



Tüm Metin Arama Mimarisine ve KNN Sınıflandırma Modeline Dayalı Kişisel Haber Öneri Sistemi Uygulaması

Personal News Recommendation System Application Based on Full Text Search Architecture and KNN Classification Model

Sezgin Seven ^{1*}, **Adil Alpoçak** ²

¹ Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, İzmir, TÜRKİYE

^{2,3} Dokuz Eylül Üniversitesi Fen Fakültesi Fen ve Mühendislik Dergisi, İzmir, TÜRKİYE

Sorumlu Yazar / Corresponding Author *: sezgin.seven@ceng.deu.edu.tr

Geliş Tarihi / Received: 21.08.2019

Kabul Tarihi / Accepted: 30.09.2019

Araştırma Makalesi/Research Article

DOI: 10.21205/deufmd.202022627

Atıf şekli/How to cite: SEVEN, S., ALPOÇAK, A. (2020). Tüm Metin Arama Mimarisine ve KNN Sınıflandırma Modeline Dayalı Kişisel Haber Öneri Sistemi Uygulaması. DEUFMD 22(64), 277-285.

Öz

Bu çalışmada metin tabanlı veriler olan haberler için öneri sistemi tasarlanması hedeflenmiştir. Haber kaynakları olarak internetteki haber kanallarından toplanan içerikler kullanılmıştır. KNN sınıflandırma yaklaşımından üretilen bir algoritma sunulmuş ve haber metinleri üzerinde tüm metin arama mimarisi uygulanmıştır. Öneri sistemi yöntemlerinden içerik tabanlı filtreleme ve işbirlikçi filtreleme yöntemlerinin yanında anahtar kelimeye dayalı öneri sistemi de açıklanmıştır ve uygulanmıştır.

Anahtar Kelimeler: Kişisel Haber Öneri Sistemi, Öneri Sistemleri, Bilgi Getirme

Abstract

In this study, it is aimed to design recommender system for news which is text based data. Recommended news content is collected from online news channels. The algorithm generated from the KNN classification approach is presented and the full text search architecture is applied on news texts. In addition to content-based filtering and collaborative filtering methods, the keyword-based recommendation system is explained and implemented.

Keywords: Personal News Recommendation System, Recommendation Systems, Information Retrieval

1. Giriş

Bilgi günümüzde en önemli değerlerden biri haline gelmiştir. İnternetin en önemli bilgi kaynağı olması ve yaygınlaşması ile birlikte bilgiye ulaşmak oldukça kolay bir hal almıştır. İnternet ile birlikte bilgiler sadece bir kütüphaneden veya kitap sayfalarından ibaret

durumda değildir. Bilgiye ulaşmak, internete bağlı cihazların internet arama motorlarından aranacak birkaç kelime ile mümkündür. Bu işlem kullanıcıların sadece bir kaç saniyelerini alır. Bilginin yayılması da ulaşımı kadar hızlıdır, bir bilgi internet ile tüm dünyaya hızla ve kolaylıkla yayılmaktadır. Bir insanın yayınlayacağı bir bilgi

saniyeler içerisinde tüm dünyaya sunulmaktadır. Tüm insanların bilgiyi bu kadar kolay tüketilmesi de bilginin günden güne ve sürekli olarak artmasını sağlamaktadır. Bu durum insanları devasa bir bilgi yığınıyla karşı karşıya bırakmaktadır. Bu kadar fazla bilgi ve kaynak olması, insanların kendilerine en uygun ve en doğru bilgiye ulaşmalarını zor bir hale getirmiştir. Özgöbek ve Erdur de yayınladıkları çalışmalarında pek çok okuyucunun ilgisini çeken makaleleri bulabilmek için çokça zaman harcadığını belirtmektedir[1]. Bilgiye insanların kendilerinin ulaşmalarının yanında, bilginin insanlara gelmesi bilginin daha ulaşılabilir hale gelmesini sağlayacaktır. Gelen bu bilgiler de rastgele olmayacak ve kullanıcılarını ilgilendirilen ve onlar için uygun ve doğru bilgi olacaktır. Bilginin bu yöntemle getirilebilmesi için belirli bir yapı, düzeni olmayan bu veriler için sistematik bir çalışma yürütülmesi gerekmektedir. Tam bu noktada bu işi bize öneri sistemleri yapmaktadır. Öneri sistemleri ne satın alacağımızı, ne yiyeceğimizi veya kiminle arkadaş olabileceğimizi bize önererek bilgiye ulaşmamızı bir anlamda kolaylaştıran sistemlerdir. Öneri sistemlerinin diğer bir amacı da kullanıcıların herhangi bir çaba sarfetmeden tercih edebileceği ve beğeneceği bilgiye ulaşmalarını sağlamaktır. Günümüzde pek çok alanda uygulanan öneri sistemlerinde birçok yöntem bulunmaktadır. Öneri sistemlerinin kullandığı alana göre uygulanma yöntemleri de değişiklik gösterir. Günümüzde alışverişten, video ve fotoğraf içerikli uygulamalara kadar öneri sistemleri yaygınlıkla kullanılmaktadır[1]. Bu alanların yanında haber tavsiye etmek amacıyla da haber sitelerinde öneriler yapıldığı sıklıkla görülmektedir. Ancak kişisel olarak haber önerisi yapan sistemler pek yoktur. Bu çalışmada da haberlerin kişisel olarak tavsiye edilebilmesi ve kullanıcıların önerilen haberleri okumaları hedeflenmiştir. Çünkü önerilen haberler ne kadar fazla okunursa, öneri sistemi o kadar başarılı olmuş demektir[2]. Makalenin ikinci bölümünde öneri sistemlerinden ve öneri sistemlerinde kullanılan yöntemlerden bahsedilmiştir. Burada içerik tabanlı öneri sistemi, işbirlikçi filtreleme ve karma yaklaşım metodlarından ve bunların projeye uygulama yöntemlerinden bahsedilmiştir. Aynı zamanda ikinci bölümde bizim yeni bir yaklaşım getirdiğimiz metod olan anahtar kelimeye dayalı öneriden bahsedilmiştir. İkinci bölümde anlatılan yöntemleri kullanarak geliştirdiğimiz

haber öneri sistemi algoritması sunulmaktadır. Makaleler arasındaki benzerliklerin nasıl hesaplandığı da bu bölümde anlatılmıştır. Bulgular bölümünde bu çalışmayla birlikte paralel olarak geliştirilen haber öneri, okuma ve paylaşım sistemi olan OKURSS uygulamasından ve öneri sistemi algoritmasının OKURSS uygulaması üzerinde nasıl çalıştırıldığından bahsedilmiştir.

2. Materyal ve Metot

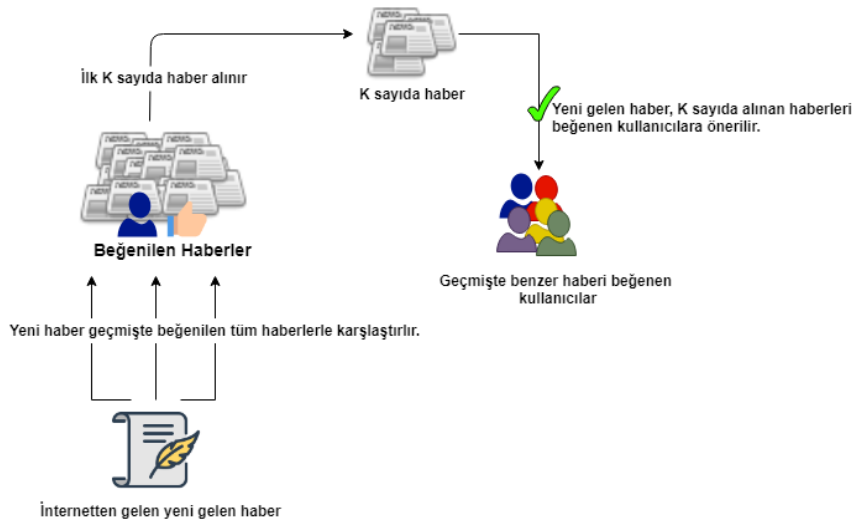
Her sistemde ve uygulamada olduğu gibi öneri sistemlerinin de farklı uygulama yöntemleri ve algoritmaları bulunmaktadır. En çok bilinen ve kullanılan yöntemler içerik tabanlı filtreleme, işbirlikçi filtreleme ve karma yaklaşım modeli yöntemleridir. Bu yöntemler yoğun olarak içerik benzerliklerini ve kullanıcı profillerini ele alır[2]. Haberler söz konusu olduğunda metin işleme ile ilgili yöntemlerin uygulamaları gerekli olacaktır. Yapılandırılmamış metinler düzensiz yapıları, sayısal olarak bir karşılıklarının olmaması, matematiksel düzlemde herhangi bir şekilde konumlandırılmamaları ve yapısal hale getirilmesindeki zorluklardan dolayı, veri madenciliği ve metin madenciliği bilgisi de bunları uygulayabilmek için bilinmesi gereken öncelikli konulardır[3].

2.1. İçerik tabanlı filtreleme

Bu yöntemde kullanıcının geçmişte okuduğu içeriklere ve tercihlerine dayanarak öneri işlemini gerçekleştirdik. İçerik tabanlı filtreleme yaklaşımını farklı metodlarla uygulayan çalışmalar mevcuttur. Örneğin, Servet Taşçı çalışmasında kullanıcıların tercihlerini belirleyip daha sonra tavsiye edilebilecek nesnelerin kullanıcı ile ilişkisini kurarak, nesne kullanıcı matrisiyle tavsiye sistemini oluşturmuşlardır[2]. Başka bir çalışmada Liu ve Dolan farklı bir yaklaşım getirerek kullanıcıların makaleler üzerinde durdukları süreleri hesaplamışlardır[4]. Pazzani çalışmasında kullanıcıların son okuduğu haberlere dikkat ederek öneri işlemini gerçekleştirmiştir[5]. Suchal ve Navrat yayınladıkları çalışmada [6] önerilecek dokümanları ve kullanıcıları birer sanal eleman olarak tanımlayıp, doküman-kullanıcı kümeleri oluşturarak birbirleriyle benzer olan kullanıcıları bulup ve benzer kullanıcılar üzerinden dokümanlara ulaşım öneri işlemini gerçekleştirmişlerdir. Çalışmalarındaki tüm metinsel işlemleri bizim çalışmamızda da kullandığımız Full Text Search kütüphanesi ile

yapmışlardır. Bu makalede anlattığımız ve uyguladığımız çalışmada ise yeni gelen bir haberi kullanıcıların daha önce beğendiği haberlerle karşılaştırdık. Yeni gelen habere en çok benzeyen haberleri okuyan kullanıcı veya kullanıcılara yeni gelen bu haberi önerdik. Öneri algoritmasını geliştirirken k -NN (k-nearest neighbor) sınıflandırma yaklaşımından yola çıktık. k -NN algoritması sınıflandırılmak istenen yeni bireyi, daha önceki bireylerden K tanesine olan yakınlığına bakarak bulmaktadır[7]. Tam da bu noktada her bir kullanıcıyı, bir sınıf olarak

düşündük. Her bir sınıfa ait elemanlar kullanıcıların daha önce beğendiği makalelerden oluşmaktadır. Bu aşamada yapmamız gereken yeni gelen haberin hangi sınıfa ait olacağını bulmaktır. Bu sebeple yeni gelen haberi kullanıcıların daha önce beğendiği haberlerle tek tek karşılaştırdık. En çok benzeyen ilk k haberi aldık. Buradaki k sayısı tarafımızca belirlenen en optimal bir sayıyı belirtmektedir. Alınan ilk k haberi beğenen kullanıcılara yeni haberi önerdik. Burada anlatılanlar aşağıdaki Şekil 1’de gösterilmiştir.



Şekil 1. İçerik Tabanlı Filtreleme Algoritması

Makaleler arasındaki benzerlikleri cosine similarity yaklaşımı ile PostgreSQL Full Text Search kütüphanesini kullanarak hesapladık. Cosine similarity N boyutlu iki vektör arasındaki açının \cos ile ifade edilmesidir ve sıklıkla doküman karşılaştırmalarında kullanılmaktadır [8]. Başak, Türkçe dokümanların benzerliğini tüm yaklaşımlar bazında makalesinde ele almıştır[8]. 4 adet benzerlik yöntemini java platformunda geliştirdiği uygulama ile gerçekleştirmiştir. Türkçe kelimeleri köklerini ayırtmak için zembek kütüphanesini kullanmıştır. Yaptığı çalışmada en iyi ayrımı cosine similarity göstermiştir. Bielikova ve Kompan de çalışmalarında cosine similarity en etkili arama yaklaşımı olduğunu belirtmişlerdir[9]. Hofmann da yayınladığı Learning of Similarity of Documents [10]

makalesinde dokümanlar arasındaki benzerlikleri bulurken sadece bag of words dediğimiz kelime haznesine odaklanmıştır. Latent Class Analysis olarak adlandırdığı yöntem ile doküman koleksiyonlarını ayırıştırıp modellemiştir. PostgreSQL Full Text Search kütüphanesi metin madenciliğinin gerekli olan metodlarını ve veri yapılarını sunmaktadır[11]. Biz metin madenciliği ile ilgili stop-word çıkarımı, kelimelere VSM(vector space model) hesaplanması, makale benzerlikler arasında rank hesaplanması gibi işlemleri PostgreSQL Full Text Search kütüphanesi ile yaptık.

2.2. İşbirlikçi filtreleme

Bu yöntemde kullanıcıları ve haberleri gruplandırarak, bu gruplar arasındaki

benzerlikleri bulduk ve benzer kullanıcılara benzer türde haberleri önerme işlemini gerçekleştirdik. İşbirlikçi filtreleme, diğer kişilerin görüşlerini kullanarak öğeleri filtreleme veya değerlendirme sürecidir[12]. Literatürde işbirlikçi filtreleme yöntemi ile uygulanan çalışmalar mevcuttur. Bunlardan StreamRec [13] sisteminde kullanıcılar okudukları benzer haberlere göre gruplandırılmıştır ve bir haberi tekrar mı yoksa ilk kez mi okuduğu üzerinde durulmuştur. Eğer tekrar bir haberi okuyorsa buna göre bir değerlendirme yapıp kullanıcı grupları yeniden oluşturulmuştur. Saranya ve Sadhasivam yayınladıkları makalelerinde kişiselleştirilmiş haber öneri sistemlerini kullanıcı profillerini sürekli olarak dinamik şekilde güncelleyerek işbirlikçi filtreleme yöntemini uygulamışlardır[14].

Oluşturduğumuz sistem içerisinde iş birliği filtrelemeye katkı sağlayabilmesi adına kullanıcıların birbirleri arasında arkadaşlık kurabildiği ve birbirlerinin okudukları, beğendikleri haberleri görebildikleri bir yapı da mevcuttur. Bu yapı göz önünde bulundurulduğunda birbirine benzer kullanıcılar da ortaya çıkacaktır. Benzer kullanıcı grupları da benzer haberleri okuyup beğendiklerinden, muhtemelen bir sonraki seferde de benzer haberleri okuyup beğeneceklerdir. Bu bağlamda benzer kullanıcı gruplarına aynı haberler önerilebilir. Bir önceki bölümde içerik tabanlı filtrelemede anlatılan yöntemle sürekli olarak aynı türde haberler de önerildiğinde kullanıcıların bu benzer haberleri okumaktan sıkılacaklarını da düşünebiliriz. Bu yüzden benzer kullanıcılara yapılan öneriler bir çeşitlilik yaratacak ve okumaya olan bağımlılığı arttıracaktır. Ancak yine de bu çalışmada işbirlikçi filtreleme ile yapılan önerilerin oranı içerik bazlı filtrelemeye göre oranı daha az tutulmuştur.

İşbirlikçi filtreleme yaklaşımında iki farklı teknik mevcuttur. Bu teknikler kullanıcı-bazlı filtreleme ve haber-bazlı filtrelemedir. Kullanıcı bazlı filtrelemede benzer kullanıcılar bir küme olarak kabul edilmiştir. Bu küme içerisindeki kullanıcılara benzer haberler önerilir. Haber bazlı filtrelemede ise kullanıcı bazının tersine haber üzerinden yola çıkılarak benzer haberleri okuyan kullanıcılar, günlük olarak analiz edilerek kullanıcı grupları yeniden oluşturulmaktadır. Böylece öneri sistemi her gün daha akıllı öneriler yapan bir sistem haline alır.

Bizim işbirlikçi filtreleme yöntemindeki algoritmamızda benzer kullanıcıları bulmak için öncelikle her bir kullanıcının her bir kategorideki okunan makale sayıları getirilir. Örneğin, A kullanıcısı X kategorisinde 156 haber, C kategorisinde 200 haber gibi. Her bir kullanıcı arasında kategoriler ve bu kategorilerin okunma sayıları alınır. Karşılaştırma işlemi sırasında aynı kategoride okunan haberler için eşleşen kullanıcılar bulunduğu, o kategorideki okunma sayıları oranlanır. Örneğin, A kullanıcısı X kategorisinde 100 haber okumuş olsun, C kullanıcısı da X kategorisinde 50 haber okumuş olsun. Bu durumda eşleşen kullanıcılar arasında aynı kategorideki haber okuma oranı 50/100 yani 0.5 olacaktır. İkili kullanıcılar arasında bulunan bu oranlar her kategori için toplanmaktadır. Sonuçta ise toplam oranları en yakın olan kullanıcılar birbirlerine en yakın okuma tarzına sahip kullanıcılar olacaktır. Yeni gelen bir haber tavsiye edilirken benzer kullanıcılara benzer kategorilerde haberler tavsiye edilecektir.

Kullanıcı gruplarını oluşturan sistemi her 15 dakikada bir yenilenecek şekilde tasarladık. Gün içerisinde okunan haberler sürekli bir değişim içinde olacağından bu grupların sürekli kendini yenilemesi ve dinamik kalması algoritmanın daha sağlıklı çalışmasını sağlayacaktır. Kullanıcı grupları her gün defalarca yeniden oluşacak ve sonunda doğruya en yakın hale gelecektir. Algoritma-1 önermiş olduğumuz işbirlikçi filtreleme algoritmasını göstermektedir.

2.3. Anahtar kelimeye dayalı öneri

Uyguladığımız bu yöntemde ise kullanıcılar kendilerine en yakın konulara ve kategorilere göre anahtar kelimeler girmektedir. Makalelerin indirilmesi aşamasında her bir doküman için aday anahtar kelimeler bulunmaktadır. Yasin Uzun, Keyword Extraction Using Naïve Bayes [15] makalesinde tüm kelimelerin bir belgede anahtar kelime olmak için önceden aynı şansa sahip olmadığını açıklamaktadır. Kelimeler metinde sıklıkla geçse bile anahtar kelime olarak işaretlenmesi gerektiğini söylemektedir. TF-IDF modeline engel olabilecek, alfanumerik olmayan karakterlerin ve sayıların kaldırılması gerektiğini anlatmaktadır. Bizim çalışmamızda anahtar kelimeler makalede en sık geçen kelimeler hesaplanarak bulunmaktadır. Hesaplamalar her bir kelime sayılarak yapılmaktadır. Sayılan bu kelimeler gruplandırılarak her birinin sayısı bulunur ve

bunlar en çok sayıdan aza doğru sıralanır. İlk 10 kelime alınıp diğerleri çıkartılır. Bu 10 kelime en önemli kelimeleri ve dolayısıyla haberin kategorisini belirleyen kelimelerdir. Aynı zamanda anahtar kelime hesaplaması yapılmadan önce boş kelimeler (stop word)

çıkarıldı. Çıkarılan bu kelimeler kullanıcıların girmiş kelimeler ile karşılaştırılır. En çok eşleşen kullanıcılara makale önerme işlemi yapılmaktadır.

```
function find_similar_users_by_collaborative_algorithm()
    reading_categories
    matched_users
    restricted_users
    score <- 0
    sum_value <- 0

    FOR each main_item on the reading_categories
        IF restricted_users not contain main_item.user_id THEN
            score <- 0
            FOR each sub_item on the reading_categories
                IF sub_item.user_id not equal main_item.user_id THEN
                    sum_value <- 0
                    FOR each main_category on the main_item.categories
                        FOR each sub_category on the sub_item.categories
                            IF main_category.id equal sub_category.id THEN
                                sum_value += get_rate(sub_category.count_read,
                                                        main_category.count_read)
                                break
                            END IF
                        END LOOP
                    END LOOP
                    IF score less than sum_value THEN
                        score.user_id <- sub_item.user_id
                        score.value <- sum_value
                        restricted_users.push(score.user_id)
                    END IF
                END IF
            END LOOP
            matched_users.push(main_item.user_id)
        END IF
    END LOOP

    return matched_users
end function
```

Algoritma-1: İşbirlikçi filtreleme

2.4. Haber öneri sistemi algoritması

Bizim çalışmamızda uygulanan tavsiye sistemi algoritması temel olarak hibrid yaklaşım üzerine kuruludur. Yapılan öneri sistemi çalışmalarında farklı algoritma yaklaşımları uygulanmıştır. Bunlardan online olarak öneri yapılan çalışmada haberlere ait tavsiye algoritmaları analiz edilmiştir.[16] Birkaç öneri sistemini ele almışlardır. Bunlardan birinde popüleritelerine göre öneriler yapılmıştır. Haberlerin popüleriteleri saatlik, günlük, haftalık, aylık olarak sınıflandırılmış. Bir de bu çalışmada bizimkine benzer şekilde kullanıcı tarafından okunan makaleler tekrar tavsiye edilmemiştir. Haberler veri madenciliği için işlenirken başlıklarına, özetlerine, anahtar kelimelerine göre ayrılıp incelenmiştir. Tavsiye edilecek haberler bir ranking algoritması ile getirilmiştir. Luostarinen ve Kohonen topic modeller üzerine

içerik bazlı öneri sistemi geliştirmişlerdir[17]. Naïve Bayes, k -NN ve Latent Dirichlet Algorithm(LDA) kullanmışlardır. Bu üç modelden en kötü performansı Naïve Bayes en iyiyi k -NN gösterdiğini tespit etmişlerdir. Aynı zamanda bu çalışmada [17] soğuk başlangıç problemi simülasyonunda, LDA'nın tüm yöntemler için istatistiksel olarak önemli gelişmeler sağladığı görülmüştür. Hot news recommendation system [18] çalışmasında da o gün yayınlanan en sıcak ve en dikkat çekici haberler önerilmektedir. Bu öneri sisteminde kişilerle veya profillerle ilgilenilmiyor. Sadece haber başlıklarındaki en önemli kelimeler seçilerek bayesian sınıflandırma algoritması ile en sıcak 20-50 sayısı haber bulunuyor. Ozgobek, Gulla ve Erdur yayınladıkları çalışmada [19] öneri sistemlerinde karşılaşılan zorlukları ele almışlardır. Bunlardan en önemlileri soğuk

başlangıç, yenilik, kullanıcıların ilgilerinin değişmesi, yapısal olmayan içerikler ve gizlilik problemleridir. Soğuk başlangıç problemi yeni bir kullanıcı hakkında herhangi bir bilgi sahibi olmadan öneride bulunmaktadır. Biz OKURSS projesinde bu problemi başlangıç aşamasında ilgisini çekebilecek haber kategorilerini seçmelerini sağlayarak aştık. Yenilik probleminde ise kullanıcıların güncel, yeni haberler okuma isteğidir. OKURSS'da her 15 dakikada bir haber kaynaklarından yeni içerikler toplayarak bu problemi çözmüş olduk. Kullanıcıların ilgilerinin değişmesine tam olarak bir çözüm olmasa da her konuda en önemli haberler önerilerek bu soruna bir çözüm olabilir. Yapısal olmayan içerikleri analiz etmek yapısal içeriklere göre oldukça zordur ve ekstra iş yükü getirmektedir. Bu da zorlukların en önemliler arasında yer almaktadır. Bu problemin çözümü için de metin madenciliği yöntemlerinden faydalanılmıştır.

Bizim tavsiye sistemi algoritmamız temelde her üç öneri yöntemini de kullanarak tavsiye sistemini çalıştıracak ve haberleri getirecek veri setlerini oluşturur. Üretilen tavsiye sistemi yaklaşımına etki eden parametreler öneri sistemine farklı derecedeki çarpanlarla etki ederler. Yani içerik tabanlı filtreleme en önemli etkiye sahip iken, bunu işbirlikçi filtreleme ve anahtar kelime yöntemi takip eder. Aşağıdaki formülasyonda t_x , t_y ve t_z olarak gösterilen içerik tabanlı filtreleme, işbirlikçi filtreleme ve anahtar kelime yöntemi ile getirilen haber listeleri farklı katsayılar(p_1, p_2, p_3) ile RSS (reading and sharing system) listesine eklenmektedir. RSS olarak gösterilen liste tüm öneri sistemleri kullanılarak oluşturulan haber listesidir. OKURSS kullanıcılara bu haber listesi gösterilecektir.

$$p_1, p_2, p_3 : N,$$

$$RSS : \{h_1, h_2, \dots, h_n\}(\text{öneri Sonuç Listesi})$$

$$t_x : \{m_1, m_2, \dots, m_n\}(\text{içerik bazlı filtreleme})$$

$$t_y : \{m_1, m_2, \dots, m_n\}(\text{işbirlikçi filtreleme})$$

$$t_z : \{m_1, m_2, \dots, m_n\}(\text{anahtar kelimeye dayalı})$$

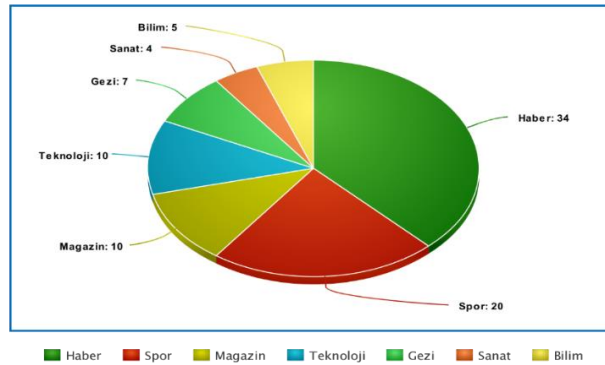
$$RSS = p_1.t_x + p_2.t_y + p_3.t_z \quad (1)$$

3. Bulgular

Çalışmada anlatılan haber tavsiyesi algoritmasına göre belirli kategorilerde haberler okuyacak şekilde kullanıcılar ile çalışılmıştır. Her gün bu kullanıcılar kendilerine belirtilen kategoriler ağırlıklı olacak şekilde haberler okumuştur. Örnek olarak seçilen 6 kullanıcıya ait 1 ayda en çok okunan haber kategorileri ve okuma sayıları tablodaki gibidir. Günlük olarak en çok tavsiye edilen haberler Grafik 1'de gösterilmiştir. Bu sonuçlar tavsiye sisteminin veritabanına yazdığı kayıtlardan alınarak çıkarılmıştır.

UserName character varying (100)	Category character varying (20)	max bigint
Adil	Haber	94
Emre	Haber	84
Eren	Spor	255
Görkem	Spor	249
Okan	Teknoloji	17
Sezgin	Teknoloji	15

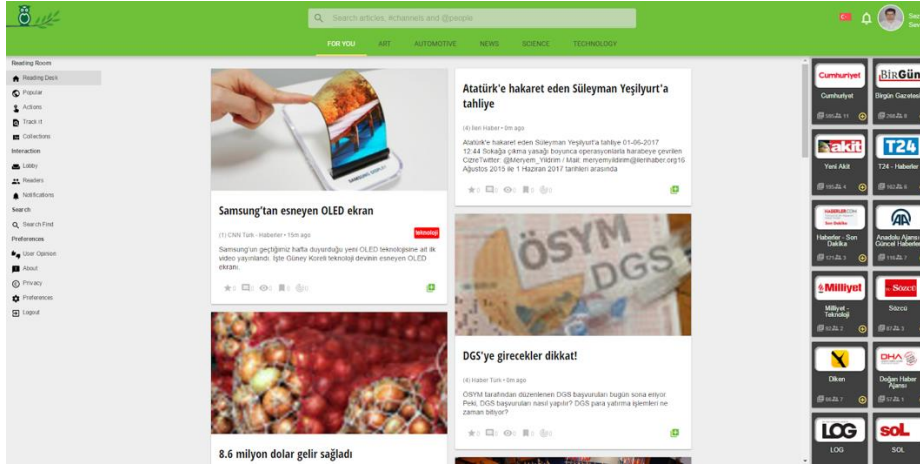
Şekil 1. 1 aylık haber okuma veritabanı kayıtları



Grafik 1. Kategorilere göre haber öneri dağılımı

Öneri sistemi algoritması OKURSS haber okuma projesi ile hayata geçirilmiştir. Bu makale çalışması ile paralelde yürütülen OKURSS uygulaması, okuyucusunu daha önceden okuyup beğendiği yayınlara bakarak onu tanıyan, internetteki yeni yayınları okuyucuları için tarayan ve okuyucularının beğenebileceği

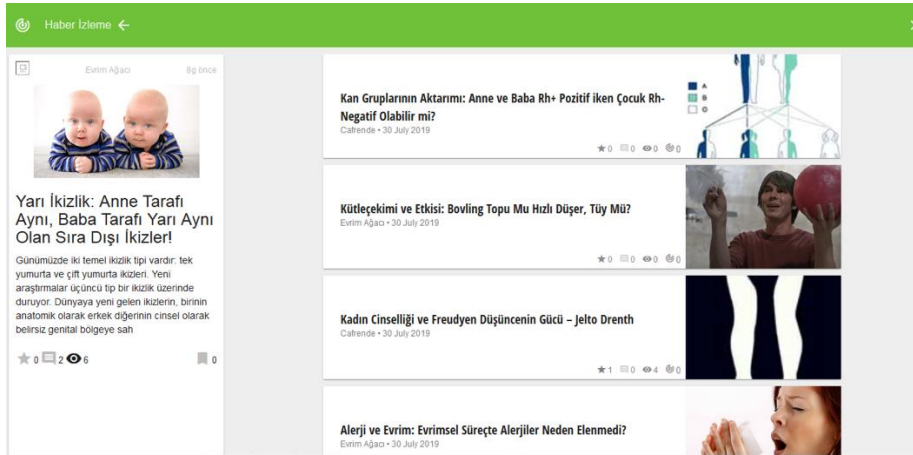
yayınları seçerek okumayı kolaylaştıracak yalın bir formatta okuyucusuna sunan, okuyucuların yayınlara yorumlar eklemesine ve arkadaşları ile paylaşmasını sağlayan, İnternet Bilgi Erişim, Okuma ve Paylaşım Sistemidir. OKURSS projesine ait bir ekran görüntüsü Şekil 2'de belirtilmiştir.



Şekil 2. OKURSS Ana Sayfa

Haber içerikleri arasındaki benzerlikleri karşılaştıran algoritma OKURSS uygulamasında haber izleme özelliği ile test edilmiştir. Haber izleme ekranını ile izlemek istediğiniz haberlere benzer diğer haberler getirilmektedir. İzleme

işlemini yapmak için haber kartında sol tarafta bulunan en sağa dayalı butona basılır. İzleme sonucu benzer haberlerin geldiği ekran Şekil 3'de gösterilmiştir.



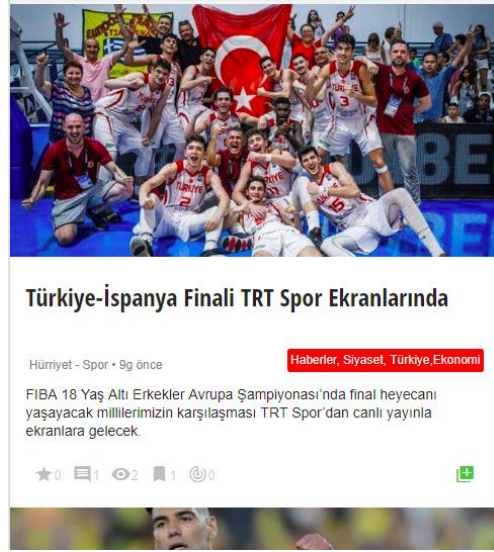
Şekil 3. OKURSS Haber İzleme

Tavsiye sistemi algoritması yöntemlerine göre bu Şekil 1'deki tabloda gösterilen veriler oldukça önemlidir. Çünkü içerik bazlı tavsiye sisteminde kullanıcının geçmişte okuduğu haberler el anılır. Aynı zamanda işbirlikçi filtreleme yöntemine göre birbirleriyle benzer haberler okuyan kullanıcıların tespiti için de okunan benzer haberlere bakılıp, benzer okuma tarzına sahip insanlara, benzer haberler tavsiye edilecektir. Buna göre aşağıdaki kullanıcılar en benzer haber okuma tarzına sahip kişilerdir. Eşleşen isimler ile birlikte gösterilen değer, kullanıcıların okuma türü anlamında birbirlerine olan yakınlık derecelerinin sonucudur. Bu eşleşmeler en başta belirlenen senaryoda beklenen sonuçlardır. Skor alanı 10 üzerinden kullanıcıların yakınlık derecelerini göstermektedir. Tablo 1'de ilk senaryoda en çok okunan kategorilerde görüldüğü gibi, aşağıdaki işbirlikçi filtreleme yöntemiyle hesaplanan sonuçta yakın kullanıcılar benzer kategorileri okumaktadır. Bu da tavsiye sisteminin doğru çalıştığının sağlamasıdır.

Tablo 1. Kullanıcı okuma türü yakınlıkları

Eşleşen Kullanıcılar	Eşleşme Skoru
Sezgin - Okan	6,06
Adil - Emre	5,31
Eren Görkem	6,12

Anahtar kelimeye dayalı öneri sisteminde kullanıcılar OKURSS platformundan kendilerine en uygun anahtar kelimeleri girmektedir. Haber tavsiye sistemi çalışırken bu anahtar kelimelerin haberlerde kullanılma sıklıklarına göre haber tavsiyesi gerçekleşir. Anahtar kelimeler sonucunda önerilen haberlerden bir görüntü Şekil 4'de gösterilmektedir. Anahtar kelimeler kırmızı çerçeveli yazı ile belirtilmektedir.



Şekil 4. Anahtar Kelimeye Dayalı Haber Öneri Kartı

4. Tartışma ve Sonuç

Bu makalede kişisel haber öneri sistemi açıklanmıştır. Bilginin kullanıcılara ulaşması, kullanıcıların bilgiye ulaşması kadar kolay olmasa da uygulanabilir yöntemleri vardır. Özellikle alışveriş ve görsel medya sektörlerinde yaygın olarak kullanılan öneri sistemleri bu alandaki büyük şirketlerin daha çok kullanıcı çekebilmek adına en önemli silahlarından biridir. Ve bu konuda da oldukça başarılı oldukları söylenebilir. Öneri sistemleri haberler ve yazılı içerikler adına çok yaygın olmasa da çok önemli bilgilerin elde edilmesini sağlar. İnsanların okuma alışkanlıkları diğer bir çok şeye olan ilgisini de gösterebilir ve kişilik analizleri buna göre yapılabilir. Aynı zaman bu sistemin okuyuculara da oldukça katkısı vardır. İnsanlara sürekli beğenecekleri makaleler sunmak, onların okuma alışkanlıklarını arttıracaktır. Gelecek çalışmalar da OKURSS uygulamasını tüm haberlerin tek bir yerde toplandığı ve yayınlandığı en önemli, en gerçek haber kaynağı yapmak olacaktır. Aynı zamanda kişisel öneri sistemi algoritmasını daha da geliştirerek daha akıllı öneriler yapılması sağlanacaktır.

Teşekkür

Bu çalışmayla birlikte geliştirilen OKURSS projesi Tubitak 1512 girişimcilik programının 2150289 proje numarası ile desteklenmiş ve Lemonsoft Ltd. Şti.' de geliştirilmiştir.

OKURSS projesine liderlik eden sayın hocam Doç. Dr. Adil Alpkoçak'a ve büyük desteklerinden dolayı sayın Dr. Okan Öztürkmenöglü'na sonsuz teşekkürlerimi sunarım.

Kaynakça

- [1] Özgöbek, Ö., Gulla, J. A., & Erdur, R. C. 2014. A Survey on Challenges and Methods in News Recommendation. In WEBIST, s. 278-285.
- [2] Taşçı, S., 2015. İçerik Bazlı Medya Takip ve Haber Tavsiye Sistemi. Hacettepe Üniversitesi, Bilgisayar Mühendisliği, Yüksek Lisans Tezi, 70s., Ankara.
- [3] Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. 2017. A brief survey of text mining: Classification, clustering and extraction techniques.
- [4] Liu, J., Dolan, P., & Pedersen, E. R. 2010. Personalized news recommendation based on click behavior. In Proceedings of the 15th international conference on Intelligent user interfaces s. 31-40, ACM.
- [5] Pazzani, M. J., & Billsus, D., 2007. Content-based recommendation systems. In The adaptive web, s. 325-341. Springer, Berlin, Heidelberg.
- [6] Suchal, J., Navrat P., 2010. Full Text Search Engine as Scalable k-Nearest Neighbor Recommendation System, In IFIP International Conference on Artificial Intelligence in Theory and Practice, s. 165-173, Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-15286-3_16.
- [7] Yong, Z., Youwen, L., & Shixiong, X. (2009). An improved KNN text classification algorithm based on clustering. Journal of computers, 230-237.
- [8] Başak, S., 2009. Türkçe Dokümanların Benzerliği, <https://www.selcukbasak.com/download/TurkceDokumanBenzerligi.pdf> (Erişim Tarihi: 12.05.2018).
- [9] Kompan, M., & Bieliková, M. 2010. Content-based news recommendation. In International conference on electronic commerce and web technologies s. 61-72. Springer, Berlin, Heidelberg. DOI : 10.1007/978-3-642-15208-5_6.
- [10] Hofmann, T., 2000. Learning the similarity of documents: An information-geometric approach to document retrieval and categorization. In Advances in neural information processing systems, s. 914-920.
- [11] Mukti, S. 2018. When to use PostgreSQL Full Text Search and Trigram Indexes. <https://www.medium.com/@sukorenomw/when-to-use-postgresql-full-text-search-and-trigram-indexes-4c7c36223421> (Erişim Tarihi : 20.01.2019).
- [12] Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S., 2007. Collaborative filtering recommender systems. In The adaptive web s. 291-324. Springer, Berlin, Heidelberg.
- [13] Chandramouli, B., Levandoski, J. J., Eldawy, A., & Mokbel, M. F., 2011. StreamRec: a real-time recommender system. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data s. 1243-1246. ACM
- [14] Saranya, K. G., & Sadhasivam, G. S., 2012. A personalized online news recommendation system. International Journal of Computer Applications.
- [15] Uzun, Y., 2005. Keyword extraction using naive bayes. In Bilkent University, Department of Computer Science, Turkey www. cs. bilkent. edu. tr/~guvenir/courses/CS550/Workshop/Yasin_Uzun. Pdf
- [16] Doychev, D., Lawlor, A., Rafter, R., & Smyth, B. (2014, September). An analysis of recommender algorithms for online news. In CLEF 2014 Conference and Labs of the Evaluation Forum: Information Access Evaluation Meets Multilinguality, Multimodality and Interaction, 2014, Sheffield, United Kingdom s. 177-184.
- [17] Luostarinen, T., & Kohonen, O., 2013. Using topic models in content-based news recommender systems. In Proceedings of the 19th Nordic Conference of Computational Linguistics, s. 239-251.
- [18] Xia, Z., Xu, S., Liu, N., & Zhao, Z., 2014. Hot news recommendation system from heterogeneous websites based on bayesian model. The Scientific World Journal, 2014.. DOI: 10.1155/2014/734351.
- [19] Özgöbek, Ö., Gulla, J. A., & Erdur, R. C., 2014. A Survey on Challenges and Methods in News Recommendation. In WEBIST, s. 278-285.