



Kahramanmaraş Sütçü İmam University

Journal of Engineering Sciences



Geliş Tarihi : 27.12.2023
Kabul Tarihi : 17.02.2024

Received Date : 27.12.2023
Accepted Date : 17.02.2024

ÖNERİ SİSTEMLERİNDE KULLANILAN PERFORMANS METRİKLERİNİN FİLTRELEME TEKNOLOJİLERİNE GÖRE DEĞERLENDİRİLMESİ: İŞ ÖNERİ SİSTEMLERİ ALANI ÜZERİNE BİR ARAŞTIRMA ÇALIŞMASI

EVALUATION OF PERFORMANCE METRICS USED IN RECOMMENDATION SYSTEMS ACCORDING TO FILTERING TECHNOLOGIES: A RESEARCH STUDY ON THE FIELD OF JOB RECOMMENDATION SYSTEMS

*Selin BİTİRİM*¹ (ORCID: 0000-0002-3575-5855)
Duygu ÇELİK ERTUĞRUL^{2*} (ORCID: 0000-0003-1380-705X)

¹Computer Engineering Department, Eastern Mediterranean University, 99628, Famagusta, North Cyprus, via Mersin 10, Turkey

*Sorumlu Yazar / Corresponding Author: Duygu Çelik Ertuğrul, duygu.celik@emu.edu.tr

ÖZET

Tavsiye Sistemleri (Recommendation Systems—RSs) sayesinde hemen hemen her sektörde (ör. e-ticaret, eğitim, eğlence, sağlık, insan kaynakları, reklamcılık, vb.) mevcut süreçlerin/operasyonların etkin bir biçimde yürütülebilmesi ve kullanıcının ilgisini çekebilecek öğelere öncelik verilmesi mümkün hale gelmiştir. RS'lerin katkısı ile, sektörel süreçlerin/hizmetlerin etkin şekilde yönetilmesi ve kullanıcılara kişiselleştirilmiş sonuçlar üretilmesi mümkündür. Bu çalışmada, RS ile ilgili araştırmaların gözden geçirilmesi, filtreleme teknikleri taksonomisinin ortaya çıkarılması ve geniş çapta rastlanan performans metriklerinin tespiti amaçlanmaktadır. Ayrıca, İnsan Kaynakları (İK) yönetiminin olmazsa olmazı olan İş Tavsiye Sistemleri bu çalışmada, araştırma sahası olarak seçilmiş olup performans metriklerinin ve öğe filtreleme yaklaşımlarının belirlenmesi planlanmıştır. RS mimarisi ve çözümleri üzerine, literatürden 2010-2023 yılları arasında yapılmış çeşitli çalışmalar ilgililik durumuna göre seçilmiş ve incelenmiştir. RS'lerde filtreleme teknikleri hiyerarşik olarak sınıflandırılmış ve performans değerlendirmelerinde kullanılan çoğunluk değerlendirme metrikleri saptanarak kategorize edilmiştir. Ayrıca, RS'lerden öğrenilen kazanımların İş Tavsiye Sistemleri'ndeki yansımaları araştırılmış ve İK alanındaki RS çözümleri/metrikleri ortaya konulmuştur. Son olarak, RS çözümleri üzerinde araştırma, geliştirme ve kalite değerlendirmeleri yapmak isteyen araştırmacılara, bu çalışmamız bir yol haritası niteliğindedir.

Anahtar Kelimeler: Tavsiye sistemleri, iş tavsiye sistemleri, taksonomi, filtreleme, performans metrikleri.

ABSTRACT

Thanks to Recommendation Systems (RSs), it has become possible to carry out existing processes/operations effectively in almost every sector (e.g. e-commerce, education, entertainment, healthcare, human resources, advertising, etc.) and to prioritize items that may interest the user. With the contribution of RSs, it is possible to effectively manage sectoral processes/services and produce personalized results for users. This study aims to review RS-related research, reveal a taxonomy of filtering techniques, and identify widely encountered performance metrics. In addition, Job Recommendation Systems, which are indispensable for Human Resources (HR) management, were chosen as the research area in this study and it was planned to determine performance metrics and item filtering approaches. Various studies from the literature on RS architecture and solutions, conducted between 2010 and 2023, were selected according to their relevance and reviewed. Filtering techniques in RSs are classified hierarchically and the majority evaluation metrics used in performance evaluations are determined and categorized. Additionally, the reflections of the gains learned from RSs on Job Recommendation Systems were investigated and RS

ToCite: BİTİRİM, S., & ÇELİK ERTUĞRUL, D., (2023). ÖNERİ SİSTEMLERİNDE KULLANILAN PERFORMANS METRİKLERİNİN FİLTRELEME TEKNOLOJİLERİNE GÖRE DEĞERLENDİRİLMESİ: İŞ ÖNERİ SİSTEMLERİ ALANI ÜZERİNE BİR ARAŞTIRMA ÇALIŞMASI. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 27(3), 706-725.

solutions/metrics in the field of HR were presented. Finally, this study serves as a road map for researchers who want to conduct research, development and quality evaluations on RS solutions.

Keywords: Recommender systems, job recommender systems, taxonomy, filtering, performance metrics.

GİRİŞ

Bilgi teknolojilerinin ve Web'in gelişmesiyle birlikte Tavsiye Sistemleri (Recommendation Systems—RS) hayatımızın önemli bir parçası haline gelmiştir. Özellikle karar alma süreçlerinde önemli bir rol oynayan RS'ler, mühendislik, eğitim, işletme, ekonomi, finans, pazarlama, güvenlik, sağlık ve hukuk gibi geniş bir alanda yer edinmeyi başarmıştır. Ayrıca kurumsal şirketlerde, hızla artan veri yığınları ve karmaşık iş süreçleri nedeniyle, çalışanların iş hayatı, yürüttükleri karmaşık operasyonlar nedeniyle daha zorlu hale gelmiştir. Bu nedenle, RS'lerin kullanımı ile, işletmelerdeki enerji tasarrufu, daha az dikkat dağınıklığı, işçilik maliyetlerinde azalma, bulaş riskinin azalması, çalışanların aileleriyle daha fazla vakit geçirmesi, kadın çalışanların çocuklarına daha iyi bakması, işe gidip gelirken harcanan zamandan tasarruf edilmesi, çalışanların sürekli denetlenmesi, iş özerkliğinin sağlanması vb. avantajları da beraberinde getirmiştir (Al-Habaibeh vd., 2021). Bu avantajların yanında, çevrimiçi e-ticaret ortamlarında (ör. Amazon.com, Netflix, Pandora, vb.) RS'ler, kişilerin bireysel özelliklerine ve ihtiyaçlarına göre, işletmelerin mevcut hizmet/ürünlerini öne çıkarma olanağı sağlar ve kişiye daha kısa sürede ona en uygun ürüne erişme olanağı sunar. Böylece, hem üretici tarafında (ör. doğru öğelerin tavsiye edilmesiyle üreticinin gelirini artırmasına) hem de ürünü alacak tarafta (ör. doğru öğe önerileri ile kişilerin yeni ürün kullanımına teşvik edilmesi, belki de hiç keşfedemeyeceği ürünlere öncelik verilmesi, vb.) olumlu sonuçlar ortaya çıkmaktadır. RS'ler, Makine Öğrenimi (ML), veri madenciliği, kullanıcı modelleme, vakaya dayalı akıl yürütme ve kısıtlama tatmini dahil olmak üzere, çok çeşitli Yapay Zekâ (AI) tekniklerini bünyesinde barındırmaktadır. Literatür incelemeleri sonuçlarına göre, RS çözümlerinde 6 farklı filtreleme türünden bahsedilmiş ve kategorize edilmiştir: (1) İşbirlikçi Filtreleme, (2) İçerik Tabanlı Filtreleme, (3) Hibrit Tabanlı Filtreleme, (4) Bilgi Tabanlı Filtreleme, (5) Bağlam Farkındalığına Dayalı Filtreleme ve (6) Diğerleri (Gunawardana vd., 2022; Palomares vd., 2021; Çelik & Elçi, 2020; Mohamed vd., 2019; Poriya vd., 2014; Lü vd., 2012; Park vd., 2012; Martin vd., 2011; Shani & Gunawardana, 2011). RS'lerin doğruluğunu, başarısını ve verimliliğini değerlendirmek için, geniş bir yelpazede metrik yöntemleri kullanılmıştır. Bu çalışmada, literatürdeki RS metrikleri, dört ana kategoride ayrılmış ve incelenmiştir: (1) Tahmine Dayalı Doğruluk Metrikleri, (2) Sınıflandırmaya Dayalı Doğruluk Metrikleri, (3) Sıralamaya Dayalı Doğruluk Metrikleri ve (4) Doğruluğa Dayalı Olmayan Metrikler olarak kategorize edilmiş ve tartışılmıştır.

Ayrıca, 2020 yılında tüm dünyayı etkisi altına alan COVID-19 salgınından sonra sektördeki birçok şirket uzaktan/evden çalışma ve dönüşümlü çalışma modelleri uygulamaya başlamıştır. Bu durum, İnsan Kaynakları Yönetimi (Human Resources Management—HRM) sahasında yürütülen süreçleri de etkilemiş ve HRM alanında çalışanların karar alma ve süreç yönetimi aşamalarında hızlı gelişen RS teknolojileri büyük bir fırsat haline gelmiştir (Vrontis vd., 2022). Bu nedenle, bu çalışmanın son katkısı olarak, JRS uygulamaları bir araştırma sahası olarak seçilmiş ve JRS çözümlerinde kullanılan öğe tavsiye sürecindeki filtreleme teknikleri ve JRS'lerde etkin performans ve kalite değerlendirme metriklerinin belirlenmesi sağlanmıştır.

Çalışmanın geri kalanı şu şekilde ilerleyecektir: Bölüm 2, RS'lerin arka planını sunmaktadır. Bölüm 3, RS'ler de kullanılan performans değerlendirme metriklerine göre kategorize eder, özetler ve sınıflandırır. Bölüm 4, JRS öğe tavsiye sürecindeki filtreleme teknikleri ve performans/kalite metrikleri üzerine yapılan çalışmalar sunulmaktadır. Bölüm 5'te çalışmadan elde edilen sonuçlar sunulmuştur.

TAVSİYE SİSTEMLERİNİN MİMARİSİ

Bu bölümde, RS'lerin tanımları, genel olarak uygulanan işlem adımları, öğe tavsiye oluşturma sürecindeki filtreleme teknikleri ve hiyerarşik sınıflandırma sonuçları sunulmaktadır. RS'ler, kullanıcının öğe seçeneklerine karşı eylemini tahmin etmeyi amaçlayan yazılım ürünleridir. Teknolojinin ve Web'in gelişmesiyle birlikte, İnternet ortamındaki Web verilerinin hacmi ve tüketimi artmış, bu da RS çözümlerinin ortaya çıkmasına neden olmuştur. RS'ler ilk 1970'ler de tasarlanmış (Goldberg vd., 1992; Negroponte, 1970) ve ilk kez 1990'ların ortalarında uygulanmaya ve ticarileştirilmeye başlanmıştır (Martin vd., 2011). RS'ler, kullanıcıların profil verilerini, demografik/kültürel özelliklerini, geçmiş/mevcut tercihlerini vb. dikkate aldığından, kullanıcının istediği hizmete/ürüne daha hızlı ulaşmasını sağlayabilmektedirler. RS'ler hem tüketiciler hem de şirketler için, kişiselleştirmeye olanak sağladığı ve süreç yönetiminde zaman tasarrufu sağladığı için günümüzde çok önemli yazılım çözümleridir. RS'ler özünde bir

arama motoru gibi davranır ve kullanıcıların tam olarak ne istediğini bilmesine gerek kalmadan, kullanıcıya en uygun tavsiyeleri sunabilir (Birjali vd., 2021). Fakat, kullanıcıların profilleri genel olarak değişmese de, zaman içinde kullanıcı tercihleri ve zevkleri değişebileceğinden, bireyleri değerlendirmek zorlaşabilmektedir. Bu zorlukla birlikte, RS'ler zaman içinde gelişmeye ve farklı endüstrilerde kullanımları popüler hale gelmeye başladığını gözlemlemekteyiz (Wu vd., 2022; Marlowe, 2021; Çelik & Elçi, 2020; Beel vd., 2016; Isinkaye vd., 2015; Lu vd., 2015; Bobadilla vd., 2013; Lü vd., 2012; Park vd., 2012; Adomavicius & Tuzhilin, 2011; Martin vd., 2011). Bu nedenle, RS'lerin zaman içindeki evrimi, güncel AI çözümlerinin RS'lerin temel bileşenlerine entegre edilmesi/uygulanmaya başlanmasıyla sağlanır (Portugal vd., 2018). RS'lerde temel süreçler aşağıda sunulmuştur:

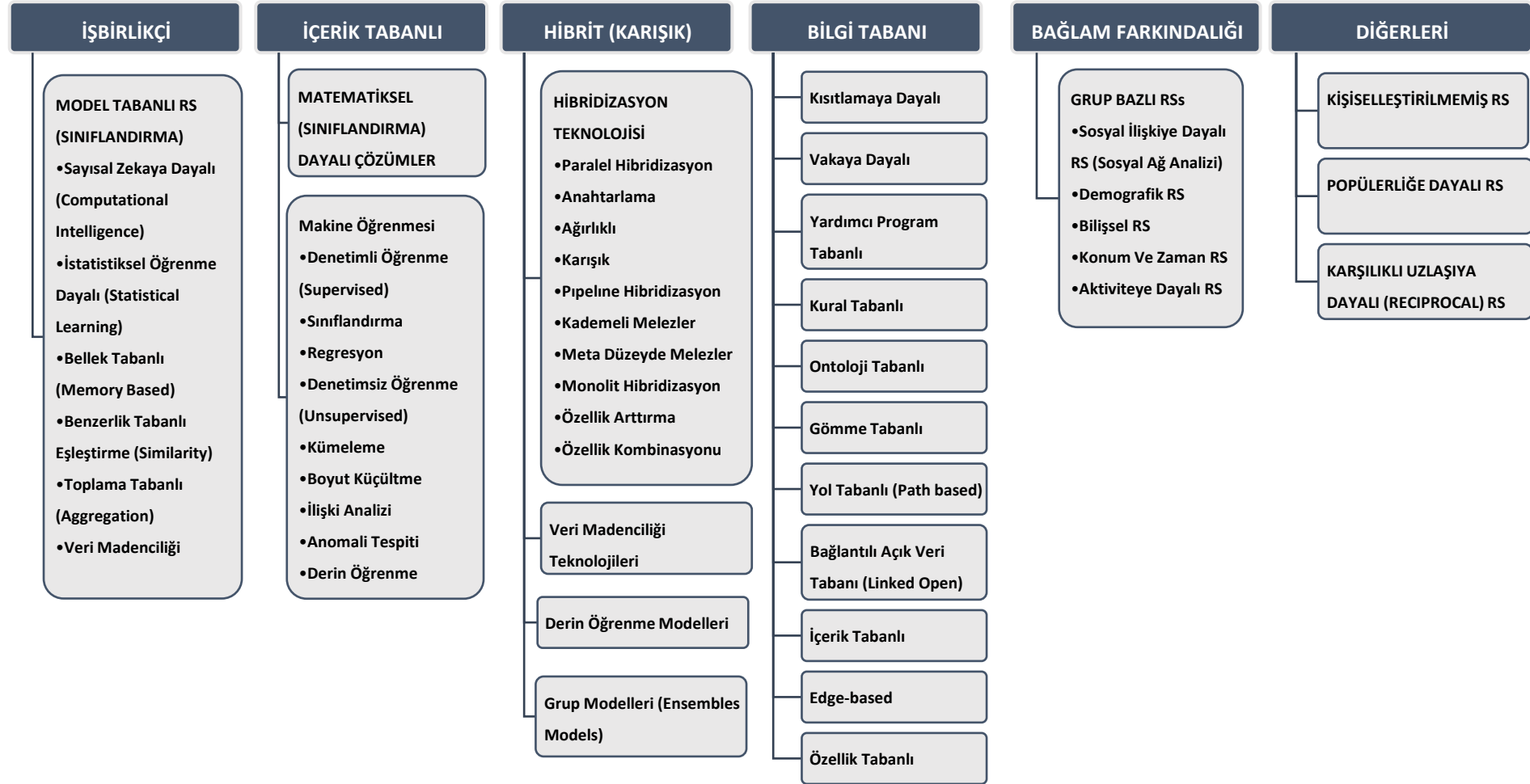
Veri Toplama: Kullanıcılar tarafından yapılan her işlem/her hareket RS'ler için birer veridir. Örneğin, kullanıcının arama motorunun kullanarak, ne tür ürünleri araştırdığı ve tıkladığı/tıklamadığı, ne aradığı, bir ürünü ne kadar süre görüntülediği, ürün yorumları, alışveriş sepetine eklenen ürünler, iade edilen ürünler, çalınan/yarıda kesilen müzikler, izlenen/izlenmeyen veya kesintiye uğrayan filmler, hatta hava durumu, vb. her detay, RS'ler için önemli bir kullanıcı verisidir.

Depolama: Toplanan veriler daha sonra kullanıcılara bir dizi tavsiye olarak sunulmak üzere saklanır.

Analiz: AI veya ML güncel metotları ve algoritmaları kullanılarak, kullanıcıların tüketebileceği bir dizi uygun hizmet/ürün çıkarılması sağlanır (Portugal vd., 2018). Örneğin, ML tabanlı RS çözümlerinde, bir veri kümesi üzerindeki kullanıcı tercih bilgileri analiz edilir ve etkili veri modelleri elde edilerek kullanıcıların tercihleri tahmin edilmeye çalışılır. Sistem daha sonra talepler ve çıkarlar arasındaki korelasyona dayalı olarak en alakalı veya faydalı öğeleri tavsiye olarak üretebilir (Portugal vd., 2018; Isinkaye vd., 2015).

Filtreleme: Bu bölümde, RS'nin farklı alanlardaki ana uygulamalarını analiz etmek ve anlamak için, RS çözümlerinde öğe tavsiye sürecinde uygulanan filtreleme teknikleri hakkında sırasıyla bilgi verilmektedir. Literatür incelemeleri sonucunda, RS çözümlerinde kişiye özgü öğe tavsiye süreci için çeşitli filtreleme teknikleri tespit edilmiş ve 6 farklı kategoriye ayrılmıştır: İşbirlikçi Filtreleme, İçerik Tabanlı Filtreleme, Hibrit Tabanlı Filtreleme, Bilgi Tabanlı Filtreleme, Bağlam Farkındalığına Dayalı Filtreleme, ve Diğerleri (Gunawardana vd., 2022; Palomares vd., 2021; Çelik & Elçi, 2020; Mohamed vd., 2019; Al-Shamri, 2016; Poriya vd., 2014; Lü vd., 2012; Martin vd., 2011; Park vd., 2012; Shani & Gunawardana, 2011).

Tespit edilen bu 6 filtreleme teknikleri Şekil 1'de gösterilmiş ve bir taksonomi formatında aşağıda sunulmuştur.



Şekil 1. Önerme Sistemlerinde (Recommender Systems-RS) Uygulanan Öğe Filtreleme ve Geliştirme Metodolojilerin Taksonomisi.

İşbirlikçi Filtreleme: İşbirliğine dayalı filtreleme tekniğinde, kullanıcıya belirli bir kullanıcının davranışına ve diğer benzer kullanıcı davranışlarına göre uygun öğeler tavsiye edilir. Bu teknikte, kullanıcının tercihlerine göre veri kümeleri oluşturulur ve bu öneriler kullanıcılara sunulur. İşbirliğine dayalı filtreleme algoritmaları bellek tabanlı ve model tabanlı olmak üzere 2 tipte sınıflandırılmaktadır.

İçerik Tabanlı Filtreleme: İçerik bazlı filtreleme tekniğinde, öneriler, yalnızca tek bir kullanıcının davranış verilerinin analiz edilmesi sonucunda üretilir. Bu filtreleme tekniğinde, öğe özellik tanımlamaları ve kullanıcının profili önemli bir rol oynamaktadır. İçerik tabanlı tavsiye sistemlerinin, temel kavramları şu şekilde ifade edilebilir: (1) Belirli tek bir kullanıcının tercih ettiği öğelerin açıklamaları analiz edilerek, bu öğelere işaret edebilecek kullanıcı tercihlerine odaklanılır. Bu tercihler, kullanıcıların profil verileri olarak saklanır. (2) Her öğenin özelliklerini, kullanıcı profilleri ile karşılaştırılır. Böylece yalnızca kullanıcı profili ile yüksek derecede benzerliğe sahip olan, ilgili öğeler söz konusu kullanıcıya tavsiye edilecektir. Yapılan araştırmalara göre, RS çalışmalarında, Şekil 1'de görüldüğü gibi, içerik tabanlı filtreleme tekniklerinde genelde ML'ne dayalı; sınıflandırma, regresyon, denetimsiz/denetimli öğrenme, kümeleme, derin öğrenme gibi çeşitli teknikler uygulanmıştır.

Hibrit Tabanlı Filtreleme: Hibrit filtreleme tekniği birden fazla öneri tekniklerini birleşmesiyle ortaya çıkmıştır. Amaç; tek bir algoritmaya göre daha iyi ve daha etkili tavsiyeler verecek olan iki farklı tekniği birleştirmektir. Çünkü bir algoritmanın dezavantajı diğer bir algoritma tarafından aşılması mümkün olabilmektedir (Isinkaye vd., 2015).

Bilgi Tabanlı Filtreleme: Bilgi tabanlı filtreleme tekniği, kullanıcılar ve ürünler hakkındaki bilgiyi, bir öneri oluşturmak için bilgiye dayalı bir yaklaşım izlemek ve hangi ürünlerin kullanıcının gereksinimlerini karşıladığını düşünmek için kullanan sistemdir. Bilgi tabanlı RS'ler genelde, kullanıcının derecelendirme geçmişine değil, kullanıcı sorgularına ve bir sistemin etki alanı bilgi bankasına dayalı öneriler sağlar. Kullanıcılarından bir dizi kullanıcı yönergesi, kural, kısıtlama veya bir öğe örneği gibi bazı bilgiler isteyebilir. Sistem, daha sonra öğe veri tabanında arama yapar ve sonuç olarak kullanıcıya en uygun öğeleri sonuç olarak döndürür (Çano & Morisio, 2017; Burke vd., 2011). Yapılan araştırmalar sonucunda, RS çalışmalarında bilgi tabanlı filtreleme, kısıtlamaya/vakaya dayalı, kural tabanlı (ör: ontolojiler), gömme tabanlı, içerik/özellik tabanlı gibi türlere ayrılarak kullanılmıştır.

Bağlam Farkındalığına Dayalı Filtreleme: Bağlam farkındalığına dayalı filtreleme teknikleri, sosyal ilişkiye dayalı, demografik, bilişsel, konum/zaman ve aktiviteye dayalı olmak üzere çeşitli filtreleme tekniklerine ayrılmaktadır. RS çalışmalarından en sık uygulanan teknik demografik filtreleme yöntemidir. Demografik filtremeye dayalı RS'ler ile kullanıcıların demografik bilgileri analiz edilir, sınıflandırılır ve buna göre önermeler oluşturulur. Daha doğrusu, tüketicilerin demografik özelliklerine dayanan ve demografik olarak hedef tüketiciye benzeyen tüketicilerden iyi geri bildirimler alan öğelerin bir listesini öneren tavsiye sistemleridir. Yeni bir kullanıcı için, önce hangi kategoriye girdiği tespit edilir ve önerilerde bulunur. Ardından önceki benzer kullanıcıların kümülatif satın alma tercihleri ait olduğu kümeye uygulanmaktadır. İşbirlikçi teknikler gibi, demografik teknikler de "insandan insana" korelasyonlar oluşturur, ancak farklı veriler kullanır. İşbirlikçi ve içerik tabanlı bir RS, demografik yaklaşımın gerektirdiği türden olmayan bir kullanıcı derecelendirme geçmişi gerektirir (Çano & Morisio, 2017; Al-Shamri, 2016).

Diğer: RS'lerde uygulanan diğer filtreleme teknikleri; kişiselleştirilmiş, popülerliğe dayalı ve iki yönlü karşılıklı uzlaşıya dayalı çeşitli teknikler tespit edilmiştir. Popülerite bazlı RS'ler, kullanıcı davranışlarını analiz etmez, diğer bir deyişle kullanıcının kişisel tercihlerine dayanmayan bir sistemdir. Tercih edilen, popüler/trend, yaygın olarak tüketilen ürünleri önerir.

Kullanıcının ilgi alanlarını genişletmek ve uygulama tercihlerini popüler hale getirmek için bir fırsat olabilir (Kumar vd., 2014). Kişiselleştirilmemiş RS'lerde kullanıcılar veya geçmiş tercihleri hakkında herhangi bir şey bilmeye gerek yoktur. Bu nedenle, RS Web uygulamalarının geliştirilmesi ve güncellenmesi kolaydır. Kişiselleştirilmemiş RS'ler, popülerlik tabanlı RS'lere oldukça benzemektedir. Kişiselleştirilmiş bir RS, bir kullanıcının geçmiş etkileşimlerine (ör. önceki satın alma geçmişine) dayalı olarak öğeler önerirken, kişiselleştirilmemiş bir RS, o zaman diliminde sistem kullanıcıları arasında popüler olan öğeleri önerir (Poriya vd., 2014). Literatürde, RS üzerine yapılan çalışmalarda, bahsi geçen 3 teknik arasından en sık uygulanan; iki yönlü ve karşılıklı uzlaşıya dayalı (Reciprocal Recommender Systems—RRS) filtreleme metodudur.

Geleneksel RS'ler de sistem kullanıcılarına önerilen öğeler cansız varlıklar olup, önerinin başarısı sistem tarafından önerilen öğelere yalnızca son kullanıcının verdiği tepkiyle belirlenmektedir. Ancak, karşılıklı öneri sistemlerinde (iki

yönlü ve uzlaşmaya dayalı RS'ler de) bu durum tam tersidir; kullanıcılar, diğer kullanıcılara tavsiye edilen öneriler haline gelir (Palomares vd., 2021). RRS çözümleri, arkadaş/eş bulma, iş bulma, doktor bulma gibi çeşitli eşleştirme Web sitelerinde çevrimiçi hizmetlerde daha yaygın uygulanmaktadır. RRS'ler de bir önerinin başarılı sayılabilmesi için, her iki kullanıcının görüşleri veya memnuniyetleri dikkate alınır. Başka bir deyişle, bir RRS'nin başarılı sayılması için, her iki tarafın da sunulan tekliften memnun olması ve uzlaşmış olması beklenir.

TAVSIYE SİSTEMLERİNDE KULLANILAN PERFORMANS DEĞERLENDİRME METRİKLERİ

Literatürde, RS'lerin başarısını ve verimliliğini değerlendirmek için çeşitli ölçüm yöntemleri ve metrikler kullanılmıştır. Literatür taramalarından, toplamda 30 adet metrik tespit edilmiş ve bu metrikler dört ana kategoride sınıflandırılmıştır: (1) Tahmine Dayalı Doğruluk Metrikleri (7 metrik), (2) Sınıflandırmaya Dayalı Doğruluk Metrikleri (10 metrik), (3) Sıralamaya Dayalı Doğruluk Metrikleri (8 metrik) ve (4) Doğruluğa Dayalı Olmayan Metrikler (5 metrik) (Gunawardana vd., 2022; Vijaysinh, 2021; Chaaya vd., 2017; Gunawardana & Shani, 2009). Bu metrikler aşağıda detaylandırılmıştır.

Tahmine Dayalı Doğruluk Metrikleri

Tahmine dayalı doğruluk metrikleri, bir tavsiye veren tahmini puanlarının gerçek kullanıcı puanlarına ne kadar yakın olduğu konusunu ele alan derecelendirme tahmin metrikleri olarak bilinir. Bu ölçü, ikili olmayan derecelendirme ölçümlerini değerlendirmek için yaygın olarak kullanılmaktadır (Vijaysinh, 2021). Tüm öğeler için derecelendirmeleri doğru bir şekilde tahmin etmenin gerekli olduğu durumlar için uygun bir yaklaşımdır.

RS'ler de yaygın olarak kullanılan ve en popüler tahmine dayalı doğruluk metrikleri; Ortalama Karesel Hata, Ortalama Hata Kare Kökü, Ortalama Mutlak Hata, Normalleştirilmiş Ortalama Mutlak Hata, R Kare ve Açıklanan Varyans'tır. Bunlara ek olarak, tavsiye sistemlerinde yapılan eşleştirme görevleri sırasında tahmin için kullanılan birkaç benzerlik tekniğinin olduğunu söylemek mümkündür. Benzerlik teknikleri ve en popüler tahmine dayalı RS doğruluk performans metrikleri aşağıda açıklanmıştır.

Ortalama Karesel Hata (Mean Squared Error—MSE): MSE, en popüler metriklerden biridir. MSE, gerçek oran ile tahmin edilen oran arasındaki farkı (hatayı) hesaplayan, karesini alan ve ardından tüm hataların ortalamasını sağlayan basit bir ölçümdür (Örneğin, bir iş ilanında bir iş adayının beklenen yetkinlik gereklilikleriyle, yetkinliklerinin tahmini hatası).

$$MSE = \frac{1}{n} \sum_{i=1}^n (R_i - P_i)^2 \quad (1)$$

Ortalama Hata Kare Kökü (Root Mean Squared Error—RMSE): Gerçek derecelendirme ile tahmin edilen derecelendirme arasındaki standart sapmaya dayanır. RMSE, MSE'nin köküdür ve hataların ölçeğini gerçek değerlere yaklaştırarak daha yorumlanabilir hale getirmeye yardımcı olduğu için faydalıdır. Hatalar toplanmadan önce karesini alınır, böylece büyük hatalara daha fazla önem verir. Normalleştirilmiş RMSE, değerlendirme hesaplamaları için kullanılabilir (Gunawardana vd., 2022; Vijaysinh, 2021; Chaaya vd., 2017).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (R_i - P_i)^2}{n}} \quad Normalized(RMSE) = \frac{RMSE}{R_{max} - P_{min}} \quad (2)$$

Burada, R_i —gerçek verilerdir, P_i —tahmin edilen verilerdir, n tahminlerin sayısıdır.

Ortalama Mutlak Hata (Mean Absolute Error—MAE): MAE, RMSE'ye benzer, ancak ortalamanın karesini ve kökünü almak yerine mutlak değer kullanır. MAE, tahmin edilen derecelendirme ile kullanıcının gerçek derecelendirmesi arasındaki ortalama mutlak sapmayı ölçer. Başka bir deyişle, MAE, tahmin edilen değerler ile gerçek değerler arasındaki mutlak fark alınarak ve ardından veri kümesi genelinde ortalaması alınarak hesaplanır. RMSE, MAE'ye kıyasla büyük hatalar için daha büyük cezalar uygular. MAE ve RMSE'nin farklılaştırılmış ve normalleştirilmiş versiyonları bulunmaktadır (Chaaya vd., 2017; Gunawardana & Shani, 2009).

$$MAE = \frac{\sum_{i=1}^n |R_i - P_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n} \quad (3)$$

Ortalama Mutlak Kullanıcı Hatası (Mean Absolute User Error—MAUE): Her bir kullanıcı için MEA bağımsız olarak hesaplanır ve ardından tüm MEA'ların ortalaması alınır. Bu yöntem, MAE'nin hesaplanmasında ortaya çıkan olası önyargıları çözmektedir. Böylece, hesaplama sırasında tüm kullanıcılar aynı ağırlığa sahip olacaktır (Chaaya vd., 2017; Gunawardana & Shani, 2009).

R Kare (R Squared—R²): Regresyon işlemi sırasında, uygulanan modelin performansını değerlendirmesinde kullanılan istatistiksel bir ölçüttür (Fkih, 2022). Bu nedenle, daha yüksek bir R² değeri, modelin uygun olduğunu gösterirken, daha düşük bir R² değeri ise modelin uygun olmadığını gösterir. Bu nedenle, R² bir hata ölçüsü olarak kabul edilir. Artıkların toplamının ortalamaların toplamına oranı ne kadar küçükse, R² o kadar yüksek olacaktır ($R^2 = SSR / SST$, burada $SSR = \{Artıkların\ Toplamı\}$, $SST = \{Ortalamaların\ Toplamı\}$). Daha düşük R² değeri, modelin daha hatalı tahminler ürettiği anlamına gelir, bu da daha az doğru tahminlerin üretildiği anlamına gelir. Başka bir deyişle, R² değeri çok düşükse, sistemin tahminleri daha az kullanışlı ve belirsiz tahminlerdir.

Açıklanmış Varyans (Explained Variance—EV): Verilerdeki varyansın ne kadarının model tarafından açıklandığını gösterir. Tahmin edilen hataların varyansı ve gerçek değerlerin varyansına oranı alındıktan sonra 1'in çıkarılmasıdır ($EV = 1 - Var(Tahmin\ Hataları) / Var(Gerçek\ Değerler)$). 1.0'a yakın puanlar oldukça arzu edilir, bu da hataların standart sapmalarının daha iyi karelerini gösterir (Mathur vd., 2019).

Benzerlik: Benzerlik ölçüm yöntemleri, kullanıcıdan kullanıcıya veya öğeden öğeye iş birliğine dayalı filtreleme RS'lerin de yaygın olarak kullanılır. Genel olarak, tüm yöntemler, bir çift kullanıcı (veya öğe) arasındaki benzerliği hesaplarken yaygın olarak derecelendirilen öğelerin derecelendirmelerini kullanır. RS'ler, aktif bir kullanıcı/öge için benzer kullanıcıları/öğeleri bulmak için Kosinüs Benzerliği, Pearson Korelasyon Katsayısı, Jaccard Benzerliği, Tanimato vb. gibi çeşitli benzerlik tekniklerini kullanır (Ayub vd., 2018; Mihalcea vd., 2006).

Kosinüs Benzerliği (Cosine Similarity): Kosinüs benzerliği, vektörler arası benzerlik metriği ya da iki vektör arasındaki açının kosinüs değerinin hesaplanmasıyla vektörler arası benzerliği ölçmeye yarayan bir mesafe ölçüsü olarak tanımlanabilir (Salton & Buckley, 1997; Salton vd., 1997; Salton & Lesk, 1971) Kosinüs benzerliği hesaplanırken, yalnızca pozitif derecelendirmelerin bir rolü vardır ve negatif derecelendirmeler atılır. Çıkış değeri 0-1 arasında değişir. 1'e yakın bir değer, iki değişken arasında güçlü bir korelasyon olduğunu gösterir. 0'a yakın bir değer, korelasyon olmadığını gösterir.

Pearson Korelasyon Katsayısı (Pearson Correlation Coefficient—PCC): PCC, iki kullanıcının ne kadar ilişkili olduğunu, yani ikisi arasındaki doğrusal korelasyonu ölçer. Bellek tabanlı iş birlikçi filtrelemeye dayalı bir yöntem olarak bu algoritma, kullanıcılar tarafından derecelendirilen öğe çiftleri arasındaki veya aynı öğeyi derecelendiren kullanıcı çiftleri arasındaki benzerliği hesaplayarak ölçeklenebilirlik sorununun üstesinden gelir. PCC, ilgilenilen her öğe için tahmini bir puan hesaplayan ve bu tahmin için özel olarak tasarlanmış bir yöntemdir. Bununla birlikte, bu yöntem literatürde tavsiye işlemi için de kullanılmaktadır (Gunawardana & Shani, 2009). Bu durumda, her olası madde için bir puan tahmin edilerek ve daha sonra tahmin edilen puanlar azaltılarak maddeler sıralanarak yapılır (Lü vd., 2012; Gunawardana & Shani, 2009).

Jaccard Benzerliği: Jaccard benzerliği, öğe sayısını sayarak iki küme arasındaki benzerliği ölçmenin bir fonksiyonudur. Bir RS tarafından döndürülen tavsiyeler arasında bir benzerlik ölçüsü tanımlayabilir. Döndürülen değer; A ve B kümelerinin benzerliği ile A ve B'nin kesişim boyutunun birleşim boyutlarına oranıdır (Ayub vd., 2018).

Tanimoto Benzerliği: Tanimoto ölçüsü, Jaccard ölçüsünün geliştirilmiş bir versiyonu gibidir. Kullanıcı tarafından derecelendirilen ortak öğelerin sayısını alır ve kullanıcılar tarafından sağlanan gerçek derecelendirme değerini yok sayar. Kullanıcılar tarafından derecelendirilen birkaç öğe ile kullanıcılardan herhangi biri tarafından derecelendirilen toplam öğe sayısı arasındaki oranı alır (Selvi & Sivasankar, 2018).

Sınıflandırmaya Dayalı Doğruluk Metrikleri

Doğruluk (Accuracy): Doğruluk, bir miktarın gerçek değerine ne kadar yakın olduğunu gösterir. Bir ML modelinin doğruluğunu ölçer ve doğru sınıflandırılmış örneklerin yüzdesini ifade eder (Chaaya vd., 2017; Gunawardana & Shani, 2009).

TN: Doğru sınıflandırılmış negatif öğe sayısı FN: Yanlış bir şekilde negatif olarak sınıflandırılan pozitif öğe sayısı (4)

TP: Doğru sınıflandırılmış pozitif öge sayısı *FP: Doğru şekilde pozitif olarak sınıflandırılan negatif öge sayısı*

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Duyarlılık (Precision): Doğru tahmin edilen pozitif ögelerin tahmin edilen tüm pozitif ögeler arasındaki oranını temsil eder (Chaaya vd, 2017; Gunawardana & Shani, 2011).

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Geri Çağırma (Recall): Doğru tahmin edilen pozitif ögelerin sayısının, gerçek pozitif ögelerin toplam sayısına bölünmesiyle elde edilen orandır (Kang vd., 2019; Gunawardana & Shani, 2009).

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Ters Duyarlılık (Inverse Precision): Bu metrik, önerilmeyen bir ögenin aslında alakasız olma olasılığını verir. Gerçekte alakasız ve tavsiye edilmeyen bir ögenin, tavsiye edilmeyen tüm ögelerin toplam sayısına oranı ile bulunur (Bhat vd., 2022; Chaaya vd., 2017).

Ters Geri Çağırma (Inverse Recall): Bu metrik, alakasız bir ögenin aslında önerilmeme olasılığını verir. Gerçekten alakasız ve tavsiye edilmeyen maddelerin toplam alakasız ögelere oranı ile hesaplanır (Bhat vd., 2022; Schröder vd., 2011).

Alıcı Çalışma Özellikleri (Receiver Operating Characteristics—ROC): Precision ve Recall arasındaki ayırım göz önünde bulundurularak doğrudan değerlendirilebilir. ROC, bir ML modelinin pozitif ve negatif sınıflarını ayırt etme yeteneğini ifade eder. ROC eğrisi yöntemi bu değerlendirme için idealdir (Shani & Gunawardana, 2011). Precision ve Recall çoğu zaman istenen sonucu vermeyebilir. ROC eğrisi ile Precision ve Recall yöntemlerini kullanmak daha doğru olabilir, ancak her ikisini de ayrı ayrı kullanmak uygun olabilir. Örneğin, bir sistemde kullanıcıya kaç haber önermeniz gerektiği üzerinde çalışıyorsanız, eşik değerini bulmak için bir grafik oluşturmanız gerekir. Bu durumda ROC eğrisinin kullanılması daha uygun olabilir (Althbiti vd., 2021; Isinkaye vd., 2015).

F1-Ölçümü (F1-Score): ROC eğrisi ile Precision ve Recall kullanılacaksa, F1-Ölçümü'nün kullanılması uygun olacaktır. Accuracy ve F1-Score ayrı bir metrik olarak uygulanabilir (Kang vd., 2019; Chaaya vd., 2017).

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

Lojistik Kayıp (Logistic Loss, Log-Loss): Log-Loss, ikili sınıflandırıcılar arasında iyi bilinen bir değerlendirme ölçüsüdür. Bir sınıflandırıcının tahminleri verilen gerçek etiketlerin negatif günlük olasılığı olarak bilinir. Log-Loss bazen Lojistik regresyon ve Sinir Ağları durumunda da optimizasyon hedefidir. Ayrıca, tahminin belirsizliğini, gerçek verilerden ne kadar farklı olduğuna bağlı olarak dikkate alır. Bu bize modelimizin performansı hakkında daha detaylı bir görüş sağlar (Painsky & Wornell, 2018; Xin vd., 2016).

Eğri Altındaki Alan (Area Under Curve—AUC): AUC, alıcı çalışma karakteristik eğrisinin altındaki ayrılmaz alandır (Chaaya vd., 2017; Liu vd., 2016; Xin vd., 2016).

Özgüllük (Specificity—SPE): Bu metrik, gerçek negatif orana atıfta bulunarak, doğru tahmin edilen gerçek negatif ögeleri gösterir (Wu vd., 2016; Xin vd., 2016; Gunawardana & Shani, 2011; Gunawardana & Shani, 2009).

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

Sıralamaya Dayalı Doğruluk Metrikleri

İstatistikte, sıralama doğruluğu veya sıralama tahmini metriği olarak kullanılmaktadır. Sistemdeki her kullanıcı için RS, sıralı bir tavsiye listesi oluşturur. RS'ler deki önerilen ögeler arasından, Precision (Duyarlılık) gibi geleneksel sınıflandırma metriklerini kullanarak, kaç tanesinin ilgili bir öneri olduğu bulunabilir. Ancak, Precision, listedeki

öğelerin doğru sırasını veya sıralamasını yakalayamaz. Kullanıcıların sınırlı bir dikkat süresi vardır ve en üstte doğru öğeleri tavsiye etmek çok önemlidir. Bu nedenle, tavsiye sıralamasının korelasyonu ölçülerek, kullanıcının tercihlerine göre önerilen öğelerin doğru sırası tahmin edilebilir (Vijaysinh, 2021; Al-Otaibi & Ykhlef, 2017; McFee & Lanckriet, 2010;).

Duyarlılık @ k (Precision@k veya P(k)): "En iyi k tane madde arasında önerilen ilgili öğelerin sayısı" anlamına gelen popüler bir metriktir. Örneğin, bir arama motorunda "plastik cerrahi" araması yaparsanız ve ilk sayfadaki 10 bağlantıdan 7'si "plastik cerrahi" ile ilgiliyse, bu sorgu için P@10 0,7'ye eşittir (Shervin, 2020).

$$Precision@k = \frac{(\# \text{ önerilen öğeler @k alakalı olanlar})}{(\# \text{ tavsiye edilen öğeler @k})} \quad (9)$$

Ortalama Duyarlılık (Average Precision—AP): 0 ile 1 arasındaki geri çağırma değeri için ortalama hassasiyet değerini hesaplar. Formülde, k: alınan öğelerin konumundaki sıralamadır, n: seçilen öğelerin sayısıdır, P(k): yukarıda açıklandığı gibi, listedeki k sırasındaki kesinlik değeridir, rel(k): bir gösterge fonksiyonudur, k derecesindeki öğe ilgili bir öğe ise 1'e eşittir, aksi takdirde 0 olacaktır. N_r: ilgili öğelerin sayısıdır (Shervin, 2020; Shakirova, 2017).

$$AP = \frac{\sum_{k=1}^n P(k) \cdot rel(k)}{N_r} \quad (10)$$

Mutlak Ortalama Duyarlılık (Mean Average Precision—MAP): MAP, tüm kullanıcılar üzerinde normalleştirilmiş her kullanıcı için ortalama duyarlılıktır. Önerilen k tane madde varsa, P(k)'nin ortalaması k=1... K, tüm veri noktalarının ortalamasını almaktadır (Shervin, 2020; Chaaya vd., 2017; Shakirova, 2017).

$$MAP = \frac{\sum_{i=1}^I AP_i}{I} \quad (11)$$

burada, I toplam öğe sayısıdır.

İsabet Oranı (Hit Ratio—HR): Bir RS'de isabet oranı, L uzunluğundaki öneri listesine, doğru öğe(ler)in dahil edildiği kullanıcıların oranıdır. Formüldeki L ne kadar büyükse, isabet oranının o kadar yüksek olduğu anlamına gelir. Çünkü doğru madde(ler)in öneri listesine alınma şansı daha yüksektir. Bu nedenle, L için uygun bir değer seçmek önemli bir adımdır (Zhang & Feng, 2016).

$$HR = \frac{|U_{hit}^L|}{|U_{all}|} \quad (12)$$

burada, $|U_{hit}^L|$, en üstteki L öneri listesinde doğru cevabın yer aldığı kullanıcıların sayısıdır; $|U_{all}|$, test veri kümesindeki toplam kullanıcı sayısıdır.

Ortalama Karşılıklı Sıralama (Mean Reciprocal Rank—MRR): İstatistik tabanlı bir metriktir ve Ortalama Karşılıklı İsabet Oranı (Average Reciprocal Hit Ratio—ARHR) olarak da bilinir. MRR, bir öneri listesinde keşfedilen ilk ilgili öğenin konumunu ölçer. Sonuçta ortaya çıkan listede birden fazla doğru öğe varsa geçerli değildir. Bir RS, döndürülen sonuç için 6 öğe alırsa ve üçüncü en yüksek noktada ilgili bir öğe olduğu ortaya çıkarsa, MRR yalnızca bunu dikkate alır (Chai vd., 2016; Luo vd., 2014).

$$MRR = \frac{1}{|U_{all}|} \sum_{u=1}^{|U_{all}|} RR(u) \quad (13)$$

$$RR(u) = \sum_{i \leq L} \frac{relevance_i}{rank_i}$$

burada, RR(u) bir u kullanıcısının karşılıklı sıralamasıdır ve karşılıklı sıraya göre ağırlıklandırılan en iyi L öğelerinin alaka düzeyi puanının toplamı ile tanımlanır. MRR, test veri kümesindeki tüm kullanıcıların ortalamasıdır.

Normalleştirilmiş Mesafeye Dayalı Performans Ölçümü (Normalized Distance-based Performance Measure—NDPM): Bir diğer sıralama ölçütü NDPM'dir. Kullanıcıya belirli sayıda öneri verilmeyecek durumda, yani

kullanıcıya sayfalar ayrılarak sonraki öğeler sırayla veya kullanıcının isteği üzerine gösterilecekse NDPM kullanılabilir (Lü vd., 2012; Yao, 1995).

Kazanç (Gain—G): Sıralama kalitesini anlamak için kullanılan bir metriktir. Bir RS çözümünün sunduğu her tavsiyenin, istenen sonuçla olan ilgisine bağlı olarak, bir geçerlilik puanı vardır. Örneğin, sonuçlar 1 ile 5 arasında derecelendirilebilir ve her tavsiye için ayrı bir ilgililik düzeyi puanı belirlenir. Kümülatif Kazanç (Cumulative Gain—CG), tavsiye kümesindeki tüm derecelendirilmiş ilgililik puanlarının toplamıdır (Shani & Gunawardana, 2011; Gunawardana & Shani, 2009).

$$\text{Cumulative Gain (CG)} = \sum_{i=1}^n \text{relevance score}_i \quad (14)$$

Ancak CG'nin dezavantajı, arama sonuçlarının sırasının önemli olmaması ve iki ayrı listenin farklı ilgililik düzeyi puanlarına sahip olsa bile CG değerinin aynı olabilmesidir. Bu nedenle, İndirgenmiş Kümülatif Kazanç (Discounted Cumulative Gain—DCG) daha kesin bir ölçü olarak önerilmiştir. DCG, en altta yer alan son derece alakalı öğeleri cezalandırmak için modellenmiştir (Lü vd., 2012).

$$\text{DCG} = \sum_{i=1}^n \frac{\text{relevance score}_i}{\log_2(i+1)} \quad (15)$$

DCG, sıralamanın önemini göz önünde bulundurduğu için başlangıçta iyi bir ölçü gibi görünüyor, fakat yine de yeterli değildir. Çeşitli faktörlere bağlı olarak, sunulan önerilerin sayısı her kullanıcı için değişebilir. DCG değeri de buna göre değişecektir. DCG'nin normalleştirilmesi, metriğin farklı uzunluklardaki sıralama listeleriyle karşılaştırılmasına olanak tanır.

$$\text{NDCG} = \frac{\text{DCG}}{\text{iDCG}} \quad (16)$$

Yukarıdaki formülde iDCG, ideal DCG (iDCG) anlamına gelmektedir. Bu ölçüm, DCG ölçümünün bir uzantısıdır ve sıralı listenin içindeki ilgili öğelerin konumunu dikkate alır. Aşağıdaki örnekte, B kümesi'nin ideal olarak sıralandığını görüyoruz. İdeal sıralama, en büyükten en küçüğe alaka düzeyi puanına göre sıralamadır. B Kümesi'nin DCG hesaplaması aynı zamanda iDCG hesaplamasıdır.

Son olarak, iDCG genellikle gerçek DCG birlikte kullanılarak Normalleştirilmiş İndirimli Kümülatif Kazanç (Normalized Discounted Cumulative Gain—NDCG) hesaplamak içinde kullanılabilir. NDCG, DCG'den farklı olarak, farklı uzunluklardaki veri kümelerinin değerlendirilmesinde kullanılabilir. Sunulan önerilerin NDCG ortalamasını hesaplanarak RS çözümlerinin verimliliğini değerlendirmek mümkündür (Shani & Gunawardana, 2011; Gunawardana & Shani, 2009).

Tablo 1. Literatürdeki RS değerlendirme metrikleri kıyaslamaları yapılmıştır. 4. sütun metrik türünü belirtirken, 5. sütun hedef değeri gösterir (ör., daha küçük RMSE, daha yüksek sıralama doğruluğu anlamındadır). 6. sütun, değerlendirme kategorilerini gösterirken, son iki sütun, metriğin bir sıralamadan elde edilip edilmediğini ve o metriğin öneri listesinin uzunluğuna ilişkisel olup olmadığını gösterir.

METRİK ADI (TR)	METRİK ADI (ENG)	SEMBOL	TÜR	HEDEF	KAPSAM	RANK	LEN (L)
Ortalama Kare Hatası	Mean Squared Error	MSE	Predictive	Small	Rating Accuracy	No	No
Kök Ortalama Kare Hatası	Root Mean Squared Error	RMSE	Predictive	Small	Rating Accuracy	No	No
Ortalama Mutlak Hata	Mean Absolute Error	MAE	Predictive	Small	Rating Accuracy	No	No
Ortalama Mutlak Kullanıcı Hatası	Mean Absolute User Error	MAUE	Predictive	Small	Rating Accuracy	No	No
R Kare	R Squared	R ²	Predictive	Large	Rating Accuracy	No	No
Açıklanan Varyans	Explained Variance	EV	Predictive	Large	Rating Accuracy	No	No
Pearson Korelasyon Katsayısı	Pearson Correlation Coefficient	PCC	Predictive	Large	Rating Correlation	No	No
Doğruluk	Accuracy	A(L)	Classification	Large	Classification Accuracy	No	Yes
Duyarlılık	Precision	P(L)	Classification	Large	Classification Accuracy	No	Yes
Geri Çağırma	Recall	R(L)	Classification	Large	Classification Accuracy	No	Yes
Ters Hassasiyet	Inverse Precision	IP(L)	Classification	Small	Classification Accuracy	No	Yes
Ters Geri Çağırma	Inverse Recall	IR(L)	Classification	Small	Classification Accuracy	No	Yes
F1 Ölçümü	F1-Measure	F1(L)	Classification	Large	Classification Accuracy	No	Yes
Ahçı Çalışma Özellikleri	Receiver Operating Characteristics	ROC	Classification	Large	Classification Accuracy	No	Yes
Lojistik Kayıp	Logistic Loss	LOC-Loss	Classification	Small	Classification Accuracy	No	Yes
Eğrinin Altındaki Alan	Area Under Curve	AUC	Classification	Large	Classification Accuracy	No	No
Özgüllük	Specificity	SPE	Classification	Large	Classification Accuracy	No	No
Hassasiyet @ K	Precision @ K	P(k)	Rank	Large	Classification Accuracy	Yes	No
Ortalama Hassasiyet	Average Precision	AP	Rank	Large	Classification Accuracy	Yes	No
Ortalama Averaj Hassasiyet	Mean Average Precision	MAP	Rank	Large	Classification Accuracy	Yes	No
İsabet Oranı	Hit Ratio	HR	Rank	Large	Robustness	No	Yes
Ortalama Karşılıklı Sıralama	Mean Reciprocal Rank	MRR	Rank	Large	Ranking Correlation	Yes	No
Normalleştirilmiş Mesafeye Dayalı Performans Ölçümü	Normalized Distance-Based Performance Measure	NDPM	Rank	Small	Ranking Correlation	Yes	No
Kazanç	Gain	G	Rank	Large	Satisfaction and Precision	Yes	Yes
Kümülatif Kazanç	Cumulative Gain	CG (b, L)	Rank	Large	Satisfaction and Precision	Yes	Yes
İndirimli Kümülatif Kazanç	Discounted Cumulative Gain	DCG (b, L)	Rank	Large	Satisfaction and Precision	Yes	Yes
Normalleştirilmiş İndirgenmiş Kümülatif Kazanç	Normalized Discounted Cumulative Gain	NDCG (b, L)	Rank	Large	Satisfaction and Precision	Yes	Yes
Spearman's Rho	Spearman's Rho	P	Rank	Large	Rating Accuracy	Yes	No
Kendall'in TAU'su	Kendall's Tau	τ	Rank	Large	Rating Accuracy	Yes	No
Sıralama Puanı	Ranking Score	RS	Rank	Small	Ranking Accuracy	Yes	No
Rütbe Tarafı Hassasiyet	Rank-Biased Precision	RBP (p, L)	Rank	Large	Satisfaction and Precision	Yes	Yes
Yenilik	Novelty	NOV	Non-Accuracy	Large	Popularity Ranking, Non-obviousness	Yes	No
Çeşitlilik	Diversity	NA	Non-Accuracy	Small	Intra-list Similarity	No	Yes
Tesadüf	Serendipity	SER	Non-Accuracy	Large	Novelty, Unexpectedness, Utility & Diversity, Non-Obviousness	Yes	No
Kapsama Alanı	Coverage	COV(L)	Non-Accuracy	Large	Coverage and Diversity	No	Yes
Hamming Mesafesi	Hamming Distance	H(L)	Predictive	Large	Intra-Diversity	No	Yes
Benzerlik İçi	Intra-Similarity	I(L)	Predictive	Small	Intra-Diversity	No	Yes
Popülerlik	Popularity	N(L)	Non-Accuracy	Small	Diversity and Novelty	No	Yes
Kendini Tanıma	Self-Information	U(L)	Non-Accuracy	Large	Unexpectedness, Surprisal & Novelty	No	Yes

TR-ENG: Sınıflandırmaya Dayalı (Classification), Tahmine Dayalı (Prediction-based), Kalite (Quality), Sıralamaya Dayalı (Ranking)

Alaka puanına göre örnek 2 öneri kümeleri verilmiş olsun: A= [2, 3, 3, 1, 2], B= [3, 3, 2, 2, 1]
Beklenen İdeal Sıralama= [3, 3, 2, 2, 1] olmalıdır. Bu öneri kümelerine göre:
CG, DCG ve NDCG skorları:
 $CG_A = 2+3+3+1+2=11$, $CG_B = 3+3+2+2+1=11$

$$DCG_A = \sum_{i=1}^n \frac{relevance\ score_i}{\log_2(i+1)} = \frac{2}{\log_2(1+1)} + \frac{3}{\log_2(2+1)} + \frac{3}{\log_2(3+1)} + \frac{1}{\log_2(4+1)} + \frac{2}{\log_2(5+1)} \cong 6.6$$

$$DCG_B = \sum_{i=1}^n \frac{relevance\ score_i}{\log_2(i+1)} = \frac{3}{\log_2(1+1)} + \frac{3}{\log_2(2+1)} + \frac{2}{\log_2(3+1)} + \frac{2}{\log_2(4+1)} + \frac{1}{\log_2(5+1)} \cong 7.14$$

$$NDCG_A = \frac{DCG_A}{iDCG} = 6.6 / 7.14 = 0.92$$

$$NDCG_B = \frac{DCG_B}{iDCG} = 7.14 / 7.14 = 1.0$$

Burada, $NDCG_B > NDCG_A$ olduğundan dolayı, B kümesinin ideal sıralamaya göre en uygun öneri düzenine sahip olduğu anlaşılmaktadır.

Spearman's Rho: İstatistiksel tabanlı sıralama ölçütüdür (Spearman, 1961). Spearman sıra korelasyon katsayısı, iki değişkenin sıraları arasındaki istatistiksel bağımlılığı ölçen parametrik olmayan bir ölçüdür (Lü vd., 2012). Monotonik bir ilişkide, ilk değişkenin değeri değişirse, ikinci değişkenin değeri de değişir, ancak bu oran doğrusal değildir. PCC ile karşılaştırıldığında Spearman katsayısı sıralamaya dayalıdır ve parametre olmayan unsurlardan ve büyüklükten bağımsızdır. PCC, iki değişken arasındaki doğrusal ilişkileri değerlendirirken, Spearman monotonik ilişkilerini değerlendirir (Kang vd., 2019).

Doğruluğa Dayalı Olmayan Metrikler

Bu ölçümler, tahminleri temel doğruluk değerleriyle karşılaştırmaz, bunun yerine tavsiyelerin aşağıdaki özelliklerini değerlendirir ve sistemin kalite kriterlerini irdelerler.

Çeşitlilik (Diversity): Bu metrik, bir kümedeki farklı öğelerin birbirlerine göre ne kadar olduğunu ölçer (Silveira vd., 2019; Zhang & Feng, 2016; Avazpour vd., 201; Lü vd., 2012; Shani & Gunawardana, 2011).

Yenilik (Novelty): Yenilik, kullanıcılar arasındaki öneri sıklıklarını hesaplayarak yeni öneri öğelerinin nasıl olduğunu ölçer (Silveira vd., 2019; Lü vd., 2012).

$$nov(R_u) = \sum_{i \in R_u} \min_{j \in H_u} d(class(i), class(j)) \quad (17)$$

Ölçeklenebilirlik (Scalability): Bu metrik, bir sistemin veya modelin artan iş yükü veya artan kullanıcı taleplerini sürdürme ve bunlarla iyi başa çıkma yeteneğini ifade eder (Lü vd., 2012).

Tesadüf (Serendipity): Bu metrik, belirli bir kullanıcı için önerilerin ne kadar şaşırtıcı olduğunu, kullanıcının daha önce etkileşimde bulunduğu öğelerle karşılaştırarak ölçer (Silveira vd., 2019; Avazpour vd., 2014; Shani & Gunawardana, 2011; Gunawardana & Shani, 2009).

Kapsam (Coverage): Bu metrik, sistem tarafından önerilen öğelerin dağılımı ile ilgili ölçümlerdir (Pan vd., 2016; Avazpour vd., 2014). RS'ler de kullanılan yukarıda sunulan tüm değerlendirme ölçütleri, performans karşılaştırmalı tercih ilişkisi, kısaltmaları, ve kapsamaları bakımından sınıflandırılması Tablo 1'de sunulmuştur.

RS Çözümlerinde Önerme Sürecindeki Rastlanan Zorluklar

RS veri tabanında bulunan veriler sayısal ve ilişkisel olarak zayıfsa sistem performansı da zayıf olabilir. Bu nedenle, RS'ler de hala çözülmesi gereken birçok zorluk olduğu söylenebilir. Bu sorunların bir kısmı aşağıda irdelenmiştir.

Tesadüf ve Yenilik: RS'ler sıra dışı olabilir. RS'ler, kullanıcıya daha önce hiç duyulmamış veya önerilmemiş olağandışı öneriler üretebilir, ancak aynı zamanda şaşırtıcı etkileri de vardır (Mohamed vd., 2019).

Seyreklik: Seyreklik, RS'lerde görülen bir diğer önemli sorundur. Bu sorun, bir kullanıcının tercih edilen öğelerden oluşan çok büyük bir matrisi olduğunda ortaya çıkar (Ör, izlenen filmler veya dinlenen şarkılar). Bu durum, kullanıcı tercihlerini derecelendirmediğinde ortaya çıkar. Çünkü RS'ler, kullanıcılarının tercihlerini diğer kullanıcılara önerebilmek için derecelendirme matrislerine ihtiyaç duyarlar (Althbiti vd., 2021; Mohamed vd., 2019).

Gizlilik: RS'lerdeki en büyük zorluklardan biri veri gizliliğidir. Kullanıcının tercihlerine/ilgi alanlarına göre uygun öğeler önermek için, RS'lerin kullanıcıları hakkında bazı önemli bilgileri önceden bilmesi gerekir. RS'ler kullanıcılarının bu kişisel bilgilerini saklarken, kullanıcı gizliliği yetkisiz kişilere karşı da korunmalıdır (Mohamed vd., 2019).

Çeşitlilik: RS'lerden elde edilen öneri sonuçlarının hepsinin aynı yerden gelmemesi sağlar, tekrar etmesini önler ve çeşitliliği artırır (Mohamed vd., 2019).

Yorumlanabilirlik: Günümüzde RS ve yapay zekâ birçok sektörde yüksek talep görmektedir. Bu disiplinlerin veya alt disiplinlerin çoğunda, yalnızca doğru tahminlere sahip olmak değil, aynı zamanda kararın arkasında son kullanıcılar için şeffaf hale getirilebilecek bir gerekçe sağlamak da önemlidir (Fusco vd., 2019; Mohamed vd., 2019). Bu nedenle, bir RS'nin yapısına entegre edilen model yorumlanabilir olmalıdır. Özellikle, model tahminlerinin arkasındaki mantığı açıklamalı ve kullanıcı-öge etkileşimlerinin altında yatan mekanizma hakkında içgörüler sağlamalıdır. Bunu çözmek için, kullanıcı-öge etkileşimlerini işlemek ve kullanıcı/öge özelliklerini korumak için sinir ağı mimarileri uygulanabilir (Fusco vd., 2019; Mohamed vd., 2019).

Soğuk başlatma: Yeni bir sistem kullanıcısına öneri sunmanın zorluğunu ifade eder. RS'lerde işbirlikçi filtrelemenin temel sorunlarından biridir. Popüler öğeleri yeni bir kullanıcıya göstererek veya yeni bir öge için bazı kural tabanlı filtreler ekleyerek önlenabilir (Mohamed vd., 2019).

Ölçeklenebilirlik: Bir RS'nin tek bir zaman diliminde çok miktarda farklı türde veriyi işleyebilmesi, ölçeklenebilirliğini gösterir. Kullanıcı veya ürün sayısı arttıkça, RS'nin performansına ve etkinliğine zarar vermeden kullanıcılarına uygun ürünü önermesi beklenir. Bu sorun birçok şekilde çözülebilir. Örneğin, veri kümesi çok büyük olduğunda, Hadoop gibi büyük veri sistemlerine geçiş sağlanabilir. Diğer bir çözüm ise örnek küme üzerinde modeller/kurallar oluşturmaktır. Başka bir çözüm, modeli aşamalara ayırmak ve çalıştırmak olacaktır (Mohamed vd., 2019; Lü vd., 2012).

İŞ TAVSİYE SİSTEMLERİNDE KULLANILAN PERFORMANS METRİKLERİ VE KIYASLANMASI

İş Tavsiye Sistemleri (Job Recommender Systems) Arka Planı

Özgeçmişler, bireylerin iş arama sürecinde kullandıkları kendilerini tanımlayan en önemli belgelerdir. Geçmişten günümüze insanlar iş başvurusunda bulunurken önce özgeçmişlerini oluşturur ve daha sonra şirketlerin İK sistemine veya iş arama sitelerine özgeçmişlerini yüklemektedirler. Özgeçmişler, bireyler tarafından yaratılırken doküman üretme araçları kullanarak yapısız bir formatta sunulabilirken, aynı zamanda iş bulma siteleri üzerinden de daha yapısal bir formatta (İK yazılım çözümlerinin kullanıcı girdi formları/ekranları aracılığıyla) üretilebilmekte ve İK özgeçmiş havuzlarında yükleyebilmektedirler. Özgeçmişler genellikle, “.doc”, “.docx”, “.pdf”, veya “.txt” gibi farklı formatlarda, farklı yazı tiplerinde, renklerde, yazı tipi boyutlarında, vb. gibi yapısal farklılıklar içerebilmekte ve İK personeli tarafından değerlendirilmek üzere İK sistemleri üzerinden yüklenebilmektedirler.

Zaman içerisinde, milyonlarca özgeçmiş biriken İK havuzlarından, İK'nın en önemli süreçlerinden bir olan işe alım sürecinin etkin bir şekilde uygulanması ve doğru adayın belirlenmesi oldukça zorlu bir İK süreci haline gelmiştir. İK uzmanı, önce iş tanımı hazırlar ve daha sonra güncel özgeçmişleri almak için çeşitli iş portalları kullanır. Adayın pozisyona başvurmasını beklemek yerine nitelikli adayların araştırılması, seçilmesi, belirlenmesi ve onlarla iletişime geçilmesi anlamına gelir. Bu noktada, RS teknolojilerinin kullanımı oldukça etkin bir çözüm olduğunu söylemek mümkündür. İş Tavsiye Sistemleri (Job Recommender Systems—JRS), son zamanlarda popüler bir araştırma sahası olmuş ve şirketlerin İK havuzunda/iş portallarında saklanan yapılandırılmamış/yarı yapılandırılmış özgeçmişlerin çeşitli yapay zekâ ve doğa dil işleme teknolojileri, algoritmaları veya modelleri ile işlenmesiyle, bireylerin özgeçmişlerindeki istenen bölümlerin (ör. demografik bilgilerin, iş tecrübelerinin ve eğitim bilgilerinin vb.) çıkarılması kısa zamanda mümkün olmuştur. Ayrıca, JRS çözümleri sayesinde bu sürecin efektif ve önyargısız yönetilmesi mümkün olmuştur. Böylece, hedef pozisyona en uygun personelin daha adaletli ve doğru olarak seçilip atanabilmesi, zaman kayıplarının önlenmesi ve İK departmanlarında iş yükünün azalması etkin JRS çözümleri ile

mümkün olmuştur. Literatürde, çeşitli bu tür JRS çözümlerine (ör. CASPER, ResuMatcher, eRecruiter, FES, vb) rastlanmaktadır (Almalis vd., 2015). Ancak, literatürde Türkçe dilinde JRS çözümlerine yeterince rastlanmamaktadır. İş adaylarına iş fırsatı bulma konusunda rehberlik edecek az sayıda etkin popüler JRS çözümü (ör. Kariyer.Net, Secretec.com, vb.) bulunmaktadır.

Literatür taramalarına göre, JRS çözümlerinde 6 filtreleme türünden bahsedilmiş ve şu şekilde kategorize edilmiştir: (1) İşbirlikçi Filtreleme, (2) İçerik Tabanlı Filtreleme, (3) Hibrit Tabanlı Filtreleme, (4) Bilgi Tabanlı Filtreleme, (5) Bağlam Farkındalığına Dayalı Filtreleme (Demografik JRS'ler), ve (6) İki Yönlü Filtrelemeye (Reciprocal JRS'ler gibi) dayalı çeşitli JRS çalışmalarına rastlanmıştır (Gunawardana vd., 2022; Zhao vd., 2021; Palomares vd., 2021; Çelik & Elçi, 2020; Mohamed vd., 2019; Arita vd., 2017; Al-Shamri vd., 2016; Wenxing vd., 2015; Poriya vd., 2014; Lü vd., 2012; Park vd., 2012; Shani & Gunawardana, 2011; Martin vd., 2011; Yu vd., 2011; Malinowski vd., 2006).

JRS çözümleri, tek taraflı veya çift taraflı (iki yönlü) tavsiyeler yapabilir. Tek taraflı tavsiye sağlayan JRS çözümlerinde, tavsiye isteyen tarafa ya "açık iş pozisyonu ilanı" önerilir ya da açık bir iş pozisyonuna "uygun özgeçmişler" önerilir. Literatürde, tek taraflı tavsiye sağlayan JRS çözümlerinde genelde işbirlikçi, içerik-tabanlı, hibrit, bilgi tabanlı, bağlam farkındalığına dayalı filtreleme JRS çözümlerine rastlanmıştır (Kwieciński vd., 2023; Bothmer & Schlippe, 2022; Prafajar vd., 2022; Liang & Wan, 2022; Zhao vd., 2021; Okfalisa vd., 2021; Palomares vd., 2021; Tamburri vd., 2020; Martinez-Gil vd., 2020; González-Briones vd., 2019; Rimitha vd., 2019; Guan vd., 2019; Arita vd., 2017; Aouadni & Rebai, 2017; Benabderrahmane & Lamolle, 2017; Yi vd., 2016; Wenxing vd., 2015; Almalis vd., 2015; Fang vd., 2013; Malinowski vd., 2006). Ayrıca, iki yönlü tavsiye, diğer deyişle "Reciprocal JRS" çözümlerinde, her iki yönde de tavsiyeler üretilir ve karşılıklı uzlaşma beklenir. Bu sistemlerde, her iki tarafın da uzlaşmasıyla ve yapılan tavsiyeye uygunluk vermesiyle süreç tamamlanır (Palomares vd., 2021; Wenxing vd., 2015; Yu vd., 2011; Malinowski vd., 2006).

Popülerliğe dayalı RS'ler doğası gereği kullanıcı davranışlarını analiz etmez, diğer bir deyişle kullanıcının kişisel tercihlerine dayalı olmayan bir sistemdir. Tercih edilen, popüler/trend, yaygın olarak tüketilen ürünleri önerir. Kullanıcının ilgi alanlarını genişletmek ve uygulama tercihlerini popülerleştirmek için bir fırsat olabilir. Örneğin bir e-ticaret sitesi, Covid-19 pandemisi döneminde çok alakasız bir ürüne bakıyor olsanız bile size maske ya da dezenfektan sunabilmesi şeklinde ifade edilebilir. Bunun nedeni, kullanılan RS'nin aslında popülerliğe dayalı filtreleme içermesidir (Kumar vd., 2014). Ayrıca, kişiselleştirilmiş bir RS, kullanıcının geçmiş etkileşimlerine (ör. önceki satın alma geçmişine) dayalı olarak öğeler önerirken, kişiselleştirilmemiş bir RS, o dönemde sistem kullanıcıları arasında popüler olan öğeleri önerir (Poriya vd., 2014). Literatürde kişiselleştirilmemiş JRS veya popülerliğe dayalı bir JRS çözümüne rastlanmamıştır.

Gelecek bölümde, JRS çözümlerinde kullanılan performans metrikleri hakkında literatür inceleme sonuçları paylaşılmaktadır.

İş Tavsiye Sistemlerinde Tercih Edilen Değerlendirme Metriklerinin Dağılımı ve Yorumlanması

RS'lerin başarısını ve verimliliğini değerlendirmek için çeşitli ölçüm yöntemleri ve metrikler Bölüm 3'te tartışılmıştır. Toplamda 30 adet metrik tespit edilmiş ve bu metrikler dört ana kategoride sınıflandırılmıştır: (1) Tahmine Dayalı Doğruluk Metrikleri (7 metrik), (2) Sınıflandırmaya Dayalı Doğruluk Metrikleri (10 metrik), (3) Sıralamaya Dayalı Doğruluk Metrikleri (8 metrik) ve (4) Doğruluğa Dayalı Olmayan Metrikler (5 metrik) (Gunawardana vd., 2022; Vijaysinh, 2021; Chaaya vd., 2017; Gunawardana & Shani, 2009;). Bu bölümde, 30 adet RS performans metriklerinin hangileri JRS çözümlerinde uygulanmış ve denenmiş olduğunu ve hangi filtreleme yaklaşımlarında uygulandığına dair tartışmalar yer almaktadır.

Tablo 2'de görüldüğü üzere JRS çözümlerinde filtreleme tekniğine göre en çok uygulanan performans değerlendirme metrikleri sınıflandırmaya dayalıdır (ör. Accuracy, Recall, Precision). Hibrit filtrelemeye dayalı JRS çözümleri genelde sınıflandırma, tavsiye sıralama ve tahminlemeye dayalı çözüm ve metrikleri işlemişlerdir. Takibinde işbirlikçi ve içeriğe-dayalı filtreleme tabanlı JRS çözümlerinde benzer performans kriterlerini işlemişlerdir. Kalite bakımından, diğer deyişle, önerilen tavsiyelerin ne kadar doğru üretildiğine, benzersizliğine, tekrara düşmemesine, kapsamına, vb. nitelik bakımından JRS çözümlerinin değerlendirilmesi neredeyse hiç ele alınmamış olması dikkat çekmiştir. Sadece işbirlikçi, içeriğe-dayalı ve iki yönlü filtrelemeye dayalı JRS çözümlerinde nitelik bakımından bazı değerlendirmelere rastlanmıştır.

Tablo 2. JRS Çözümlerindeki Değerlendirme Metriklerine Genel Bakış.

NO	DEĞERLENDİRME METRİKLERİ	METRİK TÜRÜ	Filtreleme Türleri						Kullanım Yüzde
			İşbirliğine Dayalı Filtreleme	İçerik Tabanlı Filtreleme	Hibrit Filtreleme	Bilgiye Dayalı Filtreleme	Bağlama Duyarlı (Demografik Filtreleme)	Karşılıklı Tabanlı Filtreleme	
1	Accuracy	Classification	+	+	+	+	+	+	%100
2	Recall	Classification	+	+	+	+	+	+	%83
3	Precision	Classification	+		+			+	%50
4	NDCG	Ranking	+		+				%33
5	AUC	Classification		+	+				%33
6	F-Score	Classification	+		+				%33
7	MAE/MRE	Prediction-based			+			+	%33
8	MSE/RMSE	Prediction-based	+	+	+				%50
9	Hit-Rate	Ranking	+		+				%33
10	Loss-Func.	Classification			+				%16
11	Reliability	Quality							%0
12	Coverage	Quality			+				%16
13	Security	Quality							%0
14	Robustness	Quality							%0
15	Scalability	Quality	+						%16
16	Consistency Ratio	Quality					+		%16
17	Time	Quality		+					%16
18	Cost	Quality	+	+					%33
19	Karşılıklılık (Reciprocity)	Quality						+	%16
20	Sınırlılık (Limitedness)	Quality						+	%16
21	Pasiflik (Passivity)	Quality						+	%16
22	Sparseness (Seyreklik)	Quality						+	%16

TR-ENG: Sınıflandırmaya Dayalı (Classification), Tahmine Dayalı (Prediction-based), Kalite (Quality), Sıralamaya Dayalı (Ranking)

RS'lerin İK çözümlerine entegrasyonu söz konusu olduğunda, yukarıda belirtilen zorlukların ön plana çıkması kaçınılmazdır. Örneğin, İK çözümlerinde temel kalite ve performans kriterlerinin yetkin bir RS'de sağlanmasının yanı sıra, İK sistemine özgeçmiş yükleyen kullanıcıların gizliliği yetkisiz saldırganlardan korunmalıdır. Şu anda, İK hizmetlerindeki AI tabanlı JRS'ler veya akıllı RS'ler, uygulama, güvenilirlik ve doğruluk açısından yeterince olgun değildir. İK profesyonellerinin bakış açısına göre, JRS'ler doğru eşleşme sonuçları üretebilecek güvenilir ve hızlı tavsiyeler sunma yeteneğine sahip olmalıdır (Isinkaye vd., 2015). İK uzmanları, etkin JRS'ler aracılığıyla iş adaylarının yüklenen özgeçmişlerini uygun iş tanımlarıyla eşleştirebilir, sıralayabilir ve etiketleyebilir. Bu ilişkisel bağlantı ve sıralama, zaman içinde yüksek hacimli derecelendirme matrisleri oluşturabilir, böylece önerilerde yenilik ve çeşitlilik sağlar.

SONUÇLAR

COVID-19'un etkisiyle birçok sektörde yürütülen iş süreçlerin ve operasyonların teknoloji ile güçlendirilmesi ve uzaktan nihai kullanıcılara, doğru hizmetin ulaştırılması, elzem bir ihtiyaç haline getirmiştir. Günümüzde tavsiye sistemleri, birçok alanda (ör. e-ticaret, eğitim, eğlence, sağlık, insan kaynakları, reklamcılık, vb.) yer edinmeye başlanmıştır. Tavsiye sistemleri sayesinde, kullanıcılara kişiselleştirilmiş sonuçlar üretilmesi ve kullanıcıya en uygun öğelerin önceliklendirilmesi mümkün olabilmektedir. Bu çalışmada, RS mimarisi ve mevcut çözümler üzerine, literatürde 2010-2023 yılları arasında yapılmış çeşitli güncel araştırma makaleleri ilgililik durumuna göre seçilmiş ve incelenmiştir. Yoğun araştırma çalışmaları sonucunda, tavsiye sistemlerinde öğe filtreleme teknikleri geniş çapta araştırılmış ve bir taksonomi yapısında sunulmuştur. Ayrıca, tavsiye sistemlerinde çoğunluk araştırmalarda uygulanan performans değerlendirme metrikleri literatürden tespit edilmiş, kıyaslanmış ve sınıflandırılmıştır. İnceleme çalışmalarının çıktılarında göre, literatürde RS'ler için uygulanan öğe filtreleme mekanizmaları 6 kategoriye ayrılmıştır: (1) İşbirlikçi Filtreleme, (2) İçerik Tabanlı Filtreleme, (3) Hibrit Tabanlı Filtreleme, (4) Bilgi Tabanlı Filtreleme, (5) Bağlam Farkındalığına Dayalı Filtreleme ve (6) Diğerleri. Ayrıca, RS'lerin doğrulama performans değerlendirmelerinde kullanılan, çoğunluk performans/kalite metrikleri saptanarak, şu şekilde sınıflandırılmıştır: (1)

Tahmine Dayalı Doğruluk Metrikleri, (2) Sınıflandırmaya Dayalı Doğruluk Metrikleri, (3) Sıralamaya Dayalı Doğruluk Metrikleri, ve (4) Doğruluğa Dayalı Olmayan Metrikler olarak kategorize edilmiş ve tartışılmıştır.

Son olarak, tavsiye sistemlerinin en çok ihtiyaç duyulduğu sektörlerden biri olan, İnsan Kaynakları sahasındaki süreç/operasyonlarda, tavsiye sistemleri çözümlerinin uygulanması, İş Tavsiye Sistemleri (JRS) olarak bilinmektedir. Bu çalışmanın son katkısı olarak, JRS uygulamaları bir araştırma sahası olarak seçilmiş ve JRS çözümlerinde kullanılan filtreleme teknikleri ve etkin performans/kalite değerlendirme metrikleri tespit edilmiştir. Gelecek çalışmalarda, JRS çözümlerinde kullanan güncel AI ve NLP model ve yaklaşımlarını irdeleyen bir sistematik literatür çalışmasının yapılması planlanmaktadır. İlave, literatürde İngilizce dilinde uygulanmış güncel çözümlerin, Türkçe dilinde çıkarım sağlayabilen yenilikçi JRS çözümlerine entegrasyonu hakkında gerekli bilgi birikiminin elde edilmesi amaçlanmaktadır.

Fon Kullanımı: Yazarlar araştırmanın fon desteği olmadığını bildirir.

Çıkar çatışması: Yazarlar herhangi bir çıkar çatışması olmadığını beyan ederler.

Etik onay beyanı: Bu çalışma için resmi onay gerekli değildir.

Yazar Katkıları: Her iki yazar eşit oranda katkı sağlamıştır.

KAYNAKLAR

- Adomavicius, G., & Tuzhilin, A. (2011). Context-aware recommender systems. In *Recommender systems handbook* (pp. 217-253). Springer, Boston, MA. <https://doi.org/10.1609/aimag.v32i3.2364>
- Al-Habaibeh, A., Watkins, M., Waried, K., & Javareshk, M. B. (2021). Challenges and opportunities of remotely working from home during Covid-19 pandemic. *Global Transitions*, 3, 99-108. <https://doi.org/10.1016/j.glt.2021.11.001>
- Almalis, N. D., Tsihrintzis, G. A., Karagiannis, N., & Strati, A. D. (2015). FoDRA—A new content-based job recommendation algorithm for job seeking and recruiting. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-7). IEEE. <https://doi.org/10.1109/IISA.2015.7388018>
- Al-Otaibi, S., & Ykhlef, M. (2017). Hybrid immunizing solution for job recommender system. *Frontiers of Computer Science*, 11(3), 511-527. <https://doi.org/10.1007/s11704-016-5241-z>
- Al-Shamri, M. Y. H. (2016). User profiling approaches for demographic recommender systems. *Knowledge-Based Systems*, 100, 175-187. <https://doi.org/10.1016/j.knosys.2016.03.006>
- Althbiti, A., Alshamrani, R., Alghamdi, T., Lee, S., & Ma, X. (2021). Addressing data sparsity in collaborative filtering-based recommender systems using clustering and artificial neural network. In *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 0218-0227). IEEE. <https://doi.org/10.1109/CCWC51732.2021.9376008>
- Aouadni, I., & Rebai, A. (2017). Decision support system based on genetic algorithm and multi-criteria satisfaction analysis (MUSA) method for measuring job satisfaction. *Annals of Operations Research*, 256(1), 3-20. <https://doi.org/10.1007/s10479-016-2154-z>
- Arita, S., Hiyama, A., & Hirose, M. (2017). Gber: A social matching app which utilizes time, place, and skills of workers and jobs. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 127-130). <https://doi.org/10.1145/3022198.3026316>
- Avazpour, I., Pitakrat, T., Grunske, L., & Grundy, J. (2014). Dimensions and metrics for evaluating recommendation systems. In *Recommendation systems in software engineering* (pp. 245-273). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-45135-5_10
- Ayub, M., Ghazanfar, M. A., Maqsood, M., & Saleem, A. (2018). A Jaccard base similarity measure to improve performance of CF based recommender systems. In *2018 International conference on information networking (ICOIN)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICOIN.2018.8343073>
- Beel, J., Gipp, B., Langer, S., & Breitingner, C. (2016). Paper recommender systems: a literature survey. *International Journal on Digital Libraries*, 17(4), 305-338. <https://doi.org/10.1007/s00799-015-0156-0>
- Benabderrahmane, S., Mellouli, N., & Lamolle, M. (2017). Predicting the users' clickstreams using time series representation, symbolic sequences, and deep learning: application on job offers recommendation tasks. In *2017*

- IEEE International Conference on Information Reuse and Integration (IRI) (pp. 436-443). IEEE. <https://doi.org/10.1109/IRI.2017.54>
- Bhat, S. S., Pranav, P., Shashank, K. V., Raghunandan, A., & Mohan, B. R. (2022). Comparative Performance Evaluation of Web-Based Book Recommender Systems. In 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 985-991). <https://doi.org/IEEE.10.1109/ICOEI53556.2022.9777116>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.). <https://doi.org/10.1016/j.knosys.2021.107134>
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46, 109-132. <https://doi.org/10.1016/j.knosys.2013.03.012>
- Bothmer, K., & Schlippe, T. (2022). Investigating natural language processing techniques for a recommendation system to support employers, job seekers and educational institutions. In *Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part II* (pp. 449-452). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11647-6_90
- Burke, R., Felfernig, A., & Göker, M. H. (2011). Recommender systems: An overview. *Ai Magazine*, 32(3), 13-18. <https://doi.org/10.1609/aimag.v32i3.2361>
- Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent data analysis*, 21(6), 1487-152. <https://doi.org/10.3233/IDA-163209>
- Çelik Ertuğrul, D., & Elçi, A. (2020). A survey on semanticized and personalized health recommender systems. *Expert Systems*, 37(4), e12519. <https://doi.org/10.1111/exsy.12519>
- Chaaya, G., Métails, E., Abdo, J. B., Chiky, R., Demerjian, J., & Barbar, K. (2017, December). Evaluating non-personalized single-heuristic active learning strategies for collaborative filtering recommender systems. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 593-600). IEEE. <https://doi.org/10.1109/ICMLA.2017.00-96>
- Chai, Y., Wang, C., Wen, Y., & Yuan, X. (2016). A Hadoop-Based Database Querying Approach for Non-expert Users. In *Asia-Pacific Web Conference* (pp. 449-453). Springer, Cham.
- Fang, D., Varshney, K. R., Wang, J., Ramamurthy, K. N., Mojsilovic, A., & Bauer, J. H. (2013). Quantifying and recommending expertise when new skills emerge. In 2013 IEEE 13th International Conference on Data Mining Workshops (pp. 672-679). IEEE. <https://doi.org/10.1109/ICDMW.2013.33>
- Fkih, F. (2022). Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. *Journal of King Saud University-Computer and Information Sciences*, 34(9), 7645-7669. <https://doi.org/10.1016/j.jksuci.2021.09.014>
- Fusco, F., Vlachos, M., Vasileiadis, V., Wardatzky, K., & Schneider, J. (2019). RecoNet: An Interpretable Neural Architecture for Recommender Systems. In *IJCAI* (pp. 2343-2349). <https://doi.org/10.24963/ijcai.2019/325>
- Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35, 61-70. <http://dx.doi.org/10.1145/138859.138867>.
- González-Briones, A., Rivas, A., Chamoso, P., Casado-Vara, R., & Corchado, J. M. (2019). Case-based reasoning and agent-based job offer recommender system. In *International Joint Conference SOCO'18-CISIS'18-ICEUTE'18: San Sebastián, Spain, June 6-8, 2018 Proceedings 13* (pp. 21-33). Springer International Publishing. https://doi.org/10.1007/978-3-319-94120-2_3
- Guan, Z., Yu, B., & Liu, Y. (2019). Recruitment and Recommendation System Based on Intelligent Computing. In *Proceedings of the 2019 5th International Conference on Computing and Data Engineering* (pp. 77-80). <https://doi.org/10.1145/3330530.3330532>
- Gunawardana, A., & Shani, G. (2009). A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, 10(12). <https://doi.org/10.1145/1577069.1755883>
- Gunawardana, A., Shani, G., & Yogev, S. (2022). Evaluating recommender systems. In *Recommender systems handbook* (pp. 547-601). Springer, New York, NY. https://doi.org/10.1007/978-0-387-85820-3_8

- Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian informatics journal*, 16(3), 261-273. <https://doi.org/10.1016/j.eij.2015.06.005>
- Kang, J. S., Shin, D. H., Baek, J. W., & Chung, K. (2019). Activity recommendation model using rank correlation for chronic stress management. *Applied Sciences*, 9(20), 4284. <https://doi.org/10.3390/app9204284>
- Kumar, R., Verma, B. K., & Rastogi, S. S. (2014). Social popularity based SVD++ recommender system. *International Journal of Computer Applications*, 87(14). <https://doi.org/10.5120/15279-4033>
- Kwieciński, R., Melniczak, G., & Górecki, T. (2023). Comparison of Real-Time and Batch Job Recommendations. *IEEE Access*, 11, 20553-20559. <https://doi.org/10.1109/ACCESS.2023.3249356>
- Liang, F., & Wan, X. (2022). Job Matching Analysis Based on Text Mining and Multicriteria Decision-Making. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2022/9245876>
- Liu, P., Ma, J., Wang, Y., Ma, L., & Huang, S. (2016). A context-aware method for top-k recommendation in smart TV. In *Asia-Pacific Web Conference* (pp. 150-161). Springer, Cham. https://doi.org/10.1007/978-3-319-45817-5_12
- Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32. <https://doi.org/10.1016/j.dss.2015.03.008>
- Lü, L., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., & Zhou, T. (2012). Recommender systems. *Physics reports*, 519(1), 1-49. <https://doi.org/10.1016/j.physrep.2012.02.006>
- Luo, Y., Xu, B., Cai, H., & Bu, F. (2014). A Hybrid User Profile Model for Personalized Recommender System with Linked Open Data. In *2014 Enterprise Systems Conference* (pp. 243-248). IEEE. <https://doi.org/10.1109/ES.2014.16>
- Malinowski, J., Keim, T., Wendt, O., & Weitzel, T. (2006). Matching people and jobs: A bilateral recommendation approach. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)* (Vol. 6, pp. 137c-137c). IEEE. <https://doi.org/10.1109/HICSS.2006.266>
- Marlowe, A. N. (2021). Robot Recruiters: How Employers & Governments Must Confront the Discriminatory Effects of AI Hiring. *J. High Tech. L.*, 22, 274.
- Martin, F. J., Donaldson, J., Ashenfelter, A., Torrens, M., & Hangartner, R. (2011). The big promise of recommender systems. *AI Magazine*, 32(3), 19-27. <https://doi.org/10.1609/aimag.v32i3.2360>
- Martinez-Gil, J., Paoletti, A. L., & Pichler, M. (2020). A novel approach for learning how to automatically match job offers and candidate profiles. *Information Systems Frontiers*, 22(6), 1265-1274. <https://doi.org/10.1007/s10796-019-09929-7>
- Mathur, A., Juguru, S. K., & Eirinaki, M. (2019). A graph-based recommender system for food products. In *2019 First International Conference on Graph Computing (GC)* (pp. 83-87). IEEE <https://doi.org/10.1109/GC46384.2019.00020>
- McFee, B., & Lanckriet, G. R. (2010). Metric learning to rank. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 775-782).
- Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. In *Aaai* (Vol. 6, No. 2006, pp. 775-780).
- Mohamed, M. H., Khafagy, M. H., & Ibrahim, M. H. (2019). Recommender systems challenges and solutions survey. In *2019 international conference on innovative trends in computer engineering (ITCE)* (pp. 149-155). IEEE. <https://doi.org/10.1109/ITCE.2019.8646645>
- Negroponte, N. (1970). *The architecture machine: toward a more human environment*. The MIT Press. <https://doi.org/10.7551/mitpress/8269.001.0001>
- Okfalisa, Siburian, R., Vitriani, Y., Rusnedy, H., Saktioto, & Yola, M. (2021). Job Training Recommendation System: Integrated Fuzzy AHP and TOPSIS Approach. In *International Conference of Reliable Information and Communication Technology* (pp. 84-94). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-98741-1_8
- Painsky, A., & Wornell, G. (2018). On the universality of the logistic loss function. In *2018 IEEE International Symposium on Information Theory (ISIT)* (pp. 936-940). IEEE. <https://doi.org/10.48550/arXiv.1805.03804>

- Palomares, I., Porcel, C., Pizzato, L., Guy, I., & Herrera-Viedma, E. (2021). Reciprocal Recommender Systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion*, 69, 103-127. <https://doi.org/10.1016/j.inffus.2020.12.001>
- Pan, Y., Zhang, Y., & Zhang, R. (2016). Combo-Recommendation Based on Potential Relevance of Items. In *Asia-Pacific Web Conference* (pp. 505-517). Springer, Cham. https://doi.org/10.1007/978-3-319-45817-5_55
- Park, D. H., Kim, H. K., Choi, I. Y., & Kim, J. K. (2012). A literature review and classification of recommender systems research. *Expert systems with applications*, 39(11), 10059-10072. <https://doi.org/10.1016/j.eswa.2012.02.038>
- Poriya, A., Bhagat, T., Patel, N., & Sharma, R. (2014). Non-personalized recommender systems and user-based collaborative recommender systems. *Int. J. Appl. Inf. Sys*, 6(9), 22-27. <https://doi.org/10.5120/ijais14-451122>
- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205-227. <https://doi.org/10.1016/j.eswa.2017.12.020>
- Prafajar, K. N., Vallyan, H., Candradewi, N. L. P. A., Edbert, I. S., & Suhartono, D. (2022). Multiclass job recommendation system in the IT field between classification and prediction method. In *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)* (pp. 181-186). IEEE. <https://doi.org/10.1109/GECOST55694.2022.10010659>
- Rimitha, S. R., Abburu, V., Kiranmai, A., Marimuthu, C., & Chandrasekaran, K. (2019). Improving Job Recommendation Using Ontological Modeling and User Profiles. In *2019 Fifteenth Int. Conference on Information Processing (ICINPRO)* (pp. 1-8). IEEE. <https://doi.org/10.1109/ICInPro47689.2019.9092271>
- Salton, G., and Buckley, C. (1997). Term weighting approaches in automatic text retrieval. In *Readings in Information Retrieval*. San Francisco, CA: Morgan Kaufmann Publishers. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- Salton, G., and Lesk, M. (1971). Computer evaluation of indexing and text processing. Prentice Hall, Ing. Englewood Cliffs, New Jersey. 143–180. <https://doi.org/10.1145/321439.321441>
- Salton, G., Singhal, A., Mitra, M., & Buckley, C. (1997). Automatic text structuring and summarization. *Information processing & management*, 33(2), 193-207. [https://doi.org/10.1016/S0306-4573\(96\)00062-3](https://doi.org/10.1016/S0306-4573(96)00062-3)
- Schröder, G., Thiele, M., & Lehner, W. (2011). Setting goals and choosing metrics for recommender system evaluations. In *UCERSTI2 workshop at the 5th ACM conference on recommender systems*, Chicago, USA (Vol. 23, p. 53). <https://doi.org/10.1007/s41870-018-0202-4>
- Selvi, C., & Sivasankar, E. (2018). A novel singularity based improved tanimoto similarity measure for effective recommendation using collaborative filtering. In *2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 256-262). IEEE. <https://doi.org/10.1109/CONFLUENCE.2018.8442697>
- Shakirova, E. (2017). Collaborative filtering for music recommender system. In *2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIconRus)* (pp. 548-550). IEEE. <https://doi.org/10.1109/EICONRUS.2017.7910613>
- Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook* (pp. 257-297). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_8
- Shervin Minaee. (2020). 20 Popular Machine Learning Metrics. Part 2: Ranking, & Statistical Metrics, Web Site: <https://towardsdatascience.com/20-popular-machine-learning-metrics-part-2-ranking-statistical-metrics-22c3e5a937b6>, LVD: 13-01-2023.
- Silveira, T., Zhang, M., Lin, X., Liu, Y., & Ma, S. (2019). How good your recommender system is? A survey on evaluations in recommendation. *International Journal of Machine Learning and Cybernetics*, 10(5), 813-831. <https://doi.org/10.1007/s13042-017-0762-9>
- Spearman, C. (1961). The proof and measurement of association between two things, *The American Journal of Psychology* 15 (1904) 72–101. <https://doi.org/10.1037/11491-005>
- Tamburri, D. A., Van Den Heuvel, W. J., & Garriga, M. (2020). Dataops for societal intelligence: a data pipeline for labor market skills extraction and matching. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 391-394). IEEE. <https://doi.org/10.48550/arXiv.2104.01966>

- Vijaysinh Lendave. (2021). How to Measure the Success of a Recommendation System?, in Developers Corner, October 24, , Web address: <https://analyticsindiamag.com/how-to-measure-the-success-of-a-recommendation-system/>, LVD: 13-01-2023. <https://doi.org/10.1016/j.heliyon.2023.e15108>
- Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. *The International Journal of Human Resource Management*, 33(6), 1237-1266. <https://doi.org/10.1080/09585192.2020.1871398>
- Wenxing, H., Yiwei, C., Jianwei, Q., & Yin, H. (2015). iHR+: A mobile reciprocal job recommender system. In 2015 10th International Conference on Computer Science & Education (ICCSE) (pp. 492-495). IEEE. <https://doi.org/10.1109/ICCSE.2015.7250296>
- Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 55(5), 1-37. <https://doi.org/10.1145/3535101>
- Xin, X., Wang, D., Ding, Y., & Lini, C. (2016). FHSM: factored hybrid similarity methods for top-n recommender systems. In *Asia-Pacific Web Conference* (pp. 98-110). Springer, Cham. https://doi.org/10.1007/978-3-319-45817-5_8
- Yao, Y. Y. (1995). Measuring retrieval effectiveness based on user preference of documents. *Journal of the American Society for Information science*, 46(2), 133-145. [https://doi.org/10.1002/\(SICI\)1097-4571\(199503\)46:2<133::AID-ASI6>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1097-4571(199503)46:2<133::AID-ASI6>3.0.CO;2-Z)
- Yi, P., Yang, C., Li, C., & Zhang, Y. (2016). A job recommendation method optimized by position descriptions and resume information. In 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC) (pp. 761-764). IEEE. <https://doi.org/10.1109/IMCEC.2016.7867312>
- Yu, H., Liu, C., & Zhang, F. (2011). Reciprocal recommendation algorithm for the field of recruitment. *Journal of Information & Computational Science*, 8(16), 4061-4068.
- Zhang, B., & Feng, Y. (2016). Improving temporal recommendation accuracy and diversity via long and short-term preference transfer and fusion models. In *Asia-Pacific Web Conference* (pp. 174-185). Springer, Cham. https://doi.org/10.1007/978-3-319-45817-5_14
- Zhao, T., Wuyu, C., & Zhixiang, C. (2021). Summer Job Selection Model Based on Job Matching and Comprehensive Evaluation Algorithm. In 2021 2nd International Conference on Artificial Intelligence and Information Systems (pp. 1-5). <https://doi.org/10.1145/3469213.3470394>