A Machine Learning Framework to Address Customer Churn Problem Using Uplift Modelling and Prescriptive Analysis

MSc Research Project
Data Analytics

Merve Baskan
Student ID:20238096

School of Computing
National College of Ireland

Supervisor: Paul Stynes, Musfira Jilani and Pramod Pathak

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Merve Baskan
Student ID:	20238096
Programme:	MSc Data Analytics
Year:	2021-22
Module:	MSc Research Project
Supervisor:	Musfira Jilani, Paul Stynes, Pramod Pathak
Submission Due Date:	19/09/2022
Project Title:	A Machine Learning Framework to Address Customer Churn Problem Using Uplift Modelling and Prescriptive Analysis
Word Count:	XXX
Page Count:	18

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	19th September 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Abstract

Customer churn refers to the percentage of customers who stop using a product or service in a given time period. Current research uses machine learning and deep learning models to classify customers for that purpose. However, the challenge to conventional customer churn prediction models is that they do not align with the real-life business objective. They only predict the outcome, i.e., whether a customer will churn or not. Models estimating the net effect of customer behaviour like uplift modelling, however, focus on whether a customer is intent on churning and will be retained when targeted with the campaign. This research proposes a machine learning framework to compare the uplift model with the conventional churn prediction model, using predictive and prescriptive analysis. The framework presents XGBoost(eXtreme Gradient Boosting) and Logistic Regression churn prediction models and their uplift models with two different treatments. Results of the four models are evaluated in this paper based on Qini coefficient, Qini curve, treatment correlation and accuracy. The results show that the uplift model outperforms the conventional customer churn prediction model, when it comes to targeting the right customer group for a retention strategy.

Keywords- uplift modelling, customer churn prediction, marketing

1 Introduction

Customer churn refers to the ratio of the number of customers lost in a certain period to the total number of customers.(Karvana et al.; 2019) Customer churn prediction is vital for every sector and one of them is internet service providers. If a provider wants to expand its income, it must acquire more subscribers; nevertheless, keeping existing customers is more vital than acquiring new ones. As a result, service providers want to know which customers will leave. Churn prediction enables businesses to design a strategy to lower the expenses.(Do et al.; 2017) Customer churn analysis is a topic of interest to researchers, especially in recent years, and various studies have been carried out with machine learning techniques that include Support vector machine(SVM), decision tree, naive bayes, and logistic regression.Vafeiadis et al. (2015)

However, conventional churn prediction applications have an important limitation, which is ignoring the "treatment" applied by companies. Treatment can be explained as an extra service provided to retain the customer. For instance, an internet service provider company sends an email to a customer about a new advantageous package. In conventional churn analysis, although some customers are considered in the "churn" category, they can be retained when they are offered the package. Likewise, customers classified in the "not churn" category may leave the company due to the offer. Finally, some customers' decisions do not change whether they are treated or not. Any service offered to the customer is costly, so it is to the disadvantage of the company to offer a treatment that does not influence the customer's decision. In order to address these limitations, uplift modeling is proposed. In other words, the conventional churn problem approaches can only distinguish between customers who are likely to churn and those who are not, whereas uplift modeling distinguishes between consumers who will benefit from being treated and those who will not.

The goal of uplift modeling is to determine the change in customer behavior that occurs from providing a given treatment. Uplift models can estimate the result as a consequence of a certain input variable. These variables indicate actions that the company has influence over and may thus be improved. In this way, it is possible to determine which strategy is applied to which customer with minimum effort and maximum profit. Figure 1 shows the target classes for customer behaviour and the uplift classes created from them. The aim of uplift modelling in churn application is to target the "persuadables", as they are the only group that needs treatment to stay with the company.

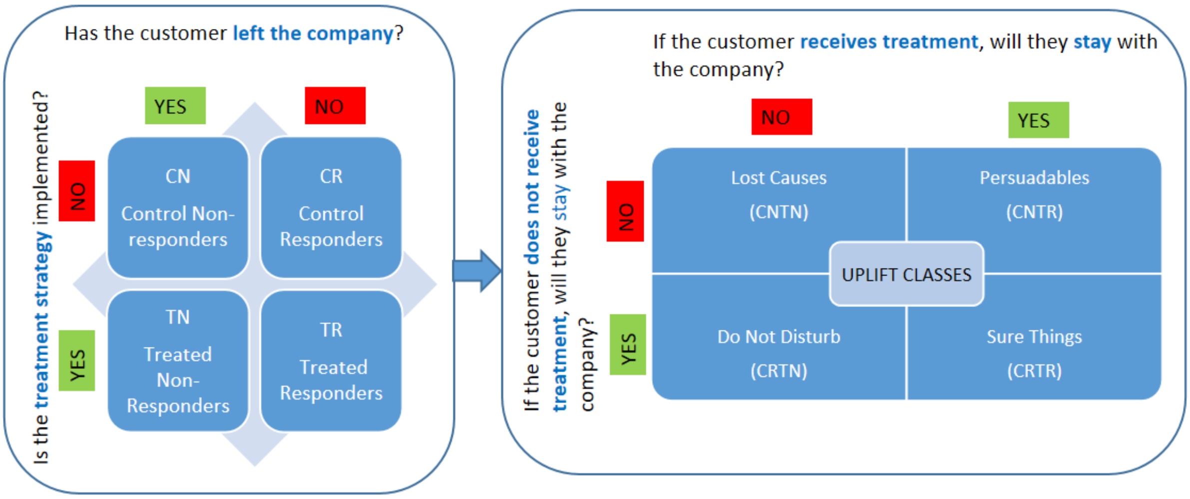

Figure 1: Target classes and uplift classes for the customer churn problem

This paper discusses whether the uplift model outperforms the conventional churn model in targeting the right customer group. The aim of this research is comparing the predictive and prescriptive performances of two approaches. Also, the research investigates whether the uplift model can reduce the churn rate over time and provide consistent reliability via targeting the correct customer group.

In order to address the explained research question, the research objectives are developed:

- Investigation of the state of the art researches about customer churn prediction applications and uplift modelling techniques using the different machine learning methodologies.
- Designing a machine learning framework for customer churn uplift model via selected uplift score formula.
- Implementing the machine learning framework using the most successful techniques in the literature : Extreme Gradient Boosting (XGBoost) and Logistic Regression.
- Evaluating the machine learning models' performance to detect the uplift models' effect on targeting the correct group of customer compared to conventional churn predicition. Utilizing evaluation metrics such as accuracy, treatment correlation and Qini coefficient.

The major contribution of this research is machine learning framework that combines conventional customer churn and uplifted customer churn together. Although a similar application is used in the field of human resource, a strategy as in this state-of-the-art research has not been followed for the customer churn problem. Also, the dataset and variables used in uplift model applications have been specifically modified. In this study, the dataset of churn has been modified so that it can also be used for the uplift model.

The rest of the paper is organized as follows: Section 2 presents literature review in customer churn prediction using machine learning algorithms and uplift models. In Section 3, research methodology is explained. Section 4 discusses the design components for the uplifted churn models with different machine learning techniques. The implementation of this research is discussed in section 5. Section 6 presents the evaluation results. Section 7 discusses the outcomes of the research and Section 8 concludes the research with future work.

2 Literature Review

The first section explains the definition, importance and machine learning applications of customer churn analysis in the literature. There is customer churn analysis of almost every field in the literature. However, state-of-the-art studies using machine learning have been highlighted. In the second part, uplift modeling studies and its effect on churn prediction problem are explained.

2.1 Customer Churn Prediction Using Machine Learning

Customer churn analysis is important in many industries, and it is often used by internet service providers. Generally, one person in each household uses the internet. Especially in the new generations, the rate of internet usage is increasing. This demand shows how open the industry is to growth.In this sector, customer losses occur quickly and in high volume. Nearly half of all Internet users quit their provider each year owing to the ease with which they may transfer.Khan et al. (2010) Churn prediction is the process of evaluating customer purchasing activity, identifying customer profiles who are likely to leave the organization, and forecasting those who are likely to leave. Churn analysis is vital, since finding a new customer is much more costly than retaining an existing customer. This analysis has now become a tool frequently used by strategic decision-making and planning officials.Çelik and Osmanoglu (2019)

Machine learning algorithms are one of the most commonly used techniques for churn prediction. Vafeiadis et al. (2015) investigated the influence of boosting methods on the classification accuracy of machine learning models for predicting customer churn. Support vector machine, decision tree, naive bayes, and logistic regression were applied, and the best overall classifier was the SVM-POLY utilizing AdaBoost, with an accuracy of over %97 and an F-measure of approximately %84. Babu and Ananthanarayanan (2016) used Hybrid models (Clustering and ANN) to increase the performance of the current classification models, and it was proven to outperform the single models. In order to achieve effective results for the customer churn prediction, Artificial Neural Networks, Decision Tree Methods, Logistic Regression, Support Vector Machines, and Naive Bayes techniques have been utilized.(De Caigny et al.; 2018) applied a hybrid classification algorithm with a decision tree model and logistic regression.(Tang et al.; 2020)suggested a customer churn prediction model based on XGBoost and Multi-layer Perceptron by combining their benefits. However, hybrid models have their own set of disadvantages in terms of application inconsistency. Ahmad et al. (2019) used decision trees, random forests, the GBM tree technique, and XGBoost. Another significant addition is the integration of consumer social networks in the prediction model using Social Network

Analysis (SNA). In comparison, XGBoost outperformed others with %93.3 of the Area Under Curve (AUC) metric. Lalwani et al. (2021) used logistic regression, naive bayes, SVM, random forest, decision trees, boosting, and ensemble approaches. Unlike other traditional techniques, K-fold cross validation has been employed to prevent model overfitting. Adaboost and XGboost Classifier were discovered to have the maximum accuracy of %81.71 and % 80.8 , respectively. The greatest AUC was % 84. Wang et al. (2019) used the Gradient Boosting Decision Trees (GBDT) model with ensembles. They assessed prediction performance in a large-scale customer data set and obtained %84.10 AUC (Area Under Curve). In addition to the studies mentioned above, Lalwani et al. (2021) used logistic regression, naive bayes, SVM, random forest, decision trees, boosting, and ensemble approaches. Unlike other traditional techniques, K-fold cross validation has been employed to prevent model overfitting. Adaboost and XGboost Classifier were discovered to have the maximum accuracy of %81.71 and %80.8 respectively.

As previously stated, the most often utilized strategies for churn prediction include Support Vector Machine (SVM), Gradient Boosting Trees (GBT), Random Forest, and Logistic Regression. Based on Kaggle competitions by using tabular data , Anthony Goldbloom (Co-Founder and CEO of Kaggle.com) claims that XGBoost (Extreme Gradient Boosting) is winning almost every competition in this field.Fogg (2016)

2.2 Uplift Modelling Applications

Uplift modeling is one of the prescriptive methods utilized in machine learning models that both predicts the result and proposes solutions according to the application.The disadvantage of conventional prediction models, as opposed to uplift models, is that they are not built to predict net response or uplift and optimize incremental impact of the applied strategy.Devriendt et al. (2018)

The uplift modeling's goal is to predict the net difference in churn likelihood as a result of a focused retention strategy. The direct effectiveness of the strategy can be examined via uplift modeling. The important part is to identify customers who are likely to churn yet could be retained by a right strategy or action. For the application, all customers are divided into two groups and random experiments are performed: Treatment group(customers who have been targeted with a retention action) and control group (customers who have not been targeted with a retention action). In order to successfully identify target customers, the diversity in customers' responses to the offer must be uncovered. De Caigny et al. (2021)

As shown in Figure 1, target class groups are formed by using the customers' responses to the applied retention strategy. Based on these groups, they are divided into 4 uplift groups:

Sure things: Customers who would never churn. Offering a retention program to this group is an unnecessary additional cost.

Lost causes: Customers who would churn regardless of the offer.

Do-not-disturbs: Customers who would churn just because they were offered the retention program.These customers will bring profit to the business as long as they are not disturbed.

Persuadables: Customers who would not churn if they were provided a retention program.The persuadables is the customer group that must be targeted via uplift modelling. Since the retention offer is successful for this group, they will churn if not targeted but stay in response to the retention campaign.

In the literature, many uplift model techniques were introduced and compared with each other which includes different data preprocessing techniques. For instance, the study of Jaskowski and Jaroszewicz (2012) evaluated a transformation technique that modifies the target variable, and defines another predictive variables. Shaar et al. (2016) introduced the Pessimistic Uplift Modeling, that minimizes disturbance effects for the process. In addition, Lai's approach applied by Kane et al. (2014). The aspect that differentiates it from other techniques is the ability to modify the target variable and convert the uplift modeling task into a binary classification. In order to show the effectiveness of the uplift modelling in employee retention Rombaut and Guerry (2020) applied causal conditional inference forest (CCIF) algorithm which improves prediction accuracy and minimizes overlearning. De Caigny et al. (2021) presented a novel uplift algorithm called logit leaf model(LLM) which includes two-step algorithm created for segmented uplift to solve customer churn prediction problem. The LLM consist of uplift decision trees and uplift logistic regressions. The model performed better than the three popular uplift algorithms in the literature which are: Uplift decision tree, uplift logistic regression , and uplift random forest. However, as a limitation, the LLM does not provide customer rankings based on treatment factor. Also, multi-treatment factor application can be performed to obtain more comprehensive results.

The common highlight of the researches of Devriendt et al. (2021) and Wijaya et al. (2021) is that they demonstrated the performance comparisons of uplift models with conventional prediction algorithms. Both mentioned studies used the Qini coefficient as the evaluation metric. This metric is evaluated via the Qini curve, which plots the total changes in churn/turnover rates between control and treatment groups as a function of a selected proportion of customers, ordered from high to low estimated uplift.(Gubela and Lessmann (2021)) Moreover, a novel metric was evaluated by Devriendt et al. (2021) aiming to solve the customer churn prediction problem with the uplift model. The maximum profit uplift (MPU) is an assessment metric that enables analyzing performance in terms of the highest profit that can be produced by using an uplift model. The proposed uplift models with Logistic Regression outperformed predictive churn models and contribute to increased retention strategy effectiveness. Despite the fact that the conventional model works well for random forest models , the uplift model achieves greater incremental improvements for smaller fractions of customers. Also the uplift curve values for conventional prediction model changes between 20% to 40%. That means the conventional prediction model is likely to target "sure things" instead of "persuadables".

To solve the employee turnover problem, Wijaya et al. (2021) compared performance by utilizing the Qini coefficient evaluation metric and accuracy, using the employee turnover prediction (ETP) and employee turnover uplift (ETU) models. Unlike other studies, it provided validation on 3 different datasets using the same methodology. According to the results, the traditional forecasting model provides an average forecast accuracy of 83.35% and the uplift employee turnover model of 70.03% which is a drawback for the proposed model. However, while the ETP model only gives a 50% consistent reliability rate to target the right employee, ETU provides close to 100% success in determining the target employee. Another point to be noted in this research is the use of Lai's generalized weighed uplift method (LGWUM)Lai (2006) was used to determine uplift classes. This method also applied by Kane et al. (2014) and Gubela et al. (2019). The generalized Lai's formula provides consistently greater performance and stability when compared to other common techniques such as the two-model approach, dummy treatment approach, and pessimistic approach. There four groups defined by this formula: Control non-responders (CN), treated responders (TR), control responders (CR) and targeted non-responders (TN). In order to predict the probability results of the four groups supervised classification can be used. Identifying these groups changes the target variable into a binary target variable, resulting in the uplift modeling becoming a binary classification prediction problem. Using binary classification makes the implementation uncomplicated.Asar (2019)

,

3 Methodology

The research methodology consists of three steps namely data preprocessing, machine learning modelling , evaluation and results as shown in Figure 2 The dataset to be used in the research is an open source dataset from Kaggle.com.[1]

In order to justify the selection of the methods, benefits and drawbacks should be considered. As described in the literature review section2, Extreme Gradient Boosting (XGBoost) is the machine learning method that gives the most successful results in recent projects. Logistic Regression, on the other hand, is the most frequently used method in customer churn problems, and there are many uplift model applications. LGWUM's formula used for Uplift model preprocessing is simple, high-performance and stable. Also, the formula has the benefit of converting any probabilistic classification model into a single model that estimates the uplift automatically.Asar (2019)

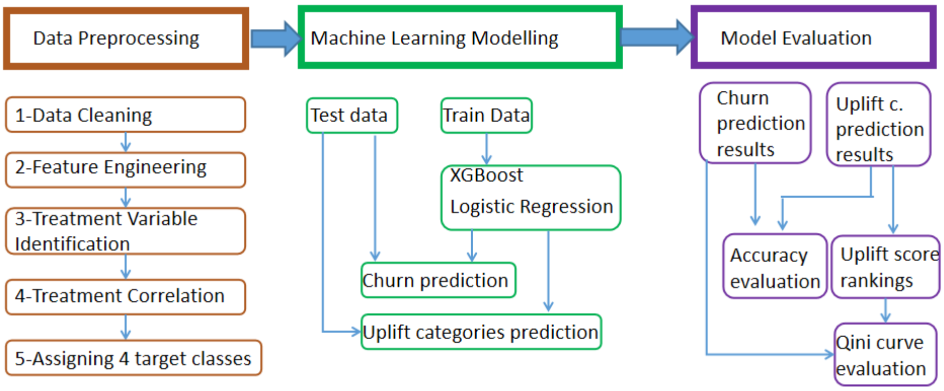

Figure 2: Research Methodology

The first step, data preprocessing, includes the steps to make the data suitable for modelling. The dataset includes 72275 rows and 11 columns. When raw data is checked; It is observed that 21572 null values in "reamining contract", 381 null values in "download avg" and "upload avg". Columns that will not be used for analysis and 381 null values are cleaned. "reamining contract" means that the customer has never preferred the contract. Therefore null values can be filled with 0. Also, a new column "has contract" is created to show whether the customer has already selected the contract or not (0 or 1). The new dataset consists of 71893 non-null rows and 11 columns.

Then the "treatment" column is determined and the treatment correlation is obtained. Treatment column should have 0 or 1 which indicates positive or negative response to the

[1]website: `https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn`

treatment. Treatment correlation(as a percentage) shows the relationship of the selected "treatment" column to "churn" and will be used in the analysis part as well. The "has contract" variable is selected as treatment 1 and "is tv subscriber" as treatment 2. (Only one treatment is used in each experiment.)

The final step in data processing is to specify four target classes (Figure 1) using the "treatment" and "churn" columns. The four target classes : CN(Control Non-responders), CR(Control Responders), TN(Treated Non-responders) and TR(Treated Responders)

The second step, machine learning modelling includes the techniques to generate results. After the dataset is separated as test and train, conventional churn prediction and uplift model prediction are made using XGBoost (Extreme Gradient Boosting) and Logistic Regression. Feature importance plots and results are obtained to see how effective the defined input variables are in estimating the target variable. This step is done for both conventional and uplift models.

For each customer, churn probability, CN, CR, TN and TR probability results are obtained. Lai's Generalized Weighted Uplift Method(LGWUM) Formula is used while training the uplifted churn model.Lai (2006) (P= Probability result)

$$UpliftScore = P(CN/C) + P(TR/T)–P(CR/C)–P(TN/T)$$

The final step is model evaluation which uses accuracy, Qini coefficient and Qini curve evaluation. While churn probability result is used for the conventional churn prediction, obtained uplift score is used for the uplift churn model. Confusion matrixes are obtained for the conventional churn model and for 4 different target classes. **'churn'** and **'prediction_churn'** are used for conventional customer churn prediction accuracy. **'target_class'** and **'prediction_target_class'** are used for the Uplift model accuracy. Therefore, while the conventional model predicts 2 outcomes, the uplift model predicts 4 outcomes.

As implemented by Radcliffe (2007), Qini curve constantly examines whether the model selects the correct target customer. (Consistent reliability) The Qini curve of the proposed model is evaluated using the random(normal) model, and a successful model has a Qini coefficient higher than 5%.(Radcliffe and Surry (2011)) The Qini-Coefficient is defined as the difference between the area under the Uplift Curve and the area under the random curve. To calculate the formula: x= population with treatment, N= total number of customers, uplift(x) = Nx[(TR/T)-(CR/C)]

$$qini\ coefficient = \sum_{n=0}^{N-1} uplift - random\ model\ curve$$

4 Design Specification

The machine learning framework architecture combines a predictive and prescriptive classification results as shown in Figure 3. Firstly, the raw data is collected and cleaned and this step followed by the feature engineering. Feature engineering includes both feature selection and construction since new variables are created and deleted. Since the data set is balanced in this state, sampling technique is not used. Also, the treatment column must be specified for the customer churn uplift model. Treatment must be actionable, (negatively) correlated with the target variable and available. Wijaya et al. (2021) Therefore, the "has contract" and "is tv subscriber" parameters are chosen as treatment, each being applied separately and treatment correlations are provided as percentages.

The "has contract" column is not in the data description. Definition of the "reamining contract" column in the dataset according to the source: *How many year remaining for customer contract. if null; customer does not have a contract. The customer who has a contract time have to use their service until contract end. If they canceled their service before contract time end they pay a penalty fare.* As explained in the methodology, the churn column (yes=1 or no=0) creates the control group in the uplift model application. For the treatment group, a binary classification with a result of 0 and 1 is also needed for the uplift model. The "reaiming contract" column has null values. As stated in the definition, each row indicates the remaining contract period. Null values mean that the customer does not have a contract. Therefore, by imputing null values with 0 and non-null values with 1, the new **"has contract"** column with a binary outcome is created.

A function is used for assigning the target classes. For instance, if the customer does not receive treatment and stays with the company, this group will be control responders(CR). Accordingly, the function has determined the target classes as follows:
'target class' = 0 (CN)
treatment == 0 & churn == 0 ,'target class' = 1 (CR)
treatment == 1 & churn == 1 ,'target class' = 2(TN)
treatment == 1 & churn == 0 ,'target class' = 3 (TR)

For the machine learning part, the new dataset is separated by 0.7 training and 0.3 testing data rates, and XGBoost and Logistic Regression classifiers are used to train the models with the data size of 71893 rows and 11 columns. Test data provides prediction and probability results using these classifiers. As mentioned in the methodology, LGWUM Formula is used while training the uplift churn model. The uplift score is utilized to observe the prescriptive performance analysis of the uplift model. *Uplift Score = P(CN/C)+ P(TR/T) – P(CR/C) – P(TN/T)* (P= Probability result)

For the conventional churn prediction model, prescriptive performance is analyzed using probability values in the Qini curve. As for the uplift model, prescriptive performance analysis aims consistent reliability and performed via Qini curve and Qini Coefficient that derived from uplift score. Finally, the accuracy results for both uplift and conventional customer churn prediction models are examined.

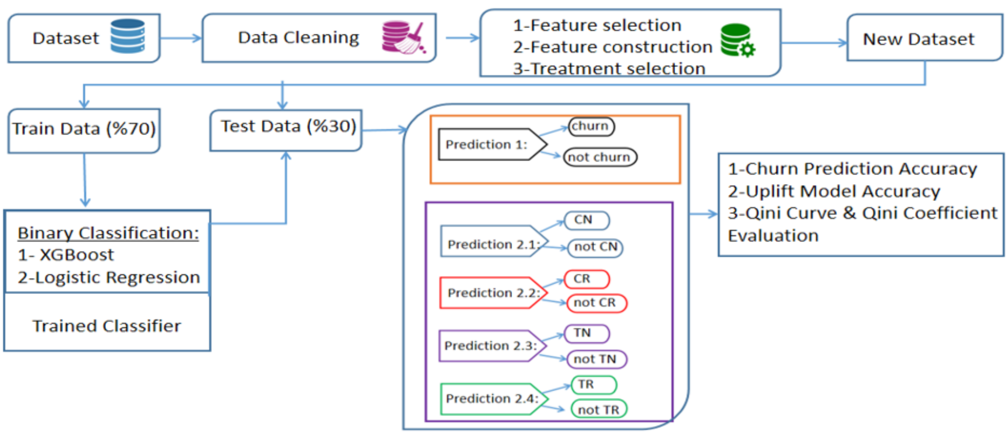

Figure 3: Research Design Specifications

Attribute Name	Description	Data Type
ID	customer ID	int64
is tv subscriber	has a TV subscription	int64
is movie package subscriber	has a movie package subscription	int64
subscription age	years of service usage	float64
bill avg	last 3 months bill average	int64
reamining contract	years left for the customer's contract	float64
service failure count	service failure calls for last 3 months	int64
download avg	last 3 months internet usage(GB)	float64
upload avg	last 3 months upload average(GB)	float64
download over limit	limit over count to be paid for the last 9 months	int64
churn	whether the customer has left the company	int64

Table 1: Research Dataset Explanation

5 Implementation

This section discusses the implementation of the machine learning models for conventional customer churn prediction and uplift customer churn prediction. Detailed explanation of this section is given in the configuration document. The data types and column descriptions before feature engineering are shown in Table 1. The produced outputs are confusion matrixes of 4 target classes, acuracies and Qini coefficients. Treatment correlation results are obtained after treatment identification. By investigating Control Non-Responders(CN) and Treated Responders(TR) "persuadables" group can be obtained. The most notable parts of this research are applying feature engineering for uplift modelling, using XGBoost and Logistic Regression Machine Learning models for uplift model and conventional model prediction.

The Integrated Development Environment (IDE) used for the implementation of this research is Google Colaboratory and the programming language used is Python (v.3.7.13). The main libraries that utilized are:Matplotlib (v.3.2.2), Pandas (v.1.3.5), Xgboost (v.0.90), Seaborn (v.0.11.2) and Sci-kit learn (v.1.0.2)

Hardware specifications : ASUS ZenBook UM425UA-AM164T, Storage: 512GB M.2 NVMe™ PCIe® 3.0 SSD, RAM: 8.0 GB, Processor: AMD Ryzen™ 5 5500U Mobile Processor (6-core/12-thread, 11MB cache, up to 4.0 GHz max boost), Operating System: Windows 10 Home.

6 Evaluation and Results

The evaluation metrics assess the performance of utilized machine learning algorithms for conventional churn prediction and uplift churn prediction. Also, the study of Wijaya et al. (2021), who used a similar method in the employee turnover problem, was repeated to compare with this research. Accuracy and Qini coefficient are the metrics to be used for evaluation. In addition, treatment correlation and Qini curve is examined for answering the research question and providing better insight.

The four target classes : CN(Control Non-responders), CR(Control Responders), TN(Treated Non-responders) and TR(Treated Responders). Since the purpose of Uplift modeling is to target persuadables, CN+TR creates the "persuadables" group.

	Treatment Correlation	Churn Accuracy	Uplift Accuracy	Churn Qini Coef.	Uplift Qini Coef.
Previous Research Data1	-6.18%	97.18%	95.16%	-138.07 %	621.87 %
Previous Research Data2	-24.61%	87.30%	61.22%	31.23 %	23.17 %
XGBoost Treatment 1	-47.28%	93.84%	81.34%	-3672.73 %	4297.1 %
XGBoost Treatment 2	-32.94%	93.90%	78.94%	-2536.44 %	2968.72 %
Logistic R. Treatment 1	-47.28%	87.25%	73.06%	-3499.51 %	-3008.26 %
Logistic R. Treatment 2	-32.94%	87.24%	71.42%	-2432.23 %	778.8 %

Table 2: Overall Evaluation Results: For previous research instead of churn, employee turnover prediction is performed.

		Extreme Gradient Boosting				Logistic Regression			
		TREATMENT 1		TREATMENT 2		TREATMENT 1		TREATMENT 2	
		P.True	P.False	P.True	P.False	P.True	P.False	P.True	P.False
Conventional Model	A.True	8864	555	8893	526	7807	1811	7902	1614
	A.False	773	11376	789	11360	939	11011	1137	10915
Uplift (CN)	A.True	13967	1582	16809	1094	13028	2652	16991	1010
	A.False	1517	4502	2055	1610	1191	4697	2548	1019
Uplift (CR)	A.True	20980	40	21139	26	20964	37	21102	48
	A.False	502	46	304	99	540	27	392	26
Uplift (TN)	A.True	13797	1641	10598	2486	13272	2234	9369	3714
	A.False	1974	4156	1741	6743	3007	3055	1935	6550
Uplift (TR)	A.True	11935	762	11616	936	11630	887	11077	1393
	A.False	32	8839	442	8574	1072	7979	1290	7808

Table 3: Confusion Matrix of Conventional Churn Prediction Model and Target Class Categories using XGBoost and Logistic Regression

6.1 Experiment 1: Replication of the State of the Art- Employee Turnover Uplift Model Using XGBoost

Replication of this research is implemented with the notebook provided by the author as open source on Kaggle.com.[2]

The aim of this experiment is to observe whether the research replicable or not. Also, this step also justifies the methodology of based research paper. The replicated research's aim is to predict employee turnover and its uplift categories. 3 independent datasets are used in the study, but since only two are available, the study is repeated with 2 datasets (Dataset 1[3] and 2[4]). Data labelling is apllied as a part of feature engineering for both datasets. As seen in Table 2 ETP(Employee Turnover Prediction) models' prediction accuracies are higher than ETU (Employee Turnover Uplift) model for both datasets. While ETP accuracies are %97.18 and %87.30, uplift model accuracies are %95.16 and %61.22.

The selected treatment column for Dataset 1 is 'promotion last 5years' and treatment correlation with target variable is -6.18. Furthermore, treatment column for Dataset 2 is 'OverTime' and treatment correlation with target variable is -24.5. A larger treatment correlation indicates that the effect of the selected variable is greater in the uplift model. A negative value indicates that it will have a negative impact on employee turnover.

[2]website: `https://www.kaggle.com/code/davinwijaya/why-you-should-start-using-uplift-modeling/notebook`

[3]website: `https://www.kaggle.com/code/davinwijaya/why-you-should-start-using-uplift-modeling/data?select=HR_comma_sep.csv`

[4]website: `https://www.kaggle.com/code/davinwijaya/why-you-should-start-using-uplift-modeling/data?select=WA_Fn-UseC_-HR-Employee-Attrition.csv`

When the results are examined, the ETP models can predict whether the employee is turnover or not, the ETU models predict four outcomes (Persuadables, Sure Things, Lost Causes, and Sleeping Dogs/Do-not-disturbs). Employees with the highest turnover probability will be targeted with a treatment.The ETU models are ranked by its uplift score with LGWUM's formulation. The Qini curves demonstrate that the ETU model outperforms the ETP model in terms of uplift value.

The Qini curves show the best proportion of employees to be treated with the employee retention strategy. The random model(grey line) shows the value of the uplift without utilizing any model. For Dataset 1,the ETP model underperforms the random model, that means the model could not target the right employees for retention program. For example, the Qini curve for Dataset 1 in Figure 4 should be interpreted as follows: a 0.1 uplift rate in 0.25 employee ratio suggests that the company will receive a 10% uplift simply by targeting the top 25% of employees with the treatment plan.

When the Qini curve of Dataset 2 examined, it is observed that both ETP and ETU provide good results for targeting the right employee. This inference can be made more comprehensively by examining the Qini coefficient scores in Table 2.

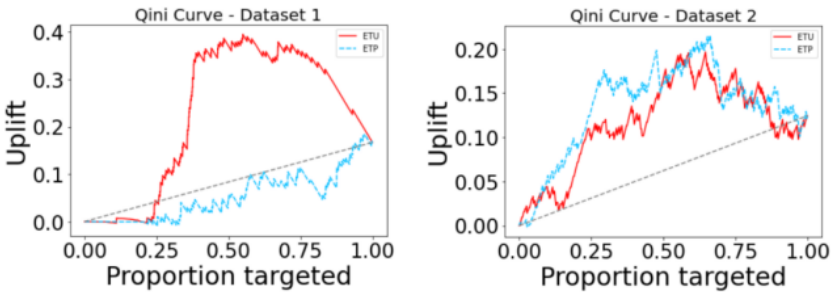

Figure 4: Qini curves of replicated research

6.2 Experiment 2 :Customer Churn and Uplift Categories Prediction Prediction Using XGBoost with 2 Different Treatments

The dataset used in Experiment 1 is changed to "internet service providers customer churn" dataset and different data pre-processing steps were applied. (Steps are explained in the methodology.) Since data labelling is not needed for this dataset, related part from Experiment 1 is also skipped. This experiment includes 2 experiments in itself. All other steps are the same except that the treatment groups are selected differently and XGBoost is applied as machine learning methodology.

For the experiment 2.1 "has contract" is chosen as treatment group.(Highest negative correlation with the target variable is "reaiming contract" and "has contract" is its derivation). The experiment 2.2 uses "is tv subscriber" as the treatment column. Confusion matrixes for Experiment 2 can be seen in Table 3 Extreme Gradient Boosting section. In order to target "persuadables" Control Non-Responders and Treated Responders should be targeted.

According to the results in Table 2, uplift model accuracy is less than conventional model accuracy in experiments with all treatments. Wheres the uplift model accuracy

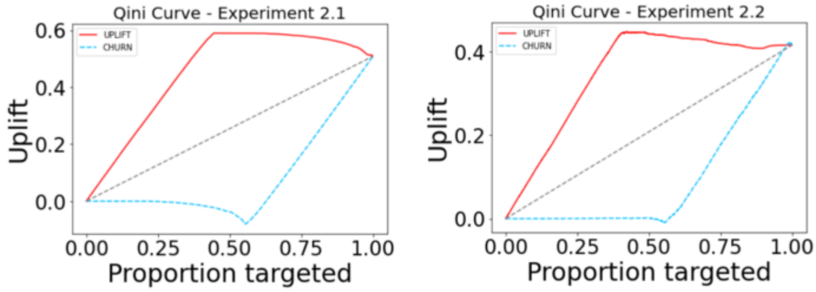

Figure 5: Qini curves of experiments 2.1 and 2.2

for treatment 1 is 81.34% , the conventional churn model accuracy is 93.84%. Treatment 1(has contract) has the highest treatment correlation with -47.28. Therefore, experiment 2.1 has the highest difference between accuracy results and highest uplift Qini coefficient with 4297.1%.

Although treatment 2 (is tv subscriber) has a lower treatment correlation(-32.94), conventional model and uplift model accuracies are, 93.90% and 78.94% respectively. As seen in Figure 5, the Qini curve of Experiment 2.2 is slightly different from the previous one. When 50% of the customers are targeted, an uplift of about 40% is achieved. However, this rate is approximately 60% for treatment 1. A more uplift rate can be reached by targeting the same customer proportion. Due to the fact that the treatment correlation is lower, uplift Qini coefficient is 2968.72% for treatment 2. It is necessary to work with more customer ratio to reach the maximum uplift rate when applying treatment 2. Finally, both uplift models are 100% successful in targeting the "persuadables" and eliminating "do-not-disturb" group. (Qini curve is above the random model and slopes upwards). On the other hand, conventional churn prediction models fail to target the correct group of customer.

6.3 Experiment 3 :Customer Churn and Uplift Categories Prediction Prediction Using Logistic Regression with 2 Different Treatments

Important Note: *The logistic regression models did not converge, however, results were stable, when the same steps as XGBoost were applied. Due to the large number of poorly fitting observations, there is a lack of convergence, that means the data does not fit the model properly. Therefore, the maximum number of iterations for logistic regression has been increased (1000 max. iteration for the conventional churn model, and 10000 max. iteration for the uplift model).After iteration change, model fitting was provided.*

As disadvantages of logistic regression: multicollinearity poses a problem, and in linearly separable datasets the probabilities are forced to the boundary (0 and 1). (These results were seen in Experiment 3.1 and specified in the configuration manual.) In order to alleviate this issue, "stratify=df['treatment']" was added in Experiment 3.1 and 3.2. The 'stratify' parameter maintains that the ratios of the main dataset match those of the train and test data. Therefore, 'stratify' was utilized to establish a relationship between the train-test datasets and to decrease sampling bias. It does not affect the main outcome since the dataset is balanced, however better probability results for 4 target classes are observed in Experiment 3.1. Also higher Qini coefficiient is obtained for Experiment 3.2 *uplift model.*

The same dataset is used in Experiment 2 and 3. Therefore, the preprocessing steps described, treatment 1 and 2 are exactly the same. Differently, Logistic Regression is applied as a machine learning methodology. Also, further applications are performed due to the convergence problem. Confusion matrixes for Experiment 3 can be seen in Table 3 Logistic Regression section. In order to target "persuadables" Control Non-Responders and Treated Responders should be targeted. When treatment 1 is applied, churn accuracy is %87.25 and uplift model accuracy is %73.06.(Table 2) Accuracy results for treatment 1 and 2 are nearly the same, due to the fact that 'stratify' parameter is used during train-test data split for logistic regression. (After defining the maximum iteration in logistic regression, uplift accuracy increased by approximately %4. However, the conventional churn accuracy score remained almost the same.) Also, the conventional and uplift accuracy results using XGBoost are higher than those obtained with logistic regression.

When Uplift Qini coefficients are evaluated, it is observed that XGBoost performs more successfully than Logistic Regression. In the logistic regression model, successful uplift was obtained when the model did not converge.Figure 6. However, even after convergence was fixed, logistic regression failed to provide positive uplift for treatment 1. Treatment 2(is tv subscriber) can reach the maximum uplift value(%778.8 Qini coefficient) with uplift model when targeting 100% of the customers. One of the purposes of the uplift model is to achieve successful results by targeting a small number of customers. Logistic regression is also not sufficient in this regard.

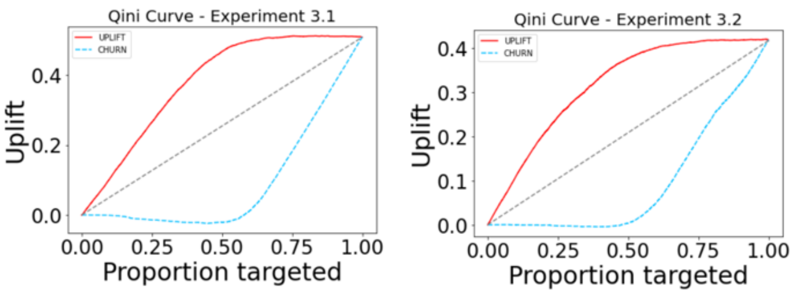

Figure 6: Qini curves of experiments 3.1 and 3.2 when the logistic regression model does not converge

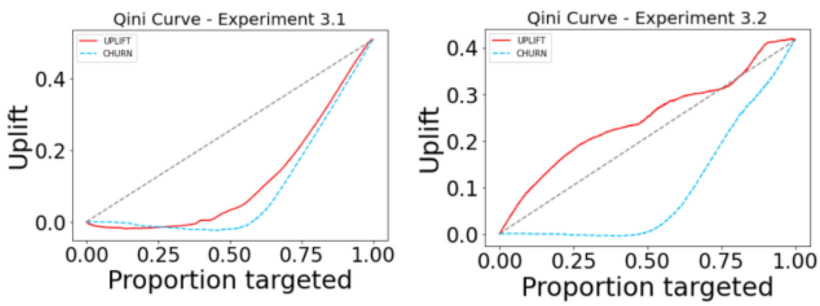

Figure 7: Qini curves of experiments 3.1 and 3.2 after maximum iteration change

When Experiment 3.2 is examined in Figure 7, the effect of treatment 2 on targeting the right customer group can be observed. If the graph is interpreted, 0.25 uplift rate in 0.5 customer proportion suggests that the company will receive a 25% uplift by targeting the 50% of customers with the treatment plan. Although it is a disadvantage that the uplift rate is low and the maximum uplift is achieved when all customers are targeted, the model for treatment 2 still shows %100 success for targeting the "persuadables" instead of "do-not-disturb" which provides consistent reliability.

7 Discussion

The conventional churn prediction approaches can only distinguish between customers who are likely to churn and those who are not, whereas uplift modeling distinguishes between customers who will benefit from being treated and those who will not. An uplift model using LWGUM's formula has been applied for employee retention before.(replicated in 6.1) In this study, uplift modeling was applied to solve the customer churn problem using a similar methodology. This study's results agree with the previous project.

The uplift model's prescriptive performance was more successful and predicted more successfully which customer group to target for the customer retention strategy. (Prescriptive performance is measured with the Qini coefficient, while predictive performance is measured with accuracy.) Although logistic regression with treatment 1 results in a negative uplift, it is still higher than the conventional churn model uplift. On the other hand, the customer churn uplift model applied in this paper underperformed the conventional customer churn prediction model in terms of accuracy. The reason is that two outcome predictions are easier than four outcomes.

However, using predictive analysis is not sufficient for customer retention strategies. For instance, let the customer be in the "churn" class according to the conventional prediction result. Whether these are Control-Responders (CN) or Treated Non-responders (TN) can not be known with the conventional model. In other words, the customer's response to the applied "treatment" can not be known. Also, applying a retention campaign (treatment) to customers who will stay with the company under any circumstances is an unnecessary expense for the company. The worst-case scenario is that the customer who stays in the company without the treatment leaves the company after receiving the treatment.(Do-Not-Disturb group) However, when the uplift model is used, persuadables (Control Non-responders and Treated Responders) can be known. The customer, who was previously in the "churn" category, can switch to the "not churn" category after the treatment is applied.

In this project two different customer retention treatments were applied for the uplift model. Treatment 1 (has contract) indicates whether the customer selected the contract or not. It has the largest negative treatment correlation. That means, when "has contract" is TRUE(1), the customer churn is likely to be decreased if the company offers retention campaign using "has contract". Also, the effect is great due to the treatment correlation. Qini curve in Figure 5 shows that the uplift model can achieve %60 uplift, targeting about % 50 of customers with the retention strategy.(Experiment 2.1) The Qini curve is also utilized to examine how the models prescribe customers in each category, since targeting the "persuadables" will cause the curve to rise above random model. However, targeting the do-not-disturb will have the reverse effect. The uplift model outperformed the conventional customer churn prediction model in all experiments, demonstrating that it correctly distinguishes between "persuadables" and "do-not-disturb" customers except experiment 3.1.(Figure 7) Conversely, the conventional churn model's Qini curve is below the random model. Therefore, it failed to target right group of customers in all experiments.

As seen in Figure 5, treatment 2 (is tv subscriber) showed similar results to treatment 1(has contract) in terms of predictive and prescriptive results with XGBoost. However, negative treatment correlation is low compared to treatment 1. For instance, the uplift model with treatment 2 can achieve %40 uplift, targeting about % 50 of customers with the retention strategy. However, both identified treatments can be used as a part of customer retention strategy, due to the fact that they both contributes to decreasing customer churn. If the internet service provider company offers its customers a campaign about having contracts or TV subscriptions, customers who are considering leaving the company can stay with the company.

When the results of different machine learning models are compared: It was observed that XGBoost outperforms logistic regression in both accuracy and Qini coefficient results. Among all experiments, the most successful results are obtained when XGBoost is used and the treatment is selected as "has contract". This experiment yields maximum uplift of approximately %60. The maximum uplift obtained with logistic regression was approximately %40 by using treatment 2(is tv subscriber). However, there is a drawback that 100% of the customers should be targeted with campaign to receive the mentioned maximum uplift. The reason is that one of the major benefits of utilizing the uplift model application is to reach maximum uplift and efficiency by targeting a small ratio of customers. XGBoost is a tree-based method for classification and prediction. Logistic Regression, on the other hand, is a linear technique that uses a generalized linear equation to define the direct relationships between a range of variables. Therefore, while logistic regression requires preprocessing in this project, XGBoost does not need scaling and intense data preprocessing. The major limitation of logistic regression is the assumption of linearity between the dependent variable and the independent variables since most datasets are not linearly separable. Even if they are, in some cases the model aims for the perfect separation and obtained probability results are nearly 0 or 1, as in this project before revision.(Figure 6) These disadvantages stated for logistic regression were observed in multiclass prediction. In this project, 4 different probabilities were used while calculating the uplift score. In the application performed with treatment 1, "CN" and "CR" probabilities were close to 0, while "TN" and "TR" had values close to 1. This situation caused the uplift score and the Qini curve to shift to negative in Experiment 3.1.(Figure 7) Although the techniques described in subsection 6.3 have been tried to solve this problem, logistic regression does not provide reliable results like XGBoost in uplift model applications.

Finally, this research's contributions to the literature are:
Investigating the customer churn problem using Lai's generalized weighted uplift method, comparing the uplift model with the conventional customer churn model by performing prescriptive and predictive analysis, showing the effect of more than one treatment in the uplift model application.

8 Conclusion and Future Work

While conventional customer churn analysis can predict whether the customer will leave the company or not, it can not predict the response to an applied customer retention strategy. At the same time, after the strategy implemented, the decision of the customer to stay in the company may also change. Uplift model can estimate the net effect of this customer response. In this study, when the uplift model and the conventional model were compared using Lai's generalized weighed uplift method (LGWUM), it was observed that the uplift model is more successful than the conventional model in targeting the persuadables(customers who are likely to churn but can be retained with the campaign). Although the conventional model gives more successful results in predictive analysis, it can not reduce the churn rate with the retention program and fails to target the correct group of customers. Also, conventional customer churn prediction model is likely to target do-not-disturb customer group. This situation causes even the customers seen in the ”churn” category to leave the company and the company to make a loss.

When machine learning techniques are compared, Extreme Gradient Boosting (XGBoost) outperformed Logistic Regression in all experiments in terms of accuracy and Qini coefficient. In addition, one experiment with logistic regression failed to target ”persuadables” for both uplift and conventional churn models. Moreover, two different treatments were used in the uplift model, one for each experiment. Having higher correlation(%), ”has contract” has more uplifts than ”is tv subscriber” using XGBoost. Therefore, it can be said that high uplift accuracy enables high Qini curve and hence high uplift value. However, the internet service provider company can apply both treatments, since both of them achieved %100 success in selecting the target customer class . Thus, customers in ”churn” can be converted to ”not churn”. Thanks to the uplift model, since they only target the persuadables group, the customer retention strategy can be successfully implemented with minimum cost and effort.

The future work for this research project can include using different sizes of datasets to see the effect of the machine learning models. Therefore, for a large dataset artificial neural network techniques with uplift modelling can be performed. Also, other common machine learning techniques can be applied for comparison. In addition, other uplift modelling approaches can be utilized such as pessimistic uplift, two-model uplift and logit leaf model(LLM). Also, different evaluation metrics can be used like maximum profit uplift(MPU). Finally, this study investigating the customer churn problem can be used in many different areas by using the uplift model. Student grade prediction, credit-worthiness assessment and fraud detection can be given as examples.

References

Ahmad, A. K., Jafar, A. and Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform, *Journal of Big Data* **6**(1): 1–24.

Asar, A. (2019). *Uplift Models: Can They Be Used to Identify and Rank Heart Failure Patients Expected to Benefit From a Clinical Telehealth Program?*, PhD thesis, Tilburg University.

Babu, S. and Ananthanarayanan, N. (2016). A review on customer churn prediction in telecommunication using data mining techniques, *International Journal of Scientific Engineering and Research (IJSER)* **4**(1): 35–40.

Çelik, O. and Osmanoglu, U. O. (2019). Comparing to techniques used in customer churn analysis, *Journal of Multidisciplinary Developments* **4**(1): 30–38.

De Caigny, A., Coussement, K. and De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees, *European Journal of Operational Research* **269**(2): 760–772.

De Caigny, A., Coussement, K., Verbeke, W., Idbenjra, K. and Phan, M. (2021). Uplift modeling and its implications for b2b customer churn prediction: A segmentation-based modeling approach, *Industrial Marketing Management* **99**: 28–39.

Devriendt, F., Berrevoets, J. and Verbeke, W. (2021). Why you should stop predicting customer churn and start using uplift models, *Information Sciences* **548**: 497–515.

Devriendt, F., Moldovan, D. and Verbeke, W. (2018). A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics, *Big data* **6**(1): 13–41.

Do, D., Huynh, P., Vo, P. and Vu, T. (2017). Customer churn prediction in an internet service provider, *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, pp. 3928–3933.

Fogg, A. (2016). Anthony goldbloom gives you the secret to winning kaggle competitions.

Gubela, R., Bequé, A., Lessmann, S. and Gebert, F. (2019). Conversion uplift in e-commerce: A systematic benchmark of modeling strategies, *International Journal of Information Technology & Decision Making* **18**(03): 747–791.

Gubela, R. M. and Lessmann, S. (2021). Uplift modeling with value-driven evaluation metrics, *Decision Support Systems* **150**: 113648.

Jaskowski, M. and Jaroszewicz, S. (2012). Uplift modeling for clinical trial data, *ICML Workshop on Clinical Data Analysis*, Vol. 46, pp. 79–95.

Kane, K., Lo, V. S. and Zheng, J. (2014). Mining for the truly responsive customers and prospects using true-lift modeling: Comparison of new and existing methods, *Journal of Marketing Analytics* **2**(4): 218–238.

Karvana, K. G. M., Yazid, S., Syalim, A. and Mursanto, P. (2019). Customer churn analysis and prediction using data mining models in banking industry, *2019 International Workshop on Big Data and Information Security (IWBIS)*, IEEE, pp. 33–38.

Khan, A. A., Jamwal, S. and Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider, *International Journal of Computer Applications* **9**(7): 8–14.

Lai, L. Y.-T. (2006). *Influential marketing: a new direct marketing strategy addressing the existence of voluntary buyers*, PhD thesis.

Lalwani, P., Mishra, M. K., Chadha, J. S. and Sethi, P. (2021). Customer churn prediction system: a machine learning approach, *Computing* pp. 1–24.

Radcliffe, N. (2007). Using control groups to target on predicted lift: Building and assessing uplift model, *Direct Marketing Analytics Journal* pp. 14–21.

Radcliffe, N. J. and Surry, P. D. (2011). Real-world uplift modelling with significance-based uplift trees, *White Paper TR-2011-1, Stochastic Solutions* pp. 1–33.

Rombaut, E. and Guerry, M.-A. (2020). The effectiveness of employee retention through an uplift modeling approach, *International Journal of Manpower* .

Shaar, A., Abdessalem, T. and Segard, O. (2016). Pessimistic uplift modeling, *arXiv preprint arXiv:1603.09738* .

Tang, Q., Xia, G., Zhang, X. and Long, F. (2020). A customer churn prediction model based on xgboost and mlp, *2020 International Conference on Computer Engineering and Application (ICCEA)*, pp. 608–612.

Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. and Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction, *Simulation Modelling Practice and Theory* **55**: 1–9.

Wang, Q.-F., Xu, M. and Hussain, A. (2019). Large-scale ensemble model for customer churn prediction in search ads, *Cognitive Computation* **11**(2): 262–270.

Wijaya, D., DS, J. H., Barus, S., Pasaribu, B., Sirbu, L. I. and Dharma, A. (2021). Uplift modeling vs conventional predictive model: A reliable machine learning model to solve employee turnover, *International Journal of Artificial Intelligence Research* **5**(1): 53–64.