

Veri Madenciliği Dersi Proje-3

Projenin teslim tarihi 15.12.2017 Cuma günü, saat 24:00'a kadardır.

Mail grubunda paylaşılan Veri dosyası üzerinde aşağıdaki işlemleri gerçekleştirmeniz beklenmektedir.

- 1- Elinizdeki veri kümesi ham veri şeklindedir. Bu veriyi analize uygun hale getirmeniz gerekmektedir. Bu nedenle veri ön-işleme adımlarından gerekli olanları ham veri kümesine uygulayarak, analize hazır işlenmiş bir veri kümesi elde ediniz. Bu analiz için WEKA değil, excel benzeri bir hesaplayıcı kullanmanız gerekmektedir.
- 2- Bu işlenmiş veri kümesini kullanarak WEKA üzerinde KNN sınıflayıcısı ile modellemeniz gerekmektedir. Bunu yaparken şu noktalara dikkat edin:
 - a. Kuracağınız modelleri eğitmek ve performansını test etmek üzere, elinizdeki veri kümesini test ve eğitim olarak ayırın.
 - b. En optimal KNN modelini bulmanız gerektiğinden farklı K değerlerine göre farklı KNN modelleri kurun
 - c. Oluşan her model için performans değerlerini karşılaştırarak size göre en iyi olan modeli belirleyin.
 - d. En iyi modeli "**öğrenci.model**" kaydedin.
- 3- Birinci ve ikinci adımda gerçekleştirdiğiniz işlemleri raporlayın. Bu rapor ilgili adımlarda sizden istenen detayları içermelidir. Örneğin:
 - a. İlk adım için çeyreklikler nelerdir, aykırı değer var mı, varsa hangileri, eksik değerler yerine hangi değerler neye göre girildi, en son oluşan veri kümenizin **Öğrenci_no.xml** veya **Öğrenci_no.csv** uzantılı hali (xml veya csv uzantılı dokümanları rapora eklemeyin, belirtilen formatta mail atın)
 - b. İkinci kısımda seçtiğiniz K değeri, bu değer içi oluşan modelin performansını ve bu değerlerin yorumlanmasını raporunuza ekleyiniz. Hangi modeli en iyi model olarak seçtiğinizi, bunun sebeplerini tartışınız.
- 4- WEKA'nın kurulumu, verinin WEKA'da kabul edilen formatlara çevrilmesi, WEKA'da KNN yönteminin uygulanması, sonuçların alınması-yorumlanması, modelin kaydedilmesi vb. adımların tümünü çeşitli kaynaklardan sizin araştırmanız gerekmektedir. Bu nedenle, ancak birkaç farklı çözüm üretip yine de başarılı olamadığınız noktalarda bana danışabilirsiniz.
- 5- Sizden beklenen dokümanlar:
 - a. Öğrenci_no ile başlayan metin editöründe yazılmış raporunuz (doc- docx – pdf – odt vb uzantılarda olabilir)
 - b. Öğrenci_no ile başlayan xml – csv – ods uzantılı verinizin analize hazır halini gösterir şekli
 - c. **Öğrenci_no.model** formatındaki KNN modeliniz