JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Identify prospects customers using lead score strategies and provide dynamic solution.

[1] Sandeep Shelar, [2]Prof. Dr.Uma Gurav

Department of Computer Science and Engineering
KIT's College of Engineering (Autonomous), Kolhapur, India.

Abstract: Lead scoring models are widely used by financial organizations to identify prospect customers. These models are developed according to business needs and strategies such as logistic regression, SVM or naïve Bayesian are applied according to data and leads to be focused. Here we are proposing logistic regression model over other models for lead score evaluation when we know exact dependent variables for profile scoring which will be useful to increase customer conversion ratio.

Users are having financial plans w.r.t certain goals like Car, House, Children Education, retirement.

Users are visiting to sites, but they are not completing registration or investing any amount w.r.t goals. We will collect ratio of prospect customers to actual customer conversion.

We will identify such data of users and based on type of goals or amount of invest

Mode of invest will suggest benefits plans like direct money points deposited to user wallet, reward system, concession on processing fees, Concession on product price. Based on proposed solution will identify increase in percentage of prospect customers ratio.

Also, we can feed online and offline data to for optimal performance.

IndexTerms – ***Lead score, Prospect customer.***

I. Introduction

Risk based Authentication also known as adaptive authentication solutions assign a risk score based on user login behaviour. Additional rules also can be applied to assign risk score. Machine learning algorithms can be used to learn user behaviour to build a user profile for login patterns.

Risk based authentication asks users less authentication information whose behaviour is in certain expected way (same login device/IP for most of the logins, same geo locations for most of the logins). This will result in lesser friction for user authentication without compromising on security. If user login behaviour is not similar to usual, user can be requested with more authentication information in addition to general login information (MFA, OTP etc). Machine learning classification algorithms can be used to classify whether a user attempting to login is genuine or fraud based on risk analysis.

K nearest neighbor algorithm for classification

The K-nearest neighbors (KNN) algorithm is supervised machine learning algorithms. KNN is easy to implement in its basic form, also performs complex classification tasks. It is called lazy learning algorithm which means it is not having a specialized training phase. It uses the entire data for training while classifying a new data point or instance. KNN is a non-parametric learning algorithm, it doesn't assume anything about the underlying data. This is useful feature because the realworld data doesn't really follow any theoretical assumption e.g., linear-separability, uniform distribution, etc. For this type of problem, KNN will be useful. With the help of K-NN, we can easily identify the category high risk /low risk authentication of a given login attempts dataset.

II. Literature review

Literature has been conducted to find key knowledge of leads and to gain a better understanding in the existing findings and challenges in new implementation.

First level of leads before they are scored is we must create the leads based on attributes. There are number of ways leads can be created online or offline.

Giving away gifts, hot and cold leads and marketing are possible offline methods to generate leads (Sabnis, Chatterjee, Grewal, & Lilien, 2012; Tittle, 1990). Online leads can directly be generated from website data via number of ways SEO Search engine optimization and social media, Search Engine Advertising.

Once leads are created, they are ready to be scored. These leads must be pursued first and ranked accordingly. Next to that, Hornstein come up with a lead scoring system based on four variables: source, need, timing, and budget. we can allocate three points per variable to one lead. This automatically results in a minimum score of four and a maximum score of twelve for every lead (Hornstein, 2005).

Online lead scoring can used in combination with offline data and only lastly updated data can be used as feed for lead score model. Four factors that used in calculation of the lead grade. First, keep track of referral sources to find out which ones converted best. Second, track the number of times a lead has visited the company's websites, when a lead visited the website, and which pages a lead visited. A page can indicate how far the lead is in the buying journey

Companies can use different criteria and rank and weighs for making purchase decisions and therefore it is difficult to decide a unique quantitative lead score model.

There are few models developed based on offline approach which are completely different from online approach.

III. THEORETICAL UNDERSTANDING

There are different ways to calculate a lead score:

Calculate Manual Lead Score

1.First Calculate lead to customer conversion rate of all your leads.

Our lead to actual customer conversion rate is equal to the number of newly onboarded customers you acquire, divided by the total number of leads created. We can use this conversion rate as our benchmark.

2. Select multiple attributes customers who you believe were higher quality leads.

Attributes could be customers who requested a free trial for duration, or customers in the insurance, finance industry, or customers with 40-50 employees.

We can use various type of skillset while selecting attributes to include in your model. We will choose attributes base on those conversations you had with sales team, analytics, and so on -- but overall, it's a prediction call. You could have ten different people do the same exercise, and they could come up with ten different models. If your scoring is based on the data, we mentioned historically.

3. Calculate the individual close rates of each of those attributes.

To generate the close rates of each type of activity a user takes on site or the type of user taking that activity incorporated because it decides the activity you'll take as part of result.

To calculate how many users become qualified leads (and ultimately, peoples) based on the activity they take or who they are in relation to your other core customer. We can use these close rates to generate lead score as below

4. Evaluate the close rates of each attribute with your all-close rate and allocate point values accordingly.

Observe the attributes with close rates that are significantly more than your close rate. Then, select which attributes you'll assign points to, and if so, how many points. Baseline the point values of each attribute on the mean of each close rates.

The actual point values will be a little different, we will try for more reliable way. For e.g, if close rate value is 2% and our model close rate is 21%, then the close rate of the proposed score model attribute is 21 overall close rate.

IV. PROPOSED WORK

1. Logistic Regression Lead Scoring

The basic method, above, for calculating a lead score is a great start. However, the most mathematically expert method is one that uses a data mining technique, such as logistic regression.

Data mining techniques are more complicated, and often more like your actual close rates as a result. Logistic regression includes and generate a formula in Excel that'll divide the probability that a lead will close into a customer. It's more precise than the technique we've outlined above since it's a broader approach that considers how all the customer attributes like company, company size, and whether or not requested an attempt -- interact with each other.

2. Prediction Lead Scoring

Generating a lead score can be great things for your business: improve the lead transition process, rise in lead conversion rate, improve productivity, and more. We can conclude from the two methods above, coming up with a scoring system can be a time taking task when used manually.

We must evaluate model with scoring criteria it's not like we set it and not trained it ss you get result from your team and stress test your scores; you'll need to modify lead-scoring system on a regularly to ensure it remains accurate.

That's where predictive scoring pitch in. Predictive scoring uses machine learning to compile thousands of data set to identify your best possible leads, so you don't have to. Predictive scoring scans what information your customers have in common, and different offset and with a methodology that simplify your contacts by importance based on their calibre to become customers.

3. Online Lead Scoring Model:

Existing credit scoring models need offline training, which makes it difficult to predict online learning and changes in the models. These models are generally developed and validated offline. These models cannot be updated online when they are executing. They usually are retrained offline with new data after a period time (such as one month, one quarter or even longer).[2] However, especially for P2P lending, the transactions are very frequent. Many new data transactions will be generated which will cause the data distribution of lending to change before retraining the model. If updated data is not available, it will affect the accuracy of a updated credit scoring model. A credit scoring model must be able to be trained and updated online to be suitable for scenarios where P2P data grows rapidly and changes frequently [2].

To solve above problems, we can propose online lead scoring model on machine learning methods. Online lead scoring model includes gradient-based decision tree and neural network. Such lead scoring model has online training and update capabilities and can handle various types of features. Inspired from [3], we propose several ways to obtain these parameters:

4. Elicitation method:

we propose here to estimate each parameter by asking questions to the domain experts, The use of such a limited set of statements are often makes it easier for the domain expert to provide assessments of conditional probability.

With this solution, we need to ask p + 1 questions, without directly asking expert' preferences. Also, we can create reward system based on domain which can be configurable. Reward points will be credited to user wallet which is time bound. Reward points are convertible to currency ratios like 1-point equals to 1 Rupee or equivalent type of currency. We can modify reward system in such way that measure the lead score in iterative manner which will result in improvement in business.

5. Ranking method:

In the opposite here, reference [3] we propose to simply obtain expert' preferences by asking him to sort the parent factors according to their importance. We then assign weights Wr Rrank i to each parent (1 for the worst parent, and p + 1 to the best one). Normalizing these weights (by dividing by the max. value)) give us values between 1 p and 1. Using probability va¬-lues can lead to bias our model by providing importance to the leak. When the number of parents is minimal, and a huge importance to the best factor. For this reason, we propose to re-calibrate these weights, by asking two questions to the expert, corresponding to the probabilities P_{min} of the worst factor (the one with Rank i = 1 and Pmax for the best factor (the one with wr Rank i = p + 1).

$$P \text{ Ranking } i = P_{min} + \text{WrRank } i - 1 \, p \, (P_{max} - P_{min})$$

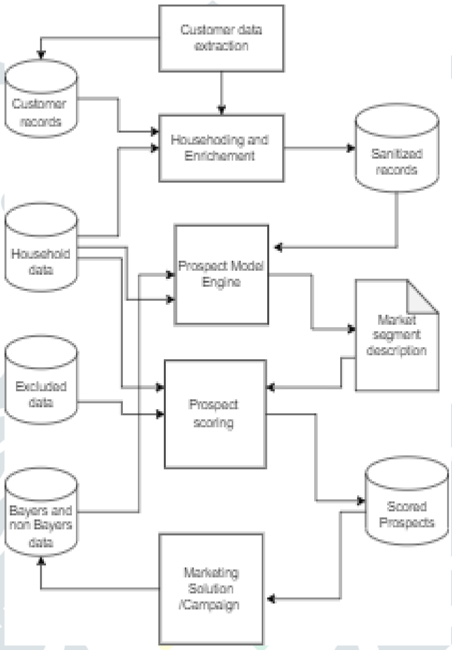

Fig.a Block diagram Lead score Model

Table 1 Lead Score Attributes

Attributes
Prospect ID
Lead Number
Lead Origin
Lead Source
Do Not Email
Do Not Call
Converted
Total Visits
Total Time Spent on Website
Page Views Per Visit
Last action
Country
Specialization
How did you hear about company
Occupation

Table 2. Rank Calculation

Profile	Rank
>80%	A1
50- 80%	A2
25 - 50%	A3
< 25%	A4

Table 3. Conversion result based on LR Model

No	Conversion	Conversion Probability	Prediction	Lead Score
13	0	0.017378	0	3
19	0	0.023034	0	4
30	0	0.177731	0	17
66	1	0.860935	1	80
88	0	0.268573	0	28
94	0	0.14934	0	12

Table 4. Reward points system for user Wallet:(Own Construction)

Goal	Reward Type	Reward Points	Duration
Sign Up	Short	50	1 month
Registration Step 1	Medium	100	2 months
Registration Step 2	Medium	100	2 months
Car	Long	300	6Month
Education	Long	500	Configurable
Home	Long	800	Configurable
Festivals	Short	200	1month/15days
Discount on processing fees	Short	200	1 month
Discount on product price	Short	Configurable	1 month
Electronic communication	Long	50	Configurable

Reward System for prospect conversion:

Based on user choice or selection over website particular points will be credited to user wallet for configured period [7].

Like if user who only signed up in such case 50 reward points will be credited to user wallet which will again increase when he complete nest registration forms with detailed information then next 100 points will be created to user wallet. User can avail those wallet points while purchasing product or services from business.

VI Conclusion

This paper described some results and observations based on a recent study with an insurance company for identifying prospects for life product families like retirement, house goal, education, car etc. We have experimented several techniques during the study and devised a model that exploits the existing user data in addition with data collected from a sample of the population. In combination with offline data if we add online lead score prediction it will provide better results. Also based on various data sets if we use mixture of algorithms like logistic regression and Bayesian network it will provide accurate lead prediction. Here we have proposed and designed dynamic reward system. We can modify this reward system based on user experience and measure conversion ratio at each iteration. We can design and implement generic wallet API which will be invoked or consumed by any organization. As part of future scope, we can implement microservice based architecture for wallet API system and reward system which will be plug and play for any business. This model is effective to increase conversion ratio form prospect to actual customer, and it is being deployed by an insurance industry client to support its sales team.

VI REFERENCES

[1] Classification Methods Applied to Credit Scoring with Collateral Germanno Teles;Joel J. P. C. Rodrigues;Kashif Saleem;Sergei A. Kozlov Year: 2020.

[2] A Deep Learning Based Online Credit Scoring Model for P2P Lending Zaimei Zhang;Kun Niu;Yan Liu Year: 2020

[3] Customer Relationship Management and Small Data — Application of Bayesian Network Elicitation Techniques for Building a Lead Scoring Model Youssef Benhaddou; Philippe Leray Year:2017

[4] Behavioural Scoring Based on Social Activity and Financial Analytics Anmol Gupta, Sandhya Pandey, Harsh Krishna, Subham Pramanik, P. Gouthaman

[5] Feature Selection Using Multiple Ranks with Majority Vote-Based Relative Aggregate Scoring Model for Parkinson Dataset

[6] Adaptive Softmax Regression for Credit Scoring Lkhagvadorj Munkhdalai, Khishigsuren Davagdorj, Van-Huy Pham, Keun Ho Ryu

[7] Proposing a Generic Online Lead Scoring Model for a B2C Market C.W.J.M. (Caro) Swelsen July 2019