

Süreç madenciliğine genel bakış: süreç akış keşfi için alfa algoritması

Overview of process mining: Alpha algorithm for process flow discovery

Onur DOĞAN^{1*}

¹İzmir Bakırçay Üniversitesi, İzmir, Türkiye.
onur.dogan@bakircay.edu.tr

Geliş Tarihi/Received: 15.11.2019
Kabul Tarihi/Accepted: 03.02.2020

Düzeltilme Tarihi/Revision: 03.01.2020

doi: 10.5505/pajes.2020.57418
Araştırma Makalesi/Research Article

Öz

Mevcut iş akışı yönetim sistemleri belirli bir iş sürecini uygulayabilmek için önceden bütün adımların ve diğer gerekli tanımların yapılmasını gerektirir. İş akışlarını tasarlamak karmaşık ve zaman alıcı bir iştir. Çoğunlukla gerçekleşen iş süreçleri ile olması istenen süreçler birbirinden farklıdır. Bu nedenle, ideal süreçlerden ziyade gerçekleşen süreçlerin keşfi için çeşitli algoritmalar geliştirilmiştir. Bu algoritmalar için ilk adım, uygulamada olan süreç hakkında detayları içeren olay kayıtlarının elde edilmesidir. Bu çalışmada bilişim teknolojileri (IT) sistemlerinde kayıtlı olan olay kayıtlarını kullanan alfa algoritması tanımlanmıştır. İşe alım süreçlerinde gerçekleştirilen temel faaliyetler ile süreç modelleri oluşturulmuştur. Alfa algoritmasının bir iş akışını başarılı bir şekilde modelleyebildiği gösterilmiştir.

Anahtar kelimeler: Süreç madenciliği, Olay kayıtları, Alfa algoritması, Süreç keşfi, İş akışı

Abstract

Current process management systems require a fully defined workflow design to implement a specific business process. Designing a business process is a complex and time-consuming task. Mostly workflow processes are different from those that are desired. Therefore, various algorithms have been developed for the discovery of actual processes rather than ideal processes. The first step for the discovery algorithms is to extract event logs that contain details about the process. In this study, Alpha algorithm which is used to generate a process model in event logs recorded in information technology (IT) systems is introduced. Process models are discovered using the basic activities performed in the recruitment processes. It has been shown that the alpha algorithm can successfully model a business process.

Keywords: Process mining, Event log, Alpha algorithm, Process discovery, Workflow

1 Giriş

Firmalar yapılan işlerle ilgili birçok veri kaydetmelerine rağmen, veriden anlamlı bilgi çıkarma konusunda sorun yaşamaktadır [1]. Süreç madenciliği, otomatik olarak süreç modelini oluşturmak gibi süreçle ilgili bilgileri çıkarmak için olay kayıtlarını (event logs) kullanmaktadır. Bu açıdan bakıldığında Şekil 1'de gösterildiği gibi süreç madenciliği veri bilimi ve süreç bilimi arasındaki kayıp bağlantı olarak düşünülebilir [2]. Veri bilimi, sadece veri odaklı olarak çözümler üretirken, süreç ile ilgili bilgi vermez [2], [3]. Diğer taraftan süreç bilimi de gerçek verilerden uzak bir şekilde genel anlamda süreç ile ilgili bilgi verir. Süreç madenciliği, veri bilimi tekniklerini kullanarak süreç ile ilgili darboğazlar, iyileştirme önerileri, takip, kontrol gibi bilgiler üretirek iki bilim dalı arasında bir bağ kurar [4], [5].

Süreç madenciliği, bir taraftan makine öğrenmesi ve veri madenciliği diğer taraftan süreç modelleme ve analizi arasında yer alan kısmen yeni bir araştırma disiplindir [6], [7]. Temel fikir, bilişim sistemleri hazır bulunan olay kayıtlarından anlamlı bilgiye ulaşarak varsayılan süreçlerin aksine gerçek süreçleri keşfetmek, gözetlemek ve geliştirmektir. Olay kayıtları, süreç madenciliği için girdi olarak kabul edilen olayların toplamıdır. Endüstri 4.0 çağında dijital dünya ve fiziksel dünya artık iç içe geçmiş durumdadır [8], [9]. Günümüzdeki bilişim sistemleri aşırı miktarda olay kayıtları tutmaktadır.

Literatürde olay kayıtları süreç madenciliğinin üç farklı çeşidi için kullanılmaktadır. Bunlar süreç keşfi, uygunluk kontrolü ve geliştirmedir [10], [11].

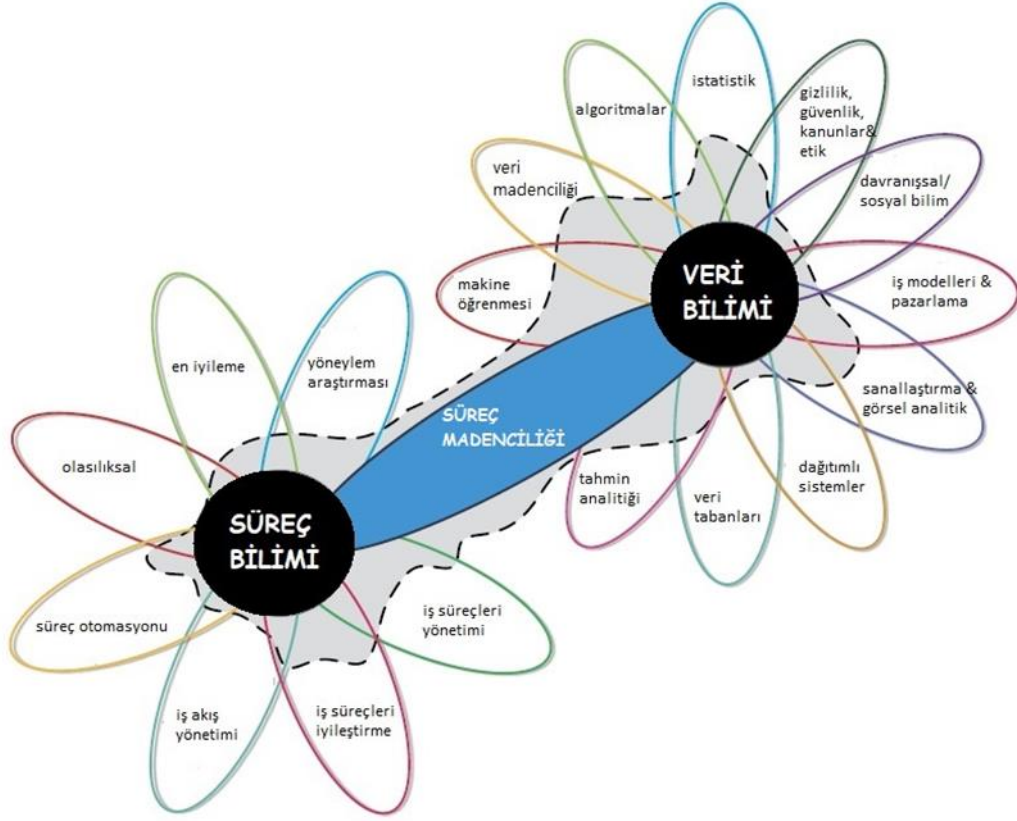
Süreç Keşfi (Discovery): Süreç keşif teknikleri olay kayıtlarından yararlanarak herhangi bir ön bilgi kullanmadan bir süreç modeli üretir. Keşfedilen model Petri ağlar gibi kayıtlardaki davranışları açıklayan bir süreç modeli gösterim notasyonu ile temsil edilir.

Uygunluk Kontrolü (Conformance Checking): Bir sürece ait var olan bir süreç modeli ile aynı sürece ait olay kayıtlarını karşılaştırarak süreç değişkenliğini ölçer. Yani kayıtlardaki olayların model ile olan uyumunu test eder. Uygunluk kontrolü değişkenliği açıklamak, ortaya çıkarmak ve yerini bulmak ve bu değişkenliğin şiddetini ölçmek için kullanılabilir.

Geliştirme (Enhancement): Amaç, olay kayıtlarını kullanarak gerçek bir süreci iyileştirmek ve geliştirmektir. Uyumluluk kontrolü model ile olay kayıtlarını karşılaştırır iken, geliştirme mevcut süreci ileriye götürür. Geliştirmenin bir çeşidi *tamir (repair)*dir. Tamir, gerçeği daha iyi yansıtması için modeli düzenleme çalışmasıdır. *Uzatma (extension)* ise diğer geliştirme çeşididir. Sürece farklı bir perspektif katarak onu geliştirme çalışmalarını içerir.

Genel olarak süreç madenciliğinde 4 temel perspektif kullanılır [11], [12].

*Yazışılan yazar/Corresponding author



Şekil 1: Veri bilimi ve süreç bilimi arasında bir köprü olarak süreç madenciliği [10].

- ✓ Akış Kontrolü perspektifi (Control-Flow Perspective) faaliyetlerin sırasını dikkate alır. Bu perspektifteki madencilik amacı, tüm olası yolları temsil eden iyi bir akış bulmaktır.
- ✓ Organizasyonel perspektif (Organizational perspective) olay kayıtlarındaki insan, sistem, rol veya departman gibi kaynaklara odaklanır. Bunların birbirleriyle olan ilişkileri incelenir. Buradaki amaç, kaynaklar arasındaki sosyal ağı göstermek veya organizasyonel birim ve roller açısından kaynakları sınıflandırmaktır.
- ✓ Olay perspektifi (Case perspective) olayların özelliklerine odaklanır. Bir olay, süreçteki yoluyla veya olayın başlangıç faaliyetiyle temsil edilebilir. Örneğin, bir malzeme yenileme siparişinde, tedarikçiyi bilmek veya sipariş verilen ürün sayısını bilmek önemli olabilir.
- ✓ Zaman perspektifi (Time perspective) olayların zamanı veya sıklığı ile ilgilidir. Olayların başlangıç ve bitiş zamanları bilindiğinde süreç içindeki darboğazları belirlemek, hizmet seviyesini ölçmek, kaynakların kullanımını görüntülemek, sürecin kalan sürecini tahmin etmek mümkün olur.

Süreç keşfi, süreç madenciliğinin en zor konularından biridir [13]. Kayıtlarda görülen davranışların yakalanarak bir süreç modelinin oluşturulması amaçlanır. Süreç keşfi için birçok algoritma geliştirilmiştir. Alfa (α) algoritması süreç keşfini anlamaya yardımcı olan ve çoğu süreç keşfi algoritması tarafından kullanılan bir algoritmadır [14]. Bu bölümde daha

açıklayıcı olması için akış kontrolü perspektifinden süreç keşfi anlatılmıştır. Diğer perspektifler ihmal edilmiştir. Çalışma, alfa algoritmasının işe alım süreçlerinde uygulaması bakımından literatüre katkı sağlamaktadır. Ayrıntıları "Literatür Taraması" başlığında verilen çalışmalar, farklı algoritmalarla sağlık, eğitim ve insan davranışları gibi farklı alanlara odaklanmıştır. Her uygulama alanı kendine özgü özellikler içermektedir. Bu nedenle alfa algoritması, daha önce uygulaması olmayan bir alanda adım adım açıklanarak süreç modelinin keşfedilmesi amacıyla kullanılmıştır.

2 Alfa (α) Algoritması

Alfa algoritması, ilk süreç keşif algoritmalarından biridir. Bu algoritma eş zamanlı faaliyetlerin modele aktarılmasında iyi çalışmaktadır. Ancak, α algoritması madencilik konusunda çok pratik değildir [14]. Çünkü gürültü (noise), seyrek ve tamamlanmamış (infrequent and incomplete) davranışlar ve karmaşık rota yapılarını çözme konusunda yetersizdir [15]. Buna rağmen süreç keşfinin mantığını anlamak için iyi bir algoritmadır.

Alfa algoritması belli bir olay için bütün kayıtları tarar. Eğer a faaliyeti b faaliyeti tarafından takip edildiyse ve b faaliyeti a faaliyeti tarafından hiç takip edilmediyse a faaliyeti ile b faaliyeti arasında bir *nedensel bağımlılık* (causal dependency) olduğu varsayılır. Bu bağımlılığı göstermek için Petri ağı, a ve b'yi bağlayan bir yere sahip olmalıdır.

Tanım: Log tabanlı sıralama ilişkisi

L, A faaliyet kümesi üzerindeki olay kayıtlarını gösterebilir. $a, b \in A$ olsun. Dört farklı log tabanlı sıralama ilişkisi tanımlanabilir.

- ✓ $a >_L b$; yalnız ve yalnız $\sigma = \langle t_1, t_2, t_3, \dots, t_n \rangle$ izi varsa ve $i \in \{1, \dots, n-1\}$ ise, $\sigma \in L$, $t_i = a$ ve $t_{i+1} = b$ olduğunda
 - ✓ $a \rightarrow_L b$; yalnız ve yalnız $a >_L b$ ve $b \not\#_L a$ olduğunda
 - ✓ $a \#_L b$; yalnız ve yalnız $a \not\#_L b$ ve $b \not\#_L a$ olduğunda
 - ✓ $a ||_L b$; yalnız ve yalnız $a >_L b$ ve $b >_L a$ olduğunda
- $L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$ L_1 olay logunun içerdiği izleri gösterebilir. Bu olay logu için aşağıdaki log tabanlı sıralı ilişkileri türetilebilir [13]:
- $>_{L_1} = \{(a, b), (a, c), (a, e), (b, c), (b, d), (c, b), (c, d), (e, d)\}$
 $\rightarrow_{L_1} = \{(a, b), (a, c), (a, e), (b, d), (c, d), (e, d)\}$
 $\#_{L_1} = \left\{ \begin{array}{l} (a, a), (a, d), (b, b), (b, e), (c, c), (c, e), (d, a), (d, d), \\ (e, b), (e, c), (e, e) \end{array} \right\}$
 $||_{L_1} = \{(b, c), (c, b)\}$
 $>_{L_1}$ ilişkisi birbirini "direkt" olarak takip eden bütün faaliyet çiftlerini içerir. Örneğin, $c >_{L_1} d$ ile gösterilir. $\langle a, b, c, d \rangle$ izinde, d direkt olarak c 'yi takip eder. Başka bir ifadeyle, c, d tarafından direkt olarak takip edilir.
 \rightarrow_{L_1} ilişkisi tüm "nedensellik ilişkisi" içinde olan faaliyet çiftlerini içerir. $c \rightarrow_{L_1} d$ ile gösterilir. Bazen d direkt olarak c 'yi takip eder ancak c asla d 'yi direkt olarak takip etmez. Yani $c >_{L_1} d$ ve $d \not\#_{L_1} c$.
 $\#_{L_1}$ ilişkisi birbirini direkt olarak takip etmeyen, yani "seçim" yapılan tüm faaliyet çiftlerini içerir. Örneğin, $b \#_{L_1} e$ ilişkisinde b asla e 'yi ve e asla b 'yi direkt olarak takip etmez. Yani $e \not\#_{L_1} b$ ve $b \not\#_{L_1} e$.
 $||_{L_1}$ ilişkisi paralel zamanlarda gerçekleşen tüm faaliyet çiftlerini içerir. Örneğin, $b ||_{L_1} c$ ilişkisinde, bazen b, c 'yi bazen de c, b 'yi direkt olarak takip eder. Yani, $c >_{L_1} b$ ve $b >_{L_1} c$.

Faaliyetler çiftleri arasındaki bu ilişkileri kolayca görebilmek için dijital ayak izi (footprint) matrisi oluşturulur. L_1 logu için oluşturulan dijital ayak izi matrisi Tablo 1'de gösterilmiştir.

Tablo 1: L_1 logunun dijital ayak izi matrisi.

	a	b	c	d	e
a	$\#_{L_1}$	\rightarrow_{L_1}	\rightarrow_{L_1}	$\#_{L_1}$	\rightarrow_{L_1}
b	\leftarrow_{L_1}	$\#_{L_1}$	$ _{L_1}$	\rightarrow_{L_1}	$\#_{L_1}$
c	\leftarrow_{L_1}	$ _{L_1}$	$\#_{L_1}$	\rightarrow_{L_1}	$\#_{L_1}$
d	$\#_{L_1}$	\leftarrow_{L_1}	\leftarrow_{L_1}	$\#_{L_1}$	\leftarrow_{L_1}
e	\leftarrow_{L_1}	$\#_{L_1}$	$\#_{L_1}$	\rightarrow_{L_1}	$\#_{L_1}$

Süreç keşfi için temel fikir anlatıldıktan sonra, α algoritmasını anlamak daha kolay olacaktır.

$L, T \subseteq A$ üzerinde olay logu olsun. $\alpha(L)$ aşağıdaki adımlarla tanımlanır [1]:

1. $T_L = \{t \in T | \exists \sigma \in L t \in \sigma\}$
2. $T_I = \{t \in T | \exists \sigma \in L t = ilk(\sigma)\}$
3. $T_O = \{t \in T | \exists \sigma \in L t = son(\sigma)\}$
4. $X_L = \left\{ \begin{array}{l} (A, B) | A \subseteq T \wedge A \neq \emptyset \wedge B \subseteq T \wedge B \neq \emptyset \wedge \\ \forall a \in A \forall b \in B a \rightarrow_L b \wedge \forall a_1, a_2 \in A a_1 \# a_2 \wedge \forall b_1, b_2 \in B b_1 \# b_2 \end{array} \right\}$
5. $Y_L = \left\{ (A, B) \in X_L | \forall (A', B') \in X_L A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B') \right\}$
6. $P_L = \{p_{(A,B)} | (A, B) \in Y_L\} \cup \{i_L, i_O\}$

7. $F_L = \{a, p_{(A,B)} | (A, B) \in Y_L \wedge a \in A\} \cup \{p_{(A,B)}, b | (A, B) \in Y_L \wedge b \in B\} \cup \{i_L, t | t \in T_1\} \cup \{t, o_L | t \in T_O\}$
8. $\alpha(L) = (P_L, T_L, F_L)$

L olay logu, T ise faaliyetler kümesidir. Algoritma adımları aşağıda sırayla açıklanmıştır.

1. L olay logundaki her faaliyet, $\alpha(L)$ 'de bir geçiş (transition) sahiptir. Algoritma T_L logunda görülen faaliyetleri kontrol ederek başlar.
2. T_I , başlangıç faaliyetleri kümesidir. Başlangıç faaliyeti bir izdeki ilk faaliyettir.
3. T_O , son faaliyetlerin kümesidir. Son faaliyet bir izdeki en sonda gerçekleştirilen faaliyettir.
4. Adım 4 ve Adım 5 algoritmanın en önemli aşamalarıdır. Bu adımda aşağıdaki şartları sağlayan T_L logunda tanımlanmış ve boş kümeden farklı olan ($A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset$) (A, B) çiftleri belirlenir.
 - a. A kümesinden alınan herhangi bir faaliyet ile B kümesinden alınan herhangi bir faaliyet arasında mutlaka nedensellik ilişkisi olmalıdır. ($\forall a \in A \forall b \in B a \rightarrow_L b$) ile anlatılan budur.
 - b. A (veya B) kümesinden alınan iki faaliyet (aynı faaliyet olsa bile) asla birbirini takip etmemelidir. ($\forall a_1, a_2 \in A a_1 \#_L a_2 \wedge \forall b_1, b_2 \in B b_1 \#_L b_2$) ile anlatılan budur.
5. Adım 4'teki şartları sağlayan (A, B) faaliyet çiftlerinin herhangi bir alt kümesi otomatik olarak bu şartları yine sağlayacaktır ($\forall (A', B') \in X_L A \subseteq A' \wedge B \subseteq B'$). Bu, gereksiz olan birçok yerin (place) belirleneceği anlamına gelmektedir. Bu nedenle, tüm faaliyet çiftlerinden en geniş kapsama sahip olan çiftler kalır, alt kümelerden oluşan diğer çiftler silinir ($(A, B) = (A', B')$). Adım 5'te yapılan da budur.
6. Süreçteki yerler bulunur. ($p_{(A,B)} | (A, B) \in Y_L$) ifadesi Adım 5'ten elde edilen en geniş kapsamlı faaliyet çiftlerinin ağıdaki yerleri ifade ettiğini göstermektedir. Bulunan bu yerlere başlangıç ve bitiş yerleri de eklenir (i_L, i_O).
7. Ağıdaki akışlar (flow) oluşturulur. Adım 6'da bulunan yerlere bu akışlar ilave edilir. Girdisi a faaliyeti ($a, p_{(A,B)} | (A, B) \in Y_L \wedge a \in A$), çıktısı b faaliyeti ($p_{(A,B)}, b | (A, B) \in Y_L \wedge b \in B$), başlangıç yeri i_L ($i_L, t | t \in T_1$), bitiş yeri o_L ($t, o_L | t \in T_O$) olan bir akış çizilmiş olunur.

Adım 7, algoritmanın son adımıdır. Yerlerin (P_L), geçişlerin (T_L) ve akışların (F_L) birleştirilmesiyle süreç modeli son halini alır.

3 Literatür Taraması

Süreç madenciliği literatürde önem kazanan ve çekici bir konudur. Sağlık [16], [17], [18] eğitim [19], finans [20], bilgi yönetimi [21], insan davranışları analizi [3], [4], [6], üretim [22], belediye süreçleri [23] gibi pek çok alanda geniş uygulama alanları vardır. Endüstri 4.0 çağında, dijital dünya ve fiziksel dünya iç içedir. IT sistemleri, süreçleri yönetmek için kullanılacak çok sayıda olay kayıtlarını saklamaktadır. Süreç madenciliği, kaydedilen olay kayıtlarını, iş süreçlerini çeşitli amaçlar için girdi olarak kullanır. Geçmiş çalışmaların özeti Tablo 2'de sunulmaktadır.

Fernandez-Llatas ve meslektaşları [17] iç mekan sistemleri kullanarak süreç madenciliği tabanlı bir yöntem sunmuşlardır. Bu yöntemle veriler, hemşireleri zaman ve süreçlerle ilgili diğer bilgileri ölçmeye zorlamadan RFID teknolojisi ile otomatik olarak toplanarak süreç madenciliğinde farklı süreçlerin akışlarını elde etmek için kullanılmıştır. Mans ve diğerleri [24] süreç madenciliğini dış hekimliğinde uygulamıştır. Önce sezgisel madencilik (heuristic mining) algoritması ile süreç modelini keşfedilmiş ve sonra ortaya çıkarılan süreci

geliştirilmiştir. Sedrakyan ve diğerleri [19] kullanıcıların öğrenme davranışlarını içeren olay kayıtlarıyla süreç madenciliği tekniklerini uygulamıştır. Çalışmada, öğrencilerin sınav notları ile yaş ve cinsiyet gibi temel özellikler arasında bir ilişki kurularak öğrencilere yönlendirme ve geri bildirim sağlamaktadır. Atastina ve Kurniati [25] her dönemin başında yapılan öğrenci kayıt sürecini bulanık madencilik ve alfa algoritmalarıyla değerlendirmektedir.

Tablo 2: Geçmiş çalışmalar

Çalışma	Yıl	Algoritma	Perspektif	Açıklama
[3]	2017	Tümevarım algoritması	Akış Kontrolü	Çok karmaşık, dinamik ve önceden tanımlanması mümkün olmayan insan davranışlarının analizi geçmiş verilerin süreç madenciliğiyle yorumlanmasıyla yapılabilir.
[14]	2012	Alfa algoritması	Akış	Ardaşık olaylardaki nedenselliğin yeniden yapılandırılması için alfa algoritması kullanılarak öğrenci kayıt sürecinin yeniden yapılandırılmıştır.
[17]	2015	PALIA	Akış Kontrolü	Hastanelerde verilerin toplanması genellikle rahatsız edici bir şekilde yapılır. Sağlık personelleri süreç için gerekli bilgiyi ve zaman verisini kaydetmesi gerekir ve bu durum süreç analizini süreci etkiler. İç mekan sistemlerinin kullanılmasıyla süreç verileri otomatik olarak toplanarak ve toplanan büyük miktardaki zaman-mekan kayıt verileriyle başa çıkamayan sağlık profesyonellerinin kolaylıkla anlayabileceği bir şekilde sunmak gerekir.
[18]	2018	PALIA	Akış Kontrolü	Felç hastalığının kritik zamanlamasını dikkate alarak acil durum süreçlerinin interaktif analizinde sağlık profesyonellerinin desteklenmesi amacıyla süreç madenciliği uygulanmıştır.
[19]	2016	-	Akış Kontrolü	Öğrenme kalitesini artırmak için öğrenme davranışı gözlemlerine dayalı bir geri bildirim sistemi süreç madenciliği kullanılarak kurulmuştur.
[20]	2013	-	Akış Kontrolü	Finans sektöründeki çoğu işin çalışanlar tarafından yapıldığı arka oda faaliyetlerinin analizi için süreç madenciliği uygulanmıştır.
[24]	2012	Sezgisel algoritma	Akış Kontrolü, Geliştirme	Dış hekimliğinde, sürece dahil olan çok sayıda bağımsız insan olduğundan süreç performansının iyileştirilmesi karmaşık bir iştir. Uygulamadaki mevcut süreçler hakkında doğru bir görüş elde etmek ve süreç performansını iyileştirmek için süreç madenciliği uygulanmıştır.
[27]	2015	Sezgisel algoritma	Akış Kontrolü	Yapısal olmayan verilerle müşteri memnuniyetinin ölçümünde süreç madenciliği uygulanmıştır.
[30]	2016	-	Akış	İşçilerin hastalanması süreci ve işsizlik sürecindeki süreç karakteristikleri farklı açılardan (kontrol veri akışı, süre, maliyet, firma,...) incelenmiş ve sınıflandırma ve tahmin etme gibi veri madenciliği yöntemleriyle de sonuçlar üretilmiştir.
[31]	2016	Sezgisel algoritma	Akış	Ayakta tedavi edilen hastalara ait olay kayıtlarıyla çevresel değişiklikler, bekleme süresi ve tedavi için harcanan süre gibi değişen durumlardaki süreçler analiz edilmiştir.
[32]	2013	Alfa ve sezgisel algoritmaları	Akış	Çoğu web analitik araçları, işletmeler için etkili öngörüler sunabilen temel müşteri süreçleri ve davranışları hakkında bir bilgi üretmemektedir. E-ticaret web sitesi olay kayıtlarına iş süreci yönetimi (BPM) metodolojilerinin uygulanması önerilmiştir. Yaklaşımın zorlukları, sonuçları ve potansiyel faydaları sunulmuştur.
[33]	2019	Alfa ve tümevarımsal algoritmaları	Akış	Akıllı Öğrenme Ortamında (SLE) öğrencilerin otomatik geri bildirim tepkilerini keşfetmenin olası yolları araştırılmaktadır.

Geleneksel ve özelleştirilmiş süreç madenciliği algoritmaları, büyük ölçekli çevrimiçi web verilerinden iş süreç modellerini çıkarmak için de değerlendirilir. De Weerd ve arkadaşları [20], süreç madenciliğine dayalı gerçek yaşam olay kayıtlarının çok yönlü analizi için metodolojik bir çerçeve önermektedir. Çalışma, büyük ölçekli bir Belçikalı sigorta şirketinin arka ofis süreçlerindeki örgütsel verimsizlikleri keşfetmek için süreç

madenciliği tekniklerinin kullanılabilirliğini ve esnekliğini göstermektedir. Zorlukları, sonuçları ve potansiyel faydaları araştırmak için e-ticaret web sitesi kayıtlarına iş süreci yönetimi tabanlı bir yaklaşım önerilmiştir. Başka bir çalışmada Fernandez-Llatas ve arkadaşları [26], dokuz kişiye ait olan ve kablosuz teknoloji ile toplanan 25 haftalık veriye sahip kişilerin hareketlerini analiz etmek için süreç madenciliğini uygulayarak

kişiselleştirilmiş davranışları ortaya koymuştur. Mahendrawathi ve arkadaşları [27] karakteristik olarak yapılandırılmamış müşteri memnuniyet stratejilerinden tipik bir süreç keşfetmiştir. Benzer bir çalışma, Dogan ve arkadaşları [2] tarafından bir çağrı merkezindeki müşteri memnuniyetinin süreç odaklı değerlendirilmesinde yapılmıştır. Sarno ve diğerleri [28] dolandırıcılık tespiti için birliktelik kuralı ve süreç madenciliği algoritmalarını birleştiren hibrit bir yöntem geliştirmiştir. Park ve Kang [29], Hollanda'daki belediye iş süreçlerinin yeniden yapılandırılmasında süreç madenciliğini bir araç olarak kullanmıştır.

Ancak, işe alım süreçlerinde alfa algoritmasının uygulandığı bir çalışmaya rastlanılmamıştır. Bu çalışmada, belirli bir işe alım sürecinde karşılaşılan temel faaliyetlerin oluşturduğu süreç, alfa algoritmasıyla keşfedilmiştir. Alfa algoritmasının adımlarının anlaşılabilir olması için küçük bir olay kaydı kullanılmıştır. Aynı olay kayıtlarında değişiklik yapılarak ikinci bir akış çıkarılmış ve süreç modelleri gösterilmiştir. Böylece olay kayıtlarındaki küçük değişikliklerin süreç modellerinde değişiklik meydana getirebileceği gösterilmiştir.

4 Süreç Keşfi

Bir olay kaydındaki temel gereklilik, örnek olay kodu (Case ID) ve faaliyet bilgisidir. Olay kaydının tip, kaynak, maliyet gibi ek bilgiler içermesi, süreç analizini zenginleştirir ve farklı bakış açılarından bakabilme olanağı sağlar [17], [34]. Tablo 3 bir işe alım sürecinin keşfi için kullanılacak örnek olay kaydını göstermektedir.

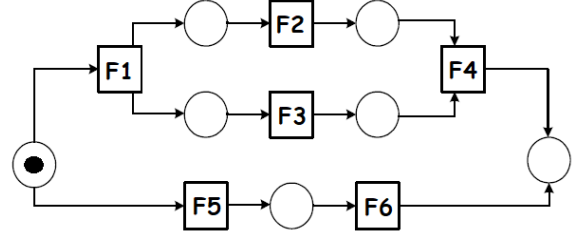
Tablo 3: Örnek Olay Kaydı 1.

Örnek Olay	Kod	Faaliyet
Örnek Olay 1	F1	İş Görüşmesi Ayarla
Örnek Olay 2	F1	İş Görüşmesi Ayarla
Örnek Olay 3	F1	İş Görüşmesi Ayarla
Örnek Olay 3	F2	Adayı Bilgilendir
Örnek Olay 1	F2	Adayı Bilgilendir
Örnek Olay 1	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 2	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 4	F1	İş Görüşmesi Ayarla
Örnek Olay 2	F2	Adayı Bilgilendir
Örnek Olay 2	F4	Adayı Değerlendir
Örnek Olay 5	F5	Olumsuz Değerlendirme
Örnek Olay 4	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 1	F4	Adayı Değerlendir
Örnek Olay 3	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 3	F4	Adayı Değerlendir
Örnek Olay 4	F2	Adayı Bilgilendir
Örnek Olay 5	F6	CV Havuzuna At
Örnek Olay 4	F4	Adayı Değerlendir

Örnek süreç basit olduğundan pratik olarak Şekil 2 elde edilebilir. Beş örnek olay incelendiğinde $F1 - F2 - F3 - F4$ (Örnek Olay 1), $F1 - F3 - F2 - F4$ (Örnek Olay 2), $F1 - F2 - F3 - F4$ (Örnek Olay 3), $F1 - F3 - F2 - F4$ (Örnek Olay 4) ve $F5 - F6$ (Örnek Olay 5) faaliyetlerinin gerçekleştiği söylenebilir. Dolayısıyla, üç temel yol vardır: $F1 - F2 - F3 -$

$F4$, $F1 - F3 - F2 - F4$, ve $F5 - F6$. Olay kaydında $F1 - F2 - F3 - F4$ sırası ve $F1 - F3 - F2 - F4$ sırası iki defa gerçekleşmiştir. Bu yüzden $\langle F1 - F2 - F3 - F4 \rangle^2$ ve $\langle F1 - F3 - F2 - F4 \rangle^2$ olarak gösterilir. Olay kaydı $L = \{ \langle F1 - F2 - F3 - F4 \rangle^2, \langle F1 - F3 - F2 - F4 \rangle^2, \langle F5, F6 \rangle \}$ olarak tanımlanabilir. Bu olay kaydı için aşağıdaki log tabanlı sıralama ilişkileri oluşturulabilir.

$$\begin{aligned} >_L = & \{ (F1, F2), (F1, F3), (F2, F3), (F2, F4), (F3, F2), \\ & (F3, F4), (F5, F6) \} \\ \rightarrow_L = & \{ (F1, F2), (F1, F3), (F2, F4), (F3, F4), (F5, F6) \} \\ \#_L = & \{ (F2, F3) \} \\ ||_L = & \{ (F1, F3), (F3, F2) \} \end{aligned}$$



Şekil 2: Alfa algoritmasıyla keşfedilmiş süreç akışı 1.

Log tabanlı sıralamalar da basit süreçlerin keşfi için yeterli olabilir. Ancak unutulmamalıdır ki, gerçek dünyadaki iş süreçleri sıklıkla çok karmaşıktır. Bu yüzden algoritma adımlarının takip edilmesi önemlidir.

Adım 1'de süreci oluşturan tüm faaliyetler $T_L = \{F1, F2, F3, F4, F5, F6\}$ olarak belirlenir. Adım 2'de ve Adım 3'te sırasıyla süreçteki ile faaliyet kümesi $T_1 = \{(F1, F5)\}$ ve son faaliyet kümesi $T_0 = \{(F4, F6)\}$ olarak bulunur. Adım 4'te boş küme olmayan A ve B gibi iki kümeden oluşan bir X_L kümesi oluşturulur. A ve B kümeleri tek elemanlı kümeler için aralarında nedensellik ilişkisi ($a \rightarrow_L b$) olan; çift elemanlı kümeler için aralarında seçim ilişkisi ($a \#_L b$) bulunan faaliyetlerden oluşur.

$X_L = \{ \{(F1), \{F2\}\}, \{(F1), \{F3\}\}, \{(F2), \{F4\}\}, \{(F3), \{F4\}\}, \{(F5), \{F6\}\} \}$ kümesi elde edilir. Adım 5'te üretilen Y_L kümesi, önceki adımda bulunan X_L kümesini kapsayacak kümeden oluşturulur. Ancak X_L kümesinde iki elemanlı A ve B kümesi olmadığı için daha geniş bir Y_L kümesi üretilmemektedir. Yani $X_L = Y_L$ olarak devam edilir. Adım 6 ile, Y_L kümesine ilk ve son faaliyetin eklenmesiyle oluşturulan P_L yer kümesi elde edilir. Böylece süreçteki yer kümesi olarak

$P_L = \{T_1, T_0, p_{\{(F1), \{F2\}\}}, p_{\{(F1), \{F3\}\}}, p_{\{(F3), \{F4\}\}}, p_{\{(F5), \{F6\}\}} \}$ kümesi bulunur. Adım 7, F_L akışının oluşturulduğu adımdır. Başlangıç ve bitiş faaliyetleri arasındaki yerleri birbirine bağlayarak akışı tamamlar ve Şekil 2 elde edilir.

$$F_L = \left\{ (i_L, F1), (i_L, F5), (F1, p_{\{(F1), \{F2\}\}}), (F1, p_{\{(F1), \{F3\}\}}), (p_{\{(F1), \{F2\}\}}, F2), (p_{\{(F1), \{F3\}\}}, F3), (F2, p_{\{(F2), \{F4\}\}}), (F3, p_{\{(F3), \{F4\}\}}), (p_{\{(F2), \{F4\}\}}, F4), (p_{\{(F3), \{F4\}\}}, F4), (F5, p_{\{(F5), \{F6\}\}}), (p_{\{(F5), \{F6\}\}}, F6), (F4, i_o), (F6, i_o) \right\}$$

Keşfedilen süreç göstermektedir ki her örnek olay ya F1 ya da F5 ile başlar. F1 ile başlayan süreçler, işe alım için pozitif anlam ifade ederken, F5 faaliyeti adayın iş için uygun olmadığını göstermektedir. İş için uygun görülen adaylar belirlendikten sonra hem aday hem de şirket içindeki ekip bilgilendirilmektedir. Ardından, adayın değerlendirilmesiyle süreç tamamlanmaktadır. Dikkat çeken bir husus, ilana çıkılan iş için değerlendirilmeyen adaylar, daha sonraki iş ilanları için

havuza atılmaktadır. Ancak, başvurusu ön incelemede uygun bulunan adaylar için havuza atma işlemi görülmemiştir. Bu da sürecin geliştirilmesi gereken bir taraftır. Uygun bulunan adaylar değerlendirmeye alınmış ancak işe alınmayanların özgeçmişleri, herhangi bir havuzda toplanmamıştır.

Farklı bir süreç olarak Tablo 4'te verilen olay kayıtları incelenebilir. Bu süreçteki temel farklılık, "CV Havuzuna At" faaliyetinin olmamasıdır.

Tablo 4: Örnek Olay Kaydı 2.

Örnek Olay	Kod	Faaliyet
Örnek Olay 1	F1	İş Görüşmesi Ayarla
Örnek Olay 2	F1	İş Görüşmesi Ayarla
Örnek Olay 3	F1	İş Görüşmesi Ayarla
Örnek Olay 3	F2	Adayı Bilgilendir
Örnek Olay 1	F2	Adayı Bilgilendir
Örnek Olay 1	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 2	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 4	F1	İş Görüşmesi Ayarla
Örnek Olay 2	F2	Adayı Bilgilendir
Örnek Olay 2	F4	Adayı Değerlendir
Örnek Olay 5	F1	İş Görüşmesi Ayarla
Örnek Olay 4	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 1	F4	Adayı Değerlendir
Örnek Olay 3	F3	Görüşme Ekibini Bilgilendir
Örnek Olay 3	F4	Adayı Değerlendir
Örnek Olay 4	F2	Adayı Bilgilendir
Örnek Olay 5	F5	Olumsuz Değerlendirme
Örnek Olay 5	F4	Adayı Değerlendir
Örnek Olay 4	F4	Adayı Değerlendir

Algoritma adımının çıktıları aşağıda gösterilmiştir.

Adım 1: $T_L = \{F1, F2, F3, F4, F5\}$

Adım 2: $T_I = \{F1\}$

Adım 3: $T_O = \{F4\}$

Adım 4: $X_L =$

$\left\{ \begin{array}{l} (\{F1\}, \{F2\}), (\{F1\}, \{F3\}), (\{F1\}, \{F5\}), (\{F2\}, \{F4\}), \\ (\{F3\}, \{F4\}), (\{F5\}, \{F4\}), (\{F1\}, \{F2, F5\}), \\ (\{F1\}, \{F3, F5\}), (\{F2, F5\}, \{F4\}), (\{F3, F5\}, \{F4\}) \end{array} \right\}$

Adım 5: $Y_L =$

$\left\{ \begin{array}{l} (\{F1\}, \{F2, F5\}), (\{F1\}, \{F3, F5\}), (\{F2, F5\}, \{F4\}), \\ (\{F3, F5\}, \{F4\}) \end{array} \right\}$

Adım 6: $P_L =$

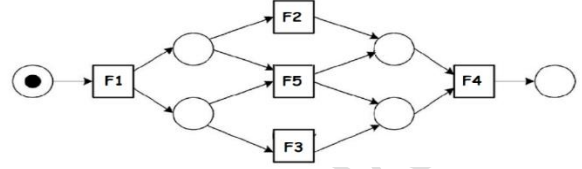
$\{T_I, T_O, P_{(\{F1\}, \{F2, F5\})}, P_{(\{F1\}, \{F3, F5\})}, P_{(\{F2, F5\}, \{F4\})}, P_{(\{F3, F5\}, \{F4\})}\}$

Bu adımın önceki süreç keşfi adımında farklı olduğuna dikkat edilmelidir.

Adım 7: $F_L =$

$\left\{ \begin{array}{l} (i_L, F1), (F1, P_{(\{F1\}, \{F2, F5\})}), (F1, P_{(\{F1\}, \{F3, F5\})}), \\ (P_{(\{F1\}, \{F2, F5\})}, F2), (P_{(\{F1\}, \{F2, F5\})}, F5), (P_{(\{F1\}, \{F3, F5\})}, F3), \\ (P_{(\{F1\}, \{F3, F5\})}, F5), (F2, P_{(\{F2, F5\}, \{F4\})}), (F5, P_{(\{F2, F5\}, \{F4\})}), \\ (F3, P_{(\{F3, F5\}, \{F4\})}), (F5, P_{(\{F3, F5\}, \{F4\})}), (P_{(\{F2, F5\}, \{F4\})}, F4), \\ (P_{(\{F3, F5\}, \{F4\})}, F4), (F4, i_o), \end{array} \right\}$

Böylelikle Şekil 3'te verilen süreç akışı keşfedilmiş olunur.



Şekil 3. Alfa algoritmasıyla keşfedilmiş süreç akışı 2.

Süreç madenciliğinde değerlendirme yapabilmek için bazı durumlarda bir faaliyet analizden çıkarılır. Şekil 2'de keşfedilen süreçte, F6 (CV Havuzuna At) faaliyetiyle ilgili yolunda gitmeyen bir durum bulunmuştu. Bu faaliyet olay kayıtlarından göz ardı edildiğinde, süreç modelinin nasıl değiştiği Şekil 3'te gösterilmiştir. Bu modelde, aday ilk incelemede olumsuz değerlendirilse bile (F5 faaliyeti) CV havuzu olmadığından ekip tarafından diğer ilanlar için değerlendirilmektedir. Böylece süreçlerin ilk faaliyeti her zaman F1, son faaliyeti de her zaman F4 olarak tespit edilmiştir.

5 Sonuç

Süreç madenciliği, süreç yönetimi ve iş analitiğini kullanarak olay kayıtlarından çeşitli keşif algoritmaları tarafından otomatik olarak oluşturulan süreç modelleri aracılığıyla anlaşılabilir bilgileri elde etmeyi amaçlayan bir tekniktir. Bu çalışmada iş akış süreçlerinin yeniden keşfedilme sorununu ele alınmış ve IT sistemlerinde zaten tutulan olay kayıtları ile süreç akışının yeniden keşfedilebileceği bir algoritma tanıtılmıştır. Alfa (α) algoritması, ilk süreç keşif algoritmalarından biridir. Bu çalışmada süreç madenciliğinin genel tanıtımı yapıldıktan sonra, küçük farklılıklarla iş başvuru süreciyle ilgili iki örnek uygulama yapılmıştır. Alfa algoritması adım adım uygulanarak, iki örneğin süreç modeli üretilmiştir. Keşif için geliştirilen algoritmalarından alfa algoritması (α), iki faaliyetin bağlanması olgusunu, yalnızca olay kayıtlarını inceleyerek nedenselliğin tespit edildiği durumlarda kullanmaktadır. Alfa algoritması, kendi içinde oluşturduğu yerler, geçişler ve ağlar ile süreç modelini üretmektedir. Örneklerle, algoritmanın iş akış süreçleri için ilginç analiz sonuçları sağladığı gösterilmiştir. Ancak alfa algoritmasının gürlüğü, seyrek ve tamamlanmamış verilerde ve karmaşık rota yapılarını çözme konusunda yetersiz olduğu unutulmamalıdır. Bu nedenle bazı çalışmalar alfa algoritmasını farklı keşif algoritmalarıyla birlikte kullanmaktadır.

6 Kaynaklar

- [1] Van der Aalst WM, Weijters T, Maruster L. "Workflow mining: Discovering process models from event logs". *IEEE Transactions on Knowledge and Data Engineering*, 16(9), 1128-1142, 2004.
- [2] Dogan O, Ayyar B, Cagil B. "Process-oriented evaluation of customer satisfaction: Process mining application in a call center". *10th International Symposium on Intelligent Manufacturing and Service Systems*, Sakarya, Turkey, 9-11 September 2019.

- [3] Hwang I, Jang YJ. "Process mining to discover shoppers' pathways at a fashion retail store using a wifi-base indoor positioning system". *IEEE Transactions on Automation Science and Engineering*, 14(4), 1786-1792, 2017.
- [4] Dogan O, Bayo-Monton JL, Fernandez-Llatas C, Oztaysi B. "Analyzing of gender behaviors from paths using process mining: A shopping mall application," *Sensors*, 19(3), 557, 2019.
- [5] Van der Aalst WM, "The application of petri nets to workflow management," *Journal of Circuits, Systems and Computers*, 8(1), 21-66, 1998.
- [6] Dogan O, Fernandez-Llatas C, Oztaysi B. *Process mining application for analysis of customer's different visits in a shopping mall*. Editors: Kahraman C, Cebi S, Cevik Onar S, Oztaysi B, Tolga A, Sari I. *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making*, 151-159 Springer, Cham, 2019.
- [7] Valle AM, Santos EA, Loures ER. "Applying process mining techniques in software process appraisals," *Information and Software Technology*, 87, 19-31, 2017.
- [8] Oner M, Oner C, Dogan O. "Internet of things and intelligent systems in digital learning," *International Conference on Engineering and Natural Sciences*, Budapest, Hungary, 1-5 May 2017.
- [9] Oner M, Oner SC, *Data analytics in industry 4.0: In the perspective of big data*. Editors: Yilmaz OF, Tufekci S. *Handbook of Research on Applied Optimization Methodologies in Manufacturing Systems*, 375-392, IGI Global, 2012.
- [10] Van Der Aalst WM. *Process mining: data science in action*, Springer, 2016.
- [11] Van Der Aalst WM. *Process mining: discovery, conformance and enhancement of business processes*, Springer, 2011.
- [12] Van Der Aalst WM, Adriansyah A, De Medeiros AKA, Arcieri F, Baier T, Blickle T, Bose JC, Van Den Brand P, Brandtjen R, Buijs J, vd. "Process mining manifesto" *International Conference on Business Process Management*, Clermont-Ferrand, France, 30 August-2 September 2011.
- [13] Van der Aalst WM, Van Dongen BF, Herbst J, Maruster L, Schimm G, Weijters AJ. "Workflow mining: A survey of issues and approaches," *Data & Knowledge Engineering*, 47(2), 237-267, 2003.
- [14] Weerapong S, Porouhan P, Premchaiswadi W. "Process mining using α -algorithm as a tool (a case study of student registration)". *Tenth International Conference on ICT and Knowledge Engineering*, (Bangkok, Thailand 21-23 November 2012).
- [15] Evermann J, Assadipour G. "Big data meets process mining: Implementing the alpha algorithm with map-reduce," *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, (Gyeongju, Republic of Korea, 24-28 March 2014).
- [16] Dogan O, "Process mining for check-up process analysis," *IIOABJ*, 9(6), 56-61, 2018.
- [17] Fernandez-Llatas C, Lizondo A, Monton E, Benedi JM, Traver V. "Process mining methodology for health process tracking using real-time indoor location systems," *Sensors* 15(12), 29821-29840, 2015.
- [18] Fernandez-Llatas, Carlos, et al. "Analyzing medical emergency processes with process mining: the stroke case." *International Conference on Business Process Management*, Sydney, Australia, 9-14 Eylül 2018.
- [19] Sedrakyian, G, De Weerd, J, Snoeck, M. "Process-mining enabled feedback: "tell me what I did wrong" vs. "tell me how to do it right". *Computers in human behavior* 57, 352-376, 2016.
- [20] De Weerd, J, Schupp, A, Vanderloock, A, Baesens, B. "Process Mining for the multi-faceted analysis of business processes—A case study in a financial services organization", *Computers in Industry* 64(1), 57-67, 2013.
- [21] Li, M, Liu, L, Yin, L, Zhu, Y. "A process mining based approach to knowledge maintenance", *Information Systems Frontiers* 13(3), 371-380, 2011.
- [22] Ou-Yang, C, Winarjo, H. "Petri-net integration—An approach to support multi-agent process mining", *Expert systems with applications*, 38(4), 4039-4051, 2011.
- [23] Park, S, Kang, Y S. "A study of process mining-based business process innovation", *Procedia Computer Science*, 91, 734-743, 2016.
- [24] Mans, R, Reijers, H, van Genuchten, M, Wismeijer, D. (2012, January). Mining processes in dentistry. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, Miami, USA, 28 Ocak 2012.
- [25] Atastina, I, Kurniati, AP. Student registration process evaluation using process mining case study: It telkom. *Ninth International Conference on Digital Information Management*, Bangkok, Thailand, 29 Eylül- 1 Ekim 2014.
- [26] Fernández-Llatas, C, Benedi, JM, García-Gómez, J, Traver, V. "Process mining for individualized behavior modeling using wireless tracking in nursing homes", *Sensors*, 13(11), 15434-15451, 2013.
- [27] Mahendrawathi, ER, Astuti, HM, Nastiti, A. "Analysis of customer fulfilment with process mining: A case study in a telecommunication company", *Procedia Computer Science*, 72, 588-596, 2015.
- [28] Sarno, R, Dewandono, RD, Ahmad, T, Naufal, MF, Sinaga, F. "Hybrid Association Rule Learning and Process Mining for Fraud Detection", *IAENG International Journal of Computer Science*, 42(2), 1-14, 2015
- [29] Park, S, Kang, YS. "A study of process mining-based business process innovation", *Procedia Computer Science*, 91, 734-743, 2016.
- [30] De Leoni, M, van der Aalst, WM, Dees, M. "A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs", *Information Systems*, 56, 235-257, 2016.
- [31] Yoo, S, et al. "Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital", *International journal of medical informatics*, 88, 34-43, 2016.
- [32] Poggi, N, Muthusamy, V, Carrera, D, Khalaf, R. "Business process mining from e-commerce web logs", *International Conference on Business Process Management*, Beijing, China, 26-30 Ağustos 2013.
- [33] Deeva, G. De Weerd, J. "Understanding automated feedback in learning processes by mining local patterns", *International Conference on Business Process Management*, Sydney, Australia, 9-14 Eylül 2018.
- [34] Fernandez-Llatas C, Pileggi SF, Traver V, Benedi JM, "Timed parallel automaton: A mathematical tool for defining highly expressive formal workflows," *Fifth Asia Modelling Symposium*, Kuala Lumpur, Malaysia, 24-26 May 2011.