Er komt veel kijken bij het maken van een recommendation engine, voornamelijk erg veel data. Gezien deze data te groot is en vaak onnodig om een goeie aanbeveling te maken, is het van belang om alleen de belangrijke informatie eruit te filteren en te gebruiken. Hieronder vertellen wij voor welke data er is gekozen met een korte redenering erbij.

ID: Het id van de bijbehorende producten moeten bekend zijn om deze makkelijk te kunnen vinden en aanteraden.

Brand: Hiervoor is voor verschillende redenen gekozen. Een merk heeft vaak een specifieke doelgroep, zoals Kruidvat zich bijvoorbeeld focust op wat goedkopere algemene producten. Iemand die kiest voor een goedkopere Kruidvat shampoo zal ook eerder kiezen voor een goedkopere huid crème van Kruidvat vergeleken met een Nivea crème. Veel merken hebben ook producten die elkaar aanvullen, zoals conditioner en shampoo of een koptelefoon en een beschermhoes voor de koptelefoon. Als laatste hebben mensen ook vaak een positieve associatie met een bepaald merk.

Category: Dit is van belang omdat er anders aanbevelingen kunnen ontstaan die totaal niet relevant zijn. Indien een klant kijkt naar speelgoedauto's voor kinderen moeten er geen aanbevelingen zijn voor autobanden of nieuwe stuurknuppels van echte auto's.

Gender: Veel producten trekken eerder een geslacht aan of zijn gemaakt met een specifiek geslacht in gedachten. Weinig mannen dragen make up en rokken, en weinig vrouwen willen baardtrimmers (hoewel dit aantal steeds meer groeit tegenwoordig). Door rekening te houden met het geslacht kunnen er betere aanbevelingen gemaakt worden.

Sub\_category: Hetzelfde met category, nog specifiekere aanbevelingen maken die dicht bij het orginele product liggen. Als een klant opzoek is naar een shampoo kunnen dit bijvoorbeeld shampoos zijn voor baby’s, dit betekent dat een klant daarbij geen shampoo’s voor gekleurd haar wil zien maar speciefieke baby shampoo’s.

Sub\_sub\_category: Hoewel dit overbodig kan lijken is deze informatie soms nog aardig toepasselijk. Denk hierbij aan Iphone hoesjes, deze vallen onder de category telefoonhoesjes, vervolgens onder iphone hoesjes maar een klant wil niet hoesjes zien van 23 Iphone modellen. De klant wil de hoesjes van zijn/haar/het/weet ik veel telefoon model dus de sub\_sub\_category iphone 12 hoesjes zou erg handig kunnen zijn in dit geval om een goeie aanbeveling te kunnen maken.

Viewed before: Het is verstandig om te onthouden wat klanten in het verleden hebben bekeken omdat je dan al gelijk een goed overzicht krijgt van hun interesses, en ze dan kunnen worden herinnerd aan mogelijke aankopen die zij in het verleden hebben uitgesteld. Dit werkt goed voor bedrijven zoals bol.com.

Sessions en Profile Buid: Dit is een van de belangrijkste data. Hiermee wordt een klant met een profiel aan eerdere sessions gekoppeld en onstaat er een goed beeld van de interesses van de klant en ook voor een groot deel klantgedrag. Denk hierbij aan hoelang een klant gemiddeld op de webshop zit in een sessie en welke paginas worden bezocht.

Profile\_id: Deze is achteraf gezien minder belangrijk dan gedacht.

Sessions\_id: Deze data bleek achteraf ook niet van belang.

Profile\_preferences: Hierin kan worden gevonden wat de klant met het profiel eerder heeft bekeken, dit is een grote bouwsteen voor de reccomendation engine. De voorkeur van een klant voor bijvoorbeeld een bepaald tv merk wordt snel gezien en kan daardoor snel worden gematcht met specifieke accesoires die meer trekken.

Session\_recommendations: Alle eerder bekeken producten in de sessie en vergelijkbare producten worden opgeslagen in dit stukje data. Dit koppelt goed met Profile\_preferences en de buid.

Psuedo code voor algoritme:

* Alle benodigde data uit MongoDB halen.
* De opgehaalde data filteren en de relevante producten en informatie bewaren.
* De data ter controle opslaan in een csv bestand.
* De informatie opslaan in de sql database.