Enhancing Sales of Digital Platforms using ML

Abhishek Upadhyay [1], Himanshu Chaturvedi [2], Keshav Goyal [3], Kushagra [4]
[1,2,3,4]*Department of Computer Engineering, Raj Kumar Goel of Institute of Technology, Ghaziabad*

Abstract **- In the competitive scenario companies used to collect a large set of data, as most of the decision are related with the customer choice only. The data which is collected that can be the main source of important competitive advantage by utilizing it in estimating a contact's purchase probability using predictive analytics using supervised as well as unsupervised algorithms respectively. The result that came after the estimation of purchasing probability which can be useful for the company to estimate their problem of selling the course to the one who have maximum chance to be a lead.**
We have used two types of data which are as follows:
The first one is the historical data which can be used as to training the data for the classification algos such as random forest, k-nearest neighbour [1], support vector machine [2] and many more which tells that the particular customer have purchased. the second data is provided by the developer side to train the data.

I.INTRODUCTION

In this pandemic situation digital platforms like e-learning platforms, e-commerce platform gaining lots of rush as everyone are getting close and close to online platforms more and more since going out for learning or buying something is not a good idea in this situation. This rush can create a great revenue for these platforms if managed smartly as maintain this amount of people every day is not easy for sales teams but this problem can be solved if technology comes into play.
In today's world where time is a commodity of sales. One of the most critical business decisions of a company is related to customer acquisition. During the acquisition phase of the customer life cycle, companies try to convert leads into customers through different methods. One useful way to save time while Increasing sales is to make sure that we focus on the best available leads and not waste time on inactive leads. In order to increase the conversion rate, so we have used the customer specific data like time spent on a website, total visits, occupation etc. and the customers are then ranked in order of the probability of conversion, which are then pursued by salespeople.

II. LITERATURE REVIEW

In present scenario, various companies specially an online platform needs to generate and collect a huge amount of proper data. As for this, organizations increasingly depend on data-driven decision to support the user for the data collection. Lead scoring, or marketing and customer relationship management is used by the companies are some nearby common to these processes.
"A dedicated team of around 100 people, whose work is supplemented by content written by Nobel laureates, historians, curators and professors named as Britannica Academic stated, that: "Machine learning, in artificial intelligence [3], the discipline concerned with the implementation of computer software that can learn autonomously" the nature and many things. But there are some glitches to tackle the machine learning models.
A huge area of marketing, presently termed as relationship marketing, is 'the ongoing process of engaging in collaborative activities in programs with immediate and end-user customers to create or enhance mutual economic, social and psychological value, profitably for the company.' A dedicated team of around 100 people, whose work is supplemented by content written by Nobel laureates, historians, curators and professors named as Britannica Academic stated, that: "Machine learning, in artificial intelligence (a subject within computer science) [4], the discipline concerned with the implementation of computer software that can learn autonomously" the nature and many things. But there are some glitches to tackle the machine learning models. It suggests that, if the necessary information is available then it is possible data available, there may be a way to optimize sales and certain approaches can be taken by the teams respectively.
Basically, supervised learning algorithms are used when we have both data i.e. data consists the input values as well as the output result by which efficiency

is maximised. By calculating the error from the difference between the model's prediction and the actual given value it is possible to change the model's prediction to minimize the prediction error. Supervised learning [4] solves regression [5], fraud and classification [6] problems. By using Regression and Classification, it is possible to estimate the output result.

Supervised learning is used when there is input as well as output data available. Instead, the algorithm tries to find the predictions data on its own. The output result also provided as the training set. Supervised learning is used to find clusters and put unseen data in a suitable manner.

Tracing these activities and applying various advanced business analytics tools or machine learning algorithms for predicting the sales strategy that can enhance company-customer relation.

Collecting the useful information(data) takes place via online input from the user side and also the direct input from the developer sides. In general, the collection of input data and the conclusion depend on the graph.

To reformulate the observations calculated in our specific context, we can state that automated marketing is the process of utilizing data from tracking online actions of potential leads to learn about the factors.

These potential buyers could estimate in identifying the ones who are more likely to turn into actual customers and neglecting the ones who cannot be a lead customer, while the tools to support these processes in an automated way are readily available, there are very few.

Based on this brief discussion, we present a brief literature review on lead scoring and machine learning applications in automated customer relationship management

III. METHODOLOGY

Predictive analytics is also a process employed within the business to customer marketing to rank lead supported their activities within the study, the general recommended process from for predictive analytics in information systems research is applied. With focus of the research being on the event and evaluation of possible predictive machine learning models for automated lead scoring, data understanding focuses on examining the knowledge and identifying and correcting potential problems present in it. The calculated purchase probability can then be used by companies to resolve different business problem. within the information preparation process, the knowledge is transformed so on address missing values and outliers, and to create a variable structure utilizing feature extraction, filtering and have selection that's appropriate for further machine learning model building.

A. Data Description

The data on which analysis the goal of the information description is to record all information about the knowledge files and their contents so as that somebody can use the knowledge in a very future research and understand the information content and structure. Documentation and more specifically metadata both provide information about the information at hand. Describing your data is significant. Systematically described research data is that the key to making your data findable, understandable and reusable. Overall data quality improves with clear data description and detailed documentation and metadata.

B. Data Preprocessing

In any Machine Learning process, data pre- processing is that step within which the information gets transformed, or encoded, to bring it to such a state that now the machine can easily parse it. In other words, the features of the information can now be easily interpreted by the algorithm.

Data pre-processing may affect the way during which outcomes of the ultimate processing is interpreted. This aspect should be carefully considered when interpretation of the results may be a key point, such within the multivariate processing of chemical data.

C. Data Cleaning

Data cleaning is that the process of preparing data for analysis by removing bad data, organizing the data, and filling within the null values. Ultimately, cleaning data prepares the information for the method of knowledge mining when the foremost valuable information will be pulled from the information set.

Data cleaning is one in every of the important parts of machine learning. It plays a big part in building a model. Data Cleaning is one in every of those things that everybody does but nobody really talks about. It surely isn't the fanciest a part of machine learning and at the identical time.

D.POINT DISTRIBUTION

Visitor 1

A person from a working company (+5) visits your site and reads a page for more than a minute (+5) and the counts to the no. of visits (+2 for each visit).

Total score = 12

Visitor 2

A student (+4) visits your site and scrolls the page for 5 times (+2) for more than a minute (+5), and last activity (+2) and preview the course (+1) and downloads the preview lecture (+2)

Total score = 16

Visitor 3

A random person visits the course for less than 20 seconds and then cancels the page and no other visit, scrolling the page. Hence the person had mistakenly visited the page / not interested in that particular course.

Total score = 0

The company will not contact him/her.

IV. RESULT

A. Results based on Algorithms:

Based on the analysis of the dataset we've seen that a lot of columns aren't adding any information to the model also these columns are leading to the degradation of the accuracy of the model hence we will drop these tables for more precision. Recursive Feature Elimination [7] and Principal Component Analysis [8] are used for the reduction of the dimensionality. High dimensionality results to the matter of overfitting, therefore help within the inefficiency of the model. After reducing the ineffective columns, the models showed high accuracy.

Table

Model	Accuracy
LR	91.36
KNN	91.06
SVM	92.37
NB	87.54
RF	91.11

The above table shows that Support Vector Machine model showed the highest accuracy, it is also seen that other models are also very much close in terms of accuracy.

V. CONCLUSION

After the work on "Enhancing sales strategy of e-learning platform using ML", it's concluded that it's possible to estimate the sales strategy using supervised learning algorithms on the training set data. plenty of conclusion and result were extracted during the functioning on the info given by the user and from the developer sides also. we've used some unsupervised algorithm which could be a statical process which helps to convert the correlated features observations into a collection of linearly uncorrelated features with the assistance of orthogonal transformation for betterment of the result, because it reduces the info which don't seem to be useful (unwanted data) which increases efficiency within the result.

Recursive Feature Elimination are often used for feature selection algorithm which effectively choose or work on those data column and rows which are likely to supply results of the targeting result.

The model which supplies the very best percentage of efficiency comes from using Support Vector Machine of 92.37% and also the other models like Logistic Regression (accuracy of 91.36%), K Nearest Neighbour (accuracy of 91.06%) and Random Forest (accuracy of 91.11%) all showed an accuracy very near Support vector machine. Thus, we've selected Support Vector Machine for training the info for the estimation of sales efficiency for e-learning platform.

REFERENCES

[1] https://www.analyticsvidhya.com/blog/2021/04/simple-understanding-and-implementation-of-knn-algorithm/

[2] Massimiliano Pontil, Theodoros Evgeniou, "SVM Theory and Their Application", Jan 2001.

[3] Keith D. Foote, "A Brief History of Machine Learning", March 2019

[4] Vlamdir Nasteski, "Overview of Supervise Machine Learning", Dec 2017.

[5] Aqeel Anwar, "Regression Analysis in Machine Learning", towards data science, April 2020.

[6] Showan Chowdhury, "Classification using Supervised ML", ieee explore, Oct 2020.

[7] Jason Brownlee, “Recursive Feature Elimination (RFE) for Feature Selection”, machinelearningmastery.com, March 2020.
[8] https://builtin.com/data-science/step-step-explanation-principal-component-analysis