

BİL3003 Veri Madenciliği'ne Giriş – 1. Ödev

Dersin Öğr. Elemanı: Dr. Öğr. Üyesi Mete Eminağaoğlu

Ödev Konusu: Öğrenciler, aşağıda **kısaca açıklanan ödevi** kodlayacak ve raporla birlikte kaynak kodları teslim edecektir. Aşağıda özet bilgiler verilmiştir. Detaylı açıklamalar, sadece bir kere, ders saatinde yapılacaktır. O derse gelemeyen, vb. öğrencilerin ödevi yanlış / eksik vb. anladığı için yetersiz bir ödev ilemesinden sadece o öğrencinin kendisi sorumludur.

Ödevin Son Teslim Tarihi: 15 Kasım 2019, 10:00

Ödevin Teslim Şekli:

CSC ÖBS (Moodle) sistemindeki ders sayfasında açılacak olan ödev yükleme (assignment) alanına; tüm dosyalar, rapor, vb. **zip / rar sıkıştırılmış tek bir dosya olarak yüklenecektir.**

Bu ödev **tek kişiliktir. Her türlü yardım, vb. kopya / intihal olarak değerlendirilecektir ve yardım eden / yardım alan öğrencilerin hepsi bu ödevden 0 (sıfır) alacaktır.**

Ödevin geç teslimi durumları:

Her ne nedenle olursa olsun, geç teslim edilen ödevler ileilmemiş olarak değerlendirilecek ve **0 olarak notlanacaktır.**

Ödev Konusu:

Bu dersin Moodle sistemindeki Ödev-1 kısmında size iletilmiş olan .txt uzantılı veri kullanılarak **Apriori algoritması ile association (birliktelik / ilişkilendirme) kuralları** oluşturulacaktır.

• Veri Seti ile ilgili bilgiler:

- 12 değişkenden oluşan, 1075 kayıttan oluşan, bir ağ (network) sistemindeki veri iletimi sırasında kayıtlanmış ve hesaplanmış çeşitli özellik ve bilgileri içeren bir veri setidir.
- 12 değişkenin adları aşağıdadır:
 - Node
 - Full_Bandwidth
 - Average_Delay_Time_Per_Sec
 - Packet_Size_Byte
 - Packet_Transmitted
 - Packet_Lost
 - Transmitted_Byte
 - AVG-Drop-Rate
 - AVG-Bandwidth-Use
 - Node Status
 - Flood Status
 - Class
- Her değişken , (virgül işareti) ile ayrılmıştır.
- Bazı değişkenler sayısal değişkendir. Diğerleri ise nominal tipte kategorik değişkenlerdir.
- ? (soru işareti) ile gösterilen değerler, kayıp (missing values) değerlerdir.

- **Veri işleme, veri düzenleme, veri madenciliği, Apriori, vb ilişkin her türlü hazır fonksiyon, kütüphane, hazır araç, vb. kullanımı kesinlikle yasaktır.**
- Kodlama kısmında, **sadece aşağıdaki programlama dillerinden istediğiniz birini seçip kullanabilirsiniz: C, C++, C#, .Net, Java, Python.**
- **DİKKAT: Java programlama dilinde kodlayacak öğrenciler Eclipse üzerinde çalıştırmalıdır. Python programlama dilini kullanacaklar da Python 2.7 veya daha üst bir sürümü (version) kullanmak zorundadır.**

Veri setinde, öncelikle bazı çeşitli ve farklı veri düzenleme / düzeltme / temizleme (data preprocessing) işlemlerini yapmanız gerekecektir. Verileri, yazacağınız kodla okutmalı ve bu data preprocessing işlemlerini manuel veya Excel vb ile değil, programınız içinde kodlayarak yapmalısınız.

Bir sonraki aşamada, programınız kullanıcının ekrandan **“Minimum Support”** değerini girmesini isteyecektir. Girilen geçerli bir min. support değerinden sonra, kullanıcı ekrandan önce **Confidence, Lift veya Leverage**

seeneklerinden birisini semesi istenmeli, ve hangi ly (metric) seerse, **buna iliřkin minimum deėeri** girmesi istenmelidir.

Bu bilgiler de ekrandan girildikten sonra, bu bilgiler iřıėında, **kodunuz nce “Frequent Item Sets” oluřturacak**, sonra da **buna gre Apriori algoritması alıřarak kurallar oluřturulacaktır**. Her bir kurala ait **Support deėeri** ve **seilen diėer l (metric) deėeri** yazılacaktır. rnek basit bir ekran grnts ařaėıda verilmiřtir.

Minimum Support deėerini giriniz:

0.85

Confidence / Lift / Leverage llerinden birini seiniz:

Lift

Setiėiniz lnn minimum deėerini giriniz:

4.8

Sonular:

Frequent Item Sets:

...
...
...

Apriori’ye gre oluřan kurallar:

- | | | |
|--|-------------|----------|
| 1. If AND Then AND | Support:1 | Lift:5.5 |
| 2. If AND Then | Support:0.9 | Lift:5.3 |
| 3. If Then | Support:1 | Lift:5.0 |
| 4. ... | | |
| 5. ... | | |

Dikkat: Ekran a gsterilen sonular aynı zamanda, programınız tarafından oluřturulacak bir txt dosyasına da kaydedilecektir.

Ekrandan geersiz bir Support, Confidence, Lift veya Leverage deėeri girilirse, programınız uyarı vermeli ve ana ekrana geri dnmelidir. (Geersiz deėer ne demek? rneėin, -0.5 ya da 2.7 gibi Support deėeri olamaz. Benzer řekilde, -1.8 gibi Confidence, Lift ya da Leverage olamaz. Leverage deėeri 0.75 olamaz, vb.)

dev Raporunun ieriėi:

MS Word ya da benzer basit bir metin belgesi ierisinde ařaėıdaki bilgiler kısa notlar halinde yazılacaktır:

- Veri temizleme / dzenleme amacıyla xxxx deėiřkenini yntemiyle iřlemini yaptım.
- Veri temizleme / dzenleme amacıyla yyyy deėiřkenini yntemiyle iřlemini yaptım.
- Ekrandan girilen xxxxx bilgisine ve deėerlere gre iřlemini yaptım.
-

devde Teslim Edilecekler:

- 1-Programın tm kaynak kodları, baėlantılı ktphane, dizinler, vb.
- 2-Kullanılan yntemler, iřlemler, vb. ile ilgili kısa bilgiler / notlar (kaynak kod iinde de kısa aıklamalar olarak eklenmelidir).
- 3- dev Raporu