|  |  |
| --- | --- |
| Oliver Gade Nielsen  Study number - 20205968 | 31.646 characters  ≈ 13,2 pages |

|  |
| --- |
| Aalborg University Business School |
| Leveraging machine learning to reduce customer churn |
| How can machine learning help customer-focused companies reduce customer churn |

Table of content

[Abstract 2](#_Toc185282551)

[Intro 3](#_Toc185282552)

[Research question 3](#_Toc185282553)

[Theory and methodology 4](#_Toc185282554)

[Understanding customer churn 4](#_Toc185282555)

[Business perspective 5](#_Toc185282556)

[Methodology 6](#_Toc185282557)

[Dataset overview 6](#_Toc185282558)

[Technical Implementation 8](#_Toc185282559)

[Feature engineering and data preparation 8](#_Toc185282560)

[Model training and evaluation 11](#_Toc185282561)

[Deployment 12](#_Toc185282562)

[Model demonstration 12](#_Toc185282563)

[Analysis 12](#_Toc185282564)

[Results 14](#_Toc185282565)

[Discussion 14](#_Toc185282566)

[Conclusion 16](#_Toc185282567)

[Future developments 17](#_Toc185282568)

[References 18](#_Toc185282569)

# Abstract

Today the telecommunication market is highly competitive, where customer churn can have significant impacts on revenue, customer acquisition costs, and market share. This project aims to explore if and how predictive machine learning models can help address customer churn by identifying key drivers and predict how likely a given customer is to churn. With the use of a dataset including information on customers such as demographics, service details, and account information, a supervised machine learning model, CatBoostClassifier, was implemented. This particular model was chosen for its ability to handle categorical data without encoding it first. Key churn drivers were identified through EDA and feature importance was calculated using Random Forest Classifier, another model commonly used for predictive tasks. According to results customers with shorter tenure, higher monthly charges, flexible contracts, and manual payment methods showed a greater risk of churning.

The model achieved promising performance metrics, including 77.93% accuracy, 83.69% recall, and 79.77% ROC-AUC, demonstrating its potential for supporting data-driven retention strategies. Although, with a precision score of 55.60% it also showed room for improvement, indicating further fine-tuning and possibly inclusion of external factors could benefit the model’s performance.

This project highlights the added value of predictive analytics when reducing churn and enhancing retention strategies, which can offer telecommunications companies a valuable competitive advantage.

# Intro

Today, the telecommunication market is filled with companies all fighting for a share of market that is not rapidly growing. Based on a review of publicly available information, I managed to find a list of telecommunication companies in Denmark, which shows a list of thirty-four (!) different companies(Mobil-daekning.dk). Such a high number of companies offering the same or similar services results in a highly competitive industry, where companies consistently strive to have the cheapest prices, best service, or a combination of both, to attract new customers. In such a highly competitive industry customer churn is a critical metric for a few reasons such as revenue loss, customer acquisition cost, loss of market share, and brand perception(B. Svendsen & K. Prebensen, 2013).

According to a Harvard Business Review article, acquiring new customers can be 5 to 25 times more expensive than retaining customers, depending on what industry you are in. The article also suggests that increasing customer retention by as much as 5% can increase profits by 25% to 95%. So, minimizing customer churn is something telecommunication companies would be very interested in (Gallo, 2014).

The aim of this project is to produce a tool that, to some extent, can predict this customer churn by identifying key drivers for customer churn. Such a tool can be a powerful tool for addressing and understanding customer churn, and in turn hopefully help develop retention strategies.

For this project, machine learning plays a crucial role since it, with the right models, is able to uncover certain patterns in customer behavior, identify key drivers that can lead to customer churn and predict a possible outcome based on the data. All of this makes such a tool very valuable for data driven decision making in customer-focused industries such as telecommunication.

While some of the tasks done by machine learning models can overlap with human capabilities, humans often underperform in speed, and in some cases accuracy, compared to machine learning models which can lead to higher efficiency.

## Research question

*How can machine learning help customer-focused companies reduce customer churn?*

* *What are key drivers for customer churn in the telecommunication market, and how can machine learning help identify these for future improvements in the service/product provided?*

# Theory and methodology

## Understanding customer churn

Before diving into this project, it is important to know what customer churn is. Customer churn refers to the rate at which customers will stop doing business with a company in a given period of time, where the “doing business” part differs from industry to industry. For this project, the focus will be on the telecommunication industry, where customer churn typically represents the percentage of customers (also referred to as subscribers) who either cancel their service contracts (phone plan, internet service etc.) or switch to a competitor.

As discussed in the aforementioned article from the Harvard Business Review, there is a direct link to be made between customer churn and loss of revenue. In this case the revenue is subscription revenue, which is the type of revenue the telecommunication industry relies on, unlike other industries where revenue might come from one time sales. If a telecommunication company loses a significant number of customers, it can lead to a significant shortfall in revenue, which will impact the company’s financial projections. Acquiring a new customer can be rather expensive since it involves marketing, advertising, promotions, possibly also discounts and more to attract new customers from competitors.

If a company, in any market really, is experiencing a consistent customer churn, it can signal issues customers might what to avoid, such as poor service, inadequate customer service/support, or prices that are higher than similar product on the market. Such issues can lead to dissatisfaction among customers, which in turn can lead to bad word-of-mouth and a damaged reputation, which can make it harder to attract new customers. On the other hand, retaining customers and offering good service, good customer support and competitive pricing strategies can foster loyalty, which in turn can lead to good reputation and good word-of-mouth. Good reputation can also make it easier to both retain customers and attract new customers.

Traditionally customer churn is analyzed with traditional statistical methods and labor-intensive manual approaches(Celik & Osmanoglu). Some of the methods are effective to some extent, but struggle to handle larger amounts of data, thus not getting the full grasp of what underlying patterns there are in the data. Some of the traditional methods include:

* Descriptive analysis where a company would rely on descriptive statistics in which metrics like churn rate, average customer lifespan, and revenue per customer is calculated to give the company a snapshot of customer churn trends over time. This method is only useful if the company is interested in identifying general trends and does not care about specific insights.
* Surveys and feedback analysis where classic surveys were sent to customers, where they would be asked about experiences with the company and reasons for leaving. This approach gives some valuable qualitative insights for the company, it is a more reactive approach than it is proactive since it most of the time handles data from customers who have already left.

With the introduction of machine learning it has significantly transformed the methodologies used in customer churn prediction, since it now offers scalable, data-driven, and proactive solutions that extend beyond the possibilities of traditional customer churn prediction(Vafeiadis et al., 2015). Machine learning models are able to handle and analyze large amounts of data and they can identify complex patterns in the data that might be almost invisible to human eyes. The ability to handle large amounts of data and to be predictive are some of the big benefits of using machine learning in tasks such as customer churn analysis.

As mentioned before, it enables companies to be proactive rather than reactive, and possibly prevent churn before it even happens. Furthermore, the large amount of data enables companies to get more complex and detailed insights into why customers might leave them. On top of that, machine learning applications are also more adaptive and scalable compared to the job done by a human, which can benefit a growing company so they can scale it to meet their demand.

With all the benefits from using machine learning models, it is quite clear that it is a really good match for the task, working with customer churn analysis.

## Business perspective

In a data-driven business context customer churn prediction is so close to strategic decision making and the company's ability to deliver services and value, that it could be argued that it is a natural step in the business modelling process. Some of the points where the company can benefit from customer churn predictions are addressing customer pain points, data-driven decision-making, and using data as a strategic asset. Being proactive at identifying customers at risk of churning, the company can quickly and efficiently identify and resolve pain points, which can improve overall satisfaction and thus the likelihood of customers staying with them, which can lead to more brand loyalty.

When implementing predictive insights into the business model, the company can leverage the predictiveness to transition from intuition-based decision-making to data-driven decision-making, allowing for more optimal resource allocation, which in the end can help maximize overall profit. As mentioned in the intro, the telecommunication market is quite competitive, and customer retention is more cost-effective than acquiring new customers, so leveraging predictive analytics in customer retention can provide significant competitive advantages. This advantage can help the company deliver better customer experience, thus differentiating themselves from competitors on the market.

## Methodology

For this project, I chose to work with supervised machine learning, namely CatBoostClassifier, since the prediction requires a binary outcome - Will a customer churn yes (1) or no (0). The data used in this project would be considered categorical, making CatBoostClassifier a good choice for the predictive part, since it is optimized for working with categorical data while not needing extensive preprocessing. On top of that, its ability to naturally integrate both numerical and categorical features, combined with high efficiency, further proves CatBoostClassifier to be a solid choice for customer churn prediction.

CatBoostClassifier has previously proven to be highly effective when predicting customer churn, as shown in a study published in 2023 by Kalasalingam Academy of Research and Education in India, where they also used CatBoostClassifier to predict customer churn. In their research, they achieve an impressive 95% accuracy (Jane Rubel Angelina et al., 2023-05-17)

### Dataset overview

The dataset used for this project consists of different types of information about the customers such as demographics, service usage patterns, and service subscriptions. First, I will divide the columns into three categories based on what they contain, and then I will describe them. The dataset contains 7043 rows of data.

The dataset contains the following columns:

|  |  |
| --- | --- |
| **Category** | **Columns** |
| **Demographics** | customerID, gender, SeniorCitizen, Partner, Dependents |
| **Service Details** | tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies |
| **Account Information** | Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges, Churn |

Table 1

In the category **demographics,** we find information such as age, gender, whether the customer is a senior citizen or not, whether the customer has a partner or not, and if the customer has any dependents such as kids, elderly parents or other family members that rely on the customer for financial support.

In the category of **service details,** we find details about the type of services the customer receives. This section includes details such as the type of internet service they receive (DSL, Fiber optic or no internet service), whether they have any streaming services included in their contract, if the customer has phone service with the company, and whether or not they have multiple lines for the phone service.

In the category **account information,** we find information about what type of account the customer has. This section includes information about tenure (i.e. how long the customer has been with the company), what type of contract they signed (month-to-month, one year, or two years), billing/payment method (paperless or not, paying with electronic check, mailed check, automatic bank transfer, or automatic credit card payment)

Lastly there is the target variable Churn which indicates whether a customer has stopped their services with the company.

# Technical Implementation

### Feature engineering and data preparation

When building accurate and reliable machine learning models it is important to do effective feature engineering and data preprocessing. This section will go through the steps taken in this project.

* *Handling missing values*

In the dataset there is a column called “TotalCharges” that contains missing values. To address the issue, the missing values was imputed as a product of the columns “MonthlyCharges” and “tenure”, to ensure a complete and meaningful dataset for the machine learning model.

Since it is a rather significant feature in my prediction logic, it can skew the model’s prediction if not handled properly, which can lead to reduced accuracy or bias.

* *Standardizing categorical data*

Multiple columns contain inconsistent categorical data, which makes it harder to train a model in a meaningful and effective way. Because of that, categorical values were converted to “Yes” and “No” to simplify the data and improve the model’s ability to interpret the data. For instance, “No phone service” was changed to “No” in the column “MultipleLines”. These columns also had categorical data converted “OnlineSecurity“, “OnlineBackup“, “DeviceProtection“, “TechSupport“, “StreamingTV“, “StreamingMovies“.

Inconsistent data can have a significant effect on the mode’s ability to generalize effectively. When data is standardized, it is ensured that the model understands categorical variables correctly, which makes the model better at generalizing the data.

* *Encoding target variable*

Churn, as a target variable, was converted from a categorical Yes/no to a numeric 1/0, to ensure that the model had the right data type for the classification tasks at hand.

When a target variable is encoded, in this case to 1/0, it ensures better compatibility with machine learning models. This ensures that the model can work with the target variable and predict correctly.

* *Stratified sampling for train-test splitting*

Since churned customers are fewer the non-churned customers in the data set, it is considered imbalanced and thus needs some work before it can be split for the training and testing of the model. To do this, a StratifiedShuffleSplit was done to ensure a more proportionally correct representation of the “Churn” data in both the train and test split. This helps by improving the model’s performance in generalizing when applied to unseen data. If the data was not stratified it would impose a risk of overfitting the majority class (non-churned customers) and it might not be able to generalize the minority class (churned customers).

* *Identifying categorical features for modelling*

Since this project is using the CatBoostClassifer model, which natively handles categorical data, there is no need for traditional encoding such as One-Hot encoding. What it needs is for the categorical columns in the dataset to be explicitly defined. This approach the preprocessing steps while not negatively impacting computational efficiency.

By performing EDA on the data at the very start of the project, I identified the features that are most relevant when predicting churn.

The highest scoring features are Tenure (0.1542), MonthlyCharges (0.1421), and TotalCharges (0.1351). This can also be seen in the correlation matrix (figure 1) made for the dataset.

Et billede, der indeholder tekst, skærmbillede, linje/række, nummer/tal

Automatisk genereret beskrivelse

Figure 1

Et billede, der indeholder tekst, skærmbillede, Kurve, nummer/tal

Automatisk genereret beskrivelse

Figure 2

### Model training and evaluation

In an attempt to maximize performance of the CatBoostClassifier, I performed hyperparameter tuning using a custom grid search. This approach helps optimize the model for the specific dataset and prediction task at hand. The following hyperparameters, and the corresponding values, were considered in the grid search:

* Iterations: The number of boosting iterations to perform (100, 200)
* Learning rate: Step size for the gradient decent (0.01, 0.1)
* Depth: The maximum “depth” of the decision trees (4, 6, 8, 10, 12)
* Scale\_pos\_weight: This parameter is used to adjust the importance of minority classes in imbalanced data (1, 3)

By doing this grid search, the process evaluates all possible combinations of the abovementioned hyperparameters. For each combination it creates a new instance of the CatBoostClassifier with specific parameter values based on the grid search. Then, the model is trained on the training dataset (X\_train and y\_train), where categorical features are passed via the cat\_feature parameter which leverages the CatboostClassifier’s ability to handle categorical values. Lastly the model is tested on the test dataset (X\_test and y\_test), where the performance is evaluated using the F1 score. F1 score is chosen as evaluation metric since it combines precision and recall, which is important when working with imbalanced data.

When the process is iterated through all possible combinations, it will select the best model based on highest F1 score. This approach to hyperparameter tuning ensures that the best combination of hyperparameters is found and used, granted it can search within the selected values for each hyperparameter.

The model is then trained with the best combination of hyperparameters found in the grid search, and finally evaluated using the following metrics:   
**Accuracy**: The percentage of correct predictions.

**Recall**: The number of actual churned customers that were correctly identified.

**Precision**: The reliability of churn predictions (i.e., the proportion of predicted churned customers that were actual churners).

**ROC-AUC**: The model’s ability to differentiate between churned and non-churned customers.

**F1 Score**: A mean of precision and recall, reflecting the balance between the two metrics.

The following show the evaluated metric for my model:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Accuracy | Recall | Roc\_Auc | Precision | F1 Score |
| CatBoost\_Model | 0.779276 | 0.836898 | 0.797676 | 0.55595 | 0.66809 |

Table 2

### Deployment

For the backend development of this application, considerations regarding deployment options have been made for ease of use by the company, such as exposing the application with FastAPI. This deployment method can allow for the application to be deployed in the company’s backend workflow, where it can continuously scan new data and save the output/prediction in a database further down the line for future analysis. Then, based on the predicted churn rate, the company can segment their customers and make some more personalized retention campaigns.

### Model demonstration

For the sole purpose of demonstration of the model’s performance, I have also developed a simple Streamlit app/dashboard. The dashboard is not considered at valuable tool for the company, but it does make it easier to demonstrate and visualize the performance and features of the model.

# Analysis

The analysis of customer churn began with uncovering patterns in the dataset and identifying key drivers of customer churn. Through exploratory data analysis (EDA), multiple influencing factors were uncovered. In this chapter, I will go through the key driver that was identified and try to analyze how these key drivers influence customer churn.

One of the most significant drivers that was identified was **Tenure**. Customers with longer tenure were found to show a lower likelihood of churning. This shows that the early stages of the customer lifecycle are critical, since short tenure and dissatisfaction show a higher likelihood of a customer churning. This highlights the importance of early engagement strategies and a strong onboarding process to address any dissatisfaction before it becomes a reason to leave.

Two other key drivers identified were **Monthly Charges** and **Total charges**, where the analysis showed a strong correlation with customer churn. This indicates that customers with higher monthly charges were more likely to cancel their services, suggesting that customer retention is significantly influenced by the perceived value of the service. This, in turn, may suggest that customers who pay higher prices may switch to a competitor who offers similar services at a lower price. Here, the findings suggest that transparent pricing and value-driven services may increase customer satisfaction.

Another key driver found to play a significant role in customer churn were **Contract type**. A higher churn rate was seen with customers on month-to-month contracts compared to customers on one- or two-year contracts. Customers may get greater flexibility with month-to-month contracts, but it also makes it easier for them to switch to a competitor if they experience dissatisfaction.

Lastly, **Payment Method** was also found to be a significant influence on customer churn. Customers paying with electronic checks were found to show higher churn rates compared to customers paying with automated methods such as bank transfer or credit card. This could suggest that customers prefer convenient payment processes, which is likely due to inefficiencies in manual payment methods.

During the EDA phase a correlation matrix was made, see figure 1 in technical implementation. This showed a clear correlation between variables such as the negative correlation between churn and tenure, confirming that longer tenure lowers the churn rate. Furthermore, a histogram was made to analyze the link between prices and churn rate, where it show an increasing churn rate at higher price points, see figure 3. This also indicates that there is a threshold where the service is perceived as too expensive compared to its value.

Et billede, der indeholder diagram, Kurve, skærmbillede, linje/række

Automatisk genereret beskrivelse

Figure 3

# Results

Findings in the analysis are further validated by the CatBoostClassifier used for the churn prediction. During the feature importance analysis is identified tenure, monthly charges, and total charges as the most significant features when predicting churn. The three features contributed with 15,42%, 14,21%, and 13,51% to the decision-making process, respectively. These results were seen to align with the patterns uncovered in the EDA.

As seen in figure 3, the model’s performance was evaluated using the metrics Accuracy, Recall, Precision, F1 score, ROC-AUC. The model achieved 77.93% accuracy, meaning in almost 8 out of 10 customers it predicted correctly. 83.69% recall, demonstrating the model’s ability to identify customers who actually churned. However, with 55.60% precision a significant portion of predicted customer churn is not actually customer churn. For the company, this could lead to increased retention costs, but it still ensures that most at-risk customers are caught. 66.81% F1-score, showing how well the model handles imbalanced data, indicating a less than ideal performance since there a risk it would predict to many wrong “churners”. Lastly, the model scored 79.77% ROC-AUC, showing how well the model distinguishes between churned and non-churned customers.

Despite precision and F1-score being less than ideal, these insights can still provide a robust foundation for strategies to reduce customer churn. For example, the company could target short-tenure customers with retention campaigns offering personalized deals to make them stay longer. Furthermore, the company could encourage long-term contracts with discounts or additional benefits, while also promoting automatic payment with either bank transfers or credit cards.

Discussion

The analysis and results of this project highlights the value of understanding what drives customer churn. It also demonstrates the benefits of using predictive models as a tool in strategic business decisions, such as creating different retention strategies for different customer segments. It also emphasizes how important optimized price strategies are, as well as focusing on customer engagement in the early stages of customer lifecycle and addressing pain points customers might have that could lead to churn.

A predictive model, such as the CatBoostClassifier used in this project, can give companies a data-driven and scalable tool to help develop new strategies. As opposed to traditional methods, these models can process large amounts of data, helping the company to be more proactive than reactive in their decision-making.

Although the project showed promising results, a few issues were observed during the project work, issues that could potentially lead to biases in the data and limitations in the model’s performance. As mentioned earlier, the dataset used in this project is imbalanced, with less churned customers than non-churned customers. This is an issue because, although the data was stratified, it can still significantly impact the model’s performance in generalizing minority classes in the dataset. By incorporating external variables such as competitor pricing or market trends, the model’s performance could be enhanced, as demonstrated in a study on Predicting Demand for Fast-Moving Consumer Goods, published in 2019 by researchers from Centro Paula Souza in São Paulo, Brazil (Tarallo et al., 2019).

There is also a risk of certain customer groups being underrepresented in the data, such as specific demographics, niche customer segments, and some rare service combination. These combinations might not be adequately captured, thus leading to limitations in the model’s ability to generalize.

In the end, there is a direct link between the model’s performance and the data it is trained on, so if any variables are underrepresented or outright missing, it will have an impact on the model’s ability to perform well on unseen data. This misrepresentation of certain features could also lead to biases in the data, where the model might overemphasize or neglect certain features due to how they are represented in the training data.

Lastly, there is the external factors, which the model does not account for. These are factors such as competitor actions, economic conditions, or market trends, which could influence customer churn. With more advanced setups with more real time data about beforementioned factors, the model could probably take it into account, but with the current data and setup, it is not possible.

I would argue that the findings and results from this project fit well with the theme of the course, being data driven business modelling and strategy. This project demonstrates that leveraging data as a strategic asset and using it in predictive analysis, as it is done in this project, can help companies transition from intuition-based decision-making to be more data-driven, which can lead to improved profits by enhancing customer experience and optimizing resource allocation. This helps the company understand churn drivers and focus their efforts on retention strategies aimed at at-risk customers. Additionally, insights like the ones seen in this project serve as a guide for product development, for example highlighting what additional services and what payment options customers prefer.

From a strategic standpoint, this project highlights the value of incorporating data-driven tools, such as predictive analytics, in the business model. In this case it is to address customer churn, which undeniably is a key factor in the company’s long-term sustainability and competitive advantage. When a company is proactively managing customer churn, not only can they strengthen their market position but also build solid customer loyalty and thus achieve better financial outcome.

Conclusion

The aim of this project was to look into how, or if, machine learning can help address the issue of customer churn, and if yes, what are key drivers for customer churn in the telecommunication market, and how can machine learning help identify these for future improvements in the service/product provided. By utilizing machine learning, the project’s aim was to identify key drivers of customer churn, and additionally demonstrate how predictive models can be used for such tasks.  
This led to the research question:

*How can machine learning help customer-focused companies reduce customer churn?   
- What are key drivers for customer churn in the telecommunication market, and how can machine learning help identify these for future improvements in the service/product provided?*

The analysis showed that the top drivers for customer churn in the telecommunication market were tenure, monthly charges, contract type, and payment method, meaning customers with high monthly charges, shorter tenure, or flexible contracts (i.e. month-to-month) showed aa higher risk of churning. As discussed, these potential issues highlights the importance of focusing on customer engagement in the early stages of the customer lifecycle, encouraging long-term contracts and optimizing the pricing strategy.

The machine learning model used in this project, CatBoostClassifier, proved to be quite effective in identifying key churn drivers and predicting customer churn. The model showed a fairly strong performance with 77.93% accuracy, 83.69% recall, 55.60% precision, 66.81% F1-score, and 79.77% ROC-AUC, although strong it still showed room for improvement, especially in the precision score.

To answer the research question, machine learning can help companies by identifying key drivers for churn but also predict at-risk customers, building a solid base for strategic decisions such as retention strategies targeted these at-risk customers, in an attempt to reduce churn and improve profits. Machine learning also helps companies to adapt customer needs directly into their strategies. As a result of this project, it is also proved that machine learning models are very powerful tools in reducing customer churn.

# Future developments

In the chapter I will talk about what I see as potential future developments for this project, if I were to keep working on it.

**Additional data sources:**

I would explore the option of adding more data sources, to hopefully give the model better performance, and to give the company more detailed insights. Additional data could be:

* Competitor analysis, an analysis on what the competitor’s prices, offerings and market trends.
* Customer interaction data, more detailed data on call logs, support tickets, and customer feedback.
* Real time data feeding allowing for continuous churn predictions when new data “arrive.”

**More advanced models**

An idea could also be to explore more advanced models. Maybe deep learning models such as Recurrent Neural Networks or transformer models could help by analyzing time-series data or sequential data. I would also explore the use of multiple models, such as CatBoost, Random Forest, and XGBoost to get a more nuanced prediction. Lastly, I would explore more advanced options for hyperparameter tuning with a much bigger grid search, to hopefully find a better performing model.

**Integrate in bigger workflow**

Integrating this model into a feedback loop with systems that analyze effectiveness of retention strategies and do cost-benefit analysis could be a valuable asset in a market like the telecommunication market. On top of that, it could also be integrated into a marketing automation tool that sends personalized offers, discounts and other reminders based on the customer churn analysis.

References

B. Svendsen, G., & K. Prebensen, N. (2013). The effect of brand on churn in the telecommunications sector. *European Journal of Marketing, 47*(8), 1177. 10.1108/03090561311324273

Celik, O., & Osmanoglu, U. O.*Comparing to Techniques Used in Customer Churn Analysis*

Gallo, A. (2014, -10-29T13:00:41). The Value of Keeping the Right Customers. *Harvard Business Review,* <https://hbr.org/2014/10/the-value-of-keeping-the-right-customers>

Jane Rubel Angelina, J., Subhashini, S. J., Harish Baba, S., Dheeraj Kumar Reddy, P., Sudheer Kumar Reddy, P. V., & Sameer Khan, K. (2023-05-17). A Machine Learning Model for Customer Churn Prediction using CatBoost Classifier. Paper presented at the 166. 10.1109/iciccs56967.2023.10142823

Mobil-daekning.dk. . *Teleselskaber i Danmark.* <https://mobil-daekning.dk/oversigt-danske-mobilselskaber/>

Tarallo, E., Akabane, G. K., Shimabukuro, C. I., Mello, J., & Amancio, D. (2019). Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research. *IFAC-PapersOnLine, 52*(13), 737. 10.1016/j.ifacol.2019.11.203

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory, 55*, 1. 10.1016/j.simpat.2015.03.003