

Hukuk Metinleri iin Anahtar Kelime Kullanımı ile Otomatik zetleme

Erol GDUR^{1, a, }

¹Bitlis Eren niversitesi, Tatvan Meslek Yksekokulu, Bilgisayar Teknolojileri Blm, Bitlis, Trkiye
^aORCID: 0000-0003-3356-4980

 Sorumlu Yazar: egodur@beu.edu.tr

zet: Mahkemelerde, benzer davalar iin nceki mahkemelerin verdiĐi kararlar mevcut mahkemenin karar verme srecinde baĐlayıcı bir etkiye sahiptir. Bu nedenle avukatlar yer aldıkları davalar ile benzerlik ieren davalara ve sonulara ulařmak iin arařtırma yaparlar. Hukuki metinlerde aranan duruma iliřkin farklı emsal kararların varlığı, kullanıcının yaptıĐı aramaya gre birden fazla mahkeme sonu metninin incelenmesini gerektirir. Metin sayısının fazla olması ve metinlerin yazı miktarlarının byk olması, kullanıcı tarafından nemli bir zaman ve emek harcanmasına neden olur. Kararlarda zet metinlerin kullanılması, bu sreyi ve abayı daha makul bir kullanım dzeyine indirecektir. Bu nedenle, bu alıřma, bu yasal metinler iin otomatik bir zetleme sistemi geliřtirmeyi amalamaktadır. Bu amala zet ıkarımı ile anahtar kelime frekans temelli zetleme, ynsz izge kullanımı ile zet ıkarma, aĐırlıklı izge kullanımı ile zet ıkarma ve frekans temelli zetleme ile izge tabanlı zetlemenin hibrit olarak kullanıldıĐı karma model uygulanmıřtır. Bu modellerden en bařarılısı karma model olarak belirlenmiřtir.

Anahtar szckler: DoĐal dil iřleme, Otomatik zet ıkarma, Hukuki metinler.

Automatic Summarization with Keyword for Legal Texts

Abstract: In legal cases, the decisions made by the previous courts for similar cases affect the determination of current courts. Therefore, lawyers do research to reach similar cases and results. The existence of different precedent decisions regarding the situation sought in the legal texts requires the examination of more than one court result text according to the search made by the user. The high number of texts and the high size of the texts cause a significant time and effort to open by the user. The use of summary texts for decisions will reduce this time and effort to a more reasonable level of use. Therefore, this study aim to develop an automatic summarization system for these legal texts. For this purpose, frequency base, undirected graph base, weighted graph base summarization extraction methods have applied. In addition, a hybrid method that consists frequency base and weighted graph base summarization has applied. This method has best results.

Keywords: Natural language processing, Automatic summarization, Legal texts.

1. Giriř

Hukuki metinlerde aranan durum ile ilgili farklı emsal kararının bulunuyor olması kullanıcının yapmıř olduĐu arama sonucuna gre birden ok karar metninin incelemesini gerektirmektedir, metinlerin sayısının fazla olması ve metin boyutların yksek olması kullanıcı aısında nemli bir zaman ve emek kullanımına neden olmaktadır. Kararlar iin zet metinlerin oluřturulması bu zaman ve emek kullanımı daha makul seviye indirecektir. Mevcut durumda Trke hukuki metinler iin otomatik zetleme yapan bir sistem bulunmadıĐından dolayı metinler iin zetleme iřlemi ancak uzman kiřiler tarafında manel olarak gerekleřtirilmektedir. Emsal karar sayısının fazla olması mevcut sistemlerde kararların tamamın zetlemesini bu sebeple engellemektedir. Otomatik zetleme sistemi ile bu iřlemi daha makul srelere indirgemek ve emek israfını nlemek mmkn olacaktır. Bu sebeple bu alıřma kapsamında hukuki metinlerin otomatik zetlemesinin amalanmaktadır. Bu kapsamda hukuki metinler zerinde uygulanmak zere eřitli otomatik zetleme teknikleri ilgili arařtırmalar ve uygulamalar gerekleřtirilmiřtir.

Otomatik zet ıkarma sistemleri kullandıkları dilsel bilgi miktarına gre genel olarak abstraktif ve ekstraktif olarak ikiye ayrılabilir. Abstraktif sistemlerde orijinal metin zerinde yorumlama yapılarak

çıkarılan bilgilerden yeniden bir metni oluşturulacak özetleme işlemi gerçekleştirilmektedir. Ekstraktif sistemlerde ise orijinal metnin cümlelerini veya cümle parçalarını kullanarak metin için bir özet oluşturmaktadır. Bu sistemlerde işlemin en önemli adımı özet metnini oluşturacak olan, asıl metinden seçilen cümlelerin belirlenmesi işlemidir. Seçilen cümleler anlamlı bir bütün oluşturacak şekilde bir araya getirilmektedir. Ekstraktif sistemleri Abstraktif sistemlerden ayıran en önemli özellik oluşturulacak olan özet metnin yeniden oluşturulması yerine mevcut metin parçalarını kullanıyor olmasıdır. Abstraktif sistemlerin uygulanması, bilgi çıkarımı için orijinal metnin yorumlamasını gerektiriyor olması ve özet metnin doğal dil oluşturma (Natural language generation) teknikleri ile yeniden oluşturulması gibi sebeplerden dolayı daha zordur. Türkçe için bir WordNet bulunmuyor olması, sözcük ilişkilerinin (lexical chains) elde olmaması, Türkçe metinlerde etkili yorumlama ve bilgi çıkarımı yapan bir sistemin mevcut olmayışı ve metin oluşturma için Türkçe doğal dil üretiminin mevcut durumda olmayışından dolayı Türkçe metinler için Abstraktif bir sistem oluşturulması için gerekli alt yapılar mevcut durumda bulunmamaktadır. Bu sebeple bu çalışmada hukuki metinler için otomatik özet oluşturulması için Ekstraktif özetleme teknikleri kullanılmaktadır. Geliştirilen sistem, metin içinde önemli cümleleri belirleyerek bu cümlelerden metin için tutarlı bir özet çıkarmayı temel almaktadır. Sistemin işleyişinin en önemli adımları özeti oluşturacak cümlelerin seçilmesi ve seçilen cümlelerin bir kısmının kullanılması ile özet metnin oluşturulması işlemidir.

Hukuki metinler için otomatik özetleme sistemi geliştirilirken farklı yöntemler çeşitli metrikler değiştirilerek denenmiştir. Denenen yöntemler ve metrikler aşağıda detaylandırılacaktır. Sistem geliştirilmesi aşamasında sistem çıktılarının değerlendirilmesi için metin içinde geçen davanın konusu, davanın sonucu, tarafların iddia ve savunmaları, Yargıtay kararı, yerel mahkemenin kararı gibi bölümler daha önceden belirlenmiş olan Yargıtay'ın internet üzerinden yayınladığı, 200 farklı doküman kullanılmıştır (Yargıtay, 2021). Sistem çıktılarının başarı durumları metnin tutarlılığı ve akıcılığıyla beraber metin için seçilen yapıların ne kadarını içerdiğine göre değerlendirilmiştir.

2. İlgili Çalışmalar

Doğal dil işleme teknikleri araştırma alanı olarak yaygınlaşmaya başladığından beri farklı çalışma disiplinlerinde doğal dil işleme uygulamaları hızla çoğalmaya başlamıştır. Hukuk alanında bu çalışmalar artan bir ivme ile devam etmekte, farklı dillerdeki hukuk metinleri için çalışmalar yapılmaktadır (Lenci v.d., 2007). Hukuk Metinleri üzerinde yapılan çalışmalar için öncelikle bir hukuk ontoloji/leksikonu oluşturulur. Oluşturulan ontolojiler; oluşturulan verilerin organize edilmesi ve yapısal hale getirilmesi, muhakeme ve problem çözümü, semantik indeksleme ve arama, semantik uyum ve ilişki ve 5.kapsam çıkarımı olmak üzere beş temel problem için veri sağları (Breuker v.d., 2004).

Lame) Fransızca hukuki metinler üzerinde yaptığı çalışmada, hukuki metinlerin ilgili olduğu hukuki kavramların, metin içinde bulunma durumundan yola çıkarak sözdizimsel analizle Fransızca için bir hukuki terim ontolojisi oluşturmuştur. Çalışma kapsamında sözdizimsel analiz ile bulunan kelimeler için kullanım sıklığı ve tf-idf (Term Frequency-Inverse Document Frequency) değerleri ile hukuki terim olma durumları incelenmiştir. Ayrıca bu çalışma kapsamında hukuki terimler arasındaki ontolojik ilişki (anlamsal ilişki) belirlenmeye çalışılmıştır (Lame, 2005).

İtalyanca hukuk metinleri üzerinde gerçekleştirilen başka bir çalışmada ise doğal dil işleme yöntemleri ve makine öğrenme teknikleri kullanılarak iki farklı hukuki metin derlemi (corpus) üzerinden ontoloji öğrenme sistemi geliştirilmesi gerçekleştirilmiştir. Bu çalışmada T2K (Text to Knowledge) ile metinlerden doğal dil işleme, istatistiksel yaklaşım ve makine öğrenmesi yöntemleri birleştirilerek hukuki terimler çıkarılmıştır. Bu çalışma elde edilen sonuçlar değerlendirildiğinde makine öğrenme teknikleri ile doğal dil işleme tekniklerinin birlikte kullanımı bu tarz sistemler için daha başarılı sonuçlar elde edilmesinde etkili olduğu görülmüştür (Lenci v.d., 2009).

Hukuksal metinler için yapılan bir başka çalışma İsviçre'de yasama metinleri üzerindeki hataların tespiti ve metinlerin sınıflandırılması amacı ile gerçekleştirilmiştir. Projenin amacı yasama metinleri üzerinde Federal Büro'nun gerçekleştirdiği yazım denetimi ve metin formatı tespiti çalışmalarını otomatik hale getirmektir. Projedeki işlem temel anlamda iki basamaktan oluşmaktadır. Birinci adımda

metin üzerinde önışlem (Pre-Processing) gerçekleştirmektir. Bu adım metni dizgeçlere ayırma (tokenization), metnin bölümlendirilmesi (text segmentation), morfolojik analiz ve kelime görev çözümlenme (part-of-speech tagging) işlemleri gerçekleştirilir. Bu adımda elde edilen çıktı hata modelleme (Error modelling) aşamasında kullanılır. Hata modelleme adımında ilk olarak metinler ile ilgili yazım ve şablon yapı kuralları belirlenir. Bulunan kurallar metin üzerine uygulanması ile hatalı yapılar belirlenen formatlar dâhilinde otomatik olarak tespit edilir. Bu çalışma sonucunda yasama metinleri için format belirlenme işlemi otomatize edilir. Yapılan çalışma sayesinde Federal Büro'nun yapmış olduđu yazım denetimi ve metin tipi belirleme işlemlerinin daha az çaba ve küçük ekiplerle yapılması amaçlanmıştır (Höfler ve Kyoko, 2012).

Bu çalışmada otomatik özetleme için atanan anahtar kelime kullanılmıştır. Bu nedenle, anahtar kelime ataması bu çalışmanın önemli bir parçasıdır. Bu bölümde, anahtar kelime atama ile ilgili önceki çalışmalar gözden geçirilecektir. Anahtar sözcükler, metinlerdeki bilgilere erişimi kolaylaştırmak, metinlerini içerdikleri içeriğe göre sınıflandırmak ve metinlerden istenen bilgilere erişmek için kullanılan tanımlayıcı sözcük veya ifadelerdir (Liu v.d., 2009). Anahtar kelime atama sürecini otomatikleştirme fikri, doğrudan anahtar kelimeyi veya bilgi çıkarma yöntemiyle elde edilen bilgileri içeren metni belgeleme mantığına dayanır. Turney'in 1992 yılında yapmış olduđu çalışmada metinlerin %75'e yakının atanması gereken anahtar kelimeleri içerdığını göstermiştir. Bu da atanması muhtemel kelimelerin büyük çoğunluğunun metin üzerinde işlem yapılarak çıkarılabileceğini göstermektedir (Turney, 1992). Anahtar kelime ataması, Metin kategorizasyonu olarak da tanımlanır. Bu, önceden oluşturulmuş bir anahtar sözcük listesindeki sözcükleri metinlerle ilişkilendirme mantığıyla çalışır. Ön işlem olarak bir anahtar kelime havuzu oluşturulur. Metin, metindeki kelime yapılarının bu anahtar kelime havuzundaki kelimelere yakınlığına göre sınıflandırılır (Uzun, 2005). Bu süreçte Naive Bayes, Logistic Regression Model, Support Vector Machine teknikleri birçok çalışmada kullanılmıştır (Joachims, 1998; Ikonomakis v.d., 2005; Aseervatham, 2011).

Metin Boyutların büyüklüğü, insanların bilgiye hızlı erişim talebi ve zaman kısıtlamaları nedeniyle metinlerden anlamlı bilgiler çıkarma ihtiyacına yol açmıştır. Bu bağlamda ilk akademik çalışmalar 1990'larda başlamış ve bugün geniş bir çalışma alanı olarak devam etmektedir. Otomatik özetleme sistemleri, kullandıkları dil bilgisinin miktarına göre genel olarak abstraktif (Soyutlayıcı) ve ekstraktif (Çıkarımsal) olarak ikiye ayrılabilir. Çıkarım sistemleri, bir metindeki bir özet için cümle seçip seçmemeye karar veren sistemlerdir. Soyutlayıcı olanlar genellikle metni işledikten ve bir söylem yapısı oluşturduktan sonra özeti sıfırdan oluştururlar. İkinci yöntem, insanların yaptıklarına yakın özetler üretmesine rağmen, birçok işleme aşamasını içerdığı için birincisine göre daha anlamlı ve çok daha zordur (Erkan ve Radev, 2004). Soyutlayıcı özetleme tekniğindeki çalışma sayısı, çıkarımsal özetlemeye göre oldukça sınırlıdır (Ganesan, 2010). Bu yöntemde özetlemenin birincil adımı bilgi çıkarma sürecidir. Soyutlayıcı özetleme için Bilgi Çıkarma (Information Extraction) sürecini kullanma fikri, hikâyelerin kısa özetlerini oluşturan FRUMP sistemine dayanmaktadır (Dejong, 1982). Bilgi çıkarımı sürecinden sonraki adım, oluşturulan bilgilerin özetleme için kullanılmasıdır. Bu aşamada Doğal Dil Üretimi teknikleri kullanılarak metin için bir özet oluşturulur. Dil oluşturma süreçleri için yapılan bazı çalışmalarda doğal dil üretimi basit yöntemlerle gerçekleştirilebilir. (Gatt ve Reiter, 2009). Çıkarım yöntemi, metinde bulunan bilgileri bu alt kümeden kullanıcıya aktarmak için Metindeki cümlelerden bir alt küme seçerek daha küçük boyutlu bir metin oluşturma işlemidir. Bu işlem gözetim altında veya denetimsiz olarak gerçekleştirilebilir. Sürecin ana kısmı, alt kümeyi oluşturacak metin parçalarının seçimidir. Bu alt küme öğeleri, metindeki pasajlardan veya cümlelerden oluşur. Alt küme elemanlarını belirlerken farklı teknikler uygulanır. Bu yöntemlerden bazıları; Cümle sıralaması temel özetleme tekniği; Bu yöntemde cümle belirlenirken kelimelerin metne yönelik kullanım oranı belirlenir ve belirli bir oranın üzerindeki kelimeleri içeren cümleler, içerdikleri kelimelerin ağırlığı ile orantılı olarak puanlanır. Bu puanlamanın bir sonucu olarak, yüksek puanlı cümleler, bir özet oluşturmak için alt öge setine eklenir. Çoklu belge özetlemede, kelime ağırlık oranları, terim frekansı ters belge frekansı ($Tf * Idf$) olarak bulunur (Erkan ve Radev, 2004). Derece merkezilik tekniği; Birden fazla belgede çıkarımsal özetleme işlemi gerçekleştirilirken aynı belge konusu ile belgedeki cümlelerin benzerlik oranları belirlenir, özetleme işlemi için benzerlik oranı yüksek cümleler kullanılır (Zhang v.d., 2011). LexPageRank yöntemi; yakınlık vektörü, cümlelerin birbirine yakınlığı belirlenerek oluşturulur. Eigenvector, Power yöntemi uygulanarak oluşturulur. Kosinüs benzerlik grafiğine bu yöntem

uygulandığında yönsüz grafikte cümlelerin benzerlik oranı belirlenir, özetleme işlemi için benzerlik oranı yüksek cümleler seçilir (Erkan ve Radev, 2004).

3. Hukuki Metinler İçin Otomatik Özetleme

Bu çalışmada otomatik özetleme için dört farklı yöntem uygulanmıştır. Kullanılan yöntemler ve ölçümler aşağıda ayrıntılı olarak açıklanacaktır. Sistem geliştirilmesi aşamasında sistem performansının değerlendirilmesi amacıyla, metin içinde geçen davanın konusu, davanın sonucu, tarafların iddia ve savunmaları, Yargıtay kararı, yerel mahkemenin kararı gibi bölümler daha önceden belirlenmiş olan Yargıtay'ın internet üzerinden yayınladığı, 200 farklı doküman kullanılmıştır (Yargıtay, 2021). Sistem çıktılarının başarı durumları metnin tutarlılığı ve akıcılığıyla beraber metin için seçilen yapıların ne kadarını içerdiğine göre değerlendirilmiştir.

3.1. Konu başlıkları (Topic Identification) ile Frekans Temelli Özetleme

Bu yöntemde öncelikle metin üzerinde 6'lı bir pencere boyutu (n-gram window size) kullanılarak metindeki cümleler için n-gram analizleri gerçekleştirilmiştir. Burada 6'lı pencere boyutunun kullanılmasının temel sebebi metin için konu belirlemesi yapılırken belirlenecek konunun (Anahtar Kelime) her zaman bitişik kelimelerden oluşmuyor olmasından kaynaklıdır. Metin içinde geçme ihtimali olan en uzun anahtar kelime dizisinin (konu başlığı), oluşturulan ontoloji incelendiğinde 6 kelimedenden oluşan bir terim olacağı tespit edilmiştir. Bu sebeple pencere boyutu 6 kelime yani en uzun terimin kelime uzunluğuna olarak belirlenmiştir. Bu sayede kelimelerin bütün bulunma olasılıkları elde edilerek konu belirlenmesi işlemi daha etkile hale getirilmiştir. Sonraki adım metnin içeriğinin ilgili olduğu kavramların belirlenmesi işlemidir. Konu başlıkların ve ilgili olduğu kavramları belirleyerek (Concept Identification) metnin kapsamı belirlenir. Belirlenen bu kavramların geçtiği cümleler belirlenir (Saggion ve Lapalme ,2002). Belirlenen bu anahtar kelimelerin ve kavramların geçtiği cümlelere içerdikleri anahtar kelime frekansların değerleri toplamı kadar bir ağırlık değeri atanır. Değer atama işlemi sırasında önemli metriklerden biride frekans değerleri cümlelere ağırlık değeri olarak atanırken atanacak olan frekans değerinin cümlede geçen anahtar kelimenin bütün metindeki frekans değeri olarak mı, yoksa sadece cümledeki frekans değeri olarak mı ekleneceği işlemidir. Bu seçim sonuçlarda belirgin farklılığa neden olmaktadır. Metnin devamında sonuçlarla ilgili kısımda bu durumlar detaylandırılacaktır.

$$G(S) = \sum_{k=0}^n (p(T, k))$$

G(S)= Cümlenin ağırlık değeri

n=Cümlenin içerdiği anahtar kelime sayısı

P(T)=cümlenin içerdiği anahtar kelimelerin frekans değerleri

Ağırlık değerleri belirlenen cümleler ağırlık değerlerine göre sıralanır. Daha sonra sıralanan cümlelerin ne kadarının özetleme işlemi için kullanılacağına karar verilir. Bu aşamada ilk olarak orijinal metnin cümle sayısının %20'si oranında cümlenin özetleme için kullanılması denenmiştir. Bu oran ile yapılan sonuçlarda karar metinlerdeki cümlelerin uzunluğundan dolayı seçilen bu orandaki cümlelerinden oluşturulan özet metnin orijinal metnin yaklaşık olarak %50'si oranında olduğu görülmüştür. Özet metinlerini boyut olarak daha makul seviyeye çekmek için orijinal metindeki cümle sayısının yaklaşık olarak % 17 si oranında cümle seçildiğinde metnin tutarlılığının %20'lik seçime göre düşük bir miktarda değiştiği ama özet metnin boyutunun orijinal metin boyutunun yaklaşık olarak %40'ı kadar olduğu belirlenmiştir. Bu sebeple frekans değerleri için %17'lik oran kullanılmıştır.

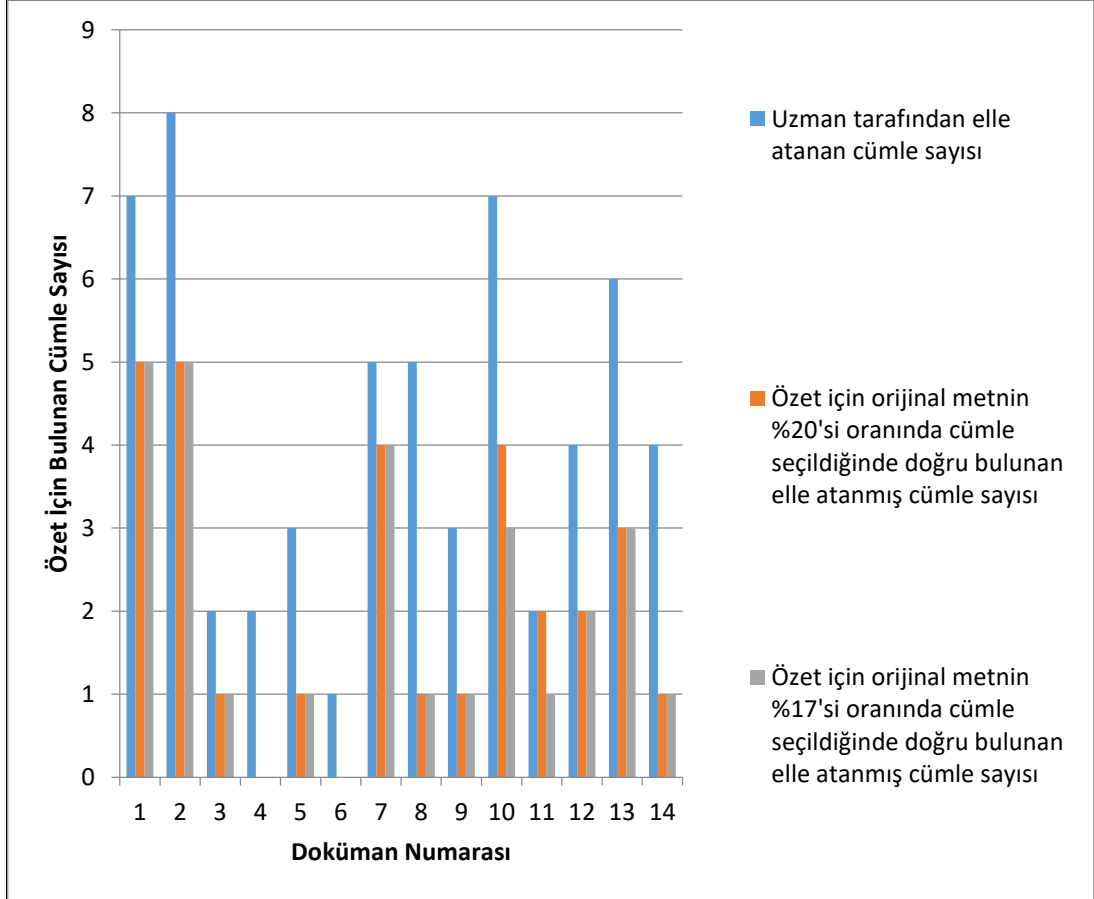
Bu yönteme ilave olarak metin için önemli bazı cümlelerin özetlemeyi daha tutarlı hale getirmek için eklenmesi amacı ile metinler içinden bazı önemli cümlelerin ağırlık değerleri düşük olsa dahi özetleme için seçilmesi sağlanmıştır. Bu işlem için bazı işaretleyici yapılar (cue phrase) belirlenerek bu yapıların geçtiği cümleler özetlemeye eklenecek şekilde ağırlık değerleri artırılır. Örneğin "Sonuç:"

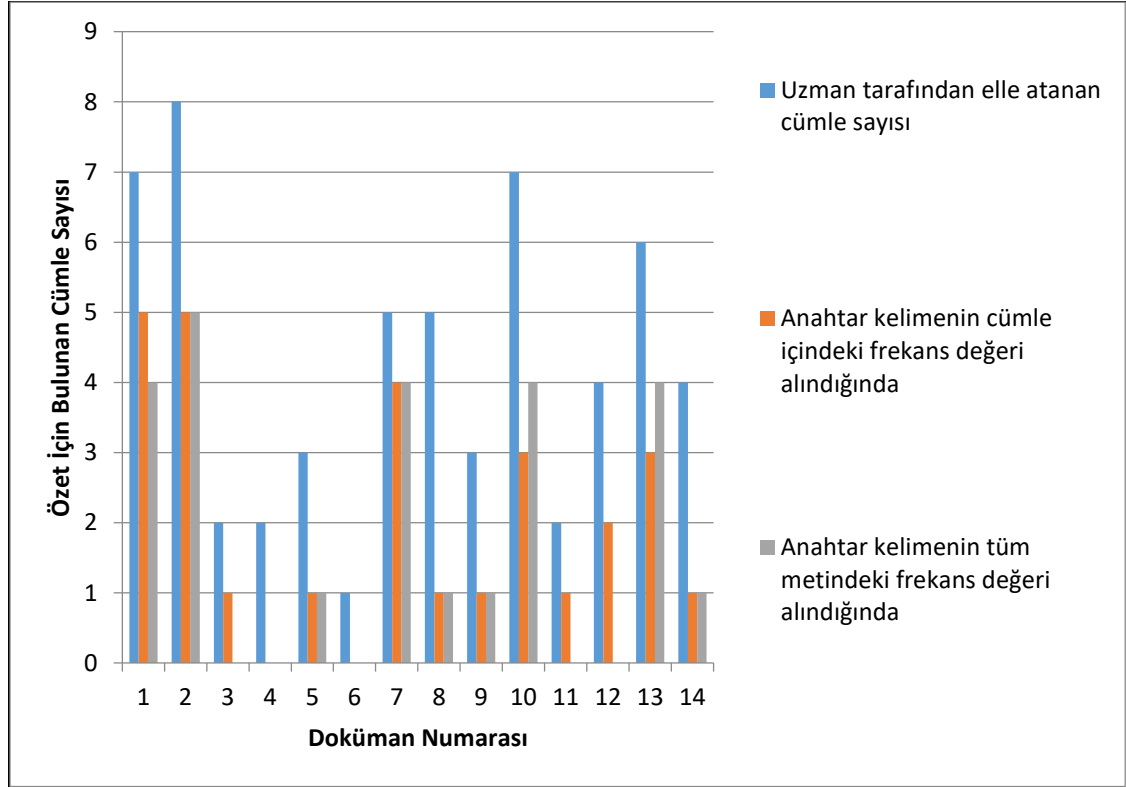
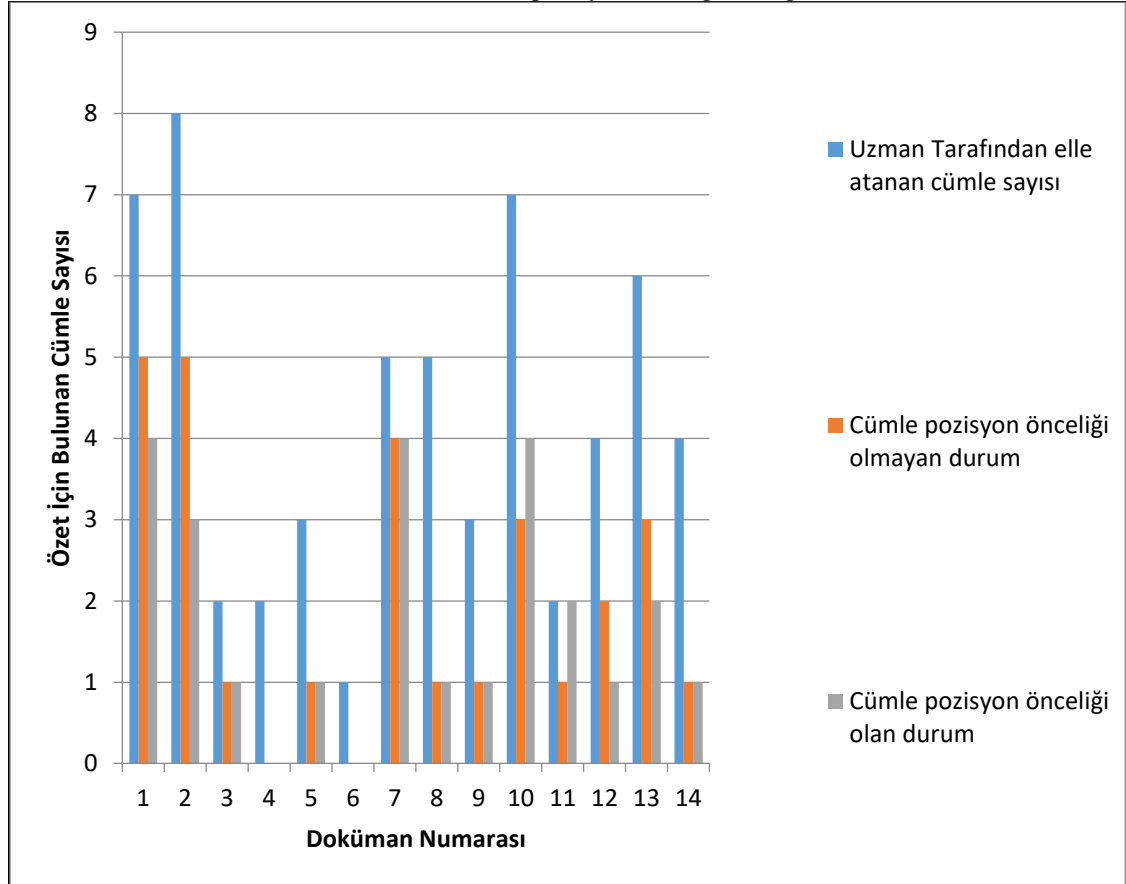
şeklindeki bir başlangıç kavramı karar cümlesini nitelediği için karar cümlesinin özetlemede olması özetleme için önemlidir. Bu yüzden karar cümlesine yüksek bir ağırlık değeri atanarak özet için seçilmesi sağlanır (Hovy ve Lin, 1999). Bu yöntem ile önemli cümlelerin özet metin içinde yer alması sağlanmaktadır.

Bu yönteme ilave olarak metinlerin başında bulunan cümleler ile sonunda bulunan cümlelerin metin ile ilgili daha çok bilgi içerdiği varsayımı (sentence positions) ile metin belirli orandaki ilk cümlesi ve son cümlesinin (ilk 5 cümle ve son 5 cümle-yaklaşık olarak metnin %10-20 oranındaki cümleler) ağırlık değerleri çok uç değerlere neden olmayacak şekilde artırılır (7 ile 15 arasında bir aralıkta artırılır) böylece metnin başındaki ve sonundaki değerlerin seçilme ihtimalleri artırılmaktadır (Erkan ve Radev, 2004).

Yukarıda belirtildiği gibi cümleye atanacak ağırlık değeri için kullanılacak frekansın anahtar kelimenin tüm metindeki frekans değeri olması veya anahtar kelimenin cümle içindeki frekans değeri olması, seçilecek cümle sayısının uzunluğu, n-gram window-size, cümle pozisyonu durumunun uygulanması gibi metriklerin değişmesi özetleme sonuçları üzerinde değişikliklere sebep olmaktadır. Bu metriklerin sonuçlar üzerindeki etkilerinin daha iyi görülebilmesi için örnek amaçlı 14 doküman üzerinden metriklere göre sonuç karşılaştırmaları aşağıdaki gibidir.

Tablo 1: Özet için seçilen cümle oranı metriği

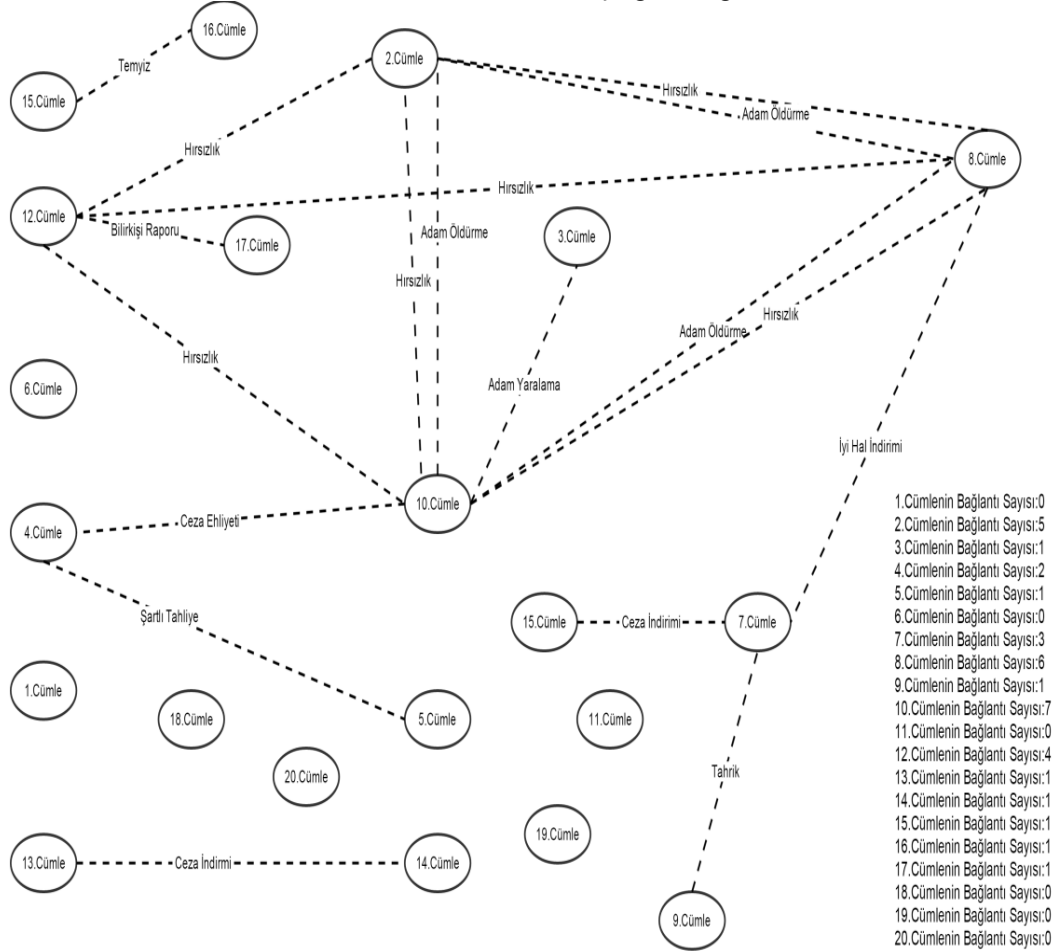


Tablo 2: Anahtar kelimenin tüm metindeki frekansı ve anahtar kelimenin cümle içindeki frekansı metriği**Tablo 3 :** Cümle pozisyon önceliği metriği

3.2. Yönsüz Çizge (Undirected Graph) Modeli ile Özetleme

Özetleme işlemi sırasında uygulanan bir başka yöntem ise cümleler arasında bir ilişki çizgesi oluşturularak özetleme için önemli cümlelerin çizgedeki düğümlerin (cümleler) önem derecesine göre belirlenmesidir (Erkan ve Radev, 2004). Bu işlem için daha önceki yöntemde de olduğu gibi öncelikle cümlelerine ayrılmış olan metin üzerinde her cümle için n-gram analizi gerçekleştirilir. Gerçekleştirilen n-gram analizleri üzerinde konu başlıkları (anahtar kelimeler) belirlenir. Bulunan her anahtar kelime bulunduğu cümle ile ilişkilendirilir. Bu işlem bütün cümleler için uygulandıktan sonra cümleler arasından bulunan ortak anahtar kelimeler üzerinden cümleler arasında yönsüz bir çizge oluşturulur. Böylece metin içindeki ilişkili cümleleri belirleme aşamasına geçilir. Bu aşamada oluşturulan çizge için bir komşuluk matrisi (Adjacency matrix) oluşturulur bu matriste ayrıca komşuluk ilişkisi olan düğümler arasındaki ortak anahtar kelime sayısı bağlantı (edge) olarak tutulur. Oluşturulan matris üzerinden her düğümün hangi cümleler ile bağlantılı olduğu ve düğümün sahip olduğu ilişki sayısı belirlenmiş olur. Her düğüme sahip olduğu edge sayısı ağırlık değeri olarak atanır. Daha sonra düğümler arasından özetleme için seçilecek cümle sayısı kadar cümle ağırlık değeri sıralamasına göre seçilerek özet metni oluşturulur. Örneğin;

Şekil 1: Yönsüz Çizge Örneği



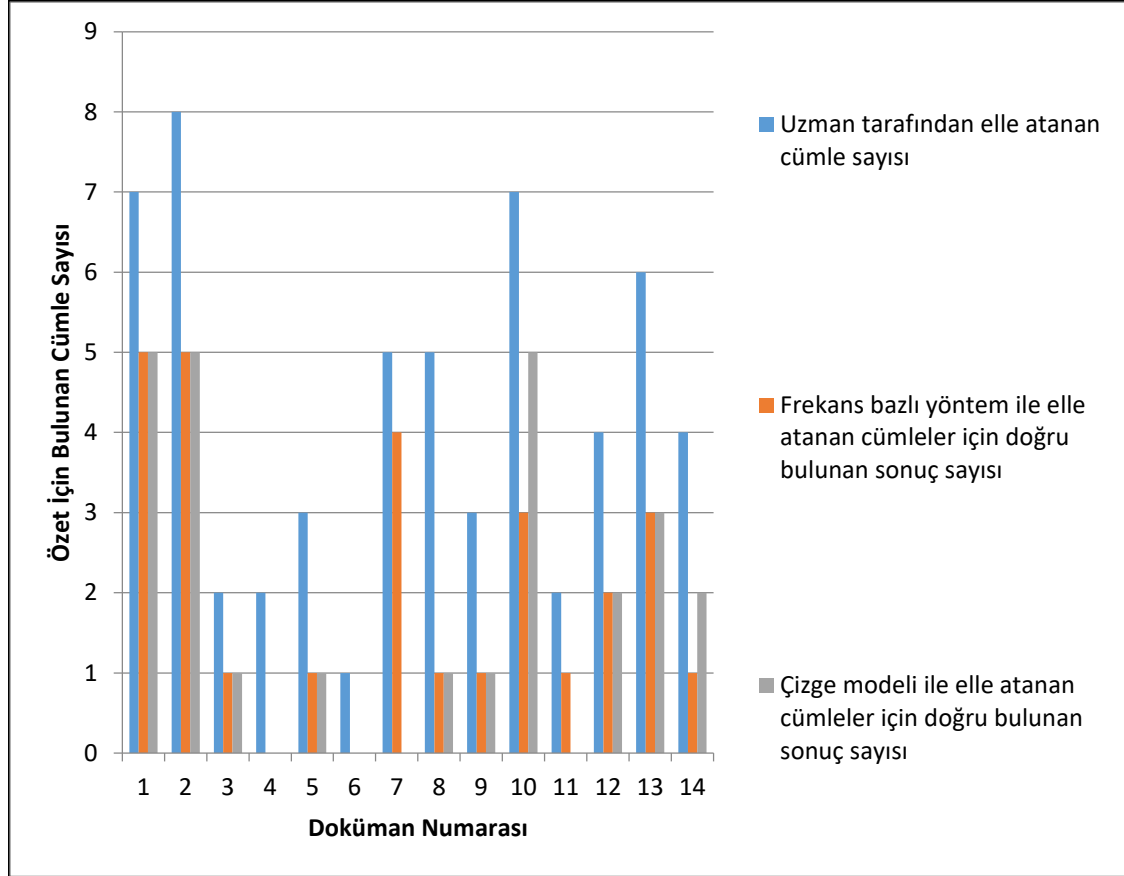
Şekilde cümleler arasında oluşturulan ilişki çizgesinde cümlelerin sahip olduğu ortak anahtar kelimeler üzerinde ilişkiler tanımlanmıştır. Her cümle için sahip olduğu bağlantı sayısı ağırlık değeri olarak atanmıştır. Bu cümleler arasından %20 oranında cümle özet için seçilecek olur ise seçilecek cümleler şu şekilde olacaktır; 10. Cümle (ağırlık değeri 7), 8. Cümle (ağırlık değeri 6), 2.Cümle (ağırlık değeri 5), 12. Cümle (ağırlık değeri 4). Bu cümlelerden oluşturulacak özet metnin sıralaması ise akışı bozmamak için 2.cümle,8. Cümle,10. Cümle ve 12.cümle şeklinde olacaktır.

Bu yöntemle ilave olarak daha önceki yöntemde de uygulanan Cue Phrase yöntemi ile önemli cümleler grafta ağırlık değeri yüksek olmasa da özet metnine eklenir.

3.3. Çizge Modeli ile Frekans Bazlı Model Sonuç Karşılaştırma

Çizge model ile frekans bazlı yöntem sonuçlarının karşılaştırılması için belirli sayıda doküman için sonuçlar şekildedeki gibidir.

Tablo 4: Çizge Modeli ile Frekans Bazlı Model Sonuç Karşılaştırma

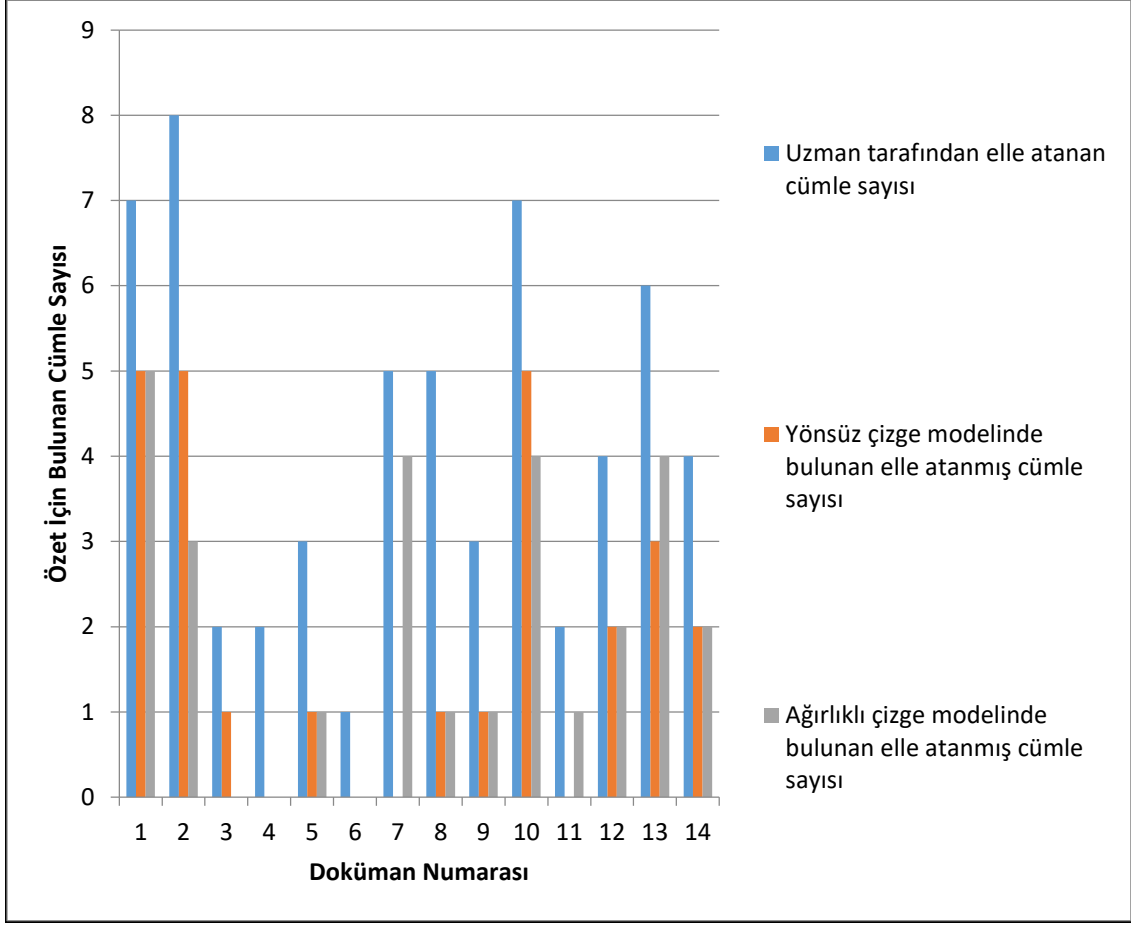


3.4. Ağırlıklı Çizge (weighted graph) Modeli ile Özetleme

Yukarıdan detaylandırılan yönsüz çizge modelinde düğümlere değerler atanırken sahip oldukları bağlantı sayısı değer olarak atanmaktadır. Bu modele ilave olarak düğümleri oluşturan cümlelerin metin içindeki pozisyonuna bağlı olarak, bağlantılarına değer atanarak düğüm için hesaplanacak olan ağırlık değerine bağlantı değerleri toplamı değeri olarak atanmaktadır. Bu yöntem için metnin belirli orandaki ilk cümlesi ve son cümlesinin (ilk 5 cümle ve son 5 cümle-yaklaşık olarak metnin %10-20 oranındaki cümleler) oluşturduğu düğümlere 1 ile 1.5 arasında bir bağlantı değeri atanır, diğer cümlelere bağlantı değeri 1 olarak atanır. Geliştirmelerde öncelikli cümleler için değer atanırken 1 ile 1.5 arasında değerler atanarak denemeler yapılmıştır. Sonuçlar arasında çok büyük değer farkları oluşturmamak için farklı bu aralıkta farklı değerler atanmıştır. Atanan değerler arasında en uygun sonuçlar öncelikli cümlelere 1.2 değeri atandığında elde edilmiştir. Bu durumda bağlantılara atanan değerlerin sadece düğümü oluşturan cümleyi etkilediği için atanan bağlantı değerleri toplanırken bağlantı yönü dikkate alınarak ağırlık değeri hesaplanır. Örneğin; 2. cümle ile 10.cümle arasındaki bir bağlantı olduğu durumda bu bağlantının 2. Düğüme olan etkisi 1.2 iken 10.Düğüme etkisi 1'dir.

Ağırlıklı (ırlılmış çizge model ile yukarda detaylandırılan yönsüz çizge modeli için elde edilen sonuçlar aşağıdaki gibidir.

Tablo 5: Ağırlıklı Çizge Modeli Karşılaştırma



3.5. Karma Metot ile Özetleme

Hukuki metinlerde özetleme işleminde yukarıda denenen yöntemlerde, özet metinleri için elle atanan cümlelerin bulunma sayılarının ve bulunan cümlelerin farklı olması özetleme yönteminde farklı metotların bir arda kullanılması durumunda daha etkili sonuç alınabileceği ihtimalini doğrulamıştır. Yapılan araştırmalarda Türkçe metinlerde özetleme için yapılan çalışmalarda daha öncede metotlar tek başına kullanıldığı gibi birden çok metodun karma olarak kullanılarak sonuçların daha anlamlı olmasının sağlandığı görülmüştür (Güran v.d., 2010). Farklı metotların bir arda kullanılması metotların olumlu yönlerinin kullanılması ile sonuçların daha başarılı olmasına imkân vermektedir. Hukuki metinler için yapmış olduğumuz özetleme çalışması kapsamında daha önceden denenmiş olan Konu başlıkları ile frekansı Temelli Özetleme ve Ağırlıklı Çizge Modeli ile Özetleme metotların karma olarak çalıştırılması ile daha etkili bir özetleme sistemi elde edilmeye çalışılmıştır. Bu metot, iki yönetimin ayrı ayrı özet için seçmiş olduğu cümlelerin tümü üzerinde işlem yapılarak bu cümlelerden özet için en uygun olanların özet metninin oluşturulmasında kullanılması mantığına dayanmaktadır. Metotların birlikte kullanılması sürecinde farklı parametreler ve teknikler kullanılarak oluşturulacak olan özette daha başarılı olması hedeflenmiştir. Yöntem ile ilgili detaylar alt başlıklarda verilecektir.

3.5.1. Öncelik Değeri Yüksek Cümlelerin Seçilmesi ile Özet Oluşturulması

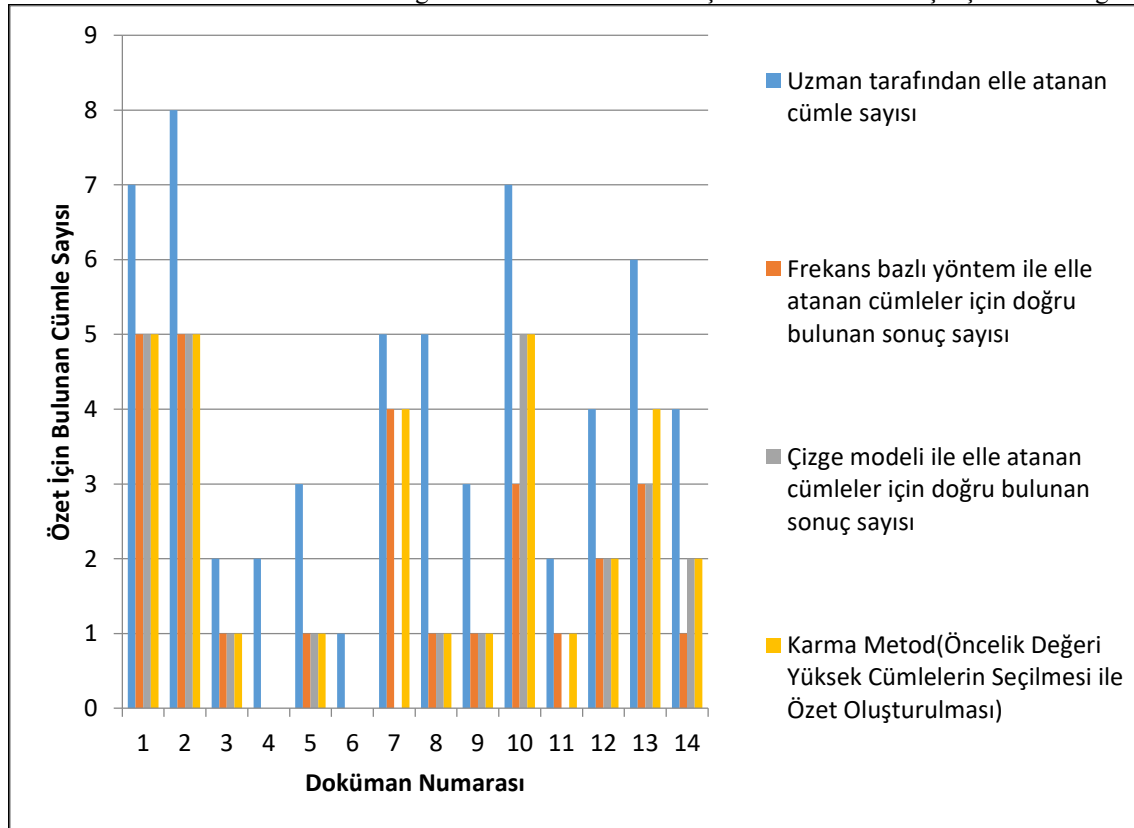
Bu teknikte iki metodun önermiş olduğu aday cümleler arasından öncelik değeri yüksek olanların seçilmesi ile yeni özette oluşturulması amaçlanmaktadır. Bunun için öncelikle özetlenecek metin önce Konu başlıkları ile frekansı Temelli Özetleme ve Ağırlıklı Çizge Modeli ile Özetleme metotları ayrı ayrı çalıştırılarak her iki yöntem içinde cümle ağırlıkları belirlenir ve bu cümleler ağırlıklarına göre sıralanır.

İki yöntem için de sıralan cümleler içinde metin boyutunun %25 oranında ağırlıkları yüksek olan cümleler seçilerek iki ayrı öncelikli cümle seti elde edilir. Bu setler metnin geri kalanında SET1 ve SET2 olarak adlandırılacaktır. Bu setlerin oluşturulmasında tüm metnin %25'i oranının kullanılmasının sebebi oran arttığında son oluşturulacak olan özete sadece yöntemlerden birinin çıktısı ile daha çok benzerlik göstermesidir, bu oran daha düşük tutulduğu durumda aynı problemler geçerlidir. Yapılan denemler sonucu başarı oranlarının en yüksek olduğu durumların %25 oranının seçildiği durumlar olduğu anlaşılmıştır. Daha sonra bu iki setten yeni bir özet oluşturulması aşaması başlatılmaktadır. Yeni oluşturulacak olan özete bulunacak cümleler bu iki setten seçilen cümlelerle oluşturulmaktadır. Cümle seçimi gerçekleştirilirken aşağıdaki algoritma kullanılmaktadır. (Oluşturulacak özet miktarı tüm metnin %17,5'i oranında olacaktır. Bu set Özet Seti olarak adlandırılacaktır.)

1. SET1 ve SET2 içinde ağırlığına göre sıralı olan cümlelerden her iki set için ayrı ayrı setlerin ilk %50'lik kısmında olan cümleler eğer diğer sette de mevcutsa özet seti içine eklenecektir. Bu işlem SET1 ve SET2 için aynı anda yapılmaktadır. Örneğin: SET1'in ilk %50'lik kısmında 3. cümle olduğunu varsayalım eğer SET2 de 3. cümleyi içeriyorsa 3. cümle özet için seçilir. Bu işlem Özet Seti dolana ve ya iki set içinde ilk %50'lik kısım için işlem uygulanıncaya kadar devam eder.
2. Özet seti eğer birinci adımda tamamlanmadıysa SET1 ve SET2 içinde setlerden birinde olup diğerinde olmayan cümleler baştan başlamak suretiyle her iki set için aynı anda yapılarak bu şartı sağlayan cümleler özet setine eklenir. Bu işlem özet seti dolana kadar devam eder.
3. Özet setindeki cümleler, cümlelerin metin içindeki konumuna göre hiyerarşilendirilerek özet metni akışına uygun olarak sıralı olarak oluşturulur.

Bu yöntem ile metinlere elle atılan cümlelerin özet metninde bulunma oranına diğer metotlara göre %10 oranında artmıştır. Özet metnin oranın %17 olarak seçilmesindeki amaç özet metnin anlamlı minimum boyutta tutulmasıdır. Önceki metotlar ile karşılaştırmaları grafikteki gibidir.

Tablo 6: Karma Model Öncelik Değeri Yüksek Cümlelerin Seçilmesi Yöntemi Karşılaştırma Grafiği



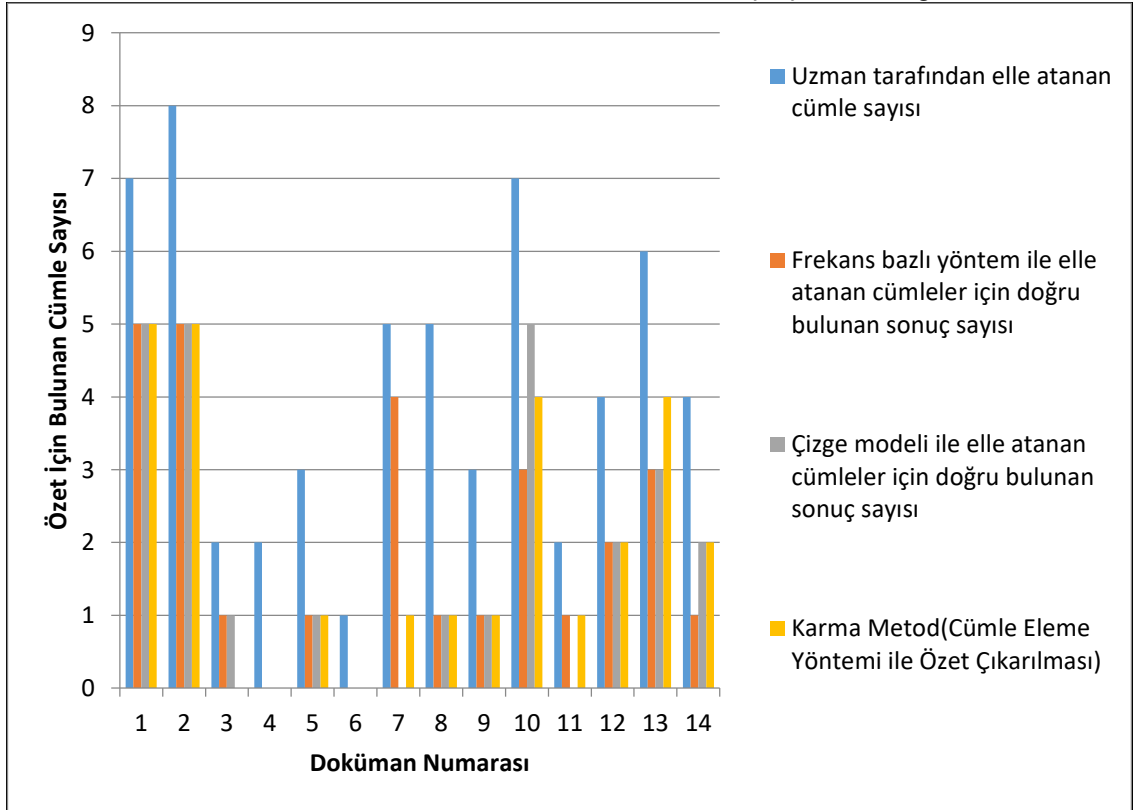
3.5.2. Cümle Eleme Yöntemi ile Özet Çıkarılması

Bu teknikte iki metodun önermiş olduğu aday cümleler arasından bazı cümlelerin elenmesi ile geri kalan cümlelerden özet oluşturulmaktadır. Bunun için öncelikle özetlenecek metin önce Konu başlıkları ile frekansı Temelli Özetleme ve Ağırlıklı Çizge Modeli ile Özetleme metotları ayrı ayrı çalıştırılarak her iki yöntem içinde cümle ağırlıkları belirlenir ve bu cümleler ağırlıklarına göre sıralanır. İki yöntem için de sıralan cümleler içinde metin boyutunun %25 oranında ağırlıkları yüksek olan cümleler seçilerek iki ayrı öncelikli cümle seti elde edilir. Bu setler metnin geri kalanında SET1 ve SET2 olarak adlandırılacaktır. Daha sonra;

1. SET1 ve SET2 içindeki her cümle için ayrı ayrı diğer sette olup olmadığı kontrol edilir. Eğer aday cümle diğer sette yok ise aday cümle elenir. Örneğin; SET birde bulunan 3.Cümle eğer SET2’de yoksa bu cümle özet için elenir.
2. Eleme işlemi tamamlanan setlerde öncelik sırasına göre cümleler özet setine eklenir. Bu işlem özet setini boyutu istenen seviye (%17) ulaşana kadar devam eder.
3. Özet setindeki cümleler, cümlelerin metin içindeki konumuna göre hiyerarşilendirilerek özet metni akışına uygun olarak sıralı olarak oluşturulur.

Bu yöntem sonucunda elde edilen sonuçlar yöntemlerin tek başına gösterdikleri başarı oranına yakın seviyededir. Başarı oranında bir miktar artmış olsa da *Öncelik Değeri Yüksek Cümlelerin Seçilmesi ile Özet Oluşturulması* yöntemindeki başarı oranını yakalamamıştır. Bu sebeple özetleme için *Öncelik Değeri Yüksek Cümlelerin Seçilmesi ile Özet Oluşturulması* yönteminin kullanılması daha uygun görülmüştür. Bu yöntem için sonuç grafiğine ve Karma Metodların kendi aralarındaki sonuç karşılaştırma grafikleri aşağıdaki gibidir.

Tablo 7: Karma Model Cümle Eleme Yöntemi Karşılaştırma Grafiği



4. Sonuç

Bu çalışma kapsamında çizge (Graph) temelli özetlemenin Frekans bazlı özetlemeye oranla kısmen daha başarılı sonuçlar vermekle beraber frekans bazlı ve çizge bazlı modellerin birlikte kullanımı olan karma modelin her iki yöntemin tek başına kullanılmasına oranla daha başarılı sonuçlar verdiği tespit edilmiştir. Karma Modelin başarısını artırmak için denenen “*Öncelik Değeri Yüksek Cümlelerin Seçilmesi ile Özet Oluşturulması*” yöntemin “*Cümle Eleme Yöntemi ile Özet Çıkarılması*” yöntemine göre daha başarılı sonuçlar verdiği tespit edilmiştir.

Kaynaklar

Aseervatham, S., Antoniadis, A., Gaussier, É., Buret, M., ve Denneulin, Y. (2011). A sparse version of the ridge logistic regression for large-scale text categorization. *Pattern Recognition Letters*, 32(2), 101-106.

Breuker, J., Valente, A., ve Winkels, R. (2004). Legal ontologies in knowledge engineering and information management. *Artificial intelligence and law*, 12(4), 241-277.

Carenini, G., ve Cheung, J. C. K. (2008, June). Extractive vs. NLG-based abstractive summarization of evaluative text: The effect of corpus controversiality. In *Proceedings of the Fifth International Natural Language Generation Conference* (pp. 33-41).

Carenini, G., Cheung, J. C. K., ve Pauls, A. (2013). Multi-document summarization of evaluative text. *Computational Intelligence*, 29(4), 545-576.

Cortes, C., ve Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297..

DeJong, G. (1982). An overview of the FRUMP system. *Strategies for natural language processing*, 113, 149-176.

Erkan, G., ve Radev, D. (2004, July). Lexpagerank: Prestige in multi-document text summarization. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 365-371).

Ganesan, K., Zhai, C., ve Han, J. (2010). Opinosis: A graph based approach to abstractive summarization of highly redundant opinions.

Gatt, A., ve Reiter, E. (2009, March). SimpleNLG: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)* (pp. 90-93).

Güran, A., Bekar, E., ve Akyokuş, S. (2010). A comparison of feature and semantic-based summarization algorithms for Turkish.

Hovy, E., ve Lin, C. Y. (1999). Automated text summarization in SUMMARIST. *Advances in automatic text summarization*, 14, 81-94.

Höfler, S., ve Sugisaki, K. (2012, April). From drafting guideline to error detection: Automating style checking for legislative texts. *Association for Computational Linguistics*.

Ikonomakis, M., Kotsiantis, S., ve Tampakas, V. (2005). Text classification using machine learning techniques. *WSEAS transactions on computers*, 4(8), 966-974.

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer, Berlin, Heidelberg.

Lame, G. (2005). Using NLP techniques to identify legal ontology components: concepts and relations. In *Law and the Semantic Web* (pp. 169-184). Springer, Berlin, Heidelberg.

Lehman, A. (2010). Essential summarizer: innovative automatic text summarization software in twenty languages. In *Adaptivity, personalization and fusion of heterogeneous information* (pp. 216-217).

Lenci, A., Montemagni, S., Pirrelli, V., ve Venturi, G. (2007). NLP-based ontology learning from legal texts. A case study. *LOAIT*, 321, 113-129.

Lenci, A., Montemagni, S., Pirrelli, V., ve Venturi, G. (2009). Ontology learning from Italian legal texts. *Law, Ontologies and the Semantic Web*, 188, 75-94.

Liu, F., Pennell, D., Liu, F., ve Liu, Y. (2009, June). Unsupervised approaches for automatic keyword extraction using meeting transcripts. In *Proceedings of human language technologies: The 2009 annual conference of the North American chapter of the association for computational linguistics* (pp. 620-628).

Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information retrieval*, 2(4), 303-336.

Saggion, H., ve Lapalme, G. (2002). Generating indicative-informative summaries with sumum. *Computational linguistics*, 28(4), 497-526.

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](http://www.cs.bilkent.edu.tr/~guvenir/courses/CS550/Workshop/Yasin_Uzun.pdf).

Yargıtay içtihat. (2021). Erişim Tarihi: 25 Nisan 2021, from <https://karararama.yargitay.gov.tr/YargitayBilgiBankasiIstemciWeb/>

Zhang, H., Fiszman, M., Shin, D., Miller, C. M., Rosemblat, G., ve Rindfleisch, T. C. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. *Journal of biomedical informatics*, 44(5), 830-838.