



Predicting and Defining Business to Business Sales Success with Machine Learning

Dr. B. Sujatha | P. Sandeep Chandra | N. Manaswini | | G. Ashok Reddy

Department of Computer Science and Engineering, Godavari Institute of Engineering and Technology (A), JNTUK, Kakinada.

To Cite this Article

Dr. B. Sujatha, P. Sandeep Chandra, N. Manaswini and G. Ashok Reddy. Predicting and Defining Business to Business Sales Success with Machine Learning. International Journal for Modern Trends in Science and Technology 2022, 8(S03), pp. 132-136. <https://doi.org/10.46501/IJMTST08S0332>

Article Info

Received: 26 April 2022; Accepted: 24 May 2022; Published: 30 May 2022.

ABSTRACT

The main aspect of the project is that to provide an accurate application that can detect the sales of the business and the business profits, so that the particular organization can improve the sales and start that business at that place, attain maximum business sales, this entire process will be carried out under the supervision of the experts who will be working in the industry for more than 10 years and they will identify whether the business is profitable to start or not. We will be using various machine learning algorithms like decision tree, random forest, advanced-boost, gradient boost algorithms to identify the profits of the sales and the probability of the sales success of our business that we are intended to start at a particular place. This project will be used for processing for both structured and unstructured data and it can be used by both big scale and small scale industries who are in the intentions to start a business in a particular area.

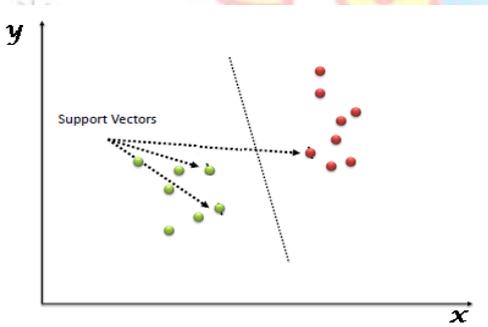
KEYWORDS: Gradient boost, Random Forest, increasing sales, structured and unstructured data

1. INTRODUCTION

The organization which has given the information for this development purpose has a wide range of experience in dealing sales and the people who are involved in this research have more than 10 years of experience in dealing this project. It is not easy for them to make intuitions for producing accurate results that we expect to come. So, the people who really have experience are involved in this process, normal people are not involved in this process, mostly this work is performed by the private companies and they will be responsible for developing the technologies and related sale models. As the entire process is being carried out by the private companies only, the data related to this work is not available to the people publicly, so that is the reason why the research in this field is very under growth as compared to other fields in the research

process, the work in this field is can be related to either to forecasting the sales aggregate or it can be based on custom algorithms instead of attaining simple results for the given data. And this algorithms are very different from the tools that are used in the industry by the data scientists. Other related research has focused on integrating very precise Due to a lack of visibility into work that predicts, machine learning algorithms based on sales pipeline data into an organization's procedures [4], as well as explaining the sales success tendency, this study establishes a foundation knowledge on the subject. This research compares and applies a number of well-known approaches for identifying and rating propensities, the bulk They belong to the decision tree model category. Salesforce.com, the company's customer relationship management system, provided the data for this project (SFDC).

SFDC is a software-as-a-service (SaaS) application. tool that lets salespeople keep track of customer contacts and sales possibilities as they progress through the sales funnel. A static picture of sales staff, Individual customer opportunities, sales representative actions, and contact information were all contained in the customer accounts and account histories. data. Some of the system's inputs were created automatically and were machine-readable. This is beneficial in a decision support application when learning the best categorical choice variables for classification is desired. Overall, these findings show that using an inverted classification strategy can improve data quality, and that this benefit improves as the number of attributes is increased. replaced. In the context of a decision support application, this indicates that when there are a high number of decision factors, the inverted classification strategy is most successful. In a real-life situation, this is also the most intriguing case. One of the most fundamental reasons why ANNs do not generate the essential interest in the industry is a lack of explanatory capability. As a result, it is vital for an ANN to be able to explain itself.

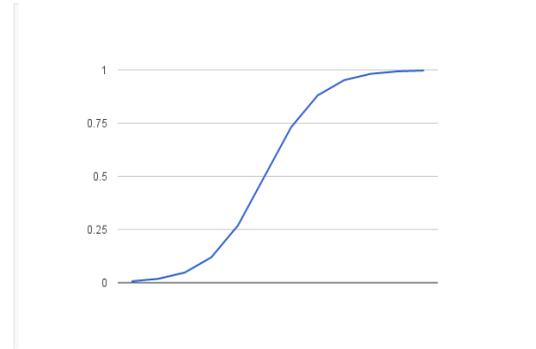


The above diagram represents the support vector machine under two dimensional plane

LOGISTIC-REGRESSION

Logistic regression is a supervised learning technique. Logical regressions become a classification approach when a decision threshold is used, and they are used to solve classification problems. It's used to create educated estimations about target variables' probabilities. The main distinction between logical and linear regression is that linear regression is used to solve regression problems. Logistic regression is used to solve classification difficulties. The three types of logistic regression are binomial, multinomial, and ordinal

logistic regression. The binomial form is the most fundamental in logistical regression.



LOGISTIC FUNCTION

Logistic function is the function that is used in logistic regression process.

NAIVE BAYES CLASSIFIERS ALGORITHM

For binary and multi-class categorization problems, Naïve-Bayes is a classification algorithm. When it is presented with (binary) or (categorical input values), this method is simple to understand. As the possibilities for each and every hypothesis it is simplified to make the computations tractable,, so it is defined as (naïve-Bayes) or (foolish Bayes). Rather of trying to figure out what each attribute value $P(a_1, a_2, a_3|h)$ means, they're attempting to figure out what each attribute value $P(d_1, d_2, d_3|h)$ means. Rather of trying to figure out what each attribute value $P(d_1, d_2, d_3|h)$ means, they're attempting to figure out what each attribute value $P(d_1, d_2, d_3|h)$ means. $P(a_1|h) * P(a_2|H) = P(a_3|H) = P(a_4|H) = P(a_5|H) = P(a_6|H) = P(a_7|H) = P$. The assumptions that the attributes do not interact is a strong one that is difficult to persist true in real data. Nevertheless, for data when this assumption does not apply, the technique performs quite well.

RANDOM-FOREST ALGORITHM

Random Forest is a well-known machine learning algorithm which uses the supervised learning method. In machine learning, it can be applied for both regression and classification problems. It is based on ensemble learning, which is a method of integrating several classifiers to solve complicated problems and improve the performance of the model. As the name suggest that,, "Random Forest is a classifier that contains a

number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than relying on a single decision tree, the random forest collects the forecasts from each tree and predicts the final output based on the majority votes of predictions.

DECISION-TREE ALGORITHM

By ordering instances along the tree from the root to a leaf node, which offers the classification, decision trees are used to classify them. As seen in the figure above, an instance is categorized by starting at the tree's root node, checking the attribute specified by this node, and then progressing along the tree branch based on the attribute value. The new node's rooted subtree is then handled in the same way.

XG-BOOST ALGORITHM

XG-Boost is a decision-tree-based efficient Machine Learning technique that uses gradient boosting. In predicting problems involving unstructured data, deep neural networks surpass all existing algorithms or frameworks (pictures, text, etc.). However, for small-to-medium structured/tabular data, decision tree-based algorithms are now rated best-in-class. The graph below demonstrates the evolution of tree-based algorithms over time. The XG Boost algorithm was developed as part of a research initiative at the University of Washington. Tianqi Chen and Carlos Guestrin spoke at the SIGKDD conference, where they were acknowledged for not only winning many Kaggle challenges, but also for being the brains behind some cutting-edge industry.

THE PROJECT'S STRUCTURE (SYSTEM ANALYSIS)

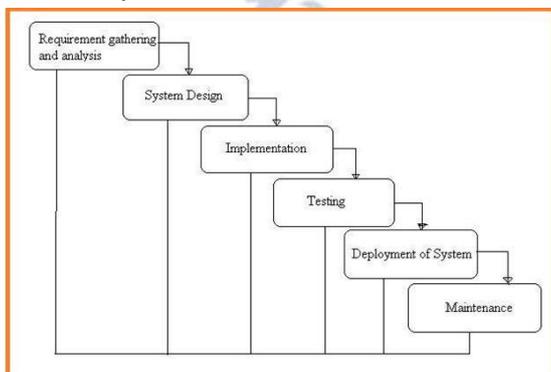


Figure-1: Software Development Life Cycle Requirements Accumulation and research Application Design System

Implementation

Testing

Deployment

Maintenance

ACCUMULATION AND ANALYSIS ARE REQUIRED

Because ours is an academic leave, it is the primary and most important stage of any undertaking. We followed Institute of electrical and electronics Journal articles for basic requirements gathering up and accumulated so many IEEE Relegated papers that we finally culled a paper titled "Individual web revisitation by setting and substance importance input" and for the analysis phase, we took referees from the paper and did a literature study of some papers and collected all of the design specifications in this stage.

DEVELOPMENT OF A SYSTEM

GUI design, Data flow diagrams design, and class diagram are the three types of system design. The use case model, the project's flow using The Class Diagram, the Sequence Diagram provide the information on the project's numerous classes and methods that must be used. The use case diagram, the project flow utilizing sequence, and the Class Diagram all provide information on the many classes in the project, as well as the methods that must be employed and the Class Diagram provides information about different classes in the project with methods that must be used in the project. by which we will implement to develop a (data-base dependent) on the types of parts in the application development.

THE UNIT- TESTING

Unit testing is a type of testing process in which is used in software development life cycle process, which divides the program application into smaller parts and starts testing on each and every part of the project separately , unit testing is implemented at all the development levels of the project it mostly performed by the developers and some times by the question and answers team.

THE MANUAL-TESTING

Manual testing, as the name implies, is a test technique in which a Q/A manually reviews a software application to find problems.Q/As do this by following a documented test strategy that sets out a set of specified test cases. The Q/A is in responsibility of examining the mobile and web

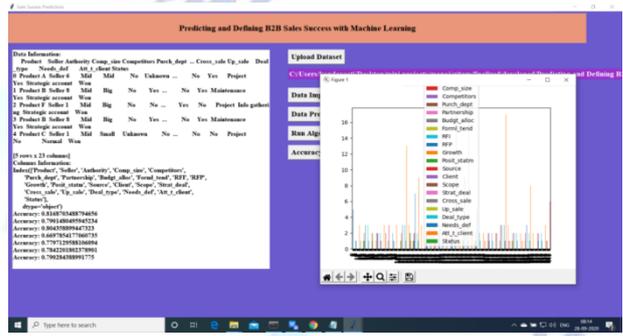
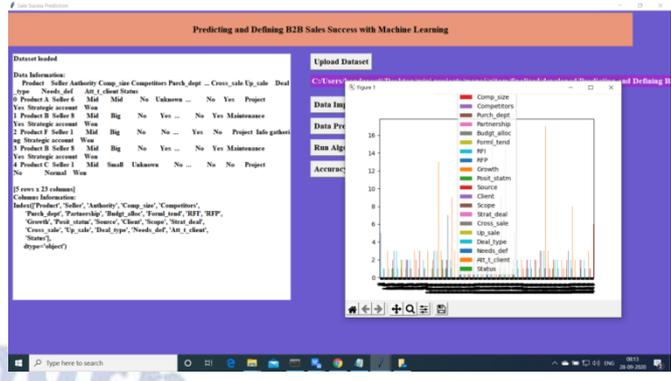
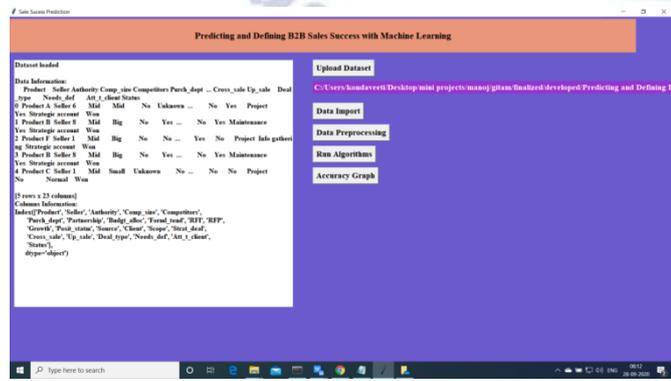
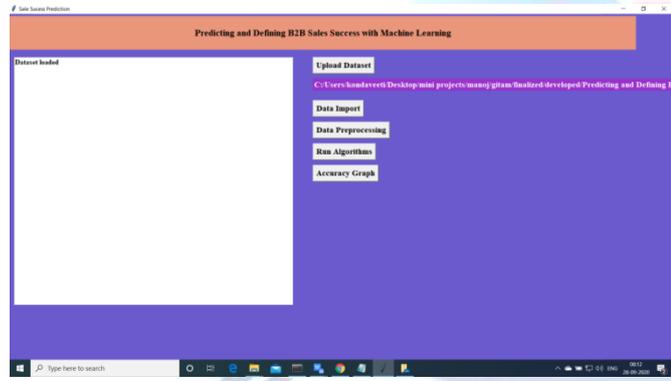
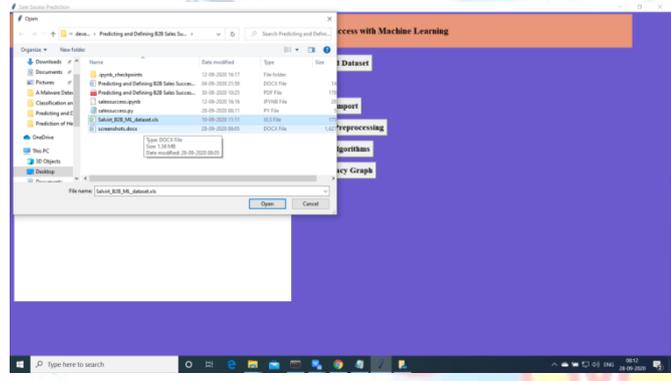
application's performance from the perspective of the user.

MAINTENANCE AND DEPLOYMENT OF SYSTEM

After the application has been completed, we have to deploy In the actual world, the client system. During the academic break, we will simply installed the client system in the college laboratory, replete with the required software, and the Windows-operating-system.

RESULTS AND DISCUSSIONS

Step 1 : Double click on the file name so, that the program interface opens



CONCLUSION

This study was the initial stage in the development of a larger endeavor to operationalize predictive modelling on success of sales for packaging paper and company. As a result, any large corporation has obstacles such as developing deep knowledge of the data and enlisting the help of big companies. To assist in the collection of trustworthy data. Despite some early data discrepancies, overall accuracy was encouraging, indicating that large increases might be developed with good data quality and quantity, additional feature-dependent analysis and tuning and for new approaches like neural networks. The study also yielded new insights into the characteristics that drive sales success. New questions, on the other hand, frequently accompany new insights: What kinds of data should be collected, for example, to improve the model's prediction abilities? What cultural changes are required to improve data collection? Given that the project as a whole being a highly incremental process, this is expected. In this instance, there may appear to be an unlimited number of possible next steps to take. With this in mind, the team has identified a handful that they believe are smart to be considered. Currently, the organization might use a variable version to make predictions on chances in the pipeline for the divisions where the precision is sufficient. It is better to accomplish and the goal of forecasting available possibilities, it will

be good to document and simulate how opportunity fields evolve over time, maybe via periodic snapshots. It will enable the organization to forecast at various steps of the process of opportunity lifecycle mechanism. Another key using these types of predicting models is to help determine where to focus sales time and money in order to maximize profits. Predictions from accurate models are also worth rolling up into aggregate sales forecasts and adjusting existing “bottom-up” methods. Before these applications would be addressed however, data ops resources would be required to perform a number of critical tasks: continue building

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Implicit (Sales Cloud by Salesforce.com).[Online].Available: <https://www.salesforce.com/blog/2014/08/infographic-7-powerfulpredictors-closed-won-opportunity-gp.html>
- [2] Insight Squared. [Online]. Available: <https://www.insightsquared.com/features/sales-forecasting/>
- [3] J. Yan, C. Zhang, H. Zha, et al, “On Machine Learning towards Predictive Sales Pipeline Analytics.” Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pp. 1945-1951,2015.[Online].Available: <https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/download/9444/9488> [Accessed: Mon. 24 Sept. 2018].
- [4] M. Bohaneca, M.K. Borstnarb, M. Robnik-Sikonja, “Integration of machine learning insights into organizational learning: A case of B2B sales forecasting.” 28th Bled conference, June 7-10, 2015. [Online].Available:[https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF238AB59EEC1257E5B004B39B7/\\$File/2_Bohanec.pdf](https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF238AB59EEC1257E5B004B39B7/$File/2_Bohanec.pdf) [Accessed: Tue. 25 Sept. 2018].
- [5] M. Bohaneca, M.K. Borstnarb, M Robnik-Sikonja, “Explaining machine learning models in sales predictions.” Expert Systems with Applications, no. 71, pp. 416-428, 2017. [Online]. Available: <http://lkm.fri.uni-lj.si/rmarko/papers/Bohanec17-ESwA-preprint.pdf>