IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006222|

Predictive Lead Scoring Using Machine Learning

Dr. S.A Sahaaya Arul Mary[1], Kavi Kumaran S[2], Krishna Kumar V[3]

Professor, Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, India[1]

Department of Computer Science and Engineering, Saranathan College of Engineering, Trichy, India[2, 3]

ABSTRACT: In the last decades, machine learning has become quite popular for solving business problems as it often delivers high-quality and efficient solutions. Moreover, the amount of data collected by companies has grown substantially, which has contributed to this trend. Companies do not have enough resources to contact every lead, so contact prioritization is essential.Predictive lead scoring is a data-driven lead scoring methodology that uses historical and activity data and predictive modelling to identify the sales leads that are most likely to convert. It takes out or reduces the element of human error and increases the accuracy of identifying quality leads. Predictive lead scoring uses predictive modelling, a common statistical technique to predict future behavior based on past behavior.Lead scoring supports contact prioritization as companies do not have enough resources to contact every lead by assigning a value to each lead based on their actions or characteristics. Lead scoring contributes to higher lead conversion rates.

I. INTRODUCTION

Selling something can be a hard work. A business might have many potential customers' leads but most of those customers won't turn into actual paying customers in the end. A sales team has to sort through a long list of potential customers and figure out how to spend their time. That's where lead scoring comes in. This is a system that analyses attributes about each new lead in relation to the chances of that lead actually becoming a customer and uses that analysis to score and rank all of the potential customers. With that new ranking, The sales team can then prioritize their time and only spend time on the leads that are highly likely to become paying customers.

Predictive lead scoring is a system that involves the use of algorithms to score leads instead of arbitrarily decided frameworks. The algorithms take the existing data and use it to determine which of your leads are sales qualified. The algorithms tend to use the same data that traditional lead scoring uses and matches this data against how leads have performed in the past. One of the key advantages of predictive lead scoring is that it removes the need for humans to create the scoring system themselves. Predictive lead scoring can look deeper into past data and identify patterns that humans are likely to miss, helping to make the leads more accurately qualified. More accurately qualified leads, in turn, means more sales.

II. LITERATURE SURVEY

[1] Indian higher education sector is characterized by severecompetition. The market size is 91.7 USD and its contribution to GDP is 3.8%.The admissions market is an evolving one with digital players' significantrole. Students depend on these portals for making decisions pertaining to aselection of an institution. These companies make revenue by providing variousdigital marketing services to the educational institutions. Selling leads is oneof the critical activities as part of their package offer to various educationalinstitutions. The major problem faced by these companies is thelow conversion rate of the leads at the client end. The main objective of thisstudy is designing and developing a lead score model for these companies toqualify the leads. Methods: The research focuses on developing a lead scoremodel for these companies to qualify the leads based on three parameters— explicit parameters, implicit parameters, and negative parameters. Theseparameters are chosen because they represent the lead behaviour andengagement level. The overall lead score is calculated by assigning scores foreach parameter. The

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006222|

value for each parameter is determined based on theirimportance and also in consultation with the sales team. Then the leads areclassified into hot, warm, and cold leads based on the lead score and a leadscore matrix is created based on explicit and implicit scores. Themajor finding of the study shows that out of a sample of 1900 leads, 21% are hotleads which are sales qualified, 35% are warm leads also called as marketingqualified and 44% are cold leads. The lead score matrix is used to qualify theleads. The result of the matrix shows that 60% of the leads are qualified. Thus, ithelps the digital companies to filter out unqualified leads and manage the leadsin a better way which improves the quality of the leads delivered to the clients.This will raise the conversion rate at customer level. Such improved conversionrates reinforce the business model of digital marketing companies. The lead score model is designed with customization and applied for digitalmarketing firms in education vertical in India.

[2] Competition between organizations urges the need of using new approaches. A lead is described by an individual or organisation with interest in the [products or services that an entity sells, and machine learning appears to be a very effective way of improving the performance of the corporate leads. Predicting the success of the new leads is a critical aspect to enhance the effectiveness and efficiency of marketing, and to tackle this problem, we propose in this paper a solution in the machine learning approach. This Solution was implemented to increase the winning probability of proposals, preliminary results also suggest that by sing this solution, managers could improve the quality of business decisions.

III. EXISTING SYSTEM

Traditional lead scoring is the process of manually creating a scoring point system to a contact based on both explicit and implicit information to qualify leads that should be passed to your sales team.Essentially, a marketer will identify a series of qualifying factors that indicate whether or not a lead should be pursued or not.For example, if you are a B2B business, when a lead submits an email that ends with @yahoo.com or @gmail.com, this lead will likely be unqualified. On the flip side, you may find that a particular industry serves as a better fit for your product or service. These are all examples of qualifying factors that need to be identified to run lead scoring.Next, you will need to assign values to how important or unimportantthese factors are to qualify a lead. For example, if someone submits an email that ends with @yahoo.com, you may want to deduct 10 points to their leads score. But, if someone belongs to a particular industry, you may want to add 5 points to their lead score.

IV. PROPOSED SYSTEM

Lead scoring is a methodology used by sales and marketing departments to determine the worthiness of leads, or potential customers, by attaching values to them based on their behavior relating to their interest in products or services. The "value" of each lead varies from company to company, but generally is characterized by the interest shown in the company or their places in the buying cycle. Companies assign point-based systems in qualifying leads or simply refer to them as "hot," "warm" or "cold" based on the history of interactions.

Companies need to figure out which leads are most likely to become customers. This is fundamental for efficiency, because if these leads are prioritized and contacted first, Sales will save a lot of time on leads that are not ready to purchase yet. The relevant leads are those that have the need and the budget now and lead scoring is crucial because it helps identifying those ready for sale leads. Lead scoring is a subtask that consists of calculating and assigning a value to each lead based on its actions and attributes. Its main goal is to prioritize the leads, respond to them adequately, and increase the conversion rate. Lead scoring allows Sales representatives to identify leads that are ready for immediate follow-up, contributing to lower costs per lead and higher conversion rates. Lead scoring can be based on a wide range of features. The data is usually behavioral data, such as reactions to marketing campaigns and visits to the company's website, or demographic information. Furthermore, the data collected by companies can be classified as explicit data, if it is directly obtained from leads' input, or implicit, if it is obtained by data collected on the leads' actions.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006222|

The Proposed model is developed by using logistic regression algorithm. The overall architecture based on logistic regression is light weight and presents higher accuracy. We have a used a dataset from Kaggle which consists of 9000 leads data and 37 fields. In this dataset we will use the above algorithm in order to get the accuracy and provide lead score to new dataset. Relevant fields are considered after data cleaning and data analysis process. Modelling is based on these fields. Now upcoming leads dataset is provided to the model. Based on the conversion probability a lead score is assigned to each lead. Users can make use of this lead score to make a decision on whether to approach a lead or not.

V.ARCHITECTURAL DESIGN

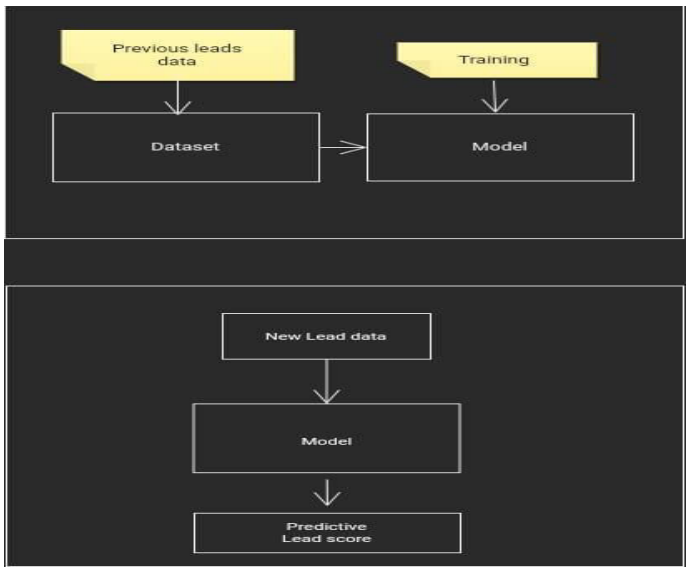

Fig.1 Architecture Diagram

The Architecture Diagram displays how the leads dataset is used to produce a model using machine learning techniques. With this model, the upcoming leads data is inputted to determine the lead score.

VI. MODULE DESCRIPTION

- Logistic Regression Model:

Logistic regression model calculates the probability of lead conversion (which is 0 -1) and converts them to 1 - 100 and assigns them as lead score.

- Decision Tree:

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether a lead is from Delhi or Mumbai), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes).

- Random Forest:

The random forest algorithm establishes the outcome based on the predictions of the decision trees. It predicts by taking the average or mean of the output from various trees. Increasing the number of trees increases the precision of the outcome.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006222|

VII. VALIDATION

The performance of the training model and the overall application can be validated using the classification report. The precision and recall values are above 0.65 and the weighted average accuracy is more than 70%. This shows that our model is capable of predicting the appropriate lead score. The confusion matrix shows the situations when the model's output is wrong

```
print(classification_report(y_test['Converted'],df))
```

	precision	recall	f1-score	support
0	0.80	0.87	0.84	1662
1	0.77	0.66	0.71	1061
accuracy			0.79	2723
macro avg	0.79	0.77	0.77	2723
weighted avg	0.79	0.79	0.79	2723

Fig.2Decision Tree Classification Report

```
print(classification_report(y_test['Converted'],df))
```

	precision	recall	f1-score	support
0	0.79	0.81	0.80	1662
1	0.69	0.67	0.68	1061
accuracy			0.76	2723
macro avg	0.74	0.74	0.74	2723
weighted avg	0.75	0.76	0.75	2723

Fig.3Random Forest Classification Report

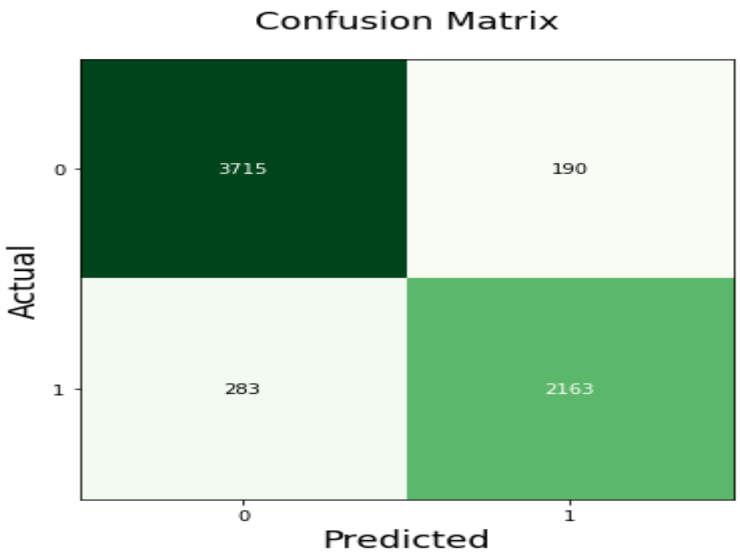

Fig.4 Confusion Matrix

VIII. RESULTS

. Lead score is predicted using the trained model and historical dataset for new leads. The screenshot of produced lead score is displayed below.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006222|

Index	Lead Score
0	7.33
1	85.22
2	32.1
3	95.6
4	5.6

Fig.5Predicted Lead Score

IX. CONCLUSION

In this work, an approach is proposed to use machine learning techniques to predict lead score. It is more consistent in determining the lead score based on the historical dataset. Based on the experimental results, it is shown that the system is working fine and produces desired results such as:

- Extracting the historical dataset
- Producing models based on regression techniques
- Predicting the lead score over a lead.

REFERENCES

[1] Kumar, G. and Hariharanath, K., 2021. Designing a Lead Score Model for Digital Marketing Firms in Education Vertical in India. *Indian Journal of Science and Technology*, 14(16), pp.1302-1309.
[2] Custódio, João & Costa, Carlos & Carvalho, Joao. (2020). Success Prediction of Leads – A Machine Learning Approach. 1-6. 10.23919/CISTI49556.2020.9141002.
[3] (The Basic Science Behind Lead Scoring, 2022)) Salesforce.com. 2022. *The Basic Science Behind Lead Scoring*.
[4] Peng, C., Lee, K. and Ingersoll, G., 2002. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), pp.3-14
[5]de Ville, B., 2013. Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), pp.448-455.
[6] Chen, J., 2020. An Introduction to Machine Learning for Panel Data: Decision Trees, Random Forests, and Other Dendrological Methods. *SSRN Electronic Journal*,.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details