

Yayın Geliş Tarihi (Submitted): 29/05/2020

Yayın Kabul Tarihi (Accepted): 21/06/2020

Makele Türü (Paper Type): Araştırma Makalesi – Research Paper

TÜRKİYE’DE LİSANS DÜZEYİNDEKİ İSTATİSTİK DERS İÇERİKLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ

Fikriye ERDOĞAN¹ ve Güzin YÜKSEL²

ÖZET

Günümüzde teknolojinin gelişmesi ile birlikte veri miktarlarında ve çeşitliliğinde hızlı artışların olduğu gözlenmektedir. Veri artışları büyük ve karmaşık veri kümelerini oluşturmaktadır. Veri madenciliği, büyük ve karmaşık olan bu veri kümelerini veri madenciliği yöntemleri ile kolay bir şekilde analiz ederek yararlı olan anlamlı bilgilerin ortaya çıkarılmasını sağlamaktadır. Bu çalışmada, Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümü bulunan diğer üniversitelerin lisans programlarında yer alan derslerin ders içerikleri karşılaştırılmıştır ve veri madenciliği yöntemleri ile analizleri yapılmıştır. Yapılan analizlerde; istatistik bölümlerinin birbirleri ile olan ilişkileri ve istatistik bölümlerinde birlikte alınması gereken derslerin veri madenciliği ile belirlenmesi amaçlanmıştır. Uygulamada veri madenciliği yöntemlerinden kümeleme analizi ve birliktelik kuralları analizi kullanılmıştır. İstatistik bölümleri bulunan üniversitelerin birbirleri ile olan benzerlikleri kümeleme analizi ile gösterilmiştir. İstatistik bölümlerinin ders programlarında hangi derslerin birlikte alınması gerektiğini belirlemek için ise birliktelik kuralları analizi kullanılmıştır. Analizler için KNIME veri madenciliği programı kullanılmıştır. Kümeleme analizi, Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer 19 üniversitenin birbirlerine benzerlik gösterdiğini ifade etmiştir. Birliktelik kuralları analizi sonucunda ise; 21 tane birliktelik kuralı oluşmuştur. Bu kurallar incelenerek; istatistik bölümünün lisans programlarında bulunması gereken önemli derslerin Olasılığa Giriş, Matematik 1, Kombinatorik, Aktüerya Modelleri, Lineer Cebir 1 ve Lineer Cebir 2 olması gerektiği tespit edilmiştir.

Anahtar Kelimeler: Veri Madenciliği, İstatistik, Ders İçeriği

¹ Sorumlu Yazar, Çukurova Üniversitesi, İstatistik Bölümü, Adana, Türkiye, ORCID ID: <https://orcid.org/0000-0001-6972-5155>.

² Prof. Dr., Çukurova Üniversitesi, İstatistik Bölümü, Adana, Türkiye, ORCID ID: <https://orcid.org/0000-0002-1644-3696>.

ANALYSIS OF THE COURSE CONTENTS OF STATISTICS AT BACHELORS LEVEL IN TURKEY WITH DATA MINING TECHNIQUES

ABSTRACT

Today, with development of technology, it is take noticed that there is a rapid increase in the amount and diversity of data. Data increases are large and complex data sets. Data mining can easily analyze these large and complex data sets with data mining methods to reveal useful information. In this study, Çukurova University Department of Statistics and other universities with statistics department were compared and data mining methods were analyzed. In the analyzes; The relation ship between the statistical sections was examined. However, it is aimed to determine the courses that should be taken together in statistics departments with data mining. In this application, cluster mining and association rules analysis, one of the data mining methods, were used. The similarities of the universities with statistical departments were shown by clustering analysis. The association rules analysis was used to determine which courses should be taken together in the curriculum of the departments of statistics. KNIME datamining program was used for analysis. Clustering analysis indicated that Çukurova University Department of Statistics and 19 other universities with statistics department were similar to each other. As a result of the association rules analysis; There are 21 association rules. By examining these rules; It has been determined that the major courses required in the undergraduate programs of the department of statistics should be Introductionto Probability, Mathematics 1, Combinatorics, Actuarial Models, Linear Algebra 1 and Linear Algebra 2.

Keywords: Data Mining, Statistics, Course Contents

1. GİRİŞ

Son yıllarda bilgisayar teknolojisindeki gelişmeler ile veri miktarlarında ve çeşitliliğinde hızlı bir artış olduğu gözlenmektedir. Bu veri artışları büyük ve karmaşık veri kümelerini oluşturmaktadır. Büyük ve karmaşık olan veri kümelerinde bulunan bilgilere ulaşmak hem zor

hem de zaman almaktadır. Ulaşılması zor olan ve zaman alan bu bilgilerin oluşturulması veri madenciliği ile sağlanmaktadır. Veri madenciliği büyük ve karmaşık veri kümelerindeki yararlı olan anlamlı bilgilerin ortaya çıkarılmasını kolay bir şekilde yapmaktadır. Veri madenciliği analizleri, elde bulunan veri kümelerinin durumuna ve verilerin kullanım amacına göre veri madenciliği yöntemleri ile yapılmaktadır. Veri madenciliği yöntemlerini uygulamadan önce analiz için önemli olacak değişkenlerin belirlenmesi gerekmektedir. Değişkenler belirlendikten sonra veri madenciliği yöntemleri ile analizler gerçekleştirilmektedir. Böylece, veri madenciliği ile doğru bilgilere hızlı ve güvenilir bir şekilde ulaşılmaktadır.

Veri madenciliğinin kullanımı belirli uygulama alanlarıyla sınırlandırılmaz. Veri madenciliğini verinin üretilip kayıt altına alındığı her alanda kullanmak mümkündür. Sağlık, endüstri, mühendislik, pazarlama, bankacılık ve eğitim alanları veri madenciliğinin yoğun olarak kullanıldığı başlıca uygulama alanlarıdır (Han ve ark., 2006).

Eğitim, veri madenciliğinin yaygın olarak kullanılan alanlarından biridir. Eğitim alanlarında yapılan analizlerde; öğrencilerin başarı durumlarının tespit edilmesi, ders içeriklerinin incelenmesi ile ders programlarının hazırlanması gibi eğitimin kalitesini arttıracak konular veri madenciliği ile gerçekleştirilmektedir.

Veri madenciliği yöntemlerinin eğitim alanında kullanılmasıyla birlikte ortaya çıkan, literatürdeki adıyla eğitsel veri madenciliği (educational data mining), öğrencilere ve okullara ait çoklu verinin işlenmesine olanak sağlayan ve kullanım alanı oldukça yaygın olan bir yöntemler topluluğudur. Özellikle eğitim alanındaki kullanımı, eğitimcilere ve eğitim planlamacılarına farklı açılardan ışık tutarak eğitim stratejilerinin belirlenmesinde ihtiyaca yönelik çözümler sunabilmektedir. Eğitsel veri madenciliğini, eğitim ortamından elde edilen verilerin daha önceden bilinmeyen ya da oluşturulmamış yapısını keşfetmek için yöntem geliştirme ve geliştirilen yöntemlerin öğrencileri daha iyi tanımada kullanan gelişmekte olan bir disiplin alanı olarak tanımlamak mümkündür (Bilen ve ark., 2014).

Üniversiteler ile ilgili yapılan veri madenciliği çalışmaları şu şekildedir:

Kabra ve Bicher tarafından (2011) tarafından ele alınan diğer çalışmada ise; karar ağacı kullanılarak model oluşturulmuş, bu modelde öğrencilerin akademik performansları kullanılması ve karar ağacının sonuçları mühendislik sınavlarında öğrencilerin performansının tahminde kullanılması ele alınmıştır (Kabra vd., 2011).

Hırvat ve İskoç öğrencilerinin davranışlarının karşılaştırılması ve öğrenci algılarına göre homojen iki grubu ilişkisel olarak belirlenmesi amaçlanan çalışmada, veri madenciliği bağlamında istatistiksel ve küme analizleri kullanıldığında elde edilen sonuçlar gösteriyor ki, İskoç öğrencileri genellikle Hırvat öğrencilerinden daha olumlu ve e-öğrenme yeteneğine sahiptir (Darko, 2011).

Delavari ve Beikzadeh veri madenciliğinin kullanımı için genel kurallar vermiş olup, farklı veri madenciliği tekniklerinin eğitim verilerine ne kadar uygulanabilirliğini tartışmıştır (Delevari, 2008).

Bilgisayar mühendisliği lisans programlarının ders içerikleri ile Erasmus Programıyla gidilen Avrupa ülkelerindeki bilgisayar mühendisliği lisans programlarının ders listelerindeki dersler karşılaştırılarak veri madenciliği çalışması yapılmıştır. Bu tez çalışmasında Türkiye'deki bilgisayar mühendisliği lisans programı eğitim kalitesini artırabilmek amacıyla Türkiye'deki üniversitelerden çoğunun bilgisayar mühendisliği lisans programı dersleri ve bu üniversitelerin Erasmus anlaşması yaptığı üniversitelerdeki dersleri toplanmıştır. Elde edilen bu dersler, veri madenciliği algoritmalarından J48 algoritması ve Naive Bayes algoritmaları ile incelenmiş ve bu yöntemlerden Naive Bayes algoritması ile %98'lik bir başarı elde edilmiştir. Bu tez çalışmasının sonucunda, Türkiye ve Avrupa'daki bilgisayar mühendisliği lisans programı ders içeriklerinin dağılımları incelenerek, ağırlık verilmesi gereken ders grupları belirlenmiş ve ders listelerine eklenilmesi gereken dersler tespit edilmiştir. Ayrıca derslerin kredi üzerine yapılan çalışması sonucunda hem Erasmus programı ile yurt dışına giden öğrencilerin mağduriyetini önlemek hem de dersin öğrenci gözündeki önemini artırmak için kredilerde düzenleme yapılması gerektiği sonucuna varılmıştır (Kılıçer, 2018).

Türk Üniversitelerindeki Bilgisayar Mühendisliği Bölümleri müfredatlarında yer alan derslerin ders içerikleri incelenerek bir eğitsel veri madenciliği uygulaması gerçekleştirilmiştir. Bu tez çalışması ile üniversitelerdeki Bilgisayar Mühendisliği bölümlerinin müfredatlarının karşılaştırılarak benzerliklerinin ortaya çıkarılması, veri madenciliği tekniklerinden olan Kümeleme yönteminin Bilgisayar Mühendisliği ders planları üzerinde uygulanması, Karar Ağacı yöntemi ile Bilgisayar Mühendisliği bölümlerinin başarısına etki eden derslerin belirlenmesi, Birliktelik Analizi algoritmaları ile bölüm derslerinin birlikteliklerinin bulunması amaçlanmıştır. Bu çalışma, Bilgisayar Mühendisliği bölümlerinin müfredatlarına göre üniversitelerin benzerliğini ortaya çıkarmasına yardımcı olmuştur. Aynı zamanda

üniversitelerde yeni kurulacak olan Bilgisayar Mühendisliği bölümleri için ders planı ve ders içeriği oluşturma konusunda yardımcı olacaktır (Güven, 2016).

Üniversitelerin aynı bölümlerinin ders içerikleri incelendiğinde, derslerin birbirinden farklı olduğu görülmüştür. Bu yüzden, farklı üniversitelerin aynı bölümündeki ders içeriklerinin birbirleriyle olan benzerlikleri, bu benzerlikler ile üniversitelerin birbirleriyle ne kadar ilişkili olduğu durumlara çözüm bulmak amacıyla bu çalışmada, Çukurova Üniversitesi İstatistik Bölümü ile ülkemizde bulunan diğer istatistik bölümlerinin birbiriyle olan ilişkilerini belirlemek amacıyla istatistik bölümlerinin ders içerikleri incelenmiştir.

Kılıçer (2018) Türkiye'deki bilgisayar mühendisliği lisans programlarının ders içerikleri ile Erasmus Programı ile gidilen Avrupa ülkelerindeki bilgisayar mühendisliği lisans programlarının derslerini karşılaştırarak veri madenciliği yöntemlerinden Naive Bayes ve J48 algoritmasını kullanarak WEKA programında analizini gerçekleştirmiştir.

Güven (2016) ise ülkemizdeki bilgisayar mühendisliğinin ders içeriklerini inceleyerek hiyerarşik kümeleme, birliktelik kuralları ve karar ağaçları veri madenciliği yöntemlerini kullanarak KNIME programında analizi yapmıştır.

Kılıçer (2018) ve Güven'in (2016) yaptığı çalışmalardan yararlanılarak bu çalışmada ülkemizde bulunan 24 istatistik bölümünün ders içerikleri karşılaştırılarak hiyerarşik kümeleme ve birliktelik kuralları veri madenciliği yöntemleri ile analizler KNIME programında yapılmıştır.

Derslerin birbirleriyle olan ilişkilerini belirlemek için 2 aşamada analizler gerçekleştirilmiştir.

İlk analizde, istatistik bölümü bulunan üniversitelerin birbirine olan benzerliklerini bulmak için kümeleme analizi yöntemlerinden seçilen hiyerarşik kümeleme analizi yöntemi kullanılarak analiz yapılmıştır. Kümeleme analizi istatistik bölümleri derslerine göre birbirilerine benzer olan üniversitelerin hangileri olduğunu göstermiştir.

İkinci analizde ise birliktelik kuralları analizi için istatistik bölümlerinin eğitim programlarında bulunan derslerin hangilerinin birlikte alınması gerektiğini tespit etmek amacıyla Apriori algoritması kullanılarak analiz gerçekleştirilmiştir. Apriori algoritmasında düşük destek, yüksek güven ve kaldıraç (lift) değerleri seçilerek birliktelik kuralları oluşturulmuştur. Bu birliktelik kurallarının incelenmesi ile istatistik bölümlerinde bulunması gereken dersler belirlenmiştir.

Kılıçer'in (2018) çalışmasının sonucunda, Türkiye ve Avrupa'daki bilgisayar mühendisliği lisans programı ders içeriklerinin dağılımları incelenerek, ağırlık verilmesi gereken ders grupları belirlenmiş ve ders listelerine eklenilmesi gereken derslerin tespit edilmesi, Güven'in (2016) yaptığı çalışmada da Bilgisayar Mühendisliği bölümleri için ders planı ve ders içeriği oluşturması literatüre katkı sağlamıştır.

Kılıçer'in (2018) ve Güven'in (2016) yaptığı bu çalışmalar göz önünde bulundurulduğunda istatistik bölümleri için yapılan veri madenciliği yöntemleri üniversitelerin istatistik bölümlerinin birbirine olan benzerliklerini ve birlikte bulunması gereken derslerin belirlenmesini başarılı bir şekilde analiz etmiştir. Analizler sonucunda ortaya çıkan bilgilerin kullanılması ile hem istatistik bölümlerinin hem de üniversitelerin başarılarının artması amaçlanmıştır.

2. İSTATİSTİK BÖLÜMLERİ DERS İÇERİKLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ İÇİN KULLANILAN YÖNTEMLER VE PROGRAMLAR

İstatistik bölümleri ders içeriklerinin analizi için veri toplama işlemi, üniversitelerin istatistik bölümlerini belirlemek amacıyla lisans programlarının bilgisinin yer aldığı YÖK'ün Lisans Atlası ve Yüksek Öğretim Bilgi Yönetim Sistemi web sayfaları referans alınarak gerçekleştirilmiştir. Referans alınan YÖK'ün Lisans Atlas'ında yer alan tüm devlet ve vakıf üniversiteleri bu çalışmaya dâhil edilmiştir. Uygulamada üniversitelerin istatistik bölümlerinin ve ders içeriklerinin tespit edilmesi 2 aşamada yapılmıştır.

Birinci aşamada, Çukurova Üniversitesi İstatistik Bölümü'nün ders programında yer alan dersler Excel dosyasına girilmiştir. Referans alınan YÖK'ün Lisans Atlas'ında tespit edilen istatistik bölümü bulunan 25 üniversitenin isimleri belirlenerek Excel'e kaydedilmiştir. Bu üniversitelerin, istatistik bölümlerinin ders içeriklerine erişmek için her üniversitenin kendi web sayfasına giriş yapılarak ders programlarındaki ders içerikleri incelenmiştir. İstatistik bölümlerinin ders içerikleri incelenirken üniversitelerin web sayfalarındaki AKTS bilgi paketinden yararlanılmıştır.

İkinci aşamada ise istatistik bölümlerinin ders içerikleri incelenerek Excel'e istatistik bölümü bulunan 24 üniversitenin isimleri ile Çukurova Üniversitesi İstatistik Bölümü'nün ders programında yer alan dersleri yazılmıştır. Dersler Excel'e yazıldıktan sonra, Çukurova

Üniversitesi İstatistik Bölümü'nün ders içeriği ile diğer üniversitelerin istatistik bölümlerinin ders içerikleri karşılaştırılmıştır. Ders içerikleri aynı ise 1, aynı değil ise 0 yazılarak derslerin veri madenciliği analizleri yapılmıştır.

Yapılan çalışmada, ülkemizde bulunan istatistik bölümüne sahip 24 üniversitenin ders içerikleri incelenmiştir. Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer üniversitelerin lisans programlarındaki derslerin ders içerikleri karşılaştırılmıştır ve 2 farklı veri madenciliği yöntemi ile analizler yapılmıştır.

Üniversitelerin istatistik bölümlerinin birbirilerine olan benzerliklerini belirlemek için, hiyerarşik kümeleme yöntemi kullanılmıştır. Hiyerarşik kümeleme yönteminde küme sayısı önceden belli değildir ve bu yöntem tüm gözlemleri birer küme olarak dikkate alıp gözlemleri birleştirerek daha az sayıda küme elde etme yöntemidir (Altunkaynak, 2017).

İstatistik bölümlerinin birbirilerine olan benzerliklerini göstermek için küme sayısının 5 belirlenmesi ve KNIME programında hiyerarşik kümeleme algoritmalarının seçilmesi ile analiz gerçekleştirilmiştir.

İstatistik bölümlerinin ders programlarında yer alan derslerin hangilerinin birlikte alınması gerektiğini belirlemek için ise Apriori algoritması kullanılmıştır. Apriori algoritması, sık geçen öge kümeleri seçilen destek eşik değerine göre birliktelik kuralları belirlemektedir. Bu kurallar için güven değerleri hesaplanarak belirlenen güven eşik değerine göre birliktelik kuralları oluşturmaktadır. İstatistik bölümlerinin ders programlarında yer alan derslerin hangilerinin birlikte alınması gerektiğini belirlenmesi amacıyla ders birlikteliklerinin güven değerleri %70 ve üzeri alınarak birliktelik kuralları analizi KNIME programında yapılmıştır.

Bu çalışmada yapılan analizlerde, istatistik bölümleri bulunan üniversitelerin birbirleriyle olan ilişkileri, istatistik ders programlarında bulunması gereken derslerle ilgili sonuçlar elde edilmiştir. Analizler sonucunda ortaya çıkan bilgilerin kullanılması ile hem istatistik bölümlerinin hem de üniversitelerin başarılarının artması amaçlanmıştır.

2.1. KNIME

KNIME (Konstanz Information Miner) adlı veri madenciliği geliştirme yazılımı Konstanz Üniversitesi görsel veri madenciliği araştırma grubu tarafından Eclipse Rich Client Platformu üzerinde geliştirilmiştir. KNIME, modüler ve görsel veri akış sistemi geliştirme ortamı, eğitim ve araştırma amaçlı ortak çalışma ortamlarını içermektedir. Veri madenciliği

yöntemlerinden destek vektör makinaları, bayes, multidimesional scaling (MDS) gibi yöntemleri içerir. KNIME, Regresyon, Korelasyon ve Korelasyon Filtresi gibi istatistik tabanlı yöntemlerin veri akış tasarımında kullanılmasına imkân sağlar. Diğer yazılımlara göre en zengin görselleştirme araçlarına sahiptir.

Scatter Plot, Parallel Coordinates, Box Plot ve Histogram gibi birçok görselleştirme araçları ve JFree Chart bileşeni sayesinde çok daha ayrıntılı Java tabanlı görselleştirme düğümleri bulunmaktadır (Bilgin, 2009).

2.2. Veri Madenciliği Yöntemleri

2.2.1. Hiyerarşik Kümeleme

Küme sayısı önceden belli değildir. Bu yöntemde mümkün küme yapıları bir ağaç şeklinde gösterilir. Bu ağaca dendogram adı verilir. Dendogramın oluşturulmasında iki yaklaşım söz konusudur. Bunlardan birincisi tüm gözlemleri birer küme olarak dikkate alıp gözlemleri birleştirerek daha az sayıda küme elde etme yöntemidir. Bu yöntem, birleştirici hiyerarşik kümeleme (agglomerative hierarchical clustering) yöntemi olarak adlandırılır. İkincisi ise, tüm gözlemleri tek bir küme olarak ele alıp bu kümeye alt kümeleri de daha alt kümelere ayırma şeklinde uygulanan yöntemdir. Bu yöntem, ayrıştırıcı hiyerarşik kümeleme (divisive hierarchical clustering) yöntemi olarak adlandırılmaktadır (Altunkaynak, 2017).

KNIME programından seçilen hiyerarşik kümeleme düğümlerinin çalıştırılması ile kümeleme analizi yapılmıştır ve istatistik bölümleri derslerine göre birbirilerine benzer olan üniversitelerin kümeleri analiz sonucunda gösterilmiştir.

2.2.2. Apriori Algoritması

Birliktelik kuralı için en çok bilinen strateji Apriori'dir (Liao, vd., 2007). Algoritmanın ismi, yaygın nesnelerin önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından "önceki (prior)" anlamında aprioridir (Döşlü, 2008). Agrawal ve Srikant (1994) tarafından geliştirilen Apriori algoritması, veri madenciliği tarihinde birliktelik kurallarının çıkarılması konusunda elde edilmiş büyük bir başarıdır. Apriori algoritması veri tabanındaki verileri tekrarlayarak kaydeder ve her kayıttan sonra geniş veri kümelerini oluşturur. İşlemleri indirgemek için aday veri kümeleri için sadece destek seviyeleri hesaplanır (Liao, vd., 2007). Apriori algoritmasında k ögeli sık geçen öge küme adayları, (k-1) ögeli sık geçen öge kümelerinden faydalanılarak bulunur. Ancak bu algoritma veri tabanının pek çok kere taranmasını gerektirmektedir. Takipteki taramalarda bir önceki taramada bulunan sık geçen öge

kümeleri aday kümeleri adı verilen yeni potansiyel sık geçen öge kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek metriğini sağlayan kümeler o geçişte üretilen sık geçen öge kümeleri olur. Sık geçen öge kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öge kümesine rastlanıncaya kadar devam eder (Bilen, 2009).

Apriori algoritmasında sık geçen öge kümeleri seçilen destek eşik değerine göre birliktelik kuralları belirlemektedir. Bu kurallar için güven değerleri hesaplanarak belirlenen güven eşik değerine göre birliktelik kuralları oluşturulmaktadır.

Apriori algoritmasının kullanılması için KNIME programından seçilen düğümler ile birliktelik kuralları analizi gerçekleştirilmiştir ve istatistik bölümlerinde bulunması gereken dersler analiz sonucunda tespit edilmiştir.

3. BULGULAR

Bu çalışmada, Çukurova Üniversitesi İstatistik Bölümü lisans programlarında bulunan derslerin ders içerikleri karşılaştırılarak, üniversitelerin birbirilerine olan benzerliklerini belirlemek için hiyerarşik kümeleme analizi yöntemi, istatistik bölümlerinin ders programlarında yer alan derslerin hangilerinin birlikte alınması gerektiğini belirlemek için ise Apriori algoritması kullanılmıştır.

Uygulamada ilk olarak kümeleme analizi yapılmıştır ve istatistik bölüm derslerine göre birbirine benzerlik gösteren üniversitelerin hiyerarşik kümeleme analizi ile kümelmesi sağlanmıştır. Bu analiz sonucunda; küme sayısının 5 seçilmesi ile oluşan Tablo 1'de üniversitelere ait küme numaraları 0, 1, 2, 3 ve 4 şeklinde Cluster Number sütununda gösterilmiştir. Küme numaraları 0, 1, 2 ve 4 olan kümelerin eleman sayıları toplamı 4'tür. Bu durum istatistik bölüm derslerine göre birbirine benzemeyen üniversite sayısının toplamda 4 olduğunu göstermektedir. Küme numarası 3 olan kümenin eleman sayısı ise 20'dir. Bu ifade ise istatistik bölüm derslerine göre birbirine benzerlik gösteren üniversite sayısının 20 olduğunu açıklamaktadır.

Tablo 1. Kümeleme Analizi Sonucunda Üniversitelerin Kümelenmesi

Row ID	S Cluster number
Çukurova Üniversitesi	3
Afyon Kocatepe Üniversitesi	3
Amasya Üniversitesi	3
Anadolu Üniversitesi	3
Ankara Üniversitesi	3
Bartın Üniversitesi	3
Dokuz Eylül Üniversitesi	3
Ege Üniversitesi	3
Eskişehir Osmangazi Üniversitesi	0
Fırat Üniversitesi	3
Gazi Üniversitesi	3
Giresun Üniversitesi	3
Hacettepe Üniversitesi	1
İstanbul Aydın Üniversitesi	2
Marmara Üniversitesi	3
Muğla Sıtkı Koçman Üniversitesi	4
Necmettin Erbakan Üniversitesi	3
Ondokuz Mayıs Üniversitesi	3
Orta Doğu Teknik Üniversitesi	3
Selçuk Üniversitesi	3
Sinop Üniversitesi	3
Sivas Cumhuriyet Üniversitesi	3
Ufuk Üniversitesi	3
Yıldız Teknik Üniversitesi	3

Çalışmada ikinci olarak ise birliktelik kuralları analizi yapılmıştır. Bu analizde, istatistik bölümü bulunan üniversitelerde hangi derslerin birlikte alınması gerektiği Apriori algoritması kullanılarak ortaya çıkarılmıştır. Bu algoritma düşük destek değerlerine sahip olan ders birlikteliklerinin güven değerlerini %70 ve üzeri olarak yüksek kaldıraç (lift) değerleri ile öge kümeleri için derslerin birliktelik kurallarını belirlemiştir. Analiz sonucunda 21 tane birliktelik kuralı oluşturulmuştur ve Tablo 2’de gösterilmiştir. Tablo 2’de gösterilen bu kurallar 2, 3 ve 4 ders bulunan birlikteliklerin öge kümelerini ifade etmektedir.

Tablo 2. Birliktelik Kuralları Tablosu

	Antecedent	Consequent	Support%	Confidence%	Lift
1.Kural	İstatistiksel Sonuç Çıkarım	Olasılığa Giriş	12,5	100	1,6
2.Kural	Matematik 2	Olasılığa Giriş	12,5	100	1,6
3.Kural	Optimizasyon Teknikleri I	İstatistik İçin Matematik I	12,5	100	8
4.Kural	Kombinatorik	Matematik 1	12,5	100	4,8
5.Kural	Kombinatorik	Olasılığa Giriş	12,5	100	1,6
6.Kural	Aktüerya Modelleri	Matematik 1	12,5	100	4,8
7.Kural	Aktüerya Modelleri	Olasılığa Giriş	12,5	100	1,6
8.Kural	Lineer Cebir 1	Lineer Cebir 2	12,5	75	3
9.Kural	Lineer Cebir 1	Olasılığa Giriş	16,667	100	1,6
10.Kural	Matematik 1	Olasılığa Giriş	16,667	80	1,28
11.Kural	Çok Değişkenli İstatistiksel Analiz 2	Olasılığa Giriş	29,167	100	1,6
12.Kural	Lineer Cebir 2	Olasılığa Giriş	20,833	83,3	1,333
13.Kural	Kombinatorik Matematik 1 Matematik 1	Olasılığa Giriş	12,5	100	1,6
14.Kural	Kombinatorik Olasılığa Giriş	Matematik 1	12,5	100	4,8
15.Kural	Matematik 1 Olasılığa Giriş	Kombinatorik	12,5	75	6
16.Kural	Aktüerya Modelleri Matematik 1	Olasılığa Giriş	12,5	100	1,6
17.Kural	Aktüerya Modelleri Olasılığa Giriş	Matematik 1	12,5	100	4,8
18.Kural	Matematik 1 Olasılığa Giriş	Aktüerya Modelleri	12,5	75	6
19.Kural	Lineer Cebir 1 Lineer Cebir 2	Olasılığa Giriş	12,5	100	1,6
20.Kural	Lineer Cebir 1 Olasılığa Giriş	Lineer Cebir 2	12,5	75	3
21.Kural	Çok Değişkenli İstatistiksel Analiz 2 Lineer Cebir 2	Olasılığa Giriş	12,5	100	1,6

4. SONUÇ VE TARTIŞMA

Veri madenciliği yöntemlerinin eğitim alanında kullanılmasıyla birlikte ortaya çıkan, literatürdeki adıyla eğitsel veri madenciliği (educational data mining), öğrencilere ve okullara ait çoklu verinin işlenmesine olanak sağlamıştır. Özellikle eğitim alanındaki kullanımı, eğitimcilere ve eğitim planlamacılarına farklı açılardan ışık tutarak eğitim stratejilerinin belirlenmesinde ihtiyaca yönelik çözümler sunabilmektedir (Bilen, vd., 2014).

Bu çalışmada; ülkemizde bulunan istatistik bölümüne sahip 24 üniversite incelenmiştir. İstatistik bölümlerinin lisans programlarında yer alan derslerin ders içerikleri kullanılarak farklı üniversitelerin istatistik bölümlerindeki ders içeriklerinin birbirleriyle olan benzerlikleri, bu benzerlikler ile üniversitelerin birbirleriyle ne kadar ilişkili olduğu kümeleme analizi ile, istatistik bölümlerinin ders programlarında bulunan derslerden hangi derslerin birlikte alınmasını belirlemek için ise birliktelik analizi kullanılmıştır.

Çalışmada yapılan hiyerarşik kümeleme analizinde, istatistik bölüm derslerine göre birbirine benzerlik gösteren üniversiteler kümelenmiştir. Küme sayısının 5 seçilmesi ile oluşan Tablo 1'deki kümeleme sonuçları incelendiğinde; mor renkte gösterilen 3 numaralı kümede bulunan Çukurova Üniversitesi İstatistik Bölümü ile istatistik bölümüne sahip diğer 19 üniversitenin kümeleme analizi sonuçlarının birbirlerine benzerlik gösterdiği ifade edilmiştir.

Birliktelik analizi ise istatistik bölümlerinin eğitim programlarında yer alan derslerden hangi derslerin birlikte alınması gerektiğini belirlemiştir. Bu analiz istatistik bölümü bulunan üniversitelerde hangi derslerin birlikte alınması gerektiğini Apriori algoritması kullanarak düşük destek değerlerine sahip olan ders birlikteliklerinin güven değerlerini %70 ve üzeri olarak yüksek kaldıraç (lift) değerleri ile öge kümeleri için derslerin birliktelik kurallarını belirlemiştir ve Tablo 2'de gösterilen 21 birliktelik kuralının oluşmasını sağlamıştır. Bu kurallar incelendiğinde; istatistik bölümünün lisans programlarında bulunması gereken önemli derslerin Olasılığa Giriş, Matematik 1, Kombinatorik, Aktüerya Modelleri, Lineer Cebir 1 ve Lineer Cebir 2 olması gerektiği tespit edilmiştir. Ayrıca, 21 birliktelik kuralının 17 tanesinde Olasılığa Giriş dersi yer almaktadır. Bu durum istatistik bölümü olan üniversitelerin ders programlarında bulunması gereken en önemli dersin Olasılığa Giriş dersi olacağını ifade etmiştir.

Kılıçer'in (2018) çalışmasında Türkiye ve Avrupa'daki bilgisayar mühendisliği lisans programı ders içeriklerinin dağılımları incelenerek, ağırlık verilmesi gereken ders grupları belirlenmiş ve ders listelerine eklenilmesi gereken dersler tespit edilmiştir.

Güven'in (2016) yaptığı çalışmada ise Bilgisayar Mühendisliği bölümlerinin müfredatlarına göre üniversitelerin benzerliği ortaya çıkarılmış ve bu analiz üniversitelerde

yeni kurulacak olan Bilgisayar Mühendisliği bölümleri için ders planı ve ders içeriği oluşturmuştur.

İstatistik bölümleri için yapılan veri madenciliği yöntemleri üniversitelerin istatistik bölümlerinin birbirine olan benzerliklerini ve birlikte bulunması gereken derslerin belirlenmesini başarılı bir şekilde analiz etmiştir. Kılıçer'in (2018) ve Güven'in (2016) yaptığı çalışmalar özellikle üniversitelerin bölümlerindeki ders müfredatına yönelik yapılacak yorumlara kolaylık sağlayacaktır. Aynı zamanda istatistik ders içerikleri ile yapılan bu çalışmada ortaya çıkan bilgilerin kullanılması ile hem istatistik bölümlerinin hem de üniversitelerin başarıları arttırılabilir.

Eğitim alanında yapılan bu çalışmada üniversitelerin istatistik bölümlerinde bulunan derslerin analizi yapılarak üniversitelerin birbirine olan benzerlikleri ve istatistik bölümlerinde birlikte bulunması gereken dersler belirlenmiştir. Yapılan bu analizler istatistik bölümleri için yapılabilecek çalışmalara yol göstermesi ve bu çalışmaların yorumlanabilmesi açısından önemlidir. İstatistik bölümleri ile bu bölümlerin anlaşmalı olduğu Erasmus programına bağlı üniversitelerin ders içeriklerinin karşılaştırılarak istatistik lisans programlarındaki eksikliklerin tespit edilmesi ve tamamlanması, istatistik bölümlerinin öğrenci işleri otomasyonu üzerinden elde edilecek lisans öğrencilerinin aldıkları dersleri gösteren veriler ile yapılacak olan analizler sonucunda öğrencilerin başarılarına etki eden faktörlerin belirlenmesi, yeni açılacak üniversitelerdeki bölümlerin ders planlanması hazırlanırken ders programlarında hangi derslerin yer alması gerektiği ve ders içeriklerinin nasıl olacağı hakkındaki bilgilerin bulunması gibi eğitim alanındaki araştırma konularının analizlerinin veri madenciliği yöntemleri ile yapılması önerilmektedir.

TEŞEKKÜR

Bu Çalışma Çukurova Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenmiştir. (Proje No: FYL201810542)

KAYNAKÇA

Agrawal, R. ve Srikant, R. (1994), *Fast Algorithms for Mining Association Rules in Large Databases. Proceedings of the 20th International Conference on Very Large Data Bases*, Santiago de Chile, 487-499.

Altunkaynak, B. (2017), *Veri Madenciliği ve R Uygulamaları*. Seçkin Yayıncılık, Ankara.

- Bilen, H. (2009). *Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirilmesine İlişkin Veri Madenciliği Uygulaması*. Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, 10.
- Bilen, Ö., Hotaman, D., Aşkın E., Ö., Büyüklü H., A., (2014). LYS Başarılarına Göre Okul Performanslarının Eğitsel Veri Madenciliği Teknikleriyle İncelenmesi: 2011 İstanbul Örneği. *Eğitim ve Bilim*, 39(172), 78-94.
- Bilgin, T. T. (2009), Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları, *11.Akademik Bilişim Konferansı*, Şanlıurfa, 807-814.
- Darko Dukic, Kat I Penny, (2011), *Analysis of Croatian and Scoottish Students Attitudes towards e-learning*, ITI 2011 33rd International Conference on Information Technology Interfaces, Cavtat, Croatia, June 27-30.
- Delevari N, Beikzadeh M.R. (2008), *Data Mining Application in Higher Learning Institutions*, Informatics in Education, 7(1), 31-54.
- Döşlü, A. (2008), *Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, 100.
- Güven, B. Z. (2016), *Türk Üniversitelerindeki Bilgisayar Mühendisliği Bölümleri Müfredatları Kullanılarak Veri Madenciliği Uygulaması Gerçekleştirilmesi*, Yüksek Lisans Tezi, Maltepe Üniversitesi, İstanbul, 68.
- Han, J. and Kamber, M., (2006), *Data Mining Concepts and Techniques*. Morgan Kaufmann, 770.
- Kabra, R.R. and Bichkar, R. S. (2011), *Performance Prediction of Engineering Students using Desicion Trees*, International Journal of Computer Applications (0975-8887), 36(11).
- Kılıçer, S. (2018), *Veri Madenciliği ile Türkiye'deki ve Avrupa Birliği Ülkelerindeki Bilgisayar Mühendisliği Bölümleri Ders İçeriklerinin Karşılaştırılması*. Yüksek Lisans Tezi, İstanbul Üniversitesi, İstanbul, 103.
- Liao, S. H. and Wen, C. H. (2007), *Artificial Neural Networks Classification and Clustering of Methodologies and Applications–Literature Analysis from 1995 to 2005*, Expert Systems with Applications, 32(1-11).