

VERİ MADENCİLİĞİNDE VERİ TEMİZLEME VE TÜRK MÜZİĞİ VERİLERİ ÜZERİNDE UYGULAMA ÇALIŞMALARI

Geliş Tarihi: 10.04.2019

Sinan DURU¹

Kabul Tarihi: 30.06.2019

Doç. Dr. Oya H. YÜREĞİR²

Makale Türü: Derleme

Özet

Veri kümelerinin giderek artması, bu verilerin kontrol altında tutulmasıyla ve etkin bir şekilde kullanımıyla ilgili çalışmaların yapılmasını, verilerin yönetilmesini bir gereklilik hâline getirmiştir. Verilerin anlamlı ve yararlı bir hâle getirilebilmesi için buradaki verilerin işlenmesi ve bilgi hâline dönüştürülmesi gerekmektedir. İyi yönetilen bilginin veriye değer katması gibi bir yararı olmasıyla beraber, yeni adımlar atmak için de ana kaynak olduğu ortadadır. Günümüzde veri bilimi daha çok üretim, ekonomi, bilişim, eğitim, sağlık vs. gibi alanlarda kullanılsa da bu bilimin verinin olduğu her yerde kullanılabilmesi mümkündür. Bu çalışma kapsamında bir java programıyla internetten alınan Türk Müziği repertuarındaki 43.936 adet eserden oluşan veri tabanının istatistiksel analizi yapılmıştır. Yapılan veri analizi çalışmasında müziğin temellerini oluşturan ritm, Türk Müziğindeki kullanılan ismiyle “usûl” ve buna doğrudan bağlı olduğu düşünülen güftelerdeki arûz unsuruyla ilgili tespitler yapılmaya çalışılmış ve veri temizlemenin süreç içerisindeki önemine değinilmiştir. Veri analizi sonucunda 20. yy’dan önce bestelenen eserlerde arûz vezninin daha çok kullanıldığına istatistiksel analizi ortaya koyulmuştur.

Anahtar Kelimeler: Veri Madenciliği, Müzik Madenciliği, Türk Müziği, Usûl, Arûz.

Jel Kodları: C15, C80.

DATA CLEANING FOR DATA MINING AND APPLICATIONS ON TURKISH CLASSICAL MUSIC DATA

Abstract

As a result of increasing data sets, keeping data under control and analyzing them have become compulsory. It is a necessity to convert the raw data into knowledge with data mining. Managing knowledge adds value to data. Furthermore, it is a basis for innovation. Even though data science is generally used in production, economics, informatics, health etc., it is possible to implement it in all sectors where data exist. In this study, a statistical analysis is conducted on a data set of Turkish Music repertoire which was downloaded with a Java script programme. This data set consists of 43.936 data. For the statistical analysis the topic is selected in rhythmic patterns which is the structure of music called “usûl” in Turkish Classical Music. This study concerns with the relationship between “usûl” and prosodic patterns called “arûz” and the importance of data cleaning in data mining process. The result of the statistical analysis shows that composers who lived in the early 20th century used “arûz” more than others.

Keywords: Data Mining, Music Mining, Turkish Music, Usûl, Arûz.

Jel Codes: C15, C80.

¹ Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Endüstri Mühendisliği, E-posta: sinanduru@hotmail.com, ORCID ID: 0000-0003-2232-0710

² Çukurova Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, E-Posta: oyuregir@yahoo.com, ORCID ID: 0000-0002-9607-8149

1. GİRİŞ

Günümüz gelişen teknolojiyle birlikte, çok büyük ölçeklerdeki verileri yıllarca saklayabilmek ve veri tabanlarında tutabilmek kolay bir işlem hâline gelmiştir. Daha önceleri kullanılan verilerin birimlerinde “byte” gibi ölçüler konuşulurken, günden güne gelişen teknolojiyle birlikte bilgisayarların “tera byte” ölçüsündeki verileri işlemesi artık saniyelik zaman dilimlerinde gerçekleşebilmektedir.

Teknolojinin ilerlemesiyle birlikte verilerin kayıt edilmesi, depolanması, saklanması ve tekrâr kullanımı kolaylaşmış ve giderek daha da kolaylaşmaktadır. Bilgilerin depolanmasındaki kolaylık, her ne kadar veriyi uzun süre ve düşük kapasitelerle saklanmasını kolaylaştırır da; veri sayısındaki artışla birlikte verilerin analiz edilmesi ve doğru verilerin doğru yerde kullanılmasıyla ilgili çalışmalar yapılması bir zorunluluk hâline gelmiştir. Verilerden çıkarılacak bilgi tıpkı bir cevher gibidir. Bu cevheri ortaya çıkarmak için veri madenciliği teknikleri ve algoritmaları geliştirilmiştir. Veri madenciliği teknikleri genel olarak, istatistiksel hesaplamaların temelleriyle bilgisayar biliminin birleştirilmesine dayanmaktadır. Böylece disiplinler arasında bir bilgi keşfi yapılmasını sağlayan yeni bir disiplin olarak ortaya çıkmıştır denilebilir. Günümüzde kamuda, şirketlerde ve özel kuruluşlarda birçok veri toplanmakta ancak bu verilerin ne amaçla ve nerede kullanılacağıyla ilgili bilgileri iyi analiz edilmediği için büyük bir veri yığını ortaya çıkmakta, daha sonra ise bu verilerin imhâ edilmesi için tekrâr bir iş yükü harcanmaktadır. Bu verilerin olduğu gibi saklanması ise verinin içerisindeki bilginin hapsedilmesine sebebiyet vermektedir.

Bilginin ortaya çıkması veriyle başlar. Veri kavramı basit bir düzeyde karakterler, şekiller ya da sayılardan oluşan tek başına anlam ifâde etmeyen bir bulgu iken, bu verilerin işlenmesiyle ve anlamlandırılmasıyla birlikte bilgi oluşur. Bilgi, işlemden geçmiş veriler topluluğudur (Yüreğir, 2003, s. 40-41).

Bu çalışma kapsamında Türk Müziği repertuarında bulunan 43.936 adet esere ait bilgilerle oluşturulan veri tabanı ile bir veri analizi çalışması yapılmıştır. Türk Müziğindeki meşk halkalarının kopmasıyla birlikte çok önemli bir yapısal unsur olan usûl-arûz münâsebetinin de koptuğu ve bu konunun günümüz şartlarında artık önemsizleştiği görülebilmektedir. Çalışmada arûz vezniyle bestelenen sözlü eserlerin bestelendiği zamanlarla ilgili bir analiz yapılmıştır ve Türk Müziğinde tartışılabilen usûl-arûz münâsebetiyle ilgili yapılacak olan veri madenciliği çalışması için bir ön çalışma oluşturulmuştur.

2. VERİ VE MÜZİK MADENCİLİĞİ

2.1. Veri Madenciliği

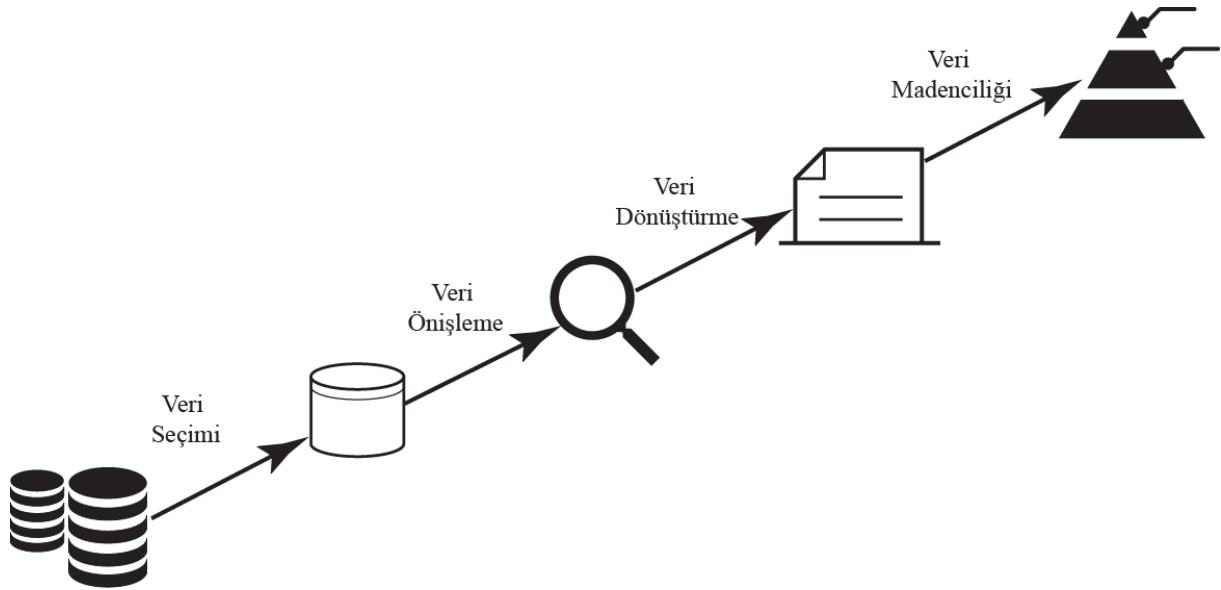
Veri madenciliği, 21. yüzyılın başlarından itibaren veri depolama araçlarının, barkod teknolojilerinin, veri işleme teknolojilerinin gelişimine de paralel olarak büyük bir hızla gelişmeye devam etmektedir ve birçok farklı alanda kullanımı giderek artmaktadır. Buna bağlı olarak da veri madenciliğinin ne olduğuyla ilgili çok farklı tanımlar yapılmıştır. Bu tanımlar zaman içerisinde, yeni ihtiyaçlara bağlı olarak da sürekli değişmektedir (Silahtaroglu, 2012, s. 12). Veri madenciliğiyle ilgili literatürdeki bazı tanımlamalara örnek vermek gerekirse;

- Veri madenciliği ya da bilgi keşfi büyük veri setlerinin bilgisayar destekli olarak analiz edilmesi sonucunda anlamlı verinin çıkarılması işlemidir (Sumathi & Svendam, 2006, s. 9).

- Veri madenciliği; istatistik, makine öğrenmesi ya da yapay zeka gibi çeşitli yöntemler kullanarak tek başına anlam ifade etmeyen veri topluluklarından bilgiyi ortaya çıkaran süreçler topluluğudur (Sarıman, 2011, s. 192-193).

Veri Madenciliği, son zamanlarda yaygınlaşan ismiyle “big data” olarak bilinen çok büyük veri kümelerinden bilgi çıkarımı sağlar. Standart yöntemlerle çok uzun zaman alacak araştırmalara ve analizlere dayandırılarak çıkarılacak olan bilgiyi, teknolojiyle kısa sürede ve daha net bir şekilde elde edilir. Elde edilen bilgiler ise yeni kararlar verilmesinde ya da geleceğe yönelik tahminlerde bulunulmasında yardımcı olacaktır. Diğer bir tanımla şirketlerin stratejik kararlar verirken olası ihtimalleri göz önünde bulundurmasına ve farklı alternatifleri değerlendirirken olası yolların görülmesine yardımcı olur (Alpaydın, 2000). Veri madenciliği kavramı, verilerin arasında bir ilişki kurma ya da matematiksel, istatistiksel olarak bu ilişkilerin tanımlanmasının bilgisayar yoluyla yapılmasıyla alakalıdır. Verilerin arasındaki ilişkinin belirlenmesini, kurulan çeşitli algoritmalara bağlı olarak yazılım sağlamaktadır. Buradaki amaç kolayca fark edilemeyen veri desenlerini belirlemektir (Arslan, 2008, s. 4).

Veri madenciliğindeki bilgi keşfi süreci sırasıyla verinin temizlenmesi, dönüştürülmesi, veri madenciliği süreci ve çıkan sonuçların yorumlanması olarak tanımlanabilir.



Şekil 1: Veri madenciliği aşamaları

Veri madenciliği çalışmaları incelendiğinde verinin kalitesi ve bütünlüğünün büyük önem arz ettiği görülmektedir. Birçok veri madenciliği çalışmasında en fazla zaman alan kısım verilerin temizlenmesi ve veri bütünleştirme kısmıdır demek yanlış olmayacaktır. Yapılan araştırmalara göre veri madenciliği uygulamalarında harcanan sürenin yaklaşık %80'inin veri önleme aşamalarından oluştuğu bildirilmektedir (Piramuthu, 2003: Aktaran: Oğuzlar, 2003, s. 70). Kullanılacak olan verilerin eksik ya da yetersiz olması verilerin analizini zorlaştıracak ve hatta hatalı bir analiz yapmaya sebep olacaktır. Toplanan verilerin hatalı olmasına sebep olarak; yanlış donanımla veri toplama, veri girişlerinin yanlış yapılması, kullanıcı hataları, verinin iletiminden kaynaklı oluşan hatalar, teknolojik sınırlamalar, verilerin yapılarından kaynaklanan uyumsuzluklar gösterilebilir (Artinyan, 2019, s. 1-2).

2.2. Müzik Madenciliği

Veriye ulaşımın kolaylaşması ve bilgi çağının avantajları birçok alanda etkisini gösterdiği gibi müzik alanında büyük miktarlardaki verilere ulaşımın sağlanmasını beraberinde getirmiştir. Müzikle ilgili verilerin artışıyla birlikte bu alanda yapılan çalışmalar ve analizler günden güne artış göstermektedir (Li, Ogihara & Tzanetakis, 2011, s. 4-7). Müzik insanların hayatında gün geçtikçe daha da önemli bir rol almaktadır. Dijitalleşme ve teknolojinin gelişmesine bağlı olarak da sadece kişilerin raflarındaki plak, kaset ya da cd'lerde kalmaktan çıkmış dijital ortamda ciddi yer kaplayan veri yığınları hâline gelmiştir. Artık istenilen her yerde ulaşılabilen açık kaynaklar söz konusudur. Bu veri yığınlarının artmasıyla birlikte müzikle ilgili tanımlamaların yapılması, müzik verilerinin birbiriyle ilişkilerinin kurulması ihtiyacı, bu alanda veri madenciliği ve makine öğrenimi çalışmaları yapılmasını bir zorunluluk hâline getirmiştir (Jensen, 2006, s. 2).

Müzik verisi tek yönlü olmayıp birçok farklı yönlerden incelenmesi gereken türdendir. Müzik ile ilgili olarak müziğin tınısı, ritmi, sözleri, melodik yapısı, türü, frekans aralıkları vs. gibi birçok yönden farklı veriler elde edilebilir ve gruplandırma yapılabilir. İnsanlar için müziği algılama, duylara ve hislere bağlı olarak gerçekleşmektedir. Bu algıyı bilgisayarlarla yakalamak daha zordur ancak farklı yönlerden müzik verilerinin elde edilebilmesine yönelik çalışmaların gelişimi, müziğe yönelik veri toplama ve istatistiksel analiz ihtiyaçlarının doğması ve disiplinler arası çalışmaların da artmasıyla Müzik Bilgisi Edinme (Music Information Retrieval) çalışmaları ortaya çıkmıştır (Downie, 2003; Aktaran: Karatana & Yıldız, 2018).

Son yıllarda özellikle farklı veri türlerinin ortaya çıkmasıyla birlikte veri madenciliği ve müzikle ilgili birçok çalışma yapılmaktadır. Bunun yanında müzik madenciliği sosyal medyada ve internet uygulamalarında da görülmektedir. Kullanıcıların daha önce yaptığı aramalara göre sınıflandırma yaparak yeni önerilerde bulunulması, müzik türlerine göre sınıflandırma yapılarak çalma listesi oluşturması gibi çalışmalar veri madenciliği çalışmalarına örnek olarak gösterilebilir.

Müzik madenciliğiyle ilgili dünyada yapılan farklı çalışmalar olsa da henüz Türk Müziğiyle ilgili yapılan analizler çok azdır. Klasik Türk Müziği notasyon sistemi, armonik yapısı, makâmsal yapı, ses sistemleri vs. gibi birçok özelliğinden dolayı Klasik Batı Müziğinden farklılık göstermektedir (İmik, 2017). Bu sebeple mevcut çalışmalar ve yazılımlar Klasik Türk Müziği analizlerini yapmak için yetersiz kalmıştır. Klasik Türk Müziği notasyon sistemindeki farklılıklar, Batı Müziğinden daha karmaşık yapıda olması ve yazılım eksiliğinden dolayı Türk Müziği eserlerinin sayısal ortama aktarılmasıyla ilgili de problemler bulunmaktadır (Abidin vd., 2017, s. 1222-1223).

Tarih boyunca Klasik Türk Müziğinde farklı nota yazım sistemleri geliştirilmiş ancak bu nota yazım sistemleri genellikle notaya bakarak eser icrâ etmek için değil, eserlerin kayıt altına alınması için kullanılmıştır. Eserlerin icrâ edilmesiyle ilgili kurallar ise “meşk” dediğimiz yöntemle nesiller boyunca aktarılmıştır. Bu sebeple mevcuttaki notasyon sistemi Klasik Türk Müziğinin makâmsal yapısına uygun değildir. Mevcuttaki sistemde notalar belirli bir frekansı ifâde etmekte iken fiiliyatta makâmın melodik yapısına ya da o anda icrâ edilmekte olan melodinin oluşturduğu müzikal duyuşa göre bir frekans bandını ifâde eder (Çevikoğlu, 2017). Ancak tüm bu olumsuzluklara rağmen son zamanlarda Türk Müziği alanında da çok ciddi çalışmalar yapılmaya devam etmektedir.

3. USÛL VE ARÛZ KAVRAMI

3.1. Türk Mûsikîsinde Usûl Kavramı

Müzik dili yapısal bakımdan melodik ve ritmik yapılar olarak iki kısımda incelenebilir. Melodik yapının karşılığı olan seslerin sayısal birimi frekansken, ritmik yapının sayısal karşılığı zamandır şeklinde tanımlanabilir. Türk Müziğinde ses dizilerinin belirli bir âhenk içerisinde kullanılmasından doğan kavrama makâm denilirken, ritmik yapıların zaman kavramının sadece sayısal olarak değil yapısal olarak da kalıplaşmış şekline usûl denir. Usûl sözcüğü dilimizde “yol, yordam, yöntem, metod” anlamlarında kullanılmaktadır. Kelimenin kökeni Arapça “asıl” anlamına gelen “asl” sözcüğünün çoğuludur. Türk Müziğinde kullanılan usûl kavramını açıklayabilmek için öncelikle ritm ve ölçü kavramlarını açıklamak gerekecektir. Söz konusu olan zaman kavramı yani ritm, zamandaki düzgün ve kesintisiz akıp giden düzene denir. Buradaki zaman kavramı “düzüm” ya da “ikâ” denilen birimlerce tanımlanmasıyla oluşmaktadır. Buradaki ikâların düzgün oranlı sürelerden düzenlenmiş kümelere bölünmesiyle birlikte de ölçüler oluşur. Usûl kavramı da bu ölçülerin belirli amaçlarla kalıplaştırılmış şekli olarak tanımlanabilir (Tanrıkorur, 2018, s.137-142). Literatürdeki diğer bazı usûl tanımlarını inceleyecek olursak, Hurşit Ungay’a göre usûl tanımı; “*Belirli düzümlerden yapılarak kalıp hâlinde saptanmış ölçülerdir*” şeklindedir (Ungay, 1981, s.4-5). 17. yüzyılda Osmanlı Sarayında yetişmiş bir bilgin olan Dimitrius Kantemir ise, Kantemiroğlu Edvârı olarak tanınan eserinde usûl kavramını; “*Usûl Türk müziğinin tartısı ve ölçüsüdür. Usûlün gücü kullanılarak nağmenin anlamlandırılmasının tam kıvamında olması sağlanmalıdır.*” şeklinde ifâde etmektedir (Tura, 2001, s.158).

3.2. Arûz Vezni

Kelime olarak Arapça’da çadırın orta direği anlamına gelmektedir. Arapçadaki uzun ve kısa vurguların ardı ardına gelmesine bağlı olarak iç âhenk oluşturmuş bir şiir ölçüsüdür. 7. yüzyılda yaşamış Mekkeli bir dilbilimci olan İmam Halil tarafından bilim hâline getirilmiştir. İslâmiyetle birlikte özellikle Orta Asya’da yaygınlaşan Arap kültürü; İran, Afgan, Pakistan, Türk ve Hint şâirlerinin de etkilenmesine sebep olmuştur. Böylece bu dillerde kullanılan şiir formlarının da vezni hâline gelmiştir (Tanrıkorur, 2003, s.87-89).

Arûz kavramı, şiirde kullanılan ölçü kalıpları olarak kısa bir tanımla tanımlanabilir. Bu kalıplar peş peşe eklenen ve “tef’ile” olarak adlandırılan parçacıkların birleşmesiyle meydana gelmektedir. Tef’ileler en az bir en çok beş heceli olabilmektedir (Tanrıkorur, 2018, s.359-366).

Arûz vezninde 3 hece tipi bulunmaktadır. Bunlar; “Açık-Kısa Heceler”, “Kapalı-Uzun Heceler” ve “Kapalı ve Açık-Uzun ve Kısa Heceler” şeklinde sınıflandırılabilirler. Bununla birlikte mısra sonundaki hecelerin çeşidine bakılmaksızın daima kapalı sayılması gerektiği bir kuraldır (Timurtaş, 1994: Aktaran: Bildik, 2010, s. 12).

4. ÇALIŞMANIN YÖNTEMİ

4.1. Materyal

Bu çalışmada kullanılan veri tabanı, TRT nota repertuarı ve bazı özel repertuarların müzik arşivindeki eserler bir Java yazılımı kullanılarak internetten indirilmiş ve MS Excel programına aktararak bir veri tabanı oluşturulmuştur. Analiz çalışmasından önce veri kümesinde temizleme ve veri bütünleştirme işlemi yapılmıştır. Veri sayısının çok olması, verilerdeki yazım hatalarının çok olması, aynı bilgilerin farklı şekilde girilmesi, tekrâr eden veri sayısının çok olması vb. gibi durumlardan dolayı veri kümesi dağınık bir yapıdadır. Bununla birlikte yapılacak analizleri olumsuz etkileyecek, veri gürültüsü olarak nitelendirilebilecek

veriler ve herhangi bir usûl kalıbı içermeyen ya da sözlü olmayan eserlere ait verilerin tamamı veri kümesinden çıkarılmıştır. Bu veri temizleme işleminden sonra analiz yapılmaya uygun veri sayısı 43.936 adet veriden 28.300'e düşmüştür. Veri tabanındaki bulunan veri kümelerine ait satır başlıkları Tablo 1'de gösterildiği gibidir.

Tablo 1: Veri tabanındaki veri kümelerinin başlıkları

Sıra No	Sütun Başlığı
1	Eserin İlk Dizesi
2	Eserin Bestekârı
3	Eserin Söz Yazarı
4	Eserin Makâmı
5	Eser Formu
6	Eserin Usûlü

4.2. Metot

Öncelikli olarak veriler bir Java yazılımı kullanılarak veriler internetten indirilmiş ve MS Excel programında bir veri tabanı oluşturulmuştur. Analiz için uygun veriyi elde edebilmek için SQL ve MS Excel programları kullanılmıştır. Veri temizlenmesi ve bütünleştirilmesi aşamasında öncelikli olarak tekrâr eden verilerin tespit edilmesi ve veri tabanından çıkarılması, analize uygun olmayan verilerin temizlenmesi, verilerdeki yazım yanlışlarının tespit edilmesi ve düzeltilmesi, veri gürültüsü olarak nitelendirilebilecek verilerin tespit edilmesi ve temizlenmesi işlemleri yapılmıştır. Veri temizleme işlemleri için SQL ve MS Excel yazılımlarındaki arama ve başvuru, metin yerine koyma, özet tablo gibi formüller ve Excel için makro kodları kullanılarak verilerin temizlenmesi ve bütünleştirilmesi işlemleri sağlanmaya çalışılmıştır.

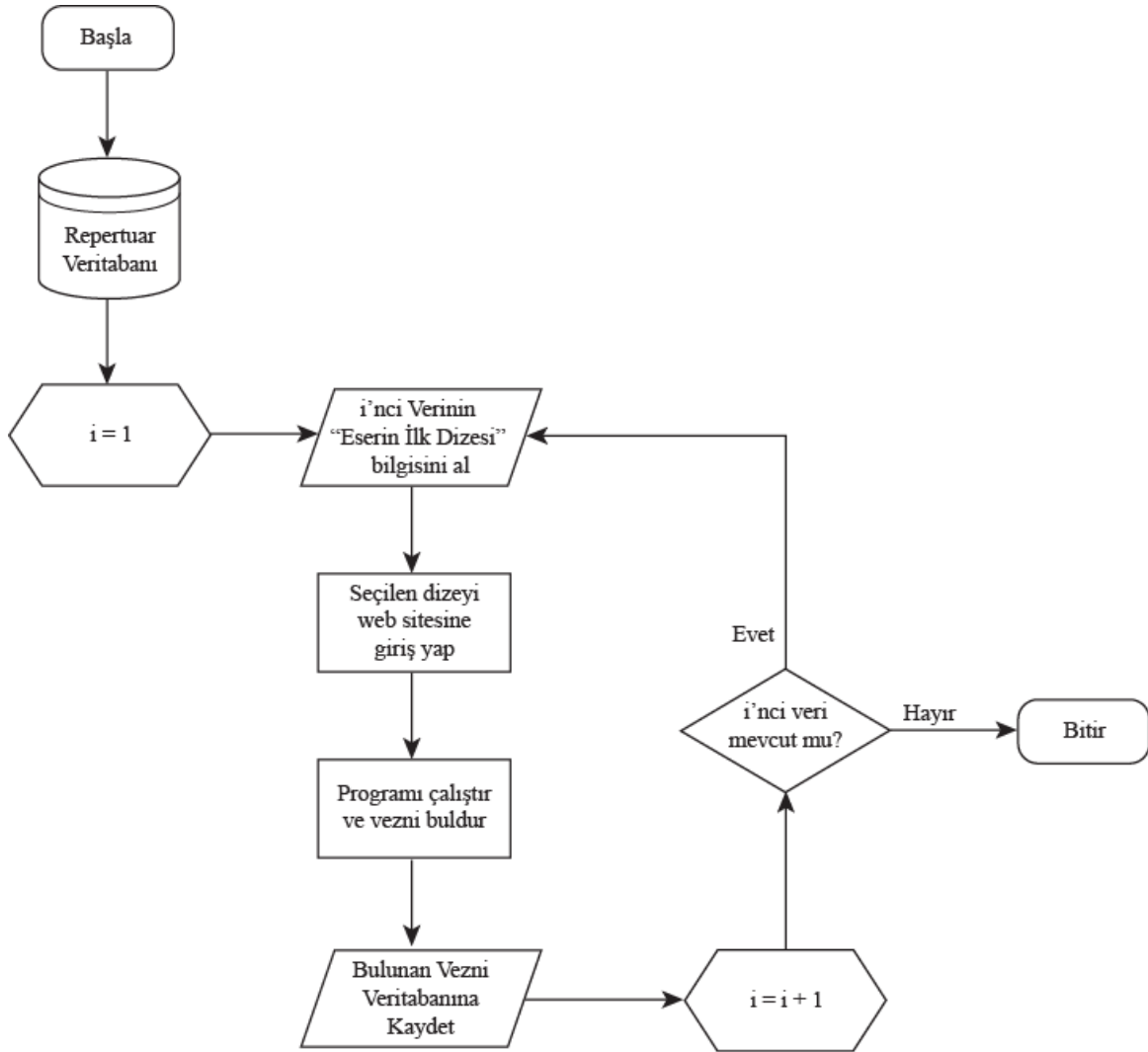
Söz konusu olan veri kümesinin farklı kaynaklardan toplanmış olması ve daha da önemlisi veri yapısının çok dağınıktır. Verilerin düzenlenmesi işlemi için veri içeriklerinde öncelikli olarak “Eserin İlk Dizesi” kolonundaki verilerdeki yazım yanlışları bulunmuştur. Bu kolonda yer alan veriler içerisinde şarkı isimlerinde yeri olmayan parantez işareti, yıldız işareti, soru işareti, ünlem işareti gibi değerleri içeren veriler SQL sorgusuyla getirilmiş ve buradaki yazım yanlışlarında manuel düzenlemeler yapılmıştır. Düzeltmeleri yapılan veriler Excel veri tabanına aktarılmıştır. Excel yazılımında bulunan yazım denetimi eklentisini kullanarak veri kümesindeki hatalı verilerin hangi verilerde olduğu makro koduyla tespit edilmiş ve ayrı bir veri tabanına kaydedilmiştir. Alınan verilerdeki imlâ hatalı kısımlar kontrol edilerek manuel girişlerle düzeltmeler yapılmıştır. Bu düzeltmeden sonra tekrâr aynı olan veriler için arama yaptırılmış ve birbiriyle aynı olan ancak yazım hatalarından ya da verinin düzgün olarak tutulmamasından kaynaklanan sebeplerden dolayı farklı veriymiş gibi görünen birçok veri tespit edilerek veri kümesinden çıkarılmıştır.

Veri kümesinde veri gürültüsü olarak nitelendirilebilecek verilerin, veri kümesinden temizlenmesi doğru bir analiz açısından önem arz etmektedir. Bu tip verileri saptayabilmek için öncelikli olarak ihtiyaç olan frekans tabloları çıkarılmış ve bu aşamada analizde kullanılabilecek olan veriler ayrı ayrı incelenmiştir.

Frekans tabloları “Eserin Makâmı”, “Eserin Formu”, “Eserin Usûlü” ve “Eserin Bestekâri” sekmelerindeki bilgilerinde ayrı ayrı oluşturulmuş ve bu veriler arasında veri gürültüsü olarak nitelendirilebilecek veriler ya da analiz için uygun olmayan veriler saptanmış ve veri kümesinden çıkarılmıştır. Yine bu işlem yapılırken de Excel yazılımındaki eklentiler kullanılmıştır.

Yapılacak olan analiz çalışmasındaki hipotez için ihtiyaç olan zaman verisinin belirlenmesi için bestekâra göre bir sıralama yapılmış ve eksik olan veriler bestekârların doğum tarihlerine göre manuel giriş yapılarak veri tabanına dahil edilmiştir.

Veri kümesinde geçen eserlerin arûz veznini bulmak içinse “Aruz Vezninin Bilgisayarla Çözülmesi” projesi kapsamındaki yazılım kullanılmıştır (Kara & Kurt, 2007). Eserin ilk dizisine göre alınan sonuç Visual Basic tabanlı bir yazılımla internet sitesi üzerinden arama yaptırılmış ve gelen sonuçlar yine veri tabanına kaydedilmiştir. Arûz vezninin bulunmasıyla ilgili yapılan yazılıma ait algoritma Şekil-2’de verilmiştir.



Şekil 2: Arûz vezninin bulunması ve veri tabanına eklenmesi prosesine ait akış diyagramı

5. BULGULAR

Veri temizleme işlemi tamamlandıktan sonra eserlerin bestekârlarına göre bestelendiği dönemler yüzyıl olarak tespit edilmeye çalışılmıştır. Söz konusu eserlerin dönemleri bestekârların doğum tarihlerine ve yaşadıkları dönemlere göre tertip edilip veri kümesine eklenmiştir. Eserlerin bestelendiği tarihler bestekârların dönemine göre verildiği için yüzyıl sonlarında yaşamış olan bestekârların yüzyılında sapma söz konusu olabilmektedir.

Tablo 2: Veri kümesinin dönemlere göre dağılımı

Bestelendiği Dönem (yy)	Veri Adedi	Frekans
14. yy	23	0,08%
16. yy	12	0,04%
17. yy	114	0,40%
18. yy	690	2,44%
19. yy	5.223	18,46%
20. yy	21.578	76,25%
Bilinmeyen	660	2,33%
Toplam	28.300	100,00%

Verilerin dönemlere göre dağılımı Tablo 2’de gösterilmiştir. Geçmiş yüzyıllardan gelen eserlerin frekansının az olmasının sebebi önceki yüzyıllarda Türk Müziğinde nota yazım sistemlerinin aktif olarak kullanılmayışı, eserlerin meşk silsilesiyle nesilden nesile aktarılması ve eserlerin bir kısmının da tahrip olması gösterilebilir. Bununla birlikte bestekârı belli olmayan 660 adet eserin bestelendiği dönem bilinmeyen olarak kaydedilmiştir. Çalışma kapsamında aruz vezni kullanılarak yazılan bestelerin yüzyıllara göre dağılımının belirlenerek yapılacak olan veri madenciliği için bir ön analiz oluşturulmuştur.

Veri temizliği yapıldıktan sonra eserlerin arûz vezninin belirlenmesi için eserin ilk dizesine ait güftelere bağlı olarak eserin yazıldığı şiirin arûz vezni bulunmuştur. Söz konusu vezin bulma işlemi için “Aruz Vezninin Bilgisayarla Çözülmesi” isimli projeden yararlanılmıştır. Veri tabanındaki verilerin vezinlerinin bulunması ve veri tabanına kaydedilmesi işlemi Visual Basic tabanlı bir yazılım yapılarak sağlanmıştır. Mevcut veri tabanında bulunan eserin ilk dizesinden vezin işlemi buldurulmuş ve elde edilen veriler ayrı bir sütun olarak kaydedilerek gerekli veri dönüştürme işlemi yapılmıştır.

Oluşturulan yeni veri tabanında web sitesinden alınan verilerin doğruluğu için vezinlerin kontrolü yapılmıştır. Arûz vezni yazılımı kelimelerin hecelerine göre ayırıştırma yaparak uygun vezin bulma işlemi yaptığı için cümle yapısı uygun olmasa da bir vezin bulma işlemi yapılmış ve bir sonuç vermiştir. Bulunan vezinler arasında edebi literatürdeki vezin kalıpları arasında yeri olmayan vezinler veri tabanından silinmiştir. Veznin bulunamamasının eserin güftesinin gerçekten arûz vezniyle yazılmaması dışında iki sebebi olabilir; birincisi yazılan güftenin birinci dizesinin doğru bir şekilde yazılmaması ikincisi ise güftenin eksik yazılması sebebiyle vezin kalıbına uygun bir hâle gelmemesi. Genel sonuç değerlendirilirken bu hata payı göz önünde bulundurularak değerlendirilme yapılmalıdır.

Veri temizleme işlemi yapıldıktan sonra yapılan veri analizi sonucunda çıkarılan frekans tablosu Tablo 3’de gösterilmiştir.

Tablo 3: Arûz vezniyle yazılmış güftelerin dağılım oranı tablosu

Bestelendiği Dönem (yy)	Toplam İncelenen Eser Sayısı	Vezinle Yazılan Eser Sayısı	Vezinle Yazılan Eserlerin Oranı
14. yy	23	20	87,0%
16. yy	12	9	75,0%
17. yy	114	113	99,1%
18. yy	690	503	72,9%
19. yy	5.223	4.748	90,9%
20. yy	21.578	2.604	12,1%
Bilinmeyen	660	165	25,0%

Frekans tablosunda da görüldüğü gibi 20. yy öncesinde bestelenen verilerin arûz vezniyle bestelenmesinin oranı çok yüksektir. Her ne kadar incelenen veri kümesindeki veri gürültüsü olarak nitelendirilebilecek değerler veri kümesinden çıkarılmış ve veri bütünlüğü sağlanması için veri temizleme ve birleştirme işlemi yapılmışsa da veri kümesindeki bilgilerin genişletilmesiyle daha sağlıklı analiz sonuçlarına ulaşmak mümkün olacaktır.

6. SONUÇ

Bilgi özellikle son yıllarda her alanda ve herkes için çok önemli bir kavram hâline gelmiştir. Ancak çok fazla verinin olması ve bu verilerin kullanılabilir bilgiye dönüştürülmesi iyi yönetilmesiyle mümkün olabilecektir. Bilgi hâline dönüştürülmeyen ham verinin herhangi bir faydası bulunmayacaktır. Her ne kadar teknolojiye ilerleme hızlı bir şekilde gerçekleşiyor olsa da mevcut verilerin iyi bir şekilde kullanılması ve analiz edilmesi yani iyi yönetilmesi organizasyonların ayakta kalabilmeleri için en önemli unsur hâline gelmektedir.

Bununla birlikte yapılan veri madenciliği çalışmalarında verinin hazırlanma aşamasının süreç içerisinde en çok zaman alan kısım olması, veri kalitesi ve bütünlüğünün önemini göstermektedir. Verilerin yapılarındaki bozukluk, farklı kullanıcıların farklı biçimlerde ya da eksik veri girişi, tutarlı olmayan ya da dinamik olmayan veri yapıları, veri kalitesini ve bütünlüğünü bozan unsurlar olarak gösterilebilir. Bunların tamamının veri ve verinin bilgiye dönüşüm süreciyle ilgili bilinç eksikliğinden kaynaklandığını söylenebilir.

Yapılan analizin sonucuna göre arûz vezniyle yazılan eserlerin büyük bir çoğunluğunun 20. yüzyıldan önce bestelenen eserlerde olduğu görülmektedir. Bu durum son yüzyıllarda meşk halkalarının da kopmasıyla birlikte kullanılan güftelerdeki şiirlerin arûz vezniyle yazılması dikkate alınmaya başlanmıştır. Bu sonuç için gelenekte çok önemli bir unsur olan usûl-arûz münasebetinin artık önemsenmediğini gösteren bir unsur demek yanlış olmayacaktır. Her ne kadar bu alanda daha önce münferit çalışmalar yapılmış olsa da bilgisayar yazılımları kullanılarak geniş veri kümeleriyle yapılan veri çalışması bulunmamaktadır. Türk Müziği alanındaki verilerin veri analizine daha uygun hâle getirilerek daha detaylı veri madenciliği çalışmaları yapılmasının bir ihtiyaç olduğu ortaya çıkmaktadır.

KAYNAKÇA

- Abidin, D., Öztürk, Ö. ve Öztürk, T.Ö. (2017). Klasik Türk müziğinde makam tanıma için veri madenciliği kullanımı, *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 32(4), 1221-1232.
- Alpaydın, E. (2000). *Zeki veri madenciliği: Ham veriden altın bilgiye ulaşma yöntemleri*, Bilişim 2000 Eğitim Semineri.
- Arslan, H. (2008). *Sakarya Üniversitesi web sitesi erişim kayıtlarının web madenciliği ile analizi* (Yayınlanmış yüksek lisans tezi). Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Artinyan, E. N. (2019). Veri analizi – Veri kalitesi ve bütünlüğü, Erişim tarihi: 25.03.2019, http://www.denetimnet.net/UserFiles/Documents/Makaleler/BT%20Denetim/Veri_Analizi_Veri_Kalitesi_ve_B%FCt%FCnl%FC%F0%FC.pdf
- Bildik, Ö. (2010). *Râkım Elkutlu'nun eserlerinde usûl arûz vezni ilişkisi* (Yayınlanmış yüksek lisans tezi). Afyon Kocatepe Üniversitesi Sosyal Bilimler Enstitüsü, Afyon.
- Çevikoğlu, T. (2017). *Mevlevî müziği eserlerinin yazımında kullanılan eski yazım sistemleri ve çeviri sorunları*, Mevlana Araştırmaları Enstitüsü Uluslararası Mevlana Sempozyumu, Konya.
- İmik, Ü. (2017). “Klasik” kavramı ekseninde Türk Müziği ve Batı Müziği kültüründeki benzerlikler/farklılıklar üzerine genel bir değerlendirme, *İnönü Üniversitesi Kültür ve Sanat Dergisi*, 3(1), 199-205.
- Jensen, B.S. (2006). *Exploratory data mining in music* (Yayınlanmış yüksek lisans tezi). Technical University of Denmark Department of Informatics and Mathematical Modelling, Kongens Lyngby.
- Kara, M. ve Kurt, A. (2007). Aruz vezninin bilgisayarla çözülmesi, Erişim tarihi: 05.01.2019, <http://161.9.143.104/aruz/?Home>
- Karatana, A. ve Yıldız, O. (2017). Müzik türlerinin makine öğrenmesi yöntemleri ile sınıflandırılması, Erişim tarihi: 20.07.2018, <https://www.researchgate.net/publication/320281584>
- Li, T., Ogihara M. ve Tzanetakis G. (2011). *Music Data Mining*, Boca Raton: CRC Press.
- Oğuzlar, A. (2003). Veri ön işleme, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21, 67-76.
- Sarıman, G. (2011). Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-means ve K-medoids kümeleme algoritmalarının karşılaştırılması, *SDÜ Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202.
- Silahtaroglu, G. (2016). *Veri Madenciliği: Kavram ve Algoritmaları*, İstanbul: Papatya Yayıncılık.
- Sumathi, S. ve Svendam, S.N. (2006). *Introduction to Data Mining and It's Applications*, Berlin: Springer.
- Tanrıkorur, C. (2003). *Osmanlı Dönemi Türk Müsîkîsi*, İstanbul, Dergâh Yayınları.
- Tarıkorur, C. (2018). *Müzik Kültür Dil*, İstanbul, Dergâh Yayınları.
- Tura, Y. (2001). *Kantemiroğlu, Müsîkîyi Harflerle Tesbît ve İcrâ İlminin Kitabı*, İstanbul, Yapı Kredi Yayınları.
- Ungay, H. (1981). *Türk Müsîkisinde Usûller ve Kudüm*, İstanbul, (b.y.).
- Yüreğir, O.H. (2003). Türkiye'deki tekstil işletmelerinde bilgi ve bilişimin yeri, *TSE Standart Dergisi*, 497, 39-49.