

# NETFLIX

**Competència Transversal d'Intel·ligència Artificial**  
**Desembre 2014**

Baradad, Manel  
Dalmasso, Marc  
Villalobos, Albert

# Índex

1. Què és Netflix?	p.3
2. Història	p.4
3. Sistemes de recomanació	p.4
4. Netflix Prize	p.5
5. El sistema de recomanació de Netflix	p.6
5. 1 Una base primordial	p.6
5. 2 aprenent a ordenar	p.8
5. 3 Perfeccionant el model	p.8
5. 4 Innovació constant	p.9
6. Riscos i beneficis per a l'empresa i l'usuari	p.10
6. 1. 1 Riscos per a l'empresa	p.10
6. 1. 2 Beneficis per a l'empresa	p.11
6. 2. 1 Riscos per a l'usuari	p.12
6. 2. 2 Beneficis per a l'usuari	p.13
7. El futur de Netflix	p.14
8. Bibliografia	p.16

## 1. Què és Netflix?

Netflix és un proveïdor de continguts audiovisuals “a la carta” o “sota demanda”, els quals s’ofereixen en streaming a través d’internet. Actualment, Netflix té cobertura al seu país d’origen, els Estats Units, que representa amb diferència el seu mercat principal, tot i que darrerament ha decidit expandir-se a Sud-Amèrica i a alguns països del mercat europeu. A dia d’avui, el servei té subscriptors a 40 països, tot i que la intenció d’arribar a nous mercats no ha fet més que començar.



Fig 1: Creixement de subscriptors al servei de 2003 a 2010.

## 2. Història

Netflix va ser fundada l'any 1997 a Scotts Valley, Califòrnia, per Mark Randolph i Reed Hastings, amb intenció de reparar les mancances que ofería el lloguer de pel·lícules tradicional. Tot i tenir seriosos problemes econòmics en els seus inicis, d'on en podríem destacar una venda fallida a Blockbuster per 50 milions de dòlars, el 2003 la companyia registra els primers guanys en els seus 6 anys d'història, i des de llavors, no ha parat de créixer. En l'actualitat disposa de més de 100.000 títols i 50 milions de subscripcions actives a tot el món.

Aquest creixement quasi bé desmesurat va obligar a l'empresa a desmarcar-se en el seu camp, i això ho va fer amb la creació d'un sistema de recomanació de vídeos personalitzat sota el nom de Cinematch.

## 3. Sistemes de recomanació

Els sistemes de recomanació formen part d'un tipus específic de tècnica de filtre d'informació, els quals presenten diferents tipus de temes o ítems d'informació (pel·lícules, música, llibres, notícies, imatges, pàgines web, etc.) que són de l'interès d'un usuari en particular. Generalment, un sistema recomanador compara el perfil de l'usuari amb algunes característiques de referència dels temes, i busca predir el "ràncing" o ponderació que l'usuari li donaria a un document que encara el sistema no ha considerat. Aquestes característiques poden basar-se en la relació o acostament de l'usuari amb el tema o en l'ambient social del mateix usuari.

Netflix, a partir dels inicis del seu creixement exponencial, ha pogut desenvolupar un sistema de recomanació de pel·lícules i sèries per als seus subscriptors. Aquest, s'inicià a partir d'un model molt simple, però amb els anys s'ha anat perfeccionant. Sempre amb la presència d'un dels algoritmes més utilitzats en sistemes recomanadors, com és el Nearest Neighbour, Netflix ha tingut clar que les preferències i gustos van més enllà de l'individu. Si bé el producte no s'encara al consumidor individual, sinó a un de més comunitari, és a dir, la família o la llar en general, els algoritmes de recomanació tenen en compte els gustos de l'entorn de l'individu, permetent així calcular més precisament allò que pot agradar al client.

Així doncs, amb les ganes d'afinar més els resultats, Netflix va acabar creant un dels recomanadors de continguts més complexos del món, tot i que, insatisfets amb això, l'any 2006 és crea el Netflix Prize.

## 4. Netflix Prize

L'octubre del 2006, Netflix, abocant gran part dels seus recursos en l'obtenció d'un sistema de recomanació tan bo com fos possible per l'estat de l'art en aquell moment, decideix crear una competició anomenada el "Netflix Prize". Aquesta, volia premiar amb un milió de dòlars aquell que pogués construir un programa per predir les preferències de l'usuari i que superés el llavors existent sistema de recomanació de Netflix, anomenat Cinematch, per, com a mínim, un 10% en la precisió de la predicció.

Al llarg dels anys, i després de que nombrosos equips de programació decidissin participar al concurs, el 2009 tan sols dos equips (el "BellKor's Pragmatic Chaos" i el "The Ensemble") van ser els únics en aconseguir millores de

10,05% i de 10,10%, respectivament. Tot i això, va ser el "BellKor's Pragmatic Chaos" qui va ser designat guanyador del concurs, ja que presentava un RMSE (root-mean-square error) lleugerament millor.

El constant augment de clients i la presència de Netflix en diferents plataformes, ara no tan sols a l'ordinador, sinó a SmartTV's, tauletes, consoles, etc... ha estat obligant a Netflix a anar refent mica en mica el seu algoritme. Tot i plantejar-se un nou concurs, anys després, aquest no es va acabar realitzant.

## 5. El sistema de recomanació de Netflix

### 5.1 Una base primordial

L'objectiu dels sistemes de recomanació és el de presentar una sèrie d'elements que interessin a l'usuari. Això, generalment s'aconsegueix mitjançant la selecció d'alguns d'aquests elements i ordenant-los en l'ordre de gaudi esperat (o utilitat).

Tot sistema de classificació ha de partir d'una criba inicial. En el cas de Netflix, una base lògica per començar a decidir quin tipus de pel·lícules poden agradar a l'usuari és la popularitat. La raó és clara: de mitjana, un membre té més probabilitats de veure el que la majoria dels altres estan mirant. No obstant això, la ordenació per popularitat és tot el contrari de la definició de personalització, ja que es produiria el mateix ordre de les partides per a cada membre. Per tant, l'objectiu es converteix en trobar una funció de rànkning personalitzat que pugui satisfer millor als membres amb diferents gustos.

Així doncs, amb la popularitat com a punt de partida però no com objectiu, es necessita utilitzar la qualificació teòrica que l'usuari donaria al contingut com un complement al punt popularitat. En definitiva, necessitem trobar un equilibri entre aquests dos indicadors.

Hi ha moltes maneres de construir una funció de classificació. Des de simples mètodes de puntuació a preferències per parelles. Per tenir una primera imatge d'un d'aquests sistemes, començarem amb un enfocament molt simple de puntuació triant com a funció de rànkung una combinació lineal de popularitat i una predicció de qualificació. Això dona una equació de la forma  $F_{\text{rank}}(u, v) = w_1 p(v) + w_2 r(u, v) + b$ , on  $u$  = usuari,  $v$  = l'element a valorar,  $p$  = popularitat i  $r$  = la predicció de la qualificació. Aquesta equació defineix un espai de dues dimensions com la que es representa a continuació.

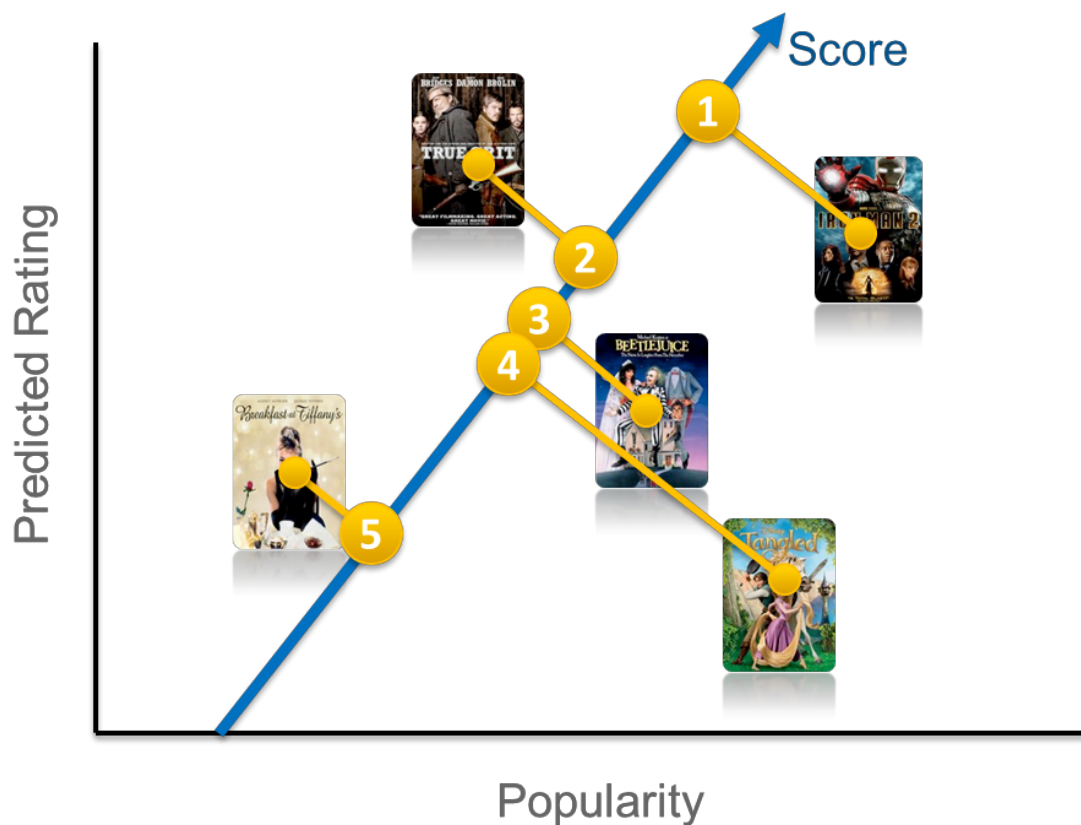


Fig 2: Combinació lineal de popularitat i predicció de valoració

## 5.2 Aprenent a ordenar

Un cop establerta la funció, caldrà determinar els pesos que  $w_1$  i  $w_2$  tindran al model, és a dir, com es determina si la popularitat és més o menys important que la predicció de la valoració. S’han considerat dues possibles aproximacions a aquest problema. La primera, segurament molt descartable i poc eficient, seria la de fer una cerca binària amb diferents pesos. Tot i que es trobaria potencialment una resposta molt precisa en cada cas, l’eficiència seria prohibitiva. Així doncs, com a alternativa ens endinsem en el món del Machine Learning. Aquest, es basa en tenir en compte resultats positius i negatius de dades passades, i permetre a l’algoritme de Machine Learning “aprendre” mica en mica a “saber” ordenar. Actualment usem aquest tipus d’algoritmes en motors de cerca i personalització de publicitat, pel que no és difícil entendre que és perfectament aplicable en aquest cas.

## 5.3 Perfeccionant el model

Certament, en un marc purament teòric n’hauriem de tenir prou amb aquest algoritme que hem descrit. Tanmateix, i com passa en la gran majoria de problemes que es resolen amb intel·ligència artificial, no afrontem un marc teòric, sinó que resolem problemes del món real. Això doncs, implica aplicar desviacions pròpies del problema que es vol resoldre. El gràfic a continuació, mostra com aquestes petites modificacions poden arribar a tenir un valor molt més important que simplement profunditzar en les dues dades amb les que principalment treballem, la popularitat i les valoracions.



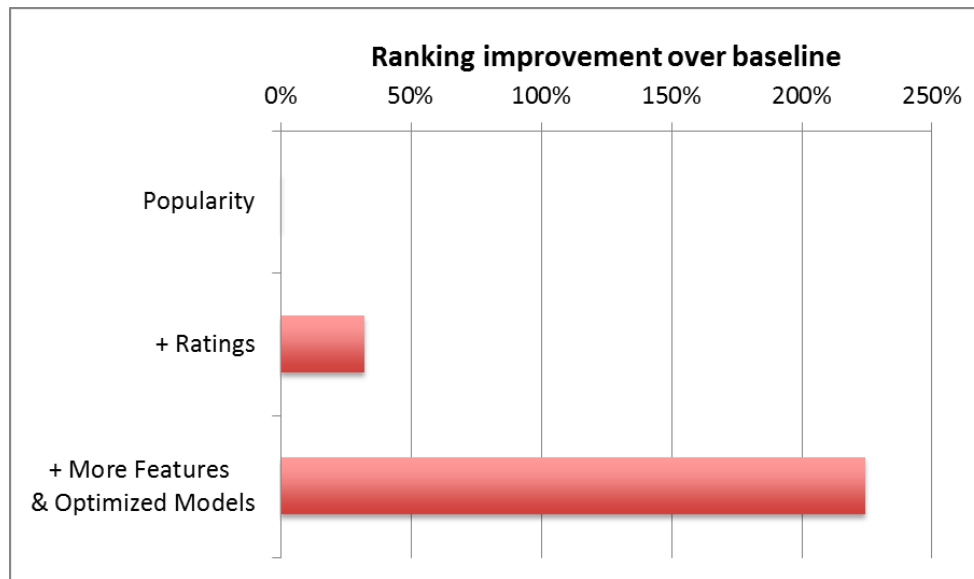


Fig 3: Impacte de les millores en la ordenació de pel·lícules

Malgrat tot, són infinites les proves que es poden fer per a poder millorar aquest rànquing. Tanmateix, sovint ens trobem amb algoritmes similars, com la regressió logística, màquines de vector de suport (SVM), xarxes neuronals, o arbres de decisió, dels quals en deriven moltes tècniques. No podem decidir ni saber amb antelació quines d'aquestes tècniques serviran més o menys. Sovint es decanta per la simplicitat, ja que tindrà un impacte menor en el model que aprengui el Machine Learning, o simplement, que tindrà una menor eficiència, una característica cada cop més preuada.

#### 5.4 Innovació constant

Si bé a Netflix es van veure obligats a reformular la majoria d'algoritmes amb la diversificació de continguts i plataformes, la pressió per oferir un millor servei als consumidors i la simple competitivitat amb altres empreses del sector, la intenció de millorar el sistema és permanent.

Seguint la tècnica del A/B test, el procediment a seguir a l'hora d'introduir nous canvis és tant simple com el següent:

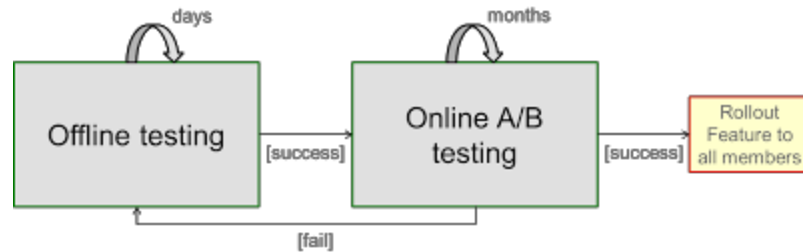


Fig 4: Procediment a l'hora d'introduir canvis al sistema

Com podem veure, es procedeix a realitzar un test offline intern i en el cas de que funcioni comparar online la versió antiga i la versió nova en dos grups d'usuaris diferents.

## 6. Riscos i beneficis per a l'empresa i l'usuari

### 6.1.1 Riscos per a l'empresa

El principal risc que ha d'afrontar una empresa d'aquestes característiques a l'hora de desenvolupar un servei com aquest és el trade-off entre els beneficis que traurà del servei i els recursos que hi invertirà. En el cas de Netflix, al organitzar un concurs amb un premi d'1 milió de dolars va aconseguir que múltiples equips treballessin per desenvolupar un sistema que, tot i probablement no ser final i requerir més inversions, millorava els sistemes dels que es disposaven llavors. Així, podem concloure que el risc econòmic no va ser massa elevat.

Tot i que amb aquest concurs van aconseguir millorar els sistemes que existien fins llavors, tot el codi desenvolupat durant aquest concurs és públic. Així, hi ha un risc inherent en que terceres empreses utilitzin part d'aquest codi per desenvolupar els seus propis sistemes. Tot i així, en tant que actualment Netflix és el principal distribuïdor de continguts d'streaming, i els seus competidors o bé tenen un sistema de recomanacions propi o bé no són suficientment grans com per competir amb aquesta empresa, el fet que el codi pugui ser utilitzat per altres empreses no resulta massa rellevant.

A més a més, és possible que per l'estat de l'art en els sistemes de recomanacions l'experiència de l'usuari no fos suficientment bona, ni invertint-hi diners il·limitats. Tot i així, com es demostra en el cas de Netflix (i en altres sistemes de recomanacions), aquests sistemes han assolit un nivell de maduresa en el que l'experiència d'usuari és suficientment bona per l'usuari mitjà.

També cal tenir en compte els riscos de no implementar un sistema de recomanacions: si la resta d'empreses del sector (com per exemple Amazon, que tot i no ser un clar competidor en serveis de streaming de vídeo, si que competeix indirectament en la venda de productes audiovisuals) si que implementen aquests sistemes, la imatge de Netflix es pot veure perjudicada.

### **6.1.2 Beneficis per a l'empresa**

Els beneficis directes sobre l'empresa al desenvolupar un servei com aquest són clars: com que l'experiència global de l'usuari al utilitzar-los millora, la marca és veurà reforçada en múltiples àmbits.

En primer lloc, l'usuari probablement farà un major consum dels seus continguts. En ser Netflix una empresa que ofereix una tarifa mensual, el benefici de l'empresa no és proporcional al nombre de pel·lícules que consumeix l'usuari. Tanmateix, cal esperar que la fidelitat de l'usuari sigui major com més pel·lícules consumeixi, ja que veurà que la inversió en la tarifa mensual és menor que la que hauria de fer amb un consum igual però pagant pel volum del seu consum.

En segon lloc, la marca Netflix és veurà reforçada: al tenir un sistema de recomanacions bo i del mateix nivell o inclús superior que el de la resta d'empreses del sector, serà vista com una marca sòlida i de qualitat. Si la bondat dels seus sistemes, entre els quals es troba el de recomanacions, no fossin excel·lents, el prestigi de la marca es podria veure afectada.

Tot i que cal tenir en compte que l'usuari difícilment acabaria pagant pel servei de recomanacions en sí mateix, és probable que el valor i que, per tant, augmenti la seva fidelitat envers la marca.

### **6.2.1 Riscos per a l'usuari**

Els principals riscos per l'usuari són a nivell de privacitat: com hem vist, Netflix necessita emmagatzemar dades del consum audiovisual de cada usuari pel bon funcionament del sistema de recomanacions, i aquest fet li pot servir com a justificació per guardar qualsevol tipus de dades. A més de la informació que directament es pot extreure d'aquestes i gràcies a tècniques d'anàlisi de dades, a partir d'aquestes es podrien inferir característiques més privades de l'usuari, com per exemple ideologia política o religió.

Així, tan en aquest cas com en qualsevol altre en la que l'empresa pugui recollir dades dels usuaris, aquests afronten el risc de que aquestes siguin venudes a tercers o utilitzades per la mateixa empresa amb fins no lícits.

Tanmateix, en ser Netflix un servei de pagament i que, per tant, la major part dels seus ingressos difícilment provindran de la venda d'aquestes dades, cal esperar que la seva política en quan a venda de dades sigui la de no fer-ho.

També cal tenir en consideració que, en casos extrems, l'usuari pot acabar fent un ús per sobre del normal del sistema, en part per culpa d'aquest sistema de recomanacions. Tot i així, el nombre de casos en que això es produeixi únicament per culpa dels sistemes de recomanació és probablement, bastant reduït.

### **6.2.2 Beneficis per a l'usuari**

Els beneficis per l'usuari també són clars: l'accés als continguts del seu interès resulta més fàcil, ràpida i eficient. Així, l'usuari tindrà una millor experiència en utilitzar el servei: serà eficient a l'hora de decidir que val la pena veure i que no, l'accés al que vol veure és més ràpid i la seva experiència final serà més fluida i entenedora.

A més, l'usuari també tindrà un benefici econòmic amb la implantació d'aquest sistema: al augmentar el consum i ser aquest més eficient, la inversió en la tarifa mensual és veu compensada amb la que hauria de fer per un mateix consum i adquirint els productes audiovisuals per separat.

## 7. El futur de Netflix

Aquesta imatge ens mostra la fluctuació de les accions de Netflix des de que va sortir a borsa:



Fig 5: Preu de les accions de Netflix al Nasdaq

Dues coses són certes que podem provar amb aquesta imatge: el futur de l'entreteniment domèstic passa pel consum de continguts audiovisuals “a la carta”, i Netflix lidera aquest canvi. Els canals de televisió potser han tardat a reinventar-se, i es queden enrere davant companyies com Netflix, la qual ha començat no només a distribuir sinó també a produir contingut.

Per altra banda, en un món on el tractament del Big Data encapçala les preocupacions de les grans empreses, la necessitat de Netflix de treballar-hi des del principi, encara que fos per necessitat, li donarà segurament un avantatge respecte altres grans empreses. Productes com el propi Cinematch són eines molt valuoses, que no es limitaran a conèixer només els gustos fílmics de la població, sinó que poden adaptar-se a altres béns de consum.

Amb tot això, podem concloure que, tot i no ser la base principal del negoci, el sistema de recomanacions és una de les parts fonamentals en com l'usuari interacciona amb els sistemes d'aquesta empresa. Per tant, la bondat del sistema de recomanacions comporta diferents implicacions empresarials (com per exemple el nombre total de visualitzacions per usuari o la fidelitat d'aquest amb l'empresa), i totes aquestes implicacions acaben contribuint al bon funcionament de Netflix .

## 8. Bibliografia

### Articles i pàgines web

<http://techblog.netflix.com/2012/06/netflix-recommendations-beyond-5-stars.html>

[http://en.wikipedia.org/wiki/Netflix\\_Prize](http://en.wikipedia.org/wiki/Netflix_Prize)

<http://techblog.netflix.com/2013/03/system-architectures-for.html>

[http://www.netflixprize.com/assets/GrandPrize2009\\_BPC\\_BellKor.pdf](http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf)

<http://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf>

### Imatges i gràfics

Figura 1:

[http://ericsadblog.files.wordpress.com/2011/02/02fea\\_netflix\\_subscribers\\_a\\_l\\_hollywoodreporter\\_sourcenetflix.jpg](http://ericsadblog.files.wordpress.com/2011/02/02fea_netflix_subscribers_a_l_hollywoodreporter_sourcenetflix.jpg)

Figura 2:

[http://3.bp.blogspot.com/-u1wrwtggj7o/T-IVRAydO7I/AAAAAAAAAR4/ot\\_aHu0FLjc/s1600/TwoDimensionalRanking-final.png](http://3.bp.blogspot.com/-u1wrwtggj7o/T-IVRAydO7I/AAAAAAAAAR4/ot_aHu0FLjc/s1600/TwoDimensionalRanking-final.png)

Figura 3:

[http://2.bp.blogspot.com/-rSdjs1ipQGU/T-Iz\\_lFVcbI/AAAAAAAAAS0/xKc5jkYtKYg/s1600/Ranking-FeaturesPerformance.png](http://2.bp.blogspot.com/-rSdjs1ipQGU/T-Iz_lFVcbI/AAAAAAAAAS0/xKc5jkYtKYg/s1600/Ranking-FeaturesPerformance.png)

Figura 4:

[http://4.bp.blogspot.com/-o\\_xmCVGr66I/T-I1OoKU9oI/AAAAAAAAATM/UIamph87TJ4/s400/Offline-online-simple.png](http://4.bp.blogspot.com/-o_xmCVGr66I/T-I1OoKU9oI/AAAAAAAAATM/UIamph87TJ4/s400/Offline-online-simple.png)

Figura 5:

<https://uk.finance.yahoo.com/echarts?s=NFLX#symbol=NFLX;range=my>