Artificial intelligence-based lead propensity prediction

Aissam Jadli[1], Mustapha Hain[1], Anouar Hasbaoui[2]
[1]AICSE Laboratory, ENSAM Casablanca, University Hasan II, Casablanca, Morocco
[2]Department of Communication and management, Faculty of Science and Techniques, University Hassan II, Mohammadia, Morocco

Article Info

Article history:

Received Feb 3, 2022
Revised Jun 12, 2022
Accepted Jul 16, 2022

Keywords:

Artificial intelligence
Client relationship management
Machine learning
Marketing management
Predictive lead scoring

ABSTRACT

Lead propensity prediction is a data-driven method used to define the value of prospects, by assigning points to them based on their engagement with the business's digital channels, based on multiple key attributes correlating to their attraction to the proposed services or items. The resulting score is closely related to the financial worth of each lead and may be revealing its position in the buying cycle. The marketing teams can then focus on generated leads and prioritize the most prominent ones to improve the conversion rates, using the assigned score on the lead scoring step. The authors investigated using a combination of a data-driven approach and Artificial intelligence (AI) techniques for the lead-scoring process. The experimentation shows that the random forest (RF) is the most suitable model for this task with an accuracy score of 93.04% followed by the decision tree (DT) model of 91.47%. In contrast, when considering the training time, DT and logistic regression (LR) needed a shorter time to learn from the dataset while maintaining decent performances. In contrast, these models represent promising alternatives to the RF model especially in the case of a huge volume of transactions and prospects or in a big data context.

Corresponding Author:

Aissam Jadli
AICSE Laboratory, ENSAM Casablanca, University Hassan II
Casablanca, Morocco
Email: jadliaissam@gmail.com

1. INTRODUCTION

The marketing activities in business-customer interactions may belong to two categories: Lead generation and lead conversion. For many business to consumer (B2C) companies, the collection of prospects' data is undoubtedly crucial to keep their competitive advantage by enhancing their understanding of customers. In the first step, the lead is approached by the business via their marketing channels (e.g., social media, influencers, ads) and encouraged to engage with the business' website through various actions (i.e., newsletter, forms, discounts). these interactions are saved using an internal system and added to their information technology (IT) system as unstructured data.

Secondly, the lead is assigned to a sales agent who will help him take down any hurdles and lead him to make the purchase decision using different techniques such as personalized discounts or payment installments. The efficiency of this process depends largely on the good matching between the agent and the lead's propensity (i.e., the likelihood of the conversion to a customer). Since sales operations are expensive, businesses must focus on the most fitted and engaged prospects to maintain a profitable return on investment (RoI) while exchanging with their visitors. The process of predicting accurately, using data analytics, the weight of each prospect is called lead scoring [1]. The finding of fixed importance to the features is a milestone phase in the scoring strategy. In a classical procedure, the weights associated to the features are deduced using a "try and guess" technique alongside the contribution of an expert marketer to find their optimal weights.

Artificial intelligence (AI) has allowed researchers in several disciplines to optimize different applications, predictive modeling [2] has opened new horizons for lead scoring techniques. Machine learning algorithms can differentiate between cold leads (just interested or passing by) and hot ones (motivated buyers with pre-made purchase determination) by examining common details of existing customers to define a successful profile that will allow to pinpoint them to the marketing department. This experiment compares the performance of different machine learning models in a lead scoring context using a publicly available dataset utilized in lead scoring research.

In the second section, the paper presents a quick overview of the internal working of the lead scoring system of a business to consumer (B2C) company, as well as the recent related works. In the third section, will explain the experimentation details starting with the data description, the machine learning algorithms utilized, and the conducted steps. Section 4 will show and debate the observed results with the review of models' performances using several metrics such as accuracy, precision and F1-score. In the end, the authors come up with the marketing usages of such a system and elaborate on the next continuation to this work.

2. CONTEXT

2.1. B2C lead scoring

Lead scoring is a common commerce process that allows decision-makers to pinpoint the more worthwhile possible buyers among the generated leads. If a potential customer visits a business's website, interacts with a live chatbot, or fills a contact form, these actions should increase their overall lead score because he is revealing a boosted interest. Once the lead reaches a pre-defined threshold, he is handed to the sales department for an individual contact or to take more aggressive sales actions through digital channels. As a result, the salesforce experts will not waste time aimlessly reaching leads and can focus their actions on the most promising opportunities.

Ideally, the business needs to determine matching scores for all potential prospects based on how their features align with the already defined profile of a valuable client. Leads reaching a score higher than a pre-established threshold are regarded as perfect targets. The issue in this system is the determination of profile-pertinent features with their respective importance.

2.2. Classical lead scoring system

Traditionally, a marketing professional or experienced salesforces leader tackles the mission of discovering the most relevant customer characteristics and marks assignation. Where anterior experiences can help select the most pertinent characteristics among several other possible traits. Usually, leads' scores are established on how satisfactorily the prospect resembles the business's target profile (demographic points) and their engagement to the brand (behavioral points).

This system presents several flaws and limitations due to human nature and limited processing power. For instance, the expert cannot consistently determine exactly which characteristics are more pertinent because his judgment can be based on prior prejudices and preconceptions, and then tend to be biased. He also tends to maintain gained experience as ground truth and rarely revise the selection process.

2.3. AI-based lead scoring system

Propensity modelling is a statistical approach that tries to foretell the likelihood of an event [3], which is the foundation for predictive lead scoring. This approach attempts to foretell the probabilities that a website guest to the business website will execute particular activities (e.g., purchase, reservation) on the website. It attempts to combine machine learning and text mining to predict the engagement of the targets and the probability of successful conversion [4].

Machine learning algorithms can glimpse the invisible relationships and patterns in recorded (demographic and behavioral) sales data [5], [6] to choose pertinent characteristics and uncover valuable patterns that reveal a lead's tendency to take the buying decision. The introduced model is trained and assessed to minimize the recall score (i.e., deliver the lowest false-positive predictions). If a mistake/irregularity is detected in a prediction, it can be labeled and re-integrated into the training dataset allowing the model to adapt and update its parameters and weights to remain pertinent, particularly in developing industries where customers' habits change rapidly without prior notice. Figure 1 illustrates the design of an AI-based propensity scoring system.

The main distinction between the two systems is the capability to handle big datasets and collect more beneficial understandings for performance gain [7]. In addition, human limited processing resources can reduce considerably our capacity to find rules and significance in thousands of individual cases and find hidden value in opportunities. AI-based lead propensity prediction allows substituting an experienced marketer (probably costly) with an automated system with equivalent appraisal aptitudes, which will accomplish excellent progress over time with the rise of big data usage and integration [8] in businesses. Read Table 1 for fundamental disparities between lead scoring systems.

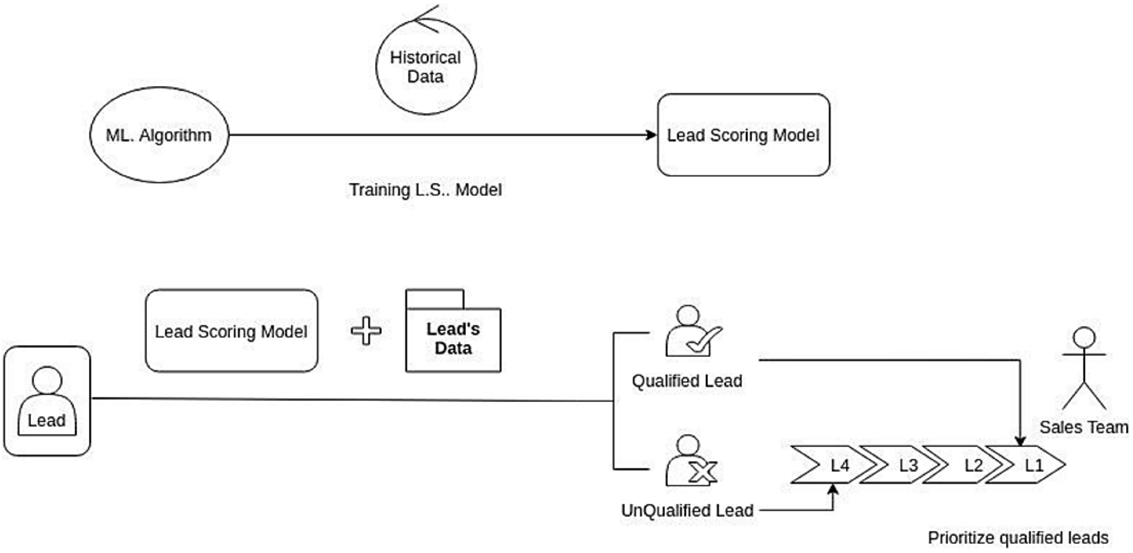

Figure 1. AI-based propensity scoring system

Table 1. Fundamental disparities between lead scoring systems

	Traditional Scoring	Predictive Scoring
Rules	Biased rules selected by experienced marketing professionals	Deduced by algorithms
Supervision	Needs Manual management and periodic adjustments and updates	Minimal maintenance
Data Size	Tiny datasets and restricted processing capacity	All sizes can be used (performance will improve with size increase)
Result	Matching score	Propensity Score

2.4. Related works

The predisposition of prospects to conversion to customers is a prosperous field for research works, due to its outstanding effect on revenue growth and inner marketing optimization, it is was thoughtfully examined by researchers using several machine and deep learning approaches for this purpose. Nygård and Mezei [9] proposed a supervised learning method to score incoming leads, based on algorithms such as logistic regression (LR) and decision trees (DT), to foretell the buying likelihood (utilizing preceding wisdom and behavioral data). The authors discovered that random forest (RF) produces the best performance for this task. Similarly, Singh *et al.* [10] suggested modeling search patterns of retail website visitors, employing supervised machine learning models, to catch and extract the buying habits, and the change in tendencies among this population, while Prasad and Anjaneyulu [11] suggested a comparative breakdown between support vector machine (SVM) algorithm and LR, for constructing models for conversion likelihood, and assessed their results.

Mortensen *et al.* [12] utilized structured and unstructured data from the system of a paper-box manufacturer to foresee B2C deals win. The authors compared different algorithms such as LR, and RF with different variations. The winning model demonstrated a propensity accuracy of 80%, with a recall score of 77% and a precision score of 86%.

Zhang [13] suggested fi the most worthwhile leads via machine learning algorithms. The authors explored the predictive ability of RF and LR and found that the RF model demonstrated a superior accuracy score to the LR model. The latter outperformed RF in the other metrics (e.g., recall, F1, receiver operator characteristic (ROC)).

Benhaddou and Leray [14] proposed an answer in training models with tiny datasets in the context of lead scoring with a Bayesian network. In their article, the authors proposed building the model from expertise, and then applying typical heuristics to decrease the complexness of the model. In addition, they propose three ways of estimating the parameters of their NoisyOr submodels.

Etminan [15] strived to assess the impact of characteristic significances by evaluating some attributes ranking. These ranking are then used in a predictive lead scoring context to evaluate current methods usefulness, leading to optimization opportunities. The author concluded that the combination of classifier and resampling method with the highest average precision score has been selected as the best model.

Yan *et al.* [16] suggested a harmonious, machine learning based framework for e-commerce opportunities propensity analysis. The authors seek answers to various challenges in the business to business (B2B) field (e.g., the disparity in deals volume between B2B and B2C businesses, noisy data, and the high-speed fast-changing demand conditions). The authors are further interested at modeling and forecasting of the customers' purchasing event sequence, such as self/mutual-exciting point process models and the sales resource optimization problem.

Rezazadeh [17] proposed a solution to the issue of predicting the outcome of B2B and B2C deals by suggesting a data-driven, machine learning based pipeline in a cloud environment. The author deduced that decision-making based on machine learning is more precise and carries a higher financial significance than the standard, human-based, strategy. However, the author mentioned that machine learning solutions should not be overwhelmingly used to rule out sensible or justifiable sentiments of salespersons in evaluating a sales opportunity.

Sabbani and Haddadi [18] suggested a new method for seller-buyer matching when visiting a business show event founded on machine learning. The authors proposed an automatic system by substituting the syntactic study of the stakes of the customer with implied user feedback on a frontend smart application (website). They suggest that the integration of this method in the business sales pipeline will allow matching score improvement in B2C match forecasting, resulting in better planning for sellers and buyers and more effective meetings.

Whereas, the added value of this article can be recapitulated as the following:

- The authors tested diverse algorithms (six in total) to affirm that the best model of lead scoring is RF as stated per the literature review.
- The authors utilized several metrics and validation techniques than using only the accuracy metric (as several other works do) to assess the superiority of constructed models.
- The authors raised the execution time and processing power as useful benchmarks for the model choosing to keep durable performance over bigger datasets.

3. METHOD

A general framework for predictive lead scoring is proposed in this paper using predictive modeling and data analytics. In the first step, familiarity with the dataset characteristics is a fundamental phase that concentrates on exploring the dataset and correcting potential issues due to missing or malformatted values. Firstly, there is a data preparation operation that starts with working on missing values, outlier's detection, and forming a pertinent characteristic vector utilizing diverse methods, such as feature selection and extraction. This strategy is essential to enhance the performance of the machine learning model. Similarly, various models are executed and assessed. Aftermath, the performances of each model are examined and assessed.

3.1. Dataset

3.1.1. Dataset description

The principal objective of this article is to illustrate the advantages of machine learning in improving the lead scoring approach using predictive modeling. For instance, the authors experimented with utilizing a publicly available dataset called "X Education" used by researchers in lead propensity prediction. This dataset includes several features covering the following facets,

- The result of the interactions in terms of financial value (Converted or not).
- Unstructured data relating to users' actions on digital channels (e.g., visit count, forms, actions duration).
- The prospect origin (e.g., ads, organic search, referral).

dataset has 9,240 enteries with 37 features for each potential prospect to represent his attributes. These features can be numerical data (e.g., website's visit time and duration and frequency) or categorical (e.g., keywords, origin, and mailbox domain).

3.1.2. Dataset cleaning and feature selection

Using different pre-processing procedures, we extracted 89 features from the bare dataset from initially 37 features. In machine learning tasks, input data can be optimized for better results by applying suitable processing pipelines depending on type and values distribution. For instance, in this experiment,

- Regarding missing values, those with a missing value proportion higher to 70% were excluded from input data due to small values spreading (diminishing variance gain) and the risk of getting a skewed distribution. Likewise, entries with multiple missing features can be excluded for the same reasoning. The remaining features may need distinct processing to replace the missing values with suitable values (i.e., using descriptive statistics)
- For categorical features (e.g., origin, professional status, address), a one hot encoding technique is employed to create a one-hot encoded vector representing an encoded shape of the feature, therefore expanding the available features from 37 to 89.

the resultant dataset has a shape of (9,074.89) with no missing values.

3.2. Techniques

After the data pre-processing phase with several procedures and enhancements. The authors selected six machine learning algorithms widely used in consumer classification tasks and predictive modeling as stated by the literature review. In this section, we present these algorithms and explain their characteristics.

3.2.1. k-nearest neighbors

k-nearest neighbors (kNN) algorithm is a non-parametric, simple yet powerful supervised learning algorithm, introduced by Fix and Hodges in 1951, and later extended by cover [19]. The prediction of a class for a data point in the dataset is achieved by a majority voting of its nearest neighbors selected by the chosen distance function. Various distance functions exist for different use cases and sample distributions. The distance employed in the experimentation was Minkowski's distance: $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$.

3.2.2. Naive Bayes

The Naive Bayes algorithm is a simple yet effective conditional probabilistic classifier [20]. Based on the Bayesian theorem with a Naïve assumption about strong features independence (thus the naive part). Provided a single sample characterized by a vector $x = (x_1, x_2, \ldots, x_n)$ with n pertinent independent characteristics, it gives the probability $\rho(C_k|x_1, \ldots, x_n)$ for each of the K's possible results or category $\mathrm{C_k}$ (i.e., converted or not) [21].

3.2.3. Support vector machine SVM

SVM is a supervised machine learning classifier introduced by Cortes and Vapnik [22]. it tries to locate a hyper-plane belonging to higher N-dimensional features space to differentiates the two classes. It is a usual classifier in various situations because it can deliver significant performances with lower processing needs. Even though SVM is a binary classifier, it can accomplish multi-class prediction by merging multiple executions in a "one-against-all" system.

3.2.4. Decision tree (DT) and random forest (RF)

Decision tree (DT) is a supervised machine learning algorithm used for both regression and classification situations. The algorithm employs training data to create several rules that can be represented by a graph or a tree structure. DT classifier begins by splitting the input dataset into smaller subsets and the classification rule in each level is learned to create incrementally a DT. Each node at each level is tagged with a pertinent feature (based on different metrics) and a probability over the classes.

The RF illustrates a machine learning approach called integrated learning, which attempts to resolve a single prediction task by constructing multiple models for the same problem and then using a majority vote to select the best prediction. These predictions are merged into a global prediction which will outperform naturally most classic classification models. For instance, an RF model is a mixture of multiple DT models and its result is defined by the voting of each individual DT model output.

3.2.5. Logistic regression (LR)

Logistic regression (LR but also known as logit) is a straightforward yet practical binary classifier used for predictive analytics. It is widely utilized in practice and in research works to estimate the probability of an event. It maps an independent input features vector to a value ranging between 0 and 1 using a sigmoid function. LR is characterized by its training speed faster than other algorithms while preserving similar performances.

3.3. Model evaluation

3.3.1. First level indicators

To evaluate objectively the performance of a machine learning model, different indicators can be utilized to assess model's generalization ability. By splitting the dataset into a training set and a testing set, and making predictions on the testing set, we get four fundamental indicators named first-level indicators. Using these indicators, we can create a table named the confusion matrix (CM). Table 2 shows the layout of a confusion matrix.

Table 2. The structure of the confusion matrix

Confusion Matrix		Actual Value	
		Positive	Negative
Model's Prediction	TP	FP	FP
	FN	TN	TN

Confusion Matrix is a tabular visualization of the ground-truth labels versus model forecasts. Each row of the confusion matrix symbolizes the samples in a forecasted class and each column designates the instances in an actual class. The first-level indicators provide a clear vision regarding the model's performance and are straightforward to analyze. Generally, a good model needs to improve its true positive (TP) and true negative (TN) scores and minimize its false negative (FN) and false positive (FP) scores.

3.3.2. Second level indicators

For more objective and accurate benchmarking for selected models, researchers defined additional refined metrics founded on the earlier indicators to interpret trained models' objectively depending on specific needs. These indicators are named secondary indicators and are computed on the first-level indicators. The secondary indicators widely exploited are,

- Precision: The ratio of correct positive predictions to all positive predictions.
- Accuracy: The ratio of correct predictions to the entire predictions.
- Sensitivity or Recall: The ratio of correct positive predictions to all positive predictions.
- Specificity: The ratio of correct positive predictions to all negative predictions.
- F1-Score: The f1-score is a prevalent indicator commonly used when the dataset is unbalanced (the number of samples belonging to a category is exceptionally greater than in the other categories).

3.3.3. K-folds cross-validation

Cross-validation (CV) is a widespread and usefull statistical technique exploited to assess the generalization ability of a machine learning model to make correct predictions. The method is based on resampling the total dataset after dividing it into multiple subsets of the same length (k) and assigning each time a set of them as training data and the remaining subsets as testing data. K-fold cross-validation produces a less prejudiced model (reducing biais) because it permits every instance to play a part in the construction of the model enhancing tthus he overall variance.

4. RESULTS AND DISCUSSION

Following data preprocessing and cleaning, the prepared dataset was split into a training set and a validation set. After a comprehensive investigation of possible alternatives, the authors selected the proportion training/validation of 80/20 due to its good performance as stated in similar research works [23]–[25]. Furthermore, the authors used a cross-validation approach to assess the model's ability to generalize over the dataset without introducing bias. This experience was executed using the Python Programming Language using open-source machine learning libraries (i.e., pandas, matplotlib, NumPy). The goal of this experiment is to predict the probability of conversion from a business visitor to a customer based on the demographic and behavioral data of the users, allowing the business to identify worthwhile prospects from a huge volume of website visitors. The experimental results are presented in Table 3.

When analyzing the results of constructed models, it is evident that the RF model is performing exceptionally well and surpassing the other models, attaining an accuracy score of 93.04%, seconded by the DT and the LR models with 91.49% and 89.90% respectively. While SVM and NB, due to their straightforward approach, reached respectively 80.12% and 73.20%. Figure 2 illustrates the confusion matrix of the best four constructed models.

In a propensity prediction task, the recall score is very interesting because it can intuitively denote the capacity of the model to find false positive (those labeled as negative by the model while being positive). From the outcomes of the experiment, the RF model had the most satisfactory recall score (89.05%), followed tightly by the DT model (87.24%) and SVM (85.23%).

Table 3. The experiment results

Metric	Models					
	NB	RF	KNN	DT	LR	SVM
TP	1,467	1,651	1,471	1,610	1,622	1,477
TN	712	877	693	872	806	537
FP	265	83	261	115	110	255
FN	275	110	251	122	181	450
Training Time (s)	0.02	0.92	0.11	0.06	0.26	8.98
Accuracy (%)	80.12	93.04	80.06	91.49	89.90	73.20
Precision (%)	75.12	92.24	75.55	90.36	90.29	69.26
Recall (%)	74.34	90.06	73.15	87823	84.33	55.03
Specificity (%)	83.63	94.20	83.60	92.24	92.52	86.20
F1-Score (%)	88.90	85.64	88.98	72.69	72.46	60.33

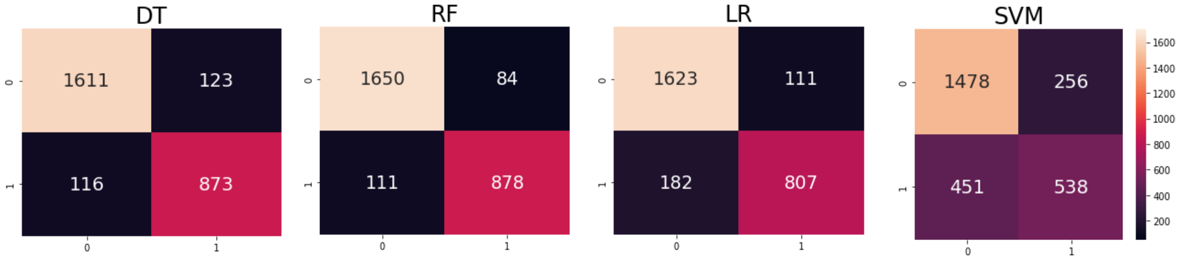

Figure 2. The confusion matrix of the four best models

The A useful metric when predicting the likelihood of a binary output is the receiver operating characteristic (ROC) curve. It plots the false positive rate or (FPR) on the x-axis versus the true positive rate (TPR) on the y-axis for some different candidate threshold values ranging in [0, 1]. In other terms, it shows the false alarm rate versus the hit rate. The true positive rate is calculated as TP divided by the sum of TP and FP. It describes how accurate the model is at predicting the positive class when the actual result is positive. When used in this experiment, the RF model got the best ROC score (0.92) while SVM came last (0.69), which confirms that RF is the most suitable model for this task, shows in Figure 3.

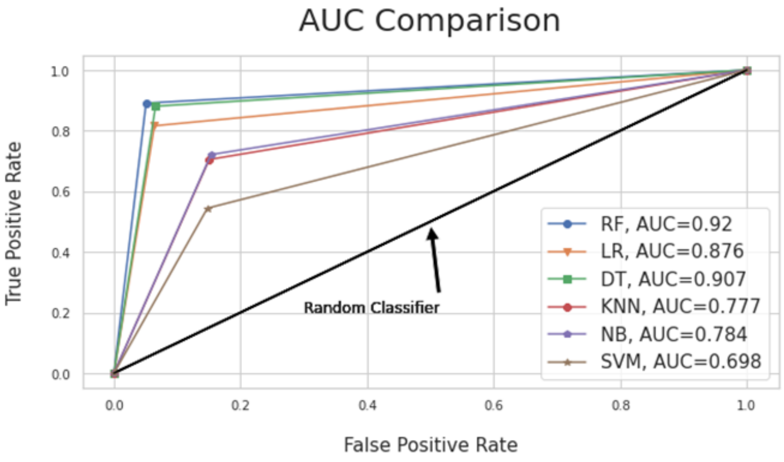

Figure 3. AUC comparison

Training time is a useful criterion when selecting a machine learning algorithm. In this experience, due to its simple concept, the NB model took only 0.02s to build, seconded by the DT model (0.06s) followed by the LR model (0.26), while the RF model took a considerable extra time (0.92s) for training. This disparity in learning duration between chosen models can influence the choice for processing power limitations if applying this approach on very large datasets or in big data context. The SVM model was the slowest model to train (over 8s) presumably because of the non-linearity of the kernel utilized in the default implementation of this model in SKLearn library. Eventually, the RF model had the most satisfactory results accuracy-wise, whereas the DT model demonstrated good performance while holding down a smaller training duration. Shows in Figure 4.

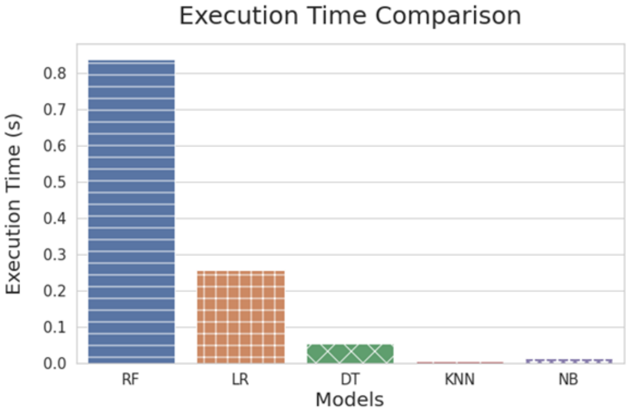

Figure 4. Comparison of execution time

K-fold cross validation (CV) has been performed on the dataset to evaluate the model's generalization capability over the entire dataset. By changing the number of subsets (k), we can experiment the impact of the proportion train/test on the prevailing bias of the model. Table 4 shows the outcome of CV.

Table 4. K-folds cross-validation results

Cross-Validation		k-value			
		4	6	8	10
Accuracy (%)	DT	91.71	90.52	90.65	91.54
	RF	92.86	91.91	92.97	93.06
	KNN	79.85	80.39	80.22	80.36
	NB	78.10	77.07	79.18	80.21
	SVM	73.25	74.64	72.12	74.10
	LR	90.40	90.01	90.15	89.89

The results of the k-fold CV were compatible with the expected results seen earlier in the experiment. Even when changing the k-value, the RF model had an evident edge over the other ones. Yet, when we raise the training duration as a decisive measure, DT can be a better efficient solution to exploit than RF.

5. CONCLUSION

This paper investigated the advantages of using machine learning algorithms to optimize the scoring of prospects and substitute the classic opportunity identification system. Several machine learning techniques were employed to categorize possible prospects into a winning opportunity or not. The result of the experiment demonstrated that RF and DT have both achieved satisfactory results as they can predict the lead's propensity with an accuracy score of 93.04% and 91.49% respectively, and a precision score of 92.24% and 90.36%. This prediction of the users having a buying decision is based on demographic, chronological, and unstructured data recorded on the company's website. DT and LR models were particularly faster to train and can be used in environments where available resources are limited or when dealing with a huge volume of data to process. Allowing firms to pinpoint high-return opportunities from business's digital channels and enhance their RoI. A possible issue with the proposed solution is the handling of imbalanced dataset, which is a common problem with machine learning solutions. An imbalanced dataset is defined as a dataset having significantly more instances of a certain class compared to other classes. In our context, this would translate to a dataset that has more lost sales record than won (which is the common case). To overcome this issue, we suggest either over-sampling the smaller class or under-sampling the larger one. There were no comparisons between automatic machine learning and manual lead scoring in this study, so it is accurate to say which one is better with complete confidence. However, we have shown that machine learning based lead-scoring models offer a conceivable option. Some areas of probable prospective research work would be to add client business value to lead scoring, resulting in a financial value that may seem more tangible than a simple purchase propensity. For example, one could multiply customer lifetime value by the purchase probability. Another example would be to use regression instead of classification to estimate the customer lifetime value of leads. In addition, identifying different lead types would be beneficial for companies. That way, they could treat the different types of leads with different types of marketing material, for example through nurturing campaigns. This could be done using unsupervised learning since it is unknown how many different types of leads there are. Finally, various steps in the machine learning-model building could be further optimized and finally explore the use of deep learning to solve the issue of insufficient or inadequate training data or imbalanced dataset classes, using approaches such as transfer learning and reinforcement learning.

REFERENCES

[1] E. Brynjolfsson and K. McElheran, “The rapid adoption of data-driven decision-making,” *American Economic Review*, vol. 106, no. 5, pp. 133–139, 2016, doi: 10.1257/aer.p20161016.

[2] G. Shmueli and O. R. Koppius, “Predictive analytics in information systems research,” *MIS Quarterly: Management Information Systems*, vol. 35, no. 3, pp. 553–572, 2011, doi: 10.2307/23042796.

[3] Ö. Artun and D. Levin, “Predictive marketing,” *Predictive Marketing*, 2015, doi: 10.1002/9781119175803.

[4] J. Järvinen and H. Taiminen, “Harnessing marketing automation for B2B content marketing,” *Industrial Marketing Management*, vol. 54, pp. 164–175, 2016, doi: 10.1016/j.indmarman.2015.07.002.

[5] W. K. Lin, S. J. Lin, and T. N. Yang, “Integrated business prestige and artificial intelligence for corporate decision making in dynamic environments,” *Cybernetics and Systems*, vol. 48, no. 4, pp. 303–324, 2017, doi: 10.1080/01969722.2017.1284533.

[6] C. L. Pan, X. Bai, F. Li, D. Zhang, H. Chen, and Q. Lai, “How business intelligence enables E-commerce: Breaking the traditional E-commerce mode and driving the transformation of digital economy,” *Proceedings - 2nd International Conference on E-Commerce and Internet Technology, ECIT 2021*, pp. 26–30, 2021, doi: 10.1109/ECIT52743.2021.00013.

[7] A. Algi and Irwansyah, “Consumer trust and intention to buy in Indonesia instagram stores,” *Proceedings - 2018 3rd International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2018*, pp. 199–203, 2018, doi: 10.1109/ICITISEE.2018.8721033.

[8] M. B. Adam, “Improving complex sale cycles and performance by using machine learning and predictive analytics to understand the customer journey,” 2018, [Online]. Available: https://dspace.mit.edu/handle/1721.1/118010.

[9] R. Nygård and J. Mezei, “Automating lead scoring with machine learning: An experimental study,” *Proceedings of the Annual Hawaii International Conference on System Sciences*, vol. 2020-January, pp. 1439–1448, 2020, doi: 10.24251/hicss.2020.177.

[10] S. Singh, S. Madhwal, G. Datta, and L. Singh, “Modelling search habits on E-commerce websites using supervised learning,” *Proceedings of the 8th International Advance Computing Conference, IACC 2018*, pp. 53–58, 2018, doi: 10.1109/IADCC.2018.8692113.

[11] K. V. N. K. Prasad and G. V. S. R. Anjaneyulu, “A comparative analysis of support vector machines and logistic regression for propensity based response modeling,” *International Journal of Business Analytics and Intelligence*, vol. 3, no. 1, 2015, doi: 10.21863/ijbai/2015.3.1.002.

[12] S. Mortensen, M. Christison, B. C. Li, A. L. Zhu, and R. Venkatesan, “Predicting and defining B2B sales success with machine learning,” *2019 Systems and Information Engineering Design Symposium, SIEDS 2019*, 2019, doi: 10.1109/SIEDS.2019.8735638.

[13] Y. Zhang, “Prediction of customer propensity based on machine learning,” *Proceedings - 2021 Asia-Pacific Conference on Communications Technology and Computer Science, ACCTCS 2021*, pp. 5–9, 2021, doi: 10.1109/ACCTCS52002.2021.00009.

[14] Y. Benhaddou and P. Leray, “Customer relationship management and small data - application of Bayesian network elicitation techniques for building a lead scoring model,” *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, vol. 2017-Octob, pp. 251–255, 2018, doi: 10.1109/AICCSA.2017.51.

[15] A. Etminan, “Prediction of lead conversion with imbalanced data: A method based on predictive lead scoring,” 2021, [Online]. Available: www.liu.se.

[16] J. Yan, M. Gong, C. Sun, J. Huang, and S. M. Chu, “Sales pipeline win propensity prediction: A regression approach,” *Proceedings of the 2015 IFIP/IEEE International Symposium on Integrated Network Management, IM 2015*, pp. 854–857, 2015, doi: 10.1109/INM.2015.7140393.

[17] A. Rezazadeh, “A generalized flow for B2B sales predictive modeling: An Azure machine-learning approach,” *Forecasting*, vol. 2, no. 3, pp. 267–283, 2020, doi: 10.3390/forecast2030015.

[18] A. Sabbani and A. El Haddadi, “Business matching for event management and marketing in mass based on predictive algorithms,” *Proceedings - 15th International Conference on Signal Image Technology and Internet Based Systems, SISITS 2019*, pp. 619–626, 2019, doi: 10.1109/SITIS.2019.00102.

[19] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” *American Statistician*, vol. 46, no. 3, pp. 175–185, 1992, doi: 10.1080/00031305.1992.10475879.

[20] A. Jadli, M. Hain, and A. Jaize, “A novel approach to data augmentation for document image classification using deep convolutional generative adversarial networks,” *Lecture Notes in Networks and Systems*, vol. 211 LNNS, pp. 135–144, 2021, doi: 10.1007/978-3-030-73882-2_13.

[21] A. Jadli, M. Hain, A. Chergui, and A. Jaize, “DCGAN-based data augmentation for document classification,” *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science, ICECOCS 2020*, 2020, doi: 10.1109/ICECOCS50124.2020.9314379.

[22] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/bf00994018.

[23] M. Karim and R. M. Rahman, “Decision tree and Naïve Bayes algorithm for classification and generation of actionable knowledge for direct marketing,” *Journal of Software Engineering and Applications*, vol. 06, no. 04, pp. 196–206, 2013, doi: 10.4236/jsea.2013.64025.

[24] J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, “A critical look at the current train/test split in machine learning,” 2021, [Online]. Available: http://arxiv.org/abs/2106.04525.

[25] B. Vrigazova, “The proportion for splitting data into training and test set for the bootstrap in classification problems,” *Business Systems Research*, vol. 12, no. 1, pp. 228–242, 2021, doi: 10.2478/bsrj-2021-0015.

BIOGRAPHIES OF AUTHORS

Aissam Jadli Aissam Jadli is a Ph. D student at Ensam Casablanca, university Hassan II. He graduated from the Faculty of Science of Agadir with a bachelor in Computer Science. In 2018, He graduated with a Master at Big Data and Internet of Things from ENSAM Casablanca. His researches are in the fields of ERP, computer vision, machine and deep learning. He is an active contributor in different scientific journals and conferences as an invited Board Member or guest reviewer. Besides, he is also involved in NGOs and student associations. He can be contacted at email: aissam.jadli-etu@etu.univh2c.ma.

Mustapha Hain Mustapha Hain is a Ph.D. researcher and professor at ENSAM Casablanca from University Hassan II and the Director of the AICSE Laboratory (Artificial Intelligence & Complex Systems Engineering) in University Hassan II at ENSAM Casablanca. He is also a visitor academic researcher at Arts et Métiers ParisTech. His researches are in the fields of digital systems, E-logistics, software modeling and supply chain management. He is an active contributor in different scientific journals and conferences as an invited Board Member or guest reviewer. He can be contacted at email: infohain@gmail.com.

Anouar Hasbaoui Anouar Hasbaoui is an Economist and a Doctor of Philosophy in Finance working actually with the Moroccan Government. He got his Master of Public Administration (MPA) at University of Saint Mary's of San Antonio Texas in 2005 and obtained his PhD in Hassan 1st University of Settat in 2016. His researches are in fields such as banking and support services, Risk management, and stock prediction. He can be contacted at email: anhasbaoui@yahoo.fr.