

## **Ders Projesi**

### **Ders Projesi aşağıdaki adımlardan/bileşenlerden oluşur:**

1. Veri kümesini bulun (çevrimiçi herhangi bir veri kümesi seçin). Varolan bir veri kümesini değiştirebilir ve kendinizinkini oluşturabilirsiniz.
2. Veri kümeniz için anlamlı olan bir veri madenciliği problemi formüle edin. *Muhtemel Sınıflandırma, ancak İlişkilendirme veya Kümelemeyi seçmek iyidir.*
3. Bilgi temsili için modeli düşünün (Kurallar, Karar Ağaçları, vb.).
4. Test etmek için probleminiz/bilgi temsiliniz için WEKA'da mevcut/uygulanan birkaç (bu en az üç veya daha fazla anlamına gelir) algoritmalar/yöntemler seçin.
5. **Ders projeniz için planınızı bildirin:**
  - 5.1. Proje ekibinizi,
  - 5.2. Seçilen veri kümesi – tanımlayın. (örneğin, verileri özetleyin, ancak özniteliklerin sayısını, örnek sayısını, orijinal biçimi, kökenini vb. dahil edin)
  - 5.3. Çözülecek Veri Madenciliği sorunu.
  - 5.4. Bilginin temsil edildiği modelin değerlendirmesi.
  - 5.5. Kullanılacak algoritmalar.
6. Veri kümenizi madencilik için hazırlayın:
  - 6.1. Hangi özelliklerin problemle ilgili olduğuna ve dikkate alınması gerektiğine karar veriniz. Bazılarını atıyorsanız, kararınızı açıklayınız.
  - 6.2. Eksik nitelik değerleri var mı? Varsa sorunu çözünüz. Bunun için seçilen yöntemi açıklayınız. Rasyoneli açıklayınız.
  - 6.3. Verileriniz hazır .arff formatında değilse .arff formatına, dönüştürünüz.
7. Eğitim ve test seti için birkaç boyut kullanın (veri seti=eğitim seti + test seti). En az on katı çapraz doğrulama ile de çalıştırdığınızdan emin olunuz.
8. Eğitim verileriniz üzerinde algoritmaları çalıştırınız. Oluşturduğunuz model/modellerin verimliliğini/doğruluğunu test verilerinizde test ediniz.
9. **Projeniz için bir sonuç raporu yazın ve bir sunum oluşturun:**
  - 9.1. Veri kümesini, boyutu, öznitelikleri, öznitelik türünü ve olası değerleri tanımlayınız.
  - 9.2. Veri kümesiyle ilgili sorunları tanımlayınız: çok fazla öznitelik, eksik değer vb.
  - 9.3. (9.2) ile başa çıkmak için kullanılan çözümleri tanımlayınız.
  - 9.4. Veri madenciliği probleminizi tanımlayın/açıklayınız: Ne arıyorsunuz?
  - 9.5. Bilgi Temsil Modelini Tanımlayınız.
  - 9.6. Hangi algoritmaları (en az üç) seçtiğinizi açıklayınız. (Cesur olun ve sınıfta yapılanları kapsamayanları deneyin.)
  - 9.7. Eğitim/test veri kümesi boyutlarının seçimini tanımlayınız. Her çalıştırma için verimlilik sonuçlarını raporlayınız.
  - 9.8. Öğrenilenlerin güvenilirliğini tartışınız.
  - 9.9. Oluşturulan her model için verimlilik sonuçlarını analiz ediniz. - Garip bir şey var mı? Varsa açıklamaya çalışınız.

## Veri Madenciliği Proje

9.10.Projenin olası uygulamalarını/uygulamalarını tanımlayınız.

**Raporla ilgili sadece birkaç not:** Düzeltmeyi yaptığınızdan emin olun (belki hataları yakalamak için bir arkadaşınıza da okutabilirsiniz). *Şeyler* ve *çok* gibi kelimelerden kaçınınız. Bunun yerine, spesifik ve kesin olunuz. Sözcüklerin yazılışında tutarlı olunuz, örneğin *çalışma zamanı*, *iş zamanı* ve *çalışma süresi*, hepsi doğru olabilir, ancak birini seçin ve ona bağlı kalın. Gayri resmi yazmayın, siz kelimesinden kaçınınız.

**Nihai Rapor ve tüm destekleyici materyaller, elektronik ve basılı kopya için 04 Ocak Çarşamba, 13:00, son teslim tarihi**

10. Projenizi sınıfta sunun. – **Dönemin son 2 haftası (toplam 6 saat)**

Proje grubunuzla birlikte çalışacaksınız. **Proje grubu takımı 3 öğrenciden oluşacaktır.**

**Burada mevcut, muhtemelen ilginç veri kümelerine bazı bağlantılar:**

<https://www.kdnuggets.com/datasets/index.html>

[http://www.cs.waikato.ac.nz/ml/weka/index\\_datasets.html](http://www.cs.waikato.ac.nz/ml/weka/index_datasets.html)