

CENG420 Ödev 1

<https://archive.ics.uci.edu/> sitesinden Predict Students' Dropout and Academic Success veri setini indirin.

İndirdiğiniz data.csv dosyasını Weka Gui Chooser ->Tools-> Arffviewer'da açın. Bir readToken exception error göreceksiniz. Bunun sebebi csv'nin localizationda ;/, dönüşümüdür.

Weka Explorer'da Preprocess tabında OpenFile ile data.csv'yi açın. Dosyanın yukarıdaki gibi tanımlanmayan bir csv dosyası olduğu hatasını alacaksınız. Use Converter seçin. Field separator ; enclosure characters i de sadece " şeklinde güncelleyin. Veriniz artık yüklenecektir. (37 öznitelik, Merital status- Target) veriyi Save menüsünden data_upd.arff şeklinde kaydedin.

Classify sekmesinden j48 algoritmasıyla

- 1- Cross validate 10 fold
- 2- Cross validate 2 fold
- 3- Use training set
- 4- Percentage split %80

Seçeneklerini deneyin.

Her bir seçenek için

- i. elde ettiğiniz modeli kaydedin (örn student1.model şeklinde)
- ii. Bir sonuçlar.txt dosyası oluşturarak her bir model için

Correctly Classified Instances oranını, Doğru tahmin edilmiş Dropout sayısı, Dropout olarak tahmin edilmiş ancak graduate olan yanlış tahmin sayısı ve Dropout için tahmin kesinliği (precision) değerini yazın.

Sonuçlar dosyanızın sonuna hangi modeli sonraki tahminlemelerde kullanacağınızı ve bunu seçmenizin nedenini yazın.