Artificial Intelligence and Robotics

KAN-CINTO1820U

Course professors

Daniel Hardt

Kim Normann Andersen

Group members

Edoardo Abate (130381)

Michele Franco Scarperi (130403)

Date of submission: 11-05-2020

Number of pages: 15

Number of characters: 28,320

Table of Contents

Table of Figures

Abstract

This paper explores the business value of Artificial Intelligence (AI) in predicting the likelihood of customer churn in a supervised setting. First, a dataset from a telecommunication company is explored and pre-processed. Second, the performance of different Machine Learning (ML) algorithms, ranging from k-NN to Multilayer Perceptron is evaluated. Third, after a technical discussion on the supervised classification problem, the implications of ML and AI for telecommunication businesses in general are being examined. Ultimately, reflections and concluding remarks are being presented. Even though the source code is not explicitly included within this paper, we do encourage to review it in the accompanying Jupyter Notebooks[1], as further details and explanations can be found there.

1 Introduction

Back in 2014, IBM and a bank named DBS had announced a partnership to develop one of the first Artificial Intelligence (AI) applications in a business context. Eventually, the idea, which consisted of bringing a "robo-advisor" to life, did not work out: the scope was way beyond what was possible at the time. However, it taught DBS a valuable lesson, and it triggered the spin-off of several smaller scope AI projects within the bank. Individually, these projects were less ambitious, but their aggregated benefits allowed the bank to lower their operating expenses, increase productivity, reduce human errors and increase speed-to-market (Davenport, 2018).

Despite providing undisputed business value, applications of *AI* are more prosaic than what the general public and businesses believe. The current technical state of the art consists of highly capable mapping algorithms, which often outperform skilled humans. However, these algorithms are focused on individual tasks and they do not contain any inherent "intelligence". Multi-purpose learning systems, that can imitate and augment general intelligence, are the end goal of the research in this field. Nevertheless, AI in its current forms is a valuable tool which already supports many business processes, services and products. Common business applications of AI, according to Davenport (2018, p. 6), include *"demand forecasting, product search ranking, product and deals recommendations, merchandising placements, fraud detection, translations, and much more."*

[1] The Jupyter Notebooks accompanying this work are available at the following resource: https://github.com/edab96/AI-and-Robotics-Project

Nowadays, it is widely agreed upon that Artificial Intelligence is a general-purpose technology, much like electricity (Kelly, 2014; Davenport, 2018; Ng, 2019). An additional area of application for AI is the augmentation of human intelligence. The human brain is not capable of processing and analysing high amounts of data at high speed, in the same way computers are. AI algorithms can vastly enrich the information humans have access to and enable them to make better decisions faster.

1.1 Telecommunication Industry

Commercial organizations seek to generate profits by marketing products and services. Some of them operate in highly competitive industries, such as the telecommunication sector, where they deal with extremely low profit margins. Additionally, switching costs for consumers are low, amounting to even more pressure on companies. The Danish telecommunication market is of particular interest, as it is characterized by a robust DSL and fibre optic infrastructure, a comprehensive LTE coverage and a fast-developing 5G segment (Lancaster, 2020). According to the Danish Energy Agency (Energistyrelsen, 2018) profit margins in the telecommunications industry in 2018 were rather low with 8.4%. Although appearing to be quite concentrated in terms of turnover, with Telenor AS holding 34.8% of the total revenue share[2] and HI3G Denmark ApS with 27.2% (Statista, 2019a), the hierarchy is not reflected in terms of profit margin[3], with Telenor AS not appearing in the top 20 (Statista, 2019b) of the most profitable businesses.

1.2 Customer Churn

To gain an edge over competitors, companies need to be capable of conquering customers while retaining existing customers, or in other words, limit customer attrition. Customer attrition, or churn, occurs when a company loses a customer, due to a failed repeat-purchase or a terminated contract. Churn is usually measured with a rate, defined by the number of ceased customers over the total amount of customers during a certain time period (Nokia, 2011). A high churn rate is a significant problem for businesses, since the impact can be felt directly on revenues. Customer attrition should direct companies' attention to the specific reasons that are causing customers to leave. Attrition analysis can drive continuous improvement and enable companies to outperform the competition on a continuous basis (Reichheld & Earl Sasser Jr., 1990). Ideally, companies

[2] See Appendix (A), page 18.

[3] See Appendix (B), page 19.

should leverage a proactive defection strategy, as opposed to only collecting feedback from customers who have already churned. Assessing when a customer is at risk to churn is not a simple task. This 'blind spot' impedes businesses to run strategic actions with the aim to prevent customers from churning. Nowadays, though, most businesses collect data about their customers. Analytics can assist in studying this data to develop strategies against attrition and help companies improve their internal processes as well as strengthen their value proposition.

2 Technical Perspective

This work is based on a dataset published by IBM[4]. The technical analysis presented below comprehends an Exploratory Data Analysis (EDA), an overview of pre-processing techniques used to prepare the dataset for the models' evaluation and the achieved scores. The EDA[5] has enabled us to get a thorough understanding of the data; this paper however only features a selection of the most interesting insights due to the page constraint.

2.1 Exploratory Data Analysis

The dataset consists of a list of 7043 customers of a non-specified telecommunication company. The information about those customers comprehends 21 features, ranging from gender, to the monthly fees and the type of contract that the customers have with the telecommunication company. The dataset also includes the target feature: whether a specific customer has churned or not. The distribution is skewed towards the non-churning customer class, since only 26.5% out of the total 7043 customers are attributed to the churning class.

We analysed the features of the dataset with a correlation analysis in order to find out which features account for most explanatory power. The limitation of a correlation analysis is however, that it can only help to identify linear relationships if they exist between the target feature and the rest of the data. If features are related in a non-linear way a correlation analysis won't detect those patterns in the data. A relatively strong correlation between non-churning customers and their

[4] The official IBM source is: https://developer.ibm.com/technologies/data-science/patterns/predict-customer-churn-using-watson-studio-and-jupyter-notebooks/#

[5] The full in-depth EDA Jupyter Notebook is accessible at the following resource: https://github.com/edab96/AI-and-Robotics-Project/blob/master/EDA_In_Depth.ipynb

seniority is expected. Seniority refers to how long a particular individual has been a customer of the company. This information is captured by the feature `tenure` and is measured in number of months. The computed correlation (Figure 1) shows how, `tenure` has a 35% positive correlation with the target feature and is over all the second highest correlated feature in the dataset.

Feature	Value	Correlation with target
Contract	Month-to-month	40.46 %
Tenure	*Continuous*	35.41%
Internet Service	Fiber optics	30.75%
Contract	Two year	30.16%
PaymentMethod	Electronic check	30.15%

Figure 1: Top 5 correlated features against target.

Figure 2 proposes a visualization of `tenure` clustered into five groups. Churning customers are more concentrated in the groups with lower seniority, while non-churning customers are more evenly distributed among the different `tenure` clusters. The first year of the customer lifetime appears to be the most critical in terms of likelihood of churn. It is important to point out that the lower ranges of `tenure` include for most of the customers, therefore it is expected to find most churning customers in those ranges.

Figure 2: Tenure groups against churning and non-churning customers.

Feature contract type shows a similar tendency. Most churning customers (88%), are on month-to-month contracts. This does not come as a surprise, as contracts with lower time-horizons provide less barriers to churn, as opposed to longer-term contracts. Non-churning customers are more distributed on different contract durations: month-to-month, annual, biennial (see Figure 3).

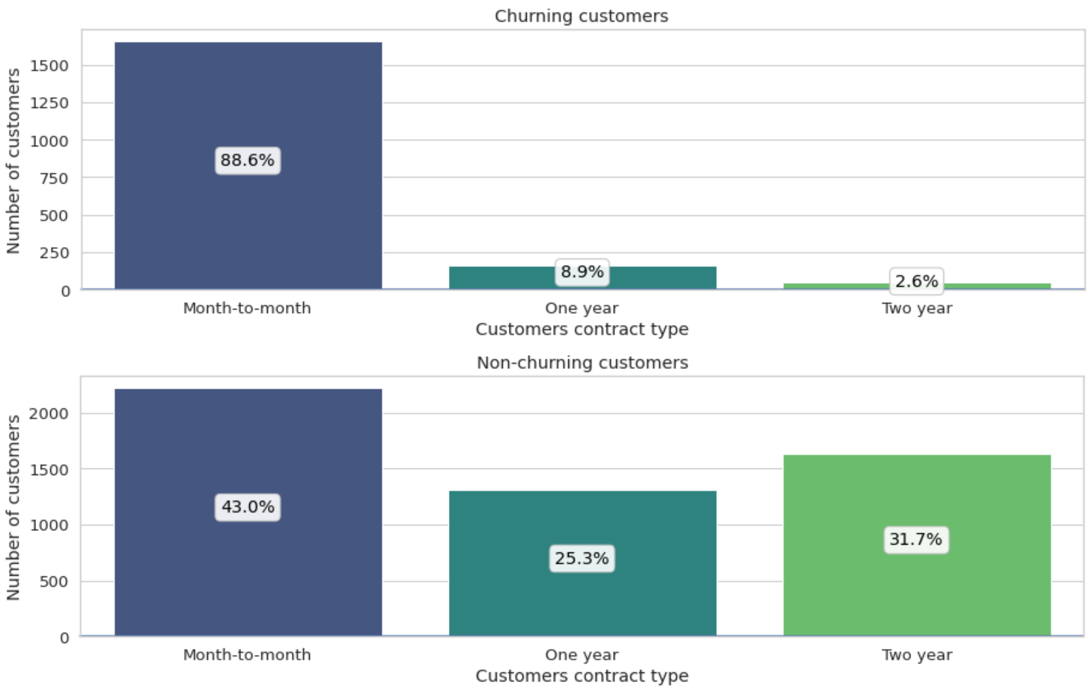


Figure 3: Contract type against churning and non-churning customers.

Feature `InternetService` (Figure 4) shows a correlation of 30% with the target. Most churning customers are on a DSL plan (69.4%), while a minority is on fibre optic. We hypothesise that individuals on DSL plans are more prone to move to either cheaper DSL options or more expensive yet faster alternatives. Customers are less likely to churn if they are on a fibre optic plan, if competitors do not either offer lower prices or better internet speeds. The company should evaluate upselling its fibre optic plan to DSL customers. An extensive list of implications inferable from all the features in the dataset can be found in Appendix C.

Figure 4: Customers with and without Internet Service and Type

During the exploration of the dataset, a Principal Component Analysis (PCA) with two principal components was conducted. Figure 5 shows that PCA does not separate the two classes of the target feature well. Churning and non-churning customers overlap a lot, although churning customers are less present per higher value of the second principal component.

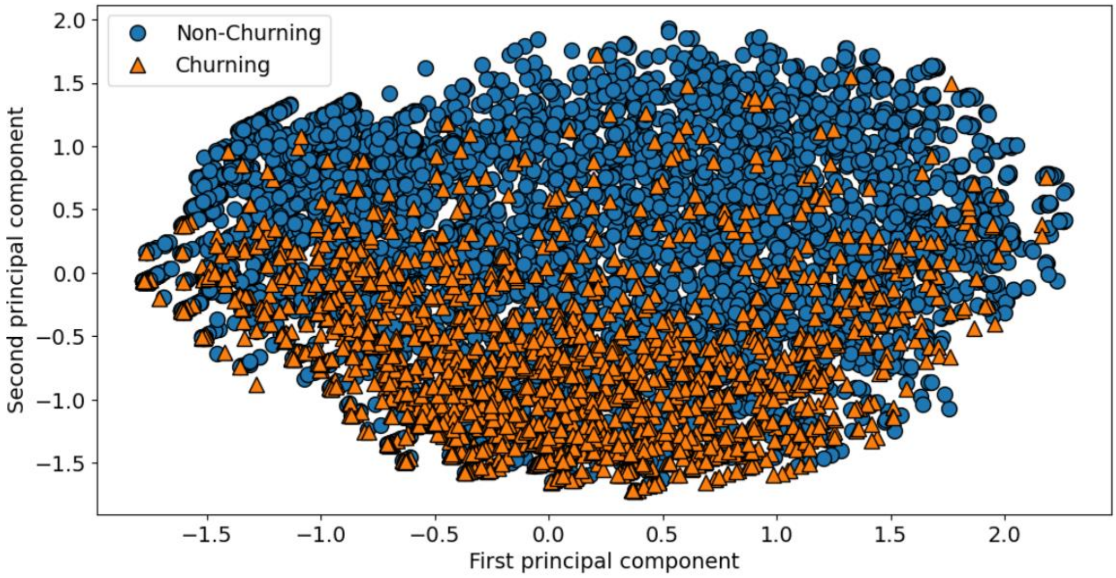


Figure 5: Principle Component Analysis.

2.2 Dataset pre-processing

The features from the original dataset come in different types: `float64`, `int64`, `object` (or `string`).

All the `object` (or `string`) features contain categorical data. First, multiple-value categories have been simplified to binary values where appropriate. Next, a single new one-hot encoded feature for binary features such as `gender` was defined, since having multiple one-hot encoded columns for each binary value would increase computational costs with no added benefit. The remaining categorical features with more than three unique values, such as `InternetService`, `Contract` and `PaymentMethod`, were pre-processed by one-hot encoding a new column for each unique feature value.

Features `tenure`, `MonthlyCharges` and `TotalCharges`[6] were normalized to numerical values between 0 and 1, since disparate value ranges have a negative effect on the performance of some Machine Learning models. The applied normalization consists of a linear transformation $x'_{f(i)} = \frac{x_{f(i)} - average(f_i)}{max(f_i) - min(f_i)}$ to each numerical feature f_i.

Feature type	Unique values	Transformation	One-hot encoded columns
Binary category (string)	2	/	1
Polynary category (*string*)	3, 4	/	One for each unique value
Continuous number (*float64*)	inf	Linear scale	None

Figure 6: Summary of the data preparation process.

One final remark on the pre-processing. Features `MonthlyCharges`, `TotalCharges` and `tenure` do not present a normal distribution. To adjust this, we considered applying a logarithmic transformation. This is a widely used method in the data science community to achieve symmetrical distributions, which are preferable for some model building techniques (Desarda, 2018). However, logarithmic transformation has a big short coming when dealing with 0 values, since the logarithm of 0 is minus infinity. Instead of falling back on alternative methods such as square root, we opted for not changing the distributions at all.

[6] In the original dataset, `TotalCharges` appears to be of type `object` and not `float64`. We had to drop 11 entries for `TotalCharges` and then converted the type to `float64`, to be consistent with the type of `MonthlyCharges.`

2.3 Evaluation of Models

Various ML models have been implemented, fine-tuned and evaluated using the following structure. First, a grid with arbitrary ranges for two different hyperparameters was defined. Starting from this first parameter grid, we performed a grid search[7], assessing performance using a stratified five-fold cross validation. Next, the grid search output was plotted. Finally, the parameter grid was adjusted in order to move to a more promising parameter region[8]. The evaluation ended with a second Grid Search run.

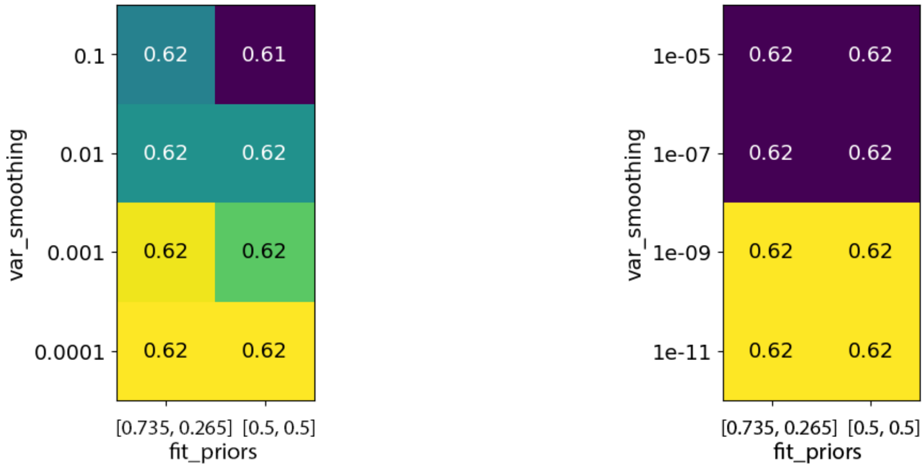


Figure 7: First and second Grid Search for Gaussian Naive Bayes.

The results are summarized in the following figure and can be replicated within the Jupyter Notebook accompanying this paper. Performance was very similar among the various tested models. Gaussian Naïve Bayes (GNB) performed the best with 61.42%, followed by Logistic Regression, while Random Forest only achieves 52.29%. Figure 7 shows how GNB plateaus at the highest validation score with different hyper setting configurations, various values of `Priors`. This is consistent with Naïve Bayes being robust to parameter settings (Müller & Guido, 2016).

[7] Compared to a standard `train_test_split`, cross-validation is more computational expensive, though it enables the model to be fit to a bigger training dataset. Since our computations were run on *Google Colab*, and the dataset is not to large, we were not worried about the extra workload on the processing unit.

[8] Visualizing parameter regions for more than three dimensions is not possible. Instead of running random grid searches, and given the educational purpose of this project, we have opted for visualizing the score against two hyperparameters only.

Model	Best hyperparameter settings	Cross-val score (F1)	Test score (F1)
Gaussian Naïve Bayes	var_smoothing = 1e-11 Priors = [0.735, 0.265]	62.15%	**61.42%**
Logistic Regression	C = 100 Solver = 'saga'	60.15%	60.00%
MLP	hidden_layer_sizes = (5, 5) learning_rate_init = 0.01	58.72%	58.96%
SVC	C = 100 gamma = 0.001	58.82%	58.73%
k-NN	n_neighbors = 40 weights = 'uniform'	58.59%	58.33%
Linear SVC	C = 0.1 tol = 1	59.81%	58.02%
Random Forest	n_estimators = 10 max_features = 50	55.93%	52.29%

Figure 8: Evaluation of Models with F1 scores.

The F1[9] score has been adopted as an evaluation metric in the cross-validation process. Different authors have shared their works on the same dataset that we have studied. Most of them used `accuracy` as a metric for evaluation, though we disagree with this decision. As mentioned during the EDA, the churning customer class accounts for only 26.5% of the total customers. A model that always predicts the majority class (non-churning customers) might have a relatively high accuracy, because the dataset is vastly comprised of non-churning customers. We have initially also used `accuracy` as an evaluation metric and all tested models achieved an accuracy of around 80%. This was however a non-meaningful evaluation method, due to the fact that accuracy does not take imbalanced classes into account as the F1 score does.

Paliwal & Kumar (2017) propose a Neural Network to predict churn using the same dataset we have worked with. Unfortunately, they also use `accuracy` as a metric, so our works are not comparable. They claim that their Artificial Bee Colony (ABC) Neural Network outperforms a classic backpropagated Neural Network for this task and achieves an accuracy of 89%. We are sceptical of this result for two reasons. First, the score seems to refer to the training score and not to the test score. This would suggest that their fast convergence to a lower error (11%) consists in a quicker way to overfit to the training data and is not proven to be reflected on the test data.

[9] F1 is defined as $F1 = \left(2 * \frac{precision * recall}{precision + recall}\right)$, where $precision = \frac{true\ positives}{true\ positives + false\ positives}$ and $recall = \frac{true\ positives}{true\ positives + false\ negatives}$. In this context, false positives refer to customers being predicted as churning when they are not; false negatives refer to churning customers being predicted as non-churning.

Second, other works show no clear benefits of using ABC (Bullinaria & AlYahya, 2014). It must be stressed that a different churn dataset can yield different scores for the tested models. Telecommunication companies should not indiscriminately adopt Naïve Bayes as a prediction model for their own datasets. It just appears to be the best performing model with the dataset object of our study. More of this, and its implications for businesses, is discussed in the next section.

3 Business Perspective

Applying AI in order to tackle customer churn in Denmark is estimated to amount for 3-5 billion DKK, and todays' AI applications can unlock DKK 100 to 160 billion in value in the Danish private sector (Lindberg et al., 2019). Since establishing endowments of data and the right organizational capabilities are required to successfully harvest the benefits of AI, prioritizing AI implementation projects can create a first-mover advantage.

Davenport (2018) suggests that in order to succeed with AI, companies need to invest steadily and make a good match between their business problems and the capabilities of AI. Decision makers within companies need to have a clear picture of the weaknesses of their business as well as a comprehensive overview of what the current technical state of the art potentially enables. This allows corporations to build a healthy *ambition*. Healthy in the sense that a vision is achievable and can drive projects that bring business value. Ambition, or vision, is deeply intertwined with the context of a company. In order to create value and provide insights, AI algorithms need to be carefully trained. Humans learn through experiences, while AI learns through data. Companies must therefore evaluate whether they possess the needed data, and if not, what strategies they can pursue in order to get hold of it. We discuss data collection in depth in a dedicated section below.

3.1 State of the Art

As mentioned in the introduction to this paper, current AI algorithms perform extremely well in associating one or multiple inputs with one or multiple outputs. Today's business applications range from sentiment analysis, entity extraction, image recognition, to speech recognition and synthesis, and many more. Many authors argue that the hardest aspect of AI is integrating it with existing systems and business processes. Davenport (2018) shows that managers often do not understand cognitive technologies and how they work, additionally some businesses cannot get enough people with relevant expertise. Chui et al. (2018) write that mastering AI requires new levels of expertise, which may be an impediment to successful adoption. They also recommend businesses to remain vigilant and responsible as they deploy AI.

According to Reichheld & Earl Sasser Jr (1990), profitability of a company increases with customers locked in for a longer time period. They argue that loyal customers are more likely to purchase more from the company, that operating cost for loyal customers will decrease over time and that they are more likely to generate referrals for the company. Ultimately, the company is able to charge a price premium, due to the established and well-known brand. Predicting churn could function as an AI pilot project within a telecommunication company: the following section outlines how telecommunication companies can strategically approach piloting AI projects such as churn predictions.

3.2 Content, Ambition, Talent and Acquisition

Since AI needs data in order to "learn", decision makers must initially identify the digital information possessed and already available within their company. This may include monthly charges, seniority of the customers and active subscriptions, just like in the dataset studied in this paper. Companies may also have data about service usage, that can be utilized. Another alternative to enrich the data of a certain customer is also by leveraging third party services.

A subsequent phase in the company strategy is to develop a vision, an *ambition*, of what to do with the possessed data. For a telecommunication company, this should not be limited to churn predictions only, but also include predictive maintenance for physical networks, service recommendations for customers, or optimizing internal operations to gain a cost advantage. Clearly, different objectives may require different input data. In this regard, a fundamental role is to be played by the expertise of the decision makers. A company's *ambition* should consist in critically matching use cases enabled by AI with the company's business goals. *Content* and *ambition*, although important, are only the start of an AI implementation process.

Additional requirements are knowledge and capabilities within machine learning and artificial intelligence. Some companies may already have those resources. For instance, Telia AB, the fifth largest telecom operator in Denmark, is applying ML and AI-powered technologies to identify the most valuable customer accounts based on internally available data (Kekäläinen, 2016). Knowledge and expertise within the field of AI, however, may not be possessed by companies. Paths to pursue, in this case, are either acquiring or renting expertise. A telecommunication company should seek to acquire talents with a background in computer science, if it has long-term ambitions to incorporate AI in its business processes. All Danish telecommunication companies have headquarters in Copenhagen or the surrounding area, where many Universities are based. Danish telecommunication companies should acquire new talents with educational backgrounds in business and IT from those Universities. Young, unexperienced talents could be coupled with external consultants that have experience with deploying AI in businesses. We argue that, although relying on consultancy agencies to work on AI solutions might seem attractive, it will not help building internal capabilities and AI culture within the company in the long term. We believe that a combination of talent acquisition and external experts' guidance would provide the most value to a telecommunication company at an early stage, until capabilities are more rooted in house. Davenport (2018) suggests that it is also possible to limit or avoid talent acquisition by training the current workforce in AI. For example, Cisco Systems has made it possible for its employees to deep dive into data science by setting up learning programs with two different universities. This alternative strategy could also be pursued by Danish telecommunication companies. A look towards the rich start-up scene in Copenhagen might also offer some solutions. It is highly likely for some of the existing start-ups to possess proprietary data, software, or algorithms that can be valuable additions to the likes of HI3G Denmark ApS or Telenor AS.

4 Reflections

The technical discussion has shown how Artificial Intelligence can help to assess the likelihood of customer attrition. For the purpose of evaluation of the different algorithms, an F-score metric is preferable over the most commonly applied metric accuracy. F1-score assigns equal importance to false positives (non-churning customers classified as churning) and false negatives (churning customers classified as non-churning). In a real industry setting, telecommunication domain experts need to assess what error type is the most critical for their businesses goal. False positives might trigger a customer retention strategy without necessity. For instance, incumbents might try to secure the loyalty of a customer by offering a one-time discount or gifting them an additional service for a limited time. In this case, the customer benefits from a better offer, while the corporation incurs unnecessary cost. Revenue from this customer is not lost.

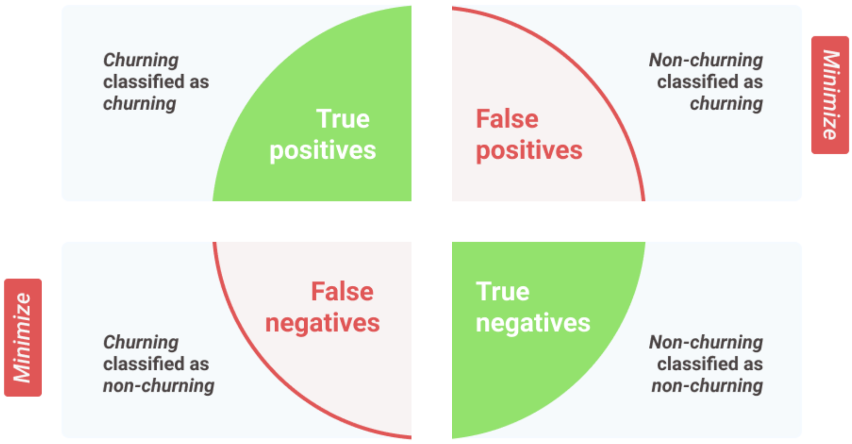


Figure 9: Model deployment evaluation matrix for churn prediction.

On the other hand, false negatives would not trigger the company to pro-actively act in securing a customer's loyalty. The company does not invest resources, at the expense of losing all potential future revenue coming from that customer. We argue that false negatives are more critical to businesses, assuming lost future revenues account for a higher amount than the proactive retention strategy. After all, loyalty strategies, in the form of one-time discounts or rewards, could be automated and triggered when the AI algorithm identifies a customer is at risk with enough confidence.

Model	Best hyperparameter settings	Cross-val score (F2)	Test score (F2)
Gaussian Naïve Bayes	var_smoothing = le-9	71.02%	**69.64%**
Linear SVC	C = 0.01 tol = 100	62.18%	60.28%
Logistic Regression	C = 1000 Solver = 'liblinear'	57.11%	57.08%
k-NN	n_neighbors = 25 weights = 'uniform'	57.14%	56.62%
SVC	C = 100 gamma = 0.001	55.69%	55.05%
MLP	hidden_layer_sizes = (10,8,4) learning_rate_init = 0.1	57.51%	49.30%
Random forest	n_estimators = 100 max_features = 10	53.41%	48.78%

Figure 10: Evaluation of Models with F2 Scores.

Assuming false negatives are more important, the F2 score is the appropriate metric to optimize for. Figure 10 shows the computed cross-validation and test scores with metric set to F2. Gaussian Naïve Bayes still performs best and achieves a much better result compared to the other models, outperforming the linear Support Vector Machine Algorithm by 9%. Results seem to suggest that the Multilayer Perceptron Neural Network and the Random Forest algorithm are overfitting the training data, since the respective test scores are considerably lower than the cross-validation scores.

5 Conclusion

Fridman (2020) advocates for less hype, as well for less anti-hype, in regard to Artificial Intelligence. His advice points towards more technical research and multi-disciplinary collaboration. Business decision makers must accept that AI in the sense of a general intelligence system, capable of reasoning and building knowledge on a variety of scenarios is not possible yet. What is possible, however, is to leverage single-purpose AI systems to create business value. A possible application is in the form of churn prediction. Churn can happen among customers as well as internally among employees. For instance, DBS introduced a system to help their human resources department to predict churn among their own sales force. This information helps DBS to avoid excessive disruption in their operations by proactively trying to keep the churning employee, or else anticipate when to recruit new personnel and trigger those activities timely (Priya, 2019).

Customer churn predictions allow businesses to operate more efficient, while having a richer information picture about their customers. This enables them to take pro-active measures to secure customer loyalty. This application also demonstrates how AI does not necessarily replace the human workforce, but instead provides support by augmenting human intelligence. Predictive maintenance has been implemented in an industrial setting to achieve cost savings. This paper has shown that AI can similarly allow companies to implement proactive strategies to secure customer loyalty, avoid the loss of market share to competitors and reduce missed future revenue.

References

Bullinaria J.A., AlYahya K. (2014) Artificial Bee Colony Training of Neural Networks. In: Terrazas G., Otero F., Masegosa A. (eds) Nature Inspired Cooperative Strategies for Optimization (NICSO 2013). Studies in Computational Intelligence, vol 512. DOI: 10.1007/s12293-014-0137-7

Chui, M., Henke, N. & Miremadim M. (2018). Most of AI's Business Uses Will Be in Two Areas. Harvard Business Review. Retrieved from: https://hbr.org/2018/07/most-of-ais-business-uses-will-be-in-two-areas

Davenport, T. H. (2018). The AI Advantage: How to Put the Artificial Intelligence Revolution to Work, MIT Press Books, The MIT Press, edition 1, volume 1, number 0262039176, August.

Desarda, A. (2018). Getting Data ready for modelling: Feature engineering, Feature Selection, Dimension Reduction (Part 1). Towards data science. Retrieved from: https://towardsdatascience.com/getting-data-ready-for-modelling-feature-engineering-feature-selection-dimension-reduction-77f2b9fadc0b

Energistyrelsen (2018). Financial Key Figures for Telecommunications Industry 2018. Retrieved from: https://ens.dk/en/our-responsibilities/telecom/telecom-numbers

Fridman, L. (2020). Deep Learning State of the Art (2020) | MIT Deep Learning Series [Video]. YouTube. https://www.youtube.com/watch?v=0VH1Lim8gL8

Kekäläinen, J. (2016). How Telia's $60 bot created €5 million in savings. 3rd Annual Intelligent Automation Nordics. Retrieved from: https://www.aiia.net/events-rpanordics/landing/telia-case-study/

Lancaster, H. (2020). Denmark - Telecoms, Mobile and Broadband - Statistics and Analyses. Paul Budde Communication Pty Ltd.

Lindberg, B., Andersen, J. R., Hansen, M. A. E., Frandsen, S., Alstrup, S., Krause, S., Duvold, T. (2019). An AI Nation? Harnessing the opportunity of artificial intelligence in Denmark. *McKinsey*. Retrieved from: https://innovationsfonden.dk/sites/default/files/2019-09/an-ai-nation-harnessing-the-opportunity-of-ai-in-denmark.pdf

Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, Inc.

Ng, A. (2019). Andrew Ng: Enterprise AI Strategy (with Landing AI) - CxOTalk #365. CXOTALK. [Video]. YouTube. https://www.youtube.com/watch?v=Y7fH2iT1m7Q

Nokia (2011). Operational Efficiency Benchmarking Insight Paper An Anatomy of Churn. Retrieved from: https://archive.vn/20141103141907/http://webcache.googleusercontent.com/search?q=cache:XSlskfssUM0J:networks.nokia.com/system/files/document/an_anatomy_of_churn.pdf+&cd=3&hl=en&ct=clnk&gl=us

Paliwal P., Kumar D. (2018) ABC Based Neural Network Approach for Churn Prediction in Telecommunication Sector. In: Satapathy S., Joshi A. (eds) Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2. ICTIS 2017. Smart Innovation, Systems and Technologies, vol 84. DOI: https://doi.org/10.1007/978-3-319-63645-0_38

Priya, S. (2019). DBS implements a new system for predicting employee resignation. *OpenGovAsia*. Retrieved from: https://www.opengovasia.com/mitigating-employee-turnover-with-predictive-analysis-and-ai/

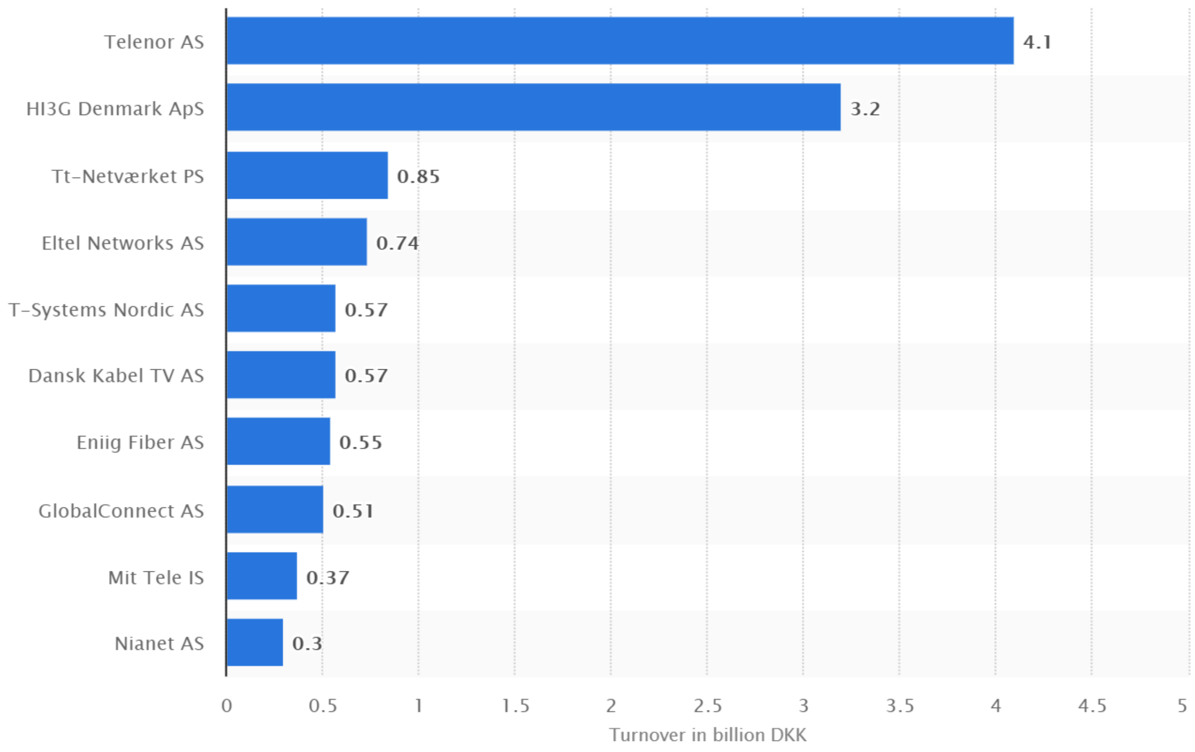
Reichheld, F.F. & Earl Sasser Jr., W. (1990). Zero Defections: Quality Comes to Services. *Harvard Business Review 68, no. 5 (September–October 1990): 105–111*. Retrieved from: https://hbr.org/1990/09/zero-defections-quality-comes-to-services

Statista (2019a). Leading 10 telecom companies in Denmark as of May 2019, by turnover. Retrieved from: https://www.statista.com/statistics/554473/denmark-top-20-telecommunication-companies-by-turnover/.

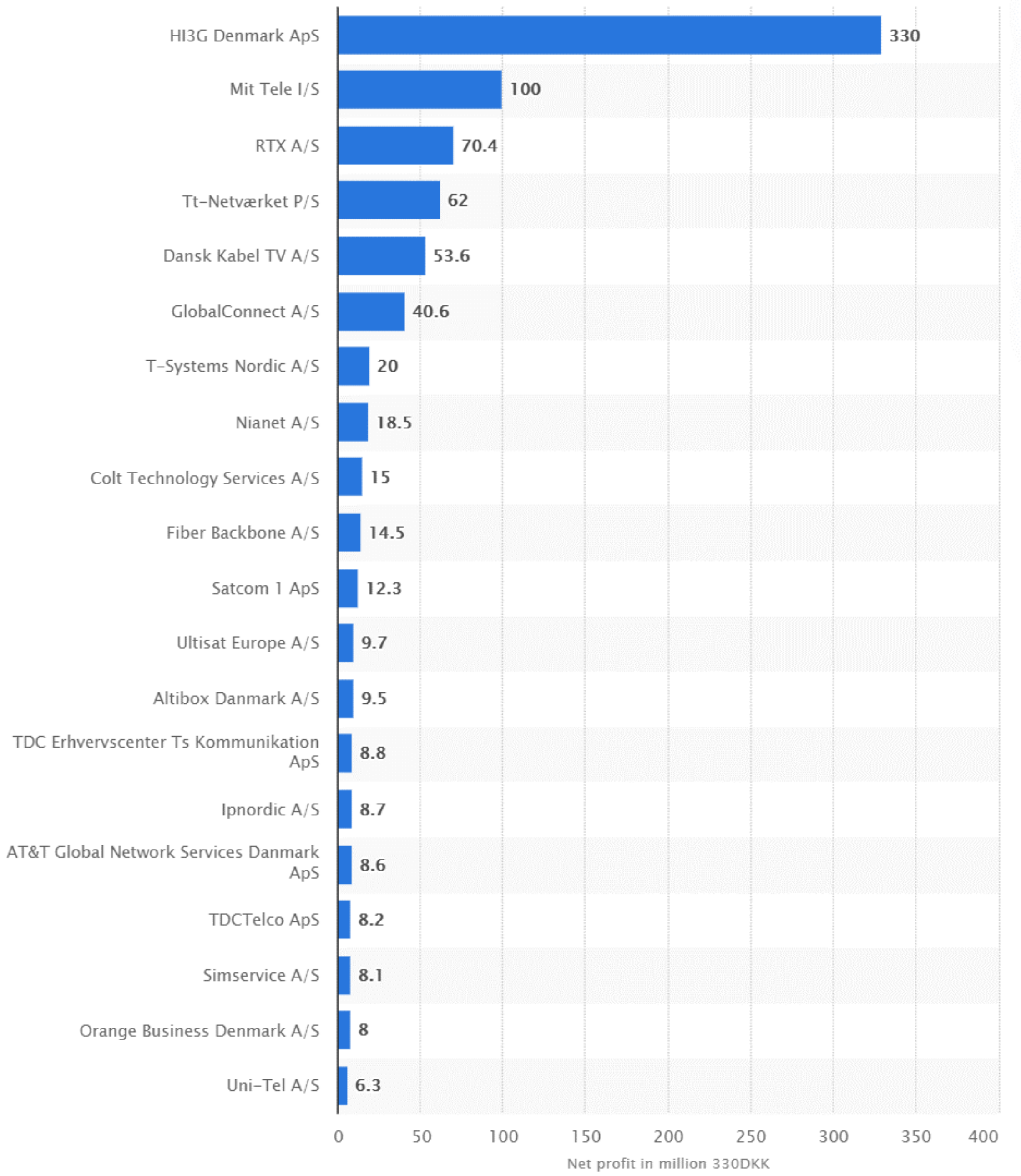
Statista (2019b). Leading 20 most profitable telecommunication companies in Denmark as of May 2019, by net profit. Retrieved from: https://www.statista.com/statistics/554655/denmark-top-20-most-profitable-telecommunication-companies/.

Appendix

A) Leading 10 telecom companies in Denmark as of May 2019, by turnover (in billion DKK).



B) Leading 20 most profitable telecommunication companies in Denmark as of May 2019, by net profit (in million DKK).



C) List of Implications per Feature

Feature	Insights	Potential impact on Churn*	Recommendation to Management
Gender	50/50 distribution	low	-
Senior Citizen	out of the 1869 customers that churned 74.5% were not Senior Citizens.	high	Address needs of younger customers in order to retain them.
Partner	64.2% of customers that churned did not have a partner.	high	Address needs of customers that are single for higher retention.
Dependents	82.6% of customers that churned did not have dependents.	mid	Follow the same strategy as with customers without partners.
Tenure Groups	TG1: 0-12 months, TG2: 12-24 months, TG3: 24-48 months, TG4: 48-60 months, TG5:	high	The overall trend is customers that churn less often the longer they are subscribed. Most churns happen in the segments 0-12 months, more than 60 months and 48-60 months.
Phone Service	90.9% of customers that churned did not have the phone service.	low	Upselling potential, unlikely to impact churn directly because 90.1% of customers that haven't churned also don't have the phone service.
Multiple Lines	54.5% of customers that churned had no multiple lines or no phone service at all.	low	This represents an opportunity for the telco company to sell more phone service and multiple line packages but is unlikely to affect the churn rate because the distribution is relatively even also within customers that did not churn.
Internet Service	69.4% of customers that churned had DSL internet service.	low	Indication of upselling opportunity for the telco company to try to switch customers from DSL to Fiber Optic. It could also indicate potential issues with the DSL internet service the telco company is providing.
Online Security	84.2% of customers that churned had no online security service.	low	Improve service, upselling opportunity
Online Backup	72% of customers that churned did not have the service online backup.	low	Improve service, cross-selling opportunity

Device Protection	70.8% of customers that churned did not have the device protection service.	low	Improve offer, upselling opportunity
Tech Support	83.4% of customers that churned had no technical support service booked.	low	Bundle with other services, justification for higher priced bundles.
Streaming TV	56.4% of customers that churned did not have the TV streaming service.	low	Bundle with other services, justification for higher priced bundles.
Streaming Movies	56.2% of customers that churned did not have the Movie streaming service.	low	Bundle with other services, justification for higher priced bundles.
Contract	88.6% of customers that churned were paying on a monthly basis.	high	This fact indicates that the telco company should try to convert customers into the more long-term contract agreements since churn is significantly lower within the one- or two-year contract types.
Paperless Billing	74.9% of customers that churned did not have paperless billing.	high	Strong indicator that this does affect the churn rate and should be therefore addressed by the telco company.
Payment Method	57.3% of customers that churned used the automatic bank transfer payment method.	mid	The telco company should try to incentivize its customers to use one of the other three options.
Monthly Charges	Bottom 25% Group 0-36$, 50% Group 36-71$, 75% Group 71-90$, Top 25% Group >90$.	mid	Most customers churned from the two groups paying monthly: 0-36$ and 71-90$.
Total Charges	Group 1: 0-402$, Group 2: 402-1397$, Group 3: 1397-3795$, Group 4: >3795$	low	40.9% of customers that churned were charged less or equal to 402$ in bills. The general trend shows the more was charged the less customers had churned which makes sense in the context of the 3 different contract types and that mostly monthly paying customers churn because they aren't as locked into the service as the one- or two-year contract customers.

*Potential impact on Churn does not describe the potential impact on the churn predictions (for that a correlation analysis will be presented in order to point out the most important features) but rather the logical conclusions that can be drawn from the EDA on why customers churn.