

Svojo nalogo bom začela s primerjavo metodologije, ki jo uporabljata Amazon Prime in Netflix, ko se lotevata vprašanja izbire nove serije, ki jo bodo predstavili svojim uporabnikom. Amazon iz nabora različnih idej za novo serijo, preko ocenjevanj izbire osem različnih idej, njihovi zaposleni nato napišejo, posnamejo, objavijo prvo epizodo vsake od serij na internet, zastonj, dostopno komurkoli. Uporabniki interneta priložnost seveda izrabijo za 'brezplačen' ogled serije in tako priskrbijo Amazonu podatke ter pomagajo pri odločitvi za nastanek novih serij. Včasih način ni najbolj uspešen. Pri Netflixu na drugi strani pa pa si prav tako pomagajo s podatki, a z drugačno metodologijo; uporabijo že obstoječe podatke o svojih uporabnikih in gledalcih. Ti podatki jim povejo skoraj vse; kaj kdo najraje gleda, kdaj gledajo serije in filme, kako jih ocenjujejo, s kom delijo, hkrati imajo pa tudi cel kup meta podatkov; torej podatkov, kot so kliki miške, premikanje miške, koliko časa izbirajo serijo ali film, kolikokrat si premislijo in tako dalje. Izkazalo se je, glede na povprečno oceno Netflixovih serij in filmov, da je sledeča metodologija uspešnejša. Pri Netflixu gre torej za nenehno zbiranje ogromnih količin podatkov, veliko tudi takih, za katere najbrž večina uporabnikov sploh ne ve. Vse to jim pomaga pri razvijanju algoritmov in modelov, s katerimi izbirajo nove vsebine in pa upravljajo personalizacijo platforme za vsakega upravnika.

Netflix je sicer doživel že veliko sprememb in razvojnih posodobitev, velja pa za, tehnološko eno močnejših platform, prav zaradi svoje usmeritve k podatkom, s katerim so ustvarili eno najbolj uporabniku prijaznih platform. Netflix v svoji prvotni obliki seveda ni bil tak, kakršen ga poznamo danes. Sprva (zgodnja 2000) je šlo za neke vrste izmenjavo oziroma izposojno DVD-jev, šele leta 2007 so predstavili t.i. »streaming service«, torej platformo, kjer lahko uporabnik izbira med naborom video vsebin (večinoma televizijskih serij in filmov) in si le-te tudi ogleda v istem trenutku. Premik iz DVD-jev do predvajanja vsebin je seveda prinesel tudi nove podatke in nove razsežnosti zbiranja teh podatkov. Približno dve leti kasneje je temu doprineslo še dejstvo, da ne gre več samo za eno spletno stran, temveč je Netflix postal dostopen na skoraj vsaki napravi, leta 2012 je bil Netflix dostopen v 47 državah sveta in je imel kot 23 milijonov naročnikov (danes še več; 200 milijonov! AHA moment; bolj »jaw dropping« moment). Zaradi tega porasta naročnikov in sprememb v delovanju, je seveda prišlo tudi do spremembe v metodologiji in načinu zbiranja podatkov. Netflix temelji na personalizaciji platforme za čisto vsakega uporabnika. To vključuje vključitev raznolikosti okusa vsakega gledalca, podobnosti z njegovim/ njenim socialnim krogom, ažurnost in dostop do najnovejših vsebin in pa usklajenost in podobnost med predlaganimi vsebinami.

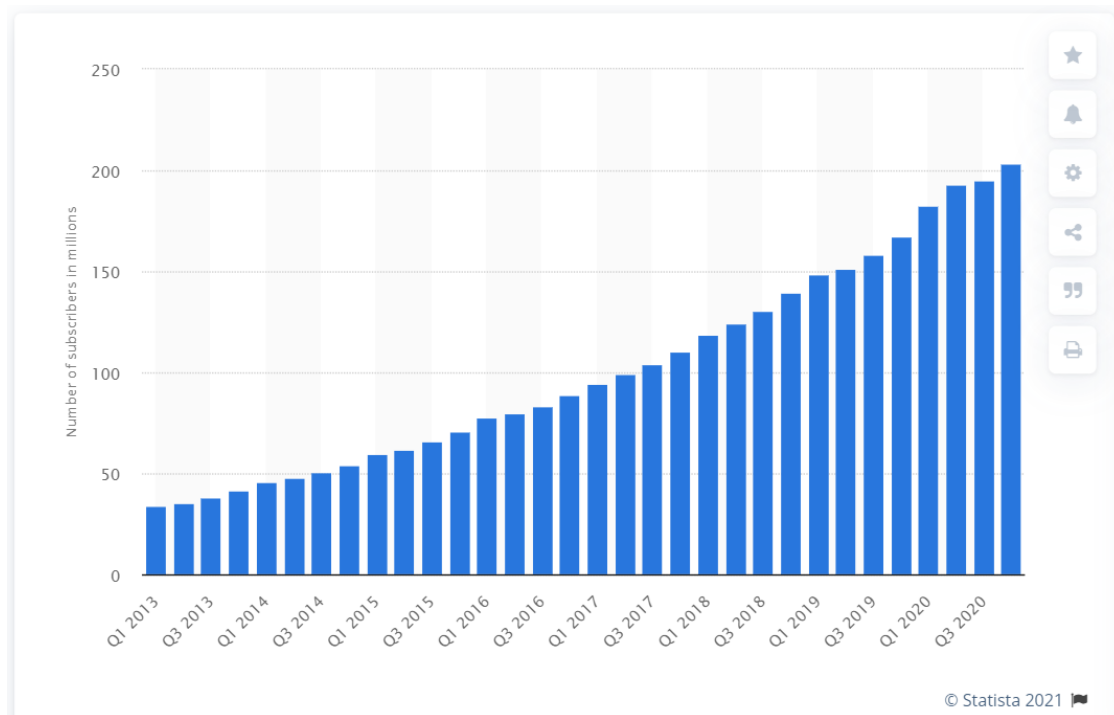
Sprva je velik del personalizacije temeljil na sistemu ocenjevanja oz. priporočanja (»recommendation system«) V zadnji četrtini leta 2011 so uporabniki Netflixa namreč generirali preko 4 milijone priporočil in tako je 75 % ogledanih vsebin tako ali drugače priporočenih. Prav zaradi pomembnosti tega sistema so pri Netflixu želeli izboljšati algoritme za napoved ocen vsebin. To so dosegli s tem, da so leta 2006 organizirali t.i. »Netflix Prize«; tekmovanje z nagrado vredno 1 milijon dolarjev, za tistega, ki bi njihov obstoječi sistem napovedovanja ocen filmov Cinematch z uporabo strojnega učenja in podatkovnega rudarjenja, izboljšal za 10 % (pri tem so za kriterij vzeli koren standardne napake ocene). To je uspelo ekipi, ki je s pomočjo dveh algoritmov (SVD; Singular Value Decomposition in RBM; Restricted Boltzmann Machines), zmanjšala napako na 0.88.

Netflix je za potrebe tega tekmovanja sodelujočim posredoval anonimizirane podatke o uporabnikih, na podlagi katerih so tekmovalci razvijali algoritme. Podatki o uporabnikih so bili, sicer anonimni, a zelo podrobni. Dva sodelujoča raziskovalca iz Univerze v Teksasu sta namreč uspela na podlagi ocen in objav na IMDB-ju identificirati kar nekaj Netflixovih uporabnikov (vključno s političnimi prepričanji in seksualno usmeritvijo). Prav zaradi tega je imel Netflix kar nekaj težav in tožb. Ena od njih je bila tožba ameriške homoseksualne (ne razkrite) mame, ki je Netflix tožila zaradi (skorajšnjega) razkritja njene identitete, predvsem razkritja njene spolne usmerjenosti, kar bi 'negativno vplivalo na njeno zmožnost podpiranja družine in na njeno življenje in posledično življenje njenih otrok'.

Paul Ohm, profesor prava iz Univerze v Coloradu, je v kontekstu tožb pod vprašaj postavil prakso anonimiziranja podatkov, ter zaključil, da so lahko, če so podatki koristni in informativni za raziskovalce, ti tudi re-identificirani. Na primer, če podatki vključujejo posameznikovo pošto številko, rojstni datum

in spol, obstaja 87 % verjetnost, da je lahko ta posameznik uspešno identificiran (spet; AHA moment). Taisti profesor ne podpira prej omenjene tožbe, saj naj bi Netflix naredil dovolj za to, da bi bile identitete uporabnikov prikrite (s procesom petrubitacije).

Večina podanih virov se je nanašala na zgodnejša obdobja Netflix, a je Netflix od takrat še naprej uspešno oblikoval svoje algoritme, se širil po svetu in nabiral nove uporabnike. Kot že omenjeno, je imel Netflix konec leta 2020 kar 200 milijonov naročnikov, dostopen pa je že v več kot 190 državah. Raste njihov letni prihodek, uspevajo pa tudi v primerjavo s konkurenco; na Oscarjih leta 2020 so bila dela Netflix nominirana kar 24-krat, več kot katerikoli drugi studio.



Slika 1: <https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming-subscribers-worldwide/>

Pri pripravi te naloge, se moje večje ugotovitve ne tičejo toliko samega načina zbiranja podatkov s strani Netflix, vsekakor so me šokirale številke, ki prikazujejo uspeh in rast Netflix tekom let, predvsem zaradi članka, ki je bil napisan leta 2011 in opisuje metodologijo Netflix, ki mi je dal misliti, če so lastniki in razvijalci takrat pričakovali tak uspeh, ali so slučajno pričakovali večji uspeh? Prav tako me zanima, kakšne metode in načini zbiranja podatkov so bili predstavljeni kasneje, predvidevam namreč, da nenehno stremijo k izboljšavam. Kot že omenjeno je bil AHA moment tudi del o profesorju prava Ohmu in njegovo stališče o praksi anonimizacije podatkov.