikleri ya da günlük satış grafiklerinin görülebilmesi için her gün sadece o güne özel verilerin çekilmesi dolayısıyla incremental(artan) extract kullanılması daha mantıklıdır.

3.2. PreProcessing(temizleme)

Bazı kaynaklarda temizlemek anlamına gelen Scrub/Cleansing olarak da geçmektedir. Yazım hataları, tarih hataları, eksik veriler, imkânsız veriler ve tekrarlanan veriler PreProcessing aşamasında düzeltilir veya ayıklanır.

3.3. Transform(Dönüşüm)

Transform işlemi satır(Record level) veya sütun(Field level) olmak üzere iki boyutta ele alınır. Satır boyutunda; verinin parçalanması, verinin birleştirilmesi ve ortalamanın alınması(avg), toplamın bulunması(sum) gibi işlemler ele alınır. Sütun boyutunda; birden fazla sütunun birleştirilmesi veya tek bir sütunun birden fazla sütuna bölünmesi işlemleri ele alınır.

3.4. Load(Yükleme)

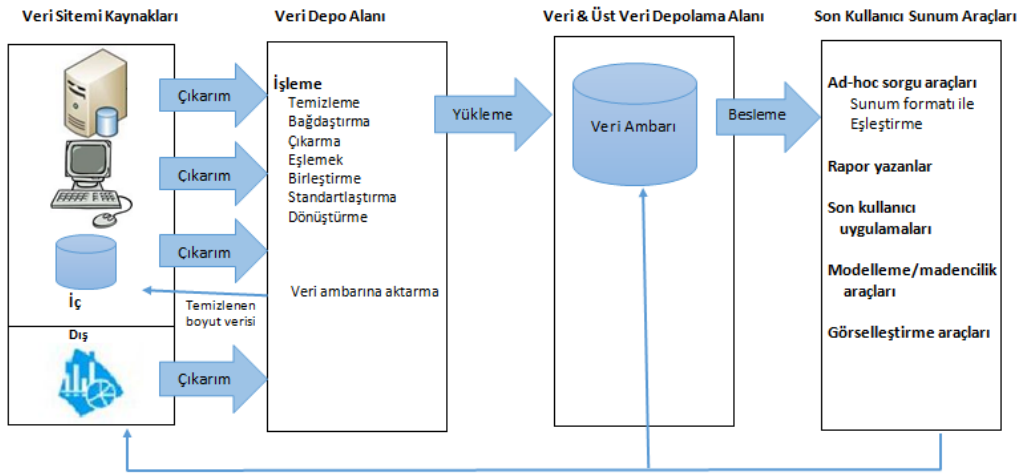
Refresh Mode(yenileme) ve Update Mode(güncelleme) olmak üzere iki durumda incelenmektedir. Refresh Mode; verinin yenilenmesi amacıyla belli aralıklarla tekrar alınmasıdır ve verinin sürekli olarak yenilenmesi sağlanır. Update Mode ise; verinin güncellenmesi amacıyla sadece değişen kısımlarının alınmasıdır.

4. Veri Ambarı Mimarileri

Birden fazla veri ambarı mimarileri (Jarke, 1999) bulunmaktadır. Bunlar; iki seviyeli mimari, Bağımsız Veri Marketi, Bağımlı Veri Marketi ve Operasyonel Veri Kaydı(Operation Data Store, ODS), Mantıksal Veri Marketi ve Aktif Ambar, Üç Katmanlı Mimari'dir.

4.1. İki Seviyeli Mimari

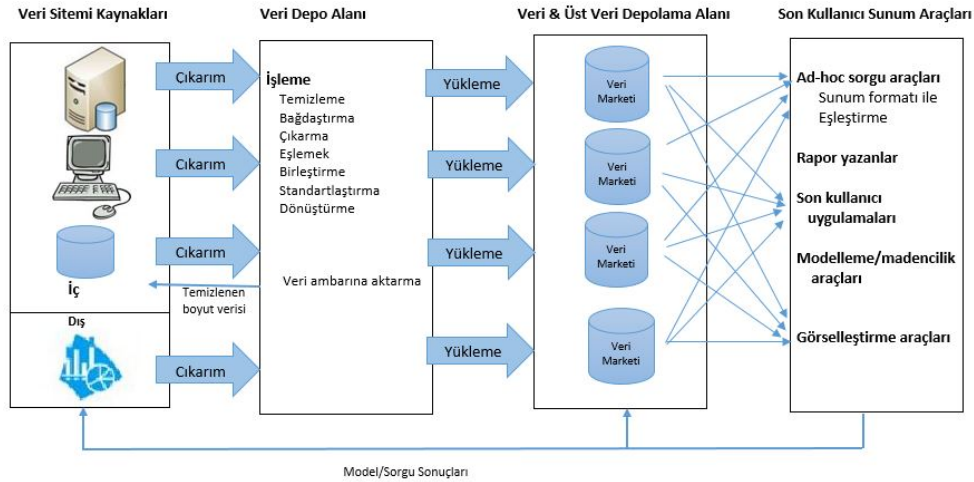
Şekil 4' teki iki seviyeli mimaride ETL süreçleri verilmiştir. Veri kaynağından(Source Data System) alınan veriler extract(çıkartım) işleminden geçirilerek; transform(dönüşüm) kısmında temizleme, ayrıştırma, birleştirme gibi işlemlerden geçerek veri ambarına yüklenir ve son kullanıcı(End-User) aşamasında sorgular, raporlar ve uygulamalar bu veriyi kullanır.



Şekil 4 İki Seviyeli Mimari

4.2. Bağımsız Veri Marketleri

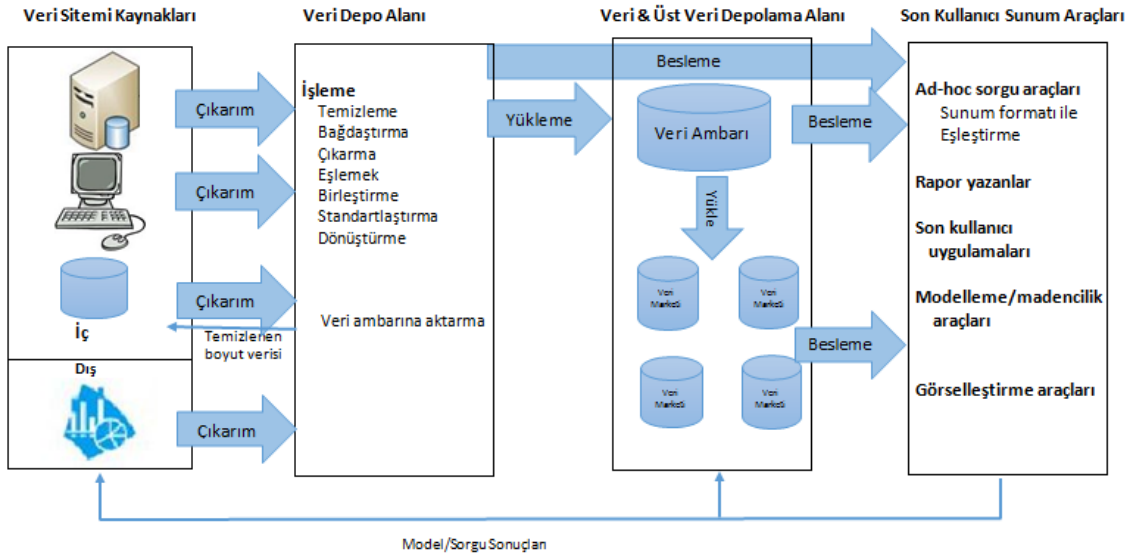
Bağımsız veri marketlerinde de ETL süreçleri Extract(çıkartım), Transform(Dönüşüm) ve Load(yükleme) olarak gerçekleştirilmektedir. Ancak işlenmiş verilerin yüklendiği yer veri ambarı yerine birden fazla veri marketinden oluşan bir sisteme yüklenir ve bu veri marketlerinden sorumlu kişiler veriler üzerinde kendi bölümleri ile ilgili çalışmalar yapar.



Şekil 5 Bağımsız Veri Marketleri

4.3. ODS ve Bağımlı Veri Marketleri

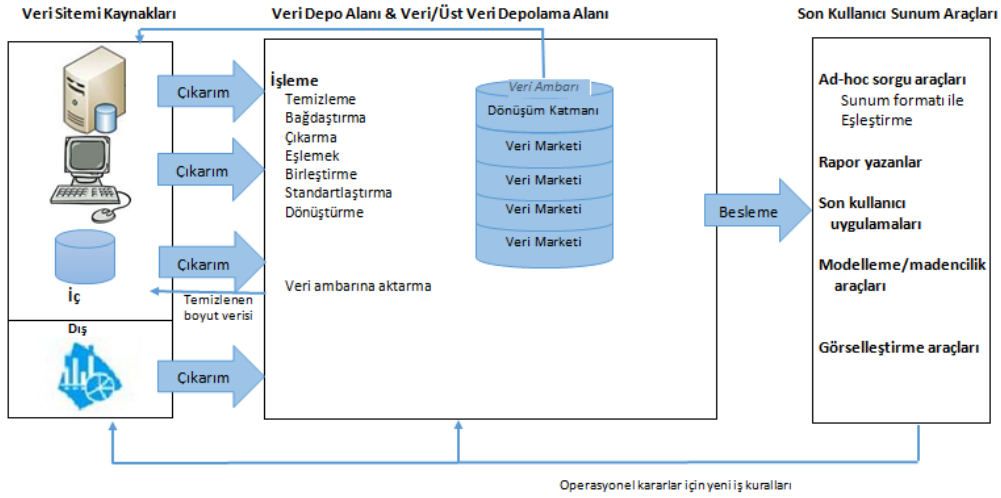
Veriler ETL süreçlerinden geçtikten sonra içinde bir tane veri ambarı ve birden fazla veri marketlerinin bulunduğu bir sisteme yüklenir. Bu sistemde işlenmiş veri ilk olarak veri ambarına ve sonrasında veri ambarından farklı bölümlere ayrılmış veri marketlerine yüklenir. Dolayısıyla ETL süreçlerinin sonunda veri tek bir yerde tutulurken sonrasında işlemlere göre parçalanmaktadır. Bağımsız veri marketlerinde iki farklı kullanıcının kullandığı aynı verilerin farklı veri marketlerinde tutulması söz konusu olabilir. Yani bu iki kullanıcı tek bir veri marketi yerine farklı veri marketlerinden aynı veriyi kullanabilir. Aynı durum bağımlı veri marketleri içinde geçerli olabilir ancak bağımsız veri marketlerinde veriler her bir veri marketi için tekrar tekrar ETL süreçlerinden geçerken, bağımlı veri marketlerinde ETL sürecinden geçen işlenmiş veriler tek bir yerde toplandığı ve sonrasında veri marketlerine parçalandığı için ETL süreçlerinin tekrarlanmasına gerek kalmaz.



Şekil 6 Bağımlı Veri Marketleri

4.4. Aktif Ambar Teknolojisi

Veri ambarı ve veri marketleri bir dönüşüm katmanında tutulur. Her bir son kullanıcının ilgilendiği bölümler ayrı veri marketlerinde tutulan ve aynı veri ambarından beslenen yapılardır.



Şekil 7 Aktif Ambar Teknolojisi

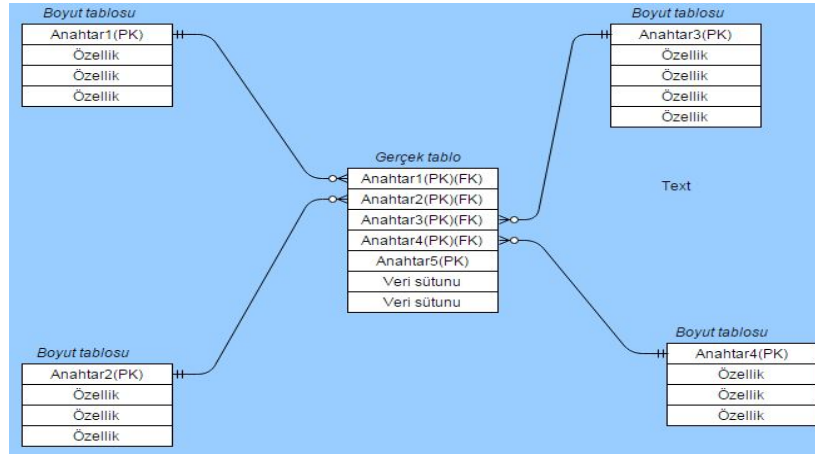
4.5. Gerçek Zamanlı Veri Ambarı(Realtime Data Warehouse)

Adaptive(uyumlu) bir yapıdadır, veri kendini uyumlu bir şekilde ayarlayabilmektedir. Kullanım alanları şu şekildedir:

- Sistemlerin monitör edilmesi
- Kredi kartları ile yapılan harcamaların kullanıcıya bu harcamaların kendisi tarafından yapıp yapılmadığını uyararak anlık uyarılar
- Ağ(network) güvenliğinin sağlanması gereken yerler
- Müşterilerin bir sorun bildirdiğinde anlık olarak geri dönüş yapılması ve bununla ilgili verinin toplanıp raporlanması
- Fiyatların otomatik olarak değişmesi
- Bir ürüne karşı talep artışının olduğu yerlerde bu ürün ile ilgili üretimin artırılması gibi işlerin anlık olarak tetiklenmesi
- Verinin doğru olup olmadığının kontrolünün yapılması gibi alanlarda kullanılabilmektedir.

5. Yıldız Şema(Star Scheme)

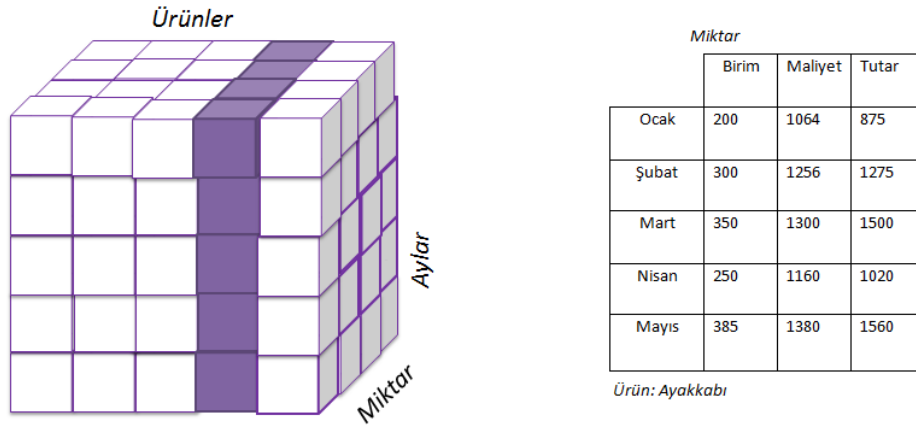
Veri ambarı tablo yapılarının en çok bilinen modelidir. Merkezde tutulan ve veri ambarının işlenmiş, temiz verilerinin bulunduğu tablo, boyut tabloları olarak adlandırılan küçük tablolar ile ilişkilendirilmiştir. Bu ilişkilendirme temel olarak veri tabanlarındaki varlıklar arası ilişki(Entity Relationship) modeline benzemektedir.



Şekil 8 Yıldız Şema Tablosu

6. Veri Küpü(Data Cube)

Verinin iki boyuttan üç boyuta çıkması ve her bir dilimlerin tabloları temsil etmesi durumudur. Dolayısıyla tablolar küp şeklinde bir araya toplanır. Şekil 9'da ürünler(products) arasından ayakkabıların(shoes) aylara göre satışını gösteren dilim tablo halinde ifade edilmiştir.



Şekil 9 Veri Küpü

Referanslar

- Bonifati, A. (2001). *Designing data marts for data warehouses*.
 Hoffer, J. A., Prescott, M., & McFadden, F. R. (2001). *Modern Database Management*. Prentice Hall.
 Jarke, M. (1999). *Architecture and quality in data warehouses: An extended repository approach*.
 Kimball, R. a. (2011). *The data warehouse toolkit: the complete guide to dimensional modeling*.
 Seker, S. E. (2015). Büyük Veri ve Büyük Veri Yaşam Döngüleri. *YBS Ansiklopedi* , 2 (3), 10-17.
 Turban, E. (1990). *Decision support and expert systems: management support systems*.

Doğal Dil İşleme(Natural Language Processing)

Sadi Evren SEKER

^aIstanbul Medeniyet University, Department of Business

Özet

Doğal dil işleme konusuna giriş içeriğinde olan bir yazı olmuştur. Doğal dil işleme ingilizce literatürde *natural language processing* ya da NLP olarak kısaltılmış hali geçmektedir. Doğal dil işleme süreçleri, doğal dil işlemeye neden gereksinim duyulmakta, karşılaşılan sorunlar nelerdir ve son kullanılan yerlere değinilerek giriş anlamında bir anlatım olmuştur.

Anahtar Kelimeler: Doğal dil işleme , Makine Çevrimi ,Kelime işleme ,internet arama motoru , Metin işleme , Finans ,Zamansal Mantıklar, Argüman Ağları

Summary

This article aims to make a brief introduction to natural language processing concept and terms. Some problem types and some application areas are introduced in the article.

Keywords: Natural Language Processing , Machine Translation, Word Processing ,NL/DB interface, AI Bots (Siri), Text Processing, Text Summarization, Argument Aggregation

1. Doğal Dil İşlemeye Genel Bakış

Diller bilindiği üzere ikiye ayrılır; makine dili(machine language) ve insanlar tarafından kullanılan doğal dildir(natural language). Bilgisayarların insanların dillerini anlaması , onlarla iletişime geçmeleri için doğal dil işleme bilimi kullanmaları gerekmektedir. Kısacası, bilgisayarların doğal dilleri işleme sürecidir. Bunu insan-bilgisayar etkileşimi (human computer interaction-HCI) (Preece, 1994.) biliminin altında da görülebilir veya hesaplamalı dil bilimi(computational linguistics-CL) olarak da görülebilir. Burada dilin hesaplanması işlem görmesi durumudur.

Veri bilimiyle doğal dil işleme arasındaki ilişki düşünüldüğünde, aslında bu konuyu metin işleme(Text Processing) ve metin madenciliğinin(text mining) (Seker, Metin Madenciliği(Text Mining), 2015) altında düşünmek gerekir. Çünkü veri biliminin genelde ilgilendiği kısım buralardır. Bunun sebebi , eğer veri kaynakları metin şeklindeyse bunların ilk önce analiz edilmesi ve işlenebilir hale getirilmesinde doğal dil işleme önemli bir rol oynamaktadır.

Doğal dil işleme girdilere göre ikiye ayrılır. Yazılı metinlere(text processing) göre ve ses (speech processing) üzerine yapılanlardır. Doğal dilin genellikle iki kaynağı bulunmaktadır. Birinin konuşması durumunda ona sesli yanıt verilmesi veya yazılı bir metnin analiz edilmesi olarak düşünülebilir. Doğal dil işleme çalışmalarını şu şekilde olmaktadır; ses üzerinden çok fazla çalışma yapılamamakla birlikte daha çok sesli anlatımların yazılı hale getirilip ve daha sonra yazılı hale getirilen metinler işleme dahil edilmeleriyle oluşmaktadır.

Doğal dil işlemenin çalışma seviyelerine bakıldığında dört ana madde görülmektedir. Kelime bilimi(Morphological-Lexical), söz dizimsel (syntactic), anlamsal(semantic) ve söylevdir(pragmatic-discourse). Doğal dil işleme aşamaları bu dört ana madde üzerinden anlatılmaya çalışılacaktır. Kelime bilimi(lexical), kelimelerin anlamlarının anlaşılmasıyla ilgilenir. Bir kelimenin kökünün hangi ekler alarak hangi anlama dönüştüğünü inceler ve bu sayede kelimenin ne olduğu anlaşılır. Söz dizimsel (syntactic), cümledeki kelimelerin dizilim şeklidir. Bir cümlede kelimelerin nasıl dizildiğiyle ilgilenir. Anlamsal(semantic), kullanılan cümlelerin anlamlarını inceler. Çünkü doğal dilin bilgisayar tarafından doğru işlenebilmesi için ilk, bu cümlelerin doğru bir şekilde anlaşılmasıdır. Söylev(pragmatics-discourse), bir konuşma sırasında kullanılan kelimeler ve anlamlarıyla ilgilenir. İlk önce bilgisayar tarafından konuşma anlaşılmalı ve nelerden bahsedildiği anlaşılan cümlelere doğru kelimeler kullanılarak cevap verilmesi durumudur.

Doğal dil işlemenin bir diğer kısmı üretim(Generation) kısmıdır. Burada bir dilin anlaşılması için gerekli zamanı yine bu dili üretmek için de kullanılmalıdır. Örneğin bilgisayarın ürettiği bir türkçe kelime olarak düşünülebilir.

2. Doğal Dil İşlemede Karşılaşılan Zorluklar

2.1. Belirsizlik-Muğlaklık Durumu (Ambiguity)

2.1.1. Sözcük Seviyesi(Lexical)

Diller ailesinde bitişik ,sondan eklemeli(Agglutinative) olarak bir dil ailesi bulunmaktadır. Yani bir kelimenin sondan eklemeli olarak ,birden fazla ek almasıyla bir çok anlama gelebilir. Bu sebepten dolayı bu kelimelerin anlamını bulmak zorlaşmaktadır. Örneğin Türkçe dili de sondan eklemeli bir dildir.

2.1.2. Sözdizimsel(Syntactic)

Cümlelerin içinde kelimelerin yer değiştirilmesiyle oluşturulan cümleler bulunmaktadır. Bu kelimelerin cümlede bulundukları yere göre anlamı değişebilmektedir. Kelimenin vurgusu değişebilir. Bu durumun bilgisayar tarafından anlaşılması gerekmektedir.

2.1.3. Kısmi Bilgi Yorumlanması(Interpretation of Partial Info)

Bir tercüme ,çeviri yapılması durumunda bir cümlede bulunan örneğin bir zamirin hangi kelimeyi karşıladığı bulunması durumudur. İngilizce zamirlerden “he,she,it” durumu gibi.

2.1.4. Bağlam(Context)

Bilgisayara girilen bir cümlelerin hangi alandan veya konudan bahsedildiği bilgisayar tarafında anlaşılmalıdır. Örnek olarak ‘Türkiye de ege bölgesinde dağlar denize diktir. ’cümlesi bilgisayara girildiğinde bunun coğrafi bir bilgi olduğu anlaşılmalı ve içerisindeki kelimelerin coğrafi terimler olabileceği varsayabilmelidir. Birşeyi ifade eden çok şey, çok şeyi ifade eden bir şey vardır cümlesi bu durumu açıklayan çok güzel bir ifadedir. Bir kişi bir cümle ile karşısındaki kişiye bir çok şey ifade edebilir ya da bir kelimeyle her şeyi açıklayabilir. İşte bu durumda bilgisayarın bu cümleyi nasıl anladığı önemli bir sorun olabilir ya da bir kelimeyi ifade eden bir çok kelime olabilir, bilgisayarın en doğru karşılığı bulması da burada yine önemli bir sorundur.

3. Dil Bilgisi Bilimi(Knowledge)

3.1. Ses bilimi(Phonology-Phonemes)

Kelimelerdeki vurgu durumudur. Her kelimenin, her harfin bir söyleniş şekli bulunmaktadır. Sesler ve bunlara bağlı fonetik olma durumları değişebilmektedir. Kısacası her kelime her zaman aynı şekilde ifade edilemez.

Türkçede her kelime yazıldığı gibi okunur fakat her ses her zaman aynı şekilde söylenmez. Örneğin ‘k’ harfi ele alındığında bazen ‘ke’ bazen de ‘ka’ şeklinde seslenilir. Yani yeri geldiğinde vurgular değişmektedir. Bu durum ingilizce de daha kompleks durumdadır.

3.2. Kelime Bilimi(Morphology-Morphemes)

Kelime bilimindeki en küçük parçalar olarak ifade edilebilir. Yapım ekleri ,çekim ekleri bunlara örnek verilebilir. Örneğin -cı,-ci gibi gelen yapım ekleri bir kelime parçalanırken bölünen en küçük parçadır.

3.3. Sözdizimi(Syntax-Phrases,sub-phrases)

Kelime gruplarının belirlenmesi ,tamlamalar sözdizimi için örnek olarak gösterilebilir. Tamlamadaki her kelime kendine ait bir anlamı varken birlikte de bir anlam ifade edebilir. Örnek olarak kırmızı evin kapısı isim tamlamasında her kelime tek başına bir nesneyi ifade ederken birlikte farklı bir şeyi ifade etmeleridir.

3.4. Anlamsal(Semantics-Representation)

Kelime gruplarının bilgisayarın anlayabileceği, kelimeleri birbiriyle karşılaştırabileceği gruplayabileceği eklemeler yaparak yeni kelime türetebileceği ,üretirken kullanacağı formüller halinde ele alınabilecek bir gösterim durumudur. Kısacası mantıksal gösterine ihtiyaç duyulmaktadır.

3.5. Yorumsal (Pragmatics)

Sadece kelimelerin anlamları ya da sadece cümlede nerede oldukları değil aynı zamanda pratik anlamda hangi anlama geldikleri ,güncel hayatta nasıl kullanıldığının yorumlanmasıdır.

3.6. Söylev(Discourse)

Bir kelimenin hangi alanda ne anlama gelir bunun bilinmesi gerekir. Örnek olarak bir top kelime üzerinde düşünüldüğünde futbolda kullanılan ‘top’ kelimesi başka bir şey ifade ederken basketbol kullanılan ‘top’ kelimesi farklı bir şeyi ifade etmektedir. Ya da bambaşka bir alan olan askeri terminoloji de bambaşka bir ifade olur.

3.7. World Knowledge

İnsanların yıllardır edindikleri bilgi birikimi ,okudukları ,araştırdıkları geçmiş birikimler sayesinde kelimelerin ifade ettikleri kültürel anlamlar değişebilmektedir. Bahsedilen konuların değişmesiyle bir kelime farklı anlamlar ifade edebilir. İşte bu kelimelerin doğal dil işleme ile incelenip bilgisayarların anlamasını sağlamak başka bir problem olarak görülmektedir.

Dil bilgisi, kelimenin anlaşılmasından başlanarak, gelen eklerin yorumlanması, seslerin incelenip anlam yüklenilmesiyle hangi konuya dahil olduğu tahmin edilip insan bilgi birikimi ve kültürel anlam ile birlikte en yakın anlamı bulmaya çalışmaktadır.

4. Karşılaşılan Örnek Problemler

Koyun kelimesi ele alındığı; Koyun kelimesi tek başına anlamlı bir kelimedir. Yani ‘koyun’ kelimesi görüldüğünde ya da duyulduğunda direk akla gelen anlam hayvanlar aleminden bir hayvan olma ihtimalidir. Fakat ‘koyun’ kelimesi biraz parçalandığında koy-un ve buradan ilk kelime olan ‘koy’ ifadesi koymak olan bir fiil olur. Bu sefer koyun kelimesine tekrar bakıldığında bir nesneyi bir yere koy anlamında karşıdaki şahsa emir anlamını ifade eder. İşte burada da görülüyor ki bir kelime biraz incelendiğinde başka anlamlar barındırabilir. Bu kelimenin anlaşılır olması bu sebeple önemlidir. Bir kelime hakkında ; fiil ya da ek olarak türemiş bir kelime ya da söz öbeği şeklinde anlaşılması morfolojik seviyede birden fazla anlamını ortaya çıkarmaktadır.

Literatürde geçen art gönderim çözümleme sorunu (anaphora resolution problem) (Van der Sandt, 1992)

olarak da geçen başka bir problem ;

Ali okula otobüsle geldi. O çantasını unuttu .

Cümleleri üzerinde düşünüldüğünde ;ikinci cümlede bulunan ‘O’ ifadesi hangi ifadeyi temsil ediyor bu sorun ortaya çıkmaktadır. Bu durumun sorun olma sebebi ,bu cümlelerin başka bir dile çevrilme durumunda ‘o’ ifadesini kime göre çevirme yapılacaktır, bunun belirsizliğidir. Örnek üzerinden incelenmeye devam edildiğinde Ali kelimesi türkçede erken ismi olarak geçer ve bu ‘o’ ifadesi Ali ismini vurgulayan bir zamirdir. Bu durumda başka bir dile çevrildiğinde ‘O ’ kelimesi erkek(masculen) takısı almalıdır. İngilizcede olan ‘he’ karşılığı gibi.

Başka bir örnek üzerinden devam edildiğinde ;

Elektrik arızası cebe girdi. (Haber, 2015)

Cümlesi üzerinden düşünüldüğünde, bir insan tarafından bu cümle analiz edilebiliyor ve anlaşılıyor. Nedir bu, elektrik arızasının cep telefonuna indirilen bir uygulama sayesinde kolay bir şekilde ihbar edilebilmesidir. Fakat bunun bilgisayar için anlamına bakıldığında bu cümlelerin anlaşılabilmesi için bilgisayarın ciddi bir bilgi birikimine ihtiyacı vardır. Bilgisayarın cep kelimesinden kısaca ‘cep telefonu ’ndan bahsedildiğini , girmek kelimesinden bir uygulamanın cep telefonuna indirilme işlemini vurguladığını tahmin etmelidir.

5. Uygulama Alanları

Doğal dil işleme(DDI),günümüz teknolojilerinde bir çok alanda kullanılmaktadır bunlardan bazılarını şu şekilde sıralamak mümkündür.

5.1. Soru cevaplama(Question answering)

Doğal dil işleme konusunun ilk problemlerinden biridir. Verilen bir metin üzerinden düşünüldüğünde, bu metnin analiz edilerek anlamlı soru ve cevap çıkarılması ya da verilen sorular üzerinden o soruların cevaplarının bir metin üzerinden bulunması istenilir. Bu sorunun çözümü aynı zamanda arama motorlarının da bir problem olan ‘konuşarak (sesli)arama yapabilme sorunu’ çözülmesine de sebep olacaktır.

5.2. Otomatik tercüme(machine translation)

Bir makine tarafından bir metnin tercüme bir dilden başka bir dile çevirme işlemi yapılması istenilmektedir.

5.3. Bilginin Getirilmesi(Information retrieval)

Aranılan bir bilginin metin kaynakları arasında aranıp bulunması ve getirilmesi işlemidir.

5.4. Kelime işleme (Word Processing (Grammer checking, spell checking ,spell correction,Find/Replace))

Sıkça kullanılan bir işlem olan kelime işleme bir çok alanda kullanılan bir takım işlemlerden oluşur. Metin üzerinde yazımsal hataları kontrol etmek için dil bilgisi kontrol işlemi (Grammer checking) kullanılır. Ya da bir kelime için heceleme kontrol ve düzeltme (spell checking and spell correction) işlemleri kullanılır. Kemal Oflazer (Oflazer, (1996))tarafından yazılan bir makale de ,word programında bir kelime arama (Find/replace) işlemi yapıldığında, bul ve değiştir işleminin uygulanması istenildiğinde; örneğin bir metinde okul kelimesinin ev ile değiştirilmesi istenildiğinde ,metinde bulunan bütün okul kelimelerinin ev kelimesi ile değiştirildiği görülmektedir. Fakat burada okullar kelimesi düşünüldüğünde bu kelimenin evlar olarak değiştiği görülecektir. Bu kelime hatasının evler olarak düzeltilme durumu da hece düzeltme-kontrol(spell checking-correctoin) işlemi uygulanmalıdır.

5.5. Doğal Dil İşleme-Veri Tabanı Arayüzü (NL/DB interface)

Veri tabanı üzerinden sorgulama işlemi yapıldığında kullanılan bir işlemidir. Bir üretim departmanında soru olarak gelen ‘Bu yıl ki satış oranı nedir?’ sorusunu veri tabanında uygulamak için bir sorgu cümlesine çeviren bir sistemdir.

5.6. İnternet Arama Motorarı (Web Search)

Doğal dilde arama olarak internet arama motorları sayesinde arama yapma işlemidir. Örnek olarak ,İstanbul kayseri arası kaç km’dir? Everest Dağ’ının yüksekliği ne kadardır? Gibi sorulara cevap aranabilmesi.

5.7. Anti-Intelligence (AI) Bots(Siri)

Konuşularak verilen bir soru üzerine makine tarafından yine konuşarak karşılık verilmesi. İstenilen sorulara arama yaparak bularak karşılık vermesi ve günümüz teknolojilerinde sıkça rastlanan akıllı sistemlerde görülen durumdur. Bu durumun en bilinen örneği SİRİ uygulamasıdır.

5.8. Metin İşleme(Text processing)(Categorization,clustering,IE)

Bir metin üzerinde yapılan kategorize etme, bölütleme ve kümeleme işlemidir. Bir önceki konu olan text mining (Seker, Metin Madenciliği(Text Mining), 2015)konusunda bu konuları daha detaylı görmek mümkündür.

5.9. Finance,Chat Rooms,Telefon Cevaplama

Finansal raporlama ,sohbet odaları ve çokça rastlanılan telefon cevaplama sistemleri. Arama sonucunda karşı tarafta bir makinenin olması ve sorulan sorulara aldığı cevaplar ile yönlendirme yapması ya da sohbet odalarında sohbet diyaloglarına karşılık veren makinelerin bulunması.

5.10. Bilgi Deposu Olarak İnternet(Internet as a repository)

İnternet ortamının en büyük bilgi kaynağı ve ulaşılabilirlik açısından en kolay yol olduğu bilinmektedir. Düşünün her gün milyonlarca kişi tarafından gerek kendi sitelerinde ,gerek bloglarında gerekse kendi kişisel sayfalarında bir şeyler paylaşmaktadırlar. Paylaşılan bu bilgilerin anlaşılıp ,modellenip ve kümelenmesi durumunda bütün internet ağı üzerinde yani insanların ürettiği en büyük bilgi kaynağında ,sonuçta insanlık tarihi bir çok bilgi kaynağı üretmiştir ,geçmiş zamanlarda ,ansiklopediler ,yazıtlar gibi ve bu kaynakların hemen hemen hepsi internet kaynağına çevrilmiştir ,bilgisayarlar tarafından doğal dil işleme yöntemi sayesinde ortaya çıkması diğer insanların bu bilgiye çok kolay bir şekilde ulaşmaları demektir.

5.11. Metin Özetleme(Text summarization)

İnternet üzerindeki bir milyon tane kaynak üzerinden düşünüldüğünde bunlar üzerinden yapılan bir sonucunun bu sayfalar üzerinden bulunması ve özetleme yapılmasıdır.

5.12. Argümanların Birleştirilmesi(Argument aggregation)

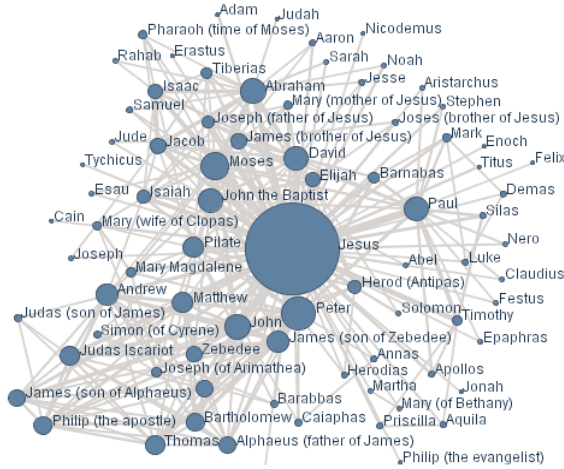
Sosyal ağlar üzerinden yapılan tartışmalar sonucu ortaya çıkan argümanların analiz edilmesi tekrarlı olanların ya da boş olanların temizlenmesi ve kalan argümanlardan anlamlı verilerin üretilmesiyle sonucun bir argümanda birleştirilmesi gerekmektedir. Bu sayede başlatılan bir tartışmaya daha kolay dahil olunması sağlanır. Daha önce hangi argümanlar konuşulmuş neler tartışılmış bunların bilinmesi durumudur.

Güzel bir örnek olarak, sosyal medya da sıkça görülen bir örneği paylaşmakta yarar var.

“A ,B’yi beğendiysen ama C’yi beğenmediyse, C ise D ve E’yi beğendiysen ve hem D hem de E ,A’yı beğendiysen ama A’nın D’nin varlığından haberi bile yoksa, E’nin, C’nin D’yi ve E’yi beğendiğini bilebilmesi için F, G’nin B ile konuşmasını teşvik etmeli mi ki bu sayede C , H’nin çevresini biraz da olsa etkileyebilsin.”

Bu sorunun size sosyal medya da sorulduğunu düşünün ve bunu bilgisayar tarafından sorulması durumunda cevap çıkmasının aranması durumu ortaya çıkacaktır.

6. Metin Madenciliği



Şekil 1-Metin Madenciliği Grafik-1

Yukarıdaki şekilde, Eski Ahit'ten yani Tevrat olarak bilinen kısımdan çıkarılmıştır. Önce metinde ismi geçen kişiler çıkarılmış ve daha sonra da bu kişiler arasındaki ilişkiler doğrultusunda grafiksel görünümü ortaya çıkmıştır.

7. Zamansal Mantık

Bir metindeki zamansal olarak olayların sıralanması hangi olay hangi olaydan sonra ortaya çıkmaktadır bunun incelenmesi ve bunu görsel olarak ortaya konulması işlemidir. Bir roman düşünüldüğünde bu romandaki tüm olayların bir zaman göre sıralanması ,hangi olayın hangi olaydan sonra geldiği konular birbirleriyle ilişkilendirilir.

Bir örnek üzerinde anlatılması gerekirse,

“ Ali odaya girdikten sonra masadaki elmayı yedi.”

Bu cümle incelendiğinde;

Dışarıda	Girdi	Odanın içerisinde
Masadan Uzakta	Yaklaştı	Masanın Yanında
Elması yok	Aldı	Elması Var
		Yedi

Şekil 2-Zamansal Mantık

şeklinde bir tablo ortaya çıkmıştır.Bu tablo üzeri

Bir morfoloji(morphology) örneği verilmesi gerekirse,

* uygarlaştıramayabileceklerimizdenmişsinizcesine

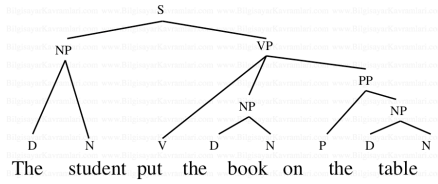
uygar/civilized las/BECOME tir/CAUS ama/NEG yabil/POT ecek/FUT ler/3PL imiz/POSS-1SG

den/ABL mis/NARR siniz/2PL cesine/AS-IF

Bu kelime üzerinden analiz yapıldığında ilk önce kelimenin kökü olan ‘uygar’ kelimesi bulunup daha sonra bu kelimeye gelen eklerin çeşitleriyle ayrıldığı görülmektedir.

8. Parçalama(Parsing)

Aşağıdaki örnek üzerinden cümlelerin parçalanma durumu incelenecektir.



Şekil 3-Parçalama İşlemi

The student put the book on the table.(öğrenci kitabı masanın üzerine bıraktı)

-S(sentence-cümle)→ ilk adım bu bir cümledir.

-NP(Noun phrase-Kelime grubu)

VP(Verb phrase-Fiil grubu)→ kelime gruplarının ayrılması

-D ve N(article(belirli bir nesne) ve noun(isim))→kelime grubundaki ifadenin ayrıştırılması

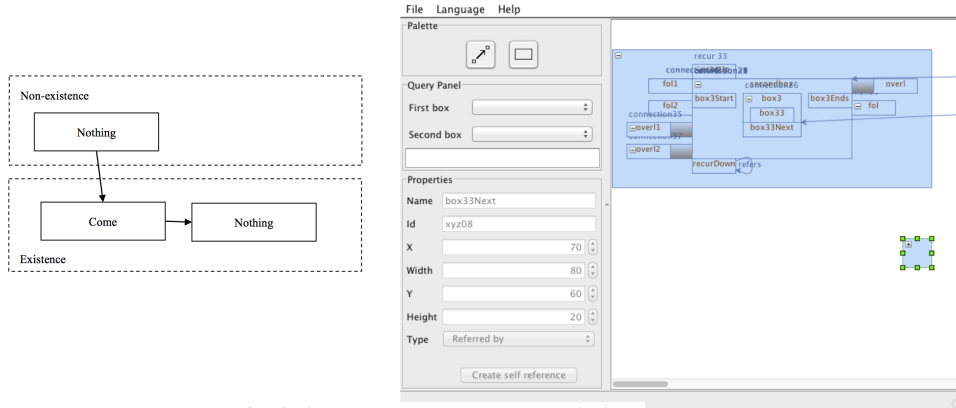
-V-Np-D-N-PP-P-Np-D-N-Son durumun anlaşılması

Buradan da görüldüğü gibi bir parçalama ağacı oluşturuluyor ve bu ağaç üzerinden cümlelerin anlamı çıkarılmaya çalışılıyor.

9. Mantık

Mantıksal bir şekilde gösterilmesi de söz konusudur.

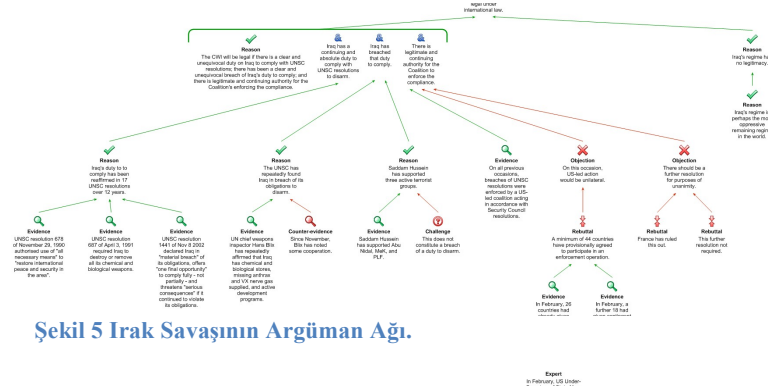
Nothing comes from nothing. Cümlesi düşünüldüğünde ” hiçbir şeyden hiçbir şey çıkmaz” anlamı çıkarılmaktadır. Cümle incelendiğinde ilk önce var olmayan bir durumdan(non-existence),var olan(existence) bir duruma dönüştüğü düşünülmektedir (Seker, Temporal logic extension for self referring, non-existence, multiple recurrence and anterior past events, 2015) .



Şekil 4- Zamansal mantık gösterimi

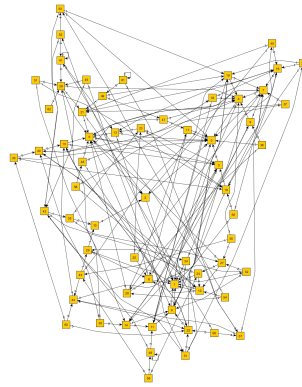
10. Argüman Ağları

Bir tartışmada karşı argümanlarını bulunma ve bu şekilde dallanmaların olma durumudur.



Yukarıdaki tablo Irak'ta bulunan nükleer silahın durdurulması argümanıdır. Burada Amerika'nın bu duruma müdahale etmesi gerekmektedir bu sebeple ortaya argümanlar oluşmaktadır. Çünkü Amerika'nın bu hareketini onaylamayanlar da bulunmaktadır.

11. Argüman Birleştirme(Argument Aggregation)



Şekil 6- Argüman Birleştirme

Sosyal ağ üzerinde tartışan kişiler tarafından gelen argümanlar üzerinden oluşturulan Delphi programıyla oluşturulan bir haritadır (SEKER, Computerized Argument Delphi Technique, 2015). Sosyal medyadaki argümanlar ortaya çıkarılıp ilk önce doğal dil işlemeden geçerek , bunlar arasındaki ilişkiler belirlenerek ortaya çıkmaktadır.

Referanslar

- Brualdi, R. A. (1992). Introductory combinatorics.
- Eryiğit, G. J. (2008). Dependency parsing of Turkish. *Computational Linguistics* 34.3. 357-389.
- Gray, J. e. ((1997)). Data cube: A relational aggregation operator generalizing group-by, cross-tab, and sub-totals.
- Haber, H. (2015, Nisan 25). *Hürriyet Haber*. From hurriyet.com: <http://www.hurriyet.com.tr/elektrik-arizasi-cebe-girdi-28807674>
- Oflazer, K. ((1996)). Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction." *Computational Linguistics* 22.1. 73-89.
- Plattner, H. (2009). A common database approach for OLTP and OLAP using an in-memory column database. *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM* .
- Preece, J. e. (1994.). Human-computer interaction.

- SEKER, S. E. (2015). Computerized Argument Delphi Technique. *IEEE Access* ,v. 3, 368 - 380.
- Seker, S. E. (2015). Metin Madenciliği(Text Mining), YBS Ansiklopedi v. 2, is. 3, pp. 30-32.
- SEKER, S. E. (2015,). Veri Ambarları, YBS Ansiklopedi, v. 2, is. 4, pp. 6 - 13.
- Van der Sandt, R. (1992). Presupposition projection as anaphora resolution. . *Journal of semantics* , 333-377.

Zaman Serisi Analizi (Time Series Analysis)

Sadi Evren SEKER

^aIstanbul 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 extration 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.

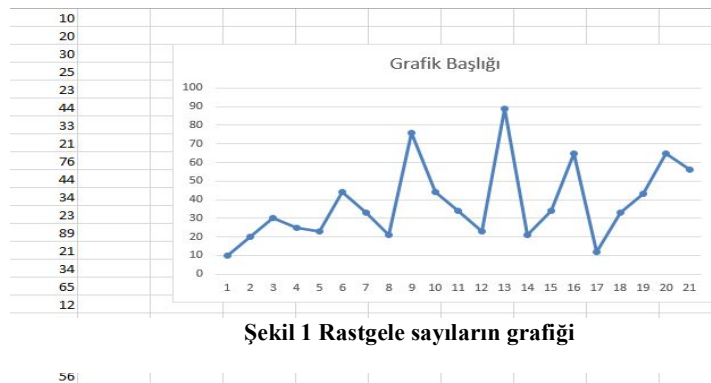
3.1.1. Basit Hareketli Ortalama(Simple Moving Average, SMA)

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

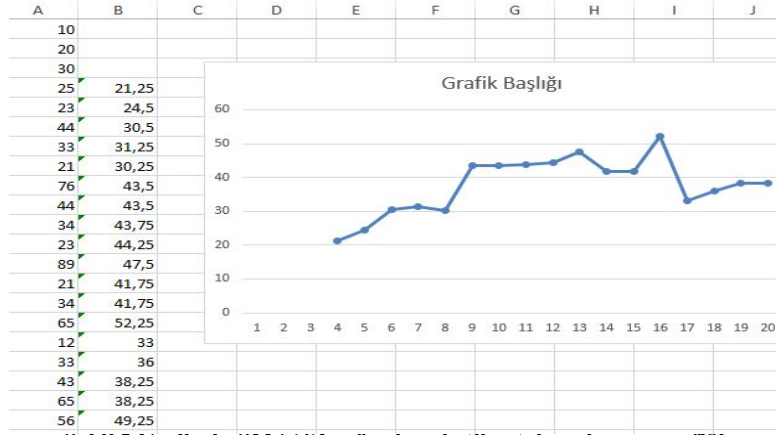
Formül 1 Hareketli ortalama

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.

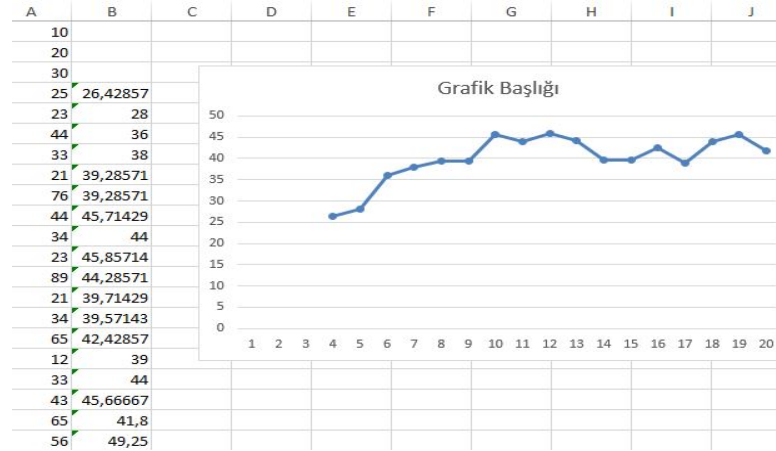


Şekil 1'deki verilerin hareketli ortalamasını $n=4$ değerine göre almak için Excel'de ORTALAMA fonksiyonunu ilk 4 veri üzerinde kullandıktan sonra aynı işlemi diğer veriler üzerinde de uygulayarak bütün verilerin hareketli ortalaması bulunur. Bulunan hareketli ortalamaların grafiği Excel'de çizdirildiğinde 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.



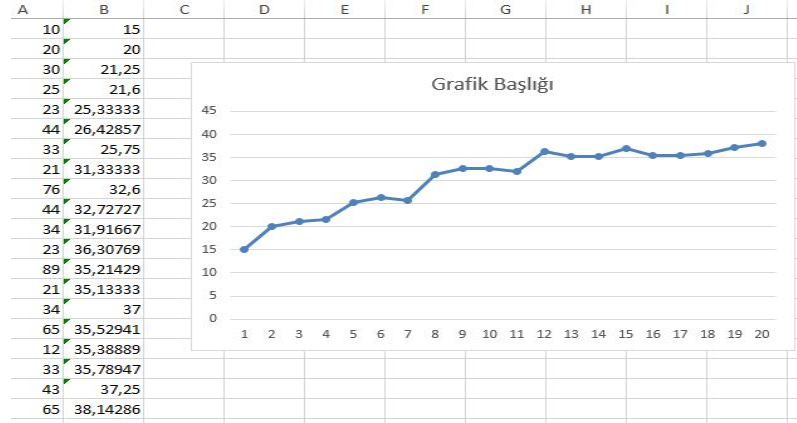
Şekil 3 Verilerin SMA(7) değerine göre hareketli ortalamaları ve grafiği

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($(4*A_4 + 3*A_3 + 2*A_2 + 1*A_1)/(4+3+2+1)$) Şekil 8'deki grafik oluşmaktadır.



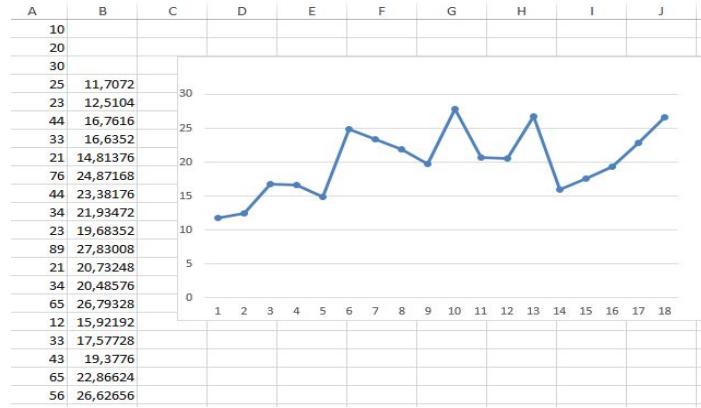
Şekil 5 Verilerin WMA formülüne göre hareketli ortalaması ve grafiği

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.

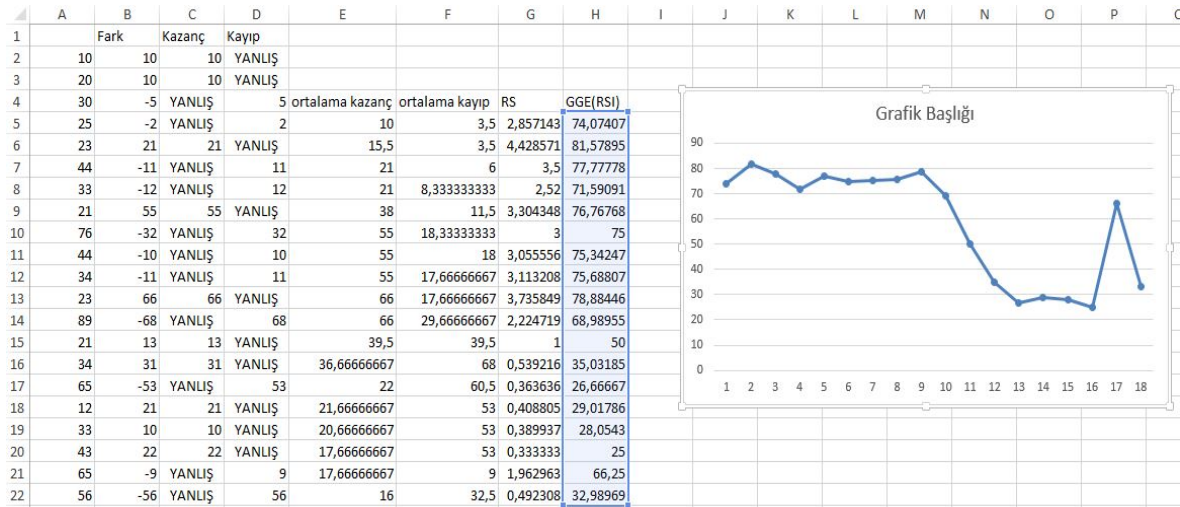
3.2. Göreceli Güç Endeksi(Relative Strength Index,RSI)

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

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

Formül 5 Göreceli Güç Endeksi

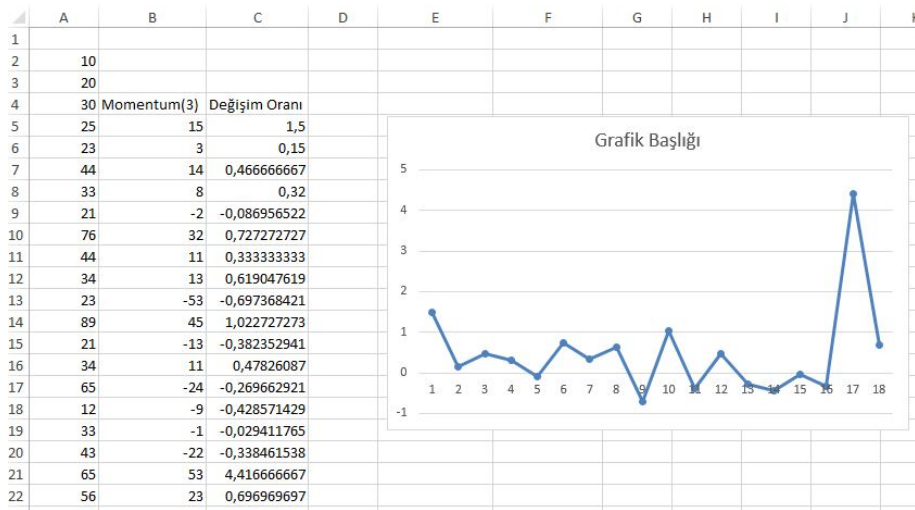
Göreceli Güç Endeksi ile verilerin güç endeksleri(RS) bulunur ve bulunan bu değer formülde yerine yazılır. Literatürde güç endeksleri Formül 5'teki gibi bulunmaktadır (Kim, 2003). Ancak kullanımı daha kolay olan RS=ortalama kazanç/ortalama kayıp formülü de kullanılmaktadır. Ortalama kazanç ve ortalama kayıp değerleri bir sonraki verinin değerinden bir önceki verinin değeri çıkartılarak bulunmaktadır. Çıkarma işleminin sonucu pozitif bir değerse; 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 hesaplandıysa 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*.