Ders Bilgi Formu

Ders Adı	Kodu	Yerel Kredi	AKTS	Ders (saat/hafta)	Uygulama (saat/hafta)	Laboratuar (saat/hafta)
Veri Madenciliğine Giriş	IST4412	3	5	3	0	0

Önkoşullar	Yok

Yarıyıl	Bahar

Dersin Dili	İngilizce, Türkçe
Dersin Seviyesi	Lisans Seviyesi
Ders Kategorisi	Uzmanlık/Alan Dersleri
Dersin Veriliş Şekli	Yüz yüze

Dersi Sunan Akademik Birim	İstatistik Bölümü
Dersin Koordinatörü	Gülhayat Gölbaşı Şimşek
Dersi Veren(ler)	Gülhayat Gölbaşı Şimşek, Coşkun Parim
Asistan(lar)ı	

Dersin Amacı	Bu dersin amacı, öğrencilere veri madenciliğinin temel kavramlarını, yöntemlerini ve uygulama alanlarını öğretmektir. Öğrenciler, büyük veri kümelerini analiz etme, sınıflandırma, kümeleme, birliktelik kuralları çıkarımı, boyut indirgeme ve yapay zekâ tabanlı veri analizi teknikleri hakkında bilgi sahibi olacak; istatistiksel yazılım ve programlama araçlarını kullanarak uygulamalar yapabilecektir.
Dersin İçeriği	Veri madenciliğine giriş; veri ön işleme teknikleri, sınıflandırma algoritmaları (Karar Ağaçları, Naive Bayes, KNN, Destek Vektör Makineleri); kümeleme yöntemleri (K-means, Hiyerarşik kümeleme, Yoğunluk temelli yöntemler); birliktelik kuralları (Apriori, FP-Growth); boyut indirgeme (PCA, LDA); yapay zekâ ve makine öğrenmesi tabanlı yaklaşımlar; Python/R uygulamaları ve örnek vaka çalışmaları.
Opsiyonel Program Bileşenleri	Yok

Ders Öğrenim Çıktıları

1	Bu dersi başarıyla tamamlayan öğrenciler,
2	Veri madenciliğinin temel kavramlarını ve yöntemlerini açıklayabilecekler.
3	Veri ön işleme ve veri hazırlama tekniklerini uygulayabileceklerdir.
4	Sınıflandırma, kümeleme ve birliktelik kuralları algoritmalarını karşılaştırır ve uygulayabileceklerdir.
5	Boyut indirgeme yöntemlerini veri setleri üzerinde uygulayabilecekler.
6	Yapay zekâ ve makine öğrenmesi yöntemlerini istatistiksel problemlere uyarlayabileceklerdir.
7	İstatistiksel programlama dilleri (R/Python) ile veri madenciliği uygulamaları geliştirebileceklerdir.
8	Gerçek hayat verileri üzerinde proje geliştirir ve sonuçları yorumlayabileceklerdir.

Haftalık Konular ve İlgili Ön Hazırlık Çalışmaları

Hafta	Konular	Ön Hazırlık

1	Konu Anlatımı: Veri Madenciliğine giriş, veri bilimi ve istatistik ilişkisi Sınıf-içi Uygulama (40 dk): Python/R tanıtımı ve basit veri örnekleri Sınıf-içi Tartışma (10 dk): Veri bilimi-istatistik ilişkisi üzerine tartışma	1. Giriş bölmünün okunması Kaynak: Ders Kitabı[1],1–20 2. Python temel veri yapıları tekrar edilmesi , Ders Kitabı[2],1–25
2	Konu Anlatımı (30 dk): Veri türleri, veri kalitesi ve veri ön işleme Sınıf-içi Uygulama (40 dk): Pandas ile veri yükleme, eksik değerlerin temizlenmesi Sınıf-içi Tartışma (10 dk): Veri kalitesi sorunlarının sonuçları üzerine tartışma	1. Pandas ile veri seti yükleme, eksik değerleri incelemesi, Kaynak: Ders Kitabı [2], 1-25 2. Bölüm 2 kısmının okunması, Kaynak Ders Kitabı, 35-60
3	Konu Anlatımı: Veri temizleme, dönüştürme ve öznitelik seçimi Sınıf-içi Uygulama (40 dk.): Python ile normalizasyon ve standardizasyon uygulamaları) Sınıf-içi Tartışma: Öznitelik seçiminin model başarısına etkisi (10 dk)	1. Veri standardizasyonu ve normalizasyon yöntemlerini araştırması bölümlerinin okunması, Kaynak: Ders Kitabı [1],101-140, Ders Kitabı [2], 61-85
4	Konu Anlatımı: Sınıflandırmaya giriş, Karar ağaçları Sınıf-içi Uygulama (40 dk.): ID3/CART ile sınıflandırma örnekleri Sınıf-içi Tartışma (10 dk.): Karar ağaçlarının avantaj /dezavantajları	1. Entropi ve bilgi kazancı kavramlarının incelenmesi, Kaynak: Ders Kitabı [1], 323,350; Kaynak Ders Kitabı [2], 145-170
5	Konu Anlatımı: Naive Bayes, k-En Yakın Komşu Sınıf-içi Uygulama (40 dk.):Naive Bayes ve k-NN ile uygulama Sınıf-içi Tartışma (10 dk.): Küçük veri setlerinde yöntem karşılaştırması	1. Koşullu olasılık ve Bayes teoremini tekrar edimesi, Kaynak: Ders Kitabı [1],351-370,;Kaynak: Ders Kitabı [2], 180 -205
6	Konu Anlatımı: Destek Vektör Makineleri Sınıf-içi Uygulama (40 dk.): Python sklearn ile SVM uygulaması Sınıf-içi Tartışma (10 dk.): Doğrusal/Doğrusal olmayan sınıflandırıcı tartışması	1. Doğrusal ayrılabilirlik ve hiper-düzlem kavramlarının gözden geçirilmesi. Kaynak: Han [1], Bölüm 8, ss 380–400 2. Support Vector Machines konusunun incelenmesi., Kaynak: Ders Kitabı, 337-364
7	Konu Anlatımı: Yapay sinir ağlarına giriş Sınıf-içi Uygulama (40 dk.): Basit yapay sinir ağı eğitimi Sınıf-içi Tartışma (10 dk.): Küme sayısı seçimi tartışması	1. Doğrusal ayrılabilirlik ve hiper-düzlem kavramlarının gözden geçirilmesi. Kaynak: Han [1], Bölüm 8, ss 380–400 2. Support Vector Machines konusunun incelenmesi., Kaynak: Ders Kitabı, 337-364
8	Midterm 1 / Practice or Review	
9	Konu Anlatımı: Kümeleme: K-means algoritması Sınıf-içi Uygulama (40 dk.): K-means uygulamaları Sınıf-içi Tartışma (10 dk.): Küme sayısı seçimi tartışması	1. Öklid mesafesi ve centroid kavramlarının araştırılması. Kaynak: Ders Kitabı [1], ss. 443–470, Kaynak: Ders Kitabı [2],500-520
10	Konu Anlatımı: Hiyerarşik kümeleme, DBSCAN Sınıf-içi Uygulama (40 dk.): Python ile dendrogram ve DBSCAN uygulaması Sınıf-içi Tartışma (10 dk.): Yoğunluk tabanlı yöntemlerin avantajları	1. Dendrogram örneklerinin incelenmesi, yoğunluk kavramının okunması. Kaynak: Ders Kitabı [1], Bölüm 10, ss. 471–500, Kaynak: Ders Kitabı [2], Bölüm 8, ss. 521–540
11	Konu Anlatımı: Birliktelik kuralları: Apriori algoritması Sınıf-içi Uygulama (40 dk.): Market sepeti analizi uygulaması Sınıf-içi Tartışma (10 dk.): Apriori kurallarının ticari kullanım tartışması	1. Market sepeti analizine dair kısa makalenin okunması.Kaynak: Ders Kitabı [1], Bölüm 6, ss. 227–260;Kaynak: Ders Kitabı [2], Bölüm 6, ss. 350–370

12	Konu Anlatımı: FP-Growth yöntemi Sınıf-içi Uygulama (40 dk): Python ile FP-Growth uygulaması Sınıf-içi Tartışma (10 dk.): Apriori ile FP-Growth kıyaslaması	1. Apriori ile FP-Growth arasındaki farkların araştırılması.Kaynak: Ders Kitabı [1], Bölüm 6, ss. 261–280; Kaynak: Ders Kitabı [2], Bölüm 6, ss. 371–390
13	Konu Anlatımı: Boyut indirgeme: PCA, LDA Sınıf-içi Uygulama (40 dk.): PCA ve LDA uygulamaları Sınıf-içi Tartışma (10 dk.): Boyut indirgemede bilgi kaybı tartışması	1. Matrislerde özdeğer ve özvektör kavramlarının tekrar edilmesi, feature reduction konusunun incelenmesi.. Kaynak: Ders Kitabı [1], Bölüm 3, ss. 141–160;Kaynak: Ders Kitabı [3], Bölüm 6, ss. 215–245
14	Konu Anlatımı: Yapay zekâ tabanlı veri madenciliği yöntemleri Sınıf-içi Uygulama (40 dk.): Random Forest / XGBoost uygulaması Sınıf-içi Tartışma (10 dk): Klasik yöntem vs. yapay zekâ tabanlı yöntem tartışması	1. Güncel makine öğrenmesi algoritmaları (Random Forest, XGBoost vb.) hakkında makale okunması. Kaynak: Ders Kitabı [1], Bölüm 12, ss. 501–540 ; Kaynak: Ders Kitabı [3], Bölüm 8, ss. 295–336
15	Proje sunumları ve genel değerlendirme	Proje raporlarının hazırlanması, sunumların yapılması
16	Final	

Değerlendirme Sistemi

Etkinlikler	Sayı	Katkı Payı
Devam/Katılım	12	5
Laboratuar		
Uygulama		
Arazi Çalışması		
Derse Özgü Staj		
Küçük Sınavlar/Stüdyo Kritiği		
Ödev		
Sunum/Jüri	1	15
Projeler	1	20
Seminer/Workshop		
Ara Sınavlar	1	20
Final	1	40
Dönem İçi Çalışmaların Başarı Notuna Katkısı		60
Final Sınavının Başarı Notuna Katkısı		40
TOPLAM		100

AKTS İşyükü Tablosu

Etkinlikler	Sayı	Süresi (Saat)	Toplam İşyükü
Ders Saati	14	3	42

Laboratuar			
Uygulama			
Arazi Çalışması			
Sınıf Dışı Ders Çalışması	14	4	56
Derse Özgü Staj			
Ödev			0
Küçük Sınavlar/Stüdyo Kritiği			0
Projeler	1	20	20
Sunum / Seminer	1	10	10
Ara Sınavlar (Sınav Süresi + Sınav Hazırlık Süresi)	1	12	12
Final (Sınav Süresi + Sınav Hazırlık Süresi)	1	20	20
		Toplam İşyükü	160
		Toplam İşyükü / 30(s)	5.33
		AKTS Kredisi	5

Diğer Notlar	Yok