Towards Explainable Direct Marketing in the Telecom Industry Through Hybrid Machine Learning

Russel Petersen and Olawande Daramola(✉)

Department of Information Technology,
Cape Peninsula University of Technology, Cape Town, South Africa
203132173@mycput.ac.za, daramolaj@cput.ac.za

Abstract. Direct marketing enables businesses to identify customers that could be interested in product offerings based on historical customer transactions data. Several machine learning (ML) tools are currently being used for direct marketing. However, the disadvantage of ML algorithmic models is that even though results could be accurate, they lack relevant explanations. The lack of detailed explanations that justify recommendations has led to reduced trust in ML-based recommendations for decision making in some critical real-world domains. The telecommunication domain has continued to witness a decline of revenue in core areas such as voice and text messaging services which make direct marketing useful to increase profit. This paper presents the conceptual design of a machine learning process framework that will enable telecom subscribers that should be targeted for direct marketing of new products to be identified, and also provide explanations for the recommendations. To do this, a hybrid framework that employs supervised learning, case-based reasoning and rule-based reasoning is proposed. The operational workflow of the framework is demonstrated with an example, while the plan of implementation and evaluation are also discussed.

Keywords: Machine learning · Direct marketing · Explainable AI · Telecommunication · Case-based reasoning

1 Introduction

Majority of industries today face greater challenges in selling the right product to the right customer and at the right time. With the vast amounts of data available through big data technologies, a direct marketing campaign can be constructed to analyse customer characteristics to recommend the right product offering to a customer at the right time [1, 2].

Machine learning (ML) is a substantially better way of making predictions using complex sets of data from data sources to derive results. It utilises specially designed algorithms to identify customer patterns from large data sources, which are more difficult to handle for traditional statistical methods, expert systems-based approaches, or manual approaches [3].

O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12254, pp. 471–486, 2020.
https://doi.org/10.1007/978-3-030-58817-5_35

The disadvantage of ML algorithms is that they are typically black-boxes that produce results without explanations. Often, the user must be an expert to interpret the result to utilise it [4]. This lack of explanation reduces the level of confidence in ML predictions, and adoption by the management of many organizations, particularly in critical sectors. [4, 5]. This problem makes it necessary to have highly skilled personnel to interpret the results produced by ML algorithms. This increases capital and operational expenditure of organisations on personnel significantly [4, 6].

The generally poor explanation attribute of ML systems has led to increased interests in the field of Explainable AI (XAI) [6]. The information available on product review websites such as *Softwareadvice* (www.softwareadvice.com), *capterra* (www.capterra.com), and *goto crowd* (www.g2crowd.com) revealed that most ML-based direct-marketing tools lack the capability to provide a detailed explanation. ML predictions accompanied by detailed explanations would allow for valuable insight into understanding customer behaviour [7].

The telecommunication (telecom) industry has seen a continuous decline in revenue in core areas such as voice and text messaging [8]. At the same time, the domain has also absorbed a variety of disruptive technologies, message services and over-the-top streaming services. These factors have resulted in the high need to continuously increase product sales within the domain. With the rich customer data available, a data-driven direct marketing strategy enables the right product to be offered to the right customer and at the right time [8]. This makes the concept of direct-marketing with explanations quite relevant to the telecom industry.

This paper presents the design of a process framework that explores a hybrid approach that is able to combine supervised machine learning, and an intelligent reasoning model (IRM) that consist of case-based reasoning, and a set of expert defined knowledge rules to generate explanations for direct marketing predictions. This type of framework facilitates the use of experience from previous use case scenarios, and domain knowledge to create the basis for rich explanations that justifies ML predictions. As a contribution, an adaptable process framework that enables effective direct-marketing with explanations in the telecom industry is proposed, which will be an improvement on previous approaches for direct marketing in the telecom domain that lack explanations.

The remaining part of this paper is described as follows. Section 2 presents an overview of the background and related work. Section 3 presents the description of the process framework, while an example was used to describe how the framework can be applied is presented in Sect. 4. Section 5 discusses the plan for implementation and evaluation, while the paper is concluded in Sect. 6 with a brief note.

2 Background and Related Work

This section provides an overview of important topics that are relevant to this paper. It reviews the subject of data-driven direct product marketing, machine learning, rule-based reasoning, and case-based reasoning. It also presents an overview of previous work on ML-based direct marketing.

2.1 Data-Driven Direct Marketing

Direct marketing is a method whereby customer features such as spending, geographical details and past product purchases are used to offer a product directly to a specific customer. Across industries, this method has been proven to increase sales significantly [1, 9]. The technological advancement in data processing and storage has allowed organisations to store large datasets. A larger set of data allows an organisation to know more about a customer and execute a more accurate prediction. The right data allows customers to be segmented accurately and products can be marketed to identified customer segmentations [2].

Over time technological advancements have enabled the use of data-driven direct marketing. Various data analytics technologies can be used to build customer characteristics and identifying customer patterns from data. It is now possible to predict the probability that a customer would respond to a specific product offering [10, 11]. Studies have shown that top-performing companies rely more on large detailed datasets and data analytics than low performing companies [9, 10].

2.2 Machine Learning

Machine learning (ML) is a computer-based method which looks at existing data as input and produces a prediction as output by applying an algorithmic model. An algorithmic model is a sequence of instructions used to convert the input into the output [12]. At a high level, ML consists of two phases. The first phase is applying an algorithmic model to a dataset to train the model on a dataset. The second phase is to take a dataset and make a prediction of future occurrences.

ML has proven to be very effective in direct marketing. It allows a company to take a large dataset, and then establish detailed customer characteristics based on the dataset, and apply an algorithmic model to make a probabilistic prediction [3]. The most accurate ML models are usually in a nested non-linear structure. Nested non-linear models such as Artificial Networks (ANN), Random Forest (RF) are mostly applied in a black-box manner where no explanation is provided on how the model arrived at a particular prediction. This lack of explanation makes it difficult to identify flaws in ML models and biases in the data [4].

2.3 Case-Based Reasoning

Case-based reasoning (CBR) uses the record of past occurrences to provide the solution to a new occurrence. The approach has been found to have strong explanation mechanisms because it derives its explanations from similar previous cases [13, 14]. Typically the CBR problem-solving process entails case retrieval, case reuse, case adaptation, and case retention. CBR can be applied to address the lack of explanation and simplify the interpretation of results. CBR can be used to substantiate the predictions made, which can complement a machine learning model in a good way. CBR examines past occurrences with similar output and provides a detailed explanation of why the current output occurred. Real evidence is used from a set of relevant cases to

explain the task at hand. CBR Explanations are simplistic regardless of the complexity of the current problem [14].

2.4 Related Work

A number of efforts that focussed on the use of machine learning techniques to improve direct marketing have been reported in the literature. Of these efforts, the banking sector, and telecommunication appears significantly, hence they have been considered for review.

In [15], an attempt was made to determine the best classification technique for data-driven direct marketing. Four commonly used classification techniques were selected and evaluated. The results showed that decision trees produced a reasonably high sensitivity, specificity and accuracy. In [16] a similar attempt was made with respect to the banking sector where four different machine learning (ML) algorithms were compared to investigate the effectiveness in predicting potential customers for banking products. The results showed that the use of ML algorithms allows the processing of large amounts of data and gathering a greater historical view of customer spending patterns. The emphasis was on accuracy of results and no explanation. In [17] banking data was applied to the SVM and Random Forest (RF) Regression model to predict potential banking customers. Both models produced a good performance for classification but the RF Regression showed slightly improved results for accuracy and sensitivity. The study also noted that both models lacked explanation that can enable the understanding of results. In [18] the Naive Bayes and RF Regression model were compared by using a large banking dataset to determine customers for long term deposits. The RF Regression model performed better in terms of accuracy, specificity and sensitivity, but offered no explanation.

The authors in [19] showed that data classification techniques are effective in determining which customers will subscribe to term deposits in the banking industry. A banking dataset was labelled with relevant attributes, and customers grouped by region. Decision tree and Naïve Bayes classification techniques were used to categorise the customers. The decision tree was found to have a higher level of accuracy but offered no explanation.

In [20], it was found that the application of ML models improves overall marketing significantly. This was tested by applying a ML models to customer dataset containing various customer features to obtain predictions. In [21] a study to identify customers for upselling of products within the telecommunication industry using a ML was conducted. A Support Vector Machine (SVM) model was applied to a customer dataset to classify the customers as either a churner or a non-churner. The classification was then used to determine the probability to upsell or not. The prediction had high accuracy with minimal errors but no explanation was provided to justify the classification.

In [22] a deep learning model was applied to a telecommunication dataset to determine customer churn. The findings indicated that accurate predictions were made but the lack of explanation prohibited a greater understanding of customer behaviour to reduce the rate of churners in the long term proactively. A study by [23] recommended that extraction of data from the resource, service and customer layers within a telecommunication company could provide an accurate and insightful customer profile.

In addition to this, customer billing information which includes consumption, spend limit and usage data were found to be valuable in predicting product offerings. There is a relationship between a customer's past spending pattern and current consumption. This relationship can be exploited to recommend meaningful products to a customer at the right time. This paper extends the recommendation made by [23] by adding billing data as an extra dimension for generating prediction on customer choices, also focusses on providing explanations. A study by [24] revealed that combining a CBR system with a ML model can enhance the predictive accuracy as it enabled domain expert knowledge to be used to formulate cases in the case base. The hybridized system was applied to predict prices for internet domain names and it produced a better prediction. In [25] a CBR module was used to select the specific ML model that should be used to solve a problem. The CBR-ANN hybrid system showed improved results when compared to a situation when a basic ANN model is used. The aim of the CBR-ANN is not to improve explanation of results. A study to predict skin disease by using a real life skin disease dataset was reported in [26]. The CBR-ANN was adjudged to be of acceptable performance but the primary objective of the hybrid system was not explanation. In [27] a hybrid CBR-ANN system that can be used for different types of classification problems was presented. A trained ANN model was used to extract feature weights that were used to improve the performance of the CBR module within the hybrid system. use in case-based reasoning (CBR) system. The focus of the hybrid system is not explanation. In [28] a study that compared the performance and usability of a hybrid CBR-ANN system, a CBR system, and a ANN system for the prediction of the valuation of residential properties was presented. The hybrid CBR-ANN system was found to have better performance and usability. The emphasis of the CBR-ANN system was not primarily on explanation. Generally, although some hybrid systems have been proposed, and the concept of explainable machine learning is currently attracting increasing attention of researchers, so far, this has not been applied to the telecommunication domain, which makes the intended contribution of this paper to be unique [14, 29].

3 Methodology

In this section, the requirements and conceptual design of the proposed process framework for direct marketing with explanation are presented.

3.1 Requirements

In order to identify the requirements that the proposed approach must satisfy, the Joint Application Development (JAD) technique was used. JAD allows the inclusion of all project stakeholders, developers and users so that the requirements are scrutinized by all participants [30]. Requirements were derived from all functional and non-functional aspects that pertain to the system and all the identified use cases. Sessions were held over four days in scheduled daily sessions until all issues have been discussed and information collected [31]. The JAD session included 9 participants described as follows:

- Specialist in Predictive Analytics (2 persons)
- Customer Care End User (3 persons)
- Specialist in Customer Experience Management (2 persons)
- Customer Care Supervisor (1 person)
- Specialist in Business Intelligence (1 person)

During the JAD sessions, participants were given the opportunity to mention different types of interactions with customer data whether it is with the front-end resources or back-end resources. They were also asked to identify how a customer's purchasing pattern can be influenced in a positive way. They were also tasked to list any problem which they think might be a challenge in customer interaction or influence a customer's purchasing pattern. By using a brainstorming technique involving all participants, solutions were identified for all the problems, which were then converted to system requirements. At the end of the JAD sessions, the following requirements of the system were derived (Table 1):

Table 1. List of requirements

Req. No.	Description
1	The system shall predict customers that are most likely to be interested in a product and provide explanation of why the customers were selected
2	The system shall predict which products can be used to upsell specific customers
3	The system shall predict the customers that should not be considered for a particular product, with an explanation of why the customer were selected
4	The system shall produce its recommendations in a speedy and reliable manner
5	The system shall be able to classify customers into low and high spending customers based on the amount they spend per month
6	The system shall be able classify customers into low and high usage subscribers based on the amount they use per month
7	The system shall generate explanations that are easy to understand and of a high quality
8	The system shall enable use of specific computational operations in different scenarios through APIs or web services
9	The system shall ensure that customer data is protected through adequate security measures such as user authentication and user login access functions to disallow unauthorised users
10	The system shall display a list of available operations to enable the user to select an operation of choice or which prediction is required
11	The system shall display prediction results and explanation to the user
12	The system shall allow the user to load products to be sold to customers onto the system
13	The system shall allow a user accounts to be created and deleted

3.2 A Hybrid ML Approach for Explainable Direct Marketing

Based on the identified requirements, a hybrid ML approach that will enable direct marketing with explanation was conceived. The approach will enable data collection from 4 different layers of a telecommunication organisation, which are the resource, service, customer and billing layers, and to discover customers that should be targeted for direct marketing by using a hybrid ML approach (see Fig. 1). The resource layer includes subscriber and device data to verify the technologies available to the subscriber. The service layer includes the voice, data and video services utilized by the subscriber during a specific period of time. The customer layer includes the service provisioning data and any interaction the subscriber had with the organization; while the billing layer includes data on spending and user consumption data. After the data collection is complete, the data is passed to the hybridized machine learning system that does the prediction and generates an explanation.

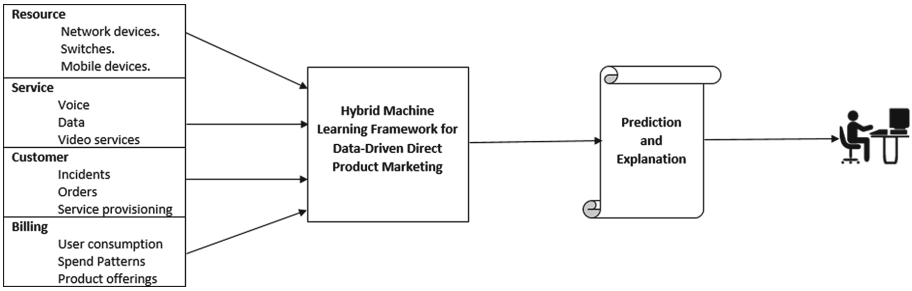

Fig. 1. An overview of the Hybrid ML approach for explainable direct marketing

3.3 A Process Framework for Explainable Direct Marketing (PROFEDIM)

We have formulated the process framework for explainable direct marketing (PROFEDIM), which consists of two phases. PROFEDIM is composed of a hybrid framework that integrates supervised machine learning (ML) and case-based reasoning (CBR). The activities of PROFEDIM, which can be classified into two phases are depicted in their sequential order of (1) – (15) in Fig. 2. The first phase consists of activities of data gathering – (1), data wrangling – (2), data labelling – (3), and data selection and model training – (4) which are all mostly semi-automated, and offline procedures. Data selection entails identifying the specific attributes that are required for the direct product marketing task at hand, while model training is the process of using supervised learning – (5) to train a nested non-linear ML model such as ANN, SVM, or RF on the selected dataset. Model training is an offline activity that is separated from the prediction task.

The second phase entails the generation of a prediction and an explanation in response to a query. For prediction, when data of new query case/instance – (6) are passed into the PROFEDIM, feature selection – (7) takes place so that relevant data are extracted from the new query case. To do this, case matching and case retrieval from the case base – (8) takes place to determine if the new case is similar to some old cases if the new query case is significantly similar to some old cases, then the CBR module– (9) by using the predefined rules by domain experts – (10) is used to generate a prediction and a relevant explanation – (11).

When significant similarity with cases in the case base cannot be established, the ML-based prediction module – (12) performs the prediction. This offers three possibilities, which are to determine if customers could be interested in a new product, not interested in a new product, and to determine if a customer can be a candidate for upselling. After the prediction by the supervised ML model – (13) the case-based reasoning module – (15) will search the case base – (14) for cases that have somewhat similar attributes to the current case, and use them to construct a relevant explanation for the new prediction – (11) by relying on the predefined rules by domain experts that are stored in the repository – (11). The two alternate paths of computation that can be explored in the course of operation of PROFEDIM is shown in Fig. 3.

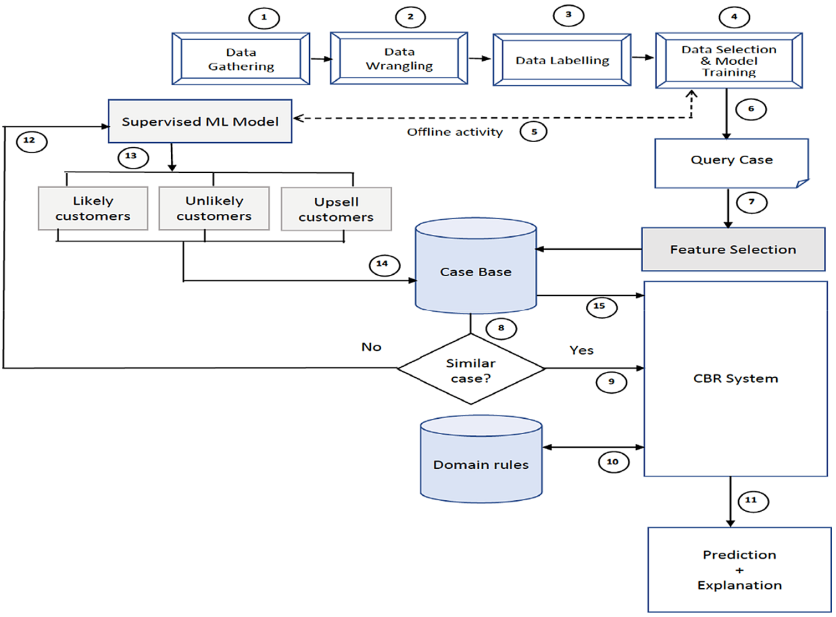

Fig. 2. Overview of the process workflow of the Hybrid ML framework

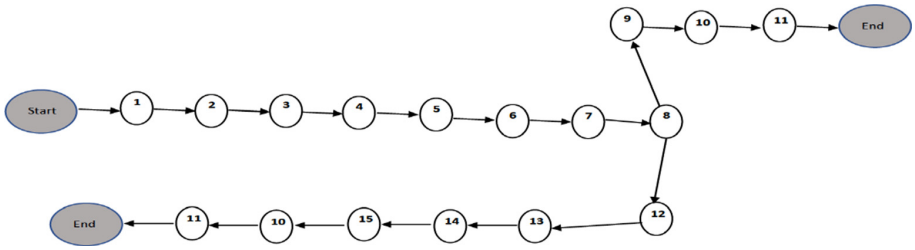

Fig. 3. Alternative computational paths of Hybrid ML framework

4 An Illustrative Example

In this section, we present a scenario example that deals with identifying subscribers that should be targeted for direct marketing for a new YouTube data product by a Telecom company (herein referred to as XC). The task is to select the subset of customers of XC that should be targeted for direct marketing from a large pool of subscribers. We shall use demo subscriber data that is derived based on data attributes that emulate the schema of an unnamed Telecom company.

i. Data Gathering Process

The data gathering activity will involve extracting data from various sources at the different layers of XC organization, which are pooled into a central database. The database will include the following:

- Subscriber Profile (A1) – the subscriber's main offering and any data package.
- Device Profile (A2) – the details of the device which was used by the subscriber this will include the device capability, streaming functionality and the technology supported by the device (2G, 3G, or 4G).
- Data Add-on Service (A3) – any other additional services such as a video streaming, WhatsApp or YouTube service that the customer uses.
- Billing Profile (A4) – the subscriber's billing data from the billing management system. This will include the total amount spent for the month, the monetary products attached to the profile and the billing cycle.
- Data Allocation (A5) – the amount of data allocated to the subscriber through purchase during a particular time period.
- Data Usage (A6) – the amount of data used by the subscriber during a particular time period.
- Transfer Activity (A7) – the amount of data received by the subscriber and the amount of data sent to other subscribers during a particular time period.
- Streaming Application Usage (A8) –This is the amount of data used on the 3 most used applications by the subscriber on these applications during a particular time period.
- Total Spend (A9) – the total amount spent by the subscribers during a particular time period based on the billing system (Table 2).

Table 2. Sample data on subscribers

Subscriber ID	A1	A2	A3	A4	A5	A6	A7	A8	A9
CUST9581111	Select + Package	iPhone 11	1 GB	2000 MB	19500	0 MB	500 MB	WhatsApp (300); Youtube (100); Facebook (450)	850.00
CUST9581001	Smart XL	Samsung S9	250 MB	250 MB	500	0 MB	0 MB	Youtube (100);	600.00
CUST9781011	Smart L	Samsung A1	1 GB	0 MB	1200	0 MB	0 MB	WhatsApp (300)	300.00
CUST3561221	Prepaid	Vodacom T1	0 MB	0 MB	200	0 MB	0 MB	WhatsApp (300)	110.00
CUST6481215	Smart Red	iPhone 10	2 GB	1000 MB	10500	500 MB	500 MB	Facebook (200)	1100.00
CUST4591101	Red VIP	iPhone 11	5 GB	1000 MB	2000	0 MB	0 MB	Youtube (100)	950.00

ii. Data Wrangling Process

Data wrangling process will entail cleaning the data and removing duplicate, and null values, and missing data that affects good data quality.

iii. Data Labelling Process

The data labelling process will allow the gathered data to be categorized and used as a basis to derive relevant data attributes that could be used as a basis for prediction (see Table 3). For this example case, the data attributes that were derived are the following:

- Spending Category (L1) – This is categorised as High, Medium, and Low based on A7, and A9 according to a predefined threshold (e.g. High $\geq$ R150; Medium R50 $\leq$ R149); or Low < R50)
- Data usage category (L2) – This will be categorised as High, Medium and Low based on A5, A6, and A7 according to a predefined threshold (e.g. High $\geq$ 300 MB; Medium $\geq$ 100 $\leq$ 299 MB; or Low < 100 MB).
- Period of connection (L3) – this is the measure of the period of years when the subscriber is connected to the network. This number of days will be divided by 365 days to convert it to years. This is derived based on A3.
- Spend rate per day (L4) – this is the ratio of the total monthly spend divided by the number of days in the month(s). This is derived based on A9
- Data usage rate per day (L5) – this is the ratio of the total amount of data used in the month divided by the number of days. This is derived based on A6
- Data add-on usage rate per day (L6) – the total monthly data add-on utilized divided the number of days in the month. This is derived based on A3.
- Device capability (L7) – This is categorized into 2G (Low), 3G (Medium) or 4G (High). This is derived based on A2.
- Streaming capability (L8) – This classified as streaming (1) or non-streaming (0) to indicate if a subscriber's device has the streaming capability or not.
- Streaming usage per day (L9) – This classified as Low, Medium, or High based on A8, and according to a predefined threshold.

Table 3. Subscriber profile after the data labelling process

Subscriber ID	L1	L2	L3	L4	L5	L6	L7	L8	L9
CUST9581111	Medium	Medium	28,33	66,66	33,33	10,33	High	1	High
CUST9581001	High	Medium	1,36	20,00	1,36	1,36	High	1	Medium
CUST9781011	High	Medium	3,28	10,00	33,33	0	High	1	Low

iv. Data Selection and Model Training

This activity involves selecting labelled data attributes that are relevant to the prediction task at hand from Table 3. For the current task L1, L2, L6, L7, L8, L9 are the most important to predict whether a customer would be interested in a new YouTube product or not. L4 and L5 are less important because their effects are subsumed by L1 and L2, which captures the total amount spent, and the total data usage by subscribers hence they were excluded. The selected attributes lead to a regression problem that seeks to predict the likelihood (True/False) that a customer would be interested in a new YouTube product as follows:

$$Y = a + L_1x_1 + L_2x_2 + L_6x_6 + L_7x_7 + L_8x_8 + L_9x_9 + e \tag{1}$$

Where Y is the dependent variable to be predicted, a is the intercept, L is the slope, X is the observed score on the independent variable, and e is an error value. This regression task can be solved by using a supervised learning model such as ANN, Support Vector Regression (SVR) or Random forest (RF). A dataset K consisting of normalized vectors can be used to train the RF regression model.

$$K = \left\{L_1^*, L_2^*, L_6^*, L_7^*, L_8^*, L_9^*, Y^*\right\} \tag{2}$$

Where L_i^* *and* Y^* *are the normalised values of* L_i *and* $Y \in K$

In a scenario where RF is selected as the ML model to use, then a RF model can estimate the value of Y for every vectorized data of L_i that is presented to it. The most discriminant attributes that are found to be critical for the prediction by the RF model are also stored in the case base. For example after training and testing of the RF, the attributes L_1 – Spending category, L_2 – Data usage category, and L_9 – streaming usage per day could have been found to be the most significant attributes for the prediction made. These attributes are then stored in the case base for subsequent reuse in order to generate explanation in future prediction scenarios.

v. Random Forest (RF) Regression

Random Forest (RF) Regression performs prediction by using an ensemble learning approaches that returns the mean of the predictions of several decision trees to produce its prediction. The decision trees are made up of questions which branches the learning sample into smaller and smaller parts. The regression algorithm will search all possible variables and values to find the best available branch which would be returned as the prediction [32]. Studies have shown that RF Regression is an ideal technique in

handling both categorical and numerical data because it produces a high level of accuracy of predictions and uses minimal overheads to perform predictions. The procedure the RF prediction will typically entail the following:

i) Import appropriate RF libraries, and load the dataset for training
ii) Split the dataset into a training set and test set (the 80/20 or 70/30 or 65/35 rules are options to consider depending on the nature of data)
iii) Create a RF regression model and fit it to the dataset.
iv) Experimentally adjust hyperparameters to obtain a good-fit of the training and validation data set.
v) Visualize the results.

vi. Prediction and Explanation via the CBR Module
The CBR module will work by comparing features extracted from a new case with existing cases in the case base. The retrieval of similar cases to a query (new) case will be done by using a K-Nearest Neighbour algorithm that fetches the set of most similar cases to the new case [13]. For our example, a similarity threshold in the upper percentile of 0.8 and above is considered significant. This is to ensure that case adaptation by the CBR module is based on the solution parts of existing cases that are significantly similar to a new case. The case adaptation process is used to generate a prediction for a current case. The predefined rules by domain experts are used together with the case attributes of the most similar cases to generate an explanation to justify the prediction. On the contrary, if similarity of less than 0.8 is established, the RF model will be used to generate the prediction because of its superior predictive ability compared to CBR when there are no significantly similar cases in the case base.

vii. Prediction and Explanation via the RF Model
When the set similarity threshold between a new case and existing cases is not met (< 0.8) then the pre-trained RF regression model is used to generate a prediction. After obtaining a prediction, the CBR module will use the knowledge contained in the case base to construct an explanation for the new prediction. To do this, the identified most discriminant attributes for the RF prediction are used as a basis to identify the existing cases that are most similar to the current case and retrieve them from the case base. The basis for this retrieval will normally be based on a lower similarity threshold so as to retrieve sufficient multiple cases. By using a heuristic search operation, individual discriminant attributes of these multiple cases are examined to find instances that are similar to those of the query case. These are then used together with the predefined domain rules to generate an explanation for the prediction obtained from the RF model for the query case. The examples of predefined rules to aid the generation of meaningful explanations to support the predictions made by the RF model are shown in Table 4.

Table 4. Sample domain rules that can influence explanation generation.

Rule No	Condition	Offer
1	If low spender and high data receiver and streaming capability	YouTube product
2	If medium spender and high data consumption and WhatsApp user	WhatsApp product
3	If low spender and long-time user	Data product
4	If high spender and high data consumption and streaming capability	Data product
5	If high spender and high data consumption and high streaming	Video streaming product
6	If high spender and medium data consumption and high streaming	Offer YouTube product
7	If low consumer and low spender and short-time user	No offering
8	If high spender and high data sender	Data product
9	If high spender and high streaming	Video streaming product
10	If low spender and high data consumption and streaming capability	Video streaming product

5 Implementation Plan

The recent advancement in the fields of machine learning (ML), and the existence of reliable software development frameworks for case-based reasoning (CBR) makes the actual implementation of process framework that can facilitate explainable direct marketing plausible. We intend to develop an integrated system that will be able to support the full scope of activities that spans the 2 phases of the process framework for explainable direct marketing (PROFEDIM) as outlined in Sect. 3. We plan to leverage Scikit-learn a Python-based ML framework to realise the supervised learning capabilities for prediction such as Artificial Neural Networks (ANN), Random Forest (RF), and Support Vector Machines (SVM). The myCBR Restful API [33] will be used as the building block to realise all case-based reasoning functionalities. The two development frameworks will be integrated within a single hybrid system architecture that can support end-to-end activities of PROFEDIM. A prototype software will be developed with distinct interfaces that support all forms of user activity such as selecting a specific direct marketing task of interest, pre-processing, the specific type of computation by using appropriate middleware algorithms, obtaining predictions, and obtaining explanations by relying on appropriate knowledge resources. The outlook of the components of the integrated hybrid ML system architecture is shown in Fig. 4.

From Fig. 4, it is obvious that data selection and feature selection can be invoked from both the Scikit-learn component and the myCBR component, while prediction and explanation capabilities are also enabled by these middleware components.

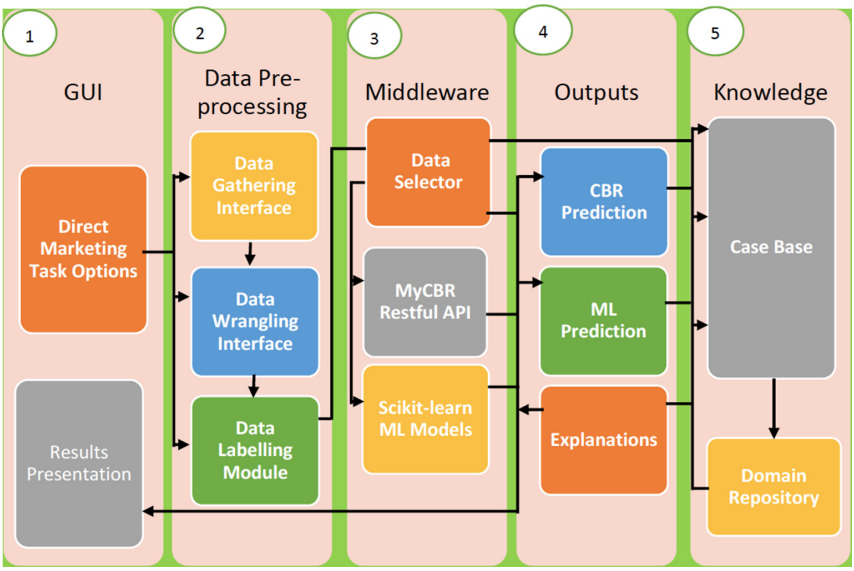

Fig. 4. Components of the Hybrid ML system architecture

Post-implementation, the evaluation of PROFEDIM will be done from two perspectives. First will be to assess the performance of the framework in terms of the accuracy of its predictions and the quality of its explanations. The process framework allows either of ANN, RF, and SVM to be selected as the ML model for generating a prediction, hence standard regression metrics will be used to do this. The quality of explanations that are generated to justify predictions will also be evaluated in terms of its understandability. The second perspective will be to assess the usability of PROFEDIM as it fits into the operational workflow of a Telecom company. These two evaluation perspectives will be essential in order to derive a valid conclusion of the plausibility of the proposed framework.

6 Conclusion

In this paper, the description of a process framework for explainable direct marketing (PROFEDIM) is presented. PROFEDIM is based on a hybrid architecture that integrated both supervised machine learning and CBR. The sequence of activities of the process framework, and its capabilities, and plan of implementation and evaluation was discussed. This is further demonstrated by using an illustrative example of how the process framework can be applied in a real problem scenario. The contribution of this paper is that it offers a new perspective on direct marketing in the telecom domain through the use of a hybrid AI architecture. This is because existing direct marketing tools mostly lack explanation capability that is able to justify their predictions/recommendations. In further work, we shall take steps to implement the proposed framework, and also conduct an evaluation.

References

1. Alanen, A.: Efficient direct marketing: Case: Valtapinnoite Oy (2016)
2. Yu, C., Zhang, Z., Lin, C., Wu, Y.J.: Can data-driven precision marketing promote user ad clicks? Evidence from advertising in WeChat moments. Ind. Mark. Manage. (2019). https://doi.org/10.1016/j.indmarman.2019.05.001
3. Erel, I., Stern, L.H., Tan, C., Weisbach, M.S.: Selecting directors using machine learning (No. w24435). National Bureau of Economic Research (2018)
4. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM Comput. Surv. (CSUR) **51**(5), 1–42 (2018)
5. Gordon, J., Perrey, J., Spillecke, D.: Big data, analytics and the future of marketing and sales. Digital Advantage, McKinsey (2013)
6. Bonacina, M.: Automated reasoning for explainable artificial intelligence. In The First International ARCADE (Automated Reasoning: Challenges, Applications, Directions, Exemplary Achievements) Workshop (in association with CADE-26), Gothenburg, Sweden (2017)
7. Goebel, R., et al.: Explainable AI: the new 42? In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2018. LNCS, vol. 11015, pp. 295–303. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99740-7_21
8. Vantara, H., Kalakota, R., Partner, LiquidHub: Transform Telecom: A Data-Driven Strategy for Digital Transformation, White Paper, Hitachi Vantara (2019)
9. Beheshtian-Ardakani, A., Fathian, M., Gholamian, M.: A novel model for product bundling and direct marketing in e-commerce based on market segmentation. Decis. Sci. Lett. **7**(1), 39–54 (2018)
10. Flici, A.: A conceptual framework for the direct marketing process using business intelligence (Doctoral dissertation, Brunel University Brunel Business School Ph.D. theses) (2011)
11. Buttle, F., Maklan, S.: Customer Relationship Management: Concepts and Technologies. Routledge, Abingdon (2019)
12. Pereira, F.C., Borysov, S.S.: Machine learning fundamentals. In: Mobility Patterns, Big Data and Transport Analytics, pp. 9–29. Elsevier (2019)
13. Daramola, O., Stålhane, T., Omoronyia, I., Sindre, G.: Using ontologies and machine learning for hazard identification and safety analysis. In: Maalej, W., Thurimella, A. (eds.) Managing requirements knowledge, pp. 117–141. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-34419-0_6
14. Vásquez-Morales, G., Martínez-Monterrubio, S., Moreno-Ger, P., Recio-García, J.: Explainable prediction of chronic renal disease in the colombian population using neural networks and case-based reasoning. IEEE Access **7**, 152900–152910 (2019)
15. Wisaeng, K.: A comparison of different classification techniques for bank direct marketing. Int. J. Soft Comput. Eng. (IJSCE) **3**(4), 116–119 (2013)
16. Nachev, A.: Application of data mining techniques for direct marketing. In: Computational Models for Business and Engineering Domains, pp. 86–95 (2015)
17. Lawi, A., Velayaty, A.A., Zainuddin, Z.: On identifying potential direct marketing consumers using adaptive boosted support vector machine. In: 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), pp. 1–4. IEEE (2017)
18. Ruangthong, P., Jaiyen, S.: Bank direct marketing analysis of asymmetric information based on machine learning. In: 2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 93–96. IEEE (2015)

19. Karim, M., Rahman, R.: Decision tree and Naïve Bayes algorithm for classification and generation of actionable knowledge for direct marketing. J. Softw. Eng. Appl. **6**, 196–206 (2013)
20. Bayoude, K., Ouassit, Y., Ardchir, S., Azouazi, M.: How machine learning potentials are transforming the practice of digital marketing: state of the art. Period. Eng. Nat. Sci. **6**(2), 373–379 (2018)
21. Lian-Ying, Z., Amoh, D.M., Boateng, L.K., Okine, A.A.: Combined appetency and upselling prediction scheme in telecommunication sector using support vector machines. Int. J. Mod. Educ. Comput. Sci. **11**(6), 1 (2019)
22. Castanedo, F., Valverde, G., Zaratiegui, J., Vazquez, A.: Using deep learning to predict customer churn in a mobile telecommunication network (2014)
23. Chen, C.: Use cases and challenges in telecom big data analytics. APSIPA Trans. Sig. Inf. Process. **5**, e19 (2016)
24. Dieterle, S., Bergmann, R.: A hybrid CBR-ANN approach to the appraisal of internet domain names. In: Lamontagne, L., Plaza, E. (eds.) ICCBR 2014. LNCS (LNAI), vol. 8765, pp. 95–109. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11209-1_8
25. Hegdal, S.S., Kofod-Petersen, A.: A CBR-ANN hybrid for dynamic environments. In: CEUR Workshop Proceedings (2019)
26. Dabowsa, N.I.A., Amaitik, N.M., Maatuk, A.M., Aljawarneh, S.A.: A hybrid intelligent system for skin disease diagnosis. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. IEEE (2017)
27. Biswas, S.K., Sinha, N., Purakayastha, B., Marbaniang, L.: Hybrid expert system using case based reasoning and neural network for classification. Biol. Inspired Cognit. Archit. **9**, 57–70 (2014)
28. Musa, A.G., Daramola, O., Owoloko, E.A., Olugbara, O.O.: A neural-CBR system for real property valuation. J. Emerg. Trends Comput. Inf. Sci. **4**(8), 611–622 (2013)
29. Du, M., Liu, N., Hu, X.: Techniques for interpretable machine learning. Commun. ACM **63** (1), 68–77 (2019)
30. Hyseni, L., Dika, Z.: An integrated framework of conceptual modelling for performance improvement of the information systems. In: 2017 Seventh International Conference on Innovative Computing Technology (INTECH), pp. 174–180. IEEE (2017)
31. Couronné, R., Probst, P., Boulesteix, A.: Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinform. **19**(1), 270 (2018). https://doi.org/10.1186/s12859-018-2264-5
32. Ouedraogo, I., Defourny, P., Vanclooster, M.: Application of random forest regression and comparison of its performance to multiple linear regression in modelling groundwater nitrate concentration at the African continent scale. Hydrogeol. J. **27**(3), 1081–1098 (2019). https://doi.org/10.1007/s10040-018-1900-5
33. Bach, K., Mathisen, B.M., Jaiswal, A. Demonstrating the myCBR Rest API. https://iccbr2019.com/wp-content/uploads/2019/09/01_paper_Demonstrating_the_myCBR_REST_API.pdf