



Proef ingediend met het oog op het behalen van de graad van
Master of Science in Toegepaste Economische Wetenschappen

Verwachtingen over het onverwachte: een gebruikersgerichte evaluatie van muzikeraanbevelingssystemen

Daan De Witte

0566525

Academiejaar 2023-2024

Promotor: Nanouk VERHULST

Sociale Wetenschappen & Solvay Business School

VERKLARING VAN AUTHENTICITEIT EN OPENBAARMAKING

De ondertekende verklaring van authenticiteit en openbaarmaking is een integrale component van het geschreven werk dat wordt ingediend door de student.

Met mijn handtekening verklaar ik dat:

- ik de enige auteur ben van het ingesloten geschreven werk¹;
- ik dit werk in eigen woorden heb geschreven;
- ik geen plagiaat heb gepleegd zoals gedefinieerd in artikel 118 van het Onderwijs- en Examenreglement van de VUB; waarbij de meest voorkomende vormen van plagiaat zijn (niet-limitatieve lijst):
 - aard 1: tekst overnemen van andere auteurs, weliswaar met bronvermelding maar zonder gebruik van aanhalingstekens waar het om een letterlijke overname gaat;
 - aard 2: tekstfragmenten overnemen van andere auteurs, al dan niet letterlijk, zonder bronvermelding;
 - aard 3: verwijzen naar primair bronmateriaal waar de tekst en bronvermelding al dan niet letterlijk wordt overgenomen uit niet-vermelde secundaire bronnen;
 - aard 4: tekstfragmenten overnemen van andere auteurs, al dan niet met bronvermelding, met geringe en/of misleidende tekstaanpassingen;
 - aard 5: eigen werk overnemen zonder bronvermelding (zelfplagiat);
 - aard 6: tekstfragmenten genereren of substantieel aanpassen met softwaretools, zonder de uitdrukkelijke vermelding hiervan.
- ik in de tekst en in de referentielijst volledig heb gerefereerd naar alle internetbronnen, gepubliceerde of ongepubliceerde teksten die ik heb gebruikt of waaruit ik heb geciteerd;
- ik duidelijk alle tekst heb aangeduid die letterlijk is geciteerd;
- ik alle methoden, data en procedures waarheidsgetrouw heb gedocumenteerd;
- ik geen data heb gemanipuleerd;
- ik alle personen en organisaties heb vermeld die dit werk hebben gefaciliteerd, dus alle ingediende werk ter evaluatie is mijn eigen werk dat zonder hulp werd uitgevoerd tenzij uitdrukkelijk anders vermeld;
- dit werk noch een deel van dit werk werd ingediend aan een andere instelling, universiteit of programma;
- ik op de hoogte ben dat dit werk zal gescreend worden op plagiaat;
- ik alle origineel onderzoeksmateriaal onmiddellijk zal indienen op het Decanaat wanneer hierom wordt gevraagd;
- ik op de hoogte ben dat het mijn verantwoordelijkheid is om na te gaan dat ik word opgeroepen voor een hoorzitting en tijdens de periode van hoorzittingen beschikbaar te zijn;
- ik kennis genomen heb van artikel 118 van het Onderwijs- en Examenreglement van de VUB omtrent onregelmatigheden en dat ik op de hoogte ben van de disciplinaire sancties;
- (enkel voor masterproeven) ik akkoord ga met de openbaarmaking van mijn masterproef. Indien ik niet akkoord ga dien ik samen met mijn masterproef het formulier 'openbaarmaking masterproef' in (terug te vinden op de VUB website).

Student familienaam, voornaam: De Witte Daan

Datum: 31/07/2024

Handtekening: 

¹ Voor groepswerken zijn de namen van alle auteurs verplicht. Hun handtekeningen staan collectief borg voor de volledige inhoud van het geschreven werk.

Dankwoord

Het schrijven van deze masterproef was een uitdagend proces, daarom wil ik graag dit moment nemen om enkele mensen te bedanken die hierin een belangrijke rol hebben gespeeld. In de eerste plaats wil ik mijn promotor Nanouk Verhulst bedanken om de tijd te nemen om op een uitgebreide en kritische manier mijn voorlopige werk te beoordelen. Het was fijn om dit proces aan te gaan met een promotor die betrokken was. Daarnaast wil ook alle respondenten bedanken die mijn enquête hebben ingevuld. Ook wil ik mijn goeie vriend Max Van Den Bosch bedanken voor de fijne momenten in het 2^{de} semester waar wij samen op dagelijkse basis in de bib aan onze master proef hebben gewerkt. Ondanks dat onze onderzoeksdomeinen helemaal verschilden heb ik veel gehad aan onze gesprekken over de literatuur. Daarnaast was je met consistentie en drang naar excellentie een inspiratiebron. Dit motiveerde mij om alles uit de kast te halen om deze finale proef tot een goed einde te brengen. Ten slotte wil ik mijn lieve moeder bedanken voor de onvoorwaardelijke steun tijdens het schrijven en om me te motiveren om überhaupt aan deze masteropleiding te beginnen. Iets wat achteraf gezien aanvoelt als een heel goeie en leerrijke beslissing.

Abstract

Recommender systems (RS) worden dagelijks toegepast in diverse sectoren, zoals e-commerce, video en muziek. Dit heeft interesse gewekt naar het optimaliseren van RS met als doel het verhogen van satisfactie en de WTP bij de gebruiker. Ondanks het feit dat de literatuur benadrukt dat de evaluatie van RS het best gebeurt vanuit het standpunt van de gebruiker is dit in de context van muziekstreaming nog te weinig onderzocht geweest. Deze studie onderzocht wat de invloed is van gepercipieerde accuracy en serendipity op WTP voor een muziekstreamingdienst met satisfactie als mediator. Ook werd onderzocht of de mate van gebruik (lichte gebruiker vs zware gebruiker) invloed heeft op dit verband als moderator. Deze studie werd uitgevoerd aan de hand van een 2 X 2 between subject scenario experiment met 165 respondenten in België. De resultaten tonen dat gepercipieerde serendipity en accuracy via satisfactie een positieve invloed hebben op WTP. De invloed van serendipity is het grootst. Daarnaast blijkt ook uit dit onderzoek dat zware gebruikers accuracy niet meer belangrijk vinden. Hieruit concluderen we dat een evaluatie vanuit het standpunt van de gebruiker met inbegrip van serendipity noodzakelijk is om de WTP te verklaren. Muziekstreamingdiensten in de praktijk kunnen inspelen op de nood voor serendipity en kunnen hun aanbeveling differentiëren op basis van de mate van gebruik. Dit zal resulteren in een hogere WTP en satisfactie. In verder academisch onderzoek kan het onderzoek hernomen worden in een realistische setting en met andere mogelijke moderatoren zoals transparantie van het RS.

Inhoudstafel

Dankwoord.....	1
Abstract.....	2
Inleiding.....	5
Literatuurstudie	8
Recommender systeem	8
Evaluatie van Recommender Systems.....	9
Systeem gefocust analyse.....	9
Gebruiker gefocuste evaluatie	12
Perceived qualities	14
Gepercipieerde serendipity.....	15
Gepercipieerde Accuracy	17
Consumer outcomes	17
Customer satisfactie.....	18
Willingness to pay	20
Mate van gebruik.....	21
Methodologie	23
Onderzoeksdesign en procedure.....	23
Introductie vragen.....	23
Scenario's.....	24
Items	26
Resultaten.....	30
Realisme en Manipulatie check.....	30
Hypotheses testen	31
Directe invloed op WTP.....	31
Satisfactie	32
Discussie.....	36
Limitaties en toekomstig onderzoek.....	36

Praktische implicaties	37
Limitaties en toekomstig onderzoek.....	40
Referenties	44
Bijlagen.....	55
Bijlage A: Geïnformeerde Toestemming	55
Bijlage B: Demografische vragen	57
Bijlage C: Mogelijke condities.....	58
Basis conditie :	58
Conditie 1: lage serendipity en lage accuracy	58
Conditie 2: hoge serendipity en lage accuracy	59
Conditie 3: Lage serendipity en lage accuracy.....	59
Conditie 4: hoge serendipity en hoge accuracy	59
Bijlage D: Vragenlijst Theoretisch kader.....	61
Mate van gebruik	61
WTP	62
Bijlage E: Realisme en Manipulatie check	63
Realisme.....	63
Manipulatiecheck.....	64
Bijlage F: Bonus vragen.....	65
Appendix G: AI Tools.....	67

Inleiding

Recommender systems (RS) zijn booming. Er gaat geen dag voorbij waarop de consument er geen gebruik van maakt. Van het scrollen op social media tot het kopen van een boek op Bol.com, er komt een RS aan te pas. RS gebruiken data van gebruikers zoals voorkeuren en surfgedrag om bijvoorbeeld producten, nieuwsartikels, ... aan te bevelen aan specifieke consumenten (Lü et al., 2012). Hierdoor zijn RS alomtegenwoordig in verschillende sectoren zoals e-commerce, video, muziek, nieuws, online zoekmachines en gaming (Schedl et al., 2015) Meer specifiek, muziekstreamingplatformen gebruiken RS om nummers aan te bevelen aan hun gebruikers. Denk bijvoorbeeld aan de muziekstreamingdienst Spotify waar gebruikers kunnen kiezen tussen meer dan 100 miljoen nummers (Spotify, 2023). Om de gebruiker te ondersteunen bij het ontdekken van deze enorme catalogus muziek, maken muziekstreamplatformen aanbevelingen (Scheibehenne et al., 2010). Dit gebeurt aan de hand van de luistergeschiedenis van de luisteraar gecombineerd met zijn aangegeven voorkeuren (Pu et al., 2012).

De muziekstreamingmarkt is een heel competitieve markt. Er zijn een tal van spelers die allemaal dezelfde muziek aanbieden. Hierdoor zijn er weinig mogelijkheden tot een unique selling proposition (USP) (Colbjørnsen et al., 2022). Dit zorgt voor een sterke druk om een lage prijs te hanteren. Een maandabonnement kost rond de 10,99 euro (Spotify, 2024). Deze lage prijs zorgt ervoor dat de winstmarges laag tot onbestaande zijn (Wlömert & Papiés, 2016). Zelf de grootste streaming dienst ter wereld staat onder druk. Spotify maakte in 2022 ondanks een omzet van 11,7 miljard euro een verlies van 113 miljoen euro (Spotify 2022). Ondanks de lage marges is er nog maar weinig academisch onderzoek gedaan naar het zoeken van een (meer) optimale prijs (De Toni et al., 2017). Om deze moeilijke markt toch te kunnen penetreren is het belangrijk om de noden van de (potentiële gebruiker) goed te begrijpen. Voorgaand onderzoek toonde reeds aan dat de aanbevelingen invloed hebben op de evaluatie

van de aangeboden dienst (Kim et al., 2017). Het is belangrijk om te begrijpen welk aanbevelings-algoritme zorgt voor de grootste willingness to pay (WTP) en satisfactie bij de gebruiker. Veel voorafgaand academisch onderzoek gebruikte om RS te evalueren maatstaven gebaseerd op een evaluatie door het systeem zelf. Deze maatstaven waren vaak gebaseerd op de accuracy van de aanbevelingen. Hierbij is het de bedoeling om een aanbeveling te geven die zo nauwkeurig mogelijk en op verschillende dimensies overeenkomt met de ingegeven informatie (Schedl et al., 2018) maar, dit komt in de realiteit niet altijd overeen met wat consument werkelijk verkiest (Vargas et al., 2014). De consument is naast accurate aanbevelingen ook opzoek naar verrassende muziekaanbevelingen (Jannach & Bauer, 2020). Om te begrijpen wat de consument daadwerkelijk belangrijk vindt aan een RS, zou de evaluatie vanuit het standpunt van de gebruiker moeten gebeuren. Ondanks het feit dat de nood hiervan al vaak is aangehaald in de literatuur, is dit nog niet genoeg onderzocht geweest (Schedl et al., 2015). Ook is de link tussen de evaluatiecriteria van RS en WTP nog niet gelegd in de academische literatuur. De uitdaging van dit onderzoek is om deze leegte in de literatuur verder te onderzoeken. De evaluatiecriteria die in dit onderzoek gebruikt worden gebruikt zijn accuracy en serendipity. Doordat deze inhoudelijk ver uit elkaar liggen is het interessant om te weten wat de consument het “belangrijkste” vindt. Aanbevelingen die gepercipieerd worden door de gebruiker als heel accuraat of eerder aanbevelingen die gepercipieerd worden als een positieve verassing. Mogelijks speelt er ook een interactie-effect tussen beiden. Een conclusie hierover kan de ontwerpers van RS van muziekstreamingdiensten helpen om deze te optimaliseren met mogelijks hogere WTP en satisfactie als gevolg.

Samengevat: huidig onderzoek geeft de nood aan om beter te begrijpen wat gebruikers van streamingplatformen belangrijk vinden. Maar dit is toe nu toe nog te weinig onderzocht vanuit het standpunt van gebruiker in de context van muziek streaming.

Dit onderzoek gebeurt doormiddel van een scenario based experiment. De onderzoeksvraag luid als volgt: wat is de invloed van van de gepercipieerde accuracy en serendipity op de WTP (via satisfactie) van de gebruikers van streaming platformen. Bijkomend wordt ook onderzocht of de mate van gebruik (zware versus lichte gebruiker) invloed heeft als moderator op de sterkte van dit verband.

In het volgende hoofdstuk wordt een uitgebreid overzicht gegeven van de bestaande literatuur over RS en de evaluatie ervan. Vervolgens wordt in het hoofdstuk methodologie gedetailleerd beschreven op welke manier het onderzoek is uitgevoerd. Daarna worden de statistische resultaten besproken. In het discussiegedeelte worden de implicaties van de onderzoeksresultaten voor de praktijk en het academische werkveld besproken. Tot slot wordt er aandacht besteed aan de beperkingen van het onderzoek.

Literatuurstudie

Recommender systeem

Een RS maakt gebruik van gebruikersdata, zoals voorkeuren en surfgedrag, om gebruikers gepersonaliseerde aanbevelingen voor producten, nieuwsartikelen, muziek, films en andere items aan te bieden (Lü et al., 2012; Schedl et al., 2015).

RS maakt aanbevelingen op basis van algoritmen (Ekstrand et al., 2014). Algoritmen zijn een sequentie van stappen om een probleem op te lossen (Miyazaki, 2013). In de praktijk zijn er 3 soorten algoritmes die vaak gebruikt worden in de context van muziekaanbevelingen namelijk content based filtering, collaborative filtering en hybridization (Schedl et al., 2018). Collaboratieve filtering (CF) identificeert gelijkaardige nummers of gebruikers om aanbevelingen te doen (Herlocker et al., 2004; Schafer et al., 2007). Bij de benadering vanuit de gebruiker worden gebruikers met een gelijkaardige muzieksmaak geïdentificeerd. Als twee gebruikers vaak naar dezelfde nummers luisteren, worden zij als gelijkaardig beschouwd. Indien er een nummer is dat één van hen kent, maar de ander niet, kan dit nummer aan diegene die het niet kent worden aanbevolen. Bij de benadering vanuit de nummers worden nummers als gelijkaardig beschouwd wanneer ze door dezelfde gebruikers als “goed” zijn beoordeeld. Aan een gebruiker die een bepaald nummer al kent kan een gelijkaardig nummer aanbevolen worden (Sunitha & Adilakshmi, 2018). Het tweede veelvoorkomende algoritme is content-based filtering (CB). Deze filtering gaat de nummers met elkaar vergelijken. Indien twee nummers als gelijkaardig worden gezien, dan zal een gebruiker die het ene nummer al kent het andere nummer aanbevolen krijgen (Schedl et al., 2018). Het vergelijken van nummers gebeurt bij CB op basis van twee mogelijke systemen (Barrington et al., 2009). Enerzijds heb je de content based systems. Deze gaan nummers vergelijken op basis van akoestische eigenschappen, zoals het tempo of de frequentieverdeling. In dit systeem zijn nummers die gelijkaardig zijn: nummers die gelijkaardig klinken (Aucouturier & Pachet,

2002; Flexer et al., 2008). Anderzijds heb je metadata-based systemen. Het vergelijken van nummers gebeurt hier aan de hand van metadata zoals bijvoorbeeld genre, artiestennaam, regio. In dit systeem zijn nummers die gelijkaardig zijn nummers die bijvoorbeeld in hetzelfde genre zitten of van dezelfde artiest zijn (Barrington et al., 2007; Xiao et al., 2009). Bij het derde algoritme hybridization worden CB en CF gecombineerd om aanbevelingen te doen (Donaldson, 2007; Yoshii et al., 2006).

Evaluatie van Recommender Systems

Door het grote succes van RS in de praktijk is de academische interesse in dit onderwerp heel sterk gestegen (Schedl et al., 2015). Er is reeds veel onderzoek gedaan naar het verbeteren van een (muziek) aanbevelingsalgoritmes (Yoshii et al., 2008), maar het blijft onduidelijk op basis van welke maatstaven een RS het best wordt geëvalueerd (Schedl et al., 2018). Er zijn 2 perspectieven van waaruit de evaluatie kan worden gedaan: systeem gefocust (*system centered*) of gebruiker gefocust (*user centered*) (Cremonesi et al., 2013). Bij de systeem gefocuste analyse wordt er gebruik gemaakt van data vanuit het systeem zelf om maatstaven zoals precisie te berekenen. Het systeem wordt geëvalueerd door maatstaven opgesteld door het systeem zelf. Bij de gebruiker gefocuste analyse wordt rekening gehouden met de interactie van de gebruiker met het systeem. Hier evalueert de gebruiker het systeem op basis van hoe hij het zelf percipieert. In de volgende paragrafen wordt besproken hoe RS volgens de bestaande literatuur wordt geëvalueerd en hoe deze evaluatiemethoden in de loop der tijd zijn geëvolueerd.

Systeem gefocust analyse

Doorheen de literatuur is er een duidelijke evolutie van de wijze waarop RS geëvalueerd wordt. Eerst lag de focus van de evaluatie van RS op accuracy (Cacheda et al., 2011; Jannach & Bauer, 2020). Dit is een systeem gefocuste analyse.

Er zijn verschillende maatstaven om accuracy te kwantificeren zoals de mean absolute error (MAE) en de precision at top K recommendations (P@K) (Schedl et al., 2018). MAE meet de gemiddelde afwijkingen tussen de door gebruiker gegeven score en de voorspelde score. P@K is de verhouding tussen het aantal relevante items (in de top k aanbevelingen) en het aantal items (in top k).

Het evalueren van RS op basis van accuracy leent zich goed tot het vergelijken van RS omdat deze evaluatie-maatstaven makkelijk te berekenen en te herhalen zijn (A. Kim et al., 2021) maar, deze overdreven focus op accuracy heeft ook nadelen. Deze overdreven focus zorgt voor problemen zoals bijvoorbeeld een filter bubbel (Areeb et al., 2023) en “saai aanbevelingen” (Y. C. Zhang et al., 2012; Ziarani & Ravanmehr, 2021). Deze problematieken worden samengevat onder de term “McNamera fallacy” (Jannach & Bauer, 2020). Bij een filterbubbel wordt enkel gelijkaardige muziek aan de gebruiker aanbevolen waardoor deze constant hetzelfde gaat luisteren en hij zijn smaak niet gaat verruimen. De gebruiker zit dus vast in een bepaalde “bubbel van smaak” waardoor hij geen nieuwe items gaat ontdekken (Areeb et al., 2023). Hierdoor dreigen vooral populaire items te worden aanbevolen waardoor de “long tail” items minder aanbod komen. Dit zijn de 20% minst gestreamde items die 80% van het totale aantal liedjes in de muziek database uitmaken (Jannach & Bauer, 2020). In de toepassing van muziek zijn dit de meer niche muziek nummers/genres (Zhang et al., 2012). Naast de filterbubbels heb je ook het probleem van “saai aanbevelingen”. De mens heeft een verlangen naar variatie en verandering (Jannach & Bauer, 2020). Dit kan een RS dat (enkel) gefocust is op accuracy niet bieden (Y. C. Zhang et al., 2012). Een RS gefocust op accuracy zal te veel aanbevelingen maken die de gebruiker al verwachtte. Dit zal zorgen voor een lagere satisfactie waardoor deze het RS minder zal gebruiken (Abbas & Niu, 2019). Accuracy alleen als evaluatiecriterium is niet genoeg om de satisfactie van de gebruiker te verklaren

want een RS met een (door het systeem bepaalde) hogere accuracy heeft niet altijd een hogere satisfactie dan een RS met een lagere accuracy (Pu et al., 2012).

Daarom suggereerden tal van onderzoekers (Schedl et al., 2012; Adomavicius & Tuzhilin, 2005; Ge et al., 2010) om de evaluatiecriteria van RS uit te breiden met “beyond accuracy” maatstaven zoals novelty (L. Zhang, 2013), serendipity (Ziarani & Ravanmehr, 2021) en diversity (Kunaver & Požrl, 2017) Deze evaluatiecriteria zijn net als de accuracy evaluatiecriteria gebaseerd op het systeem en bepaald door data verzameld door het systeem zelf. Er wordt geen rekening gehouden met wat de gebruiker daadwerkelijk van deze aanbevelingen vindt. Net als de accuracy evaluatiecriteria zijn de beyond accuracy evaluatiecriteria *system centered*. (Cremonesi et al., 2013). Novelty wordt bepaald door de verhouding van aanbevolen items die een gebruiker nog niet kende tegenover het totale aantal aanbevolen items (Baeza-Yates et al., 1999). Novelty heeft een grote impact op satisfactie (L. Zhang, 2013). Serendipity is gelijkaardig aan novelty maar word uitgebreid met enkele voorwaarden. Om tot serendipity te horen moeten de aanbevolen items moeten niet enkel nieuw maar ook onverwacht zijn en een positieve ervaring opwekken bij de gebruiker (Abbas & Niu, 2019). Het is moeilijk om serendipity te definiëren omdat het een gevoelsmatige dimensie heeft. De gebruiker moet het gevoel hebben positief verrast te zijn (Kotkov, Veijalainen, et al., 2016; Zheng et al., 2015). Hierdoor is er discussie over de exacte definiëring van serendipity (Kotkov, Wang, et al., 2016). Hier wordt verder op ingegaan in deze literatuurstudie bij het onderdeel “gepercipieerde serendipity”. Serendipity kan (system centered) gekwantificeerd worden als de verhouding van onverwacht en relevante aanbevolen items tegenover het aantal items in zijn geheel (Schedl et al., 2018). Diversity wordt gezien als het verschil in items die aanbevolen worden (Kunaver & Požrl, 2017). Bij een hoger niveau van diversity gaan de aanbevelingen een breder beeld geven waardoor er meer kans is

dat de aanbeveling relevant is voor de gebruiker (Kaminskas & Bridge, 2016). Een hogere diversity heeft dus een positieve invloed op satisfactie van de gebruiker (Yadav et al., 2021).

De accuracy en beyond accuracy evaluatiecriteria zijn system centered gekwantificeerde criteria waarmee verschillende RS op een duidelijke, goedkope en gemakkelijke manier vergeleken kunnen worden (Cremonesi et al., 2013). Hoewel dit zeer positieve eigenschappen zijn worden deze evaluatiecriteria niet als voldoende gezien en is het niet duidelijk welk niveau voor deze criteria ideaal is (Schedl et al., 2012). Dit kan per gebruiker verschillen (Schedl et al., 2012). Daarnaast kunnen de door gebruiker gepercipieerde serendipity, novelty een accuracy aanzienlijk verschillen van de kwantitatief gemeten, door het systeem zelf bepaalde waardes voor serendipity, novelty en accuracy (Vargas et al., 2014; Bollen et al., 2010). Deze negatieve elementen zorgen ervoor dat de accuracy en beyond accuracy evaluatiecriteria niet diepgaand genoeg zijn om RS te kunnen evalueren (Schedl et al., 2012) en uitsluitsel te maken wel RS in de praktijk voor meer satisfactie bij de gebruiker zal zorgen (Knijnenburg et al., 2012; Pu et al., 2012). Ze missen het standpunt van de gebruiker (Schedl et al., 2018)

Gebruiker gefocuste evaluatie

De effectiviteit en het succes van RS in de praktijk hangt af van de perceptie van de klant (Cremonesi et al., 2013; Pu et al., 2012). Om de satisfactie van de gebruiker bij het gebruik van RS te kunnen onderzoeken is het cruciaal om gebruik maken van user centrische meeteenheden. Dit wordt in de literatuur nog te vaak genegeerd (Schedl et al., 2018; Kim et al., 2021). De evaluatie gebeurt hier vanuit een op de gebruiker gefocust perspectief (Cremonesi et al., 2013).

De 2 voornaamste modellen, met op de gebruiker gefocuste evaluatiecriteria, die de satisfactie bij het gebruik van RS modeleren zijn het RESQUE model (Recomender System, Quality of User Experience) van Pu et al. (2012) en het Objective System Aspects model

(OSAs) van Knijnenburg et al. (2012). In de deze modellen evalueren gebruikers een RS systeem op basis van verschillende *perceived qualities* van het RS zoals gepercipieerde accuracy en gepercipieerde serendipity. Een *perceived quality* is de subjectieve evaluatie van een bepaald aspect van RS door de gebruiker, gebaseerd op hun persoonlijke ervaringen. De gebruikers schatten in hoe goed het RS scoort, bijvoorbeeld het RS heeft volgens de perceptie van gebruiker een hoge accuracy en een lage serendipity. Op basis daarvan zal de gebruiker inschatten wat zijn consumer outcomes zijn. Dit is de inschatting van hoe tevreden (satisfactie) hij is bij het gebruik van het RS. Deze modellen zijn niet specifiek op muziekaanbevelingen gebaseerd.

Het OSA-model is gericht op aanbevelingen voor films en korte video's, wat sterk verschilt van muziek. Deze verschillen zijn te wijten aan de aard van de content. Muzieknummers zijn veel korter dan films, waardoor tijdens één luistersessie meerdere nummers worden afgespeeld en aanbevolen. Bij films wordt per kijkbeurt meestal maar één film bekeken, wat leidt tot slechts één aanbeveling. Muziek wordt vaak herhaaldelijk beluisterd, waardoor hetzelfde nummer meerdere keren aanbevolen kan worden, in tegenstelling tot films. Bovendien wordt muziek vaak passief geconsumeerd, bijvoorbeeld als achtergrondgeluid. Dit kan het RS beïnvloeden, aangezien het niet doorspoelen van een nummer verkeerd geïnterpreteerd kan worden als een positieve reactie, terwijl dit bij films wel als positief wordt gezien (Schedl et al., 2012). Het RESQUE model werd naast op film en videostreamingdienst ook gebaseerd op aanbevelingen van items bij e-commerce platformen zoals Amazon. Deze zijn ook sterk verschillend met muziekaanbevelingen. Bij e-commerce bijvoorbeeld moet je een aanbeveling daadwerkelijk kopen terwijl je in de context van muziekstreaming kosteloos naar een aanbeveling kan luisteren. Mogelijks heeft dit invloed op de verbanden uit deze modellen. Bij muziekstreaming ga je mogelijks door het luisteren van meer items veel meer opzoek naar variatie als bij filmstreaming, waardoor serendipity in de

context van muziekstreaming een grotere invloed heeft op satisfactie als bij film streaming (Schedl et al., 2018). Daarom is het interessant om deze modellen ook te testen in de context van muziek.

Het RESQUE model van Pu et al. (2012) geeft een link tussen *perceived qualities* en satisfactie met als mediators de attitude en de overtuigingen van de gebruiker. De overtuiging van de gebruiker is de korte termijn perceptie van hoe effectief het aanbevelingssysteem (RS) aanbevelingen kan maken. De gebruikersattitude is een soortgelijk concept maar dan op lange termijn. *Perceived qualities* zijn bepaalde eigenschappen die een RS kan hebben en waarop ze kunnen geëvalueerd worden. Deze evaluatie gaat dan bekijken hoe aanwezig deze eigenschappen zijn in het RS. De *perceived qualities* die we in dit model terugzien zijn vergelijkbaar met de meeteenheden uit de accuracy en beyond accuracy evaluatie zoals aanbeveling accuracy, novelty en diversity, maar in het RESQUE model worden deze niet kwantitatief berekend vanuit een systeem centrisch perspectief zoals bij de accuracy en beyond accuracy evaluatiecriteria. Deze worden in het RESQUE model bepaald door de consument. Dit is volgens het perspectief van de gebruiker.

Het OSA model van Knijnenburg et al. (2012) bouwt verder op het RESQUE model maar stelt dat ook de eigenschappen van de gebruiker (bv. domein kennis), situationele eigenschappen (bv. het doel van gebruiker), de interactie tussen de gebruiker en het RS een positieve impact hebben op satisfactie.

Perceived qualities

In dit onderzoek wordt bij de evaluatie van RS gefocust op de door de gebruiker *perceived qualities*: gepercipieerde accuracy en gepercipieerde serendipity. Deze symboliseren de tweestrijd in de literatuur. Gepercipieerde accuracy gaat over het gebruik maken van data om aanbevelingen te maken die overeenkomen met het huidige luistergedrag terwijl men bij serendipity op zoek gaat naar de positieve verrassing die komt door een

verruiming van het huidige luistergedrag. Deze analyse gebeurt vanuit het oogpunt van de gebruiker in de context van muziekstreaming. De link tussen deze *perceived qualities* en satisfactie is in het verleden al onderzocht geweest maar nog niet in de specifieke context van muziekstreaming. Ook is tot op het heden het onderzoek over de link tussen deze *perceived qualities* en WTP nog heel summier. Deze leegte biedt mogelijkheden voor verder onderzoek en analyse in dit werkstuk.

Gepercipieerde serendipity

Serendipity is het positief gevoel dat een gebruiker krijgt wanneer hij, in geval van muziekstreaming, een nummer voorgesteld krijgt dat hij nog niet kende maar wel leuk vindt (A. Kim et al., 2021). Herlocker et al. (2004) beschrijven als voorwaarde voor serendipity dat een aanbeveling nieuw en een “positieve verassing” moet zijn voor de gebruiker. Schedl et al. (2012) en Kim et al. (2021) leggen de nadruk op het relevant zijn van een aanbeveling naast het nieuw zijn voor de gebruiker. In de meeste modellen worden serendipity en novelty samengevoegd (A. Kim et al., 2021). Novelty is de kwaliteit van een RS die draait rond het aanbevelen van een item dat de gebruiker nog niet kende. Omdat “nieuw” zijn voor de gebruiker ook een voorwaarde is voor serendipity wordt er vaak geen onderscheid gemaakt tussen deze variabelen om verwarring te vermijden (Pu et al., 2011; Schedl et al., 2012). In dit onderzoek gebruiken we een combinatie van deze definities om een brede basis te vormen. Een aanbeveling die serendipity opwekt definiëren we in dit onderzoek als een aanbeveling die zowel a) nieuw, b) relevant is voor de gebruiker, en c) als een positieve verassing wordt ervaren. Een relevante aanbeveling is een aanbeveling waar je iets mee bent. Je zou dit nummer opnieuw willen beluisteren en zelf in je playlist zetten. Een aanbeveling die nieuwer (Herlocker et al., 2004), relevanter, verrassender is zorgt voor meer serendipity. Er zijn verschillende manieren waarop serendipity vanuit een system centered perspectief gekwantificeerd kan worden zoals de eerder besproken coëfficiënten: MAE en P@K. Maar in

dit onderzoek ligt de focus op hoe de gebruiker zelf serendipity percipieert. We definiëren dit als de gepercipieerde serendipity. Dit is nodig omdat een playlist die door de systeem centered evaluatie wordt beoordeeld als een playlist met een hoge serendipity door een luisteraar beoordeeld kan worden als een playlist met lage serendipity (Schedl et al., 2018). Als een gebruiker aangeeft dat zijn aanbevelingen heel nieuw, relevant en een heel positieve verrassing zijn dan zal de gepercipieerde serendipity voor dit RS hoog zijn. In tabel 1 vindt u een overzicht van de gebruikte definities in dit onderzoek.

Tabel 1

Overzicht definities

Term	Definitie	Bron
Recomender system (RS)	Technologische systemen die data van gebruikers zoals voorkeuren en surfgedrag gebruiken om bijvoorbeeld producten, nieuwsartikels aan te bevelen aan specifieke consumenten.	(Lü e.a., 2012)
Gepercipieerde accuracy	De mate waarin gebruikers het gevoel hebben dat aanbevelingen matchen met hun eerder interesses en voorkeuren.	(Pu et al., 2011)
Gepercipieerde serendipity	Het positief gevoel dat iemand krijgt wanneer het in geval van muziekstreaming een nummer voorgesteld krijgt dat hij nog niet kende maar wel leuk vindt.	(A. Kim et al., 2021)

Customer satisfactie	Een algemene evaluatie van een consument bij van zijn ervaring met een service of goed.	(Jannach & Bauer, 2020)
Willingness to pay (WTP)	Het maximale dat iemand bereid is om te betalen voor een goed of service.	(Ström & Martínez, 2013).
Mate van gebruik	De mate waarin muziek luistert (in dit onderzoek opgedeeld in zware en lichte gebruikers).	(Knijnenburg et al., 2012)

Gepercipieerde Accuracy

Gepercipieerde accuracy is de mate waarin gebruikers het gevoel hebben dat aanbevelingen matchen met hun eerder interesses en voorkeuren (Pu et al., 2011). De interesses van een gebruiker worden in het geval van muziek afgeleid uit hun luistergedrag (Schedl et al., 2012). Voorkeuren worden in dit onderzoek ingevuld als de genres waar iemand aangeeft graag naar te luisteren. Dit is een kwalitatieve maatstaf. De gepercipieerde accuracy (gebruiker gefocust) en de kwantitatieve benadering accuracy (systeem gefocust) zijn mogelijk sterk gecorreleerd. Als de gebruiker het RS als accuraat percipieert dan in is het onderliggende algoritme vaak ook kwantitatief (door het systeem zelf) gezien ook accuraat (Chen & Pu, 2009), maar er is ook onderzoek dat dit verband tegenspreekt (McNee et al., 2002). Een RS met gepersonaliseerd aanbevelingen zorgt voor en hogere gepercipieerde accuracy (Knijnenburg et al., 2012). In de context van muziek specifiek is gepercipieerde accuracy nog niet vaak onderzocht geweest (Pu et al., 2011).

Consumer outcomes

Het gebruik van een RS met verschillende waarden van accuracy en serendipity kan leiden tot verschillende consumer outcomes. De consumer outcome is de impact die het

gebruik van een bepaalde service heeft op de gebruiker. Deze kan uitgedrukt worden uitgedrukt in de variabelen customer satisfactie en WTP (Roth & Crane-Ross, 2002). Het doel van RS is om deze consumer outcome te verhogen/verbeteren (Ge et al., 2010).

Customer satisfactie

Customer satisfactie is een algemene evaluatie van een consument in verband met zijn ervaring met een dienst of goed (Jannach & Bauer, 2020). Dit is een evaluatie van de totale ervaring. Deze begint bij het aankoopproces van een bepaald goed of dienst en gaat tot en met de resultaten bij het gebruik ervan (Ferrentino & Boniello, 2020). Voor dienstverleners en retailers is het belangrijk om de satisfactie van de klanten te verhogen omdat de customer satisfactie bij de klant op termijn zorgt voor een grote waarde van het bedrijf (Gruca & Rego, 2005). Het verhogen van de satisfactie van de gebruikers door middel van gepersonaliseerde aanbevelingen is één van de belangrijkste doelen van RS (Ge et al., 2010; Martins & Riyanto, 2020). Ondanks het feit dat het belang van satisfactie bij RS gekend is, is de impact van satisfactie nog niet voldoende onderzocht geweest in voorgaande onderzoeken (He et al., 2024)

Satisfactie van streamingdiensten werd al onderzocht via het ESSQUAL model (electronic subscription service quality) (Staccini et al., 2006; Ström & Martínez, 2013). Hieruit blijkt dat de kwaliteit van de streambare content, de ervaring en de toegankelijkheid van de software, een positieve invloed heeft op de customer satisfactie bij het gebruik van streamingdiensten. Azzahro et al. (2020) voegt hier enkele *perceived values* aan toe zoals: monetary value (inschatting verhouding prijs/kwaliteit), perceived enjoyment (gevoel entertainend te worden) en identity salience (connectie tussen mens en product, deze wordt verhoogd als er personalisatie is). Deze aspecten hebben ook een positieve invloed op de satisfactie bij het gebruik van streamingdiensten. In de context van muziekstreamingdiensten zijn er ook verschillende zaken die invloed hebben op satisfactie. Indien een gebruiker zelf

input kan geven over wat voor soort aanbeveling hij wil krijgen (bv. het kiezen van het genre), zal dit zijn satisfactie verhogen.

Ook andere factoren uit de system en gebruiker gefocuste evaluatie hebben invloed op de customer satisfactie bij gebruik van RS. Het, in de system gefocuste evaluatie bepaalde, niveau van serendipity heeft volgens het auralist model van Y. C. Zhang et al. (2012) in de context van muziekaanbevelingen een positieve invloed op satisfactie. Maar er zijn ook aanwijzingen dat gepercipieerde serendipity (uit het gebruiker gefocust evaluatie) invloed heeft op satisfactie. In het RESQUE model van Pu et al. (2011) wordt de term serendipity niet gebruikt maar heb je wel variabele gepercipieerde novelty en diversity. Deze leunen gecombineerd dicht aan bij de beschrijving van serendipity (Schedl et al., 2012). Novelty en diversity hebben volgens het RESQUE model van Pu et al. (2011) en onderzoek van Martins en Riyanto (2020) in de context van film aanbevelingen een positieve invloed op satisfactie. Ook uit bijkomend onderzoek van A. Kim et al. (2021) blijkt dat de gepercipieerde serendipity een positieve invloed heeft op satisfactie. Het krijgen van een onverwachte positieve aanbeveling geeft het gevoel dat je “geluk” hebt. Dit kan bijdragen tot een hogere satisfactie (A. Kim et al., 2021).

Deze verbanden zouden mogelijk ook kunnen gelden in de specifieke context van muziekaanbevelingen door RS.

Op basis van de bovenstaande literatuur, stellen we de volgende hypothese op:

***H1a:** Een hogere waarde van gepercipieerde serendipity zorgt voor een hogere customer satisfactie.*

Uit het OAS model van Knijnenburg et al. (2012) en het RESQUE model van Pu et al. (2011) blijkt dat accuracy een positief effect op customer satisfactie in de context van video en film RS heeft. Ook in de context van nieuws aanbevelingen (J. Kim et al., 2021; Shin, 2020) en e-commerce (He et al., 2024) heeft een hogere accuracy een positieve impact op

satisfactie Dit positief verband zou mogelijks ook kunnen gelden in de specifieke context van muziekaanbevelingen door RS. We stellen op basis daarvan de volgende hypothese op:

***H1b:** Een hogere waarde van Gepercipieerde Accuracy zorgt voor een hogere customer satisfactie*

Willingness to pay

Willingness to pay (WTP) is het maximale dat een klant bereid is om te betalen voor een goed of service (Ström & Martínez, 2013). WTP is essentieel in het maken van core business beslissingen. (Scholz et al., 2015). De prijszetting baseren op WTP zorgt vaak voor een hogere winstgevendheid, hierdoor is het belangrijk om te begrijpen welke factoren in een dienst of goed deze gaan bepalen (De Toni et al., 2017). Opvallend is dat WTP van gebruikers van de muziekstreamingdienst hoger als de huidige maandelijkse subscriptie prijs (J. Kim et al., 2017)

Uit onderzoek van J. Kim et al. (2017) bleek dat er verschillende zaken invloed hebben op de WTP van een muziekstreamingdienst zoals exclusieve content, het zelf kunnen kiezen van nummers en de mogelijkheid om offline gebruik te kunnen maken van de streamingdienst. Het RS van de streamingsdienst werd hier niet bij vernoemd. Maar het gebruikte RS heeft daadwerkelijk invloed op WTP voor een streaming platform bij een (potentiële) gebruiker (Adomavicius et al., 2017). Dit verband werkt ook in de omgekeerde richting, een aanbeveling die als niet relevant wordt gepercipieerd door de gebruiker heeft een negatieve invloed op WTP. (Adomavicius et al., 2017). Er is een direct positief verband uit de literatuur tussen serendipity en WTP (A. Kim et al., 2021). Deze is nog niet bevestigd geweest in de context van RS in muziekstreaming. Daaromtrent stellen we de volgende hypothese op:

***H2a:** Een hogere waarde van Serendipity zorgt voor een hogere willingness to pay*

Uit het ESSQUAL model blijkt dat in de context van film RS het geven accurate aanbevelingen de flow van het gebruik van RS verbetert. Dit verhoogt de WTP (Pu et al.,

2012). In de context van e-commerce heeft accuracy (systeem gefocust) een positieve invloed op WTP. Maar, deze invloed vlakt af vanaf een bepaald niveau van accuracy (Scholz et al., 2015). Op basis hiervan stellen we de volgende hypothese op:

H2b: Een hogere waarde van gepercipieerde accuracy zorgt voor een hogere customer satisfactie

Customer satisfactie biedt een directere evaluatie van het gebruik van een service als WTP en kan leiden tot een hogere WTP (Oliver, 1980). Uit onderzoek van Azzahro et al. (2020) in de context van film aanbevelingen en Martínez-López et al. (2015) in de context van e-commerce aanbevelingen blijkt dat customer satisfactie een positieve invloed heeft op WTP. Op basis van de bovenstaande literatuur stellen we de volgende hypothese op:

H3: Customer satisfactie heeft een positieve invloed op WTP

Mate van gebruik

Volgens het model van Knijnenburg et al. (2012) hebben de persoonlijke eigenschappen van de gebruiker invloed op hoe het systeem gepercipieerd wordt. Een consument die in het kader van dat onderzoek meer films kijkt en zichzelf een filmexpert noemt gaat de films anders percipiëren en evalueren als iemand waarbij niet zo is (Knijnenburg et al., 2012). Iemand die veel gebruik maakt van de streamingdienst labelen we in dit onderzoek als een zware gebruiker. De mate van gebruik heeft in het onderzoek van Knijnenburg et al. (2012) een de impact op de perceived quality en de consumer outcomes. Er wordt niet dieper ingegaan op de aard van dit verband. Ze raden onderzoekers aan om dit verband in verdere onderzoeken verder te onderzoeken (Knijnenburg et al., 2012). Dit gaan we in dit onderzoek doen. Mogelijks is de rol van mate van gebruik de mediator tussen de perceptie van het model en de evaluatie (in satisfactie) hiervan. Doordat zware gebruikers veel meer muziek luisteren hebben ze mogelijks meer nood aan het verkennen van nieuwe

muziek waardoor ze serendipity belangrijk gaan vinden. Op basis van deze redenering kunnen we de volgende hypothese opstellen:

H4: *De mate van gebruik beïnvloedt het verband tussen gepercipieerde serendipity (H4a) en gepercipieerde accuracy (H4b) met customer satisfactie.*

In het onderzoek van He et al. (2024) over RS in de context van e-commerce blijkt dat er een interactie-effect is tussen serendipity en accuracy op het verband met satisfactie. Dit verband geldt mogelijks ook in de context van RS bij muziekstreamingplatformen.

H5: *Er speelt interactie-effect op het verband tussen gepercipieerde serendipity en accuracy met customer satisfactie.*

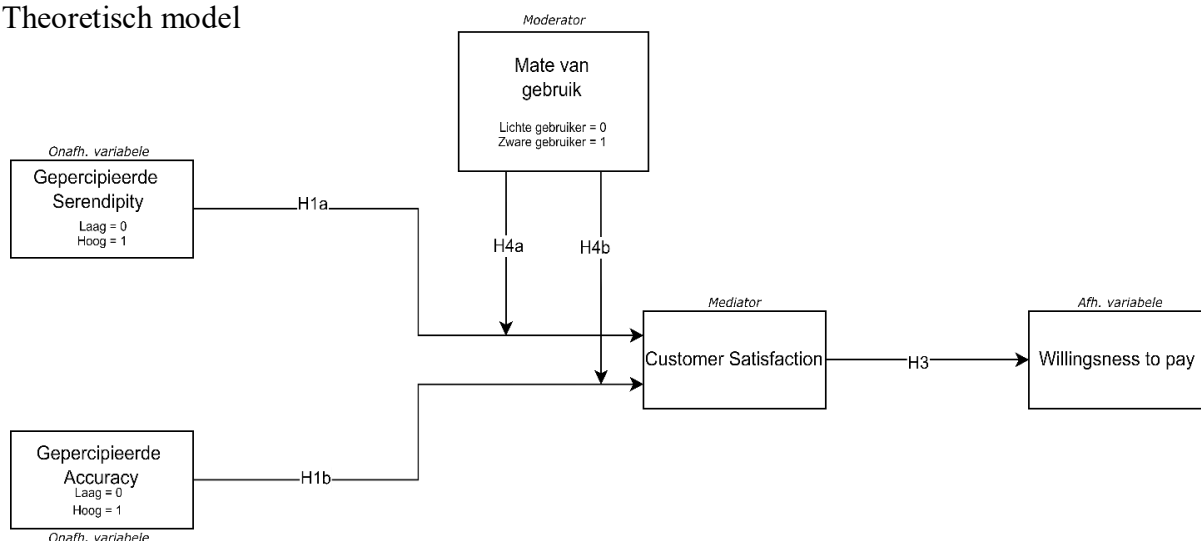
Schedl et al. (2012) hint in zijn model met beyond accuracy evaluatie criteria dat accuracy en serendipity een ander gewicht hebben. Serendipity en accuracy zouden een andere invloed hebben op satisfactie. Om te weten welke van de 2 maatstaven de grootste invloed heeft op satisfactie moet nog verder onderzoek gedaan worden (Pu et al., 2011; Schedl et al., 2012). Op basis van deze opmerkingen stellen we de laatste hypothese op.

H6: *Accuracy en serendipity hebben een verschillende invloed op satisfactie.*

Op basis van de onderstaande literatuur kunnen we het theoretisch model van dit opstellen zoals afgebeeld in figuur 1.

Figuur 1

Theoretisch model



Methodologie

Onderzoeksdesign en procedure

Om de hypotheses te testen voerde men een online scenario 2x2 between-subject experiment uit met 142 respondenten. In deze steekproef bestond 62,67% van de respondenten uit mannen, terwijl 37,32% van de respondenten vrouwen waren. De gemiddelde leeftijd van de respondenten was 28 (SD = 10.6, bereik: min = 18, max = 74). Deze respondenten werden geworven via online platforms en persoonlijke netwerken. In totaal begonnen 321 respondenten aan het experiment. Van deze groep hadden 165 respondenten niet alle vragen ingevuld en hadden 14 respondenten de controlevragen niet correct beantwoord. Deze werden uit de steekproef geweerd.

De respondenten werden geworven via online platformen en persoonlijke netwerken, dit is een gemakssteekproef. Hoewel dit type steekproef niet representatief is voor de gehele populatie en zo de generaliseerbaarheid van de conclusies beperkt zijn, blijft het een geschikte en veelgebruikte methode om op een kostenefficiënte en tijdbesparende manier een groot aantal respondenten te bereiken (Jager et al., 2017). In het volgende onderdeel wordt beschreven aan welke scenario's en items de respondent werd blootgesteld. Eerst kregen de respondenten een aantal introductievragen die peilden naar demografische informatie. Daarna kregen de respondenten één van de vier mogelijke scenario's te zien. Vervolgens kregen zij een lijst met items te zien die onder andere peilden naar satisfactie en WTP in de getoonde scenario's. Als laatst werden er enkele extra vragen gesteld die peilden naar de satisfactie en wtp van de huidige (echte) gebruikssituatie.

Introductie vragen

Aan het begin van het onderzoek ontvingen alle respondenten instructies met betrekking tot het algemene doel van de studie. Daarnaast werden belangrijke aspecten van de deelname aan de studie uiteengezet; zoals anonimiteit en privacy, vrijwillige deelname, het

recht op inzage van gegevens en de geïnformeerde toestemmingsverklaring. Deze instructies vindt men terug in bijlage A. In het eerste deel van het experiment werden enkele inleidende vragen gesteld met betrekking tot de demografische kenmerken van de respondenten, zoals leeftijd en geslacht. Daarnaast werd gevraagd of de respondenten op dat moment of in het verleden gebruik hadden gemaakt van een muziekstreamingdienst en indien van toepassing werd gevraagd hoeveel minuten ze gemiddelde per dag luisteren muziekstreamingdiensten. Een overzicht van deze demografische items vindt men terug in bijlage B.

Scenario's

Vervolgens krijgt elke respondent een scenario voorgelegd waarin één van de 4 mogelijke condities werd beschreven. De respondent werd gevraagd om zich hier zo goed mogelijk in te leven. Dit scenario gaat over een nieuwe (in het echt, niet bestaande) muziekstreamingdienst BeatBox. De respondent gaat deze volgens het scenario een maand gebruiken om daarna te evalueren wat hij/zij er ervan vond. In het scenario wordt er beschreven hoe de respondent de aanbevelingen percipieerde op de variabelen accuracy en serendipity. De respondenten werden ad random aan 1 van de vier condities toegewezen. Elke conditie bevat een manipulatie op 2 variabelen: percieved serendipity en accuracy. In elk scenario staat er beschreven hoe dat de respondenten deze percipiëren: hoog of laag. De 2 mogelijke waarden van deze variabelen worden gekruist, zo worden er 4 mogelijke condities die in het scenario staan gecreëerd. De verdeling van de waarden van de variabelen in de condities wordt getoond in de tabel 2.

Tabel 2*Overzicht condities*

Conditie	Gepercipieerde Serendipity	Gepercipieerde accuracy	N
1	Laag	Laag	31
2	Hoog	Laag	37
3	Laag	Hoog	32
4	Hoog	Hoog	42

De wijze waarop de hoge/lage waarde van serendipity of accuracy in de scenario's is beschreven werd gebaseerd op de definities van de 2 variabelen. Voor de variabele serendipity was dit het wel of niet positief verrast zijn door de aanbevelingen. Voor accuracy werd beschreven dat de aanbevelingen sterk of niet sterk leken op de nummers die de respondent nu al luistert. Uit een pre-test van de vragen (via een manipulatie check) bleek dat de respondenten deze manipulatie niet opmerkten. Daarom werd er in elke scenario een specifiek geval uit de realiteit besproken waarin de condities voorkwamen. Dit op een uitgesproken manier, zodat de respondenten de manipulatie zeker opmerkten. Om dit tastbaarder te maken wordt accuracy in de scenario's verder beschreven als ten eerste: "het gelijkaardig klinken en hetzelfde tempo hebben als nummers die je al luistert" en ten tweede: "het bevinden in hetzelfde genre waar je nu al naar luistert". Dit is gebaseerd op de definitie van accuracy waar voorop gesteld wordt dat accurate aanbevelingen gelijkaardig zijn aan ten eerste het huidig luistergedrag ("het gelijkaardig klinken en tempo hebben als nummers die je al luistert") en ten tweede de interesses van de respondent ("zelfde genre"). Dit is ook gebeurd voor serendipity. Hier wordt er ook, zoals in definitie uitgebreid uitgelegd dat een aanbeveling die serendipity bevat ten eerste onbekend moet zijn bij de gebruiker, ten tweede moet aanvoelen als een

positieve verassing en ten derde relevant moet zijn bij de gebruiker. De volledig uitgeschreven scenario's vindt men in bijlage C. Belangrijk om te vermelden is dat bij elke scenario werd beschreven dat je de aanbevelingen leuk vond. Dit om neutraal te kunnen stellen dat je een accurate aanbeveling of een aanbeveling die serendipity veroorzaakt beiden (even)leuk kan vinden. Dit werd niet gedaan bij het eerste scenario omdat de aanbevelingen daar niet accuraat zijn en geen serendipity opwekken. Volgens de opgestelde definities is het dan niet mogelijk om de aanbevelingen dan ook leuk te vinden.

Items

Na het lezen van de scenario's vullen de respondenten verschillende items (vragen) op een 7-punt Likert schaal in rond WTP en Satisfactie. Daarna werd ook het realisme van deze scenario's getoetst en werd een manipulatie check uitgevoerd. Afsluitend werden extra vragen gesteld die niet in de hypotheses staan maar mogelijks kunnen helpen om resultaten van deze hypotheses beter te begrijpen en kaderen. Er werd gepeild naar hoe belangrijk de respondenten accuracy en serendipity vonden en er wordt bevraagd wat de WTP is van de streamingdienst die ze nu gebruiken en hoe deze volgens henzelf scoren op accuracy en serendipity. In het volgende deel worden de verschillende items die bevraagd worden aan de respondent besproken.

De schalen omtrent **customer satisfactie** zijn gebaseerd op de schalen die gebruikt worden in het ESSQUAL model. Voor satisfactie worden items ingezet die gebaseerd zijn op andere veel gebruikte satisfactie studies (Brown et al., 1993; Hausknecht, 1990) Dit ESSQUAL model is ook al toegepast in de context van aanbevelingssystemen zoals in studies van Rada (2022) en Azzahro (2020). Het is een bewuste keuze om niet de schaal uit het RESQUE model (Pu et al., 2011) te gebruiken ondanks het gegeven dat de definitie van customer satisfactie hierop is gebaseerd. De schaal voor customer satisfactie in het RESQUE model bevat maar één item. Vanwege de experimentele aard van de term satisfactie is het

goed om verschillende items te hebben die naar satisfactie peilen (Azzahro et al., 2020). Daarom gaan we de schaal van ESSQUAL model gebruiken. Deze schaal bevat 3 items op een 7-punt Likert schaal. De Cronbach's alpha waarde voor deze schaal is goed ($\alpha = .84$).

Om **willingness to pay (WTP)** te meten werd er net als bij customer satisfactie gebruik gemaakt van een schaal uit het ESSQUAL model (Ström & Martínez, 2013). Het komt de interne validiteit van het onderzoek ten goede dat deze items uit eenzelfde onderzoek komen waar het verband tussen de WTP en customer satisfactie al is bewezen. In dit onderzoek wordt WTP beschreven als het maximum dat iemand wil betalen voor een dienst. Dit is een onafhankelijke evaluatie los van het extra nut tegenover een andere streamingdiensten maar je mogelijks extra voor zou willen betalen. Dit is belangrijk omdat academisch onderzoek uit het verleden hint dat de huidige prijs van streamingdiensten onder de WTP van de gebruikers ligt (J. Kim et al., 2017). Deze schaal bevat 1 item waar een bedrag moet ingevuld worden. Het item gaat als volgt: "Wat zou je maximaal per maand bereid zijn om te betalen voor BeatBox?".

De variabele **mate van gebruik** werd bevraagd door de respondenten te vragen naar hoeveel minuten per dag ze gemiddeld luisteren naar muziekstreamingdiensten. De respondenten werden door middel van het aantal minuten dat ze luisteren ingedeeld via een median split in 2 groepen: zware en lichte gebruikers. Hiermee werd verdergegaan in de analyse. In dit onderzoek maken we gebruik van een median split, omdat dit ons in staat stelt de variabele mate van gebruik te categoriseren waardoor we deze kunnen gebruiken in een ANOVA-analyse. (Iacobucci et al., 2015).

Om te testen of de respondenten zich bewust zijn van de manipulatie in de scenario's, werd er een **manipulatiecheck** opgenomen in de enquête. Hierbij werd getoetst of respondenten opmerken dat de waarden voor accuracy en/of serendipity hoog of laag zijn in de scenario's. Voor **serendipity** werd gebruikgemaakt van een schaal met 3 items op een 7-

punt Likertschaal, zoals beschreven in het onderzoek van A. Kim (2021). In het onderzoek van A. Kim (2021) wordt net als in dit onderzoek de link gelegd tussen serendipity en satisfactie. Hierdoor is de schaal ook relevant voor dit onderzoek. De Cronbach's alpha waarde voor deze schaal is goed ($\alpha = .85$). Voor **accuracy** werd gebruik gemaakt van de accuracy schaal uit het RESQUE model (Pu et al., 2011). De oorspronkelijke schaal bestond uit één item op een 5-punt Likertschaal, maar voor dit onderzoek is deze aangepast naar een 7-punt Likertschaal om consistent te zijn met de andere vragen en om de analytische verwerking te vereenvoudigen. Het is belangrijk om te vermelden dat de schalen van serendipity en accuracy niet gaan over de condities in de scenario's. Hier gaat het over vragen in de manipulatie check. In de 4 mogelijke condities zitten de 2 mogelijk waarden ingebakken voor serendipity en accuracy beschreven. De hoge en lage waarden voor serendipity en accuracy die beschreven staan in de condities, worden vertaald in de data set als 0 (lage waarden) en 1 (hoge waarden). In deze manipulatiecheck wordt onderzocht of respondenten de lage/hoge waarden voor accuracy en serendipity die zijn ingebed in de scenario's ook als zodanig interpreteren. Een overzicht van de items waarmee het theoretisch kader getest wordt vindt men in bijlage D.

Om de kwaliteit van de scenario's te waarborgen, is er een check op **realisme** uitgevoerd. Dit is courant in onderzoek met scenario's zoals in onderzoek van Verhulst et al. (2019). Hiervoor werd er gebruikt van een schaal met drie 7-punt Likert schaal items uit het marketing scales handboek van Bruner (2019). Deze schaal werd in het verleden al specifiek toegepast in een onderzoek van Adomavicius et al. (2017) en Gao et al. (2017), waar naar het realisme van een scenario uit een reclamecampagne wordt gekeken. Dit is gelijkaardig aan het beoordelen van het realisme van deze scenario's in dit onderzoek. De Cronbach's alpha waarde voor deze schaal is goed ($\alpha = .84$). Een overzicht van de items van de realisme en manipulatiecheck vindt men in bijlage E.

Aan het einde van het onderzoek werd ook rechtstreeks gepeild naar wat respondenten als het belangrijkste beschouwen: accuracy of serendipity. Deze items zijn bijzaak in dit onderzoek en kunnen alleen dienen om conclusies, gebaseerd op andere items, te versterken. Daarom is een zelf opgestelde schaal hiervoor voldoende geschikt. Dit werd bevestigd op 2 manieren. Eerst via een 7-punt Likert schaal waar gepeild wordt naar hoe belangrijk de respondenten accuracy en serendipity vinden. Deze schaal gaat van 1 (helemaal niet belangrijk) tot 7 (heel belangrijk). Ook hier was de Cronbach Alpha waarde voor de schaal van serendipity goed ($\alpha = .75$). Daarna werd er een open vraag gesteld waar rechtstreeks gepeild werd welke van de 2 variabelen (serendipity of accuracy) de respondent het belangrijkste vindt. Er werd ook gevraagd om in woorden een verklaring te geven waarom ze dat zo vinden. Dit is mogelijk interessant ter versterking van mijn conclusie, om gevonden verbanden te versterken. Een overzicht van deze items vindt men in bijlage F. De antwoorden van de respondenten werden daarna met behulp van het programma IBM SPSS Statistics (versie 29.02) geanalyseerd.

Resultaten

Realisme en Manipulatie check.

De uitgevoerde One way anova test ($CI = .95$) gaf geen significant verschil in realisme aan tussen de 4 verschillende scenario's ($p = .21$, $F(3, 138) = 1.52$) Er zijn 4 verschillende condities/scenario's waar de 2 onafhankelijke variabelen gepercipieerde accuracy en serendipity met 2 mogelijk waarden, hoog en laag, gekruist worden. Zie tabel 2 uit het hoofdstuk methodologie tabel voor de indeling van de variabelen en de frequenties waarin de condities voorkwamen.

Daarnaast worden de 4 scenario's ook gemiddeld als realistisch beschouwd door de respondent op een 7 punt Likertschaal ($(M_1 = 5.04, SE_1 = 1.06)$, $(M_2 = 5.47, SE_2 = 1.08)$, $(M_3 = 5.6, SE_3 = 1.02)$, $(M_4 = 5.45, SE_4 = 1.2)$). Dit is belangrijk want wanneer de respondenten de scenario's als niet realistisch zien, kan dit invloed hebben op hun antwoorden later in de enquête, want voor de respondent zal het moeilijker zijn om zich in te leven in een minder tot niet realistisch scenario. Dit kan het onderzoek belemmeren om de impact van de waargenomen serendipity en accuracy op satisfactie en WTP te begrijpen. Ook blijkt uit 4 uitgevoerde (1 per scenario) one sample t testen dat voor alle scenario's het niveau van realisme significant verschilt met de "neutrale" waarde 4 ($(t_1(30) = 5.5, p_1 < .001)$, $(t_2(36) = 8.24, p_2 < .00)$, $(t_3(31) = 8.86, p_3 < .001)$, $(t_4(41) = 7.827, p_4 < .001)$).

Om te controleren of de respondenten de verschillende condities in de scenario's interpreterden zoals verwacht werd er een manipulatiecheck uitgevoerd. Een independent t test toonde aan dat bij scenario's met hoge serendipity ($M = 5.37$, $SE = 0.95$) en lage serendipity ($M = 3.17$, $SE = 1.39$), de variabele serendipity door de respondenten als significant verschillend gepercipieerd worden ($t(140) = -11.15, p < .001$). Diezelfde independent t test werd ook uitgevoerd voor de variabele accuracy. Hieruit blijkt dat voor dat

de respondent de hoge ($M = 5.27$, $SE = 1.54$) en lage waarden voor accuracy ($M = 4.06$, $SE = 1.3$) als verschillend percipiëren ($t(140) = -5.06$, $p < .001$).

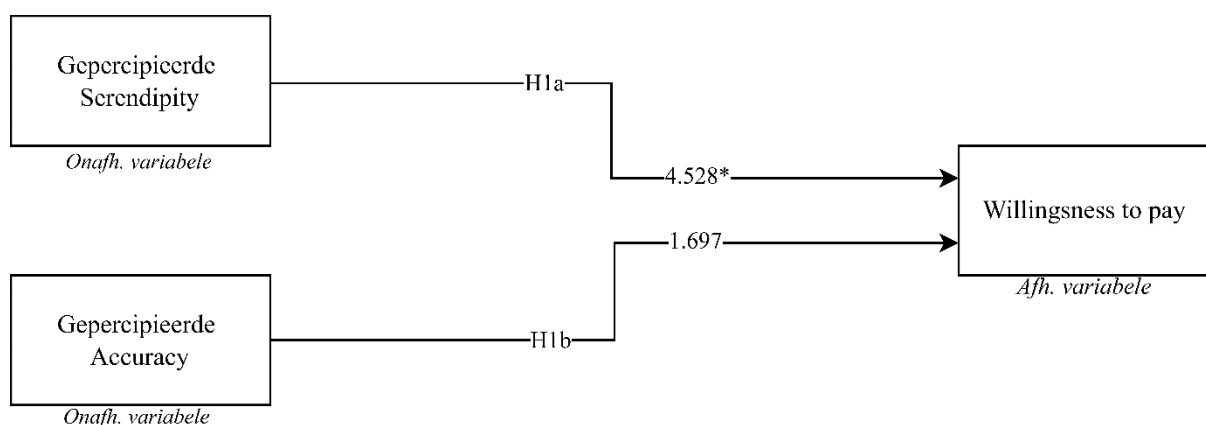
Hypotheses testen

Directe invloed op WTP

Om het directe verband tussen de 2 mogelijke waarden (hoog en laag) van serendipity en accuracy en de afhankelijke variabele WTP te bepalen werd er een 2 x 2 sided between subject ANOVA test ($N = 142$, $CI = .95$) uitgevoerd. In deze en de volgende analyses worden de waarden hoog en laag vervangen door een dummy variabelen die de waarde 0 (laag) of 1 (hoog) inneemt. Hieruit blijkt dat enkel serendipity een significant direct verband heeft met WTP ($F(1, 138) = 4.528$, $p = .035$, $\eta^2 = .03$). Op basis van dit resultaat kunnen we H1a aanvaarden. Bij accuracy is er geen significant verband met WTP ($F(1, 138) = 1.697$, $p = .195$, $\eta^2 = .01$). Hierdoor word H1b verworpen. Figuur 2 geeft overzicht van de directe verband met toebehorende statistische coëfficiënten.

Figuur 2

Theoretisch kader met de directe verbanden met bijhorende coëfficiënten



Noot: * $p < .05$

Uit diezelfde 2 way ANOVA analyse ($N = 142$, $CI = 0.95$) met satisfactie als afhankelijk variabele blijkt dat het interactie effect tussen accuracy en serendipity niet significant is ($F(1, 138) = 1.93$, $p = .167$, $\eta^2 = .01$).

Dit blijkt ook uit een 2 way ANOVA analyse ($N = 142$, $CI = 0.95$) met WTP als afhankelijke variabele. Hier is de interactie term tussen serendipity en accuracy ook niet significant ($F(1, 138) = 1.78$, $p = .185$, $\eta^2 = 0.13$). Omdat de interactie niet significant is wordt deze niet nog verder geanalyseerd. Op basis van deze info wordt hypothese H5 verworpen. Er is geen significant interactie-effect van accuracy en serendipity op satisfactie of WTP.

Uit deze ANOVA analyse blijkt ook dat serendipity ($F(1, 138) = 70.24$, $p < 0.001$, $\eta^2 = 0.34$) een grotere invloed heeft op satisfactie als accuracy ($F(1, 138) = 21.15$, $p < 0.001$, $\eta^2 = 0.13$). Hiermee wordt hypothese H6 bevestigd. Accuracy en serendipity hebben een verschillende invloed op satisfactie. Er werd ook direct gevraagd naar hoe belangrijk respondenten accuracy en serendipity vinden. Uit een paired t test op deze data blijkt dat er een significant verschil is in hoe belangrijk de respondenten serendipity ($M = 5.63$, $SD = 1.03$) en accuracy ($M = 4.25$, $SD = 1.64$) vinden. Uit deze resultaten blijkt dat de respondenten serendipity belangrijker vinden. ($t(140) = -8.727$, $p < .001$.)

Satisfactie

Daarna werd de impact van de 2 mogelijke waardes (hoog/laag) voor serendipity en accuracy uit de scenario's op WTP met customer satisfactie als mediator en mate van gebruik als moderator geanalyseerd door het PROCESS MACRO model, meer bepaald model 7 ($N = 142$, $CL = .95$). Deze analyse gebeurt in 2 stappen, eerst met customer satisfactie als afhankelijke variabele en daarna met WTP als afhankelijke variabele.

Uit deze eerste stap blijkt dat serendipity ($p < .001$, $t = 5.23$, $b = 1.66$, $CI (95\%) = 1.03 - 2.29$) een significant positief verband heeft met satisfactie. Ook accuracy ($p = .0001$, $t = 4.01$,

$b = 1.02$, $CI(95\%) = 0.52 - 1.52$) heeft een significant positieve invloed heeft op satisfactie.

Dit bevestigt hypothesen H2a en H2b.

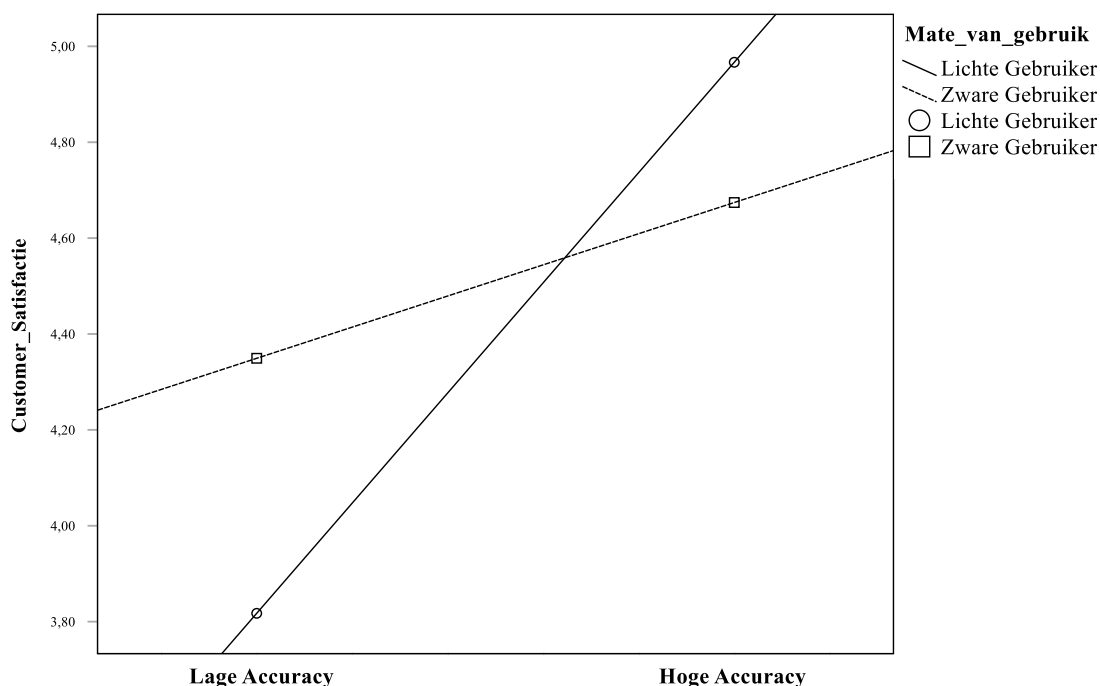
Uit de 2^{de} stap met WTP als afhankelijke variabele blijkt dat satisfactie ($p = .029$, $t = 2.2$ $b = 0.99$, $CI(95\%) = 0.41 - 1.87$) een significant positieve invloed heeft op WTP. Hieruit kunnen we besluiten dat satisfactie een mediator is tussen het verband van de 2 verschillende waarden van serendipity en accuracy met WTP. Dit bekrachtigt hypothese H3.

Moderatie van mate van gebruik

De moderator mate van gebruik is een dummy variabele. Deze variabele bestaat uit 2 groepen: zware en lichte gebruikers. Deze verdelingen gebeurt via een median split naargelang hoeveel minuten de respondenten gemiddeld per dag aangeven gebruik te maken van muziekstreaming platformen ($N_{Licht} = 76$ en $N_{zwaar} = 66$). Omdat in de PROCESS analyse een moderator maar op één variabele kan werken, werd de PROCESS analyse tweemaal uitgevoerd: één keer met moderatie op accuracy en één keer met moderatie op serendipity. Volgens PROCESS analyse (model 7, $N = 142$, $CL = 0.95$), heeft de mate van gebruik een significant moderende invloed op het verband tussen de 2 condities van accuracy ($p = .032$, $t = 2.16$, $b = 0.53$, $CI(0.95) = 0.05 - 1.1$) en satisfactie. Bij een lichte gebruiker heeft het niveau van accuracy een grotere invloed op satisfactie als bij een zware gebruiker. Dit bevestigt hypothese H4b. Dit is niet het geval voor de variabele serendipity. Hierbij heeft mate van gebruik geen significante invloed op het verband met satisfactie ($p < 0.001$ $t = 0.3$, $b = 0.08$, $CI(95\%) = -0.44 - 0.6$). Op basis hiervan wordt hypothese H4a verworpen. Moderatie van mate van gebruik op het verband tussen accuracy en satisfactie is weergegeven in Figuur 2.

Figuur 3

Moderatie Mate van gebruik op het verband tussen Accuracy en Customer Satisfactie



Om de moderatie tussen de onafhankelijke variabelen en satisfactie beter te begrijpen werd dit verband nog verder geanalyseerd. Hiervoor werden de PROCESS MACRO (model 4) analyses apart herdaan per groep gebruikers (apart voor de lichte en zware gebruiker).

Voor de zware gebruikers heeft accuracy geen significante invloed meer op satisfactie ($p = .133$, $t = 1.52$, $b = 0.3255$, $LLCL = -0.1$, $ULCL = 0$.) Dit in tegenstelling tot serendipity.

Serendipity blijft bij de zware gebruiker wel een positieve invloed hebben op satisfactie ($p < 0.001$, $t = 6.81$, $b = 1.48$, $CI(95\%) = 1.05 - 1.92$).

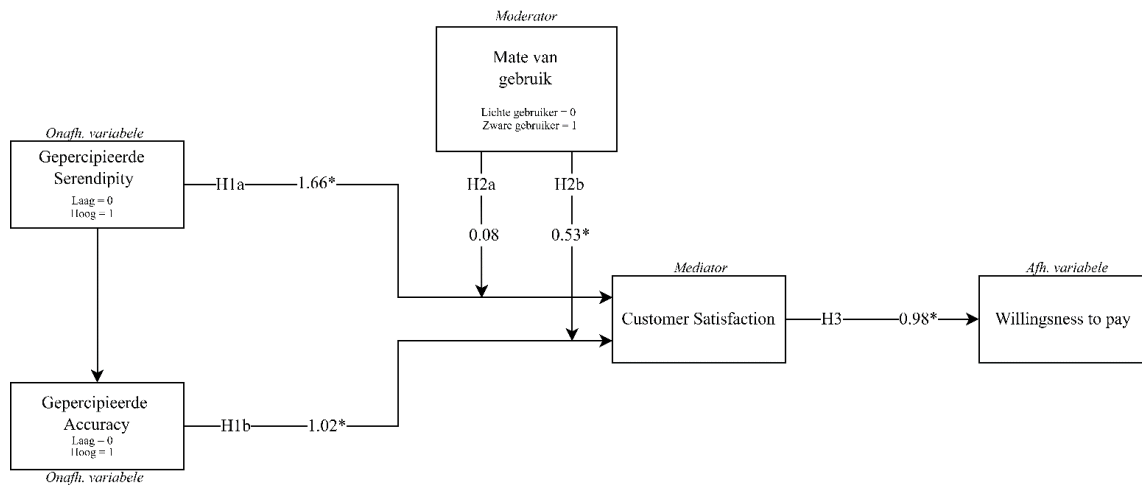
Voor specifiek de lichte gebruikers ($N = 76$) heeft accuracy nog steeds een significant positieve invloed op satisfactie ($p = .0000$, $t = 4.55$, $b = 1.15$, $CI(95\%) = 0.65 - 1.65$). Net zoals bij de variabele serendipity. Serendipity heeft ook bij zware gebruikers een positieve invloed op satisfactie ($P = 0.000$, $t = 5.13$, $b = 1.39$, $CI(95\%) = 0.8546 - 1.94$).

Hiermee concluderen we dat bij zware gebruikers, accuracy geen significant invloed meer heeft op satisfactie. Bij lichte gebruikers heeft accuracy wel nog een belangrijke impact op

satisfactie. Lichte gebruikers vinden accuracy nog steeds belangrijk. Figuur 4 geeft een overzicht van de hypothesen met de bijbehorende statistische coëfficiënten.

Figuur 4

Theoretisch kader met bijhorende coëfficiënten



Noot: * $p < .05$

Discussie

Beperkingen en toekomstig onderzoek

In dit onderzoek werd onderzocht wat de invloed is van gepercipieerde serendipity en accuracy van RS op WTP met customer satisfactie als mediator in de context van muziekstreamingdiensten. Ook werd onderzocht of er een mogelijke modererende invloed is van de variabele mate van gebruik op het verband tussen gepercipieerde serendipity en accuracy met customer satisfactie. In de literatuur was er een tekort aan onderzoek naar de evaluatie van RS vanuit het oogpunt van de consument (Schedl et al., 2018). Daarnaast zijn de bestaande modellen nog amper onderzocht geweest in de context van muziekstreaming. In dit onderzoek werd onderzocht welke factoren belangrijk zijn bij de gebruiker voor de evaluatie van RS. Dit werd onderzocht op basis van een scenario studie.

Uit deze studie bleek dat gepercipieerde serendipity en accuracy een positieve invloed hebben op WTP door middel van de mediator customer satisfactie. Gebruikers van streamingdiensten die hun aanbevelingen evalueren als het hebben van een hoger niveau van serendipity en/of accuracy zullen een hoger niveau van satisfactie voelen. Dit zal bijdragen tot een hogere WTP. Ook satisfactie heeft een positieve invloed op WTP. Deze bevindingen liggen in lijn met bevindingen uit het RESQUE model (Pu et al., 2011), het OSA model (Knijnenburg et al., 2012) en onderzoek van Kim (2021) maar, deze zijn in dit onderzoek voor het eerst specifiek in de context van muziekstreaming gevonden. Wat opvalt is dat de gepercipieerde serendipity een grotere (positieve) invloed heeft op satisfactie en WTP als gepercipieerde accuracy. Ook heeft gepercipieerde accuracy in tegenstelling met gepercipieerde serendipity geen directe invloed op WTP. Dit wijst erop dat gepercipieerde serendipity een noodzakelijk component is in het onderzoek rond de evaluatie van RS. Een evaluatie van RS die enkel gebaseerd is op (gepercipieerde) accuracy zal niet volstaan om satisfactie bij de consument te verklaren en te verhogen. Gepercipieerde serendipity en

accuracy moeten opgenomen worden in deze evaluatie. Dit alles bevestigt het advies van Schedl (2018) dat het belangrijk is om de evaluatie van RS te maken vanuit het standpunt van de consument en beyond accuracy evaluatiecriteria zoals serendipity mee te nemen in de evaluatie

Om verder op dit verband in te gaan werd de moderator mate van gebruik toegevoegd aan dit model. Dit gaf aanleiding voor de meest opvallende bevinding van dit onderzoek. Perceived accuracy heeft voor de groep zware gebruikers geen significante invloed meer op satisfactie en WTP. De zware gebruikers vinden accuracy niet meer belangrijk. Het is opvallend omdat in het bestaande onderzoek de focus sterk lag op het optimaliseren van accuracy in RS (Schedl et al., 2018). Dit blijkt voor zware gebruikers niet meer relevant te zijn. Bij de groep lichte gebruiker hebben serendipity en accuracy wel nog beide een positieve invloed op satisfactie. Zij vinden beide zaken wel nog steeds belangrijk.

Uit eerder onderzoek in de context van filmaanbevelingen (Knijnenburg et al., 2012) bleek dat zware gebruikers meer belang hechten aan serendipity dan lichte gebruikers. Dit verband werd niet bevestigd in dit onderzoek in de context van muziekstreaming. In deze context heeft mate van gebruik enkel een modererend invloed op het verband tussen gepercipieerde accuracy en satisfactie. Zware gebruikers vinden accuracy niet meer belangrijk. Terwijl bij zware en lichte gebruikers de impact van gepercipieerde serendipity op satisfactie (en WTP) gelijk blijft.

Praktische implicaties

Naast de eerdergenoemde theoretische bijdrage, heeft dit onderzoek ook belangrijke praktische implicaties voor muziekstreamingdiensten omtrent RS. De muziekstreaming markt is een heel moeilijke markt. Muziekstreamingplatformen hebben weinig unique selling points (USP) waardoor ze zich moeilijk kunnen onderscheiden. Ze bieden (grotendeels) dezelfde service aan. Dit zorgt voor een druk om de prijs laag te houden met lage winstmarges als

gevolg (Colbjørnsen et al., 2022). Als een streaming platform zijn prijs zou verhogen verliest het mogelijks een groot deel van zijn klanten aan een ander streaming platform dat praktisch hetzelfde aanbiedt zonder verhoogde prijs. De onderstaande aanbevelingen stellen aanpassingen voor aan het RS. Deze aanpassingen moeten de WTP van de gebruiker verhogen. Een RS dat voor meer satisfactie kan zorgen zal in de toekomst een belangrijk USP worden waarvoor gebruikers een premium voor willen betalen. De muziekstreamingservices die hier op inspelen zouden zo hun prijzen kunnen verhogen zonder (te veel) klanten te verliezen en zodanig hun winstmarge te kunnen vergroten.

Ten eerste blijkt uit dit onderzoek dat serendipity van RS in muziekstreamingdiensten weldegelijk invloed heeft op de WTP (direct en via satisfactie) van consumenten en dat gebruikers serendipity belangrijker vinden dan accuracy. Daarom is het dus heel belangrijk voor RS om hierop in te spelen. Het aanbevelen van nummers die de gebruiker al kent zal niet volstaan om deze tevreden te houden. Het positief verrast worden door een aanbeveling is voor de gebruiker een vereiste, geen surplus. Voor een RS die serendipity aanbiedt wil de consument meer betalen. Ten tweede blijkt uit de resultaten dat het belangrijk is om de verhouding tussen accuracy en serendipity in een RS af te stemmen op het type luisteraar waarvoor je aanbevelingen doet. Een zware gebruiker vindt serendipity belangrijk maar accuracy niet meer. De lichte gebruiker vindt deze beide zaken belangrijk. De muziekstreamingdiensten hebben al de luisterdata van hun gebruikers om deze in twee (zware en lichte gebruikers) of meer categorieën in te delen. Deze data kunnen ze gebruiken om de hoeveelheid serendipity en accuracy in hun RS te optimaliseren en te personaliseren voor een zo hoog mogelijk satisfactie en WTP bij de gebruiker. In de praktijk zou een zware gebruiker bijvoorbeeld enkel nog nummers aanbevolen krijgen die het nog niet kent en die mogelijk voor een positieve verassing kunnen zorgen. Een lichte gebruiker een mix zou krijgen van voor hem bekende en onbekende nummers.

Als derde en laatste blijkt ook uit dit onderzoek dat de satisfactie van de consument invloed heeft op zijn WTP. Daarom is het goed om het niveau van satisfactie van hun gebruikers te monitoren. Bijvoorbeeld, Streaming services kunnen op regelmatige basis de tevredenheid bevragen na het aanbevelen van nummers. Zo kan het RS beter inschatten wat voor soort aanbeveling de gebruiker fijn vindt en kan het hier dan op inspelen om de satisfactie en WTP te verhogen. Niu (2021) stelt een RS voor die zijn aanbevelingen aanpast op basis van de satisfactie die bij de gebruiker wordt gemeten. De Personalisatie van RS verhoogt de WTP (Zhao et al., 2015). De toekomst van RS is hyperpersonalisatie met behulp van data (Schedl et al., 2018). Consumenten hebben mogelijke op individueel niveau andere noden van accuracy en serendipity bij aanbevelingen (Lü et al., 2012). Een overzicht van de onderzochte hypotheses vind je in tabel 3

Tabel 3
Overzicht hypotheses

Hypothese	Status
H1a Een hogere waarde van gepercipieerde serendipity zorgt voor een hogere customer satisfactie.	Geaccepteerd
H1b Een hogere waarde van gepercipieerde accuracy zorgt voor een hogere customer satisfactie	Verworpen
H2a Een hogere waarde van serendipity zorgt voor een hogere willingness to pay	Geaccepteerd
H2b Een hogere waarde van gepercipieerde accuracy zorgt voor een hogere customer satisfactie	Geaccepteerd
H3 Customer satisfactie heeft een positieve invloed op WTP	Geaccepteerd
H4a De mate van gebruik beïnvloedt het verband tussen gepercipieerde serendipity en customer satisfactie	Verworpen

H4b	De mate van gebruik beïnvloedt het verband tussen gepercipieerde accuracy en customer satisfactie	Geaccepteerd
H5	Er speelt interactie effect op het verband tussen gepercipieerde serendipity en accuracy met customer satisfactie	Verworpen
H6	Accuracy en serendipity hebben een verschillende invloed op satisfactie.	Geaccepteerd

Beperkingen en toekomstig onderzoek

Het is belangrijk om bij het interpreteren van de resultaten, rekening te houden met de beperkingen van dit onderzoek. Ten eerste, werden de data verzameld via een gemakssteekproef via de socialmedia-pagina van de onderzoeker. De socialmedia-pagina's van de onderzoeker staan in het teken van muziek waardoor vooral mensen die geïnteresseerd zijn in muziek de oproep om de enquête in te vullen te zien kregen. Dit kan de resultaten van het onderzoek mogelijks beïnvloeden. Ondanks dat, kwam het gemiddelde aantal minuten dat een respondent in dit onderzoek muziek streamde overeen met het Europees gemiddelde (M = 114 minuten) (Statista, 2019). In toekomstig onderzoek zou het beter zijn indien men op een meer willekeurige manier respondenten voor de steekproef verzamelt. Het kan interessant zijn om respondenten in de vorm van “mechanical turkers” op crowdsourcing platformen zoals Amazon's MTurk (Shapiro et al., 2013) te bevragen. Deze methode werd al gebruikt in onderzoek van Zhao (2015). Ook kan het nuttig zijn om quota op te stellen om zo zeker te zijn dat je bijvoorbeeld ook niet gebruikers van streamingdiensten in je steekproef hebt zitten. Zo kan je je conclusie ook veralgemenen naar dat deel van de populatie (Moser, 1952). De voorkeuren tussen wel en niet gebruikers van streamingdiensten omtrent streamingdiensten verschillen (Börjeson, 2006). Misschien vinden mensen die geen streamingservice hebben het wel belangrijk om veel serendipity in hun aanbevelingen te hebben. Want dit is de surplus in nut die ze ontvangen tegenover het luisteren van de muziek die ze nu al (fysiek) bezitten en

kunnen luisteren. Dit kan in de toekomst verder onderzocht worden. In dit onderzoek kan weinig gezegd worden over de niet-gebruiker want er was maar één niet-gebruiker in de steekproef.

Ten tweede kan het ook interessant zijn om dit onderzoek opnieuw uit te voeren door respondenten effectief naar muziek te laten luisteren in plaats van dat ze zich scenario's moeten inbeelden. Dit wordt aangegeven door de respondenten van dit onderzoek. Uit een de open vraag aan het einde van dit onderzoek, waar gevraagd werd naar opmerkingen over de enquête, bleek dat de respondenten het moeilijk vonden om de scenario's in te beelden: "*De vragen met het fictief scenario maakt het wat verwarrend.*" Dit maakt het voor de respondent ook moeilijk om effectief een gevoel van satisfactie te voelen bij het lezen van de scenario's of om hun WTP in te schatten. In toekomstig onderzoek zouden respondenten werkelijk playlists met aanbevolen muziek kunnen krijgen waarin een bepaalde waarde van serendipity en accuracy zit verwerkt. Ze gaan dus echt luisteren naar de aanbevolen nummers. Dit staat ook toe om andere kwantitatieve data te verzamelen. Data over hoe lang een luisteraar naar een aanbevolen nummer luistert, als ze aanbevolen nummers opslaan in een eigen playlist of als ze het delen met andere luisteraars. Deze maatstaven worden nu in de praktijk gebruikt om playlists van muziek te evalueren (Asher Chodus, 2019). De dataverzameling kan ook in een niet experimentele setting maar in de praktijk gebeuren. Een respondent zou dan in de praktijk echt een maand lang een fictieve muziekstreamingdienst kunnen gebruiken waar bepaalde condities inzitten gemanipuleerd. Zo staan de resultaten uit het onderzoek dichterbij de realiteit.

Ten derde staat in het RESQUE model van Pu (2011), waarop het verband in het theoretisch kader van de *perceived quality's* zoals gepercipieerde accuracy en customer satisfactie zijn op gebaseerd zijn er, ook nog andere *perceived quality's* van RS die mogelijks ook in de context van muziekstreaming invloed op customer satisfactie kunnen hebben. Deze

zaken kunnen ook gebruikt worden als evaluatiecriteria bij het evalueren van een RS. In dit onderzoek van Pu (2011) is er een positief verband tussen de transparantie van het RS en customer satisfactie. Een transparant RS is een RS waar de gebruiker begrijpt hoe het werkt en kan aangetoond worden waarom iets aanbevolen wordt. Dit kan gaan doormiddel van het geven van uitleg aan de gebruikers over de werking van de RS of het geven van zichtbare labels aan muzieknnummers waaraan de gebruiker kan herkennen waarom dat ze gelinkt worden door het RS. Bij een aanbeveling zou je bijvoorbeeld via een “explanation interface” te weten kunnen komen op basis van wat een nummer is aanbevolen (Pu & Chen, 2006). Deze interface kan je bijvoorbeeld vertellen dat dit nummer aan je aanbevolen is omdat in het in specifiek muziek genre zit waar je in het weekend graag naar luisteren. Het RESQUE model werd toegepast op aanbevelingen in de e-commerce. Het is dus interessant om te onderzoeken als transparantie van RS ook een impact heeft in de context van muziekstreaming heeft op WTP en satisfactie.

Ook geeft Pu (2011) aan dat interessant kan zijn om de variabele serendipity te ontleden tot de variabele novelty (als je een nummer nog niet kende) en diversity (hoe verschillend de nummers in een playlist van elkaar zijn). Deze zaken kunnen in toekomstig onderzoek ook mee worden genomen als perceived values van RS.

Ten slotte kan in toekomstig onderzoek onderzocht worden naar welke andere moderatoren nog een rol spelen in de evaluatie. In dit onderzoek was mate van gebruik een significante moderator maar mogelijks spelen er nog andere moderatoren. Het OSA model van Knijnenburg (2012), waarop het concept mate van gebruik is gebaseerd, heeft nog andere persoonlijke en situationele dimensies die ook nog als mogelijke moderator kunnen dienen. Deze zouden streamingservices kunnen gebruiken om de ervaring van de gebruiker nog beter te personaliseren. Mogelijke moderatoren als voorbeelden zijn: introvert vs extravert of het vertrouwen dat een gebruiker heeft in het RS. Volgens Schedl (2018) zijn er ook nog

situationele omstandigheden die kunnen optreden als moderator. De situatie/reden waarom je naar een playlist luistert kan de evaluatie ervan beïnvloeden. Je wil mogelijk andere nummers luisteren terwijl je aan het sporten bent dan wanneer je aan het ontspannen bent na een lange dag op het werk. Mogelijk zijn er bepaalde activiteiten waarin je meer openstaat voor verrassing als andere. Ook heeft de emotie die voelt invloed op je muzieksmaak (Ferwerda et al., 2015). Het zou interessant zijn om in de toekomst te onderzoeken of deze zaken moderend werken op het verband tussen gepercipieerd accuracy en serendipity met satisfactie en WTP. Kennis over mogelijke moderatoren kan in de praktijk gebruikt worden op RS verder te personaliseren (Schedl et al., 2018).

Men kan dus concluderen dat de toekomst van RS in de context van muziek ligt in het verder optimaliseren en personaliseren van RS door middel van data over persoonlijkheid van gebruikers en de situatie waarin wordt gestreamd. Deze data kunnen worden aangevuld met de traditionele data over het huidige luistergedrag.

Referentias

- Abbas, F., & Niu, X. (2019). Computational serendipitous recommender system frameworks: A literature survey. *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA)*, 1-8.
https://ieeexplore.ieee.org/abstract/document/9035339/?casa_token=zwKufqOBpCIAAAAA:HHoaPUPTfaIEHRHwZ1fbBKVw0u21KGszdK9p00FvzqiUo-73rVX1-viUZQdl7vpnKei--uO7oXU
- Adomavicius, G., Bockstedt, J., Curley, S., & Zhang, J. (2017). *Effects of Online Recommendations on Consumers' Willingness to Pay* (SSRN Scholarly Paper 2982194). <https://papers.ssrn.com/abstract=2982194>
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6), 734-749.
- Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., Hussain, A., & Amira, A. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review. *WIREs Data Mining and Knowledge Discovery*, 13(6), e1512.
<https://doi.org/10.1002/widm.1512>
- Asher Chodus. (2019). *Solving and Dissolving Musical Affection: A Critical Study of Spotify and Automated Music Recommendation in the 21st Century - ProQuest*.
<https://www.proquest.com/openview/710170eea7ab205bcc705d534a549873/1?cbl=18750&diss=y&pq-origsite=gscholar&parentSessionId=9xQUIfYgu27nPk3iYr82v9YTfCR9R8iMWFncXkek2BY%3D>
- Aucouturier, J.-J., & Pachet, F. (2002). Music similarity measures: What's the use? *Ismir*, 7, 339-340. <https://archives.ismir.net/ismir2002/paper/000020.pdf>

- Azzahro, F., Ghibran, J. V., & Handayani, P. W. (2020). Customer Satisfaction and Willingness to Pay OnDemand Entertainment Streaming Service: The Role of Service Quality and Perceived Values. *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, 179-184.
<https://doi.org/10.1109/ICITSI50517.2020.9264953>
- Baeza-Yates, R., Ribeiro-neto, B., Mills, D., Bonn, O., Juan, S., Mexico, M., Taipei, C., Wesley, A., & Limited, L. (1999). *Modern Information Retrieval*.
- Barrington, L., Chan, A., Turnbull, D., & Lanckriet, G. (2007). Audio information retrieval using semantic similarity. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, 2, II-725.
https://ieeexplore.ieee.org/abstract/document/4217511/?casa_token=izGorIEKoEkAAA:AAA:FDUGs1hqK0_-7T8asGAARFNFXULWK5ZXTtOE_PlzGyCy3J5WpxdQpBm9ZKgeskY4FgFatijM24
- Barrington, L., Oda, R., & Lanckriet, G. R. (2009). Smarter than Genius? Human Evaluation of Music Recommender Systems. *ISMIR*, 9, 357-362.
<https://archives.ismir.net/ismir2009/paper/000014.pdf>
- Bollen, D., Knijnenburg, B. P., Willemsen, M. C., & Graus, M. (2010). Understanding choice overload in recommender systems. *Proceedings of the Fourth ACM Conference on Recommender Systems*, 63-70. <https://doi.org/10.1145/1864708.1864724>
- Börjeson, L. (2006). Scenario types and techniques: Towards a user's guide. *Futures*, 38(7), 723. <https://doi.org/10.1016/j.futures.2005.12.002>
- Brown, T. J., Churchill, G. A., & Peter, J. P. (1993). Improving the measurement of service quality. *Journal of Retailing*, 69(1), 127-139. [https://doi.org/10.1016/S0022-4359\(05\)80006-5](https://doi.org/10.1016/S0022-4359(05)80006-5)

- Bruner, G. C. (2019). *Marketing Scales Handbook: Multi-Item Measures for Consumer Insight Research (Volume 10)*. GCBII Productions.
- Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web*, 5(1), 1-33. <https://doi.org/10.1145/1921591.1921593>
- Chen, L., & Pu, P. (2009). Interaction design guidelines on critiquing-based recommender systems. *User Modeling and User-Adapted Interaction*, 19(3), 167-206. <https://doi.org/10.1007/s11257-008-9057-x>
- Colbjørnsen, T., Hui, A., & Solstad, B. (2022). What do you pay for all you can eat? Pricing practices and strategies in streaming media services. *Journal of Media Business Studies*, 19(3), 147-167. <https://doi.org/10.1080/16522354.2021.1949568>
- Cremonesi, P., Garzotto, F., & Turrin, R. (2013). User-Centric vs. System-Centric Evaluation of Recommender Systems. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Red.), *Human-Computer Interaction – INTERACT 2013* (Vol. 8119, pp. 334-351). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40477-1_21
- Daily time spent streaming music on Apple Music vs. Spotify by region 2018*. (2019, december). [Statista]. Statista. <https://www.statista.com/statistics/1100724/time-spent-streaming-music-by-region/>
- De Toni, D., Milan, G. S., Saciloto, E. B., & Larentis, F. (2017). Pricing strategies and levels and their impact on corporate profitability. *Revista de Administração (São Paulo)*, 52, 120-133.
- Donaldson, J. (2007). A hybrid social-acoustic recommendation system for popular music. *Proceedings of the 2007 ACM Conference on Recommender Systems*, 187-190. <https://doi.org/10.1145/1297231.1297271>

- Ekstrand, M. D., Harper, F. M., Willemsen, M. C., & Konstan, J. A. (2014). User perception of differences in recommender algorithms. *Proceedings of the 8th ACM Conference on Recommender systems*, 161-168. <https://doi.org/10.1145/2645710.2645737>
- Ferrentino, R., & Boniello, C. (2020). Customer satisfaction: A mathematical framework for its analysis and its measurement. *Computational Management Science*, 17(1), 23-45. <https://doi.org/10.1007/s10287-018-0305-1>
- Ferwerda, B., Schedl, M., & Tkalcic, M. (2015). Personality & emotional states: Understanding users' music listening needs. *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015) Dublin, Ireland, June 29-July 3, 2015*, 1388. https://bia.unibz.it/esploro/outputs/991005772990601241?institution=39UBZ_INST&skipUsageReporting=true&recordUsage=false
- Flexer, A., Schnitzer, D., Gasser, M., & Widmer, G. (2008). Playlist Generation using Start and End Songs. *ISMIR*, 8, 173-178. <https://books.google.com/books?hl=en&lr=&id=OHp3sRnZD-oC&oi=fnd&pg=PA173&dq=A.+Flexer,+D.+Schnitzer,+M.+Gasser,+and+G.+Widmer.+Playlist+generation+using+start+and+end+songs.+In+ISMIR,+2008&ots=oHKRoKky97&sig=C7Z6VLdsyyr-OC60rr6HdTKkfmo>
- Gao, H., Zhang, Y., & Mittal, V. (2017). How Does Local–Global Identity Affect Price Sensitivity? *Journal of Marketing*, 81(3), 62-79. <https://doi.org/10.1509/jm.15.0206>
- Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: Evaluating recommender systems by coverage and serendipity. *Proceedings of the Fourth ACM Conference on Recommender Systems*, 257-260. <https://doi.org/10.1145/1864708.1864761>

- Gruca, T. S., & Rego, L. L. (2005). Customer Satisfaction, Cash Flow, and Shareholder Value. *Journal of Marketing*, 69(3), 115-130. <https://doi.org/10.1509/jmkg.69.3.115.66364>
- Hausknecht, D. R. (1990). Measurement scales in consumer satisfaction/dissatisfaction. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, 3, 1-11.
- He, X., Liu, Q., & Jung, S. (2024). The Impact of Recommendation System on User Satisfaction: A Moderated Mediation Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, 19(1), 448-466.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53. <https://doi.org/10.1145/963770.963772>
- Iacobucci, D., Posavac, S. S., Kardes, F. R., Schneider, M. J., & Popovich, D. L. (2015). Toward a more nuanced understanding of the statistical properties of a median split. *Journal of Consumer Psychology*, 25(4), 652-665. <https://doi.org/10.1016/j.jcps.2014.12.002>
- Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). Ii. More Than Just Convenient: The Scientific Merits of Homogeneous Convenience Samples. *Monographs of the Society for Research in Child Development*, 82(2), 13-30. <https://doi.org/10.1111/mono.12296>
- Jannach, D., & Bauer, C. (2020). Escaping the McNamara Fallacy: Toward More Impactful Recommender Systems Research. *AI Magazine*, 41(4), 79-95. <https://doi.org/10.1609/aimag.v41i4.5312>
- Kaminskas, M., & Bridge, D. (2016). Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, 7(1), 2:1-2:42. <https://doi.org/10.1145/2926720>

- Kim, A., Affonso, F. M., Laran, J., & Durante, K. M. (2021). Serendipity: Chance Encounters in the Marketplace Enhance Consumer Satisfaction. *Journal of Marketing*, 85(4), 141-157. <https://doi.org/10.1177/00222429211000344>
- Kim, J., Choi, I., & Li, Q. (2021). Customer satisfaction of recommender system: Examining accuracy and diversity in several types of recommendation approaches. *Sustainability*, 13(11), 6165.
- Kim, J., Nam, C., & Ryu, M. H. (2017). What do consumers prefer for music streaming services?: A comparative study between Korea and US. *Telecommunications Policy*, 41(4), 263-272. <https://doi.org/10.1016/j.telpol.2017.01.008>
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4), 441-504. <https://doi.org/10.1007/s11257-011-9118-4>
- Kotkov, D., Veijalainen, J., & Wang, S. (2016). Challenges of serendipity in recommender systems. *International conference on web information systems and technologies*. <https://jyx.jyu.fi/handle/123456789/49763>
- Kotkov, D., Wang, S., & Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, 111, 180-192. <https://doi.org/10.1016/j.knosys.2016.08.014>
- Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems—A survey. *Knowledge-based systems*, 123, 154-162.
- Lü, L., Medo, M., Yeung, C. H., Zhang, Y.-C., Zhang, Z.-K., & Zhou, T. (2012). Recommender systems. *Physics Reports*, 519(1), 1-49. <https://doi.org/10.1016/j.physrep.2012.02.006>

- Martínez-López, F. J., Esteban-Millat, I., Argila, A., & Rejón-Guardia, F. (2015). Consumers' psychological outcomes linked to the use of an online store's recommendation system. *Internet Research*, 25(4), 562-588.
- Martins, M. A. J., & Riyanto, S. (2020). The effect of user experience on customer satisfaction on netflix streaming services in indonesia. *International Journal of Innovative Science and Research Technology*, 5(7), 573-577.
- McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., Konstan, J. A., & Riedl, J. (2002). On the recommending of citations for research papers. *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work*, 116-125. <https://doi.org/10.1145/587078.587096>
- Miyazaki, S. (2013). Algorhythms everywhere: A heuristic approach to everyday technologies. In *Off Beat* (pp. 135-148). Brill.
<https://brill.com/downloadpdf/book/9789401208871/B9789401208871-s010.pdf>
- Moser, C. A. (1952). Quota sampling. *Journal of the Royal Statistical Society. Series A (General)*, 115(3), 411-423.
- Niu, X., & Al-Doulat, A. (2021). LuckyFind: Leveraging Surprise to Improve User Satisfaction and Inspire Curiosity in a Recommender System. *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 163-172.
<https://doi.org/10.1145/3406522.3446017>
- Pu, P., & Chen, L. (2006). Trust building with explanation interfaces. *Proceedings of the 11th International Conference on Intelligent User Interfaces*, 93-100.
<https://doi.org/10.1145/1111449.1111475>
- Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 157-164. <https://doi.org/10.1145/2043932.2043962>

- Pu, P., Chen, L., & Hu, R. (2012). Evaluating recommender systems from the user's perspective: Survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4), 317-355. <https://doi.org/10.1007/s11257-011-9115-7>
- Rada, M. V., Santiago, M. A., Castil, F., Dayanghirang, N. D., Enoval, K., Gonzaga, R., Kazi, I., Raza, T., Rico, J., & Supangco, D. (2022). *Integrating E-S-QUAL Model and Analytical Hierarchy Process to Assess Consumers' Preferences in Selected Online Shopping Platforms in Manila, Philippines*. <https://doi.org/10.46254/EU05.20220025>
- Roth, D., & Crane-Ross, D. (2002). Impact of services, met needs, and service empowerment on consumer outcomes. *Mental Health Services Research*, 4, 43-56.
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative Filtering Recommender Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Red.), *The Adaptive Web* (Vol. 4321, pp. 291-324). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_9
- Schedl, M., Hauger, D., & Schnitzer, D. (2012). A model for serendipitous music retrieval. *Proceedings of the 2nd Workshop on Context-Awareness in Retrieval and Recommendation*, 10-13. <https://doi.org/10.1145/2162102.2162105>
- Schedl, M., Knees, P., McFee, B., Bogdanov, D., & Kaminskis, M. (2015). Music Recommender Systems. In F. Ricci, L. Rokach, & B. Shapira (Red.), *Recommender Systems Handbook* (pp. 453-492). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_13
- Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2), 95-116. <https://doi.org/10.1007/s13735-018-0154-2>

- Scheibehenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of consumer research*, 37(3), 409-425.
- Scholz, M., Dorner, V., Franz, M., & Hinz, O. (2015). Measuring consumers' willingness to pay with utility-based recommendation systems. *Decision Support Systems*, 72, 60-71.
- Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to Study Clinical Populations. *Clinical Psychological Science*, 1(2), 213-220.
<https://doi.org/10.1177/2167702612469015>
- Shin, D. (2020). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109, 106344.
<https://doi.org/10.1016/j.chb.2020.106344>
- Staccini, P., Joubert, M., & Fieschi, M. (2006). Evaluation of the usage of an object-oriented lesson model to be implemented in medical virtual learning environments. *AMIA Annual Symposium Proceedings, 2006*, 744.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839319/>
- Ström, J., & Martínez, K. B. (2013). *The determinants of customer satisfaction, loyalty and*
- Vargas, S., Baltrunas, L., Karatzoglou, A., & Castells, P. (2014). Coverage, redundancy and size-awareness in genre diversity for recommender systems. *Proceedings of the 8th ACM Conference on Recommender systems*, 209-216.
<https://doi.org/10.1145/2645710.2645743>
- Verhulst, N., Slabbinck, H., & Vermeir, I. (2019). Boosting service performance by dark chocolate seduction. *Journal of Services Marketing*, 33(5), 576-588.
<https://doi.org/10.1108/JSM-01-2019-0026>

- Wlömert, N., & Papies, D. (2016). On-demand streaming services and music industry revenues—Insights from Spotify’s market entry. *International Journal of Research in Marketing*, 33(2), 314-327. <https://doi.org/10.1016/j.ijresmar.2015.11.002>
- Xiao, L., Lu, L., Seide, F., & Zhou, J. (2009). Learning a music similarity measure on automatic annotations with application to playlist generation. *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1885-1888. https://ieeexplore.ieee.org/abstract/document/4959976/?casa_token=ZkUA96p2ZdQAAAA:DMGRnR7Ng4hXDmdWRYMxlNAIZx6HR-2lgb342DXaS5JcV2T2JAcC0ODmqhMZItZJUJnzOzcYEE
- Yadav, N., Mundotiya, R. K., Singh, A. K., & Pal, S. (2021). Diversity in Recommendation System: A Cluster Based Approach. In A. Abraham, S. K. Shandilya, L. Garcia-Hernandez, & M. L. Varela (Red.), *Hybrid Intelligent Systems* (Vol. 1179, pp. 113-122). Springer International Publishing. https://doi.org/10.1007/978-3-030-49336-3_12
- Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2006). Hybrid Collaborative and Content-based Music Recommendation Using Probabilistic Model with Latent User Preferences. *ISMIR*, 6, 296-301. <https://archives.ismir.net/ismir2006/paper/000047.pdf>
- Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2008). An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), 435-447. <https://doi.org/10.1109/TASL.2007.911503>
- Zhang, L. (2013). The Definition of Novelty in Recommendation System. *Journal of Engineering Science & Technology Review*, 6(3). http://www.jestr.org/downloads/Volume6Issue3/fulltext25632013.pdf?utm_campaign=

elearningindustry.com&utm_source=%2Ffuture-of-learning-tired-waiting-click-now&utm_medium=link

- Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, 13-22.
<https://doi.org/10.1145/2124295.2124300>
- Zhao, Q., Zhang, Y., Friedman, D., & Tan, F. (2015). E-commerce Recommendation with Personalized Promotion. *Proceedings of the 9th ACM Conference on Recommender Systems*, 219-226. <https://doi.org/10.1145/2792838.2800178>
- Zheng, Q., Chan, C.-K., & Ip, H. H. S. (2015). An Unexpectedness-Augmented Utility Model for Making Serendipitous Recommendation. In P. Perner (Red.), *Advances in Data Mining: Applications and Theoretical Aspects* (Vol. 9165, pp. 216-230). Springer International Publishing. https://doi.org/10.1007/978-3-319-20910-4_16
- Ziarani, R. J., & Ravanmehr, R. (2021). Serendipity in Recommender Systems: A Systematic Literature Review. *Journal of Computer Science and Technology*, 36(2), 375-396.
<https://doi.org/10.1007/s11390-020-0135-9>

Bijlagen

Bijlage A: Geïnformeerde Toestemming

Dit onderzoek is deel van een masterthesis voor de studierichting toegepaste economische wetenschappen aan Vrije Universiteit Brussel (VUB). Het doel van deze vragenlijst is om beter te begrijpen welke factoren de tevredenheid beïnvloeden van gebruikers van muziekstreamingdiensten. Hiermee achten we dan ook muziekstreaming diensten te helpen de tevredenheid van hun gebruikers beter te begrijpen en te verbeteren. Het invullen van de enquête duurt +- 10 minuten.

Muziekstreamingdiensten zijn platformen waarbij je tegen een maandelijkse betaling toegang krijgt tot een catalogus muziek vb: Spotify, Apple Music en Deezer.

Uw gegevens en de antwoorden die je in deze enquête deelt worden vertrouwelijk behandeld en alleen voor wetenschappelijke doeleinden in het kader van deze master thesis gebruikt. Het verzamelen en verwerken van gegevens is in overeenstemming met de wettelijke principes opgelegd door de Europese Algemene Verordening Gegevensbescherming 2016/679 (GDPR of AVG) die gelden sinds 25 mei 2018.

Door het invullen van deze vragenlijst bevestig je deel te nemen aan dit onderzoek en bevestig je de volgende instructies te hebben gelezen en aanvaard:

- (a) Deze studie wilt onderzoeken welke factoren invloed hebben op de tevredenheid van gebruikers van muziekstreamingdiensten;
- (b) Jouw deelname aan dit onderzoek is vrijwillig;
- (c) De verzamelde gegevens zullen in strikte anonimiteit verwerkt worden;
- (d) Je kan jezelf terugtrekken uit het onderzoek, zonder enige boete of consequentie;
- (e) Uw gegevens zullen enkel en alleen gebruikt worden in het kader van dit onderzoek en niet verkocht worden aan derden;

(f) U geeft de onderzoekers toestemming om de gegevens die uit deze enquête gegenereerd worden te gebruiken voor wetenschappelijke doeleinden in het kader van deze studie.

Aarzel niet mij te contacteren mocht u vragen of opmerkingen hebben via het e-mail adres: daan.de.witte@vub.be

Ik bedank u alvast voor uw deelname

Daan De Witte

Bijlage B: Demografische vragen

Wat is je leeftijd? (vul een getal in)

Wat is je geslacht

- Man
- Vrouw
- Andere

Ben je een gebruiker van muziekstreamingdiensten ?

- Ja
- Nee, maar vroeger wel
- Nee, nooit gehad

Bijlage C: Mogelijke condities

Basis conditie :

Lees het onderstaande scenario goed zodat jij je hierin kunt inleven.

Tijdens het lopen in een winkelstraat krijg u een trial abonnement van 30 dagen aangeboden voor de net opgestarte muziekstreamingdienst BeatBox. Je besliste om het aangeboden abonnement uit te proberen. Hierdoor kreeg je voor 1 maand toegang tot de BeatBox. BeatBox is gelijkaardig aan muziekstreaming applicaties zoals Spotify, Apple music, Youtube Music,.... Net als bij deze stramingdiensten kan je bij BeatBox ook luisteren naar al je favoriete muziek. Om een goed idee te krijgen van wat je van deze streaming dienst vindt luisterde je 7 dagen lang exclusief muziek op BeatBox. Bv: Bij het luisteren van muziek onderweg naar je werk/school tot het luisteren van muziek in de douche. Na het luisteren van een zelfgekozen nummer maakte BeatBox een gepersonaliseerde playlist waar je dan verder naar kan luisteren zonder dat jij zelf een nummer moet kiezen. Je totale luistergedrag was een combinatie van het zelf nummers kiezen en het luisteren naar de door BeatBox aanbevolen nummers.

Conditie 1: lage serendipity en lage accuracy

Na een maand gebruik te maken van BeatBox evalueer je de gekregen aanbevelingen om te kijken of je dit trial abonnement wilt omzetten in de betalende versie.

Je bent niet positief verrast door de aanbevolen nummers van BeatBox. Je kende de meeste aanbevolen nummers al en je had wel verwacht dat deze aanbevolen gingen worden. Je vond het ook niet leuk om ernaar te luisteren. Je gaat geen aanbevolen nummers in je persoonlijke playlist zetten om hier later zelf nog naar te luisteren.

Daarnaast heb je het gevoel dat de aanbevolen nummers totaal niet lijken op de nummers waar je al naar luistert. Ze klinken echt heel anders en hebben ook een ander tempo. De aanbevolen nummers zitten in een ander genre als het genre waar jij nu al veel naar luistert.

Conditie 2: hoge serendipity en lage accuracy

Na een maand gebruik te maken van BeatBox evalueer je de gekregen aanbevelingen om te kijken of je dit trial abonnement wilt omzetten in de betalende versie.

Je bent positief verrast door de aanbevelingen van BeatBox. Je kende de meeste aanbevolen nummers door BeatBox nog niet maar je vond ze wel leuk. Je had niet verwacht deze nummers aanbevolen te krijgen. Je zou in de toekomst nog naar deze nummers willen luisteren daarom heb je enkele aanbevolen nummers opgeslagen in je eigen playlist.

Daarnaast heb je het gevoel dat de aanbevolen nummers totaal niet lijken op de nummers waar je al naar luistert. Ze klinken echt heel anders en hebben ook een ander tempo. De aanbevolen nummers zitten in een ander genre als het genre waar jij nu al veel naar luistert;

Conditie 3: Lage serendipity en lage accuracy

Na een maand gebruik te maken van BeatBox evalueer je de gekregen aanbevelingen om te kijken of je dit trial abonnement wilt omzetten in de betalende versie.

Je bent niet positief verrast door de aanbevelingen van BeatBox. Je kende de meeste aanbevolen nummers al en je had wel verwacht dat deze aanbevolen gingen worden. Je vindt het leuk om naar de aanbevolen nummers te luisteren. Je zou er in de toekomst ook nog naar willen luisteren daarom heb je enkele aanbevolen nummers opgeslagen in je eigen playlist.

Daarnaast heb je het gevoel dat de aanbevolen nummers hard lijken op de nummers waar je al naar luistert. Ze klinken heel gelijkaardig, zitten in hetzelfde genre en zijn rond hetzelfde tempo als de nummers waar jij nu al veel naar luistert.

Conditie 4: hoge serendipity en hoge accuracy

Na een maand gebruik te maken van BeatBox evalueer je de gekregen aanbevelingen om te kijken of je dit trial abonnement wilt omzetten in de betalende versie.

Je bent positief verrast door de aanbevelingen van BeatBox. Je kende de meeste aanbevolen nummers door BeatBox nog niet maar je vond ze wel leuk. Je had niet verwacht om deze

nummers aanbevolen te krijgen. Je zou in de toekomst nog naar deze nummers willen luisteren daarom heb je enkele aanbevolen nummers opgeslagen in je eigen playlist.

Daarnaast heb je het gevoel dat de aanbevolen nummers hard lijken op de nummers waar je al naar luistert. Ze klinken heel gelijkaardig, zitten in hetzelfde genre en zijn rond hetzelfde tempo als de nummers waar jij nu al veel naar luistert.

WTP

Hoeveel euro zou je maximaal per maand bereid zijn om te betalen voor BeatBox? (vul een getal in aantal EURO in)

Bijlage F: Bonus vragen

	Helemaal niet belangrijk (1)	(2)	(3)	Neutraal (4)	(5)	(6)	Heel belangrijk (7)
Hoe belangrijk vind je dat de aanbevolen nummers hard lijken, in het zelfde tempo en genre zitten als de nummers waar jij nu al veel naar luistert?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hoe belangrijk vind je de aanbevolen nummers aanvoelen als een positieve verrassing?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Open vraag: Wat vind je het belangrijkste bij het evalueren van aanbevelingen: dat de aanbevolen nummers lijken op wat je nu al luistert of als de aanbevolen nummers voor een positieve verrassing zorgen? Leg ook uit waarom je dit vindt.

Hoeveel euro zou je maximaal per maand bereid zijn om te betalen voor de streamingdienst die jij nu gebruikt? (vul een getal in aantal EURO in)

	Helemaal niet mee akkoord (1)	(2)	(3)	(4) Neutraal	(5)	(6)	(7) Helemaal mee akkoord
Het luisteren van de door jou gebruikte streamingdienst aanbevolen nummers zorgde voor een goeie verassing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De door jou gebruikte streamingdienst aanbevolen nummers waren een onverwachte ontdekking.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De door jou gebruikte streamingdienst aanbevolen nummers matchen met jou huidig luistergedrag.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Heb jij nog een opmerking over de enquête?

Appendix G: AI Tools.

Tijdens het schrijven van deze thesis is er gebruik gemaakt van de AI tool ChatGPT voor het vertalen van vaktermen van het Nederlands naar het Engels. Ook werd in de literatuur studie op aanraden van de promotor Elicit gebruikt. Dit is een AI tool die helpt om artikels te vinden en deze samenvat. AI tools werden niet gebruik om nieuwe tekst te generen of als bron van informatie.