**Prova d’Avaluació Continuada 5**

Presentació

Aquesta Prova d'Avaluació Continuada cobreix els mòduls 6 i 8 (Avaluació de Models) del programa de l'assignatura.

Competències

Les competències que es treballen en aquesta prova són:

* Ús i aplicació de les TIC en l'àmbit acadèmic i professional
* Capacitat per innovar i generar noves idees.
* Capacitat per avaluar solucions tecnològiques i elaborar propostes de projectes tenint en compte els recursos, les alternatives disponibles i les condicions de mercat.
* Conèixer les tecnologies de comunicacions actuals i emergents i saber-les aplicar convenientment per a dissenyar i desenvolupar solucions basades en sistemes i tecnologies de la informació.
* Aplicació de les tècniques específiques d'enginyeria del programari a les diferents etapes del cicle de vida d'un projecte.
* Capacitat per aplicar les tècniques específiques de tractament, emmagatzematge i administració de dades.
* Capacitat per proposar i avaluar diferents alternatives tecnològiques per a resoldre un problema concret.

Objectius

La correcta assimilació del Mòdul 6:

En aquesta PAC treballarem la generació, interpretació i avaluació d’un model on generarem regles d'associació amb el programari de pràctiques. No perdrem de vista les fases de preparació de les dades, qualitat del model i extracció inicial de coneixement.

Descripció de la PAC a realitzar

La prova està estructurada en un total de 1 exercici teòric i 3 exercicis teòric/pràctics. Aquests exercicis estan tots relacionats. És necessari fer-los tots per a obtenir una valoració satisfactòria.

Recursos

**Bàsics**

Material docent proporcionat por la UOC.

**Complementaris**

En el wiki del aula trobareu:

* Els descrits en l’anterior PAC
* Fitxer lastfm.csv
* URL's d'ajuda per orientar l'exercici

Criteris de valoració

**Exercicis teòrics**

Tots els exercicis han de ser presentats de forma raonada i clara, especificant tots i cadascun dels passos que s'hagin dut a terme per a la seva resolució. No s'acceptarà cap resposta que no estigui clarament justificada.

**Exercicis pràctics**

Per a totes les PAC és necessari documentar en cada apartat de l'exercici pràctic què s'ha fet i com s'ha fet, quin era l’objectiu i com s’ha desenvolupat.

Format i data d’entrega

El format de lliurament és: usernameestudiant-PACn.doc/docx/odt/pdf

Data de lliurament: 09/05/2018

Cal lliurar la PAC a la bústia de lliuraments de l'aula

|  |
| --- |
| 1. **Nota:** Propietat intel·lectual   Sovint és inevitable, al produir una obra multimèdia, fer ús de recursos creats per terceres persones. És per tant comprensible fer-lo en el marc d'una pràctica dels Estudis, sempre que això es documenti clarament i no suposi plagi en la pràctica.  Per tant, al presentar una pràctica que faci ús de recursos aliens, s'ha de presentar juntament amb ella un document en que es detallin tots ells, especificant el nom de cada recurs, el seu autor, el lloc on es va obtenir i el seu estatus legal: si l'obra està protegida pel copyright o s'acull a alguna altra llicència d'ús (Creative Commons, llicència GNU, GPL ...).  L'estudiant haurà d'assegurar-se que la llicència no impedeix específicament el seu ús en el marc de la pràctica. En cas de no trobar la informació corresponent haurà d'assumir que l'obra està protegida per copyright.  Hauríeu a més, d’adjuntar els fitxers originals quan les obres utilitzades siguin digitals, i el seu codi font si correspon.  Un altre punt a considerar és que qualsevol pràctica que faci ús de recursos protegits pel copyright no podrà en cap cas publicar-se en Mosaic, la revista del Graduat en Multimèdia de la UOC, llevat que els propietaris dels drets intel·lectuals donin la seva autorització explícita |

Exemple

Trobareu a l’arxiu 75.584-PEC5.Rmd un exemple de com aplicar l’algoritme apriori a un data set.

Enunciat

Contextualitzeu els exemples de les següents preguntes respecte al projecte que heu definit a la PAC1. Si ho desitgeu, podeu redefinir o afinar el projecte.

1 Creieu que les regles d'associació són el mètode més adequat per aconseguir els objectius que us havíeu proposat? Justifiqueu la resposta tot raonant-la.

Com podria ser el model resultant?

Doneu un exemple de la interpretació que es podria derivar del model generat

2. En aquest exercici seguireu els passos del cicle de vida d'un projecte de mineria de dades pel cas d'un algorisme de generació de regles d'associació. Ho fareu amb el fitxer Lastfm.csv que trobareu a la wiki. Aquest fitxer conté un conjunt de registres. Aquests registres són l'històric de les cançons que ha escoltat un usuari («user») en un portal Web de musica. «artist» és el nom del grup que ha escoltat, «sex» i «country» es corresponen a variables que descriuen a l'usuari.

Estudieu i compreneu les dades, per exemple: Número de registres dels fitxer? Distribucions de valors per variables? Hi ha camps mal informats o buits? Atenció que hi ha molts registres!!!! Feu un petit resum de nombre d'usuaris, distribució per sexe o país... Quants grups musicals hi han? Enteneu que són propostes que us faig per orientar l'exercici.

Prepareu les dades. Atenció. El primer pas sempre és binaritzar les dades com teniu explicat al material docent. En R i per aquest fitxer jo proposo fer-ho de la següent forma:

### Volem una llista per cada usuari dels grups que escolta. Eliminen sexe i país

mba <- split(x=fitxer\_llegit(,”artist”],f=lastfm$user

### Si hi ha duplicats els eliminen

mba <- laaply(mba,unique)

## I ara binaritzem les dades

mba <- as(mba,”transactions”)

* Instal·leu, si manca, el paquet arules a R. Aquest paquet, documentat a la wiki conté una implementació de l'algorisme apriori vist al Mòdul Docent. .
* Genereu un model de regles d'associació
* Quina és la qualitat del model?
* Quines regles interessants trobeu?
* En funció del model. Quin és el coneixement que trèiem? Veieu alguna possible implementació de model extret?

3. Repetiu l’exercici 2 amb algun altre conjunt de dades. Poden ser dades reals del vostre àmbit laboral o d’algun repositori de dades a Internet. Mireu per exemple: <http://www.ics.uci.edu/~mlearn/MLSummary.html>

Compte amb les dades, heu de tenir present la teoria vista perquè l'exercici tingui sentit.

Seguiu el guió de la pregunta anterior. Els passos per binaritzar poden ser diferents en funció de les dades, no els prengueu com un dogma. De fet a R hi ha paquets per fer-ho.