

İleri Eğitim Kursu Kalite Ölçümü için Veri Madenciliği ve İstatistiksel Yöntemler

Kerem Dikbaşı ||
Sena Polat

Hacettepe Üniversitesi
İstatistik Bölümü

Sunulan:
Doç.Dr. Nilgün Özgül ÇELEBİ

İST480 Araştırma Yöntemleri
6 Haziran, 2023

- Giriş
- Veri ve Metodoloji
- Veri Analizi ve Sonuçlar
 - Box-Plot Grafikleri
 - Faktör Analizi
 - Karar Ağaçları
 - Referanslar

Özet: Enerjetik alanındaki ileri eğitim kursları, insan güvenilirliği büyümesinin çok önemli bir parçasıdır. UNESCO departmanı Enerji Teknolojileri'ndeki Teknik Eğitim Sistemleri Başkanı S.E. Magid'in deyiimiyle: "Güç santrallerindeki ekipman arızalarına bağlı zorunlu duruşma sayısı 30. Bu ihlallerde operasyonel personel hatasının payı önemli bir boyuttadır (15). Rusya Devlet Şirketi 'Birleşik Güç Sistemleri'nde personelden kaynaklanan ihlallerin toplam ihlallere oranı 2. Aynı zamanda, güç santrallerinde bu miktar yüzde 18'e ulaşır. Sibiry'a'nın enerji tedarik sistemlerinde, personelden kaynaklanan ihlallerin göreceli miktarı yüzde 50'ye kadar ulaşır." [1].

[1]: Kaynak: S.E. Magid, Enerji Teknolojileri Teknik Eğitim Sistemleri Başkanı, UNESCO Departmanı.

Veri ve Metodoloji

Bu durumda, İK departmanı öğrencilerin yanıtlarını almalı ve bunları uygun şekilde analiz etmelidir. Bu amaçla, en büyük enerji sağlayıcılarından birinin İK departmanı, özel olarak düzenlenmiş İleri Eğitim Kursları bölümünde bir kursa kayıtlı öğrenciler (elektrik mühendisleri) için bir anket hazırlamıştır. Anket formunda, öğrenciler kursların bazı parametrelerini işaretledi ve analiz amaçları için çalışmacılar bunları yeniden kodladı .

Her bir parametre, 1 (kötü) ile 10 (mükemmel) arasında bir aralıkta işaretlendi. Analiz için, kursun sonunda tamamlanan 316 anket baz alınmıştır. Burada kayıp gözlem içeren kısımlar nedeniyle bu anket sayısını çalışmacılar 301 ankete düşürmüştür.

Veri ve Metodoloji

Çalışma **yalnızca** açık kaynak kodlu programlar kullanılarak gerçekleştirildi:

- R istatistiksel programlama dili
- KNIME veri madenciliği platformu



Veri Analizi ve Sonuçlar

Kurs değerlendirilmelerini etkileyen ana faktörleri belirlemek için verilerin dağılımını incelememiz gerekir. Normallik testi için gözlem sayımız ($n=301 > 50$) olduğu için Shapiro-Wilk testi kullanılmıştır. Fakat normallik varsayımı $p = 0.05$ düzeyinde kabul edilememiştir.

Histogram grafiği incelendiğinde verilerin sola çarpık olduğu gözlemlenebilir. (Şekil 1.)

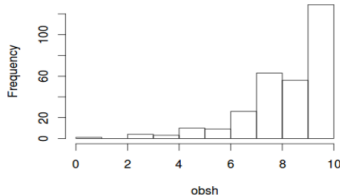
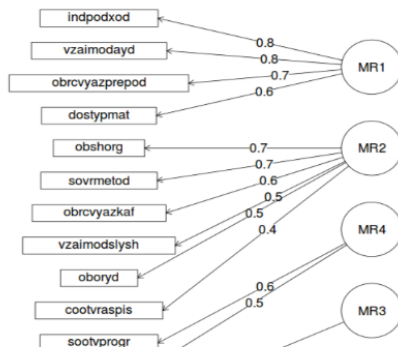


Figure 1. The histogram of the field "General impressions about provided training"

Veri Analizi ve Sonuçlar

Veri seti sıralı olduğundan, sadece parametrik olmayan istatistiksel yöntemler kullanılmıştır. Değişkenler arasındaki ilişkileri inceleyebilmek adına Spearman korelasyon katsayısı kullanılmıştır. 0.6 düzeyin filtrelendirme uygulandıktan sonra tüm katsayılar yüzde 95 güven düzeyinde anlamlı bulunmuştur.



Veri Analizi ve Sonuçlar

Sonuçlardan, "Öğrenciye bireysel yaklaşım" faktörü ile "Öğretmenin kitleyle etkileşiminin etkinliği" faktörü arasında yüksek düzeyde bir korelasyon olduğu ve "Öğretmenden geri bildirim alma" faktörü ile "Bölüm çalışanlarından geri bildirim alma" faktörü arasında da uyumlu bir korelasyon olduğu söylenebilir. Bu nedenle, öğrencilerin bireysel yaklaşım ile öğretmenin etkinliği arasında neredeyse ayrım yapmadıklarını ve öğretmenden ve bölüm çalışanlarından gelen geri bildirimleri aynı şekilde değerlendirdiklerini söylemek mümkündür.

Veri Analizi ve Sonuçlar

Bu amaçlarla küme analizi prosedürü, belirlenen 4 küme sayısı ile bulanık C-means yöntemiyle veri setine uygulanmaktadır. Sonuç olarak, yaklaşık yüzde 40'ını oluşturan biri dahil olmak üzere dört grup elde edilmiş olup ve diğer gruplar başlangıç veri setinin gücünün yüzde 17 ile yüzde 20'sini kapsamaktadır (Tablo 2). Elde edilen kümeler arasında ayrım analizi için, tüm göstergelerde medyanlar ve interquartile aralığı (IQR) kullanılmıştır.

Küme	Elementler	Yüzdelere
1	58	19%
2	117	39%
3	75	25%
4	51	17%
Total	301	100%

Tablo 2.

Sonuçlar

Genel olarak, 2. ve 3. kümelerdeki öğrencilerin tüm parametreler için yüksek puanlar verdiği belirtilmelidir. Buradan bazı sonuçlar çıkarabiliriz:

- Küme öğrencileri çoğu durumda 9-10 puan üzerinden "mükemmel" olarak 13 parametreyi işaretlemektedir. Muhtemelen bu öğrenci grubu anket doldurma sürecine resmi olarak yaklaşmış ve ileri analizlerden çıkarılması gerekmektedir.
- "Yapılan eğitimin pratik kullanışlılığı" parametresindeki farklar da istatistiksel olarak 0.05 düzeyinde anlamlıdır (Kruskal-Wallis ki-kare = 38.5622, $df = 2$, $p = 4.23e-09$); bu nedenle ilk grupta yüzde 25, kullanışlılığı 5'ten düşük olarak işaretlemiştir.

Box-Plot Grafikleri

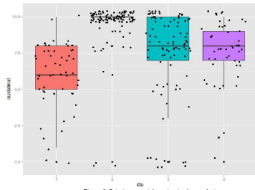


Figure 3. Printing material scoring in clusters 1-4

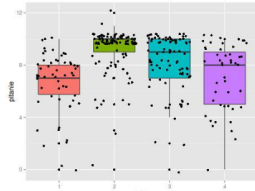


Figure 4. Maps for "Quality of food in the dining rooms" in various groups

İki değişken için box-plot grafikleri incelendiğinde 2. kümenin diğer kümelerle kıyasla daha homojen dağıldığını ve Range'inin daha düşük olduğunu söyleyebiliriz. En yüksek Range sahip kümelerin 1. ve 4. olduğunu söylenebilir. Her bir küme için aykırı değerler bulunmaktadır. 3. ve 4. Kümelerin medyanlarının birbirine eşit hatta yakın olması dolayısıyla benzerlikten şüphe edilebilir.

Faktör Analizi

Faktörlerin sıralanması açısından, öğretmenin bireysel becerisi ve bölümün genel performansı en büyük ağırlığa sahiptir. Faktör analizi sonucunda bu sorunları çözmeye yönelik yönetimin en büyük etkiye sahip olacağı sonucuna varılmıştır. Faydalılık da yeterince önemli bir faktördür ve daha önce elde edilen sonuçları dikkate alarak özel bir dikkat gerektirir.

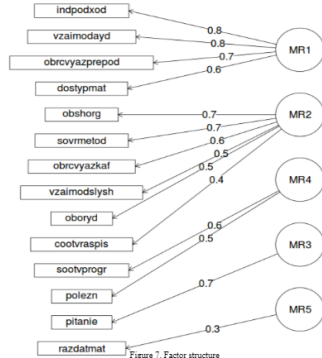
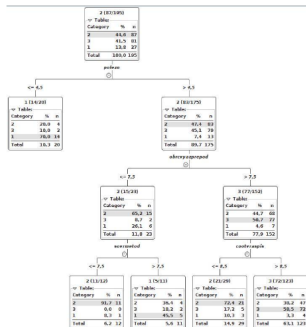


Figure 7. Factor structure

Faktör Yapısı

Karar Ağacı Algoritması

Çıkan sonucun etkileyici olduğunu söyleyemeyiz. Çapraz doğrulama yöntemi olan leave-one-out çapraz doğrulama yöntemi ile en yüksek doğruluk oranı yüzde 63 olarak bulunmuştur. F-ölçütü değeri ise yüzde 20'lik test veri seti üzerinde oldukça iyi bir sonuç verdiği görülür; "1" sınıfı için 0.8 ve "2" ve "3" sınıfları için tatmin edici sonuçlar alındığı da söylenebilir(sırasıyla 0.6 ve 0.632). Bu durum, "3" sınıfının düşük özgüllüğü (yüzde 0.565) ve "2" sınıfının düşük duyarlılığı (yüzde 0.529) ile açıklanabilir.



Karar Ağacı

Sonuçlar

- Öğrencilerin dikkatini anketi düşünceli bir şekilde doldurmaya odaklanmaları konusunda teşvik etmek önemlidir, çünkü birçok anket boşa gitmekte ve sonuç, orijinal sayıdan 40'a düşmüştür.
- Elektronik test formu kullanılmalı ve yeni alanlar eklenmelidir (yaş, deneyim, eğitim vb.), çünkü tahmin modelinin kalitesi düşüktür(yüzde 63-64). Anket anonimleştirilmesi gerekmektedir.
- İyileştirme için öncelikli parametreler belirlenebilir: "Baskı malzemelerinin varlığı, kalitesi ve kullanışlılığı", "Yemekhane kalitesi", "Yapılan eğitimin pratik faydalılığı", "Eğitim salonunun donanımı", "Kurs materyalinin anlaşılması" ve "Öğrenciye bireysel yaklaşım".
- Öğretmenleri işe alırken, çalışma saatlerine uyum sağlama, bireysel çalışmaya maksimum özen gösterme, öğrencilerle etkili etkileşim konularına dikkat etmek önemlidir.

Referanslar

- [1] Michael R. Berthold, Nicolas Cebron, Fabian Dill, Thomas R. Gabriel, Tobias Kotter, Thorsten Meinl *Studies in classification, data analysis, and knowledge organization. knime: the konstanz information miner*//Springer, 2007
- [2] Zhang, M., Mu, H., Li, G. ve Ning, Y. (2009), Forecasting the transport energy demand based on PLSR method in China, *Energy*, 34, 1396-1400.
- [3] R CORE TEAM (2013), R: A language and environment for statistical computing. R Foundation for Statistical Computing <http://www.R-project.org/>, 1-12.