Literature review

TODO: W części 02 z literature review dałabym ilustrację celów, metod i dotychczasowych badań (trochę jak w tym artykule o metodach w CRM), żeby uwypuklić o czym w ogóle mówimy.

TODO: Coś dać o tym że XAI jest ważne dla rozpowszechniania używania modelu i predykcji - tutaj link do jakiegoś badania nt. kultury w organizacji po wprowadzeniu ML: <https://www.bastagroup.nl/wp-content/uploads/2019/01/the-state-of-machine-learning-adoption-in-the-enterprise.pdf>

# Customer churn

Paragrafy do dodania ogólnie:

* ogólnie o zjawisku churnu:
  + definicja churnu
  + zalety posiadania lojalnych klientóW
  + mechanizm churn i retention - większość researchu dla subscription based services, dla retail mało a dla online retail bardzo mało. Coś o tym że klienci w internecie są mniej lojalni (jeżeli to prawda)
  + if we can predict which customers are likely to leave, we can target them with some customer attraction tools
* Churn prediction methods
  + definicja churn jako zmiennej w modelowaniu
  + “churn problem can be decomposed as choice of ML model and choice of model of variable influences”
  + review algorytmów wykorzystywanych - zacytować kilka prac o tym czego używali, ale też napisać o no free lunch theorem i że powinno się zawsze sprawdzić na swoich danych. Do tego może coś o trade offie pomiędzy LR i XGBoost - o explainability
  + review zmiennych wykorzystywanych

### General about customer retention

#### Churn in CRM

Customer Relationship Management is defined as a process, in which the business manages its interactions with the customers using data integration from various sources and data analysis (Bardicchia 2020). Oliveira (2012) provides 4 areas in which CRM approaches can be of use:

* + Customer identification (aquisition) - determining who can be a potential customer?
  + Customer attraction - how can one make this person a customer?
  + Customer development - how can one make a customer more profitable?
  + Customer retention - how can one make the customer stay with the company?

The last one is the main focus of this study.

#### Why having loyal clients is important

Improving the loyalty of customer base is profitable to the company. This has its source in multiple factors, the most important one being cost of aquisition. Multiple studies have shown that retaining customers costs less than attracting new ones (Dick and Basu 1994; Gefen 2002; Buckinx and Poel 2005). Moreover, there is some evidence that loyal customers are less sensitive to the competitor’s actions regarding price changes (Achrol and Kotler 1999; Choi et al. 2006).

#### Difference with tackling potential churners

There are 2 basic categories of approaches for the company to deal with customer churn. First category is “untargeted” approach. The company seeks to improve its product quality and relies on mass advertising to reduce the churn. The other way is “targeted” approach - the company tries to address aim their marketing campaigns at the customers that are more likely to churn (Burez and Poel 2007). This approach can be divided further, by the way in which targeted customers are chosen. The company can target only these ones that have already made a decision to resign from further relationship. For example, in contractual settings this can mean canceling the subscription or breaching the contract.

The other way to approach churn problem is to try to predict, which customers are likely to churn in the near future. This has an advantage of having lower cost, as the customers that are about to leave are likely to have higher demands from the last-minute deal proposed to them (Tamaddoni Jahromi et al. 2010).

###### Non-contractual churn prediction

As pointed out by Tamaddoni Jahromi et al. (2010) in their literature review, most of the studies concerning churn prediction were done in contractual settings. In other words, churn was defined as the client resigning from using the company’s services more by canceling the subscription or breaching the contract. Such way to specify the churn is different from the businesses in which the customer doesn’t have to inform the company about resigning.

One problem that arises in non-contractual setting is the definition of churn. As there is no clear moment that the customer decides not to use the company’s services anymore, it has to be specified by the researcher based on the goals that one has to achieve from the churn analysis. Oliveira (2012) defined partial churners as the customers not making new purchases in the retail shop for the next 3 months. A very different approach was used by Buckinx and Poel (2005). All the customers that had frequency of purchases below average were treated as “churners,” since these customers were shown to provide little value to the company.

#### The difference between one-shot and multiple purchases shopping

* + do tej pory nie widziałem takiej pracy która o tym mówiła

### Customer churn prediction

#### Why churn prediction makes sense

If the company is able to successfully predict, which customers are most likely to leave, it can target them with retention-focused campaign. Contrary to targeting all of the customers with such campaign, focusing on the customers that are most likely to leave leads to reduction of the cost of the campaign.

#### Why classification vs. ranking

Churn prediction fits well with the framework of classification, as the variable that one would like to predict is binary (churn-no churn). However, not only such binary prediction is valuable for later retention campaign efforts. As noted by Wai-Ho Au, Chan, and Xin Yao (2003), equally important is that the machine learning model can predict the likelihood of the customer leaving. After such prediction, the customers can be ranked from the most to the least likely to churn.

This has two benefits. First, the company can decide what percentage of the customers to target in the retention campaign, and not be bound by how many customers the model will predict as potential churners. Second, the company can decide how strong the targeting should be based on the likelihood to leave. For example, based on cost-benefit analysis of variuos targeting approaches, one could decide that for top 10% of the most likely to leave customers the company should offer big discounts for the next purchase, while for top 30% - only send an encouraging email.

#### Challenges to be addressed

Churn prediction task can be decomposed into 2 main important aspects that one has to tackle. First is the decision about specific Machine Learning model that gives the best performance. Second is deciding on the model formula - in other words, deciding about which variables should be included in the model and what should be the form of the relationship.

#### Review of previous churn studies for models used

In previous studies multiple machine learning algorithms for prediction were used in churn setting (for overview see Verbeke et al. (2011)). The two most widely used techniques are Logistic Regression (LR) and Decision Trees.

An important feature of both of them is that they are relatively simple, and because of that the way they make predictions can be assessed by a qualified expert (Paruelo and Tomasel 1997). However, these two methods often give suboptimal results compared to more advanced and recent approaches like Neural Networks or Random Forests (Murthy 1998; Oliveira 2012). Moreover, this was shown not only in the case of churn prediction setting, but also in more general benchmarks that used multiple datasets and comparison metrics (Caruana and Niculescu-Mizil 2006).

Recently, XGBoost algorithm (Chen et al. 2015) is gaining popularity in multiple domains in which one faces prediction tasks. XGBoost main strengths are ability to infer non-linear relationships from the data, and relative speed, which allows the researcher to try out multiple hyperparameters and decide on the best ones (Chen et al. 2015). Because of that, it is considered as a go-to standard for machine learning challenges, and very often solutions based on it achieve the best results in various competitions and benchmarks (hcho3 2020). In the context of churn prediction, XGBoost was used by Gregory (2018). It achieved superior performance compared to other techniques, like Logistic Regression and Random Forests.

##### Review of previous churn studies for variables/categories of variables used

Previous churn prediction studies used a variety of variables to include in the model formulation. Buckinx and Van den Poel (2005) divided them into 3 broad categories. These are behavioural, demographic and perception features.

#### Behavioural

First category of variables tries to describe how the customer has interacted with the company before. Typical features belonging to this category are recency, frequency and monetary value, which constitute a basis of RFM customer segmentation framework. These features are used in multiple studies (Oliveira 2012; Bhattacharya 1998), and typically accompany the more complex variables that are main focus of particular studies.

Besides these basic features, some studies focus on other areas of customer behaviour. Particular products that the customer has bought in previous purchases were shown to be useful for churn prediction (Buckinx and Van den Poel 2005; Athanassopoulos 2000).

Some of the variables studied previously are avaliable only in some of the domains. For example, one possibility of analyzing customer behaviour in the context of e-commerce shopping is analyzing the customer behaviour while interacting with the company’s website. Koehn, Lessmann, and Schaal (2020) analyzed such click-stream sequential data using Recurrent Neural Networks. They found that such information can serve as a good predictor of customer churn. Similar studies were performed by Berger and Kompan (2019) and Yu et al. (2011), however the click stream data was aggregated and preprocessed manually, instead of using sequential modeling approach.

#### Demographics

Second category of features used in churn prediction are demographic variables about the customer, such as age, gender or address. Such variables were shown to be good predictors of customer churn in multiple studies (For overview see Verbeke et al. (2011)). However, availability of such predictors to use in modeling is very often limited from multiple reasons. In non-contractual settings, customers don’t have to always provide such data to the company. Moreover, usage of such personal data can be in some cases considered unethical, and lead to predictions biased against particular age or gender.

Another way to include demographics data in the churn prediction model was shown by Zhao et al. (2005). They used the data obtained from the statistical office for particular regions that the customer is residing in. Such variables were shown to be useful in churn prediction.

#### Perception

Last category of variables used for churn prediction specified by Buckinx and Van den Poel (2005) is customer perception about the company. According to Kracklauer, Passenheim, and Seifert (2001), customer satisfaction is the most important factor driving customer retention. However, although such features could have potentially high predictive power, they are usually hard to observe and quantify meaningfuly. The one possible approach is asking the customers for direct feedback using questionaires or providing a way to post a review on the purchase. This kind of feedback can be obtained in different forms, one of them being a textual reviews. Couple of previous studies were aimed at extracting meaningful features from such reviews using different text mining methods. De Caigny et al. (2020) have used text embedding approach, while Suryadi (2020) - simple tf-idf technique. In both studies the results using such methods were superior compared to the models without including such information.

#### do methods:

* + do każdego podpunktu jak jest definiowany, jakie są wcześniejsze prace które go używają, jaka jest teoria za tym, hipoteza, opis przetwarzania w mojej pracy

Behavioural:

* + transaction value
  + no items
  + categories of products bought

Demographic:

* + wszystko z sidra
  + geoloc
  + dbscan

Perception:

* + score 1-5
  + text review

### Churn analysis do śmieci?

The most low-hanging fruit for the companies that want to start basing their business desicions on the data is usage of transaction-level data. That is because virtually every e-commerce shop is based on the the mechanism of user registration, and storing the client’s purchasing history is an industry standard.

!!! The data only about when the customer made purchases and how much did he pay are very easily translated into the framework of Recency-Frequency-Monetary value. Multiple works (Aleksandrova (2018), Yanfang and Chen (2017)) demonstrated that such data can serve as a good input to churn prediction machine learning model. In fact, most of the publications presented in this review is using RFM variables as one part of the dataset, while including more complex, engineered variables as the other part.

After the first purchase of the customer in the e-commerce shop, their exact adresses can be inferred with high probability. Usually the delivery adress would be to the home of a customer, or in worse cases to other place that the customer visits (like workplace etc.). Zhao et al. (2005) used this kind of customer location data to enrich the dataset with basic spatial characteristics of the region, that is geographic situation and demographic variables.

### Reviews analysis

TODO: do zastanowienia jakie powinny być więcej częsci w lit review, poprawić strukturę tekstu

An important source of knowledge about e-commerce customers are textual reviews. They can serve as a rich source of feedback for what in the shop or product is liked and what needs change. Also, in the textual reviews one can get to know customer’s opinions way better then using other types of feedback, for example 1-5 rating of a purchase. With these advantages, they come at the expense of increased complexity of such analysis. A big challenge is to extract meaningful information from this type of highly unstructured data.

Two most important types of text mining in text reviews is *sentiment prediction* and *topic mining* (in the context of reviews also often called *aspect mining*). Topic modeling is particularly challenging, as usually one does not have a annotated dataset with topics assigned to each text. That is why an unsupervised approach usually has to be used.

A go-to model for inferring the topic of a text is Latent Dirchlet Allocation (Blei, Ng, and Jordan 2003). The method is based on assumption, that each document is a mixture of a small number of topics. At the same time, each topic can be characterized by a distribution of words frequency.

Unfortunately, LDA approach was created with different purpose in mind. Typically reviews in the aspect of e-commerce are very short. Hong and Davison (2010) showed that LDA is not able to find informative topics in Twitter posts. These posts are bound by the rules of the platform to be shorter than 280 characters long. Possible reason that LDA does not cope well is that assumption about a document being mixture of topics is false. Short texts probably comprise of very small amount of topics, usually only one.

The drawbacks of LDA in setting of short texts were adressed by (Yin and Wang 2014) . They used Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model, which is improvement over typical LDA. The algorithm used by the authors is called Movie Group Process. Short introduction to this algorithm is included below.

Imagine a movie discussion group. There are k tables, and the goal is to assign students to tables according to their similar movie taste. There are 2 preference parameters set for each student:

1. Choose a table with students having similar movie taste. This is meant to introduce homogeneity of the clusters.
2. Choose a table with more students in this group. This rule is meant to improve completeness - so to the clusters have a reasonably high number of members.

The authors show that this algorithm provides superior performance to vanilla LDA not only when the texts are short, but also in general.

In recent years completely new approaches to Natural Language Processing emerged, thanks to improvement in the area of Neural Network algorithms. Two approaches are especially important as they serve as a baseline for the most recent findings in aspect (topic) recognition area. These two are word vector representations and attention mechanism. A short introduction of these two methods is presented in the section below.

In 2013, word2vec (Mikolov et al. 2013) was presented. The goal of this method is to learn a meaningful vector representation of each word in a corpus. Word2vec’s approach is to train a model that predicts all of the neighboring words for every occurrence of every word in an entire body of text (a corpus).

Intuitively, suppose that the model needs to learn embeddings for 3 words: “king,” “queen,” “orange.” The points in the embedding space for the first two words should lay in the proximity, while “orange” should be further. Word2vec approach is to look at the probability, that given word should be placed in particular place in the sentence, given the neighboring words. Suppose we have an incomplete sentence “XXX were usually very rich in the past.” Word2vec tries to predict what XXX should be. From the corpus it should understand, that “king” and “queen” are more probable than “orange,” that is why puts the embeddings closer.

Creating word embeddings usually serve as a preprocessing phase for next analysis steps, as with the data in numeric form one can use all tools that conventional data analysis has to offer, not being limited anymore by the complicated nature of textual data.

Another concept very helpful in the aspect recognition domain is attention mechanism (Chorowski et al. 2015). It is based on attention mechanism in psychology. When a human is trying to understand any content (visual, textual etc.) she is not using all content in the same extent, but only the relevant parts. For example, when a car driver is making a decision whether to cross an intersection, from all the visual signals that she obtains at the moment, the most important (and the only one looked at) is whether the light is red or green.

This concept can be very useful in the area of aspect prediction, as usually only couple of words from the whole sentence show the topic of it.

He et al. (2017) presented an Attention-based Aspect Extraction model. At first, words embedding using Word2Vec model is created. After that, for each text in the corpus, attention weight for each word is computed using neural network with an attention layer. Then, embedding of the whole sentence is created by computing an average for all words embedding. The words are weighted by their attention weights. Last step of the procedure is creating encoder-decoder model for learning sentence aspect embedding. The reconstruction of the sentence is the linear combination of aspect embeddings, and aspect embeddings are learned by mapping sentence embedding to a lower dimensional space.

Another work worth mentioning is by Tulkens and Cranenburgh (2020), who proposed a new type of Attention mechanism, meant especially for aspect recognition task. It’s advantage over the one presented by (He et al. 2017) is that instead of a complex neural network, a way simpler approach based on Radial Basis Function kernel is used. Another work presenting new attention mechanism is by Luo et al. (2019) - they use a use a Encoder-Decoder framework with an *Semene Attention* mechanism.

Losowe papery:

Online reviews as a feedback mechanism for hotel CRM systems Creating a geodemographic classification model within geo-marketing: the case of Eskişehir province

### XAI

While deciding on the type of Machine Learning algorithm, one usually faces the explainability-performance trade-off (Nanayakkara et al. 2018). More flexible models, like boosting or neural networks, usually present superior performance to more basic approaches. On the other hand, their predictions cannot be explained as easily as in the case of for example Decision Trees.

co możesz napisać?

* Dlaczego
* AI jest black box
* co to XAI
* <https://www.kdnuggets.com/2018/10/enterprise-explainable-ai.html>

<https://blog.goodaudience.com/holy-grail-of-ai-for-enterprise-explainable-ai-xai-6e630902f2a0>

<https://www.sciencedirect.com/science/article/pii/S1094996820300888?casa_token=Odeol3YA4U4AAAAA:NCkaJ-yDb53CB6JWnezV-MZpcI6kHb2D_XpTapz7DDih72a6U3N3Kcxn28IIiUqrmOnSnOhf5Q> challenge dla marketingu przy stosowaniu AI

<https://arxiv.org/pdf/1810.00184.pdf>

xai z perspektywy końcowego usera

Luźne opisy paperów:

<https://www.preprints.org/manuscript/202106.0063/v1> Dokładnie ten sam dataset, preprint opublikowany 5.06 (sic!)

<https://arxiv.org/pdf/1703.03869.pdf> Bardzo dokładny opis podejścia do churn prediction z użyciem DL

<https://link.springer.com/chapter/10.1007/978-981-32-9563-6_11> (Jheng and Luo 2019) - retention prediction przez CNN

<https://www.sciencedirect.com/science/article/pii/S0019850116301651?casa_token=YCUcElM8k_EAAAAA:-qJeOGXh7u2pQlqj-eyAo9k-eLgbc-m31QsDURsmpD2CEIyqtUzAjYGXUwkQRR4T0MrtkIbeWtaG>

* + sam paper średni ale ma dużo referencji

<https://sci-hub.se/https://ieeexplore.ieee.org/abstract/document/7538581>

* + fancy metody imbalance
  + używają datasetu takiego jak ja
  + mega dobra dokładność
  + ALE najpierw robią upsampling a później oceniają performance na CV - błąd! u mnie to różnica pommiędzy 0.62 a 0.78 AUC

<http://flr-journal.org/index.php/mse/article/view/10816/11113>

* + wykorzystują social network userów

<https://link.springer.com/article/10.1007/s10660-019-09383-2>

* + wykorzystują deep learning do imbalance - chyba lepiej będzie pasować do innej sekcji

<https://www.emerald.com/insight/content/doi/10.1108/17515631011063767/full/html>

dużo references

A study on factors affecting the purchasing process of online shopping: a survey in China & Japan

na podstawie kwestionariusza ocena satysfakcji

<https://d1wqtxts1xzle7.cloudfront.net/62198454/key-paper20200225-3623-15suux9.pdf?1582687757=&response-content-disposition=inline%3B+filename%3DWhat_Effects_Repurchase_Intention_of_Onl.pdf&Expires=1617187287&Signature=LfylLp7R2PXNPLFtVyCNdj~e4FhDBUz04-T152E7FSsNHjnqclWeFnnKf9C2fJskRN2q~sRx~CsXCbeuhn0zcrktL0lj8oN8GUxrWXpavIz1UaQuO~ayrylqfAH2XgIdwhDe~8FOoMNP9ZzaNz6lqYuy6DYaBNhP6G7N3sUo2spQ187dGOgRHgGafoS3Z7HZ2AgEUjgs1ldOsU1E7FXrP1delDpO7QYarp9h1euOUM6vCWCxlsDZYnRF6A-PIuQlgyP8QOyzMo2d487sDw0Jepwjrd69ocCrSMsi7dmu56Z00CUoXaUA3b~C9vyQrfYI9T1hzMcJYfQYri4lUWgblQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA>

ocena repurchase probability na podstawie kwestionariusza

<https://www.tandfonline.com/doi/abs/10.1080/08874417.2011.11645518?casa_token=33mj-Wcpw8IAAAAA%3AQUNps1MzrKJLyKf_c0Vl6gRIzoqI8wU3PbavgeTiyiJlQxwpHMi3JLUMmmGr7ZX0C2uqsrTT-TBYIw8&>

znowu kwestionariusz ale z modelem

<https://link.springer.com/article/10.1007/s10660-015-9207-2>

* + predicting repurchase intention
  + na podstawie kwestionariusza
  + ale prediction z fancy metodami

<https://ieeexplore.ieee.org/abstract/document/9325646> (Suryadi 2020)

* + repurchase jako kwestionariusz tak/nie
  + ale wykorzystują predykcję na podstawie reviews

(Ganesh, Arnold, and Reynolds 2000) definicja churn

A particularly prominent forecasting application in CRM is customer churn prediction (CCP), which is defined as a method of identifying customers who show a high inclination to abandon the company

<https://www.sciencedirect.com/science/article/pii/S0169207019301499?casa_token=kopLN0D45dwAAAAA:pARTYFQ1-0aho11qk4RpZdFdBIb1S-cJVHPb1iaggq41zU7pI-heeNpG9uK5cGThM7IWfFAkeGqU> (De Caigny et al. 2020)

-ładnie opisane profity z posiadania lojalnych customerów - dobry paper, dużo odniesień i wykorzystanie textual data

<https://www.sciencedirect.com/science/article/pii/S0957417410006779?casa_token=0C1SeJigqT8AAAAA:GCfX81AUr9p3ZfrqwTPCb23r4Slx6YijCvIOJE5xTcrxgl1nge7gjwvQnCo4c_r5fp1zaSigKjve> (Yu et al. 2011)

* + Jest o prawdziwym churnie a nie o retention
  + Jest złożona baza danych
  + minimalny wstęp o churn prediction

<https://ieeexplore.ieee.org/abstract/document/8627369> (Berger and Kompan 2019)

* + używają danych o sesji w przeglądarce

<https://cursa.ihmc.us/rid=1MYWPTN4Z-BBB2D6-30SB/Zhao_Churn_Prediction_SVM.pdf>

(Zhao et al. 2005) - Używa danych demograficznych - jest o churn

<https://ieeexplore.ieee.org/abstract/document/8284914> Yanfang and Chen (2017) - prawie nic nie ma ciekawego, tylko jako case

<https://ieeexplore.ieee.org/abstract/document/1255389> (Wai-Ho Au, Chan, and Xin Yao 2003)

* + Jest o tym że nie chodzi o predykcję tylko o ranking

##### References

(Tulkens and Cranenburgh 2020) CAT attention (Chorowski et al. 2015) word2vec (Mikolov et al. 2013) Main: (He et al. 2017) drawbacks of LDA (Hong and Davison 2010)

Achrol, Ravi S, and Philip Kotler. 1999. “Marketing in the Network Economy.” *Journal of Marketing* 63 (4\_suppl1): 146–63.

Aleksandrova, Yanka. 2018. “Application of Machine Learning for Churn Prediction Based on Transactional Data (RFM Analysis).” In *18 International Multidisciplinary Scientific Geoconference SGEM 2018: Conference Proceedings*, 18:125–32. 2.1.

Athanassopoulos, Antreas D. 2000. “Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior.” *Journal of Business Research* 47 (3): 191–207.

Bardicchia, Marco. 2020. Digital CRM-Strategies and Emerging Trends: Building Customer Relationship in the Digital Era.

Berger, P., and M. Kompan. 2019. “User Modeling for Churn Prediction in e-Commerce.” *IEEE Intelligent Systems* 34 (2): 44–52. <https://doi.org/10.1109/MIS.2019.2895788>.

Bhattacharya, CB. 1998. “When Customers Are Members: Customer Retention in Paid Membership Contexts.” *Journal of the Academy of Marketing Science* 26 (1): 31–44.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. “Latent Dirichlet Allocation.” *The Journal of Machine Learning Research* 3: 993–1022.

Buckinx, Wouter, and Dirk Van den Poel. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68.

———. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68.

Buckinx, Wouter, and Dirk Van den Poel. 2005. “Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting.” *European Journal of Operational Research* 164 (1): 252–68. https://doi.org/<https://doi.org/10.1016/j.ejor.2003.12.010>.

Burez, Jonathan, and Dirk Van den Poel. 2007. “CRM at a Pay-TV Company: Using Analytical Models to Reduce Customer Attrition by Targeted Marketing for Subscription Services.” *Expert Systems with Applications* 32 (2): 277–88.

Caruana, Rich, and Alexandru Niculescu-Mizil. 2006. “An Empirical Comparison of Supervised Learning Algorithms.” In *Proceedings of the 23rd International Conference on Machine Learning*, 161–68.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, and others. 2015. “Xgboost: Extreme Gradient Boosting.” *R Package Version 0.4-2* 1 (4): 1–4.

Choi, Duke Hyun, Chul Min Kim, Sang-Il Kim, and Soung Hie Kim. 2006. “Customer Loyalty and Disloyalty in Internet Retail Stores: Its Antecedents and Its Effect on Customer Price Sensitivity.” *International Journal of Management* 23 (4): 925.

Chorowski, Jan, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. 2015. “Attention-Based Models for Speech Recognition.” *arXiv Preprint arXiv:1506.07503*.

De Caigny, Arno, Kristof Coussement, Koen W. De Bock, and Stefan Lessmann. 2020. “Incorporating Textual Information in Customer Churn Prediction Models Based on a Convolutional Neural Network.” *International Journal of Forecasting* 36 (4): 1563–78. https://doi.org/<https://doi.org/10.1016/j.ijforecast.2019.03.029>.

Dick, Alan S, and Kunal Basu. 1994. “Customer Loyalty: Toward an Integrated Conceptual Framework.” *Journal of the Academy of Marketing Science* 22 (2): 99–113.

Ganesh, Jaishankar, Mark J Arnold, and Kristy E Reynolds. 2000. “Understanding the Customer Base of Service Providers: An Examination of the Differences Between Switchers and Stayers.” *Journal of Marketing* 64 (3): 65–87.

Gefen, David. 2002. “Customer Loyalty in e-Commerce.” *Journal of the Association for Information Systems* 3 (1): 2.

Gregory, Bryan. 2018. “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data.” *arXiv Preprint arXiv:1802.03396*.

———. 2018. “Predicting Customer Churn: Extreme Gradient Boosting with Temporal Data.” *arXiv Preprint arXiv:1802.03396*.

hcho3. 2020. “Awesome XGBoost.” <https://github.com/dmlc/xgboost/tree/master/demo#machine-learning-challenge-winning-solutions>.

He, Ruidan, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. “An Unsupervised Neural Attention Model for Aspect Extraction.” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics.

Hong, Liangjie, and Brian D Davison. 2010. “Empirical Study of Topic Modeling in Twitter.” In *Proceedings of the First Workshop on Social Media Analytics*, 80–88.

Jheng, Yang Ruei, and Zhi Chao Luo. 2019. “Customer Retention Prediction with CNN.” In Data Mining and Big Data: 4th International Conference, DMBD 2019, Chiang Mai, Thailand, July 26–30, 2019, Proceedings, 1071:104. Springer.

Koehn, Dennis, Stefan Lessmann, and Markus Schaal. 2020. “Predicting Online Shopping Behaviour from Clickstream Data Using Deep Learning.” *Expert Systems with Applications* 150: 113342.

Kracklauer, Alexander, Olaf Passenheim, and Dirk Seifert. 2001. “Mutual Customer Approach: How Industry and Trade Are Executing Collaborative Customer Relationship Management.” *International Journal of Retail & Distribution Management*.

Luo, Ling, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019. “Unsupervised Neural Aspect Extraction with Sememes.” In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 5123–29. International Joint Conferences on Artificial Intelligence Organization. <https://doi.org/10.24963/ijcai.2019/712>.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient Estimation of Word Representations in Vector Space.” *arXiv Preprint arXiv:1301.3781*.

Murthy, Sreerama K. 1998. “Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey.” *Data Mining and Knowledge Discovery* 2 (4): 345–89.

Nanayakkara, Shane, Sam Fogarty, Michael Tremeer, Kelvin Ross, Brent Richards, Christoph Bergmeir, Sheng Xu, et al. 2018. “Characterising Risk of in-Hospital Mortality Following Cardiac Arrest Using Machine Learning: A Retrospective International Registry Study.” *PLoS Medicine* 15 (11): e1002709.

Oliveira, Vera Lúcia Miguéis. 2012. “Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques.” PhD thesis, Universidade do Porto (Portugal).

Paruelo, JoséM, and Fernando Tomasel. 1997. “Prediction of Functional Characteristics of Ecosystems: A Comparison of Artificial Neural Networks and Regression Models.” *Ecological Modelling* 98 (2-3): 173–86.

Suryadi, D. 2020. “Predicting Repurchase Intention Using Textual Features of Online Customer Reviews.” In *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, 1–6. <https://doi.org/10.1109/ICDABI51230.2020.9325646>.

Tamaddoni Jahromi, Ali, Mohammad Mehdi Sepehri, Babak Teimourpour, and Sarvenaz Choobdar. 2010. “Modeling Customer Churn in a Non-Contractual Setting: The Case of Telecommunications Service Providers.” *Journal of Strategic Marketing* 18 (7): 587–98.

Tulkens, Stéphan, and Andreas van Cranenburgh. 2020. “Embarrassingly Simple Unsupervised Aspect Extraction.” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 3182–87. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.290>.

Verbeke, Wouter, David Martens, Christophe Mues, and Bart Baesens. 2011. “Building Comprehensible Customer Churn Prediction Models with Advanced Rule Induction Techniques.” *Expert Systems with Applications* 38 (3): 2354–64.

Wai-Ho Au, K. C. C. Chan, and Xin Yao. 2003. “A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction.” *IEEE Transactions on Evolutionary Computation* 7 (6): 532–45. <https://doi.org/10.1109/TEVC.2003.819264>.

Yanfang, Q., and L. Chen. 2017. “Research on e-Commerce User Churn Prediction Based on Logistic Regression.” In *2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 87–91. <https://doi.org/10.1109/ITNEC.2017.8284914>.

Yin, Jianhua, and Jianyong Wang. 2014. “A Dirichlet Multinomial Mixture Model-Based Approach for Short Text Clustering.” In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 233–42. KDD ’14. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2623330.2623715>.

Yu, Xiaobing, Shunsheng Guo, Jun Guo, and Xiaorong Huang. 2011. “An Extended Support Vector Machine Forecasting Framework for Customer Churn in e-Commerce.” *Expert Systems with Applications* 38 (3): 1425–30.

Zhao, Yu, Bing Li, Xiu Li, Wenhuang Liu, and Shouju Ren. 2005. “Customer Churn Prediction Using Improved One-Class Support Vector Machine.” In *International Conference on Advanced Data Mining and Applications*, 300–306. Springer.