A Recommender System for the Upselling of Telecommunications Products

Navin Dookeram, Zahira Hosein and Patrick Hosein
The University of the West Indies
St. Augustine, Trinidad and Tobago
ndookeram@yahoo.com, zhosein@tstt.co.tt, patrick.hosein@sta.uwi.edu

***Abstract*—Telecommunication providers are always seeking ways to upsell products to corporate customers. Traditionally, the telecommunication provider's Account Managers build a business relationship with the customers and try to persuade them to upsell. However, only some instances result in a successful upsell while others are unsuccessful. First, we focus on a binary classification framework for predicting the successful upsell of products and services, using data from one such telecommunications provider. Through this prediction model, we illustrate a recommender system for voice products/services to corporate customers of the telecommunications company. We use a logistic regression classifier to automate the selection of customers that are most likely to upsell. We also acknowledge that there may be monetary costs associated with misclassification errors. Note that minimizing losses (or maximizing revenue) may conflict with the objective of minimizing errors and so we address this trade-off. We apply our predictive model to recommend a set of target customers to approach for upsell, illustrating the different accuracy results for different cost weightings. We also show that the success rate of upselling products to the selected customers is dramatically improved when compared to the traditional approach.**

***Index Terms*—Recommender System, Binary Classification, Cost Optimization, Machine Learning, Upselling, Telecommunications**

I. Introduction

The telecommunications industry provides a technological foundation for societal communications and interconnection, and can significantly contribute to economic growth in a developing country [1]. Sustaining profits in a world of increased competition, market liberalization, and increased customer expectations is challenging, as demonstrated by [2]. Many telecommunication products have also matured and their revenue is in decline. However, a telecommunications service provider can find ways to increase their revenue in saturated markets with the aid of data analytics [3].

Here, we consider upselling as the purchase of additional products and add-ons, and/or the switch from 'legacy' products to upgraded ones. Traditional methods of upselling rely on a semi-manual approach to marketing such products to clients. One such provider assigns an Account Manager (AM) to corporate customers. This AM manages the client's sales, resolves problems, and maintains a direct line of communication, with the intention to build trust and a strong customer relationship. Each AM will meet with their customers on a scheduled basis to discuss their needs, problems, and new product offerings which could result in upsell opportunities.

This semi-manual approach aims to determine which customers are likely to purchase or upgrade their products. However, there is a deficiency in this approach since it often relies on the AM's knowledge and subjective reasoning. Some AMs may not completely understand each customer's preferences, and this could lead to inappropriate recommendations resulting in a loss of trust, decreased customer satisfaction, and ultimately the risk of losing the customer's business (churn) [4].

With the use of machine learning an upsell binary classifier addresses these problems by reducing the time spent on searching for viable customers. This is done through the automatic identification of the top customers to target for upsell. However, a binary classifier is usually designed to maximize accuracy, defined as the ratio of number of correct predictions to the total number of predictions. This results in treating all types of misclassification errors as having equal cost.

However, in the case of predicting whether a customer is likely to upsell, the cost of a false positive may be much less than that of a false negative. In other words, an instance of a false positive, given by the incorrect prediction a successful upsell, results in a waste of time, money, efforts and other resources in marketing and approaching the customer since targeting this customer would fail to result in a sale. There may be further costs associated with this false positive as approaching an uninterested customer can lead to customer dissatisfaction and can cause loss of customer trust and long term profitability [4].

Alternatively, an instance of a false negative, given by the incorrect prediction of an unsuccessful upsell, results in a missed opportunity to generate revenue from sales. This can also have an unwanted impact on the relationship between the provider and this customer and can encourage customer churn, since the customer may feel that their needs are not being satisfactorily met. In order to address this issue, we modify a standard binary classifier to incorporate class-dependent misclassification costs where the objective is to minimize the average cost of the predictions given by the model.

II. Related Work and Contributions

Previous research has been done on predicting customer purchase behaviour in the telecommunications industry, with many focusing on customer retention and product recommender systems [5]–[7]. Additionally, a few studies have focused on using machine learning to predict a customer's willingness to upsell or purchase a product in the telecommunication industry.

For example, [8] used a support vector machine (SVM) model to predict the probability of a customer falling into the classification behaviours of churn and non-churn. It was hypothesized that non-churn customers were more likely to have high appetency (willingness to buy new products/services) and/or buy upgrades or add-ons (upselling). Posterior probabilities of a customer belonging to the non-churn class were used to determine a customer satisfaction score. Customers with the highest satisfaction levels were classified as more likely to remain loyal, upgrade and buy new products.

[9] suggests that the likelihood of upsell may be determined via random forest (RF) or gradient boosted machine (GBM) machine learning techniques where multiple decision trees are used to output an indication of whether upselling a product has a probability of success exceeding a predetermined threshold. Ranking of various customers as possible upsell targets can be achieved using probabilities outputted by the model. Meanwhile, [10] developed advanced analytics tools to predict whether a customer will make purchase within a certain time frame in the near future.

Additionally, [11] attempted to predict credit card upselling using machine learning procedures. Finally, [12] analysed billing data of corporate customers in a large telecommunications company to predict high-value upsell customers. This was done by concatenating examples from multiple time periods and weighting important customers based on the dollar value of take-up of products over a given time period. However, none of these studies directly account for cost of misclassification errors when building the predictive model.

There are a number of ways to account for misclassification costs in a predictive model, including cost-sensitive approaches where the cost is directly accommodated in a model's cost function (for example [13]), or rebalancing (achieved via class weighting, under-sampling, or over-sampling techniques) where observations from the less costly class are given higher importance during training (for example [14]). However, a cost-insensitive approach was chosen for this study whereby probability estimates, outputted by the classifier that learned from the training set as given, are used to obtain a probability cut-off or threshold in order to compute cost-optimal decisions [15].

III. Description of Dataset

The dataset was provided by a telecommunications service provider and applies to corporate customers. The dataset used in this study contains basic customer demographic information, transaction data and sales records concerning legacy products that the customer purchased. Here, a legacy product refers to a version of a product that is not current or modern. There are 3,117 observations with 31 variables. A brief description of the data is given in Table I. Demographic data involved characteristics of the customer such as the type, sector and age. Transaction data involved the customer's product purchases and rentals. Sales data shows records of whether the customer was approached for upsell and if this approach was successful.

TABLE I
Data Attributes

Variable	Data Type	Description
Parent_ID	factor	Conglomerate ID
Company_ID	factor	Company Name
Sector	factor	Marketing sector
Company_Age	integer	Company Age in Years
Company_Type	factor	Industry Classification
Industry	factor	Industry Classification
Sub_Industry	factor	Industry Classification
VIP_Customer	factor	Signed a VIP contract?
MonthsBeforeVIPExpir	integer	Months until VIP expires
Last_CIS_Install	integer	Days since last CIS install
Count_CIS_Exchanges	integer	Number of CIS exchanges
A_N_PBX	integer	Count of Brand A PBX
C_PBX	integer	Count of Brand B PBX
M_PBX	integer	Count of Brand C PBX
SLine	integer	Count of Single Lines
T1	integer	Count of T1s
Trunk	integer	Count of Trunk Lines
LegacyProducts	factor	List of Legacy Products
PercentLegacyProducts	double	% of Legacy Products
PercentLegacyRental	double	% of Legacy Rental
LegacyRental	integer	Amount of Legacy Rental
LegacyMins	integer	Amount of Legacy Mins
PercentLegacyMins	double	% of Legacy Mins
Has_A_N_PBX	factor	Customer has Brand A PBX?
Has_C_PBX	factor	Customer has Brand B PBX?
Has_M_PBX	factor	Customer has Brand C PBX?
Has_SLine	factor	Customer has Single Lines?
Has_T1	factor	Customer has T1s?
Has_Trunk	factor	Customer has Trunk Lines?
Num_Products	factor	Total number of products
Approached for UpSell	factor	Approached for upsell?
Success	factor	Successful upsell?

Data cleaning and pre-processing were done before training was attempted. Uninformative variables such as Parent_ID and Company_ID were dropped since these variables had no information value. Industry and Sub_Industry were also dropped due to the overwhelming number of factor levels. In future work, the Industry and Sub_Industry variables may be retained for model building if appropriately treated, for example grouping similar factor levels into fewer factor levels.

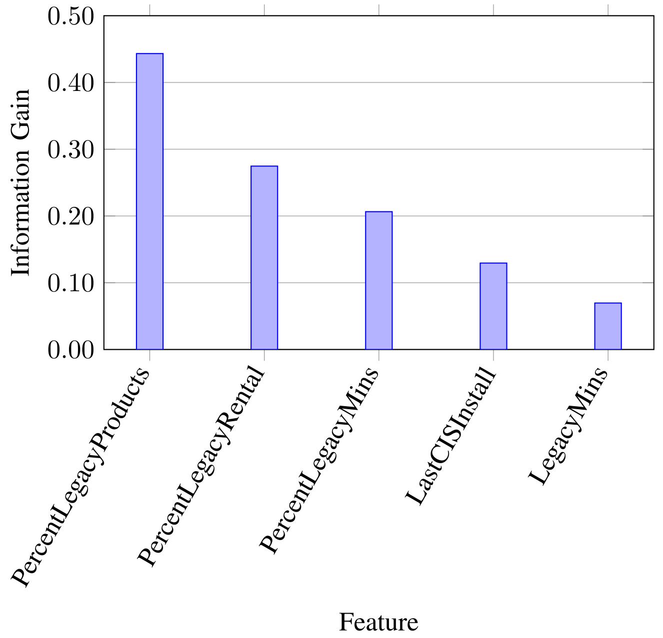

Fig. 1. Features chosen based on Importance Values (Only the five features shown had information gain greater than 0)

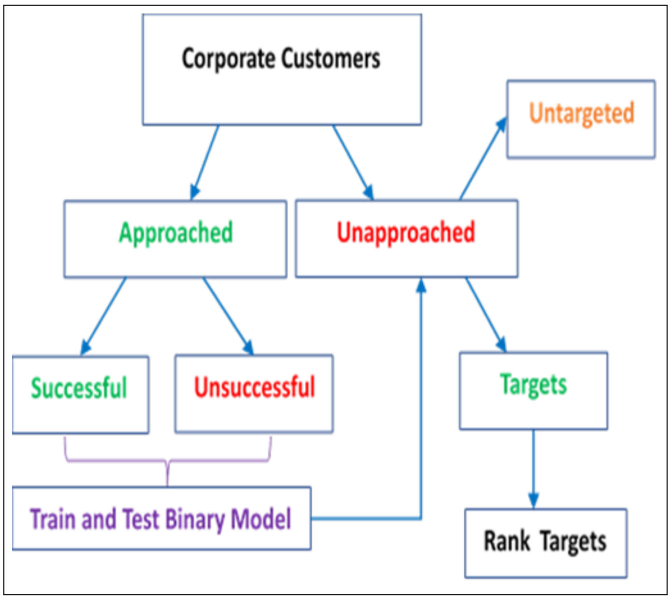

Fig. 2. Flowchart showing the general procedure for recommending upsell customers

Some missing values also occur in the numeric variables PercentLegacyRental, LegacyMins and PercentLegacyMins. All missing numerical values were imputed using the KNN (k-nearest neighbours) method with the caret package in R. The k-nearest neighbour algorithm can be used for imputing missing data by finding the k closest neighbours to the observation with missing data and then imputing them based on the non-missing values in the neighbours. Also, all numerical columns were centered and scaled, and every categorical (factor) variable was converted to numerical using dummy variable coding (one hot encoding).

IV. Feature Selection

It is not only important to determine which customers should be approached for upsell, but it is also important to determine the factors that correlate with those customers that upsell. This can aid in the further understanding of customer purchase patterns. Before model construction, feature selection was performed so as to utilize the most important features of the customer data to maximize the model's predictive power.

The generateFilterValuesData function from the FSelectorRcpp package in R was used to compute a rank of the variables by assigning an importance value to each feature with the use of information gain. The top five features were found to be the most important with a positive information gain as shown in Figure 1, while the information values for all other variables were computed to be zero. Thus only the top five important variables were used for building the logistic regression model.

V. Model Description

The data was divided into two subsets as determined by the 'Approached for Upsell' variable. There were 270 customers who were approached for upsell during February 2018 – February 2019, while the remaining 2847 customers were not approached for upsell during this time period. Figure 2 shows the general methodology for approaching customers to upsell.

We first split the data into two sets: one containing customers who have been approached for upsell in the past and one containing customers who have not yet been approached. Using the subset of customers who were already approached for upsell, the data was then split into a training and test set with a 70%/30% split, so the training set had 190 observations and the test set had 80 observations. We holdout the test set so as to give an idea of how well the model will perform on unseen data during deployment. The trained and tested model was then applied to customers who have not yet been approached. This would result in the generation of new targets which would then be ranked based on whether they have a high probability of choosing to upsell. There will of course still be a set of untargeted customers which the model will not recommend for approach.

The evaluation of performance metrics was done using repeated k-fold cross-validation with k = 10 using 3 repeats. We trained several supervised classification algorithms on the training set and monitored the performances of each algorithm, using the train function from the caret package in R. These algorithms included Naive Bayes, Support Vector Machine, Logistic Regression, Random Forest and Gradient Boosting. All algorithms gave similar performance. In the end, the Logistic Regression model was selected due to its high accuracy, simplicity and ease of interpretation. In this study, the predictive model is governed by the constraint or objective of minimizing the cost of incorrect predictions when approaching customers for upsell.

In this particular study, both short-term and long-term costs

TABLE II
COST MATRIX STRUCTURE

		Predicted	
		No	Yes
Actual	No	c_{00} (TN)	c_{01} (FP)
	Yes	c_{10} (FN)	c_{11} (TP)

TABLE III
CONFUSION MATRIX WHEN MINIMIZING ERROR

		Predicted	
		No	Yes
Actual	No	303	24
	Yes	36	207

of each classification are not available as it is difficult to measure or estimate. This is due to the fact that there is a wide range of the number and types of products a customer may purchase, and thus the price varies significantly with each instance thus making it challenging to estimate an average cost or benefit lost or gained when a product is sold to a customer. There is also a cost of a customer's dissatisfaction or benefit of a customer's satisfaction that should be considered, since this can also impact on long-term profitability. However, this variable is not easily translated into a tangible cost or benefit.

Suppose we penalize classification of false positives and false negatives by weighting with costs. Let c_{ij} denote the cost of classifying an instance of class i as class j. We assume that if $i = j$ (in which case the classification was correct) then the cost, or rather the benefit, is non-positive (i.e. $c \leq 0$) while if $i \neq j$ then the cost is positive (i.e., $c > 0$).

Let the cost of a false positive (FP), that is, incorrectly predicting an upsell as successful, be c_{01} and let the cost of a false negative (FN), that is, incorrectly predicting an upsell as unsuccessful, be c_{10}. Now c_{01} may only account for administrative costs such as the total resources expended when reaching out to the customer for upsell. This cost is assumed to be small. On the other hand, c_{01} is the foregone benefit that would have been obtained if the customer was approached for upsell, and this benefit would be the sum of revenue gained from the sale of the product. The monetary value of this foregone benefit is assumed to be much larger than the cost of approaching a customer. Note that cost associated with customer dissatisfaction due to these misclassification errors should also be acknowledged. Finally, the cost of a correct classification is assumed to be 0 since no loss is incurred. Hence the cost of a true positive (TP) is $c_{11} = 0$ and the cost of a true negative (TN) is $c_{00} = 0$. A cost matrix with classification of predicted versus actual values is structured as shown in II. Here, the classification costs and the types of classification are outlined.

Now let x denote the feature vector of a given instance. Then the binary classifier will output a continuous score $s(x)$ which will be used to determine the class in which the instance belongs. We assume that the binary classifier generally produces higher scores for class 'No' than for class 'Yes'. One must then determine some threshold t such that if $s(x) \geq t$ the instance is classified as 'No' while if $s(x) < t$ then the instance is classified as 'Yes'. For the logistic regression classifier, we denote the probability density function of the scores for class 'No' instances by $f_0(s)$ and for class 'Yes' scores by $f_1(s)$. We denote the corresponding cumulative distribution functions by $F_0(s)$ and $F_1(s)$ respectively. The prior probability of class j ϵ {'No', 'Yes'} is denoted by π_j. The expected cost is:

$$\begin{aligned}\bar{C} = c_{00}\pi_0 F_o(t) + c_{01}\pi_0(1 - F_o(t)) + \\ c_{10}\pi_1 F_1(t) + c_{11}\pi_1(1 - F_1(t))\end{aligned} \tag{1}$$

A necessary condition for minimizing $\bar{C}$ can be obtained by taking the derivative of $\bar{C}$ with respect to t and setting the result to zero. If we do this we get:

$$f_1(t)\pi_1 = f_0(t)\pi_0 \left\{ \frac{c_{01} - c_{00}}{c_{10} - c_{11}} \right\} \tag{2}$$

where t is some optimal threshold value. Note that pi_0 and pi_1, are constant for any value of $\bar{C}$. Similarly, the probability density functions of the scores for the 'Yes' and 'No' classes do not change as the same model is fitted to the data. Therefore, the threshold value t can be changed so as to satisfy the necessary condition for cost optimization by applying different cost weightings, that is, by varying η where

$$\eta = \frac{c_{01} - c_{00}}{c_{10} - c_{11}}$$

If, for example, the cost of a false negative is 10 times the cost of a false positive (with correct predictions having zero cost) then $\eta = 0.1$. Note that, since the cost of errors are likely to be greater than the cost of a correct prediction, then $\eta \geq 0$.

Hence the objective is to find the optimal threshold that will minimize the cost $\bar{C}$, for a given value of η. Note that, if no cost weightings are incorporated in the model formulation, then $\eta = 1$ and $t = 0.5$. However, empirical thresholding [16] can be used to select a cost-optimal threshold value t based on the training data, thereby minimizing $\bar{C}$.

Initially when our model is trained without incorporating costs, the resulting confusion matrix on the cross-validated training set was computed as shown in III.

This therefore gives an accuracy of 89.5%, with a success rate of 88.9%. However, there are many false positives (24) and false negatives (36) which would negatively impact the telecommunications provider financially.

To illustrate the approach, a logistic regression model was trained in a similar manner to the standard logistic regression model presented before but now incorporating costs with $\eta = 0.1$. A cost matrix with $c_{01} = 1$ and $c_{10} = 10$ was constructed to simulate this value of η. Using this model the predicted outcomes for successful upsell is shown in the confusion matrix given in Table IV. The expected cost to be

TABLE IV
CONFUSION MATRIX WHEN MINIMIZING COST

		Predicted	
		No	Yes
Actual	No	195	132
	Yes	3	240

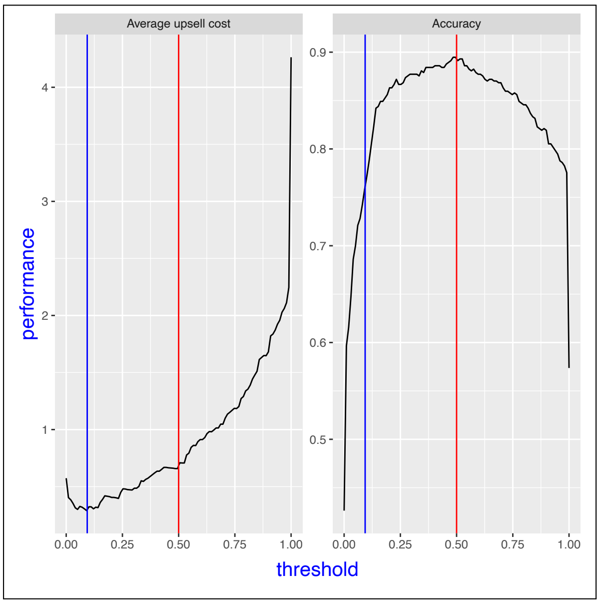

Fig. 3. Performance of Cross-Validated Training Data with Cost Optimal Threshold (blue) and Default Threshold (red) when η = 0.1

optimized was evaluated using the makeCostMeasure function in the mlr package in R.

For $\eta = 0.1$ the cost-optimal tuned threshold value was $t = 0.0935$ and the optimal expected cost was 0.2842. However, there was a reduced accuracy of 76.3% when compared to the accuracy if no costs are implemented (89.5%). In contrast, for $\eta = 1$ costs are ignored and the threshold t selected optimizes accuracy rather than cost where $t = 0.5$. The expected cost for the accuracy-based optimal result was found to be much higher with a value of 0.6737. The trade-off between accuracy and cost when $\eta = 0.1$ and when $\eta = 1$ is shown clearly in Figure 3.

Since the exact costs associated with misclassification errors are unknown, we evaluated costs using different values of η. We then compare the cost of the cost-based optimal result $\bar{C}_{\text{cost}}$ with the cost of the accuracy-based optimal result $\bar{C}_{\text{accuracy}}$. This comparison is achieved by calculating the ratio of the two costs R where the cost of the cost-based optimal result is the numerator. In Figure 4 we plot the value of this ratio R for various values of η.

Note that when $\eta = 0$ then false positive errors do not cost anything and therefore it is optimal to always choose a positive outcome in which case $R = 0$. If $\eta = 1$ then this corresponds to unit costs for both types of errors and hence the resulting

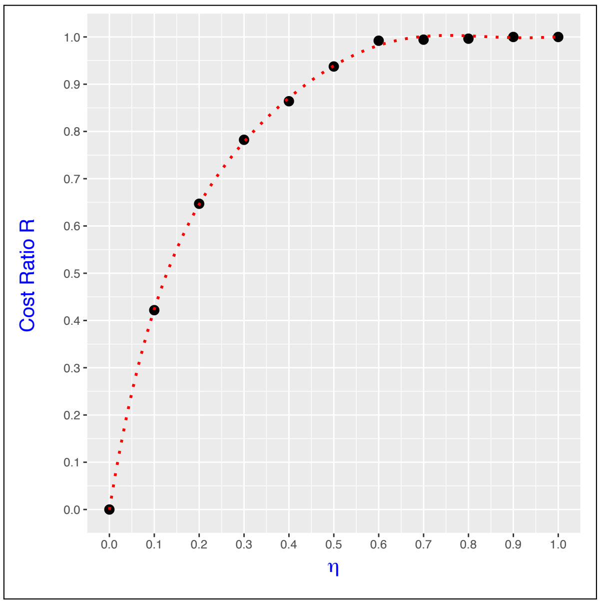

Fig. 4. Cost ratio as a function of η

optimal probabilities are the same that would be obtained for accuracy and hence in this case we get $R = 1$.

VI. PERFORMANCE RESULTS

Suppose we treat the test data with 80 observations as a set of customers that have not as yet been approached for upsell. If we use the model to select which customers to approach, the customers that would be selected will belong to the prediction = "Yes" class. A major advantage of using a predictive model to determine successful upsell is that we have an estimate beforehand of which customers will most likely refuse an upsell. We can therefore prune our selection and keep only those customers that are likely to have a successful upsell as indicated by the model (keep only the positive outcome classifications).

All customers who were predicted with an unsuccessful upsell will be ignored, and therefore l out of 80 customers will be chosen to approach for upsell. However, not all l customers that are approached will in fact have a successful upsell since some instances will be classified as false positives.

We can monitor the success rate when approaching customers. This is given by the ratio of the number of successful upsells to the total number of customers approached. This metric is really the precision or positive predictive value (PPV) since it is the number of true positives divided by the total number of predicted positives. The success rate can be written as:

$$S = \frac{\pi_1(1 - F_1(t))}{\pi_1 F_1(t) + \pi_1(1 - F_1(t))} \tag{3}$$

This simplifies to:

$$S = 1 - F_1(t) \tag{4}$$

Note that, as η is increased from zero, t is increased until it reaches $t = 0.5$ when $\eta = 1$. Hence, Equation 4 shows that with an increasing value of η, the success rate S increases since t increases and therefore $F_1(t)$, which is the cumulative probability of belonging in the positive class 'Yes', decreases. In other words, as η increases, there is relatively less penalty for false positive classifications when compared to false negative classifications and hence more positive outcomes (whether true or false) are selected by the model.

We report the results of the logistic regression models with several values of η in Table V. From this table, we can confirm that as η increases, the threshold t for optimizing cost also increases. Moreover, the accuracy of the model applied to both cross-validated training set and test set is also increased with increasing values of η. It appears that the model performs better on the test data since the accuracy values are slightly higher. The cost ratio R starts at 0 when $\eta = 0$ and approaches 1 as η approaches 1, and this is corroborated by Figure 4.

It is also interesting to note from Table V that the success rate S increases as η increases, as previously shown. Here, we see that S ranges from 65.4% to 86.1%. Note that, when the semi-manual method of approaching customers was employed, there was a total of 34 successful upsells out of a total of 80 customers approached, giving a 34/80 = 42.5% success rate. Hence, the utilization of a predictive logistic regression model to predict customer upsell for any given value of η dramatically improves the success rate that was achieved using the traditional semi-manual approach by at least 22.9%.

Finally, we use the model to predict the outcome of successful upsell on all the remaining 2847 customers not yet approached. From this, we select a subset of customers to approach for upsell by only keeping the customers with an expected positive outcome. The success rate should be similar to that generated using the approached dataset. Furthermore, we can rank the top customers who are most likely to upsell based on the generated raw probabilities used by the model to determine outcome classification. This is done by ranking instances in descending order from highest probability of belonging to the class 'Yes' to lowest probability of belonging to that class. We report the number of target customers selected for future approach for upsell in the last row of Table V. We observe that this number decreases as η increases, since the number of predicted outcomes is decreased as discussed before.

TABLE V
MODEL RESULTS FOR DIFFERENT η VALUES

Results	$\eta = 0$	$\eta = 0.1$	$\eta = 0.6$	$\eta = 1$
Threshold t for Optimal Cost	0.0475	0.0935	0.2230	0.5000
Cost-Optimal Accuracy of Cross-Validated Training Set	0.6912	0.7632	0.8719	0.8947
Cost-Optimal Accuracy of Test Set	0.7750	0.81250	0.8750	0.900
Accuracy-Optimal Upsell Cost	0.3750	0.4375	0.5625	1.0000
Cost-Optimal Upsell Cost	0.0000	0.1875	0.5500	1.0000
Cost Ratio R	0.0000	0.4286	0.9778	1.0000
Number of Customers selected for Approach from 80 in the Test Set	52	49	40	36
Number of Successful Upsells from 80 in the Test Set	34	34	32	31
Success Rate S	0.6538	0.6939	0.8000	0.8611
Number of Target Customers selected for Future Approach	397	293	89	71

VII. LIMITATIONS AND FUTURE WORK

As previously mentioned, the number of training examples was very small as data was provided for only one year. Moreover, there is a diverse number of products used by corporate customers in the telecommunications industry and there are low product take-up rates. Future work should be done to predict the probability of upsell within a defined future time period (say for example in the next quarter) using training instances over a longer historical time period. The predictive model for predicting upsell return could also solve the problem of maximizing long-term profitability.

VIII. CONCLUSIONS

Predicting the product purchase behaviour of corporate customers, specifically whether they would be likely to upsell, can contribute to increased sales and revenue generated by a telecommunications provider. We show that upsell return can be maximized by predicting whether a customer will upsell using a binary classifier that incorporates misclassification costs. Training and testing the model to predict upsell produced very good performance results and outperformed the current traditional semi-manual approach adopted by the telecommunications provider. With the use of our predictive mdoel, we were also able to create a ranked list of the top candidate customers to approach for upsell in the future.

References

[1] L. Waverman, M. Meschi, and M. Fuss, "The impact of telecoms on economic growth in developing countries," *The Vodafone policy paper series*, vol. 2, no. 03, pp. 10–24, 2005.

[2] M. M. Al-Debei and D. Avison, "Business model requirements and challenges in the mobile telecommunication sector," *Journal of Organisational Transformation & Social Change*, vol. 8, no. 2, pp. 215–235, 2011.

[3] V. J. Lall, "Application of analytics to help telecom companies increase revenue in saturated markets," *FIIB Business Review*, vol. 2, no. 3, pp. 3–10, 2013.

[4] J. Ni, Q. Shen, and T. Zhu, "Upselling versus upsetting customers? a model of intrinsic and extrinsic incentives," Working paper, Johns Hopkins University, Tech. Rep., 2015.

[5] H. Ahn, J. J. Ahn, K. J. Oh, and D. H. Kim, "Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5005–5012, 2011.

[6] D. Stanford, "Mining purchase history to predict adoption of mobile computing." *Intelligent Environments*, vol. 9, pp. 482–485, 2009.

[7] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang, and J. Lu, "A hybrid fuzzy-based personalized recommender system for telecom products/services," *Information Sciences*, vol. 235, pp. 117–129, 2013.

[8] Z. Lian-Ying, D. M. Amoh, L. K. Boateng, and A. A. Okine, "Combined appetency and upselling prediction scheme in telecommunication sector using support vector machines," *International Journal of Modern Education and Computer Science*, vol. 11, no. 6, p. 1, 2019.

[9] W. Zeng, W. Di, J. Wang, and Y. Cheng, "Machine learning for determining products to upsell or cross-sell," Nov. 30 2017, uS Patent App. 15/169,441.

[10] A. Martínez, C. Schmuck, S. Pereverzyev Jr, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *European Journal of Operational Research*, vol. 281, no. 3, pp. 588–596, 2020.

[11] S. Mitrović and G. Singh, "Predicting branch visits and credit card upselling using temporal banking data," *arXiv preprint arXiv:1607.06123*, 2016.

[12] B. Raskutti and A. Herschtal, "Predicting the product purchase patterns of corporate customers," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005, pp. 469–478.

[13] S. Sooklal and P. Hosein, "A benefit optimization approach to the evaluation of classification algorithms," in *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*. Springer, 2019, pp. 35–46.

[14] G. Cui, M. L. Wong, and X. Wan, "Cost-sensitive learning via priority sampling to improve the return on marketing and crm investment," *Journal of Management Information Systems*, vol. 29, no. 1, pp. 341–374, 2012.

[15] C. Elkan, "The foundations of cost-sensitive learning," in *International joint conference on artificial intelligence*, vol. 17, no. 1. Lawrence Erlbaum Associates Ltd, 2001, pp. 973–978.

[16] V. S. Sheng and C. X. Ling, "Thresholding for making classifiers cost-sensitive," in *AAAI*, 2006, pp. 476–481.

Navin Dookeram (M'22) was born in Trinidad and Tobago on 4th February, 1996. Navin attended the University of the West Indies, St. Augustine, Trinidad and Tobago. He holds a BSc. Petroleum Geoscience (2017) and an MSc. Statistics (2020) degree, and is currently pursuing a PhD in Statistics, all via the University of the West Indies. He has worked as an intern at Schlumberger, Trinidad (2016-2017), as an associate professional and part time demonstrator in the Department of Mathematics and Statistics, UWI (2017-2019), as a data science intern at the Telecommunications Services of Trinidad and Tobago (2019-2020).

Zahira Hosein is a Business Analyst within the Enterprise group of Telecommunications Services of Trinidad and Tobago Limited (TSTT) for the past 10 years. Through analytical dashboards and data models focused on business customers and products, she supports the executive, management and operational teams to make timely data backed decisions. Prior to joining the Enterprise group, she has 12 years IT experience at TSTT. She is experienced in several BI tools. She earned her MBA (Distinction) from Heriot-Watt University in 2021, and a Bachelor of Science Degree in Computer Science from The University of the West Indies in 1995.

Prof. Patrick Hosein attended the Massachusetts Institute of Technology (MIT) where he obtained five degrees including a PhD in Electrical Engineering and Computer Science. He has worked at Bose Corporation, Bell Laboratories, AT&T Laboratories, Ericsson and Huawei. He has published extensively with over 150 refereed journal and conference publications. He holds 41 granted patents in the areas of telecommunications and wireless technologies. He was nominated for the Ericsson Inventor of the Year award in 2004, was the Huawei US Wireless Research Employee of the year for 2007 and is a 2015 Anthony Sabga Caribbean Laureate for Science and Technology. Patrick is presently the administrative and technical contact for the TT top level domain, CEO of the TTNIC and a Professor of Computer Science at the University of the West Indies. His present areas of research include Applied Data Science, Operations Research and Performance and Pricing Optimization for Cellular Networks.