İŞ YERİ ÇALIŞAN DEVAMSIZLIĞINI BELİRLEMEYE YÖNELİK BİR VERİ MADENCİLİĞİ ÇALIŞMASI

İhsan ALPTEKİN
Teknoloji Fakültesi Yazılım Mühendisliği
Fırat Üniversitesi
Elazığ, Türkiye
alptekinihsan62@gmail.com

Özet

Günümüzde birçok kişi yoğun iş stresi altında ezilmekte ve zorlanmaktadır. Yoğun iş hayatından izin yada rapor alarak bu yoğun iş hayatından kaçmaya çalışmaktadırlar. İnsanlar iş yeri stresinden uzaklaşmak için bazı yöntemler geliştirmişlerdir. Bu yöntemler arasında yalancı hastalık durumu, konut ile iş yerindeki mesafe uzaklığı ve benzeri bahane ve davranışlarla izin almaktadırlar. Çalışanların iş yeri devamsızlıkları iş akışını ve iş performansını önemli bir şekilde etkilemektedir. Bu olumsuzlukların ileriki yıllarda tekrarlanmaması için veri madenciliği yöntemlerinden M5P algoritması uygulanmış ve hangi kişinin hangi neden ile işten devamsızlık yaptığı bu algoritma sayesinde tespit edilebilmiştir.

Anahtar Kelimeler: Veri madenciliği, doğrusal regresyon, M5P algoritması, sınıflandırma, işçi devamsızlık durumu, karar ağaçları

1.Giriş

Globalleşme ve artan rekabet, iş akışını önemli bir şekilde değiştirmiştir. Değişen iş akışı sektörleri ve iş alanları büyük ölçüde etkilenmiştir. Bu etkileşimden çalışanlar da payını almıştır. Yoğun çalışma, dinlenememe, stres ve iş gücüne karşı çıkamama sorunları ile karşılaşmaktadırlar. Sorunlar yüzünden çalışanlar işlerden kaçmak için belirli yöntem yada hastalıklar sebebi ile rahat bir nefes alabilmektedirler.

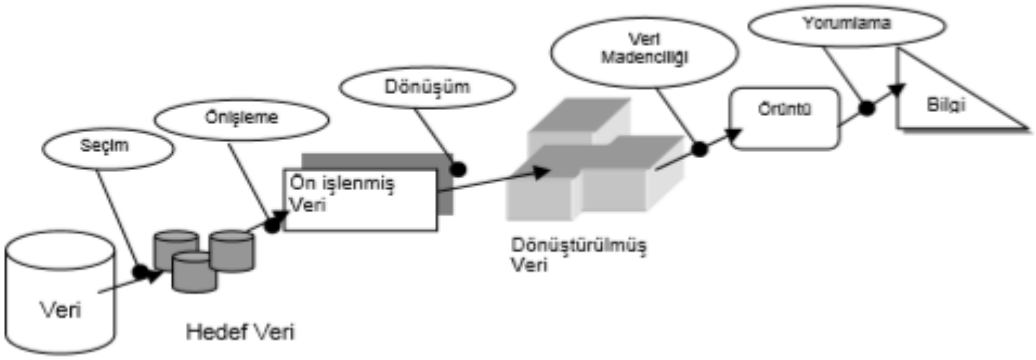
Diğer taraftan bilgisayar teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Bugün, işletmeler çalışanların hangi sebepler ile izin aldıkları gibi pek çok detayı veri tabanlarında tutmaktadır. Veri madenciliği tekniklerini kullanarak bu veriler içerisindeki anlamlı ve gizli örüntülerin ortaya çıkarılması mümkün olmaktadır. Veri madenciliği sonuçları çalışan odaklı birçok uygulamaya girdi teşkil etmektedir.

Çalışmanın 2. bölümünde veri madenciliği, çalışan devamsızlık yönetimi ve veri madenciliğinden bahsedilmiş, 3. Bölümde yapılan çalışma anlatılmış ve son bölüm olan 4. Bölümde sonuçta bulunulmuştur.

2. VERİ MADENCİLİĞİ

2.1. Veri Madenciliğine Genel Bakış

Veri madenciliği, önceden bilinmeyen ilişki ve trendlerin bulunması için bugünün endüstrisinde yaratılan büyük miktarlardaki veriyi analiz eden bir yoldur [1]. Şekil 1'de veri madenciliğinin veri işleme dönüştürme süreci içerisindeki yeri gösterilmektedir [2].



Şekil 1. Veri madenciliğinin veri işleme süreci içindeki yeri [2]

Veri tabanlarından bilgi keşfi, etkileşimli ve tekrarlanan bir süreçtir, aşağıda özetlenen çeşitli adımlardan oluşmaktadır [3]:

• Uygulama tanım kümesini öğrenme: İlişkili ön bilgiyi ve uygulamanın amaçlarını içerir

• Bir hedef veri kümesi oluşturma: Bir veri kümesinin seçimini veya keşfin yapılacağı değişkenlerin veya veri örneklerinin bir alt kümesinde odaklanmayı kapsamaktadır.

• Verilerin temizlenmesi ve ön işleme yapılması: Veri türleri, yöntem, eksik ve bilinmeyen değerlerin eşleştirilmesi gibi veri tabanı yönetim sistemine ait hususların karşılaştırılmasının yanı sıra; uygun ise normal olmayan(gürültülü) veya aykırı değerlerin çıkarılması, gürültüyü modellemek veya açıklamak için gerekli bilginin toplanması, eksik veri alanlarının ele alınması için stratejilerin kararlaştırılması, ardışık zamanlı bilgi ve bilinen değişikliklerin açıklanması gibi temel işlemleri içermektedir.

• Verilerin indirgenmesi: Görevin amacına bağlı olarak verileri temsil etmek için yararlı özelliklerin bulunmasını ve göz önüne alınan değişkenlerin etkin sayısını azaltmak ve veriler için farklı olmayan gösterimler bulmak için boyut indirgeme ve dönüşüm yöntemlerini içerir. • Veri madenciliğinin fonksiyonunu seçme: VM algoritması tarafından türetilen modelin amacını (regresyon, kümeleme, sınıflama, özetleme vb.) kararlaştırmayı kapsar.

• Veri madenciliği algoritmasını seçme: Hangi modellerin ve parametrelerin uygun olabileceğine karar verme gibi verilerdeki örüntüleri arama için kullanılacak yöntemlerin seçilmesini (örneğin kategorik veriler için modeller gerçel sayı vektörlerine dayanan modellerden farklıdır) ve belirli bir VM yönteminin Veri Tabanı Bilgi Keşfi (VTBK) sürecinin bütün kriterleriyle eşleştirilmesini içerir.

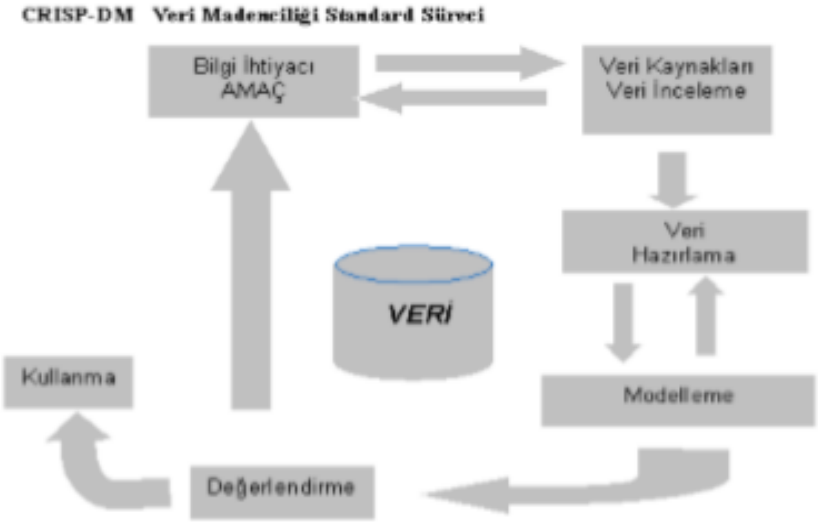
• Veri madenciliği: Sınıflama kuralları veya ağaçları, regresyon, kümeleme, ardışık modelleme, bağımlılık ve doğru analizi gibi belirli bir gösterim biçiminde veya bunların bir kümesinde, ilgilenilen örüntülerin aranmasını kapsar.

• Yorum: Çıkarılan örüntülerin görselleştirilmesi, gereğinden fazla ve ilişkisiz örüntülerin çıkarılması ve yararlı olanların kullanıcılar tarafından anlaşılabilir ifadelere dönüştürülmesinin yanı sıra, keşfedilen örüntülerin yorumlanması ve muhtemelen önceki adımlardan herhangi birine dönülmesini içerir.

• Keşfedilen bilginin kullanılması: Bu bilginin daha önceden çıkarılan veya inanılan bilgi ile potansiyel uyuşmazlıklarının kontrol edilmesi ve giderilmesine ilaveten, bilginin icra sistemine katılması, bilgiye dayanan eylemlerin gerçekleştirilmesi, basitçe belgelenmesi ve ilgili kişilere rapor edilmesini içerir [3]. Veri madenciliğinin gereksinimleri ise erişilebilir veri, etkin erişim yöntemleri, açık problem tanımı, etkin algoritmalar, yüksek tabanlı uygulama sunucusu, sonuç oluşturmada esneklik olarak sıralanabilir [3].

2.2. Veri Madenciliği Süreci

Başarılı bir veri madenciliği projelerinde, izlenmesi gereken adımlar; problemin tanımlanması, verilerin hazırlanması, modelin kurulması ve değerlendirilmesi, modelin kullanılması ve modelin izlenmesidir [4]. Veri madenciliği süreci Şekil 2 ‘de gösterilmektedir.



Şekil 2. Veri madenciliği süreci

Şekil 2’de veri kaynakları veri inceleme adımında problemin tanımlanması yapılır. Projenin hangi işletme amacı için yapılacağının ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanır. Veri hazırlama adımında toplama, birleştirme ve temizleme, dönüştürme işlemleri yapılarak veri, modelleme adımına uygun hale getirilir. Modelleme adımında, model kuruluş süreci denetimli ve denetimsiz öğrenimin kullanıldığı modellere göre farklılık gösterir. Denetimli öğrenmede sistemin amacı, verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunmasıdır. Denetimsiz öğrenmede, ilgili örneklerin gözlenmesi ve bu örneklerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanır. Değerlendirme adımında kurulan modelin doğruluğu test edilir. Kullanma adımında, kurulan ve geçerliliği kabul edilen model doğrudan kullanılabilir veya bir başka uygulamanın alt parçası olabilir. Bütün bu adımlar gerçekleştirildikten sonra kurulan model izlenir.

2.3. Çalışan Devamsızlık Yönetimi ve Veri Madenciliği

Günümüzde şirketlerin ve çalışanları ile ilişkileri büyük ölçüde değişmiştir. İşletmelerde başarılı bir çalışan ilişkileri yönetimi ile küresel bir dünyada ve giderek artan rekabet alanlarıyla işletmeler için yaşamsal önem taşıyan, çalışana için değer yaratmak, çalışan sadakatini sağlamak ve bu konularda hayat kurtarıcı önlemler almak işletmelerin sağlıklı büyümesini gerçekleştirmek mümkün olabilecektir. Diğer taraftan yazılım ve donanım teknolojilerindeki gelişmeler, işletmelerin çok miktarda veriyi saklayabilmesini ve işleyerek anlamlı bilgilere dönüştürmesini mümkün hale getirmiştir. Veri madenciliği sonuçları çalışan edinme, çalışan alan bölümlenmesi, çalışan değerlendirmesi ve çalışanlara ait sosyal ortam yaratılması gibi pek çok çalışan odaklı uygulamaya girdi teşkil etmektedir.

2.3.1. Çalışan Edinme

Yeni çalışan edinme işletmeler için çok zor bir hale gelmektedir. Çalışanın önceki işinden almış olduğu alışkanlıklar ve davranışlar işletmeler açısından hem iyi hem de kötü olabilmektedir. İşletmeler sektördeki itibarlarını çalışanlarının bilgi ve birikiminden almaktadır. Özel bir çalışan ahlakı için bulunan verilerden çıkarılmış sonuçlara göre çalışanın profilinin doğru saptanması ve bu profile uygun eylem planı hazırlanmasıdır.

2.3.2. Çalışan Alan Bölümlenmesi

Çalışan bölümlenmesi ile şirketler hangi çalışanın hangi kriterlere göre hangi bölümlerde daha verimli olabileceğini seçmekte zorlanabilir. Veri analizi ile hangi çalışan hangi alanda daha verimli olabileceği seçilmelidir. Eğer seçilmez ise çalışan iş yerinden ve işlerden kaçmanın yollarını araştırabilir. Bu da şirketler için zorlayıcı bir durum olabilir.

2.3.3. Çalışan Değerlendirmesi

Çalışan alan bölümlemesinden sonra çalışanların yaptıkları iş ve deneyimlerine göre derecelendirip hangi çalışanlar arasında izin kıyaslaması yapılarak iş ortamını düzenlemeye gitmek şirketler için iyi olacaktır. Bu çalışan kıyaslamasını elimizdeki veriler sayesinde kıyaslayabilir ve derecelendire biliriz.

3. YAPILAN ÇALIŞMA

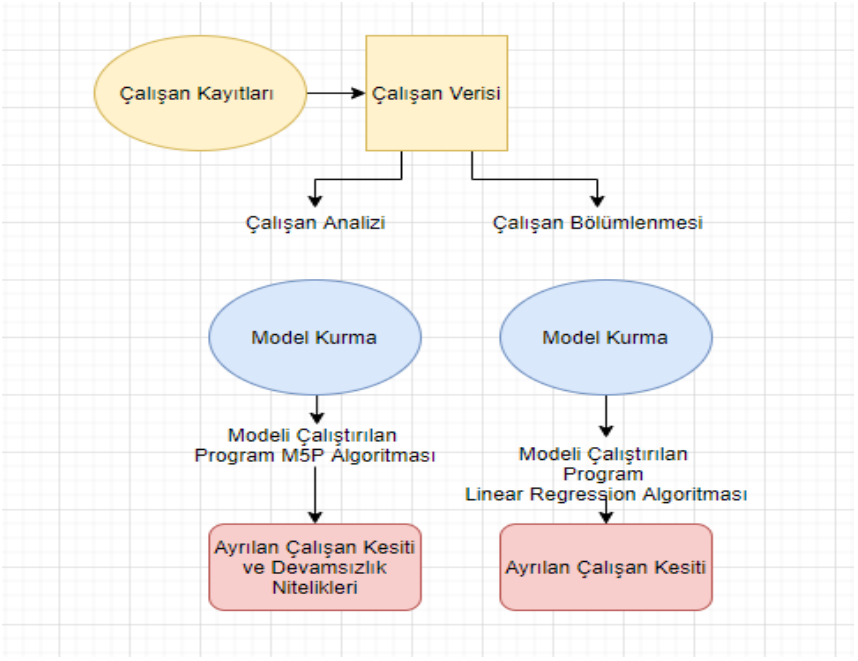
3.1. Problem Tanımı

Çalışmanın amacı , haziran 2007 ile 2010 arasında Brezilya'daki kariyer şirketine ait çalışan devamsızlık durumu veri setinden hangi çalışanın hangi nedenler ile işe gelmediğini belirlemek amacıyla bir eylem planı hazırlamaktır.

İşletme bünyesindeki bütün çalışanların yaş, çocuk sayısı, hastalık durumu gibi bir çok nitelikler ile çalışanları sınıflandırarak hangi neden ile işe gelmediğinin tespitinin yapılması amaçlanmıştır.

3.2. Yapılan Çalışmada Veri Madenciliği Süreci

İşletmenin veri madenciliği sürecinin aşamaları Şekil 5'de özetlenmiştir



Şekil 3. İşletme için veri madenciliği metodolojisi

3.2.1. Modelleme

Bu uygulamada model kurma aşamasında WEKA paket programı kullanılmıştır. WEKA, Java'da yazılmış, Windows, Linux ve Masintosh gibi farklı işletim sistemleri üzerinde çalışabilen bir programdır[11].

3.3. Materyal ve Metot

3.3.1.Veri Seti

Veri setinde 21 tane nitelik olduğundan küçük bir kesiti örnek olarak tablo 1' de gösterilmiştir[7].

Tablo 1. Veri Seti

ID	Devamsızlık Neden	Devamsızlık Aylar	Haftanın Günü	Mevsimler	Ulaşım Gideri
11	26	7	3	1	289
36	0	7	3	1	118
3	23	7	4	1	179
7	7	7	5	1	279
11	23	7	5	1	289
3	23	7	6	1	179
10	22	7	6	1	361
20	23	7	6	1	260
14	19	7	2	1	155
1	22	7	2	1	235
20	1	7	2	1	260
20	1	7	3	1	260
20	11	7	4	1	260
3	11	7	4	1	179

Veri seti Uci Machine Learning sitesinden alınmıştır 21 tane niteliği bulunmaktadır. İşçinin yaşı, çocuk sayısı, hastalık durumu, içki kullanımı ve sigara kullanımı gibi nitelikler bulunmaktadır. Bu niteliklere göre sınıflandırma yaparak kişinin hangi sebeplerden ötürü iş yerinde devamsızlık yaptığı tespit edilmesi amaçlanmıştır.

3.3.2.Linear Regression İle Çalışan Bölümlenmesi

Basit doğrusal regresyon modeli, tek bir açıklayıcı(bağımsız) değişken ile açıklanan(bağımlı) değişken arasında doğrusal(lineer) bir ilişki olduğunda, açıklayıcı(bağımsız) değişken yardımıyla açıklanan(bağımlı) değişkeni tahmin etmek(öngörmek) için kullanılan bir yöntemdir[10].

$$Y=\beta 0+\beta 1X+\epsilon \qquad (1)$$

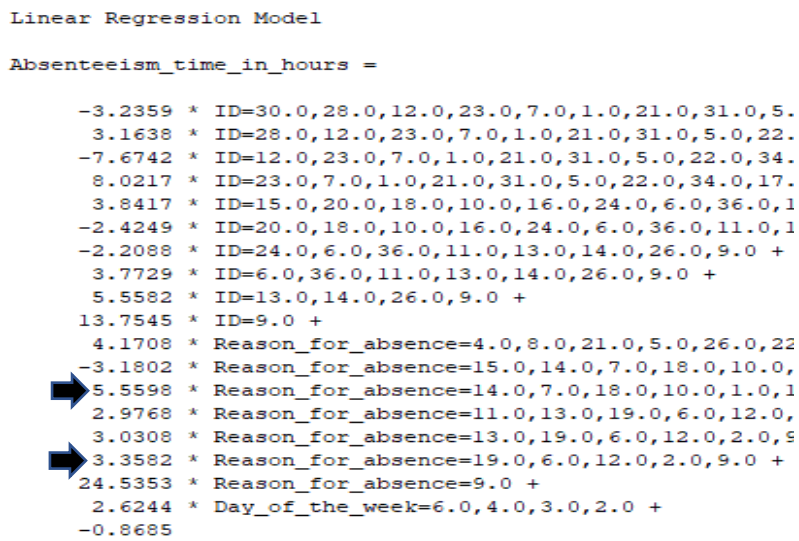
Örnek olarak işletmenin formülü aşağıdaki gibi olur;

$$\text{Çalışan İzin} \approx \beta 0+\beta 1\times \text{İD} \qquad (2)$$

Çalışan izin ve ID(çalışan) arasındaki doğrusal model yukarıdaki gibi olur.

3.3.2.1.Veri Setine Linear Regression(Doğrusal Regresyon) Uygulamak

Yaş verisi Devamsızlık Saati alındığında weka programındaki çıktımız şekil 4'de gösterilmektedir.



Şekil 4. Linear Regression (Doğrusal Regresyon) Modeli

Şekil 4'ü genel olarak ele alacak olur isek 21 nitelikten 3 tanesini değerli bulmuş ve formüle eklemiş diğer nitelikler küçük etkili yada çıkan değerlere göre etkisiz olduğu gözlemlenmektedir.

Model yukarıda devamsızlık saatine göre oluşturulmuştur. Yukarıda belirttiğimiz 1. Formüle göre işleme girmektedir.

İlk ok ile işaretlenmiş yeri ele alacak olur isek ;

5.5582 * ID=13.0,14.0,26.0,9.0 +

5.5582 = hata terimi**,**

ID=13.0,14.0,26.0,9.0 = veri setinde belirtilen yani kullanıcılar,

+ = formülün devamı

İkinci ok ile işaretlenmiş yerleri ele alacak olur isek ;

3.3582 * Reason_for_absence=19.0,6.0,12.0,2.0,9.0 +

3.3582= Hata Terimi,

Reason_for_absence=19.0,6.0,12.0,2.0,9.0 = Devamsızlık Nedeni (hastalık),

+ = formülün devamı

Kısaca bu iki durumu ele alacak olur isek doğrusal bir şekilde işlemin yapıldığı gözlemlenmektedir. Yani ID' si **19,6,12,2,9** olan ve Devamsızlık nedeni **19,6,12,2,9** kişiler birlikte işleme tutulmaktadır. Diğer niteliklerin sınıflandırmada etkin olduğu gözlenmemektedir.

3.3.3.M5P Algoritması İle Çalışan Bölümlenmesi

M5P Quinlan'ın regresyon modellerinin ağaçlarını indüklemek için M5 algoritmasının yeniden yapılandırılmasıdır.M5P geleneksel bir karar ağacını düğümlerde doğrusal regresyon fonksiyonları olasılığı ile birleştirir.

İlk olarak bir ağaç oluşturmak için bir karar ağacı indüksiyon algoritması kullanılır ancak her bir iç düğümdeki bilgi kazancını en üst düzeye çıkarmak yerine her daldaki sınıf değerlerindeki alt-grup içi varyasyonunu en aza indiren bir bölme kriteri kullanılır. Bir düğüme ulaşan tüm örneklerin sınıf değerleri çok az değişiyorsa veya yalnızca birkaç örnek kalırsa M5P'deki bölme yordamı durur.

İkincisi, ağaç her yapraktan budanır. Budama sırasında bir iç düğüm bir regresyon düzlemi olan bir yaprağa dönüştürülür.

Üçüncüsü, alt ağaçlar arasındaki keskin süreksizlikleri önlemek için yaprak modeli tahminini köke geri giden yol boyunca her bir düğümle birleştiren bu düğümlerin her birinde doğrusal modelin öngördüğü değerle birleştirerek düzleştiren bir yumuşatma prosedürü uygulanır. Breiman ve CART sistemleri için numaralandırılmış özellikler ve eksik değerlerle başa çıkmak üzere uyarlanmıştır. Tüm numaralandırılmış öznitelikler ikili değişkenlere dönüştürülür böylece M5P'deki tüm bölünmeler ikili olur. Eksik değerlere gelince M5P orijinalin yerine bölünecek başka bir özellik bulan ve bunun yerine kullanan "vekil bölme" adı verilen bir teknik kullanır. Eğitim sırasında M5P vekil sınıf değerini bölünme için kullanılan olasılıkla ilişkilendirilmesi en muhtemel olan özellik olduğu inancına bağlar. Yarma prosedürü sona erdiğinde tüm eksik değerler yapraklara ulaşan eğitim örneklerinin karşılık gelen özelliklerinin ortalama değerleri ile değiştirilir . Test sırasında bilinmeyen bir öznitelik değeri her zaman en kalabalık alt düğümü seçme etkisiyle düğüme ulaşan tüm eğitim örnekleri için bu özniteliğin ortalama değeri ile değiştirilir. M5P kompakt ve nispeten anlaşılabilir modeller üretir[8].

3.3.3.1.Veri Setine M5P Algoritması Uygulamak

Veri setimize M5P algoritmasını uygulanması şekil 5’ de gösterilmiştir.

```
LM num: 1
Age =
        0.1068 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,
        + 0.3477 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.
        + 0.3027 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        - 0.1033 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
        + 0.4161 * ID=9.0
        - 0.0004 * Month_of_absence
        + 0.0284 * Son
        + 0.0017 * Weight
        + 0.0072 * Height
        + 0.0102 * Body_mass_index
        + 26.3167


LM num: 2
Age =
        0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
        + 0.9453 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,
        + 0.5986 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        - 0.452 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
        + 0.4161 * ID=9.0
        - 0.0004 * Month_of_absence
        + 0.0094 * Son
        + 0.0248 * Weight
        + 0.0072 * Height
        - 0.0008 * Body_mass_index
        + 26.4038
```

LM num: 3
Age =
0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,
+ 2.8161 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
+ 1.0722 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0327 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
+ 0.4161 * ID=9.0
- 0.0004 * Month_of_absence
- 0.1307 * Son
+ 0.0111 * Weight
+ 0.0072 * Height
- 0.0107 * Body_mass_index
+ 26.9933

LM num: 4
Age =
0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,
+ 2.1444 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 1.0722 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0327 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
+ 0.4161 * ID=9.0
- 0.0004 * Month_of_absence
- 0.1281 * Son
+ 0.0111 * Weight
+ 0.0072 * Height
- 0.0107 * Body_mass_index
+ 27.4421

LM num: 6
Age =
0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,
+ 1.5598 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
+ 0.9819 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1231 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
+ 0.4161 * ID=9.0
- 0.0004 * Month_of_absence
- 0.0678 * Son
+ 0.0111 * Weight
+ 0.0072 * Height
- 0.0107 * Body_mass_index
+ 29.1591

LM num: 5
Age =
0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
+ 2.1576 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.
+ 1.3326 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0327 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
+ 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
+ 0.4161 * ID=9.0
- 0.0004 * Month_of_absence
- 0.1247 * Son
+ 0.0111 * Weight
+ 0.0072 * Height
- 0.0107 * Body_mass_index
+ 27.8601

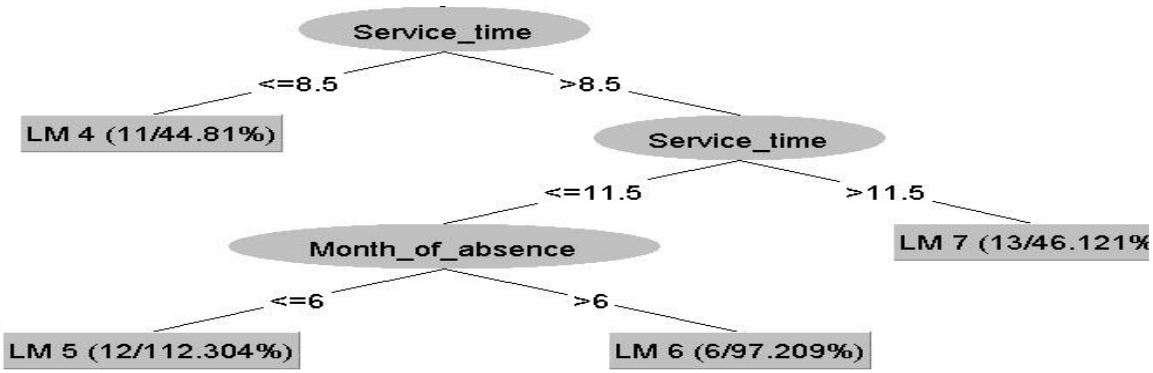
```
LM num: 7
Age =
        0.0851 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
        + 1.215 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.
        + 0.7804 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        - 0.0458 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1597 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0619 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0413 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0481 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0572 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0727 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.044 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1477 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0672 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1298 * ID=32.0,31.0,36.0,35.0,9.0
        + 0.4161 * ID=9.0
        - 0.0004 * Month_of_absence
        - 0.029 * Son
        + 0.0111 * Weight
        + 0.0072 * Height
        - 0.0238 * Body_mass_index
        + 29.5413


LM num: 8
Age =
        0.0122 * ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.
        + 0.0761 * ID=22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9
        + 0.0684 * ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        - 0.0143 * ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.1054 * ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.9312 * ID=34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.4895 * ID=3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 2.6332 * ID=7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0378 * ID=17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 1.2082 * ID=24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 2.3259 * ID=26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 1.4664 * ID=16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 3.4635 * ID=33.0,2.0,32.0,31.0,36.0,35.0,9.0
        + 0.0856 * ID=32.0,31.0,36.0,35.0,9.0
        + 9.4457 * ID=9.0
        - 0.0002 * Month_of_absence
        - 0.0033 * Transportation_expense
        + 0.0237 * Distance_from_Residence_to_Work
        + 0.0071 * Son
        - 0.0033 * Weight
        + 0.2029 * Height
        + 0.0115 * Body_mass_index
        + 1.1252
```

Şekil 5. M5P Algoritması Modelleri

Şekil 5' de görüldüğü üzere doğrusal regresyondaki formüle hemen hemen aynı fakat nitelik sayısında artış ve modeli 8 alt lineer modele bölmüş olması formül içinde 8 parçaya bölünmesi başarımı arttırmaya yöneliktir

3.3.3.2. Karar Ağacı

M5P algoritması karar ağacı oluşturmakta ve bu karar ağacına göre çalışanın işe gelmeme nedenini ortaya koymaktadır. Şekil 6' da gösterilmektedir.



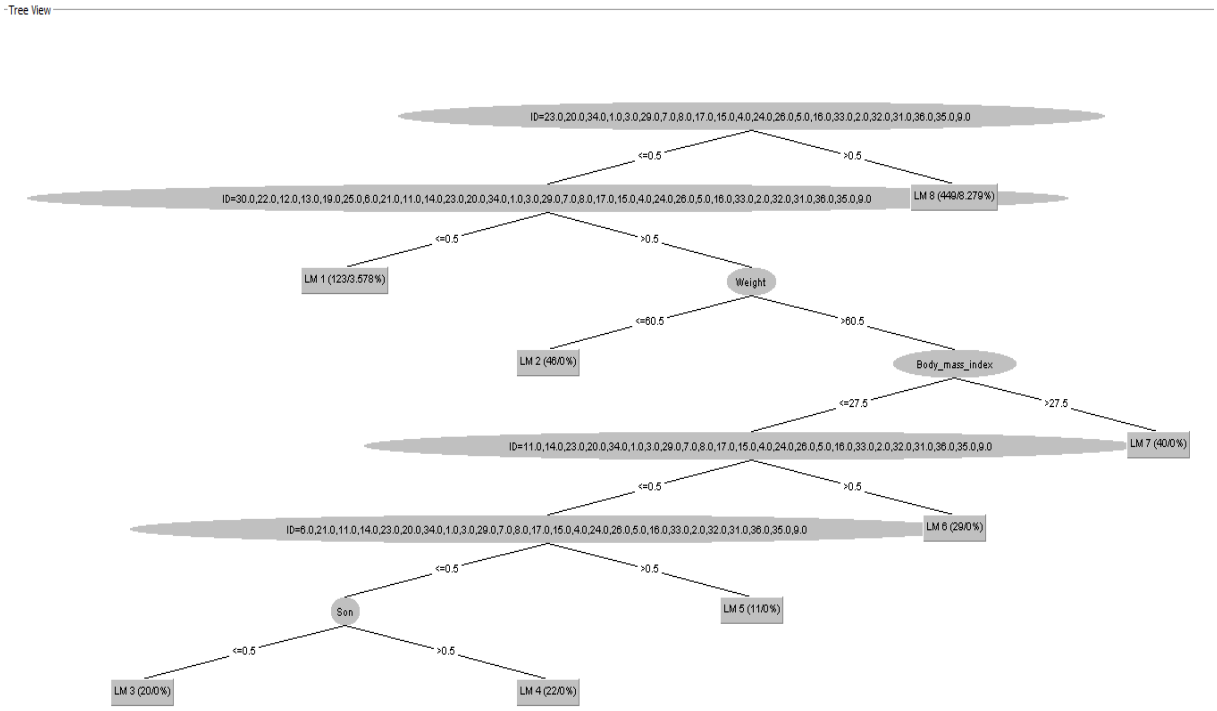
Şekil 6. M5P Algoritması Karar Ağacı İşe Geliş-Gidiş Zamanı Modeli

Şekil 6’ da görüldüğü üzere işe geliş zamanı 8.5’den küçük yada eşit ise LM4(Lineer Model) şeklinde gruplandırma yapmıştır ve LM4’de 11 kişi bulunmaktadır. Eğer işe geliş zamanı 8.5’den büyük ise yine işe geliş zamanına bakılmaktadır ve işe geliş zamanı 11.5’den küçük yada eşit ise aydaki devamsızlık sayısına bakılmaktadır ve devamsızlık 6’dan küçük veya eşit ise LM5 şeklinde gruplandırmıştır ve 12 kişi bulunmaktadır. Eğer 6’dan büyük ise LM6 şeklinde gruplandırmıştır ve 6 kişi bulunmaktadır. İşe geliş zamanı 11.5’den büyük ise LM7 şeklinde gruplandırmıştır ve bu grupta ise 13 kişi bulunmaktadır.

Genel karar ağacı ise Şekil 6.1 ve Şekil 7’deki gibi olmaktadır.

```
M5 pruned model tree:
(using smoothed linear models)

ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 <= 0.5 :
|   ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 <= 0.5 : LM
|   ID=30.0,22.0,12.0,13.0,19.0,25.0,6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 >  0.5 :
|   |   Weight <= 60.5 : LM2 (46/0%)
|   |   Weight >  60.5 :
|   |   |   Body_mass_index <= 27.5 :
|   |   |   |   ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 <= 0.5 :
|   |   |   |   |   ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 <= 0.5 :
|   |   |   |   |   |   Son <= 0.5 : LM3 (20/0%)
|   |   |   |   |   |   Son >  0.5 : LM4 (22/0%)
|   |   |   |   |   ID=6.0,21.0,11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 >  0.5 : LM5 (11/0%)
|   |   |   |   ID=11.0,14.0,23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 >  0.5 : LM6 (29/0%)
|   |   |   Body_mass_index >  27.5 : LM7 (40/0%)
ID=23.0,20.0,34.0,1.0,3.0,29.0,7.0,8.0,17.0,15.0,4.0,24.0,26.0,5.0,16.0,33.0,2.0,32.0,31.0,36.0,35.0,9.0 >  0.5 : LM8 (449/8.279%)
```

Şekil.6.1. Karar Ağacı Sonuç



Şekil .7. Genel Karar Ağacı

Şekil 7 ‘ den anlaşılacağı üzere toplam 740 kişi farklı gruplarla gruplandırılmış en çok kişinin olduğu grup ise LM1’de gözükmektedir. Bu alt grupları topladığımızda ise 740 kişinin olduğu gözlemlenmektedir.

3.3.4.M5P Algoritması İle Linear Regression(Doğrusal Regresyon) ve Full Training(Hepsinin Eğitilmesi)Karşılaştırılması

Uyguladığımız iki algoritmanın hata oranlarını başarımlarını karşılaştırıp hangi algoritmanın daha efektif sonuç verdiği tespiti amaçlanmıştır

Linear regression ve M5P Algoritması için çapraz doğrulama ve bütün verilerin eğitimi;

Tablo 2. Çapraz Doğrulama ve Full Eğitim

Değerler	**CC**	**MEE**	**RMSE**	**RAE**	**RRSE**
LR	0.3085	5.9893	13.1481	99.4635%	98.642%
M5P	0.3035	5.4969	13.3925	91.2862%	100.4752%
M5P(FTraining)	0.9978	0.1396	0.4293	2.7898%	6.6305%

Tablo 2’deki değerleri teker teker karşılaştırdığımız zaman aralarında fazla bir farkın olmadığı gözlemlenmiştir. Hata oranlarının yüksek olması LR ve M5P algoritmalarının veri setimizin verimli çalışmadığını göstermektedir(Çapraz Doğrulama için). Fakat M5P(FTraining) algoritmasında veri setimizin hata oranında ciddi azalma mevcuttur. Bu durumda verilerin hepsinin eğitime girmesi karar aşamasında başarımı arttıracaktır.

Çapraz doğrulama yapılan algoritmalar için weka programında hata oranını düşürmek için birkaç filtre uygulanmıştır. Uygulanacak algoritma M5P seçilmiştir. Bunu sebebi karar ağacı ile alt lineer modellere bölüyor olmasıdır. Buda başarımı arttırmaya yöneliktir.

3.3.5.M5P Sınıflandırma Algoritmasına Filtre Uygulamak

Tablo 3. Filtre Uygulandıktan Sonraki Durum

Değerler	**CC**	**MEE**	**RMSE**	**RAE**	**RRSE**
M5P	0.3035	5.4969	13.3925	91.2862%	100.4752%
M5P(F)	0.5866	4.5879	10.9081	76.1908%	81.8367%

Tablo3’ e baktığımız zaman kolerasyon katsayısı artmış ve hata oranlarında %30’luk bir düşüş gözlemlenmiştir. Fakat veri setimizde istenilen başarı yine gözlemlenememiştir. Lineer modelde ise filtreden önce 8 alt lineer modele bölünmüş iken filtreden sonra 1 lineer model oluşmuştur ve 740 kişiden %73.481’i bu lineer modele dahil olmuştur. Fakat tüm verilerin eğitime girmesindeki sonuçlara yaklaşamamıştır.

4. Sonuç

Sonuç olarak üç algoritmaya baktığımız zaman formüller aynı fakat M5P(FTraining) algoritmasında 8 adet Lineer Model formülü oluşturulmuş karar ağacı ile desteklenmiştir. Elde edilen verilere göre başarım olarak çok fark olduğundan dolayı M5P(FTraining) algoritması Doğrusal regresyon ve M5P algoritmalarına göre daha iyidir.Bu çalışmada kullanılan eğitim dosyaları 740 den fazla örneğe (instance) sahiptir. En fazla örneğe sahip eğitim verileri bu testte kullanılmıştır.

KAYNAKLAR

[1] Hudairy H., “Data Mining and Decision Making Support In The Governmental Sector”, Master Thesis, Louisville University, 2004.

[2] Han J., Kamber M., “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, U.S.A, 2001.

[3] Jacobs P., “Data Mining: What General Managers Need to Know”, Harvard Management Update, Cilt 4, No 10, 8, 1999.

[4] Akpınar H., “Veritabanlarında Bilgi Keşfi ve Veri Madenciliği”, İÜ İşletme Fakültesi Dergisi, Cilt 29, 1-22, 2000.

[5] https://www.uninove.br/curso/informatica-e-gestao-do-conhecimento/

[6] https://bookdown.org/ugurdar/dogrusalregresyon/basit-do%C4%9Frusal-regresyon.html

[7] https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work

[8] https://www.quora.com/What-is-a-simple-explanation-of-the-M5P-M5-model-trees-algorithm-in-machine-learning-data-mining

[9] Orhan ECEMİŞ., “PASLANMAZ ÇELĠK SEKTÖRÜ SATIŞ TAHMİNİNDE VERİ MADENCİLİĞİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI”, İÜ Sosyal Bilimler Dergisi , Cilt 7, 15, 2018.

[10] https://en.wikipedia.org/wiki/Linear_regression

[11] Witten I., Frank E., Data Mining: Practical Machine Learning Tools And Techniques With Java İmplementations, Morgan Kaufmann Publishers, USA, 267-277, 2000.