



# Sıra Desen Madenciliđi

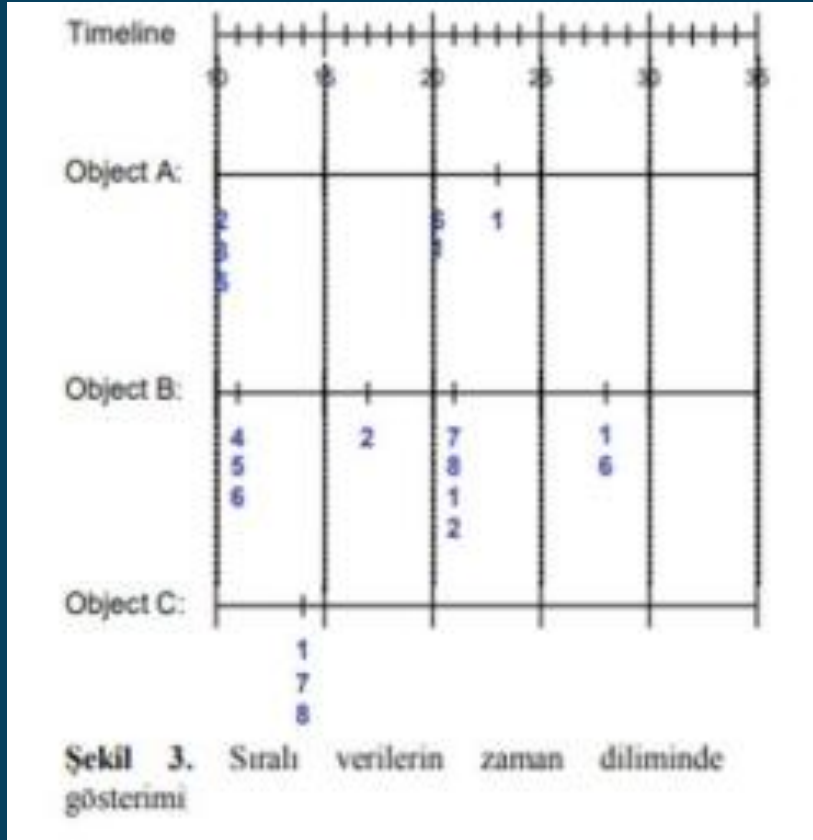
# Sıra Desen Madenciliği

- Sıra desen madenciliği, belirli bir sırayla gelen istatikselsel olarak veri örnekleri arasındaki ilgili örüntüleri bulmaya çalışır. Müşterilerin teknoloji market alışverişi verilerine göre son 3 ayda sırasıyla önce bilgisayar sonra CD-ROM son olarak dijital kamera satın almaları, tıbbi tedaviler, doğal felaketler(deprem), DNA dizilişi ve gen yapısı sıralı örüntü madenciliği ile ilgilidir. Mesela İnternet şubesinde yapılan işlemler bir sıraya göre yapıldığı için sıralı örüntü madenciliği içerisinde yer almaktadır. sıralı örüntü madenciliğinde arka arkaya yapılan işlemler göz önüne alınır. PrefixSpan: PrefixSpan (Prefix-projed Sequential pattern mining) algoritması, sıralı veri madenciliği için çok iyi bilinen bir algoritmadır. Ardışık desenleri, desen büyütme yöntemi ile çıkarır. Algoritma, küçük veri kümeleri için çok iyi performans gösteriyor. PrefixSpan sık tekrarlanan işleri tekrar eder. Örnek vererek anlatmak gerekirse bir sürecin belirli zaman dilimlerinde nasıl tekrarlandığını inceler. Bu tekrarın hangi işlemler öncesi veya sonrası olduğunda dikkate alınmaktadır Algoritmaya yönelik manipilasyon bulmada kritik öneme sahiptir. Süreç analizinde gelecekte oluşacak krizlerin izlerini aramada önem katmakta.

- Örnek: Aşağıdaki tabloda belli eylemler ve bu eylemlerin zaman dilimlerine göre gerçekleşme akışları verilmektedir. Veri madenciliğinin önemli adımlarından biri olan veri etiketleme süreci bu algoritmanında akışında rol oynamaktadır. Burada A – B – C eylemlerini bir internet bankacılığı süreçlerindeki havale(A), ödeme(B), giriş işlemleri(C) gibi etiketleme yaptıktan sonra süreci bir ID tanımlayarak sayısallaştırmaktadırlar. Bu sayıda algoritmanın temelinde yatan örüntü yakalama amacında hızlandırmış olmaktadır. Burada genel yaklaşımı anladıktan sonra, gerçek bir örnek üzerinde ilerleyebiliriz.

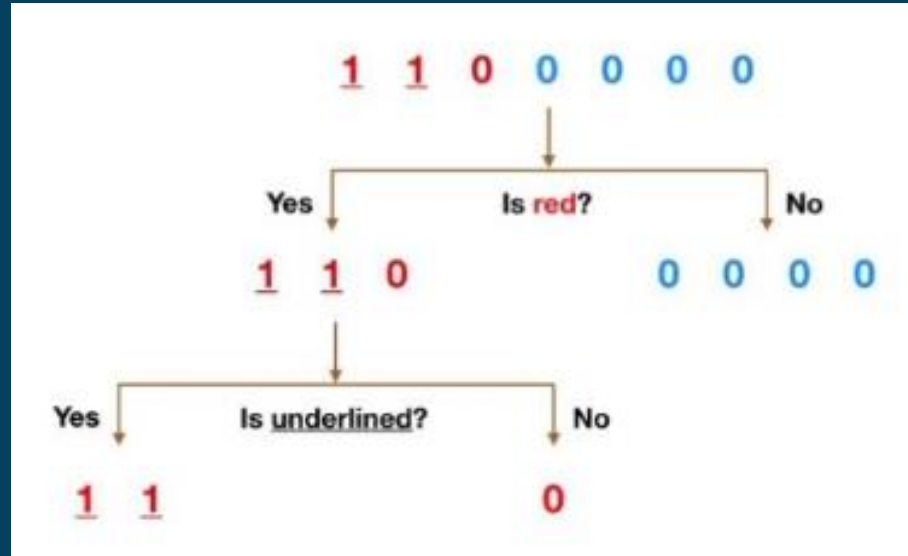
Object	Timestamp	Events
A	10	2, 3, 5
A	20	6, 1
A	23	1
B	11	4, 5, 6
B	17	2
B	21	7, 8, 1, 2
B	28	1, 6
C	14	1, 8, 7

- Aşağıdaki şekil de görüldüğü gibi sıralı örüntü madenciliği yöntemine göre müşterilerin herhangi bir t anında sırayla yaptığı işlemleri, satın aldığı ürünleri, ziyaret ettiği web sayfalarını analiz ederek müşterinin davranışına uygun çıkarımlar yapılabilir. Bir sıralı veri analizi uygulamasının başarılı olması için destek (support) değerinin en küçük destek değerinden (minsupport) başka ifade ile eşik değerinden daha büyük olması gerekir.



# Rassal Orman Modeli Algoritması :

- Rassal Orman Modeli Algoritması Rastgele ormanlar, denetimli bir öğrenme algoritmasıdır. Hem sınıflandırma hem de regresyon için kullanılabilir. Karar ağaçları: Rastgele orman modelinin yapı taşları oldukları için karar ağaçlarının bilinmesi gerekmektedir. Oldukça sezgisel yaklaşımlar içerir. Çoğu insanın hayatlarının bir noktasında bilerek ya da bilmeyerek bir karar ağacı kullandığına bahse girerim. Bir karar ağacının nasıl çalıştığını bir örnek üzerinden anlamak muhtemelen çok daha kolaydır.



- Veri setimizin soldaki şeklin üstündeki sayılardan oluştuğunu hayal edin. İki 1 ve beş 0'ımız var (1'ler ve 0'lar sınıflarımızdır) ve özelliklerini kullanarak sınıfları ayırmak istiyoruz. Özellikler renklidir (kırmızıya karşı mavi) ve gözlemin altı çizili olup olmadığıdır
- Renk, 0'lardan biri hariç tümü mavi olduğu için, ayrılması oldukça bariz bir özellik gibi görünüyor. Böylece "Kırmızı mı?" Sorusunu kullanabiliriz. İlk düğümümüzü ayırmak için. Bir ağaçtaki bir düğümü, yolun ikiye ayrıldığı nokta olarak düşünebilirsiniz - kriterleri karşılayan gözlemler Evet dalına ve Hayır dalına inmeyenler. Hayır dalı (blues) artık 0'lardır, bu yüzden orada işimiz bitti, ancak Evet şubemiz yine de bölünebilir. Şimdi ikinci özelliği kullanıp "Altı çizili mi?" Diye sorabiliriz. İkinci bir bölme yapmak için. Altı çizili iki 1, Evet alt dalına gider ve altı çizilmemiş 0, sağ alt daldan aşağı gider ve hepimiz işimiz biter. Karar ağacımız, verileri mükemmel bir şekilde bölmek için iki özelliği kullanabildi. Açıkçası gerçek hayatta verilerimiz bu kadar net olmayacak, ancak bir karar ağacının kullandığı mantık aynı kalacak. Her düğümde sorulacak - Hangi özellik, eldeki gözlemleri, ortaya çıkan grupların olabildiğince farklı olacağı şekilde (ve ortaya çıkan her alt grubun üyeleri mümkün olduğunca birbirine benzeyecek şekilde) bölmeme izin verir

- Rastgele Orman Sınıflandırıcısı: Rastgele orman, adından da anlaşılacağı gibi, bir topluluk olarak çalışan çok sayıda bireysel karar ağacından oluşur. Rastgele ormandaki her bir ağaç bir sınıf tahmini verir ve en çok oyu alan sınıf, modelimizin öngörüsü haline gelir (aşağıdaki şekle bakın)

