Recommender System for Telecommunication Industries: A Case of Zambia Telecoms

Mulizwa Soft[1,*], David Makadani Zulu[1], Ruzive Mazhandu[2]

[1]Department of Computer Science, University of Zambia, Lusaka, Zambia
[2]Department of Mathematics, University of Leeds, Leeds, United Kingdom

Abstract Recommender systems use machine learning algorithms and artificial intelligence techniques to recommend products to customers. These algorithms use historical data of purchases of other people to determine which products to recommend to a particular customer, in general recommender systems are designed in such a way that they automatically generate personalized suggestions of products to customers. With the competitiveness that is growing in the Zambian telecom industry as a result of the new fourth mobile telecom service provider, telecommunication operators are looking for ways of attracting and keeping their subscribers on their network by giving affordable products to their subscribers. This research paper describes a Recommender System for Telecommunication companies using call detail reports (CDR's), machine learning algorithms and big data concepts.

Keywords Recommender systems, Telecom products, Machine learning algorithms, Churn, Business Intelligence

1. Introduction

With the increasing number of mobile telecommunications operators in Zambia, a customer is entitled with unlimited freedom to switch from one mobile operator to another, if he is not satisfied with the services or pricing their providers are providing. This trend is not good for operators as they lose their revenues because of customers switching from one provider to another in search of cheap affordable high-quality products and services [1]. The rate at which you are losing subscribers or customer is called churn rate.

2. Statement of the Problem

Telecommunication operators in Zambia lack the ability to manage their customer retention rate because they do not have a personalized way of recommending products and services to their subscribers and most of them do not have the tactics to help reduce the churn like calling their customers and measuring customer happiness as a result subscribers tend to migrate to new providers. This trend of subscribers migrating to new providers proves to be a severe problem for Telecom providers as they experience subscriber base and revenue shrinkage, the increase in churn rate causes a loss of future incomes [2].

3. Aim of the Study

Implement a recommender system for a telecommunication company.

3.1. Research Questions

1. How do we analyze the relationship between telecommunication subscribers and telecommunication products?
2. What are the challenges telecommunication companies face that lead to high revenue loss, churn and bad customer experience?
3. How do we process approximately five billion files/day?
4. How best will Product recommender systems for mobile technology be utilized in order to assist solving the problem of low revenue, churn and fraud.

4. Significance of the Study

Recommender based systems implemented using the concept of big data, machine learning algorithms or deep learning algorithms have a lot of advantages which benefits a lot of companies, for example better user experience, increased average order value, Increased Sales and improved customer retention.

4.1. Related Works

Recommender systems are the most successful implementation of web personalization and can be defined

as personalized information filtering technology that is used to automatically predict and identify a set of interesting items on behalf of users according to their personal preferences. Recommender systems use the concept of rating to measure user's preferences and a range of filtering techniques, and can be classified in multiple ways according to the nature of the input information. The content-based (CB) methods and collaborative filtering (CF) methods are the most popular techniques adopted in recommender systems. The CB methods recommend products by comparing the content or profile of the unknown products to those products that are preferred by the target user. However, these methods tend to rely heavily on textual descriptions of items, leading to several unsolved problems such as limited information retrieval, new user problems, and overspecialization [3].

5. Methodology

Baseline study methodology and software design methodology.

5.1. Baseline Study Methodology

The data dump or raw data will be collected from the CDR's for the telecommunication company, Structured Query Language (i.e. SQL) procedures will be used to collect the SQL dump or raw CDR's will be processed using big data techniques, this data will be split into **training set**, **test set** and **validation set**.

5.2. Software Design Methodology

1. Examine current recommender systems used by Amazon, eBay, Facebook, IMDb and Netflix or Coursera.
2. Propose a recommender system for telecommunication companies.
3. Machine learning models or Algorithms.

5.2.1. How do 1 know which Model to Choose for my Problem?

You first need to figure out whether your problem is linear or non-linear, if your problem is linear, you should go for logistic regression (i.e. LR) or support vector machine (i.e. SVM) and If your problem is non-linear, you should go for k-nearest neighbor (i.e. k-NN), naive Bayes, decision tree or random forest but from a business point of view, you would rather use, logistic regression or naive Bayes when you want to rank your predictions by their probability. For example, if you want to rank your customers from the highest probability that they buy a certain product, to the lowest probability. Eventually that allows you to target your marketing campaigns. And of course, for this type of business problem, you should use Logistic Regression if your problem is linear, and Naive Bayes if your problem is non-linear.SVM when you want to predict to which segment your customers belong

to. Segments can be any kind of segments, for example some market segments you identified earlier with clustering.

5.2.2. Model Selection & Boosting

After we built our Machine Learning models, some questions remained unanswered:

1. How to deal with the bias variance trade-off when building a model and evaluating its performance?
2. How to choose the optimal values for the hyper parameters (the parameters that are not learned)?
3. How to find the most appropriate Machine Learning model for my business problem?

The above questions we will be answered using model selection techniques thus k-Fold Cross Validation and Grid Search.

5.3. Confusion Matrix

A confusion matrix illustrates the accuracy of the solution to a classification problem. Given n classes a confusion matrix is a m x n matrix, where $C_{i,j}$ indicates the number of tuples from D that were assign to class $C_{i,j}$ but where the correct class is C_i. Obviously, the best solution will have only zero values outside the diagonal a confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. Some standards and terms:

1. True positive (TP): If the outcome from a prediction is p and the actual value is also p, then it is called a true positive.
2. False positive (FP): However, if the actual value is n then it is said to be a false positive.
3. Precision and recall: Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity. Recall is nothing but the true positive rate for the class [4].
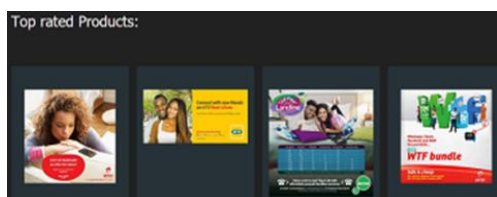
6. Discussion

The strategic use of big data and advanced analytics will help service providers in Zambia sharpen their campaign management and pre-emptive churn avoidance mechanism effectively. One of the other problems is that as of today the customer care experience for consumers is disjointed, reactive and not personalized. Service providers interested in managing the customer experience more effectively need to collect as much customer experience-related data as possible. They must be able to analyze this information to understand the customer experience and especially to detect unexpected and unpredictable events that might pose either an
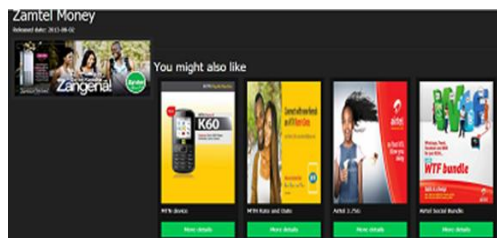
opportunity for example, to enhance or enrich the customer experience [5]. Sub-Saharan Africa is a predominantly a prepaid market for most of the telecommunications subscribers, it has been noticed that operator subscription numbers are constantly fluctuating due to the non-commitment of subscribers and the competition exists not only between operators, but also as a result of over-the-top players, i.e. social media providers such as WhatsApp. The low cost of using these over-the-top services means subscribers are less likely to use operator services such as SMS and voice [6]. On the other hand, an increase in churn rate causes a loss of future incomes. Therefore, it is profitable for telecom operators to invest in those customers that already have an experience with the service by renewing their trust rather than trying to attract new customers [7].

7. Results

The recommender system is developed for the telecom industry. It is implemented using a python programming and Machine learning algorithms. It consists of three main parts: client, web server and database server, the first time a subscriber visits the website our system will recommend the top-rated products for various telecommunication companies and when a subscriber logs in it will be able to show the recommended products based on either geographical location, gender, usage or popularity as shown in figure 1 and 2 respectively.

Figure 1. Shows the most top-rated products

Figure 2. Recommendation based on location

8. Conclusions

This research proposes a recommender system for all the telecommunication companies. With an introduction of a Recommender system, telecommunication companies can use recommender system to suggest products to their customers. The products can be recommended based on the top overall sellers on a site, based on the demographics of the customer, or based on an analysis of the past buying behavior of the customer as a prediction for future buying behavior. Broadly, these techniques are part of personalization on a site, because they help the site adapt itself to each customer.

ACKNOWLEDGEMENTS

REFERENCES

[1] Omae Malack Oteri, Mobile Subscription, Penetration and Coverage Trends in Kenya's Telecommunication Sector.

[2] Hoda A. Abdel hafez Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity.

[3] A hybrid fuzzy-based Personalized Recommender System for Telecom Products/Services.

[4] Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification Tina R. Patil, Mrs. S. S. Sherekar Sant Gadgebaba Amravati University.

[5] Big Data & Advanced Analytics in Telecom: A Multi-Billion-Dollar Revenue Opportunity.

[6] Sub-Saharan Africa june 2014 ericsson mobility report.

[7] Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity.