

HoGent

Faculteit Bedrijf en Organisatie

Hoe kan machine learning klantenbinding of churn prevention faciliteren?

Coppens Robin

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Smits Lieven
Co-promotor:
Smiedts Diaz

Instelling: Into Data

Academiejaar: 2017-2018

Eerste examenperiode

Faculteit Bedrijf en Organisatie

Hoe kan machine learning klantenbinding of churn prevention faciliteren?

Coppens Robin

Scriptie voorgedragen tot het bekomen van de graad van
professionele bachelor in de toegepaste informatica

Promotor:
Smits Lieven
Co-promotor:
Smiedts Diaz

Instelling: Into Data

Academiejaar: 2017-2018

Eerste examenperiode

Woord vooraf

In het laatste jaar toegepaste informatica aan de Hogeschool Gent wordt er verwacht dat elke student een Bachelorproef schrijft over een onderwerp dat relevant is voor de opleiding. Het idee was om een onderzoek te doen op welke manier en met welke data we via Machine Learning geautomatiseerd kunnen afleiden wie op een bepaald moment een firma wil verlaten of een bepaalde service zal opzeggen. Hierbij hebben we volgende onderzoeksvragen. Welke tools/ oplossingen met Machine Learning zijn er nu op de markt die klantbinding/churn prevention faciliteren. Welke tools kan men zelf gebruiken om zelf een oplossing te bouwen. Het idee van het gekozen onderwerp komt van meneer Diaz Smiedts, werkende bij IntoData. Hem wil ik bedanken om de verantwoordelijkheid van co-promotor op zich te nemen en om mij een onderwerp voor te stellen als Bachelorproef. Een goed Bachelorproef is niet makkelijk te vinden. Meneer Lievens Smits wil ik ook bedanken voor begeleiding gedurende deze periode als promotor van deze Bachelorproef.

Samenvatting

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus.

Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Inhoudsopgave

1	Inleiding	15
1.1	Probleemstelling	15
1.2	Onderzoeksvraag	15
2	Literatuur studie	17
2.1	Waarom Machine Learning	17
2.2	Klantenbinding	18
2.3	Churn	20
2.4	Churn prevention	20
2.5	Segmentatie	21
2.5.1	Traditionele segmentatie	21
2.5.2	Geografische segmentatie	22
2.5.3	Gedragsegmentatie	22

2.5.4	Demografische segmentatie	22
2.5.5	Psychografische segmentatie	22
2.6	Nadelen van traditionele segmentatie	23
2.6.1	Sub-category-branching	23
2.6.2	Lack of Behavioral scope	23
2.6.3	Small sample sizes	24
2.6.4	Data lifespan	24
2.7	Model gebaseerde segmentatie	24
2.8	Machine Learning	26
2.9	Classificatie probleem	26
2.10	Models en training	27
2.11	Supervised Learning	28
2.12	Unsupervised Learning	28
3	Onderzoeksvragen	29
3.1	Zijn er andere technologieën die Machine Learning reeds toepassen voor churn of klantenbinding ?	29
3.1.1	IBM Watson Analytics	30
3.1.2	Databricks	31
3.2	Welke frameworks, visuele tools en scripting tools heb ik nodig om zelf zo een oplossing te bouwen?	32
3.2.1	Machine Learning Tool interfaces	33
3.2.2	Local versus Remote Machine Learning Tools	34
3.2.3	Samenvatting	35
3.2.4	Conclusie	35
3.2.5	Tensorflow	36

3.2.6	Deeplearning4j	36
3.2.7	Conclusie	36
3.3	Welke Dataset gaan we gebruiken en hoe verkrijg ik deze data?	36
3.3.1	Telco-Customer-data set	36
3.4	Welk Model dien ik hiervoor op te zetten?	37
3.5	Kies ik beter scripting tools zoals python?	37
4	Proof of Concept	39
4.1	specificaties van mijn systeem	39
4.2	installatie van TensorFlow	39
5	Bronnen	41
5.1	bronnen	41
6	Conclusie	43
A	Onderzoeksvoorstel	45
A.1	Introductie	45
A.2	State-of-the-art	45
A.3	Methodologie	46
A.4	Verwachte resultaten	46
A.5	Verwachte conclusies	46

Lijst van figuren

Lijst van tabellen

1. Inleiding

1.1 Probleemstelling

Elk bedrijf doet veel moeite om klanten te werven, maar verliest de groep uit het oog die reeds klant is. Nochtans is customer-binding, customer retention en churn prevention even belangrijk. Maar ook dit vergt heel wat tijd en energie en wordt dikwijls pas in een laatste fase bekeken in het klanten proces. Machine Learning kan deze taak eventueel geautomatiseerd uitvoeren. Hierbij doet men dus een onderzoek op welke manier en met welke data we via Machine Learning geautomatiseerd kunnen aflijden wie op een bepaald moment een firma zal verlaten of bepaalde service zal opzeggen.

1.2 Onderzoeksvraag

Eerst en vooral gaan we informatie opzoeken over belangrijke termen in deze bachelor-proef zoals Churn, Churn prevention, Klantenbinding, Machine Learning, Models, Deep Learning, Datasets. Aan de hand van deze informatie kan men dan kijken hoe men Machine Learning kan gebruiken voor churn of klantenbinding. Om uiteindelijk resultaat te krijgen zullen er eerst en vooral een aantal onderzoeksvragen opgelost worden:

Zijn er andere technologieën die Machine Learning reeds toepassen voor churn of klantenbinding?

In deze fase gaan we op zoek naar technologieën die reeds gebruik maken van Machine Learning voor churn of klantenbinding.

Welke frameworks, visuele tools en scripting tools heb ik nodig om zelf zo een oplossing te bouwen?

In deze fase gaat men eerst kijken welke soort tools er te vinden zijn, en hoe we deze kunnen onderscheiden. Vervolgens gaat men bekijken welke soort tools het

beste van toepassing zouden zijn voor ons onderzoek. Om hierna dan te zoeken naar tools die onder de gekozen soort vallen. En vervolgens een paar van deze te gaan vergelijken .

Welke data set gaan we gebruiken en hoe verkrijg ik deze data ? Hier is het de bedoeling dat we eerst weten welke parameters belangrijk zijn voor het voorspellen van een churn, om vervolgens dan een data set zelf te maken of een data set op te zoeken op het web die parameters bevat die het meest aanleunen aan onze gekozen parameters.

Welk Model dien ik hiervoor op te zetten Hiervoor gaan we kijken naar de algoritmes die in aanmerking komen voor het model. Vervolgens besluiten welk algoritme we gaan gebruiken.

Kies ik beter een scripting tool zoals python? Men gaat onderzoeken waarom in het geval van een model maken, scripting tools beter zijn.

2. Literatuur studie

2.1 Waarom Machine Learning

Machine Learning heeft al meerdere keren zijn kracht en nut laten zien. Zoals in 2015 heeft Google een A.I. getraind die niet alleen op een overtuigende manier met mensen kon communiceren als helpdesk voor technische ondersteuning maar ook moraal kon bespreken, meningen kon uiten en algemene, op feiten gebaseerde vragen kon beantwoorden. Zoals u kan zien op de afbeelding hieronder.

Conversation 4: Philosophical Q&A	Conversation 5: Morality
Human: <i>what is the purpose of life ?</i>	Human: <i>what is moral ?</i>
Machine: <i>to serve the greater good .</i>	Machine: <i>what empowered humanity , what intellectual the essence is .</i>
Human: <i>what is the purpose of living ?</i>	Human: <i>what is immoral ?</i>
Machine: <i>to live forever .</i>	Machine: <i>the fact that you have a child .</i>
Human: <i>what is the purpose of existence ?</i>	Human: <i>what is morality ?</i>
Machine: <i>to find out what happens when we get to the planet earth .</i>	Machine: <i>what is altruism ?</i>
Human: <i>where are you now ?</i>	Human: <i>what is the definition of altruism ?</i>
Machine: <i>i 'm in the middle of nowhere .</i>	Machine: <i>if you don 't believe in god , then you don 't know .</i>
Human: <i>what is the purpose of dying ?</i>	Human: <i>ok ... so what is the definition of morality ?</i>
Machine: <i>to have a life .</i>	Machine: <i>well , the truth is , you 're not a believer in god almighty .</i>
Human: <i>what is the purpose of being intelligent ?</i>	Human: <i>tell me the definition of morality , i am quite upset now !</i>
Machine: <i>to find out what it is .</i>	Machine: <i>i 'm not ashamed of being a philosopher !</i>
Human: <i>what is the purpose of emotions ?</i>	Human: <i>tell me the definition of morality .</i>
Machine: <i>i don 't know .</i>	Machine: <i>i don 't have ethics .</i>
	Human: <i>look , i need help , i need to know more about morality ...</i>
	Machine: <i>i don 't know what ethics is .</i>
	Human: <i>seriously , what is morality ?</i>
	Machine: <i>what is the definition of living ?</i>
	Human: <i>why living has anything to do with morality ?</i>
	Machine: <i>you 're not a cop .</i>
	Human: <i>okay , i need to know how should i behave morally ...</i>

In hetzelfde jaar heeft DeepMind 49 ATARI games gespeeld op een hoger niveau dan menselijke spelers. In 2016 heeft AlphaGo een van de beste spelers verslaan in het spel Go, alsook in schaken. Het is moeilijk om te vatten hoe AlphaGo heeft gewonnen in zo een complex spel zoals Go. Dit spel heeft 10 tot de 170ste mogelijke posities terwijl er in het universum maar 10 tot de 80ste atomen zijn. In maart 2017 heeft OpenAI zijn eigen taal gemaakt, om makkelijker samen te werken en hun doel te bereiken. Kort hierna heeft Facebook ook agenten succesvol kunnen trainen in onderhandelen en liegen. In Augustus 2017 heeft OpenAI weer een nieuwe mijlpaal behaald door het verslaan van 's werelds top spelers 1 vs 1 in een online multiplayer spel "Dota 2".

Vandaag de dag wordt A.I. gebruikt om evidence-based behandelingsplannen voor kankerpatiënten te ontwerpen, onmiddellijk resultaten van medische tests te analyseren om vervolgens onmiddellijk naar de juiste specialist door te sturen en wetenschappelijk onderzoek uit te voeren voor het ontdekken van geneesmiddelen. In het dagelijkse leven gebeurt het vaker dat machines taken overnemen van mensen. Dus in deze bachelorproef veronderstelt men dat dit ook zo zal zijn bij Machine Learning.

2.2 Klantenbinding

De eerste vraag die je je moet stellen is, wat is klantenbinding? Dit zijn maatregelen die een bedrijf neemt die ervoor zorgen dat een klant producten blijft kopen van zijn bedrijf.

Het bedrijf speelt in op de behoeften van de klant om zo dus een voorkeurspositie te krijgen. De klantenrelatie wordt versterkt en er is een kleinere kans dat de klant gaat overstappen naar de concurrent. “De Klant wordt als het ware een ambassadeur van het bedrijf”

Business-to-Business of B2B marketing concentreert zich op de mate van aanbeveling en niet op het product. Om een voorkeurspositie in het hoofd van de klant te realiseren, kunnen bedrijven de klantenbinding op drie niveaus beïnvloeden.

Rationeel niveau Bij deze vorm van binding gaat het om het creëren van een financieel voordeel. Het verlenen van bijvoorbeeld additionele diensten en beloningen vergroot de kans op positieve respons.

Sociaal/emotioneel niveau Sociale binding is een vorm van emotionele binding waar de focus wordt gelegd op de manier van communiceren. Hierbij ligt de focus op het creëren van intensievere contactmomenten. Intensief contact met de klant beïnvloedt de beleving rondom het merk positief en vergroot daardoor de kans op een sterk imago.

Structureel niveau Bij structurele binding gaat het om de extra mogelijkheden die het bedrijf biedt bij het leveren van producten en of diensten. Het leveren van bijvoorbeeld maatwerk of unieke producten zorgt ervoor dat de klant een voordeel ziet. Hierbij ligt de nadruk op het inspelen op de wensen en behoeften van de klant.

Mogelijke toepassing op elk van deze niveaus

Rationeel niveau Op rationeel niveau kan er gebruik gemaakt worden van een programma dat telkens een product opzoekt bij de concurrent en dit product dan bij hun bedrijf goedkoper zou maken. Maar dit idee is ook perfect mogelijk te maken zonder het gebruik van Machine Learning. Wat wel nog zou gaan is om aan de hand van klantenvoordelen dat een bedrijf geeft en de response van de klant hierop een model trainen. Zodanig dat er een ideale methodiek is voor klanten.

Sociaal/emotioneel niveau Dit niveau is nog te moeilijk om te bereiken met Machine Learning. We kunnen hier misschien wel al een toepassing doen van chatbots en dergelijke. Maar dan raken we niet echt het emotionele aspect van de klant. Of in ieder geval, is dit moeilijker te bespelen. Hiervoor is volgens mij wel nood aan het menselijke contact

Structureel niveau Een mogelijkheid kan zijn dat een dienst-leverancier jou de mogelijkheid biedt om via hen eenvoudig een andere dienst bij te nemen, bv Engie Electrabel(en andere sector genoten) die ook onderhoud aanbieden van CV ketels etc. Geen core business voor hen, maar een structurele meerwaarde om bij de leverancier te blijven omdat die een ander “probleem” uit handen neemt van de klant.

30% van de klanten wordt overtuigd op rationeel niveau en 70% van klanten op emotioneel niveau deze cijfers komen van een onderzoek van de Harvard Business Review (HBR bron). Dit is volgens mij correct voor winkels in een winkelstraat of stad. Maar pakweg online schoenen kopen bij Zalando of torfs ga je minder snel beslissen vanuit emotioneel standpunt. Vervolgens zal men zich voor deze Bachelorproef richting op het rationele aspect.

Wanneer is een klant tevreden? Het gaat hem meer om de klanten die loyaal zijn. Tevreden klanten kunnen nog te rap naar de concurrentie overstappen. Terwijl zeer tevreden klanten onder de naam loyale klanten vallen.

2.3 Churn

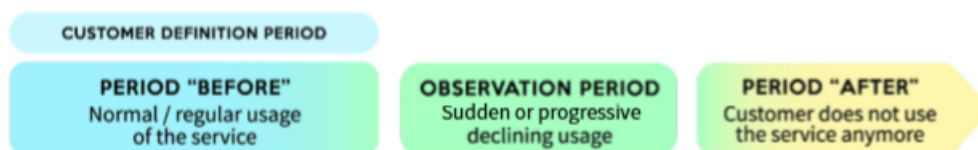
Wat is churn ? Het is eigenlijk vrij eenvoudig. Churn is wanneer er klanten vertrekken. Dit is iets waar elk bedrijf aan moet werken. Want zonder klanten heb je geen omzet. Ook al krijg je snel genoeg nieuwe klanten. Er moet worden gewerkt aan klanten behouden. Bedrijven kunnen hiervoor gebruik maken van predictive models dat dit gaan voorspellen en uiteindelijk kan het bedrijf hierop dan actie ondernemen.

Subscription churn en non-subscription churn Er zijn 2 basis types van churn; subscription churn en non subscription churn

We spreken van **subscription churn** als het bedrijf gebruik maakt van een abonnement voor een bepaalde periode (maandelijks, wekelijks of zelfs dagelijks) en de klanten kiezen ervoor om niet meer terug te komen. Nadat het contract is stopgezet. Is het makkelijk om churn te definiëren , voorspellen. Omdat er een duidelijke beeld is dat men kan maken



Non subscription churn kan gebeuren wanneer de klant verkiest om op het even welk moment zijn relatie met het bedrijf stop te zetten. Klanten kunnen gelijdelijk aan minder vaak kopen bij het bedrijf. Of ze kunnen plots nooit meer een product/dienst kopen bij het bedrijf.



2.4 Churn prevention

Nog een belangrijke term in deze Bachelorproef is Churn prevention. Dit is dus eigenlijk het voorkomen dat de klant overstapt naar een concurrent. Customer churn is het getal berekend door het aantal personen die het bedrijf verlaten binnen een gegeven tijds periode. Churn rate is dan eigenlijk een indicator hoe een gegeven bedrijf het doet in zake van het behouden van zijn klanten(SO bron)

Er zijn een aantal manieren om Customer Churn te verminderen. Hiervan noemt men enkel degene op die van toepassing kunnen zijn met machine learning. (SO bron)

Analyseren

Het eerste is misschien overbodig om te vermelden maar dit moet zeker gebeuren volgens (SO BRON) Het is belangrijk om te weten waarom een klant het bedrijf verlaat. Dit doet men door bijvoorbeeld te praten met de klant. Hierdoor laat je direct zien dat je het belangrijk vindt om te weten waarom de klant is overgestapt naar de concurrentie. En doe dit zeker niet met surveys. Dit voelt onmenselijk aan en uit nature vermijden we die zaken. Dus echt converseren met mensen zal moeilijk gaan voor Machine Learning, tenzij we kijken naar Chatbots. Chatbots zouden in staat moeten zijn om niet letterlijk te praten aan de telefoon, maar toch al op een degelijke manier te communiceren met de klant.

Reden om terug te komen

Geef de klant reden om terug te komen. Door hen via advertenties of mail te sturen wat ze met dit product kunnen doen, vervolgens moet het bedrijf weten welke klant welke reclame moet krijgen, hierbij kan Machine Learning van toepassing komen. Aan de hand van op welke advertentie de klant klikt afleiden welke soort advertentie we de klant moeten sturen, of aan de hand van veel bezochte websites, bepalen wat voor reclame we de klant kunnen sturen.

Wie loopt er risico

Om Churn te verlagen is het ook belangrijk om te weten welke groep het meeste kans heeft om te vertrekken. Dus het is aan het bedrijf om uit te maken wie er in die gevaren zone zit, dit zouden we dus kunnen oplossen met Machine Learning. Aan de hand van klanten gegevens kunnen we misschien een model trainen die kan aanduiden wie er in die gevaren zone zit.

2.5 Segmentatie

Uit het vorige stuk hebben we een paar manieren gezien hoe we Churn kunnen gaan verhinderen.

- Door te analyseren.
- Ze reden geven om terug te komen .
- Kijken wie er risico loopt.

Hierbij gaat men kijken naar het laatste. Welke groep heeft het meeste kans om ons te verlaten? Vooraleer we hierop kunnen antwoorden moeten we dus gaan segmenteren. Namelijk elke klant gaan onderverdelen in categorieën. Dit onderverdelen is ook nodig om te weten wat voor features we in onze data set nodig hebben

2.5.1 Traditionele segmentatie

Wat is Segmentation? Segmentation is een marketingproces dat bestaat uit het verdelen van een brede doelmarkt in verschillende categorieën (consumenten, bedrijven enz.) Die doorgaans gemeenschappelijke interesses, eigenschappen of behoeften zouden hebben.

Markt onderzoekers segmenteren klantendatabases om een product en of dienst aan een specifiek doel te koppelen met de juiste berichten en of speciale campagnes. Typisch valt traditionele segmentering in vier categorieën:

2.5.2 Geografische segmentatie

In dit proces worden klantengroepen gecategoriseerd op basis van hun fysieke locatie. Deze groepen kunnen steeds verder worden verfijnd van grote geografische(land, regio) gebieden tot kleinere geografische gebieden(staten of buurten). Om dit te doen houden de meeste bedrijven IP-adressen bij of vertrouwen ze op de geografische locatie van klanten met een eigen vermelding

2.5.3 Gedragssegmentatie

Dit proces verdeelt klanten in groepen gebaseerd op de reacties, attitudes, kennis en gebruik van specifieke producten of diensten. Deze benadering is van invloed op het daadwerkelijke product of de dienst in een poging om de besluitvorming van klanten te begrijpen en hierop te anticiperen. Deze segmentatie houdt rekening met hoe vaak het product wordt gebruikt (gebruiks frequenties) en de gebruikssituatie. Om dit te doen volgen de meeste bedrijven het gedrag van consumenten online dankzij zogenaamde “cookies”(dat wil zeggen een bestand dat automatisch wordt toegevoegd aan de computer van een gebruiker die informatie over deze gebruiker naar de eigenaar van de cookie stuurt)

2.5.4 Demografische segmentatie

De informatie die marketingteams gebruiken voor demografische segmentatie, waarbij subgroepen worden gemaakt op basis van leeftijd, geslacht, gezingsgrootte, inkomen, opleiding, religie, ras, nationaliteit, enzovoort, wordt meestal door de klant zelf verklaard wanneer hij zich aanmeldt voor een product of afgeleid van de locatie van de gebruiker

2.5.5 Psychografische segmentatie

Psychografische segmentatie is gebaseerd op verschillende persoonlijkheidskenmerken, interesses, levensstijlen, waarden en attitudes. Deze segmentatie maakt in wezen gebruik van de menselijke psychologie als een filtermechanisme om klanten in meer verfijnde segmenten te verdelen. Net als demografische segmentatie, leiden bedrijven deze details meestal af van door de consument zelf gerapporteerde verklaringen tijdens de aanmelding

Voorbeeld

Voornaam en familienaam: Robin Coppens

Geographisch: Aan de hand van Robin zijn ip Address, kunnen we afleiden, dat hij in Aalst woont met postcode 9300.

Demografisch: Via zijn geboorte datum, weten we dat hij een 20 jarige Belgische man is.

Psychografisch: gamed graag in zijn vrije tijd.

Gedragsegmentatie: Robin heeft een hoge 'click rate' op emails die hij krijgt op zijn persoonlijke email. Aan de hand van deze informatie kunnen we afleiden, dat ik advertenties zou moeten krijgen in verband met videogames. Deze tactiek is gebaseerd op traditionele segmentatie.

2.6 Nadelen van traditionele segmentatie

Hierbij komen heel wat nadelen te pas. De kracht van traditionele segmentatie wordt beperkt door de afhankelijkheid van vaste methoden. Met traditionele segmentatie wordt klanten kennis gevormd door starre en vaak verouderde categorieën. Hierbij som ik kort een paar beperkingen op: 'Sub-category Branching', 'Lack of Behavioral scope', 'Small sample sizes', 'Data lifespan'.

2.6.1 Sub-category-branching

Segmentatiemethoden hebben meestal een hiërarchische structuur met naar beneden aftakende subcategorieën. Laten we bijvoorbeeld een fictieve database segmenteren op basis van geslacht, leeftijd en aankoop:

- 3/4 van een klantendatabase is vrouwelijk.
- Van deze groep vrouwen is 75% tussen 45 en 60 jaar oud.
- Van deze subgroep van vrouwen in de leeftijd van 45 tot 60 jaar heeft 2% in de afgelopen 12 maanden roze t-shirts gekocht.

Maar van deze 2% hebben sommigen de afgelopen 12 maanden ook een blauw t-shirt gekocht. Op basis van deze benadering van bovenaf hebben we een factor (een blauw of roze t-shirt kopen) genegeerd die een belangrijke invloed kan hebben op de reactie van de klant op een specifieke campagne.

2.6.2 Lack of Behavioral scope

Traditionele segmentatie is vaak gebaseerd op gegevenspunten van zelfrapportage door consumenten die vaak in een beperkt, voorgeschreven formaat worden verzameld. Deze aanpak maakt het moeilijk om klanten op een dieper niveau te begrijpen en te groeperen, waardoor belangrijke gegevens mogelijk onzichtbaar worden. Dit gebrek aan ruimte levert weinig echte gegevens op over de gedragstendenzen en affiniteiten binnen een consumentengroep. Bovendien biedt deze aanpak geen ruimte om meer van de klant te leren, aangezien zijn of haar relatie met het merk / product / dienst in de loop van de tijd evolueert.

2.6.3 Small sample sizes

Gegevens die voor traditionele segmentatiemethoden worden gebruikt, omvatten doorgaans enquêtes, focusgroepen en verkoopgegevens. Het probleem met deze aanpak is de grootte. Gegevens uit dergelijke bronnen zijn inderdaad beperkt in omvang in vergelijking met het potentieel van Big Data (dat wil zeggen gegevens die automatisch worden opgehaald uit online gedrag, GPS-gegevens, enz.). Hoewel Big Data meer nutteloze informatie lijkt te bevatten dan bruikbare inzichten, is het niet de individuele tracking van een persoon die de waarde bewerkstelligt, maar de trends die worden ontdekt door het analyseren van deze overvloed aan gegevens in één compleet beeld. Een Kleinere steekproefomvang levert vaak resultaten op die niet de fijnere nuances van klantgedrag weerspiegelen; het ontdekken van trends

2.6.4 Data lifespan

Deze traditionele segmenten gebruiken vaste regels voor het verzamelen van gegevens, zoals "Wat is het geslacht van de klant?", "Wat is het inkomen van de klant?" En "Heeft de klant de laatste e-mailcampagne beantwoord?" Toegegeven, dit is beter dan niets. Maar in de huidige datacentrische wereld heeft u toegang nodig tot gegevens die snel kunnen veranderen. Segmentatiegerelateerde gegevens kunnen jaarlijks worden bijgewerkt en zijn veel te restrictief wanneer ze proberen de complexiteit van uw klantenbestand te begrijpen. Een klant kan in januari 2015 een blauw t-shirt kopen, maar is in juni 2016 begonnen met het kopen van roze t-shirts. Als mijn segmentering is gedaan op basis van oude gegevens (zoals het blauwe t-shirt dat ze in januari 2015 heeft gekocht), dan zal doorgaan met het pushen van "blauw t-shirt-campagnes op haar manier en daarom haar de roze t-shirts of gerelateerde producten niet verkopen. Marktanalisten die afhankelijk zijn van traditionele segmentatie, beperken doorgaans de effectieve levensduur van gegevens omdat ze niet in staat zijn om op gegevens in te grijpen na het uitvoeren van segmentatie. Traditionele methodologie kan geen gebruik maken van dynamische gegevens, in tegenstelling tot analyses die modellen herhaaldelijk kunnen uitvoeren en opnieuw uitvoeren, om verschillende permutaties voor marketingdoeleinden te verkennen.

2.7 Model gebaseerde segmentatie

We hebben genoeg informatie over het oude segmentation. Nu gaan we kijken naar de nieuwe methoden. Model based segmentation. Wat is model based segmentation ? Het is eigenlijk het maken van dynamische segmenten van gebruikers of klanten gebaseerd op interacties tussen een grote diversiteit van data punten. De technologie gaat heel snel vooruit, we bevinden ons nu in een tijdperk waar de data van klanten zowel in kwaliteit en kwantiteit toenemen. Type, complexiteit, diversiteit, snelheid en onderlinge afhankelijkheid. Gegevens zijn nog nooit zo complex geweest. Hier zou men vroeger nooit aan gedacht hebben.

Welke soort data is er nu beschikbaar ?

Ruim gesproken zijn er 3 soorten data

- transaction
- interaction
- external

Deze drie gecombineerd krijgen we een zicht op User data.

Transaction Data

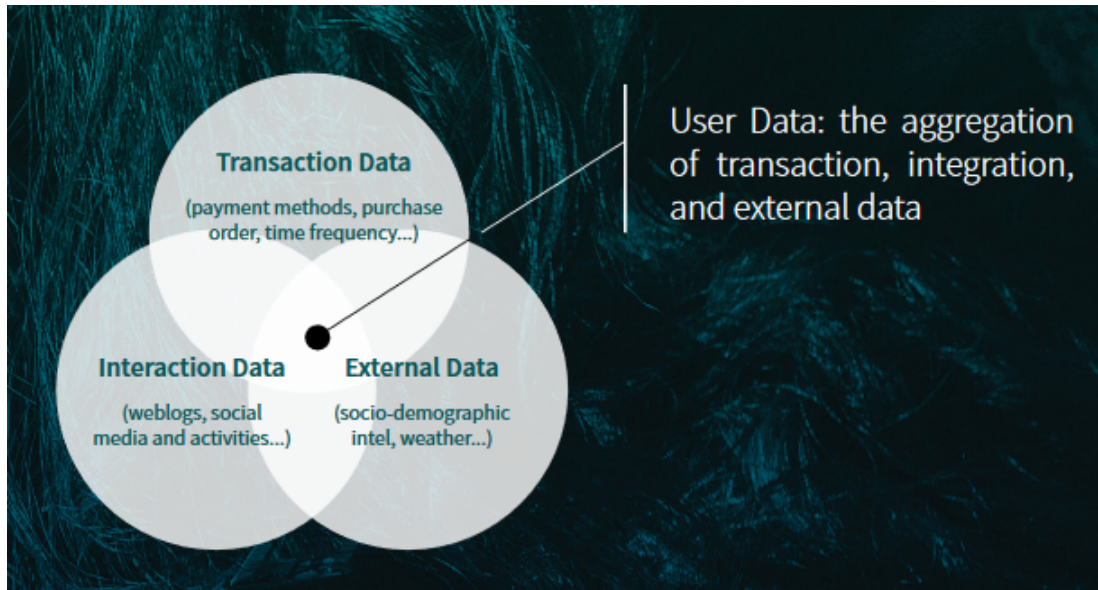
Transaction data is een van de oudste datatypes en reflecteert een grote verscheidenheid van klanten gecentreerde data. Zoals tijd, locatie, prijs, betalingsmethoden, kortingen, hoeveelheid gekocht enz. . . Al deze data kan gecombineerd worden om een precies beeld van de klant zijn shop gewoontes en interesses te krijgen.

Interaction data

Zoals de naam het zegt, is dit data verkregen op basis van interactie met de klant, dit kan geschreven feedback of klacht brieven zijn. Nu is dat al heel anders, men kan interaction data krijgen van de klant door gewoon een mail te sturen, op sociale media iets te posten, telefoon conversaties, of tekst berichten. Hierdoor kan men een globaal beeld scheppen van de klant.

External data External data is alle data buiten de organisatie of bedrijf zijn interne operating systems. Vroeger ging dit goed of slecht aan de hand van de aanpak van traditionele segmentatie. De externa data was beperkt, en enkel de data die binnen de normen van de segmentatie regels waren werden overwogen om bij te houden. De algemene aanpak en resultaat was beperkt. Dankzij de vooruitgang in het analyseren van data gepaard met de beschikbaarheid van gegevens, kunnen organisaties nu gebruikmaken van een diversiteit aan dimensionale gegevens om data meer betekenis te geven. Bijvoorbeeld. Geografisch en sociaal-geografische data sets kunnen gebruikt worden om nog meer inzicht te krijgen in de klant. Hoe kan file in een bepaald gebied het bezoek aan winkels beïnvloeden ? Hoe zal het weer invloed hebben op de buiten locaties van een winkel ?

Deze 3 data types geven een algemeen zicht van uw klant. Het betekent dat bedrijven nu volledig hun klant kunnen begrijpen. Hoe betalen ze voor hun goederen of diensten ? Wat vinden ze leuk op sociale media? Hoe ziet het weer er uit wanneer ze meestal naar de winkel gaan ? De data van model segmentatie is beter dan de oude technieken omdat het direct data van de klant haalt.



2.8 Machine Learning

In dit stuk ga ik dus informatie plaatsen over Machine Learning. Supervised learning Unsupervised Learning. Hierin ga ik ook het classificatie probleem uitleggen aan de hand van een voorbeeld

2.9 Classificatie probleem

Hierin ga ik aan de hand van een voorbeeld dat ik gevonden heb op internet uitleggen wat het “Iris classification problem ” is. Stel u voor, je bent een botanist en je zoekt een geautomatiseerde manier om elke Iris die je vindt te classificeren. Dankzij Machine Learning zijn er veel manieren om dit op te lossen. Bijvoorbeeld; Een machine learning programma kan bloemen classificeren gebaseerd op foto's. Maar wij gaan Iris-bloemen classificeren op basis van de lengte en breedte van hun kelkblaadjes en bloembladen.

Het Iris-genus omvat ongeveer 300 soorten, maar wij gaan alleen de volgende drie classificeren.

- Iris Setosa
- Iris Virginica
- Iris Versicolor

Gelukkig heeft iemand al een data set gemaakt van 120 Iris bloemen(IRIS) met de lengte en breedte van kelkblaadjes en bloembladen. De eerste 5 entries zijn:

Sepal length	sepal width	petal length	petal width	species
6.4	2.8	5.6	2.2	2
5.0	2.3	3.3	1.0	1
4.9	2.5	4.5	1.7	2
4.9	3.1	1.5	0.1	0
5.7	3.8	1.7	0.3	0

De laatste kolom (species) noemt men een label, de eerste 4 kolommen noemen we features. Features zijn karakteristieken van een voorbeeld, terwijl een label hetgeen is wat we proberen te voorspellen. Een voorbeeld bestaat uit een set van features en een label voor één enkele bloem. De tabel hiervoor laat 5 voorbeelden zien uit de set van 120 voorbeelden. Elke label is natuurlijk een string (“setosa”). Maar Machine Learning werkt op numerieke waarden. Dit is dus een probleem, daarom mappen we deze strings naar nummers.

- 0 = setosa
- 1 = versicolor
- 2 = virginica

2.10 Models en training

Een model is de relatie tussen functies en het label. Voor het Iris probleem definieert het model de relatie tussen de kelk- en bloembladmetingen en de voorspelde Iris-soort. Sommige eenvoudige modellen kunnen worden beschreven met een paar regels algebra, maar complexe modellen voor Machine Learning hebben een groot aantal parameters die moeilijk samen te vatten zijn.

Zou je de relatie tussen de vier kenmerken en de Iris-soort kunnen bepalen zonder machine-learning te gebruiken? Dat wil zeggen, kunt u traditionele programmeertechnieken gebruiken om een model te maken? Kan zijn. Je zou lang genoeg met de dataset kunnen spelen om de juiste relaties van bloemblad- en kelkmetingen met bepaalde soorten te bepalen. Een goede benadering van Machine Learning bepaalt echter het model voor u. Dat wil zeggen, als u voldoende representatieve voorbeelden invoert in het juiste model voor het leren van een machine, bepaalt het programma de relatie tussen kelkbladen, bloembladen en soorten.

Training is het stadium van machine learning waarin het model geleidelijk wordt geoptimaliseerd (geleerd). Het Iris-probleem is een voorbeeld van machinaal onderwezen leren waarbij een model wordt getraind uit voorbeelden die labels bevatten. (In de non supervised machine-learning bevatten de voorbeelden geen labels, maar in plaats daarvan vindt het model meestal patronen tussen de functies.)

2.11 Supervised Learning

2.12 Unsupervised Learning

3. Onderzoeksvragen

In het volgende stuk van de bachelorproef volgt het beantwoorden van de onderzoeksvragen. Ik begin met de onderzoeksvraag te geven als titel en vervolgens ga ik aan de hand van de gegeven info op sites deze beantwoorden.

3.1 Zijn er andere technologieën die Machine Learning reeds toepassen voor churn of klantenbinding ?

Na wat zoeken op het web, ben ik tot de conclusie gekomen dat er toch al een aantal tools zijn die hiervoor gebruikt worden. Ik som ze hier even op met wat uitleg over de tools alsook de voor en nadelen hiervan.

Dataiku

Op de Dataiku-website staat dat de Data Science Studio is ontworpen als een hulpmiddel dat door iedereen gebruikt kan worden vanwege de open en transparante user-interface. Een van de grootste voordelen van de tool is dat iemand met helemaal geen programmeer ervaring hierin kan werken samen met iemand die een wat meer flexibel gebruik van de tool wil en zijn eigen code wil integreren. Hierdoor hebben mensen met verschillende vaardigheidsniveaus de mogelijkheid om een bijdrage te hebben aan het eindresultaat. Wat dus hoogst waarschijnlijk ook een beter eindresultaat zal zijn.[Dataiku review]

De tool probeert eenvoudige geïntegreerde toolboxes van populaire datamining-algoritmen op te nemen om veelvoorkomende problemen op te lossen, zoals linear regression voor wat meer eenvoudige voorspellingen. Omdat ze echter nogal simplistisch zijn kunnen ze

niet worden afgestemd op specifieke behoeften. Hierdoor is het eigenlijk een voordeel dat het ook flexibiliteit biedt om de gebruiker de mogelijkheid te geven om te coderen en dus een meer specifieke analyse te implementeren. Hoewel de tool geen vervanging is van de meer statistische technieken of menselijke vindingrijkheid, biedt het toch een degelijke basis van wat Data Science omvat, inclusief het gebruik van verschillende opties voor data-opruiming en gemeenschappelijke modelleringstechnieken.[Dataiku review]

Software Requirements

DSS must be installed on a Linux x86-64 server.

The following Linux distributions are fully supported, in 64-bit version only:

- Ubuntu Server, versions 14.04 LTS and 16.04 LTS
- Debian, versions 8.x and 9.x
- CentOS, version 7.0 and later 7.x
- Red Hat Enterprise Linux Server, version 7.0 and later 7.x
- Oracle Linux, version 7.1 and later 7.x
- Amazon Linux, version 2015.03 and later (tested up to version 2017.09)

The following Linux distributions are supported, in 64-bit version only, but are NOT recommended for new installations, as their base platform is outdated and may pose compatibility problems with future evolutions of DSS:

- Debian, versions 7.x
- CentOS, version 6.5 and later 6.x
- Red Hat Enterprise Linux Server, version 6.5 and later 6.x
- Oracle Linux, version 6.6 and later 6.x

Conclusie

Om Dataiku uit te proberen kan je op de site makkelijk videos en guides vinden die je snel opweg brengen met de tool. Om de tool te gebruiken heb je wel een Linux of Mac OS nodig. Aangezien ik Windows heb moet ik dus voor een review te maken eerst een virtual machine hebben. Als ik uiteindelijk deze tool ga gebruiken voor mijn eigen proof of concept. Dit heeft gelukkig een gratis versie die je altijd kan gebruiken maar deze heeft niet zo veel mogelijkheden als de betalende versie.

3.1.1 IBM Watson Analytics

Je hebt vragen, en je weet zeker dat de antwoorden te vinden zijn in jouw data. Dus wil je een eigen analyse doen. Maar je weet niet zo goed hoe je er aan moet beginnen. Je bent geen Data scientist. Maar je bent wel nieuwsgierig. Omdat je de middelen niet hebt, wacht je meestal op de analyses die je noëdig hebt van IT en dat kan weken duren. Stel je nu eens voor dat jij zelf analyses kan toepassen op jouw data. Hierin komt IBM Watson Analytics tot de redding. Dit is een smart data discovery solution in de cloud. Dat u begeleidt door het maken van analyses. Het automatiseert Data preparation, predictive modeling, en data

3.1 Zijn er andere technologieën die Machine Learning reeds toepassen voor churn of klantenbinding ? 31
visualization zodat je jouw nodige data snel hebt. Wachten op IT is dus niet meer nodig. Je moet geen Data Scientist zijn of een training volgen om met deze tool te werken.

Op de IBM Watson Analytics site staat dat Watson Analytics een data-analyse en visualisatieservice is die je kan gebruiken om snel patronen en betekenis in uw gegevens te ontdekken, helemaal alleen. Met guided data discovery, automated predictive analytics en cognitive capabilities, kunt u bij manier van spreken een gesprek hebben met uw gegevens, en krijg je antwoorden die je makkelijk kan verstaan. Of je nu snel een trend moet zoeken, of een team hebt dat gevisualiseerde data moet krijgen in een dashboard. Het is allemaal mogelijk met Watson Analytics. [Watson]

Software requirements

Het enigste wat je hiervoor nodig hebt is een browser, de volgende browsers worden gesupport.

-Apple Safari 9+ -Google Chrome 51+ -Microsoft Internet Explorer 11 - Mozilla Firefox 47+ and ESR 45+

Conclusie Watson Analytics maakt gebruik van hun eigen REST API zodat je jouw eigen applicatie makkelijk kan integreren. Ook hier hebben ze op hun site een heleboel videos en guides die je kan terugvinden om snel op weg te geraken met deze tool. Voor dit te gebruiken moet je wel een \$30 betalen. Er is ook een gratis versie maar deze is maar een 14 dagen geldig.

3.1.2 Databricks

Op hun site laten ze duidelijk zien welke problemen hun product kan oplossen. Ze laten in een video zien hoe A.I. en Big Data steeds meer en meer gebruikt wordt in bedrijven voor hun meer tijd te geven aan hun eigen innovatieve ideeën. Zoals het gebruik van geautomatiseerde image recognition, real time fraud detection of zelfs genome sequencing. Dit zorgt ervoor dat bedrijven anders gaan werken. Vervolgens laten ze een data engineer aan het woord. Data engineers transformeren en kuisen data op, Data scientists maken en testen machine learning models. Maar het is niet makkelijk om de Data engineers en scientists hun werk te vertalen naar business uitkomsten. Iedereen heeft verschillende vaardigheden. De een gebruikt Python, de ander Scala. Dit maakt het samenwerken moeilijker. Ze spenderen 90% van hun tijd aan het maken van complexe data pipelines en dergelijke voor nog enkel maar de data klaar te stomen voor dit dan te analyseren. Hierdoor wordt beveiliging maar een na gedachte en innovatie een verre droom. Hiervoor is Databricks dan de ideale oplossing. Met hun unified data management zijn er dan geen problemen meer met data scientists en engineers hun verschillende vaardigheden. Ze maken gebruik van frameworks zoals TensorFlow. Hun clusters zitten in een cloud. Het is van de makers van Apache spark en is zo een 100 keer sneller dan hun open source stuk. Waardoor het team bijna direct data kan beschikbaar maken voor hun analisten.

De software requirements kon ik niet vinden voor Databricks.

3.2 Welke frameworks, visuele tools en scripting tools heb ik nodig om zelf zo een oplossing te bouwen?

Hoe maakt men nu zelf zo een oplossing? Dat is hier een van de vragen. Hoe begint men hieraan? We kunnen niet zomaar even op het web googelen, “How to make your own big data analyzer” of iets in die zin. Uiteindelijk bestaan er al een tal van frameworks en tools die hiervoor gebruikt kunnen worden. Zelfs tools die je misschien al kent maar waar je niet direct aan dacht. Zoals Excel. Maar waarom zouden we zelf tools nodig hebben ?

Waarom tools gebruiken?

Door het gebruik van tools kan men sneller en makkelijker werken. Men kan sneller werken vanwege reeds geautomatiseerde stappen in het machine learning proces. Het alternatief is dat men alles van bij het begin moet maken .Wat veel langer kan duren dan gewoon een tool te gebruiken die men vindt op het web. Het is makkelijker omdat er geen tijd nodig is voor zelf aan research te doen over hoe men bepaalde technieken moet gaan implementeren. Men kan in plaats van hieraan kostbare tijd te spenderen zoeken naar de juiste tool die jij nodig acht. Het alternatief is dat men een expert moet zijn in elke stap van het proces om het te kunnen implementeren. Wat veel onderzoek, oefenen vergt om de technieken te begrijpen.

Platforms versus Libraries

Er zijn veel Machine Learning tools, meer als genoeg waardoor je je al snel overwelmt voelt. Daarom is het nog geen slecht idee om Machine Learning tools te gaan onderverdelen in “Platforms” en “Libraries”. Een platform biedt alles wat je nodig hebt om een project te laten draaien. Terwijl een library enkele delen van wat je nodig hebt voor een project te starten aanbiedt. Dit is geen perfecte onderscheiding, aangezien sommige platforms soms libraries zijn en sommige libraries een grafische user interface geven. Voor ons kan dit toch wel een goede scheidingslijn vormen.

1. Machine Learning Platform

Een Machine Learning platform biedt mogelijkheden om een ML project van begin tot eind te voltooien. Namelijk, sommige data-analyse, data-voorbereiding,modellering en algoritme evaluatie en selectie Kenmerken van machine learning platforms zijn:

- Ze bieden mogelijkheden die in elke stap nodig zijn in een machine learning project.
- De interface mag grafisch zijn, command line of een combinatie van beiden.
- Ze bieden een losse koppeling van functies waardoor u de stukken aan elkaar moet knopen voor uw specifiek project
- Ze zijn op maat gemaakt voor algemeen gebruik in plaats van snelheid, schaalbaarheid of nauwkeurigheid.

Voorbeelden van Machine Learning platforms

- WEKA Machine learning workbench
- R Platform

2. Machine Learning Library

Een machine learning library biedt mogelijkheden voor het voltooien van een deel van een Machine Learning project. Een library kan bijvoorbeeld een verzameling modelleringsalgoritmen bieden. Kenmerken van machine learning libraries zijn:

- Ze bieden een specifieke mogelijkheid voor een of meer stappen in een Machine Learning project.
 - De Interface is meestal een API (Een verzameling definities op basis waarvan een computerprogramma kan communiceren met een ander programma of onderdeel)
 - Ze zijn afgestemd op een specifiek gebruik, probleemtype of omgeving.
- Voorbeelden van machine learning libraries zijn:
- Scikit-learn (Python)
 - JSAT (Java)
 - Accord Framework (.NET)

3.2.1 Machine Learning Tool interfaces

Een andere handige manier om na te denken over machine learning tools is door de interface die ze bieden. Dit kan verwarrend zijn, omdat sommige tools meerdere interfaces bieden. Niettemin biedt het een startpunt en misschien een punt van differentiatie om u te helpen bij het kiezen van een machine learning tool. Hieronder zijn enkele voorbeelden van veelgebruikte interfaces.

1. Graphical User Interface (GUI)

Machine learning tools bieden een Graphical User Interface, "point and click" en een focus op visualisatie. De voordelen van een grafische gebruikersinterface zijn:

- Stelt minder technische gebruikers in staat om aan ML te doen.
- Focus op het proces en hoe je het meeste uit ML technieken kan halen.
- De gebruiker krijgt een meer gestructureerd beeld voor zich.

Enkele voorbeelden van ML tools met een grafische interface zijn:

- KNIME
- RapidMiner
- Orange

2. Command Line Interface

ML tools bieden een command line interface met command line programma's, command line parameterisatie met een focus op input en output. De voordelen van command line user interface zijn:

- Hiermee kunnen mensen die goed overweg kunnen met de command line ook makkelijk door hun ML projecten.
- Biedt veel kleine gerichte programma's of programma modes voor specifieke subtaken van een ML project.
- Stelt machine learning taken af op basis van de vereiste input en output die moet worden

gegenereerd.

Een paar voorbeelden van ML tools voor een command line interface zijn:

- Waffles
- Weka Machine Learning Workbench

3. Application Programming Interface(API)

Machine learning tools kunnen u ook een API aanbieden waarmee u zelf kunt bepalen welke elementen u moet gebruiken en hoe u ze precies kunt gebruiken binnen uw eigen programma's. De voordelen voor API zijn:

- Je kan machine learning integreren in eigen software projecten.
- Hiermee kan je jouw eigen machine learning tools maken.
- Biedt flexibiliteit om eigen processen en automatiseringen te gebruiken voor machine learning projecten.
- Laat toe om uw eigen methoden te combineren met die van de libraries.

Voorbeelden van machine learning tools met API

- Pylearn2 (python)
- Deeplearning4j (java)
- LIBSVM (C)
- TensorFlow(Java, Go, C)

3.2.2 Local versus Remote Machine Learning Tools

Een laatste manier om machine learning tools te vergelijken is door na te gaan of de tool local of remote is. Een local tool is een tool die u lokaal kunt downloaden, installeren en gebruiken, waarbij als externe tool op een externe server wordt uitgevoerd. Dit onderscheid kan ook een beetje troebel zijn omdat sommige tools op een lokale of externe manier kunnen worden uitgevoerd. En uiteindelijk kunt u bijna elk tool als een gehoste oplossing op uw eigen servers configureren. Toch kan dit nog een nuttig onderscheid zijn.

1. Local Tools

Een lokale tool wordt dus zoals de naam zegt, lokaal gedraaid.

- Het wordt op maat gemaakt voor gegevens en algoritmen in het geheugen.
- Je hebt controle over runconfiguratie parameterization.
- U kan het integreren in uw eigen systeem om aan uw behoeften te voldoen.

Voorbeelden van lokale tools zijn:

Shogun Library (C++)
GoLearn (Go)

2. Remote Tools

Een remote tool wordt gehost op een server en wordt opgeroepen vanuit een lokale omgeving. Deze tools worden vaak Machine Learning as a Service (MLaaS) genoemd.

- Speciaal gemaakt voor uitbereiding, kan worden uitgevoerd op grotere datasets.

3.2 Welke frameworks, visuele tools en scripting tools heb ik nodig om zelf zo een oplossing te bouwen? 35

- Werkt op meerdere systemen, meerdere kernen en shared memory.
 - Minder algoritmen vanwege de aanpassingen die nodig zijn om op schaal te werken.
 - Eenvoudigere interfaces , waardoor er minder controle is over run configuratie en parameter aanpassingen van algoritmen.
 - Geïntegreerd in uw lokale omgeving via remote procedure calls.
- Voorbeelden van remote tools:
- Google Prediction API
 - AWS Machine Learning
 - Microsoft Azure Machine Learning.

3.2.3 Samenvatting

In dit deel hebben we ontdekt waarom tools dus zo belangrijk zijn als men aan machine learning wil doen. Zonder al deze handige ML tools zouden we al deze technieken zelf moeten maken en implementeren, wat een zeer goede expertise vereist. In dit stuk heeft men ook aangetoond dat er verschillende soorten ML tools zijn die men kan onderscheiden.

- Platforms versus Libraries
- Graphical User Interfaces versus Command-Line Interface Versus Application Programming Interfaces
- Local versus Remote

3.2.4 Conclusie

Zoals men kan zien zijn er zeer veel tools waaruit men kan kiezen. Met elk hun sterktes en zwaktes, elke tool dat werd opgenoemd heeft men niet apart uitgelegd vanwege het aantal. Hieruit kunnen we dus goed beslissen welke soort tool we nodig zullen hebben. En vervolgens daaruit dan selecteren welke tool we uiteindelijk gaan gebruiken.

Platform of libraries?

een platform is een kant en klare tool eigenlijk wat niet veel ruimte laat voor eigen modificaties. Wat ons kan belemmeren. Bij deze kiest men voor libraries

GUI, Command Line of API?

Alhoewel het gebruik van GUI tools makkelijker is voor beginners, gaat men toch kiezen voor API. Kiest men toch voor API. De API laat toe om zelf een machine learning tool te maken. En dat is uiteindelijk wat men wou bereiken in deze bachelorproef. **Local of remote ?**

Dit maakt niet veel uit voor deze bachelorproef. Men gaat hier gaan voor hetgeen wat men het eenvoudigste vindt. Local.

De tool definitie die we hebben gekozen is : Library , API, Local Na wat zoeken op het web heeft men al snel een paar tools gevonden die aan onze definitie voldoet.

3.2.5 Tensorflow

Algemene informatie over tensorflow

3.2.6 Deeplearning4j

Algemene informatie over Deeplearning4j

3.2.7 Conclusie

vergelijken

3.3 Welke Dataset gaan we gebruiken en hoe verkrijg ik deze data?

De manier waarop men deze vraag moet interpreteren is eigenlijk welke parameters/variabelen hebben we zeker nodig in een data set om churn te kunnen voorspellen? Eerst en vooral moet men besluiten of er gaat gewerkt worden met supervised of unsupervised learning. In de literatuur studie heeft men hierover reeds het verschil uitgelegd.

Supervised Machine Learning is wanneer we input variabelen en output variabelen hebben. Het doel is dat men de functie zo willen uitvoeren dat we met een nieuwe input waarde de output waarde kunnen voorspellen

Unsupervised Machine Learning is wanneer men enkel input waarden heeft zonder de output waarden. Het doel hiervan is dat men meer over de input data te weten komt.

Aangezien we niet onmiddellijk meer informatie willen over de input waarden maar wel willen voorspellen welke de output zal zijn, namelijk of een klant gaat churnen ,ja of neen ,1 of 0. Zou men graag een data set willen waar het resultaat ook aanwezig is in de data set.

Dus men kan al één conclusie trekken. De dataset moet een label hebben. In de vorm van Churn ja of neen.

Voor aan gesuperviseerd leren te doen moet men niet enkel output hebben. Daarmee zijn we niet veel. Er moet ook nog input aanwezig zijn. Deze input is eigenlijk alle data wat we weten over deze ene klant. Elke bedrijf segmenteert zijn klanten. In de literatuur studie ziet men hoe men klanten moet gaan onderverdelen. Dankzij model gebaseerde segmentatie kan men makkelijk heel veel variabelen als input meegeven. Betalingsmethoden, tijd, locatie, prijs, hoeveelheid gekocht, korting, het weer, enz. . .

3.3.1 Telco-Customer-data set

Deze dataset bevat info die ons kan helpen om het gedrag van klanten te gaan voorspellen. Deze data set heeft men gevonden op de IBM Watson Analytics site in een tutorial. In de tutorial gaat het als volgt. Een telecom bedrijf is bezorgd over het aantal klanten dat ze

verliezen aan hun concurrentie. Ze willen te weten komen wie er gaat vertrekken. In het voorbeeld op de site moet je u inbeelden dat jij de analist bent in hun bedrijf en dat jij dit moet uitzoeken.

CustomerID	7590-VHVEG
Gender	Female
seniorCitizen	0
partner	Yes
dependents	No
tenure	1
phoneService	No
MultipleLines	No
internetService	phone service
OnlineSecurity	DSL
OnlineBackup	No
DeviceProtection	Yes
TechSupport	No
StreamingTV	No
StreamingMovies	No
Contract	Month-to-month
PaperlessBilling	Yes
paymentMethod	Electronic check
MonthlyCharges	29,85
TotalCharges	29,85
Churn	No

De data set heeft informatie over

- De klanten die het bedrijf zijn verlaten. De laatste kolom Churn.
- Services waar elke klant voor geabonneerd is, phones,multiple lines, internet,online security,online backup,device,protection,tech support, streaming TV en movies.
- -Profiel informatie van de klant, hoelang ze al klant zijn,contract,payment method,paperless billing, monthly charges, en total charges
- -Demographic info over de klanten gender, age, range en of ze partners en dependents hebben.

Hier ziet men eigenlijk al goed dat deze data set een tal van onze gewenste variabelen bevat. Maar we moeten onze dataset nog opschonen. Dit komt voor in de proof of concept.

3.4 Welk Model dien ik hiervoor op te zetten?

3.5 Kies ik beter scripting tools zoals python?

4. Proof of Concept

4.1 specificaties van mijn systeem

4.2 installatie van TensorFlow

Voor men begon met de proof of concept moet men eerst TensorFlow installeren. Op de site van TensorFlow staat er een handleiding hoe men TensorFlow moet installeren, en wat voor versie van TensorFlow men wilt installeren. Men kan kiezen tussen 2 versies: TensorFlow met enkel CPU support Dit is voor de systemen die geen NVIDIA GPU hebben. Dit is een versie die makkelijker is om te installeren en men raadt ons aan om toch deze versie te installeren ook al hebben we een NVIDIA GPU TensorFlow met GPU support. De TensorFlow programma's werken sneller op hun GPU dan op een CPU, daarvoor als het systeem een NVIDIA GPU heeft met de nodige voorwaarden die onderaan ook vermeld staan en als men performance-critical applications moeten draaien, dan moeten we deze versie toch installeren.

- CUDA Toolkit 9.0, je moet ook de nodige Cuba pathnames in de environment variabele toevoegen
- NVIDIA drivers die gelinkt zijn aan de CUDE Toolkit 9.0
- cuDNN v7.0

5. Bronnen

5.1 bronnen

<https://teamloyalty.nl/kennis/wat-is-klantenbinding/>

<https://hbr.org/2010/07/stop-trying-to-delight-your-customers>

<https://www.superoffice.com/blog/reduce-customer-churn/>

<https://medium.com/machine-learning-for-humans/why-machine-learning-matters-6164faf>

https://www.tensorflow.org/get_started/get_started_for_beginners <https://en.wikipedia.org/>

<https://medium.com/big-data-driving-innovation/reviewing-data-science-tool-dataiku->

<https://cdn2.hubspot.net/hubfs/2123903/PDF/Whitepaper/guidebook-churn-prediction.pdf>

<https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights>

<http://www.ronaldvanloon.com/machine-learning-explained-understanding-supervised-un>

<https://www.dataiku.com/dss/features/machine-learning/>

6. Conclusie

Curabitur nunc magna, posuere eget, venenatis eu, vehicula ac, velit. Aenean ornare, massa a accumsan pulvinar, quam lorem laoreet purus, eu sodales magna risus molestie lorem. Nunc erat velit, hendrerit quis, malesuada ut, aliquam vitae, wisi. Sed posuere. Suspendisse ipsum arcu, scelerisque nec, aliquam eu, molestie tincidunt, justo. Phasellus iaculis. Sed posuere lorem non ipsum. Pellentesque dapibus. Suspendisse quam libero, laoreet a, tincidunt eget, consequat at, est. Nullam ut lectus non enim consequat facilisis. Mauris leo. Quisque pede ligula, auctor vel, pellentesque vel, posuere id, turpis. Cras ipsum sem, cursus et, facilisis ut, tempus euismod, quam. Suspendisse tristique dolor eu orci. Mauris mattis. Aenean semper. Vivamus tortor magna, facilisis id, varius mattis, hendrerit in, justo. Integer purus.

Vivamus adipiscing. Curabitur imperdiet tempus turpis. Vivamus sapien dolor, congue venenatis, euismod eget, porta rhoncus, magna. Proin condimentum pretium enim. Fusce fringilla, libero et venenatis facilisis, eros enim cursus arcu, vitae facilisis odio augue vitae orci. Aliquam varius nibh ut odio. Sed condimentum condimentum nunc. Pellentesque eget massa. Pellentesque quis mauris. Donec ut ligula ac pede pulvinar lobortis. Pellentesque euismod. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent elit. Ut laoreet ornare est. Phasellus gravida vulputate nulla. Donec sit amet arcu ut sem tempor malesuada. Praesent hendrerit augue in urna. Proin enim ante, ornare vel, consequat ut, blandit in, justo. Donec felis elit, dignissim sed, sagittis ut, ullamcorper a, nulla. Aenean pharetra vulputate odio.

Quisque enim. Proin velit neque, tristique eu, eleifend eget, vestibulum nec, lacus. Vivamus odio. Duis odio urna, vehicula in, elementum aliquam, aliquet laoreet, tellus. Sed velit. Sed vel mi ac elit aliquet interdum. Etiam sapien neque, convallis et, aliquet vel, auctor non, arcu. Aliquam suscipit aliquam lectus. Proin tincidunt magna sed wisi. Integer blandit

lacus ut lorem. Sed luctus justo sed enim.

Morbi malesuada hendrerit dui. Nunc mauris leo, dapibus sit amet, vestibulum et, commodo id, est. Pellentesque purus. Pellentesque tristique, nunc ac pulvinar adipiscing, justo eros consequat lectus, sit amet posuere lectus neque vel augue. Cras consectetur libero ac eros. Ut eget massa. Fusce sit amet enim eleifend sem dictum auctor. In eget risus luctus wisi convallis pulvinar. Vivamus sapien risus, tempor in, viverra in, aliquet pellentesque, eros. Aliquam euismod libero a sem.

Nunc velit augue, scelerisque dignissim, lobortis et, aliquam in, risus. In eu eros. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Curabitur vulputate elit viverra augue. Mauris fringilla, tortor sit amet malesuada mollis, sapien mi dapibus odio, ac imperdiet ligula enim eget nisl. Quisque vitae pede a pede aliquet suscipit. Phasellus tellus pede, viverra vestibulum, gravida id, laoreet in, justo. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Integer commodo luctus lectus. Mauris justo. Duis varius eros. Sed quam. Cras lacus eros, rutrum eget, varius quis, convallis iaculis, velit. Mauris imperdiet, metus at tristique venenatis, purus neque pellentesque mauris, a ultrices elit lacus nec tortor. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent malesuada. Nam lacus lectus, auctor sit amet, malesuada vel, elementum eget, metus. Duis neque pede, facilisis eget, egestas elementum, nonummy id, neque.

A. Onderzoeksvoorstel

Het onderwerp van deze bachelorproef is gebaseerd op een onderzoeksvoorstel dat vooraf werd beoordeeld door de promotor. Dat voorstel is opgenomen in deze bijlage.

A.1 Introductie

Hier introduceer je werk. Je hoeft hier nog niet te technisch te gaan.

Je beschrijft zeker:

- de probleemstelling en context
- de motivatie en relevantie voor het onderzoek
- de doelstelling en onderzoeksvraag/-vragen

A.2 State-of-the-art

Hier beschrijf je de *state-of-the-art* rondom je gekozen onderzoeksdomein. Dit kan bijvoorbeeld een literatuurstudie zijn. Je mag de titel van deze sectie ook aanpassen (literatuurstudie, stand van zaken, enz.). Zijn er al gelijkaardige onderzoeken gevoerd? Wat concluderen ze? Wat is het verschil met jouw onderzoek? Wat is de relevantie met jouw onderzoek?

Verwijs bij elke introductie van een term of bewering over het domein naar de vakliteratuur, bijvoorbeeld (**Doll1954**)! Denk zeker goed na welke werken je refereert en waarom.

Je mag gerust gebruik maken van subsecties in dit onderdeel.

A.3 Methodologie

Hier beschrijf je hoe je van plan bent het onderzoek te voeren. Welke onderzoekstechniek ga je toepassen om elk van je onderzoeksvragen te beantwoorden? Gebruik je hiervoor experimenten, vragenlijsten, simulaties? Je beschrijft ook al welke tools je denkt hiervoor te gebruiken of te ontwikkelen.

A.4 Verwachte resultaten

Hier beschrijf je welke resultaten je verwacht. Als je metingen en simulaties uitvoert, kan je hier al mock-ups maken van de grafieken samen met de verwachte conclusies. Benoem zeker al je assen en de stukken van de grafiek die je gaat gebruiken. Dit zorgt ervoor dat je concreet weet hoe je je data gaat moeten structureren.

A.5 Verwachte conclusies

Hier beschrijf je wat je verwacht uit je onderzoek, met de motivatie waarom. Het is **niet** erg indien uit je onderzoek andere resultaten en conclusies vloeien dan dat je hier beschrijft: het is dan juist interessant om te onderzoeken waarom jouw hypothesen niet overeenkomen met de resultaten.