

Veri Madenciliği ENM 424

Dersi Veren

Caner Erden

Tel

0264 295 7323

Email

cerden@sakarya.edu.tr

Ofis

M5, 5220

Ders Saati ve Yeri

1. Öğr: Çar 10:00-13:00,
5102

2. Öğr: Çar 17:00-20:00,
5102

Görüşme Saatleri

Salı 09:00-11:00

Google Class Kodu

jvobjs

Amaç

Veri Madenciliği veri tabanlarından bilgi keşfi süreci olarak tanımlanır. Veri madenciliği teknikleri sayesinde büyük veri tabanlarından gizli olan anlamlı bilgiler çıkarılır. Bu ders veri madenciliğinin istatistiksel, makine öğrenmesi ve veri tabanı yönünden temellerini içermektedir. Derste veri madenciliğinin temel kavramları gösterildikten sonra veri hazırlama, sınıflandırma, kümeleme, birliktelik analizleri gibi veri madenciliği uygulamaları anlatılacaktır. Ayrıca derste veri madenciliğinde veriden anlamlı bilgi keşfedilme sürecinden bahsedilecek ve veri madenciliğinde kullanılan istatistiksel metotlara yer verilecektir. Günümüzün trend konuları olan makine öğrenmesi ve yapay sinir ağları ile ilgili bilgiler de ders kapsamında anlatılacaktır.

1. Veri madenciliğinin temel kavramlarını, amaçlarını ve tekniklerini tanıtmak ve kullanımını yaygınlaştırılması
2. Geniş ölçekli veri tabanlarında analiz yeteneğinin kazandırılması
3. Veri madenciliği uygulamaları ve popüler konuları hakkında bilgi ve yeterlilik kazandırılması
4. Veri madenciliği uygulamaları geliştirmeye yönelik yeterlilik kazandırılması
5. Veriyi anlama, yorumlama, analiz etme ve görselleştirme gibi yeteneklerin kazandırılması

Ders Kitapları

1. Gökhan Silahtaroglu, Kavram ve Algoritmalarıyla Temel Veri Madenciliği, Papatya Yayıncılık (2008)
2. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Addison Wesley, (2005)
3. Han, J. and Kamber, M., Data Mining: Concepts and Techniques, 2nd Edition, Morgan Kaufmann, 2006 .
4. Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann, 2011. ISBN 978-0-12-374856-0.

Yazılım

Derste öğrenilen algoritma ve teknikler Python programlama dili ile geliştirilecek kodlar sayesinde uygulanacaktır. Python programlama dilinde özellikle, Numpy, scikit-learn, Scipy, pandas, matplotlib gibi kütüphanelerin bilinmesi önerilir. Python ile geliştirilen kodlar ile birlikte Orange aracı da kullanılarak veri madenciliği teknikleri uygulanacaktır. Ayrıca bu dersi programlama bilgisi olmayanların tercih etmesi önerilmemektedir.

Değerlendirme Sistemi

- Final %50
- Vize %25
- Ödev %10
- Kısa Sınav %15

Ders Akışı

Hafta	Konu
Hafta 1	Kavramsal Çerçeve
Hafta 2	Veri Hazırlama ve Önışleme
Hafta 3	Sınıflandırma
Hafta 4	Kural Tabanlı Sınıflandırıcılar
Hafta 5	Destek Vektör Makineleri
Hafta 6	Regresyon
Hafta 7	Yapay Sinir Ağları
Hafta 8	Birliktelik Analizi
Hafta 9	Kümeleme
Hafta 10	Kümeleme Algoritmaları
Hafta 11	**Ara Sınav**
Hafta 12	Veri Madenciliği Araçları
Hafta 13	Zaman Serisi Analizi
Hafta 14	İleri Veri Madenciliği Konuları
Hafta 15	**Final Sınavı**

Tarihler

Konu	Tarih
Ödev 1	12 Şub 20 - 19 Şub 20 (2. Hafta)
Kısa Sınav 1	26 Şub 20 (4. Hafta)
Ödev 2	11 Mar 20 - 18 Mar 20 (7. Hafta)
Ara Sınav	30 Mar 20 - 3 Nis 20 (9. Hafta)
Kısa Sınav 2	29 Nis 20
Final	11-24 May 2020 (15. Hafta)

Ödev Değerlendirme

Ödevler Google Class üzerinden toplanacaktır. Ödev süresini 1-2-3-4 gün geciktirmeye ceza uygulanacak daha geç gönderilmesine izin verilmeyecek.

Ek Bilgi

Bu ders bilgisayar programlama ve istatistik bilgisi gerektirmektedir.