Universiteit
Leiden
The Netherlands

Applying data mining in telecommunications

Radosavljevik, D.

Citation
Radosavljevik, D. (2017, December 11). *Applying data mining in telecommunications*. Retrieved from https://hdl.handle.net/1887/57982

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/57982

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/57982 holds various files of this Leiden University dissertation.

Author: Radosavljevik, D.
Title: Applying data mining in telecommunications
Issue Date: 2017-12-11

Samenvatting

Vanwege de digitale revolutie in de afgelopen decennia is er veel data beschikbaar. De overweldigende aanwezigheid van digitale apparaten in combinatie met platformen die het genereren en opslaan van data mogelijk maken, heeft geleid tot hoeveelheden data die in het verleden moeilijk voorstelbaar waren. Volgens IBM (2017) is 90% van de vandaag beschikbare data gegenereerd in de afgelopen twee jaar. Daarnaast worden dagelijks 2,5 triljoen bytes aan data gegenereerd en dit cijfer zal waarschijnlijk groeien, gezien de opkomst van nieuwe technologieën, apparaten en sensoren. Dit roept de vraag op welk deel van die data relevant of waardevol is. Het verkrijgen van waarde uit deze overvloed aan data is van belang voor zowel de industrie als de academische wereld.

Datamining wordt in het kort gedefinieerd als het proces van het ontdekken van patronen in data (Witten and Frank, 2005). De ontdekte patronen moeten zinvol zijn en tot een voordeel leiden, bijvoorbeeld een economisch voordeel. Grant (2003), geeft een meer gedetailleerde definitie van datamining als "een interdisciplinair veld met technieken zoals machine learning, patroonherkenning, statistieken, databases en data-visualisatie om de problemen aan te pakken met betrekking tot informatie-extractie uit grote databases."

Dit proefschrift past datamining toe in commerciële omgevingen in de telecommunicatie-industrie. Het onderzoek voor dit proefschrift is uitgevoerd bij T-Mobile Nederland B.V. en de methoden beschreven in sommige hoofdstukken zijn ook toegepast in dochterondernemingen van Deutsche Telekom in andere landen. We waren in de gelegenheid om aan echte commerciële datasets te werken en de resultaten van ons onderzoek in de praktijk toe te passen. In dit proefschrift beschrijven we enkele van de uitdagingen die data miners (of data scientists) tegenkomen bij het werken aan bedrijfsproblemen en we geven onze oplossingen hiervoor. De complexe datasets die we aan het analyseren waren bevatten in bepaalde gevallen tot honderden miljoenen records. In dit onderzoek gebruikten we datamining methoden op een innovatieve manier om resultaten te bereiken die ofwel een verbetering waren van hoe het bedrijf eerder deze problemen oploste ofwel belangrijke bedrijfsproblemen oplosten die tot nu toe niet eerder waren behandeld op zo'n gedetailleerde manier.

De inspiratie voor dit proefschrift is onze overtuiging dat het de verantwoorde-

lijkheid van de academische wereld is om niet alleen nieuwe manieren te vinden om problemen op te lossen, maar ook om het bedrijfsleven en de overheid te leren hoe ze deze ontwikkelingen kunnen gebruiken. Dit is een probleem in het geval van machine learning, waarbij de technologiereuzen van vandaag (bijvoorbeeld Facebook, Google of de NSA) op grote schaal gebruik maken van modern machine learning-onderzoek, terwijl de rest van de bedrijven achterblijft. We willen de toewijding van onderzoekers aan nieuwe en verbeterde algoritmen voor machine learning niet bekritiseren, maar hen juist aanmoedigen om een stap verder te gaan met de implementatie van deze methoden in praktische omstandigheden (verspreiding in gebruik van hun innovatie), waardoor het volledige potentieel van hun onderzoek wordt gerealiseerd. Om de hierboven genoemde redenen is dit proefschrift gericht op het toepassen van datamining in de praktijk in de telecommunicatie branche.

Churn, d.w.z. het verliezen van een klant aan de concurrentie, is een groot probleem in mobiele telecommunicatie en vele andere industrieën. Daarom wijden we drie hoofdstukken aan dit probleem.

In hoofdstuk 2 bespreken we de impact van de experimentele opzet op de voorspelling van prepaid churn in telecommunicatie. Prepaid klanten in mobiele telecommunicatie zijn niet aan een contract gebonden en kunnen daarom zonder toestemming van operator wijzigen (”churn”). Dit maakt het voorspellen van churn zowel uitdagend als financieel lonend. Het hoofdstuk presenteert een verkennend onderzoek naar prepaid churn-modellering door de experimentele opstelling te variëren op drie dimensies: data, doelvariabele definitie en steekproef.

In hoofdstuk 3 onderzoeken we de toegevoegde waarde van het combineren van reguliere tabulaire datamining met social netwerkmining, waarbij gebruik wordt gemaakt van de graaf gevormd uit communicatie tussen klanten. Hier vergeleken we de prestaties van klassieke (tabulaire) prepaid churn-modellen en klassieke sociaal netwerk modellen met twee hybride modellen. Ten eerste hebben we de dataset, die voor de tabelmodellen wordt gebruikt, verrijkt met variabelen uit de communicatiegraaf. Ten tweede creëerden we een propagatiemodel met behulp van de scores van de tabulaire churn-modellen als initiële energie van elk niet-churner-knooppunt (vergelijkbaar met boosting).

In de zeer concurrerende en geavanceerde telecommunicatiemarkt in Nederland is netwerkervaring cruciaal voor de operators. Dit is de reden waarom de hoofdstukken 4 en 5 focussen op twee verschillende manieren om het netwerk te verbeteren.

Hoofdstuk 4 beschrijft een andere toepassing van een churn-model. Dit hoofdstuk schetst een benadering die is ontwikkeld als onderdeel van een bedrijfsbreed initiatief voor churn-management. Onze aanpak van churn-preventie kan ook gezien worden als een brug tussen de disciplines marketing en mobiele netwerktechnologie, omdat we de technische oorzaken van churn identificeerden. De typische implementatiemethode voor een churn-model is een retentiecampagne waarbij klanten worden benaderd met een aanbod om hun contract voort te zetten. In dit geval was er geen campagne. Het model werd gebruikt om een reeks regels te genereren voor net-

werkoptimalisatie om de belangrijkste netwerk-gerelateerde oorzaken van churn te verwijderen en daarmee churn te voorkomen in plaats van te genezen.

In hoofdstuk 5 presenteerden we een zeer eenvoudige maar effectieve benadering van het gebruik van datamining in een commerciële omgeving. Hiervoor passen we de uitkomsten van datamining, een grote verzameling van modellen, in een simulatieraamwerk, dat niet dataminers maar domeinexperts in staat stelt niet alleen de toekomst te voorspellen, maar ook om simulaties te doen onder verschillende scenario's en condities. Na het eerste succes in T-Mobile Netherlands, waar de methode werd ontwikkeld, werd de aanpak ook gebruikt door operators van Deutsche Telekom in vier andere landen.

In hoofdstuk 6 breidden we de methode die is ontwikkeld in hoofdstuk 5 uit tot het gebied van financiën en het voorspellen van de service inkomsten. Inkomsten voorspellen in het algemeen is een van de belangrijkste financiële processen in elke bedrijfstak. Voor op diensten gebaseerde activiteiten, zoals telecommunicatie, zijn tijdige en nauwkeurige voorspellingen voor service-inkomsten essentieel, omdat ze belangrijke zakelijke beslissingen kunnen sturen, zoals wanneer en waar men kan ingrijpen om de zakelijke doelstellingen te bereiken. Door een simulatieplatform voor eindgebruikers te maken voor het voorspellen van de bedrijfsinkomsten, kan het bedrijf een beter idee krijgen van de manier waarop verschillende scenario's voor input parameters of geplande maatregelen in de praktijk kunnen uitwerken. De belangrijkste aspecten van de implementatie zijn vergelijkbaar met hoofdstuk 5: gebruik van hulpmiddelen die bekend zijn bij eindgebruikers en creatie van een platform voor simulatie van scenario's.

Vanuit een generalisatieperspectief is dit proefschrift op vele industrieën van toepassing. Het is duidelijk dat in vrijwel elke sector het verliezen van klanten aan de concurrentie (churn), het beheren van de belangrijkste resources en het voorspellen van de inkomsten erg belangrijke problemen zijn. Het vinden van kostenefficiënte middelen om deze problemen aan te pakken, is gunstig voor veel bedrijven of overheden. Het gebruik van eenvoudige algoritmen die gemakkelijk kunnen worden uitgelegd en het gebruik van implementatiemethoden die de voorkeur hebben van eindgebruikers, helpt bij de acceptatie. Dit is hoe we datamining zien worden toegepast en geaccepteerd op veel gebieden buiten de telecommunicatie, waardoor meer organisaties data- en klantgedreven worden. Vanuit een onderzoeksperspectief hopen we dat we hebben laten zien dat er veel andere interessante problemen zijn om op te lossen dan alleen het ontwikkelen van een beter voorspellend model. Hopelijk zullen we door het toepassen van datamining in telecommunicatie de academische belangstelling vergroten om betere en efficiëntere manieren te vinden voor een snellere verspreiding van machine learning onderzoek.