



DERS BİLGİ FORMU

FAKÜLTE / ENSTİTÜ ADI	Fen Edebiyat Fakültesi
BÖLÜM / PROGRAM / ANABİLİM DALI ADI	İstatistik
DERSİN ADI	Veri Madenciliğine Giriş
DERSİN KODU	IST4412
YEREL KREDİSİ	3
AKTS KREDİSİ	5
HAFTALIK DERS SAATİ	3
HAFTALIK UYGULAMA SAATİ	0
HAFTALIK LABORATUVAR SAATİ	0
ÖNKOŞULLAR	Yok
YARIYIL	Güz ve Bahar
DERSİN DİLİ	Türkçe, İngilizce
DERSİN SEVİYESİ	Lisans
DERSİN TÜRÜ	Seçmeli @İstatistik Lisans Programı
DERSİN KATEGORİSİ	Uzmanlık/Alan Dersleri
DERSİN VERİLİŞ ŞEKLİ	Yüz Yüze
DERSİ SUNAN AKADEMİK BİRİM	İstatistik Bölümü
DERSİN KOORDİNATÖRÜ	Ali Hakan Büyüklü
ASİSTAN(LAR)	Arş. Gör. Muzaffer GÖZTAŞ
DERSİN AMACI	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.
DERSİN İÇERİĞİ	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.
DERS KİTABI / MALZEMESİ / ÖNERİLEN KAYNAKLAR	Ders Kitabı: [1] Jiawei Han, Micheline Kamber, Jian Pei, <i>Data Mining: Concepts and Techniques</i> . 4th Edition, 2022. [2] Tan, P.N., Steinbach, M., Karpatne, A., & Kumar, V. <i>Introduction to Data Mining</i> . 2nd Edition, Pearson, 2018. [3] Salgado, J., <i>Machine Learning for Data Mining</i> , Packt Publishing, 2019 Zorunlu Kaynaklar: Önerilen Kaynaklar: - Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, <i>Introduction to Data Mining</i>, 2nd Edition, 2018. - Trevor Hastie, Robert Tibshirani, Jerome Friedman, <i>The Elements of Statistical Learning</i>, 2nd Edition, 2017. - Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, <i>An Introduction to Statistical Learning with Applications in R and Python</i>, 2nd Edition, 2023.



Ders Öğrenim Çıktıları	Bu dersi başarıyla tamamlayan öğrenciler, 1. Veri madenciliğinin temel kavramlarını ve yöntemlerini açıklayabilecekler. 2. Veri ön işleme ve veri hazırlama tekniklerini uygulayabileceklerdir. 3. Sınıflandırma, kümeleme ve birliktelik kuralları algoritmalarını karşılaştırır ve uygulayabileceklerdir. 4. Boyut indirgeme yöntemlerini veri setleri üzerinde uygulayabilecekler. 5. Yapay zekâ ve makine öğrenmesi yöntemlerini istatistiksel problemlere uyarlayabileceklerdir. 6. İstatistiksel programlama dilleri (R/Python) ile veri madenciliği uygulamaları geliştirebileceklerdir. 7. Gerçek hayat verileri üzerinde proje geliştirir ve sonuçları yorumlayabileceklerdir.
DEĞERLENDİRME SİSTEMİ	
Etkinlikler	Sayı Katkı Payı
Devam/Katılım: • İçerik: Öğrencilerin derse devam etmeleri ve katılmaları • Detaylı Değerlendirme Kriterleri: - Derse aktif katılım ve soru sorma - Sınıf-içi tartışmalara ve problem çözme süreçlerine katkı sağlayabilme	14 %5
Laboratuvar:	
Uygulama (Sözlü Sınav): • İçerik: Öğrencilerden lineer cebir dersinin temel kavramlarını açıklamalarının ve uygulamalı bir soruya çözüm önerisi sunmalarının istenmesi • Format: Öğrenci ile bireysel olarak gerçekleştirilecek sözlü sınav (5-10 dakika) • Detaylı Değerlendirme Kriterleri: -Kavramları açıklayabilme -Problem çözebilme -Problem çözümlerini anlatabilme	
Arazi Çalışması	
Derse Özgü Staj	
Kısa Sınavlar/Stüdyo Kritiği (Zorunlu): • İçerik: Sınav haftasına kadar işlenen konuların tümünü kapsayan kapsamlı soruların sorulması • Format: Yüz yüze. Çoktan seçmeli kısa sınav (5-10 dakika) • Detaylı Değerlendirme Kriterleri: -Derste işlenen teorik konular ile ilgili problemleri çözebilme	3 %10
Ödev: • İçerik: Derste işlenen temel kavramların eleştirel biçimde	



yorumlanmasını ve ilgili kavramların disiplin-içi ve disiplinlerarası alanlarda örneklerinin bulunmasını içeren haftalık ödevlerin verilmesi • Format: Yazılı raporlar ve grup sunumları • Detaylı Değerlendirme Kriterleri: - Bir problemin çözüm sürecini mantıklı ve doğru bir şekilde yazabilme - Kavramların uygulamadaki örneklerini bulabilme - Uygulamalı düşünebilme, yorumlama ve gerekçelendirme süreçlerinin yürütülebilmesi		
Sunum/Jüri: • İçerik: Öğrencilerin kendi öğrenme süreçlerini değerlendirmelerinin ve grup sunumları yapmalarının istenmesi • Format: Grup sunumları • Detaylı Değerlendirme Kriterleri: -Öğrenilen konuların doğru bir şekilde açıklanabilmesi -Sunum tekniklerinin doğru kullanılması	1	%10
Proje: • İçerik: Öğrencilerden akademik dönem sonunda teslim edilecek bir proje önerisi (TÜBİTAK 2209 A/B) yazmalarının istenmesi • Format: Yazılı raporlar ve grup sunumları • Detaylı Değerlendirme Kriterleri: -Özgün bir araştırma konusunun bulunabilmesi -Bir araştırma önerisinin bilimsel ilkeler ve ilgili kılavuzlar doğrultusunda yazılabilmesi	1	%15
Seminer/Workshop		
Ara Sınavlar: • İçerik: Sınav haftasına kadar işlenen konuların tümünü kapsayan kapsamlı sorular • Format: Yüz yüze. Sınav (90 dakika) • Detaylı Değerlendirme Kriterleri: -Dersin temel kavramlarının anlaşıldığının gösterilmesi -Teorik konularla ilgili problemlerin çözülebilmesi -Teorik düşünme süreçlerinin yürütülmesi	1	%20
Final: • İçerik: Dersin tüm içeriğini kapsayan kapsamlı sorular • Format: Yüz yüze. Sınav (90 dakika) • Detaylı Değerlendirme Kriterleri: -Derste işlenen tüm konuların derinlemesine kavranmış olduğunun gösterilmesi -İleri düzey problem çözme becerilerinin kullanılabilmesi	1	%40
Dönem İçi Çalışmaların Başarı Notuna Katkısı		%60
Final Sınavının Başarı Notuna Katkısı		%40



TOPLAM		%100
HAFTALIK KONULAR VE İLGİLİ ÖN HAZIRLIK ÇALIŞMALARI		
HAFTALAR	KONULAR	Ön Hazırlık
1	Konu Anlatımı: Veri Madenciliğine giriş, veri bilimi ve istatistik ilişkisi Sınıf-İçi Uygulama (40 dk): Python/R tanıtımı ve basit veri örnekleri Sınıf-İçi Tartışma (10 dk): Veri bilimi-istatistik ilişkisi üzerine tartışma	1. Giriş bölümünün 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-İçi Uygulama (40 dk): Pandas ile veri yükleme, eksik değerlerin temizlenmesi Sınıf-İçi Tartışma (10 dk): Veri kalitesi sorunlarının sonuçları üzerine tartışma Kısa Sınav 1 (15 dk): Ders sonunda, derste işlenen konuları içeren bir kısa sınavın yapılması	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-İçi Uygulama (40 dk.): Python ile normalizasyon ve standardizasyon uygulamaları Sınıf-İçi 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-İçi Uygulama (40 dk.): ID3/CART ile sınıflandırma örnekleri Sınıf-İçi Tartışma (10 dk.): Karar ağaçlarının avantaj/dezavantajları Kısa Sınav 2: (15 dk) Ders sonunda, derste işlenen konuları içeren bir kısa sınavın yapılması	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-İçi Uygulama (40 dk.): Naive Bayes ve k-NN ile uygulama Sınıf-İçi Tartışma (10 dk.): Küçük veri setlerinde yöntem karşılaştırması	1. Koşullu olasılık ve Bayes teoremini tekrar edilmesi, Kaynak: Ders Kitabı [1],351-370,;Kaynak: Ders Kitabı [2], 180 -205
6	Konu Anlatımı: Destek Vektör Makineleri Sınıf-İçi Uygulama (40 dk.): Python sklearn ile SVM uygulaması Sınıf-İçi 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-İçi Uygulama (40 dk.): Basit yapay sinir ağı eğitimi Sınıf-İçi Tartışma (10 dk.): Küme sayısı seçimi tartışması	1. Lineer regresyon ve lojistik regresyonun tekrar edilmesi, sigmoid fonksiyonunun incelenmesi. Kaynak: Ders Kitabı [1], Bölüm 9, ss. 401-430 2. Neural Networks konusunun incelenmesi. Kaynak: Ders Kitabı [3], Bölüm 10, ss. 365-390
8	Ara Sınav 1	
9	Konu Anlatımı: Kümeleme: K-means algoritması Sınıf-İçi Uygulama (40 dk.): K-means uygulamaları	1. Öklid mesafesi ve centroid kavramlarının araştırılması. Kaynak: Ders Kitabı [1], ss. 443-470, Kaynak: Ders Kitabı [2],500-520



	Sınıf-içi Tartışma (10 dk.): Küme sayısı seçimi tartışması	
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 Kısa Sınav 3 (15 dk): Ders sonunda, derste işlenen konuları içeren bir kısa sınav yapılması	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 indirgemedeki 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	İşlenen konuların tümünün tekrar edilmesi

AKTS İŞYÜKÜ TABLOSU

Etkinlikler	Sayı	Süresi (Saat)	Toplam İşyükü
Ders Saati	14	3	42
Laboratuvar			
Uygulama (sözlü Sınav)			
Arazi Çalışması			
Sınıf Dışı Ders Çalışması	14	4	56
Derse Özgü Staj			
Ödev			
Küçük Sınavlar/Stüdyo Kritiği	3	3	9



Projeler	1	15	15
Sunum / Seminer			
Ara Sınavlar (Sınav Süresi + Sınav Hazırlık Süresi)	1	12	12
Final (Sınav Süresi + Sınav Hazırlık Süresi)	1	20	20
Toplam İş yükü:			154
Toplam İş yükü / 30(s):			5.133
AKTS Kredisi:			5



COURSE INFORMATION FORM

FACULTY / GRADUATE SCHOOL	Faculty of Arts and Sciences
DEPARTMENT / PROGRAMME	Statistics
TITLE OF COURSE	Introduction to Data Mining
CODE	IST4412
LOCAL CREDIT	3
ECTS	5
LECTURE HOUR / WEEK	3
PRACTICAL HOUR / WEEK	0
LABORATORY HOUR / WEEK	0
PREREQUISITE	None
SEMESTER	Fall and
COURSE LANGUAGE	English, Turkish
LEVEL OF COURSE	First Cycle
COURSE TYPE	Elective @Bachelor Program in Statistics
COURSE CATEGORY	Major Area Courses
MODE OF DELIVERY	Face to Face
OWNER ACADEMIC UNIT	Department of Statistics
COURSE COORDINATOR	Ali Hakan Büyüklü
ASSISTANT(S)	Arş. Gör. Muzaffer GÖZTAŞ
COURSE OBJECTIVES	The aim of this course is to provide students with fundamental concepts, methods, and application areas of data mining. Students will acquire knowledge and skills in data preparation, classification, clustering, association rules, dimensionality reduction, and artificial intelligence-based approaches, and will gain experience through statistical programming (R/Python) applications.
COURSE CONTENT	Introduction to data mining; data preprocessing techniques; classification algorithms (Decision Trees, Naive Bayes, k-Nearest Neighbors, Support Vector Machines); clustering methods (K-means, Hierarchical clustering, Density-based methods); association rule mining (Apriori, FP-Growth); dimensionality reduction (PCA, LDA); artificial intelligence and machine learning-based approaches; applications using Python/R and case studies.
RECOMMENDED OR REQUIRED READINGS	Coursebooks: [4] Jiawei Han, Micheline Kamber, Jian Pei, <i>Data Mining: Concepts and Techniques</i> . 4th Edition, 2022. [5] Tan, P.N., Steinbach, M., Karpadne, A., & Kumar, V. <i>Introduction to Data Mining</i> . 2nd Edition, Pearson, 2018. [6] Salgado, J., <i>Machine Learning for Data Mining</i> , Packt Publishing, 2019 Required Readings: Recommended Readings - Pang-Ning Tan, Michael Steinbach, Anuj Karpadne, Vipin Kumar, <i>Introduction to Data Mining</i>, 2nd Edition, 2018. - Trevor Hastie, Robert Tibshirani, Jerome Friedman, <i>The Elements of Statistical Learning</i>, 2nd Edition, 2017.



- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R and Python, 2nd Edition, 2023.

Course Learning Outcomes

- Upon successful completion of the course, students will be able to
1. Explain the fundamental concepts and methods of data mining.
 2. Apply data preprocessing and preparation techniques.
 3. Compare and implement classification, clustering, and association rule algorithms.
 4. Apply dimensionality reduction methods on datasets.
 5. Adapt artificial intelligence and machine learning methods to statistical problems.
 6. Develop data mining applications using statistical programming languages (R/Python).
 7. Conduct projects on real-life datasets and interpret results.

EVALUATION SYSTEM

Activities	Number	Percentage of Grade
Attendance/Participation: Content: Student attendance and participation in the course. Detailed Assessment Criteria: -Active participation in lessons and asking questions -Ability to contribute to in-class discussions and problem-solving processes	14	%5
Laboratory		
Application (Oral Examination): Content: Students will be asked to explain fundamental concepts of linear algebra and to propose a solution to a practical problem. Format: Individual oral examination with each student (5-10 minutes). Detailed Assessment Criteria: -Ability to explain concepts -Ability to solve problems -Ability to articulate problem solutions		
Field Work		
Special Course Internship (Work Placement)		
Quizzes/Studio Critics: Content: Comprehensive questions covering all topics addressed up to the exam week Format: Face-to-face multiple-choice quiz (5-10 minutes) Detailed Assessment Criteria: -Ability to solve problems related to the theoretical topics covered in the course	3	%10
Homework Assignments: Content: Weekly assignments requiring critical analysis of fundamental concepts covered in class and identification of examples of these concepts within both intra-disciplinary and interdisciplinary		



contexts • Format: Written reports and group presentations • Detailed Assessment Criteria: -Ability to logically and accurately demonstrate the problem-solving process -Ability to find practical examples of the concepts - Ability to carry out processes of applied thinking, interpretation, and justification		
Presentations/Jury: • Content: Students will be asked to evaluate their own learning processes and deliver group presentations • Format: Group presentations • Detailed Assessment Criteria: -Ability to accurately explain the topics learned -Proper use of presentation techniques	1	%10
Project: • Content: Students will be asked to write a project proposal (TÜBİTAK 2209 A/B) to be submitted at the end of the academic term. • Format: Written reports and group presentations • Detailed Assessment Criteria: -Ability to identify an original research topic -Ability to write a research proposal in accordance with scientific principles and relevant guidelines	1	%15
Seminar/Workshop Midterms: • Content: Comprehensive questions covering all topics addressed up to the exam week • Format: Face-to-face written exam. (90 minutes). • Detailed Assessment Criteria: -Demonstration of understanding of the fundamental concepts of the course -Ability to solve problems related to theoretical topics -Ability to carry out theoretical reasoning processes	1	%20
Final: • Content: Comprehensive questions covering the entire content of the course • Format: Face-to-face written exam. (90 minutes). • Detailed Assessment Criteria: -Ability to apply advanced problem-solving skills -Demonstration of a thorough understanding of all topics covered in the course	1	%40
Percentage of In-Term Studies		%60



Percentage of Final Examination

%40

TOTAL

%100

WEEKLY SUBJECTS AND RELATED PREPARATION STUDIES

WEEKS	COURSE OUTLINE	Related Preparation
1	Lecture: Introduction to data mining, relationship with statistics and data science Practice (40 minutes): Python/R basics, small dataset exploration In Class Discussion (5 minutes) : Role of statistics in data science	1. Reading the introduction section, Source: Coursebook [1], pp. 1–20 2. Reviewing the basic data structures in Python, Source: Coursebook [2], pp. 1–25
2	Lecture: Data types, data quality, and preprocessing Practice (40 minutes): Pandas: loading data, handling missing values In Class Discussion (5 minutes): Data quality problems in real datasets Quiz 1 (15 mts.): A quiz at the end of the class covering the topics taught during the session	1. Loading datasets with Pandas and examining missing values, Source: Coursebook [2], pp. 1–25 2. Reading Chapter 2, Source: Coursebook, pp. 35–60
3	Lecture: Data cleaning, transformation, and feature selection Practice (40 minutes): Normalization & standardization examples In Class Discussion (5 minutes) : Impact of feature selection on models	1. Reading the sections on data standardization and normalization methods, Source: Coursebook [1], pp. 101–140; Coursebook [2], pp. 61–85
4	Lecture: Introduction to classification, Decision Trees Practice (40 minutes): ID3/CART examples In Class Discussion (5 minutes): Advantages/disadvantages of trees Quiz 2 (15 mts.): A quiz at the end of the class covering the topics taught during the session	1. Examining the concepts of entropy and information gain, Source: Coursebook [1], pp. 323, 350; Coursebook [2], pp. 145–170
5	Lecture: Naive Bayes, k-Nearest Neighbors (k-NN) Practice (40 minutes): Naive Bayes & k-NN applications In Class Discussion (5 minutes) : Comparing NB vs k-NN on small datasets	1. Reviewing conditional probability and Bayes' theorem, Source: Coursebook [1], pp. 351–370; Coursebook [2], pp. 180–205
6	Lecture: Support Vector Machines (SVM) Practice (40 minutes): SVM with sklearn In Class Discussion (5 minutes) : Linear vs nonlinear classifiers	1. Reviewing the concepts of linear separability and hyperplanes, Source: Han [1], Chapter 8, pp. 380–400 2. Examining the topic of Support Vector Machines, Source: Coursebook, pp. 337–364
7	Lecture: Introduction to Artificial Neural Networks Practice (40 minutes): Training a simple ANN In Class Discussion (5 minutes) : Differences between ANN and regression	1. Reviewing linear regression and logistic regression, and examining the sigmoid function, Source: Coursebook [1], Chapter 9, pp. 401–430 2. Examining the topic of Neural Networks, Source: Coursebook [3], Chapter 10, pp. 365–390
8	Midterm 1	Review of all topics covered up to the exam week.
9	Lecture: Clustering: K-means algorithm Practice (40 minutes): K-means with Python In Class Discussion (5 minutes): Choosing number of clusters	1. Studying the concepts of Euclidean distance and centroids, Source: Coursebook [1], pp. 443–470; Coursebook [2], pp. 500–520



	Quiz 3 (15 mts.): A quiz at the end of the class covering the topics taught during the session	
10	Lecture: Hierarchical clustering, DBSCAN Practice (40 minutes): Dendrograms & DBSCAN In Class Discussion (5 minutes): Density-based clustering pros & cons	1. Examining examples of dendrograms and reading about the concept of density, Source: Coursebook [1], Chapter 10, pp. 471–500; Coursebook [2], Chapter 8, pp. 521–540
11	Lecture: Association rules: Apriori algorithm Practice (40 minutes): Market basket analysis In Class Discussion (5 minutes): Business uses of association rule	1. Reading a short article on market basket analysis, Source: Coursebook [1], Chapter 6, pp. 227–260; Coursebook [2], Chapter 6, pp. 350–370
12	Lecture: FP-Growth method Practice (40 minutes): FP-Growth implementation In Class Discussion (5 minutes): Apriori vs FP-Growth	1. Investigating the differences between Apriori and FP-Growth, Source: Coursebook [1], Chapter 6, pp. 261–280; Coursebook [2], Chapter 6, pp. 371–390
13	Lecture: Dimensionality reduction: PCA, LDA Practice (40 minutes): PCA & LDA in Python In Class Discussion (5 minutes): Trade-off: info loss vs simplicity	1. Reviewing the concepts of eigenvalues and eigenvectors in matrices and examining the topic of feature reduction, Source: Coursebook [1], Chapter 3, pp. 141–160; Coursebook [3], Chapter 6, pp. 215–245
14	Lecture: AI-based data mining approaches Practice (40 minutes): Random Forest / XGBoost In Class Discussion (5 minutes): Traditional vs AI-based methods	1. Reading an article on modern machine learning algorithms (e.g., Random Forest, XGBoost), Source: Coursebook [1], Chapter 12, pp. 501–540; Coursebook [3], Chapter 8, pp. 295–336
15	Project presentations and general evaluation	Prepare project reports, rehearse presentations
16	Final	Review of all topics covered.

ECTS WORKLOAD TABLE

Activities	Number	Duration (Hour)	Total Workload
Course Hours	14	4	56
Laboratory			
Application	1	4	4
Field Work			
Study Hours Out of Class	14	4	56
Special Course Internship (Work Placement)			
Homework Assignments	1	4	4
Quizzes/Studio Critics	6	2	12
Project			
Presentations / Seminar			
Mid-Terms (Examination Duration + Examination Prep. Duration)	1	10	10
Final (Examination Duration + Examination Prep. Duration)	1	15	15
Total Workload:			157
Total Workload / 30(h):			5.23



ECTS Credit:

5

Ders Öğrenim Çıktısı & Program Çıktısı Matrisi

	<u>DÖÇ-1</u>	<u>DÖÇ-2</u>	<u>DÖÇ-3</u>	<u>DÖÇ-4</u>	<u>DÖÇ-5</u>	<u>DÖÇ-6</u>	<u>DÖÇ-7</u>	<u>DÖÇ-8</u>	<u>DÖÇ-9</u>
PC-1 İstatistik biliminin temelini oluşturan teorik ve uygulamalı alanlardaki bilgilerini, alanla ilgili problemleri tanımlama, modelleme ve çözüm üretmede kullanabileceklerdir. Use their theoretical and applied knowledge in the science of statistics to identify, model, and solve problems related to the field.									
PC-2 İstatistiksel problemlerin belirlenmesi, uygun veri toplama yöntemlerinin seçilmesi, verinin düzenlenmesi ve yorumlanması süreçlerinde yeterlilik gösterecek ve bu problemlerin çözümünde gerekli analiz ve modelleme yöntemlerini seçip uygulayabileceklerdir. Demonstrate competence in identifying statistical problems, selecting appropriate data collection methods, organizing and interpreting data, and select and apply necessary analysis and modelling methods to solve these problems.									
PC-3 Herhangi bir olgu, süreç ya da ürünü istatistiksel bakış açısıyla analiz edip yorumlayabilecek ve karşılaştıkları sorunlara uygun modern istatistiksel yöntemlerle çözüm geliştirebileceklerdir. Analyse and interpret a phenomenon, process, or product from a statistical perspective and develop solutions to encountered problems using appropriate modern statistical methods.	5	5	5	5	5	5	4		
PC-4 Disiplinlerarası bir yaklaşımla, farklı alanlarda edinmiş oldukları bilgileri sentezleyebileceklerdir. Synthesise knowledge acquired from different disciplines through an interdisciplinary approach									
PC-5 İstatistik alanında edindikleri bilgi birikimlerini ve istatistiksel düşünme becerilerini, teorik istatistik, veri bilimi, finansal istatistik, sağlık bilimlerinde istatistik gibi disiplin-içi ve disiplinlerarası uzmanlık alanlarında geliştirebileceklerdir.	5	5	4	5	4	4	5		



Advance their acquired knowledge in statistics and statistical thinking skills in both disciplinary and interdisciplinary areas of specialisation such as theoretical statistics, data science, financial statistics, and statistics in health sciences.									
PÇ-6 İstatistik alanında yaygın olarak kullanılan en az bir programlama dili ile bilgisayar ve yapay zekâ teknolojilerini, problemleri çözmek, veri analizi yapmak ve simülasyonlar gerçekleştirmek için kullanabileceklerdir. Use at least one programming language and computer and artificial intelligence technologies widely employed in statistics for problem-solving, data analysis, and simulations.	4	5	5	4	5	4	5		
PÇ-7 İstatistik ve ilgili alanlardaki bilimsel ve teknolojik gelişmeleri izleyebilecek, kariyer fırsatlarını değerlendirerek kişisel ve mesleki gelişim hedeflerini belirleyebilecek ve bu hedeflere ulaşmak için hayat boyu öğrenme stratejilerini kullanabileceklerdir. Follow scientific and technological developments in statistics and related fields, assess career opportunities, identify personal and professional development goals, and adopt lifelong learning strategies to achieve these goals									
PÇ-8 Bilimsel araştırmalarını ve mesleki faaliyetlerini yürütürken doğabilecek hukuksal sonuçları ve toplumsal etkileri dikkate alarak mesleki etik ilkeler, kalite standartları ile evrensel değerler doğrultusunda ve sosyal sorumluluk bilinci ve adalet duygusuyla hareket edebileceklerdir. Act with a sense of social responsibility and justice and in accordance with professional ethical principles, quality standards, and universal values by taking into account potential legal and societal consequences of their scientific research and professional activities.									
PÇ-9 Bireysel olarak ya da takımlarda etkin biçimde çalışabileceklerdir. Work effectively both independently and as part of a team.									
PÇ-10 İstatistik alanında güvenilir bilgi kaynaklarına ulaşarak literatür taraması yapabilecek ve akademik araştırma tasarlayıp yürütebileceklerdir. Access reliable sources of information, conduct literature reviews, and design and carry out academic research in the field of statistics.									
PÇ-11 İstatistiksel konuları, teorileri, ispatları, araştırmaları ve problem çözümlerini, istatistiksel terminoloji kullanarak tüm paydaşlara Türkçe ve İngilizcede sözlü ve yazılı olarak etkili biçimde aktarabileceklerdir.									



Effectively communicate statistical topics, theories, proofs, research, and problem solutions to all relevant stakeholders using appropriate mathematical terminology, both orally and in writing, in Turkish and in English.

--	--	--	--	--	--	--	--	--	--