



ISSN:1306-3111

e-Journal of New World Sciences Academy
2011, Volume: 6, Number: 1, Article Number: 1A0156

ENGINEERING SCIENCES

Received: October 2010

Accepted: January 2011

Series : 1A

ISSN : 1308-7231

© 2010 www.newwsa.com

Yavuz Ünal

Ufuk Ekim

Murat Köklü

Selcuk University

yunal@selcuk.edu.tr

Konya-Turkey

ÜNİVERSİTE ÖĞRENCİLERİN ORTAK ZORUNLU DERSLERDEKİ BAŞARILARININ K-MEANS ALGORİTMASI İLE İNCELENMESİ

ÖZET

Bu çalışmada veri madenciliği tekniklerinden K-Means kullanılarak 2009-2010 eğitim öğretim döneminde Selçuk Üniversitesinin 3 fakülte ve bir yüksekokulda okuyan öğrencilerin ortak zorunlu derslerdeki başarılarının analizi yapılmıştır. Çalışmada kullanılan öğrenci notları, S.Ü. Öğrenci İşleri biriminden ORACLE veri tabanı ortamında alınarak gerekli dönüşümler yapılmıştır. Elde edilen bu verilere SQL Server 2005 programı içindeki "Analysis Services" paketinde bulunan veri madenciliği modülleri, Visual Studio 2005 programı üzerinde çalıştırılarak analizleri yapılmıştır. Yapılan analizler sonucunda çeşitli fakülte ve yüksekokulda okuyan öğrencilerin ortak zorunlu derslerdeki başarıları karşılaştırılmış, elde edilmiş anlamlı sonuçlardan yola çıkılarak görüş ve öneriler belirtilmiştir.

Anahtar Kelimeler: Veri Madenciliği, Kümeleme algoritmaları, K-Means Algoritması, Ortak Zorunlu Dersler, Öğrenci Başarısı

AN INVESTIGATION OF UNIVERSITY STUDENTS SUCCESS IN COMMON COMPULSORY LESSONS VIA K-MEANS ALGORITHMS

ABSTRACT

In this study, K-Means, is one of the data-mining techniques, was used. The success analyses of college students who attend in 3 faculties and a collage of Selçuk University in 2009-2010 academic year in common compulsory lessons were performed. Data used in this study was obtained from student affairs of Selçuk University. In Visual Studio 2005 Environment, data-mining modules included in Analysis Services packet in SQL Server 2005 were run over the obtained data. Finally opinions and suggestions were specified.

Keywords: Data Minig, Clustering Algorithm, K-Means Algorithm, Common Compulsory Lessons, Student Success

1. GİRİŞ (INTRODUCTION)

2547 sayılı Yükseköğretim Kanununun 5. maddesine Yükseköğretim kurumlarında, Atatürk İlkeleri ve İnkılap Tarihi, Türk dili ve yabancı dil ortak zorunlu derslerdir [1].

Bu çalışmada Veri Madenciliği tekniklerinden K-Means kullanılarak Selçuk Üniversitesinin değişik fakülte ve yüksekokul öğrencilerinin ortak zorunlu derslerdeki başarılarının analizi yapılmıştır. Bu amaçla 2009/2010 eğitim öğretim yılı Mühendislik Mimarlık Fakültesi, İktisadi ve İdari Bilimler Fakültesi, Fen Fakültesi ve Sosyal Bilimler Meslek Yüksek okulunda okuyan 400 adet öğrencinin Atatürk İlkeleri ve İnkılâp Tarihi-I, Türk Dili-I, Yabancı Dil-I derslerine ait başarı notları incelenmiştir. Öğrenci notları, S.Ü. Öğrenci İşleri biriminden ORACLE veri tabanı ortamında alınarak gerekli dönüşümler yapılmıştır. Elde edilen bu verilere SQL Server 2005 programı içindeki "Analysis Services" paketinde bulunan veri madenciliği modülleri, Visual Studio 2005 programı üzerinde çalıştırılarak analizleri yapılmıştır.

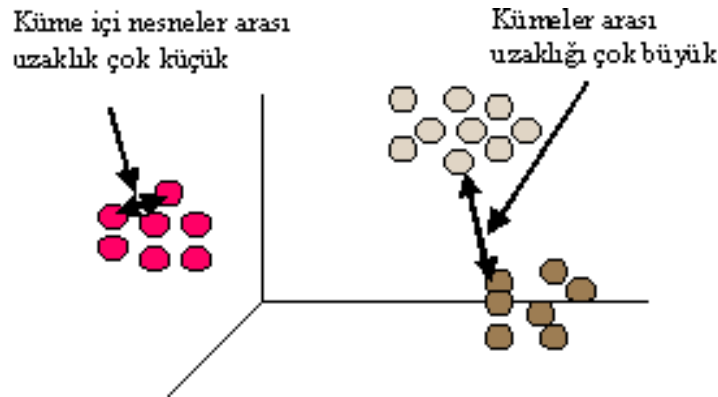
Çalışmamızda, öncelikle veri madenciliğinin ne olduğundan bahsedilmiş, kümeleme ve K-Means yöntemi kısaca özetlenmiş daha sonra ise çalışmanın yapılış aşamaları ayrıntılı olarak anlatılmıştır. Sonuç ve öneriler kısmında ise elde edilmiş anlamlı sonuçlardan yola çıkılarak görüş ve öneriler belirtilmiştir.

1.1. Veri Madenciliği (Data Mining)

Veri madenciliği; büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır [2]. Veri madenciliği uygulamaları başta pazarlama, bankacılık, tıp, mühendislik, endüstri, borsa analizleri ve ulusal güvenlik alanlarında kullanılmaktadır. Örneğin, müşteri ilişkileri yönetiminde, kredi kartı dolandırıcılıklarının tespitinde, üretim süreçlerinin iyileştirilmesinde, hatların yoğunluk tahmininde, kalite kontrol analizlerinde, hisse senedi fiyat tahmininde veri madenciliğinden etkin şekilde faydalanılmaktadır [3].

1.2. Kümeleme (Clustering)

Kümeleme birbirine benzeyen veri parçalarını ayırma işlemidir ve kümeleme yöntemlerinin çoğu veri arasındaki uzaklıkları kullanır. Örneğin Öklid, Manhattan ve Minkowski uzaklık bağıntıları kümeleme işleminde alt işlem olarak kullanılmaktadır. Kümeleme yöntemleri arasında akla ilk gelen en yakın komşu algoritması ve en uzak komşu algoritması gelmektedir; bunlar hiyerarşik kümeleme algoritmaları olarak da bilinir. Hiyerarşik olmayan kümeleme yöntemleri arasında k-ortalamalar (K-means) yöntemi sayılabilir. Çalışmamızda Microsoft SQL Server 2005 ürünü yardımıyla k-ortalamalar (K-means) algoritması kullanılmıştır [4]. Şekil 1'de küme içindeki nesnelere arasındaki benzerlik görülmektedir.



Şekil 1. Küme yapısı
(Figure 1. Cluster structure)

1.2.1. K-Ortalamlar algoritması (K-means Algorithm)

Hiyerarşik olmayan kümeleme yöntemleri arasında k-ortalamlar (k-means) yöntemi önem taşır ve yaygın biçimde kullanılır. Bu yöntemde, daha başlangıçta belli sayıdaki küme için toplam ortalama hatayı minimize etmek amaçlanır. N boyutlu uzayda N örnekli kümelerin verildiğini varsayalım. Bu uzay $\{k=1,2,\dots,k\}$ olmak üzere C_k kümesinin ortalama vektörü M_k şu şekilde hesaplanır.

$$M_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{ik} \quad (1)$$

Burada X_k değeri C_k kümesine ait olan i. örnektir. C_k kümesi için kare-hata, her bir C_k örneği ile onun merkezi (centroid) arasındaki Öklid uzaklıkları toplamıdır. Bu hataya "küme içi değişme" adı da verilir. Küme içi değişmeler şu şekilde hesaplanır:

$$e_i^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2 \quad (2)$$

K kümesini içeren bütün kümeler uzayı için kare-hata, küme içindeki değişmelerin toplamıdır. O halde söz konusu kare-hata değeri şu şekilde hesaplanır:

$$E_k^2 = \sum_{k=1}^K e_k^2 \quad (3)$$

Kare-hata kümeleme yönteminin amacı, verilen K değeri için E_k^2 değerini minimize eden K kümelerini bulmaktır. O halde k-ortalamlar algoritmasında E_k^2 değerinin bir önceki iterasyona göre azalması beklenir [4].

2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICIANCE)

Bu çalışmada veri madenciliği tekniklerinden K-Means kullanılarak 2009-2010 eğitim öğretim döneminde Selçuk Üniversitesinin 3 fakülte ve bir yüksekokulda okuyan öğrencilerin ortak zorunlu derslerdeki başarılarının analizi yapılmıştır. Çalışmada kullanılan öğrenci notları, S.Ü. Öğrenci İşleri biriminden ORACLE veri tabanı ortamında alınarak gerekli dönüşümler yapılmıştır. Elde edilen bu verilere SQL Server 2005 programı içindeki "Analysis Services" paketinde bulunan veri madenciliği modülleri, Visual Studio 2005 programı üzerinde çalıştırılarak analizleri yapılmıştır. Yapılan analizler sonucunda çeşitli fakülte ve yüksekokulda okuyan öğrencilerin ortak zorunlu derslerdeki başarıları karşılaştırılmış, elde edilmiş anlamlı sonuçlardan yola çıkılarak görüş ve öneriler belirtilmiştir.

3. BENZER ÇALIŞMALAR (SIMILAR STUDIES)

Yapılan benzer bir çalışmada apriori algoritması ile öğrenci başarı analizi yapılmıştır [5]. Diğer bir çalışmada veri madenciliği yöntemleri ÖSS verilerine uygulanarak Türkiye genelinde çeşitli istatistik sonuçlar ortaya konmuştur [6]. Diğer bir çalışmada ise ÖSS verilerine karar ağaçları ve birliktelik kuralları yöntemleri uygulanarak öğrencilerin tercih profilleri ortaya konulmuştur [7]. Başka bir çalışmada ÖSS sonuçları yıllara, bölgelere ve okul türlerine göre incelenmiştir [8]. Yine bir başka çalışmada öğrenci seçme sınavında (ÖSS) öğrenci başarısını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti yapılmıştır [9].

4. UYGULAMA (APPLICATION)

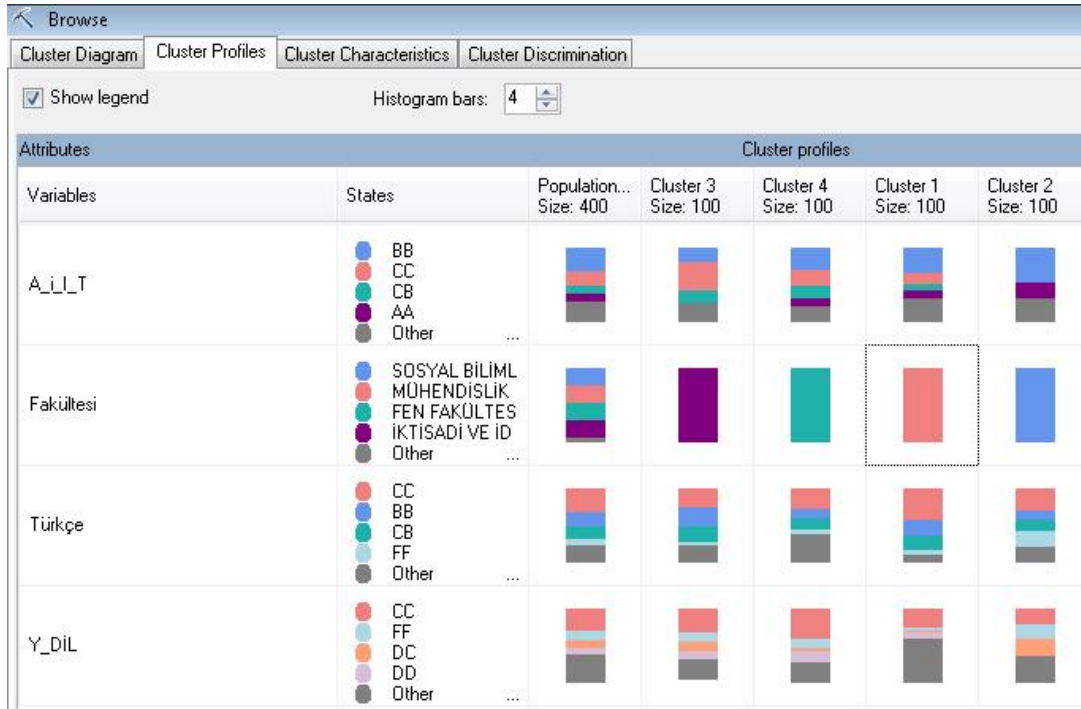
Çalışmamızda fakültelere göre öğrenci dağılımı Tablo 1'de sunulmuştur.

Tablo 1. Veri kümesinde yer alan niteliklere ait ayrıntılı dağılım
(Table 1. Detailed spectrum belonging to qualifications
exist in data set)

Fakülte /Yüksekokul Adı	Dağılım
Fen Fakültesi	100
Mühendislik Mimarlık	100
İktisadi İdari Bilimler	100
Sosyal Bilimler	100
TOPLAM	400

Buna göre Fen Fakültesinden 100, Mühendislik Mimarlık Fakültesinden 100, İktisadi ve İdari Bilimler Fakültesinden 100 ve Sosyal Bilimler Meslek Yüksek okulundan 100 öğrenci olmak üzere toplam 400 öğrenci üzerinde çalışma yapılmıştır.

Yapılan kümeleme çalışmasında küme sayısının fakülte/yüksekokul sayısı ile aynı olması amaçlanmıştır. Bu sebeple uygulama içerisinde "CLUSTER_COUNT" parametresi 4 olarak girilmiştir. Algoritma olarak K-Means'in kullanılabilmesi için uygulama içerisinde "CLUSTER_METHOD" parametresi 3 olarak girilmiştir. Analiz sonucunda dört adet öğrenci profili çıkarılmıştır (Şekil 2).



Şekil 2. Analiz sonucu elde edilen kümeler
(Figure 2. Sets obtained from analysis)

Buna göre her bir küme 100 kişiden oluşmaktadır. Küme 1'in tamamı Mühendislik Mimarlık Fakültesi, küme 2'nin tamamı Sosyal Bilimler Meslek Yüksek okulundan, küme 3'ün tamamı İktisadi ve İdari Bilimler Fakültesinden ve küme 4'ün tamamı Fen Fakültesi öğrencilerinden oluşmaktadır.

Selçuk üniversitesi eğitim-öğretim ve sınav yönetmeliğine göre; Bir dersten AA, BA, BB, CB, CC notlarından birini alan öğrenci o dersi başarmış sayılır. Çalışmamızda kümeler arası başarı mukayesesi için her bir kümenin ilgili dersten aldıkları AA, BA, BB, CB ve CC lerin yüzdeleri toplanıp

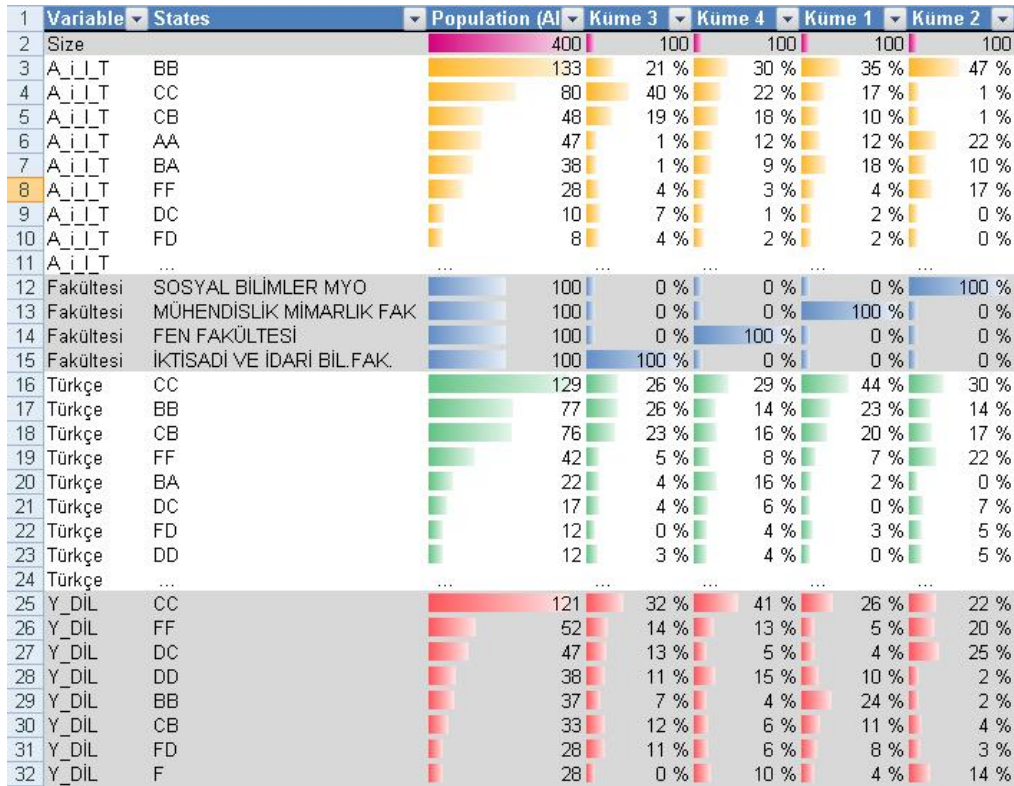
karşılaştırılmıştır. Başarı oranlarının ayrıntılı dağılımı Tablo 2 'de görülmektedir. Elde edilen sonuçlara göre; Atatürk İlkeleri ve İnkılap Tarihi -I dersinden en başarılı kümenin küme 1, ikincinin küme 4, üçüncünün küme 3 ve dördüncünün ise küme 2 olduğu görülmektedir.

Türk Dili-I dersinden en başarılı kümenin küme 1, ikincinin küme 3, üçüncünün küme 4 ve dördüncünün ise küme 2 olduğu görülmektedir.

Yabancı Dil-I dersinden en başarılı kümenin küme 1, ikincilerin küme 3 ve küme 4, üçüncünün ise küme 2 olduğu görülmektedir. K-Means algoritmasının sonucunda çıkarılan kümelerin derslerdeki başarılarının ayrıntılı dağılımı Şekil 3'te görülmektedir.

Tablo 2. Kümelerin derslere göre başarı dağılımı
(Table 2. Success spectrum of sets according to the courses)

Dersler	Kümeler	Başarı Oranı % Olarak
Atatürk İlkeleri ve İnkılap Tarihi -I	Küme 1 (Müh.Mim. Fak. Öğrencileri)	%92
	Küme 2 (Soysal Bilimler Meslek Yüksek Okulu)	%81
	Küme 3 (İktisadi ve İdari Bilimler Fakültesi)	%82
	Küme 4 (Fen Fakültesi)	%91
Türk Dili - I	Küme 1 (Müh.Mim. Fak. Öğrencileri)	%89
	Küme 2 (Soysal Bilimler Meslek Yüksek Okulu)	%61
	Küme 3 (İktisadi ve İdari Bilimler Fakültesi)	%82
	Küme 4 (Fen Fakültesi)	%75
Yabancı Dil-I	Küme 1 (Müh.Mim. Fak. Öğrencileri)	%74
	Küme 2 (Soysal Bilimler Meslek Yüksek Okulu)	%28
	Küme 3 (İktisadi ve İdari Bilimler Fakültesi)	%51
	Küme 4 (Fen Fakültesi)	%51



Şekil 3. K-Means algoritmasının sonucunda çıkarılan kümelerin derslerdeki başarılarının ayrıntılı dağılımı

(Figure 3. Detailed spectrum of successes of sets obtained from K-means algorithm in courses)

5. SONUÇ VE ÖNERİLER (CONCLISION AND RECOMMENDATIONS)

Bu çalışmada üniversite öğrencilerin ortak zorunlu derslerdeki başarıları veri madenciliği algoritmalarından k-means ile incelenmiştir. İnceleme sonucunda sayısal bölümlerden oluşan Mühendislik Mimarlık Fakültesi öğrencilerinin Atatürk İlkeleri ve İnkılâp Tarihi, Türk Dili ve Yabancı Dil gibi sözel derslerde diğer fakülte ve yüksekokul öğrencilerine göre daha başarılı oldukları görülmüştür. Üniversiteye giriş puan türüne göre özel olan Sosyal Bilimler Meslek Yüksek Okulu öğrencilerinin, sayısal ağırlıklı olan Mühendislik Mimarlık Fakültesi ve Fen Fakültesi öğrencilerine göre başarı oranlarının düşük olduğu görülmektedir. Sosyal Bilimler Meslek Yüksek Okulu öğrencilerinin incelememizdeki diğer fakültelere göre üniversiteye giriş puanlarının düşük olduğu veya üniversiteye sınavsız girdikleri bilinmektedir. Öğrencilerin üniversiteye giriş puanı ve durumlarının, ortak zorunlu derslerdeki başarıları ile doğru orantılı olduğunu düşünmekteyiz.

KAYNAKLAR (REFERENCES)

1. 2547 sayılı yüksek öğretim kanunu madde 5.
2. Alpaydın, E., (2000), "Zeki Veri Madenciliği: Ham Veriden Altın Veriye Ulaşma Yöntemleri" Bilişim Eğitim Semineri, ANKARA
3. Bozkır, A.S., Gök, B. ve Sezer, E., (2008) Üniversite Öğrencilerinin İnterneti Eğitimsel Amaçlar İçin Kullanmalarını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti, Bilimde Modern Yöntemler Sempozyumu, Eskişehir, pp 833-842.
4. Özkan, Y., (2008), Veri Madenciliği Yöntemleri, Papatya Yayıncılık, İstanbul.
5. Karabatak, M. ve İnce M.C., (2004), Apriori Algoritması ile Öğrenci Başarısı Analizi, Eleco'Elektrik-Elektronik ve Bilgisayar Mühendisleri Sempozyumu, Bursa.
6. Kılanç, B., (2008). Türkiye'nin ÖSS Tercihleri, 1.Psikolojik Danışman ve Veli Bilgilendirme-Yönlendirme Sempozyumu, İzmir.
7. Dolgun, M.Ö., Özdemir, T.G. ve Deliloğlu, S., (2007) Öğrenci Seçme Sınavında (ÖSS) Öğrencilerin Tercih Profillerinin Veri Madenciliği Yöntemleriyle Tespiti, Bilişim" 07 Kongresi, .
8. Berberoğlu, G. ve Kalender, İ., (2005), Öğrenci Başarısının Yıllara, Okul Türlerine, Bölgelere Göre İncelenmesi: ÖSS ve PISA Analizi, Eğitim Bilimleri ve Uygulama., Vol. 4(7), 21-35.
9. Bozkır, A.S., Gök, B. ve Sezer, E., (2009), Öğrenci Seçme Sınavında (ÖSS) Öğrenci Başarısını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), Karabük.