DOI: 10.21917/ijsc.2023.0427

AUTOMOBILE BONUS-MALUS SYSTEM MODELLING USING MACHINE LEARNING ALGORITHMS

Kouach Yassine, EL Attar Abderrahim and EL Hachloufi Mostafa
Department of Statistics and Mathematics Applied to Economics and Management, University Hassan II, Casablanca, Morocco

Abstract
Third-party liability is the most important sub-category in automobile insurance, that is why actuaries seek always to design the ideal price list by classifying the insureds into homogeneous classes. However, the heterogeneity persists in the priori tariffication. For that, actuaries use the Bonus Malus system to redistribute the cost of claims more equitably between insureds by rewarding good insureds with a bonus and penalizing bad insureds with a Malus. Nevertheless, the classical approach used in the conception of Bonus Malus systems is limited to the parametric methods that need to make the hypothesis of the number of claims distribution and don't consider the cost of claims. In this direction, this paper seeks to avoid this issue by using machine learning algorithms, in response to offering a fair Bonus Malus System. Two models of posteriori tariffication will be built. In addition, three algorithms will be used, in occurrence, the CART Classification And Regression Tree method, SVM the Support Vector Machine for regression, and KNN the K-Nearest Neighbor. The suggested models take into account, not only the number of claims but also the importance of the cost of claims. A numerical illustration shows the flexibility of posteriori premiums calculated by our models in relation to the risk levels. This work is a start for new actuarial research which seeks to use artificial intelligence in the design of bonus malus systems.

Keywords:
Bonus-Malus System, Machine Learning, Posteriori Tariffication, Priori Tariffication, Automobile Insurance

1. INTRODUCTION

Actuaries always seek to offer premiums that reflect the real risk and fairly distribute losses among policyholders in the auto insurance portfolio. For that, they use a priori tariffication model. This method consists of using a priori explanatory variables (Gender, age, type of vehicle, engine power, …) in order to segment the insureds into homogeneous risk classes where drivers of the same risk class pay the same premium. However, the risk classes are still heterogeneous due to other posteriori variables like the driving behavior which has not been retained in the priori pricing and has an impact on the insured claim frequency.

Some premiums are found not fair for some good drivers. With a view to adjusting the insured's annual premium, the insurers use a posterior tariffication model. The famous posterior tariffication model used is Bonus-Malus System in third-party liability automobile insurance. This system calculates the posteriori premium by considering the past claims experience. Indeed, the Bonus-Malus System takes into account the number of claims and the amounts of claims during the last period in view of the potential of past claims to better predict future claims. However, most experience rating systems use only the number of claims of a policyholder to determine its premium [1] since the number of claims and the amount of claims are often assumed to be independent [2].

The objective to introduce a Bonus-Malus System in the insurance industry is to recompense the good insured (claim-free policyholders) and to penalize the bad drivers responsible for one or more accidents. In fact, the bad driver is penalized by a Malus, a surplus of the premium, and the good driver is recompensed by a Bonus, a discount of the premium. The Bonus-Malus System aims to encourage insureds to drive carefully and to better assess individual risks [3].

Bonus-Malus Systems (BMSs) were introduced in Europe in the early 1960s [2]. This system is called merit-rating or no-claim discount system in other countries. Since July 6, 2006, in Morocco, the Bonus-Malus System has been replaced by a new system called Coefficient Reduction Increase. This system has the same logic as Bonus-Malus System by awarding a reduction on the insurance premium to good drivers and increasing the premiums of bad drivers, but the system applies a multiplier coefficient (from 90% to 250%) instead of transferring the insured between predefined levels. In fact, each two years free claims are recompensed by 10% reduction in the third liability auto basic premium, and for one or more accidents, engaging or likely to engage the insured liability totally or partially during one year, penalized by 20% increase in the premium in case of the material damage and 30% for each bodily accident without however exceed 250% of the basic premium.

Actuaries always try to offer a fair Bonus-Malus system. For this, actuaries and researchers study how to improve this system but in general, all their works relate to the analysis of the hypotheses on the probability distributions of the number of claims. The difference in the assumptions can make some difference in the posterior premium. In addition, the classical approach of Bonus-Malus does not take into consideration the cost of claims and the risk profiles determined at the priori tariffication.

This article aims to address a new conception of Bonus-Malus System or Coefficient Reduction Increase system in Morocco. In this regard, we propose to use new tools to calculate the posterior premium, in the occurrence of the Machine Learning algorithms. This developed approach makes it possible to estimate the posterior premiums of the Bonus-Malus System directly using Machine Learning models by using the CART Classification And Regression Tree method, SVM the Support Vector Machine for regression, and KNN the K-Nearest Neighbor. Furthermore, we try to take into consideration the amounts of claims and the priori risks classification. Our goal is to offer a fair posterior premium that encourages drivers to be cautious on the road and protect the insurer's customer portfolio from escaping to adversary companies in a more competitive environment.

2. RELATED WORKS

On the side of the insured, it is not clear enough how the insurance company allocates bonus-malus according to their history of claims. In this perspective, professionals in the field and researchers are trying to improve the Bonus-Malus System in order to better reflect the history of claims and to make it more clarified to policyholders. However, the main problem that researchers are trying to solve in previous work is the assumptions about the probability distributions for the number of claims.

Among others, Dionne & Vanasse [4] develop a methodology that integrates a priori and a posteriori information into a single statistical model. In order to estimate the accident distribution, they introduce Poisson and negative binomial models. They conclude that the negative binomial model with a regression component reasonably approximates the true accident distribution [4]. Morillo & Bermúdez [5] observe that the maluses are very high, and they are higher when the risks have a low frequency. To solve this problem, they propose a Poisson–Inverse Gaussian model using an exponential loss function.

Pierre-Loti-Viaud [6] proposes to enlarge the model of mixed Poisson distributions by considering mixed negative binomial distributions to adjust the observed correlation between the successive numbers of claims. Pitrebois et al. [7] work to design bonus-malus scales involving different types of claims assuming a multinomial partitioning scheme, to avoid using multivariate Beta distributions [7]. Guerreiro et al. [8] introduce an alternative approach that considers the automobile insurance portfolios as open portfolios by considering that the number of new policies entering the portfolio follows the Poisson distribution. Inoussa [9] proposes a statistical approach to Bonus-Malus Systems in order to directly estimate bonus-malus relativities using statistical tools, regression models, via the maximum likelihood method.

Ni, Constantinescu, et al. [10] employ the Bayesian approach to derive the Bonus Malus premium rate for Weibull claim severities distribution by modelling the number of claims as a Negative Binomial distribution. In another work, Ni, Li, et al. [11] focus on the modelling of claim severities using a hybrid structure, Weibull distribution is suggested for constructing smaller-sized claims, and Pareto distribution is applied to model large ones [11]. Gómez-Déniz [12] presents a statistical model which distinguishes between different types of claims, incorporating a bivariate distribution based on the assumption of dependence.

Tzougas et al. [13] employ two-component mixture distributions for the construction of Optimal Bonus-Malus Systems with frequency and severity components. In the last work quoted in this section, Oh et al. [14] propose a Bonus-Malus System that incorporates the association between frequency and severity into the posterior risk adjustment. They use a copula-based bivariate random effects model to accommodate the dependence between claim frequency and severity [14].

The issue of the classical approach used in some studies cited above is the different assumptions of the number of claims distributions that make the number of claims inadequate for the driver. Consequently, the posterior premium in its turn is inadequate for some insurers. For some studies, they did not take into account the cost of claims, as a result, the driver who makes an accident with a small cost of claims will be penalized by the same malus as the driver who has declared an important cost of claims because they have declared the same number of claims. In addition, there is no relationship between the prior tariffication and the posteriori tariffication so the Bonus-Malus System does not take into account the risk profiles determined at the first tariffication.

For all these reasons, this paper suggests building a new Bonus-Malus System based on Machine Learning algorithms. In addition, this work will introduce new techniques from artificial intelligence to the literature on the Bonus-Malus system.

3. METHODOLOGY

In a Bonus-Malus System, the insured move each year in an increasing class scale according to their loss ratio during the previous year. A year marked by no claim allows the insured to go down to a lower level (Bonus) and a year marked by claims punished the insured to climb to the next level (Malus). Consequently, the insured pays a premium according to her level occupied in the Bonus-Malus System and based on the pre-determined premium at the a priori pricing.

So, in order to develop our Bonus-Malus model, we first proceed to present the a priori tariffication to determine the premiums to be paid on subscription by the new drivers. Next, we outline the principle of the classical approach, the a posteriori tariffication. Then, we develop our approach proposed in this article, and finally, we give a numerical illustration to better understand our method.

3.1 PRIORI TARIFFICATION

Through a priori tariffication, actuaries seek to segment policyholders into homogeneous risk classes so that policyholders of the same class pay an identical pure premium. The calculated pure premium is established according to the explanatory variables known as a priori. Indeed, insurers develop models that predict the number of claims (N) and the cost of claims (X) based on risk classification variables such as gender, type of fuel, engine power, etc. The pure premium determined at a priori tariffication is the multiplication of the frequency of claims by the average cost of claims according to frequency-cost model.

$$\textit{prime pure} = E[N]*E[X] \tag{1}$$

In order to model this pure premium and determine the relationship between the number of claims or the cost of claims and the a priori variables, actuaries tend to use generalized linear models (GLM). Generalized linear models associate the response variable with a probability law which must belong to the exponential family as in Eq.(2):

$$f\left(y_i,\theta_i,\varphi,\omega_i\right)=\exp\left(\frac{y_i\theta_i-b\left(\theta_i\right)}{a\left(\varphi\right)}\omega_i\right)+c\left(y_i,\varphi,\omega_i\right) \tag{2}$$

GLM related the deterministic component $\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_kX_k$ and the random component by a link function that describes the relationship between the linear combination and the mathematical expectation μ of the response variable.

$$g(\mu) = \beta_0+\beta_1X_1+\beta_2X_2+ \cdots+\beta_kX_k \tag{1}$$

Most of the time, actuaries model the number of claims either by the Poisson distribution, the Binomial distribution, or the

Negative Binomial distribution. Thus, they model the cost of claims either by the Gamma, Log Normal, or Inverse-Gaussian distribution.

For instance, the density function of Poisson law (Eq.(4)) and Gamma law (Eq.(5)):

$$p(y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad (4)$$

$$f(y) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^{k}} \quad (2)$$

3.2 POSTERIORI TARIFFICATION: CLASSICAL APPROACH

The posteriori premium that the insured will pay according to the last claim experience is equal to the priori premium multiplied by the relativity level r_l related to the level l occupied in the Bonus-Malus System.

$$\textit{posteriori premium} = \textit{priori premium} * r_l \quad (6)$$

The insured changes his position each year in accordance with system transition rules in addition to the probability matrix of transition. In fact, before estimating the relativities it is necessary first to determine the probabilities matrix of transition. For instance, this matrix gives the probability that an insured in level 1 l_1 will be in level 2 l_2 after a given period with annual claim frequency (ϑ).

The insured changes his position each year in accordance with system transition rules in addition to the probability matrix of transition. In fact, before estimating the relativities it is necessary first to determine the probabilities matrix of transition. For instance, this matrix gives the probability that an insured in level 1 l_1 will be in level 2 l_2 after a given period with annual claim frequency (ϑ).

$$pl_1l_2(\vartheta)=\Pr[L_{t+1}(\vartheta)=l_2|L_t(\vartheta)=l_1] \quad l_1,l_2\in\{1,2,\ldots,s\} \quad (7)$$

$$= \sum_{n=0}^{\infty}\Pr\left[L_{t+1}(\vartheta)=l_2 \mid N_{t+1}=n, L_t(\vartheta)=l_1\right]\Pr\left[N_{t+1}=n \mid L_t(\vartheta)=l_1\right] \quad (3)$$

where N_{t+1} is the number of claims in t+1 and s is the maximal level.

The probabilities matrix of transition $P(\vartheta)$ is a Markovian matrix because the movement in levels only depends on the previous level and the number of claims from the previous period.

$$P(\vartheta) = \begin{pmatrix} p_{11} & \cdots & p_{1s} \\ \vdots & \ddots & \vdots \\ p_{s1} & \cdots & p_{ss} \end{pmatrix} \quad (4)$$

The transition rules are also represented in the form of a matrix $T(k)$ with values 0 or 1. $t_{ij}(k)$=1 means the insured transferred from level 1 to level 2.

$$T(k) = \begin{pmatrix} t_{11}(k) & \cdots & t_{1s}(k) \\ \vdots & \ddots & \vdots \\ t_{s1}(k) & \cdots & t_{ss}(k) \end{pmatrix} \quad (5)$$

Once the number of levels and transition rules of the Bonus-Malus System are defined, the relativity level is determined by minimizing the mean square difference between the unknown relativity Θ and the relativity level r_l related to the level l_as proposed by [15].

$$E\left[(\Theta - r_l)^2\right] = \sum_{l=0}^{s} E\left[(\Theta - r_l)^2 \mid L=1\right]\Pr\left[L=1\right] \quad (6)$$

$$= \sum_k w_k \int_0^{\infty} \sum_{l=0}^{s} (\Theta - r_l)^2 \, \pi_l(\lambda_k\theta)\, f_\theta(\theta)\, d\theta \quad (7)$$

where Pr[L=l] is the proportion of the policyholders at level l in Bonus-Malus System. λ_θ is the frequency of claims.

The solution that minimizes the equation:

$$r_l = E\left[\Theta \mid L=l\right] = \frac{\sum_k w_k \int_0^{\infty} \theta \pi_l(\lambda_k\theta)\, f_\theta(\theta)\, d\theta}{\sum_k w_k \int_0^{\infty} \pi_l(\lambda_k\theta)\, f_\theta(\theta)\, d\theta} \quad (8)$$

So, the posteriori premium will be the product of the base premium and the relativity level r_l.

3.3 POSTERIORI TARIFFICATION: MACHINE LEARNING MODEL

First, we will calculate the priori tariffication to build price classes. Then we will proceed to the posteriori pricing. We will build two models of the Bonus-Malus system; the first model consists in modeling the cost of claims according to the number of claims during the previous period, the class of risk, and the importance of risk, if it is a claim with a low amount or a very large amount. In the second model, the cost of claims is modeled based only on the number of claims and the risk class.

In both models, we multiply the estimated value of the cost of claims by the frequency of claims of the associated risk class extracted from the a priori pricing. The a priori frequency of claims is used to relativize the posteriori premiums to the risks already determined. Therefore, the calculation of the posteriori pure premium in the two models is expressed as follows:

$$pp_{post,t+1} = f_k * \textit{cost of claims}_t \quad (14)$$

In the case of 1st model:

$$pp_{post,t+1} = f_k * E(X|C=k, N=n, W=w),$$
$$k=\{1,\ldots,K\},\ n=\{0,\ldots,N\},\ w=\{H,L\} \quad (15)$$

In the case of 2nd model:

$$pp_{post,t+1} = f_k * E(X|C=k, N=n),$$
$$k=\{1,\ldots,K\},\ n=\{0,\ldots,N\} \quad (16)$$

where f_k is the frequency of claims for rate class k, n is the number of claims declared in year t, N is the maximum number of claims declared and W is the weight of the declared amount takes two indices, i.e. H: high or L: low.

By using the first model we avoid the problem of having an insured declaring a number of claims with a low amount pays an identical increase coefficient to one who declared the same number of claims with a very high amount.

The Increase Reduction Coefficient will not be a fixed coefficient but a variable coefficient depending on the number of claims, the risk class, and the extent of the claim. The coefficient will be calculated as follows:

$$CRI = \left(\frac{pp_{post,t+1} - pp_{priori}}{pp_{priori}} + 1 \right) * 100 \quad (17)$$

In order to compare the two models, we use the notion of financial equilibrium [16]. A bonus-malus system is said to have the property of financial equilibrium when the collected premiums remain sTable.over time. In other words, when there is an equivalence between the total of the premiums a priori and the total of the premiums a posteriori knowing that the average amount of total claims is not modified [16].

$$\sum_{i=0}^{N_i} pp_{post,t+1,i} = \sum_{i=0}^{N_i} pp_{priori,i} \quad (9)$$

where N_i is the total number of insureds.

3.4 MACHINE LEARNING ALGORITHMS

Machine learning is a branch of artificial intelligence that seeks to learn from data in order to predict or classify outputs. Machine Learning algorithms can be classified into three types depending on whether the data are labelled or not: supervised learning, unsupervised learning, and reinforcement learning.

The Machine Learning algorithms used in this article are K-means, KNN, SVR, and decision trees. The first algorithm is used for the classification of claim amounts into high and low amounts. However, other algorithms are used for modelling claims costs. The performance of each machine learning algorithm used for modelling claims costs will be evaluated by Root Mean Squared Error (Eq.(19)).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \quad (10)$$

where y_i is the observed value and $\hat{y}_i$ is the predicted value.

3.4.1 K-Means:

K-means is an unsupervised Machine Learning algorithm for classification. It makes it possible to group similar observations in the same cluster distinct from other clusters.

In order to group data in the same cluster, the algorithm measures the dissimilarity distances. If the distance between two observations is reduced, the two observations belong to the same cluster.

Among the most usable measures of distances, the Euclidean distance:

$$d(x,y) = \sqrt{\sum_{j=1}^{n}\left(x_j - y_j\right)^2} \quad (11)$$

The K-means algorithm tries in an iterative way to minimize the distance between the observation and the center of a cluster.

3.4.2 Support Vector Regressor (SVR):

SVR is an adapted form of the Support Vector Machine classification algorithm dedicated to regression. SVR is a non-parametric supervised Machine Learning technique that does not depend on Gauss-Markov assumptions. On the other hand, it uses Kernel functions to transform the data, which makes it possible to build nonlinear models. SVR is a convex optimization problem since it is based on a maximum margin principle.

3.4.3 K Nearest Neighbors (KNN):

KNN is a supervised machine learning algorithm that is suiTable.for both regression and classification problems. The KNN technique consists of calculating a Euclidean (as in Eq.(20)) or Manhattan (as in Eq.(21)) distance according to which the number K nearest neighbors will be retained.

$$d(x,y) = \sum_{j=1}^{n}\left|x_j - y_j\right| \quad (12)$$

3.4.4 Decision Tree:

The decision tree is the most famous machine learning algorithm in the category of supervised learning. This algorithm makes it possible to make predictions in the case of regression and a classification in the case of classification problems. The decision tree technique consists in separating at each node of the tree the explanatory variables according to decision rules in order to predict the leaves which contain the predicted response values. Data separation is based on the principle of minimizing the square sum of the errors in the two child nodes right and left.

$$SSE = \sum_{i:X_i \in R_l}\left(Y_i - \bar{Y}_l\right)^2 + \sum_{i:X_i \in R_r}\left(Y_i - \bar{Y}_r\right)^2 \quad (13)$$

The crucial advantage of the decision tree is the ease of interpretation and implementation.

3.5 NUMERICAL ILLUSTRATION

To illustrate our concept, we use an open-source claims database. The machine learning algorithms are implemented and performed using the Rstudio computer tool through the packages available on this platform such as (gplots), (rpart), and (e1071).

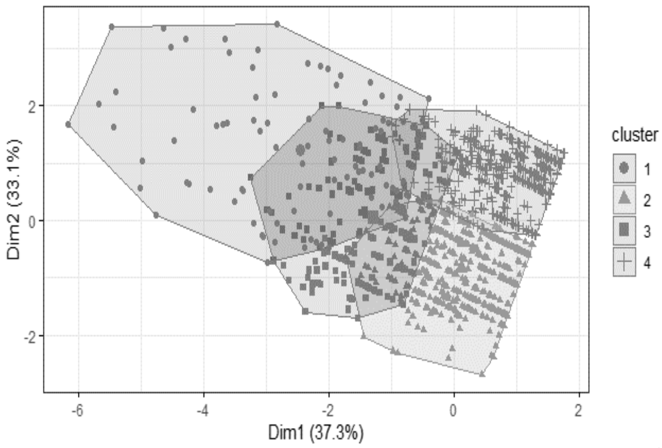

Fig.1. The Clustering of the Cost of the Claims with K-Means Algorithm

First, we construct a priori pricing. We retain two explanatory variables of the number of claims and the cost of claims most used in the motor insurance market in Morocco, namely "engine power" and "type of fuel". Thus, we use for the priori pricing the generalized linear models (appendix 1). Indeed, we use the Poisson distribution for the modelling of the number of claims and the gamma distribution for the cost modelling. After the development of the a priori tariff classes which give us 24 tariff classes corresponding to each engine power associated with each type of fuel (appendix 1), we proceed to the classification of the costs of the claims using the K-means algorithm with four clusters (figure 1). For the four clusters, the average cost is 5818,501, 1081,595, 1192,959 and 1080,819 respectively.

Then, we create a new base which contains the number of claims during the period *t* N={0,1,2,3,4}, the class of tariffs C={1,..,k,..24}, the class of claims cost (H: high or L: low) and the cost of claims.

For the first model, we model the cost of claims according to the three new variables mentioned above and for the second model, we model the cost of claims only according to the class of tariffs and the number of claims during the period t. Finally, we proceed to the calculation of the posteriori tariffs and the Coefficient of Reduction Increase.

4. RESULTS

4.1 RESULTS OF COST MODELLING: FIRST MODEL

In order to choose the model that predicts the cost of claims well, we use Root Mean Squared Error RMSE.

Table.1. Root Mean Squared Error of the First Model

Model	RMSE train data	RMSE test data
Decision Tree CART	1202.634	1195.358
SVR (kernel=radial)	1251.599	1246.902
SVR (kernel=linear)	1260.461	1263.742
SVR (kernel=polynomial)	1247.553	1253.822
KNN regression	1171.162	1240.114

According to Root Mean Squared Error RMSE, the decision tree is the model that best minimizes this error with 1202.634 at the training base and 1195.358 at the test base. The KNN algorithm has the lowest error according to the training base but does not perform well the model at the test base. The SVR with a linear kernel function is the worst performing model on both bases.

The decision tree according to the CART method retained after pruning with an error equal to 0.8994127 consists of 3 splits and 4 leaves.

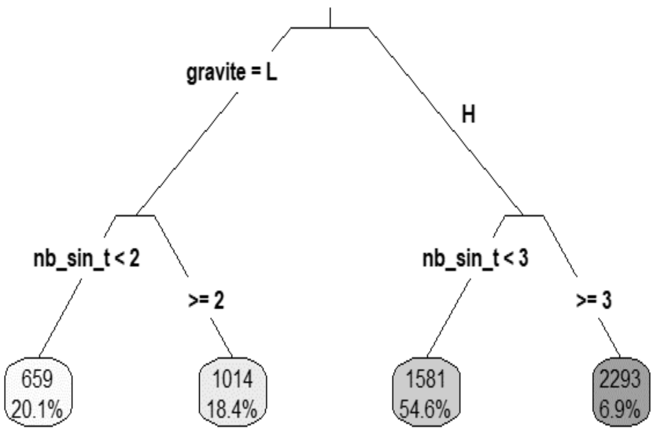

Fig.2. Decision Tree of the First Model

According to this tree, for example, for an insured who will declare three claims with a very high amount, the estimated cost of loss is 2293.

4.2 RESULTS OF COST MODELLING: SECOND MODEL

Table.2. Root Mean Squared Error of the Second Model

Model	RMSE train data	RMSE test data
Decision Tree CART	1277.404	1274.283
SVR (kernel=radial)	1303.632	1308.663
SVR (kernel=linear)	1302.894	1299.512
SVR (kernel=polynomial)	1305.019	1302.928
KNN regression	1257.516	1297.412

According to Root mean squared error, SVR with a kernel radial basis function is the model that does not minimize this error well with 1308.633 at the test base. The KNN algorithm has the lowest error according to the training base but does not perform well the model at the test base. On the other hand, the decision tree model with the CART method is the most efficient model in the two bases with an error equal to 1277.404 at the training base and 1274.283 at the test base.

The decision tree retained after pruning with a minimum error equal to 1.0134 consists of 2 splits and 3 leaves.

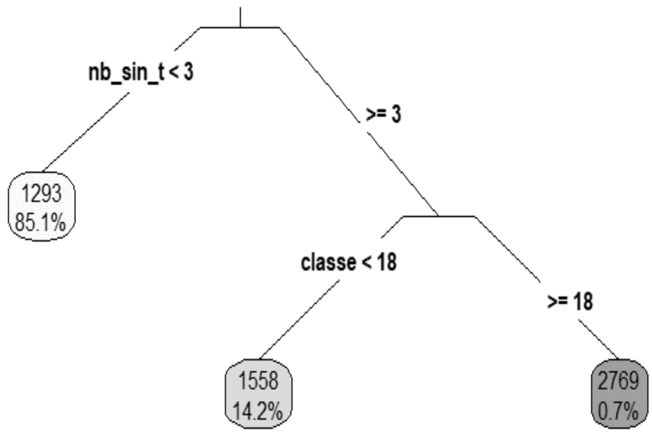

Fig.3. Decision Tree of the Second Model

The tree (figure 3) shows for example that regardless of the class of tariff occupied by the insured, the estimated cost of claims for a declared number less than or equal to 2 is 1293.

4.3 RESULTS OF POSTERIOR TARIFFICATION: FIRST MODEL

The Table.3 below presents the posteriori premiums to be paid according to the class of the insured, the number of claims during the period, and the importance of the amount of the cost of claims.

The premium presented in Table.3 is only the multiplication of the costs of claims estimated by the first decision tree model and the annual frequency of each tariff class estimated a priori (appendix 1).

Table.3. The posteriori premiums grid with the First Model

	Number of claims *N*=*n* and importance of cost claims *W*=*w*						
Class *k*	***n*=0**	***n*=1 *w*=*L***	***n*=1 *w*=*H***	***n*=2 *w*=*L***	***n*=2 *w*=*H***	***n*>=3 *w*=*L***	***n*>=3 *w*= *H***
1	148,35	148,35	355,91	228,27	355,91	228,27	516,19
2	164,52	164,52	394,70	253,15	394,70	253,15	572,46
3	167,13	167,13	400,96	257,16	400,96	257,16	581,53
4	154,06	154,06	369,61	237,06	369,61	237,06	536,07
5	168,52	168,52	404,31	259,31	404,31	259,31	586,39
6	193,25	193,25	463,62	297,35	463,62	297,35	672,40
7	186,03	186,03	446,30	286,24	446,30	286,24	647,29
8	177,97	177,97	426,96	273,84	426,96	273,84	619,24
9	174,14	174,14	417,78	267,95	417,78	267,95	605,93
10	159,29	159,29	382,14	245,09	382,14	245,09	554,24
11	206,31	206,31	494,95	317,44	494,95	317,44	717,85
12	209,58	209,58	502,80	322,48	502,80	322,48	729,23
13	193,19	193,19	463,49	297,26	463,49	297,26	672,22
14	211,33	211,33	506,99	325,17	506,99	325,17	735,31
15	242,33	242,33	581,36	372,87	581,36	372,87	843,18
16	223,17	223,17	535,40	343,39	535,40	343,39	776,51
17	218,37	218,37	523,89	336,01	523,89	336,01	759,82
18	199,74	199,74	479,19	307,34	479,19	307,34	695,00
19	198,01	198,01	475,03	304,67	475,03	304,67	688,96
20	248,24	248,24	595,55	381,97	595,55	381,97	863,75
21	248,29	248,29	595,68	382,05	595,68	382,05	863,94
22	311,29	311,29	746,80	478,98	746,80	478,98	1083,13
23	205,14	205,14	492,15	315,65	492,15	315,65	713,79
24	257,24	257,24	617,14	395,81	617,14	395,81	895,07

The a priori premiums are the initial premiums that do not correspond to a specific level in the bonus malus grid, if the insured has caused no loss the premium is readjusted, and the insured is placed at level 1 with 0 claims, and after he moves in the grid according to the number of claims he has made and their importance. By way of illustration, a policyholder from class 1 who has paid a prior premium equal to 266, if he has not claimed any accident, will enter the malus bonus grid from the 1st level, and he will pay a premium equal to 148 with a reduction coefficient equal to 55% (Table.4). If he has declared four large claims (H), he will be punished by an increased coefficient equal to 193% which corresponds to a premium of 516.

Table.4. The Coefficient Reduction Increase with the First Model

	Number of claims *N*=*n* and *W*=*w*						
Class *k*	***n*=0**	***n*=1 *w*=*L***	***n*=1 *w*=*H***	***n*=2 *w*=*L***	***n*=2 *w*=*H***	***n*>=3 *w*=*L***	***n*>=3 *w*= *H***
1	55,72	55,72	133,67	85,73	133,67	85,73	193,87
2	52,91	52,91	126,94	81,41	126,94	81,41	184,10
3	53,22	53,22	127,68	81,89	127,68	81,89	185,17
4	49,01	49,01	117,58	75,41	117,58	75,41	170,54
5	51,56	51,56	123,70	79,34	123,70	79,34	179,41
6	59,00	59,00	141,53	90,78	141,53	90,78	205,27
7	55,25	55,25	132,54	85,01	132,54	85,01	192,23
8	52,76	52,76	126,57	81,17	126,57	81,17	183,56
9	50,95	50,95	122,24	78,40	122,24	78,40	177,29
10	46,42	46,42	111,36	71,42	111,36	71,42	161,51
11	52,44	52,44	125,81	80,69	125,81	80,69	182,46
12	52,75	52,75	126,54	81,16	126,54	81,16	183,53
13	48,54	48,54	116,45	74,69	116,45	74,69	168,89
14	51,09	51,09	122,57	78,61	122,57	78,61	177,76
15	58,52	58,52	140,40	90,05	140,40	90,05	203,63
16	52,28	52,28	125,43	80,45	125,43	80,45	181,92
17	50,48	50,48	121,11	77,67	121,11	77,67	175,65
18	45,95	45,95	110,23	70,70	110,23	70,70	159,87
19	44,21	44,21	106,07	68,03	106,07	68,03	153,84
20	54,02	54,02	129,59	83,11	129,59	83,11	187,95
21	43,74	43,74	104,94	67,30	104,94	67,30	152,20
22	53,54	53,54	128,46	82,39	128,46	82,39	186,31
23	34,59	34,59	82,99	53,23	82,99	53,23	120,36
24	34,12	34,12	81,86	52,50	81,86	52,50	118,72

The Table.4 presents the increase and reduction coefficients that are not fixed and which apply to the basic premium, in this case, the a priori premium at each movement in the grid. The increase reduction coefficient varies from 34% up to 205%. And each insured moves horizontally by coefficients that are specific to their level of risk.

4.4 RESULTS OF POSTERIOR TARIFFICATION: SECOND MODEL

In the second model, we retain only the number of claims and the a priori pricing class. The premiums calculated in Table.5 are the product of the annual frequency and the cost of claims according to the second decision tree model. According to this second model, there are only two levels for each class, namely whether the insured has declared a number of claims less than or equal to two or greater than or equal to three.

Table.5. The Posteriori Premiums Grid with the Second Model

	number of claims *N*=*n*	
Class *k*	***n*<3**	***n*>=3**
1	291,07	350,73
2	322,80	388,96
3	327,92	395,13
4	302,28	364,24
5	330,66	398,42
6	379,16	456,87
7	365,00	439,81
8	349,18	420,75
9	341,68	411,70

10	312,53	376,58
11	404,79	487,75
12	411,21	495,48
13	379,06	456,74
14	414,64	499,62
15	475,46	572,91
16	437,87	527,61
17	428,46	516,27
18	391,90	839,27
19	388,50	831,98
20	487,06	1043,06
21	487,17	1043,29
22	610,76	1307,97
23	402,50	861,96
24	504,72	1080,88

According to this model, the Bonus Malus grid underestimates the a priori premium because for most classes, even if the insured has not declared any claims, the premium is readjusted upwards so that it can correspond to the associated risk. For instance, if an insured from the class of risk is equal to 4 and he does not declare any claims in the period, he will get a Bonus equal to 4% (Table.6). So, he will pay 314*0.04 = 302 as the next posteriori premium.

Table.6. The Coefficient Reduction Increase with the Second Model

	number of claims *N*=*n*	
Classes *k*	***n*<3**	***n*>=3**
1	109,32	131,73
2	103,81	125,09
3	104,42	125,82
4	96,16	115,87
5	101,17	121,90
6	115,75	139,48
7	108,40	130,61
8	103,51	124,72
9	99,97	120,46
10	91,07	109,74
11	102,89	123,98
12	103,49	124,70
13	95,24	114,76
14	100,24	120,78
15	114,83	138,36
16	102,58	123,61
17	99,05	119,35
18	90,15	193,06
19	86,75	185,77
20	105,98	226,96
21	85,82	183,79
22	105,06	224,98
23	67,87	145,35
24	66,95	143,37

4.5 FINANCIAL EQUILIBRIUM

Table.7. Financial Balance Summary Table

Number of insured		49 985
Number of claims		7 475
Total priori premium		18 377 765,55
Total posteriori premium	**Model 1**	9 755 331,957
	Model 2	18 360 510,42

The total of the a priori premiums amount to 18 377 765.55 on the other hand the total of the posteriori premiums only gives rise to an amount of 9 755 331.957 which is not equal to the total of the priori premiums. However, the second model manages to achieve the threshold of financial equilibrium since the total amount of premiums collected according to this model amounts to 18 360 510.42, which is approximately equal to the total amount of priori premiums.

5. DISCUSSION

The first model of posteriori tariffication presented in the Table.3 allows us to integrate not only the number of claims declared but also the amounts of claims and their importance. Consequently, this model makes it possible to individualize the Malus like the Bonus and to introduce more transparency since the insured is not supposed to have more increased premium if he has not made claims whose cost is very high.

So, our model has gained some scores in terms of transparency and individualization of claims but in terms of financial balance has not reached the threshold which requires, even more research and study to improve it.

In the second system (Table.5), the insured has not a lot of levels to experience according to his level of risk. For some classes, if the insured makes more than 3 claims will be severely punished but if he makes no claims, he will pay a premium less than the last year and the same or more than the first priori premium.

This system with just two levels allows the insurer to collect the same amount if the company applies the priori tariffication. As a result, our second model has gained scores in terms of financial balance but in terms of transparency still needs to be improved.

This attempt to introduce machine learning techniques into the design of the Bonus-Malus System has the advantage of offering posterior premiums more suited to the risk profile taking into account the cost of claims, but it remains to be improved in other studies to review the problem of financial balance or to propose other techniques resulting from artificial intelligence can be used in the Bonus-Malus System in substitution to the classical approach.

6. CONCLUSION

The insurer work in a competitive environment that necessitates offering a competitive product. Among these, we find the Bonus-Malus System. With this view in mind, this article presents a new conception of Bonus-Malus System or Coefficient Reduction Increase system in Morocco based on new tools, in occurrence, the Machine Learning algorithms. In contrast to the classical approach used in the literature, the most advantage of machine learning algorithms is that they do not depend on the assumptions of the number claims distributions provided in the traditional methods.

We presented as a first attempt the possibility of estimating the posterior premium of the Bonus-Malus System directly using CART Classification And Regression Tree method, SVM the Support Vector Machine for regression, and KNN the K-Nearest Neighbor. In this study, we considered not only the number of claims but also the cost of claims. In addition, both models suggested taking into account the priori risk profiles determined at the priori tariffication by considering the class risk in the modelling.

In general, the first model is optimal because it makes it possible to offer a fair Bonus-Malus System, each policyholder pays a posteriori premium that best corresponds to his level of risk taking into account the priori annual frequency of claims at each tariff class and the insured's recent personal driving experience. This will encourage insureds to drive carefully and renew their contracts with the insurer.

However, this work, which remains to be improved, is only the beginning of other research on the use of machine learning in the design of Bonus-Malus Systems which is not yet developed and remains limited to conventional approaches. Therefore, researchers are encouraged to investigate other possibilities for using machine learning in the Bonus-Malus System or improving these proposed models.

APPENDIX 1

Table.8. A Priori Tariffication Grid

Classes	Type of fuel	Power engine	Frequency	Cost	Premium
1	Essence	11	0,22511578	1182,73	266
2	Essence	6	0,24965441	1245,49	311
3	Essence	9	0,25361253	1238,29	314
4	Essence	4	0,23378475	1344,59	314
5	Essence	7	0,25572834	1278,11	327
6	Essence	12	0,29324214	1117,04	328
7	Diesel	11	0,28228986	1192,83	337
8	Essence	10	0,27005666	1249,16	337
9	Essence	5	0,26425211	1293,36	342
10	Essence	8	0,24170738	1419,72	343
11	Diesel	6	0,31306071	1256,69	393
12	Diesel	9	0,3180241	1249,36	397
13	Diesel	4	0,29316053	1357,66	398
14	Diesel	7	0,32067728	1289,91	414
15	Diesel	12	0,36771869	1126,04	414
16	Diesel	10	0,33864466	1260,42	427
17	Diesel	5	0,33136588	1305,44	433
18	Diesel	8	0,30309533	1434,29	435
19	Essence	13	0,30046325	1490,54	448
20	Essence	14	0,37669144	1220,02	460
21	Diesel	13	0,3767738	1506,62	568
22	Diesel	14	0,47236214	1230,77	581
23	Essence	15	0,31128948	1905,06	593
24	Diesel	15	0,39034964	1931,40	754

REFERENCES

[1] C. Vatter, "Simulation of Automobile Insurance Bonus-Malus Systems", *American Journal of Mathematical and Management Sciences*, Vol. 17, No. 3-4, pp. 329-368, 1997.

[2] J. Lemaire, "Bonus-Malus Systems: The European and Asian Approach to Merit-Rating", *North American Actuarial Journal*, Vol. 2, No. 1, pp. 26-38, 1998.

[3] M. Denuit, X. Marechal and J.F. Walhin, "*Bonus-Malus Scales*", John Wiley and Sons, Ltd, 2007.

[4] G. Dionne and C. Vanasse, "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information", *Journal of Applied Econometrics*, Vol. 7, No. 2, pp. 149-165, 1992.

[5] I. Morillo and L. Bermudez, "Bonus-Malus System using an Exponential Loss Function with an Inverse Gaussian Distribution", *Insurance: Mathematics and Economics*, Vol. 33, No. 1, pp. 49-57, 2003.

[6] D. Pierre-Loti-Viaud, "Parametric Credibility, Correlation and Bonus-Malus Pricing", *Comptes Rendus Mathematique*, Vol. 339, No. 4, pp. 283-286, 2004.

[7] S. Pitrebois, M. Denuit and J.F. Walhin, "Multi-Event Bonus-Malus Scales", *Journal of Risk and Insurance*, Vol. 73, No. 3, pp. 517-528, 2006.

[8] G.R. Guerreiro, J.T. Mexia and M. De Fatima Miguens, "Statistical Approach for Open Bonus Malus", *Astin Bulletin*, Vol. 44, No. 1, pp. 63-83, 2014.

[9] R. A. Inoussa, "Statistical Approach to the Calibration of Bonus-Malus Systems in Automobile Insurance", Master Thesis, Department of Computer Science, Universite Du Quebec A Montreal, pp. 1-124, 2013.

[10] W. Ni, C. Constantinescu and A.A. Pantelous, "Bonus-Malus Systems with Weibull Distributed Claim Severities", *Annals of Actuarial Science*, Vol. 8, No. 2, pp. 217-228, 2014.

[11] W. Ni, B. Li, C. Constantinescu and A.A. Pantelous, "Bonus-Malus Systems with Hybrid Claim Severity Distributions", *Vulnerability, Uncertainty, and Risk*, Vol. 45, pp. 1-14, 2014.

[12] E. Gomez-Deniz, "Bivariate Credibility Bonus-Malus Premiums Distinguishing between Two Types of Claims", *Insurance: Mathematics and Economics*, Vol. 70, pp. 117-124, 2016.

[13] G. Tzougas, S. Vrontos and N. Frangos, "Bonus-Malus Systems with Two-Component Mixture Models Arising

from Different Parametric Families", *North American Actuarial Journal*, Vol. 22, No. 1, pp. 55-91, 2018.

[14] R. Oh, P. Shi and J.Y. Ahn, "Bonus-Malus Premiums under the Dependent Frequency-Severity Modeling", *Scandinavian Actuarial Journal*, Vol. 2020, No. 3, pp. 172-195, 2020.

[15] R. Norberg, "A Credibility Theory for Automobile Bonus Systems", *Scandinavian Actuarial Journal*, Vol. 1976, No. 2, pp. 92-107, 1976.

[16] A. Charpentier and M. Denuit, "Mathematics of Non-Life Insurance: Volume 2, Pricing and Provisioning", Paris: Economica Publisher, 2005.