ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group -I) Journal Volume 12, Iss 1, 2023

USING MACHINE LEARNING ALGORITHMS TO PREDICT AND DEFINING B2B SALES SUCCESS

Angalakuduru Srinivasa Rao[1], Ghantasala Venugopal[2]

[1]Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

[2]Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

Abstract_

The goals of this undertaking are two-fold: 1) to use statistical modeling strategies to assist a Fortune five hundred paper and packaging organisation codify what drives income success and 2) to advance a mannequin that can predict income success with a sensible diploma of accuracy. The favored long-run end result is to allow the business enterprise to enhance each top-line income and bottom-line income via growing income shut rates, shortening income cycles, and reducing the value of sales. The lookup group generated numerous fashions to predict win propensities for character income opportunities, selecting the mannequin with the biggest predictive electricity and potential to generate insights to use as the spine for a customer tool. To accomplish this, the group leveraged structured and unstructured facts from the company's Salesforce.com patron relationship administration system.

Index Terms - statistical modeling; decision tree; machine learning; process improvement

1.INTRODUCTION:

The paper and packaging company that provided the data for this research has a long history of sales expertise. This expertise is captured predominantly in the intuition of sales representatives, many of whom have worked in the industry for 20 years or more. Intuition is not easy to record and disseminate across an entire sales force, however, and thus one of the company's most valuable resources is inaccessible to the broader organization. As a result, the company tasked this team with extracting the most important factors in driving sales success and modeling win propensities using data from their customer relationship management (CRM) system. Most prior work in this space has been performed by private companies, both those that have developed proprietary technologies for internal use and those that sell B2B services related to predictive sales modeling. As a result, research in the field is typically unavailable to the public. Some examples include Implisit [1]—a company recently acquired by Salesforce.com that focuses on data

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group -I) Journal Volume 12, Iss 1, 2023

automation and predictive modeling—and InsightSquared [2], which sells software that includes a capability to forecast sales outcomes. The academic work that does exist either is related to forecasting aggregate sales instead of scoring opportunitylevel propensity, or is based on custom algorithms that fall outside the standard tools used by data scientists in industry. The earliest relevant publication dates only to 2015, in which a joint team from Chinese and US universities employed a two-dimensional Hawkes Process model on seller-lead interactions to score win propensity [3]. Other relevant research has centered around applying highly accurate machine learning algorithms based on sales pipeline data to integrate the insights they produce into an organization's practices [4], and explaining the output of black-box machine learning models [5]. Considering the lack of visibility into work predicting sales outcome propensity, this research serves to create an initial baseline of understanding on the subject. This project applies and compares several well-known methods for classifying and scoring propensities, a majority of which fall into the category of decision tree modeling.

2.LITERATURE SURVEY

2.1 M. Bohaneca, M.K. Borstnarb, M. Robnik-Sikonja, "Integration of machine learning insights into organizational learning: A case of B2B sales forecasting." 28th Bled eConference, June 7-10, 2015. [Online]. Available: https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF238 1AB59EEC1257E5B004B39B7/$File/2_Bohanec.pdf [Accessed: Tue. 25 Sept. 2018]

Business-to-Business (B2B) sales forecasting can be described as a decision-making process, which is based on past data (internal and external), formalized rules, subjective judgment, and tacit organizational knowledge. Its consequences are measured in profit and loss. The research focus of this paper is aimed to narrow the gap between planned and realized performance, introducing a novel approach based on machine learning techniques. Preliminary results of machine learning model performance are presented, with focus on distilled visualizations that create powerful, yet human comprehensible and actionable insights, enabling positive climate for reflection and contributing to continuous organizational learning.

2.2 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 [Accessed: Tue. 25 Sept. 2018].

A complexity of commercial enterprise dynamics regularly forces decision-makers to make choices based totally on subjective intellectual models, reflecting their experience. However, lookup has proven that corporations operate higher when they follow data-driven decision-

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper , UGC CARE Listed(Group -I) Journal Volume 12, Iss 1, 2023

making. This creates an incentive to introduce intelligent, data-based choice models, which are complete and guide the interactive comparison of selection selections fundamental for the enterprise environment.

Recently, a new universal clarification methodology has been proposed, which helps the clarification of modern black-box prediction models. Uniform explanations are generated on the degree of model/individual occasion and guide what-if analysis. We current a novel use of this methodology inner an clever machine in a real-world case of business-to-business (B2B) income forecasting, a complicated mission regularly completed judgmentally. Users can validate their assumptions with the introduced explanations and take a look at their hypotheses the use of the introduced what-if parallel format representation. The effects show effectiveness and usability of the methodology. A enormous gain of the introduced approach is the opportunity to consider seller's moves and to define conventional guidelines in income strategy.

This flexibility of the strategy and easy-to-follow explanations are appropriate for many one of a kind applications. Our well-documented real-world case suggests how to resolve a selection help problem, particularly that the pleasant performing black-box fashions are inaccessible to human interplay and analysis. This may want to prolong the use of the clever structures to areas the place they have been so a long way ignored due to their insistence on understandable models. A separation of the laptop getting to know mannequin determination from mannequin rationalization is every other massive gain for professional and shrewd systems. Explanations unconnected to a unique prediction mannequin positively have an effect on acceptance of new and complicated fashions in the commercial enterprise surroundings thru their convenient evaluation and switching.

2.3 J. Yan, C. Zhang, H. Zha, et all, "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].

Sales pipeline win-propensity prediction is essential to high-quality income management. In distinction to the usage of subjective human rating, we advise a present day laptop mastering paradigm to estimate the win-propensity of income leads over time. A profile-specific two-dimensional Hawkes procedures mannequin is developed to seize the have an impact on from seller's things to do on their leads to the win outcome, coupled with lead's personalised profiles. It is inspired by means of two observations: i) retailers have a tendency to regularly center of attention their promoting things to do and efforts on a few leads for the duration of a rather quick time. This is evidenced and mirrored by means of their focused interactions with the pipeline, which include login, shopping and updating the income leads which are logged by way of the system; ii) the pending possibility is inclined to attain its win result rapidly after such temporally targeted interactions. Our mannequin is deployed and in persistent use to a large, global, B2B multinational technological know-how enter-prize (Fortune 500) with a case study. Due to the generality and flexibility of the model, it additionally enjoys the attainable applicability to different real-world problems.

3.PROPOSED SYSTEM

The research team employed several well-known classification models to extract important features from the data, in addition to calculating the win/loss propensity for each opportunity record. With the goal of modeling probability, the team chose different supervised machine learning algorithms that fit these criteria: Logistic Regression, Decision Tree, Random Forest, and XGBoost. In each of these supervised algorithms, the classifier was pre-defined with an iterative variable selection process. A classification model was then built with a training set split from the master table and used to predict win propensities 1 Prior to this date, portions of the company used the system, but it had not been rolled out companywide. examined by the actual win or loss of the opportunities in the testing set built from the remainder of observations. Variable selection was a critical component of this project. As previously stated, variables came directly from the SFDC system and went through a series of data processing steps. The main purpose of this research was to interpret features that gave the most useful information in terms of win propensity prediction accuracy. Both the quality and quantity of variables significantly affected the accuracy and efficiency of all algorithms. An important consideration about the current data was the widely varying quality of variable inputs. This issue created constraints on the algorithm-generated selection results. Therefore, the variable selection process also involved constant communication and validation between the team and company.

3.1 IMPLEMENTAION

1. Data Collection: Collect sufficient data samples and legitimate software samples.

2. Data Preporcessing:Perform effective data processing on the sample and extract the features.

3. Train and Test Modelling: Split the data into train and test data Train will be used for trainging the model and Test data to check the performace

4. Feature Selection: Further select the main features for classification.

3.2 ALGORITHMS

3.2.1 Logistic regression

Logistic regression is named for the function used at the core of the method, the logistic function.

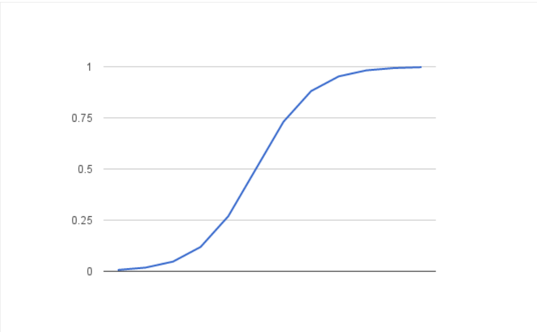
The logistic function, also called the sigmoid function was developed by

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
Research Paper UGC CARE Listed(Group -I) Journal Volume 12, Iss 1, 2023

statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

1 / (1 + e^-value)

Where e is the base of the natural logarithms (Euler's number or the EXP() function in your spreadsheet) and value is the actual numerical value that you want to transform. Below is a plot of the numbers between -5 and 5 transformed into the range 0 and 1 using the logistic function.



Logistic Function

Now that we know what the logistic function is, let's see how it is used in logistic regression.

3.2.2 Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model.*

As the name suggests, **"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 of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

Decision tree Algorithm:

Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. An instance is classified by starting at the root node of the tree,testing the attribute specified by this node,then moving down the tree branch corresponding to the value of the attribute as shown in the above figure.This process is then repeated for the subtree rooted at the new node

3.3.3. XGBoost

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group –I) Journal Volume 12, Iss 1, 2023

considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years.

XGBoost algorithm was developed as a research project at the University of Washington. Tianqi Chen and Carlos Guestrin presented their paper at SIGKDD Conference in 2016 and caught the Machine Learning world by fire. Since its introduction, this algorithm has not only been credited with winning numerous Kaggle competitions but also for being the driving force under the hood for several cutting-edge industry applications.

4.RESULTS AND DISCUSSIONS

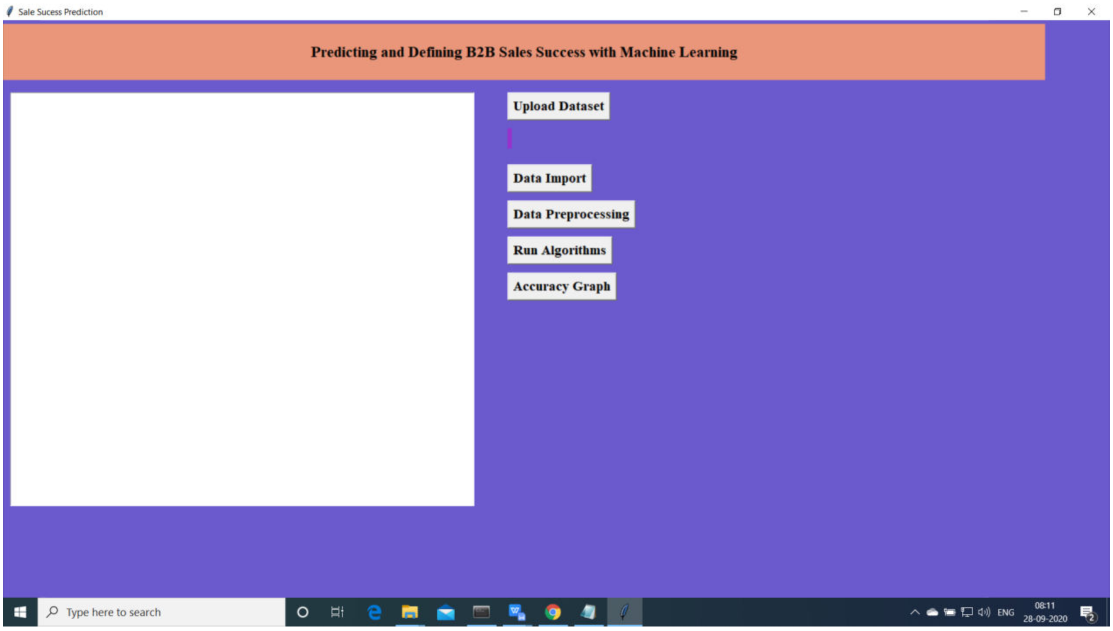


Fig 1:Home Screen

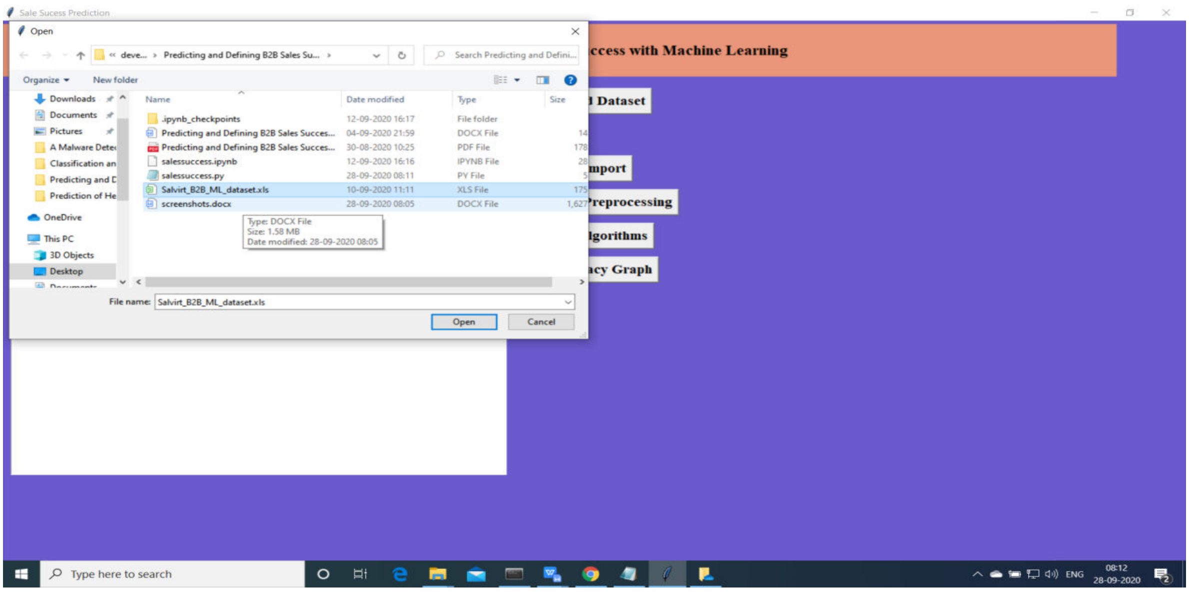


Fig 2:Upload Dataset

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
UGC CARE Listed(Group –I) Journal Volume 12, Iss 1, 2023

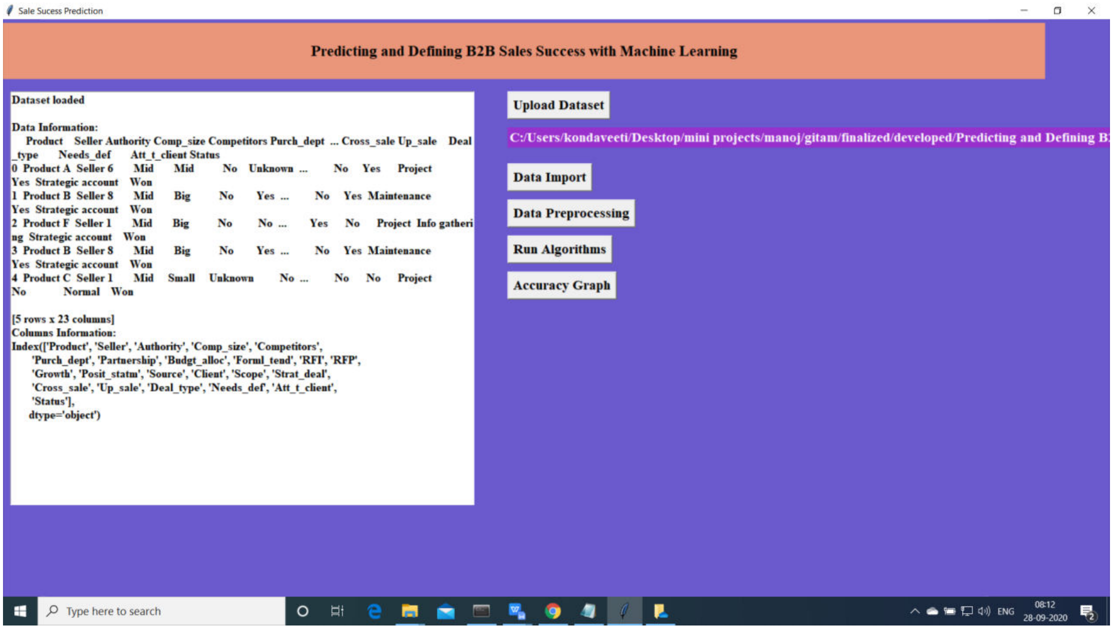


Fig 3: Data Preprocessing

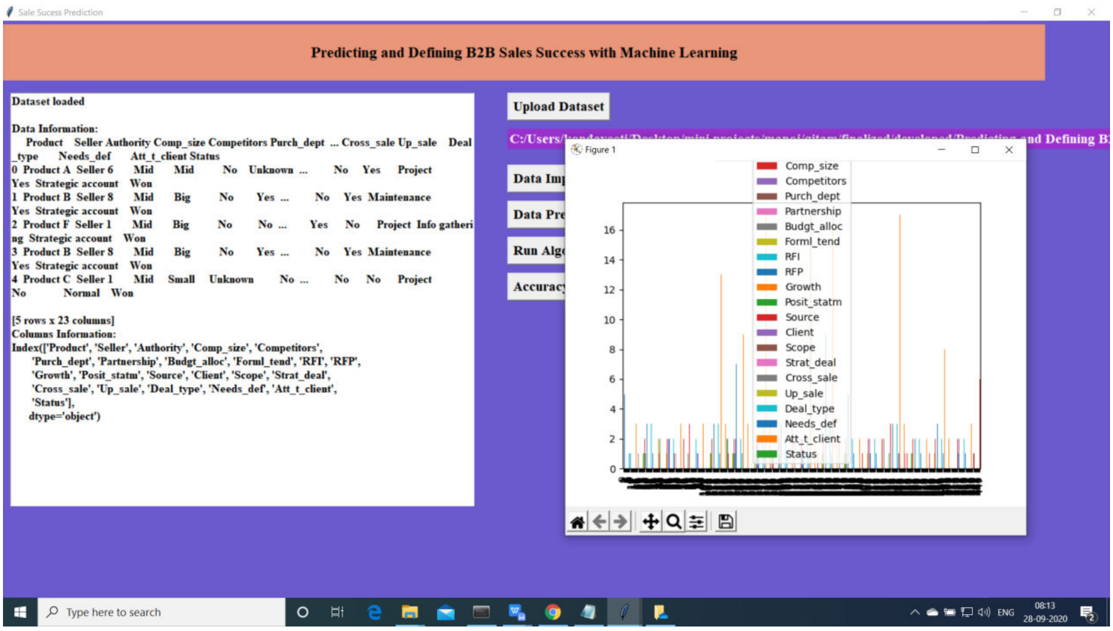


Fig 4:Now click on "Run Algoruimns". Mentioned algorithms will be run on the data

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed(Group -I) Journal Volume 12, Iss 1, 2023

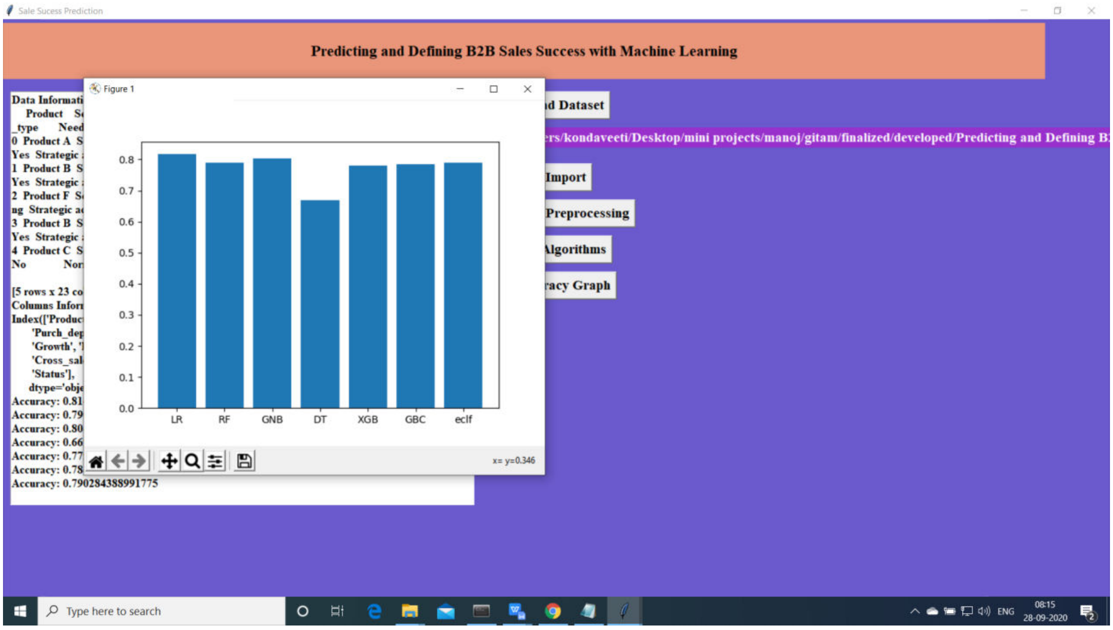


Fig 5: Accuracy Comparison for all the models

5.CONCLUSION

This lookup served as a first step in the improvement of a broader initiative for a Fortune five hundred paper and packaging agency to operationalize predictive modeling on income success. As such, the challenges with any massive employer regularly encompass requiring the constructing of deep nearby expertise of the data, in addition to corralling a massive organisation to aid with correct facts collection. Despite preliminary inconsistencies in the data, typical accuracy regarded promising and indicated in addition enhancements should be made with higher facts exceptional and quantity, greater featurerelated investigation and tuning, or possibly exceptional techniques such as neural nets. The evaluation additionally uncovered new insights into what is vital involving income success. But new insights are frequently accompanied via new questions:

REFERENCES

[1] Implisit (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 all, "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/AAA

I/AAAI15/paper/download/9 444/9488 [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 eConference, June 7-10, 2015. [Online]. Available: https://domino.fov.unimb.si/proceedings.nsf/Proceedings/B12ECF238 1AB59EEC1257E5B004B39B7/$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 [Accessed: Tue. 25 Sept. 2018].