

Analysis of the Dependence of the Performance of the Ensemble on the Accuracy of its Individual Models

Inesa Khvostina

Department of Applied Economics
Ivano-Frankivsk National Technical
University
Ivano-Frankivsk, Ukraine
inesa.hvostina@gmail.com

Andriy Matviychuk

Department of Economic and
Mathematical Modeling
Kyiv National Economic University
named after Vadym Hetman
Kyiv, Ukraine
editor@nfmte.com

Svitlana Savina

Department of Economic and
Mathematical Modeling
Kyiv National Economic University
named after Vadym Hetman
Kyiv, Ukraine
s.savina@kneu.edu.ua

Oleh Yatsiuk*

Department of Economics and
Management Theory
Ivano-Frankivsk National Technical
University of Oil and Gas
Ivano-Frankivsk, Ukraine
olegstya@gmail.com

Abstract—The article devoted to studying the problem of finding methods to increase modeling efficiency by combining the results of calculations of several models. Conditions of expediency of creation of ensemble of models for the solution of the problem of binary classification are investigated. The theoretical basis for the dependence of the ensemble efficiency on the accuracy of the models that form it is studied. Suggestions are given for creating an ensemble of models and increasing its effectiveness.

Keywords—ensemble (committee) of models, probability of a correct classification, surface of probabilities

I. INTRODUCTION

The idea of creating ensemble (committee) models is associated with the problem of solving tasks that require the processing of large arrays of data. Modern mankind produces information in such volumes that since 2007 there are not enough resources for its preservation and the amount of information continues to grow exponentially [1]. Under such conditions, traditional data processing tools are not able to analyze effectively the available streams of information. This problem is typical for most economic sectors, including banking [2; 3].

One of the most urgent tasks of banking institutions is the development of a system of scoring assessment of borrowers' creditworthiness. Today, most of these systems operate on the basis of traditional approaches, for example, such as decision trees and logistic regressions.

Such methods are not suitable for Big Data analysis. The modern theory of such analysis involves using of special algorithms that can process large data arrays more efficiently [4]. One of these approaches is the construction of ensemble models [5]. It is known that when applying mathematical models, one of the main conditions for ensuring a high level of the result's efficiency is the construction of models based on using homogeneous data.

However, in the case of working with Big Data, we come to the necessity of taking into account a significant amount of

substantially heterogeneous information. The list of indicators to be used in constructing scoring models can reach several dozen. Obviously, it's impractical to include such a heterogeneous information in one model. The best performance indicators will give some models built, for example, on data separately from each source of information. Then the synthesis of the obtained models' results is carried out with the help of ensemble technologies. Thus, each of the models will be configured on arrays of homogeneous data and their combination will ensure the correct incorporation of all necessary information. In addition, such structures will be more stable in using of poorly structured information, or in general, its partial absence. In such cases, the models from ensemble for which the data is in sufficient number will be involved.

II. LITERATURE REVIEW

Historically, the first research devoted to the development of ensemble models is the work of Robert E. Schapire (1990). The paper outlines the idea of boosting – an algorithm that improves the accuracy of the one model classification [6]. Further developments in this area, Yoav Freund and Robert E. Schapire (1997) led to the invention of a more effective implementation of this algorithm, known as AdaBoost [7].

Successive research into the development of ensemble technologies took place in the main directions influencing the accuracy of the ensemble results. The expediency of using certain types of models, for example, similar types of neural networks, was explored. Various methods of generalization of model results were considered in order to minimize the error of the ensemble.

Today one of the most popular algorithms for constructing an ensemble of models, which is used to solve classification problems, was proposed by Leo Breiman (2001) [8]. However, in the above-mentioned works, general cases of solving classification problems are considered, therefore, they have a more theoretical

orientation. Only recently, the ensemble of models began to use for constructing scoring estimates, so the number of publications on this topic is very limited.

Thus the theoretical principles of ensemble models application for solving credit scoring problems are developed by Anjali Chopra and Priyanka Bhilare [9], as well as Maciej Zieba and Jerzy Swiatek [10], but important theoretical and practical problems of ensembles technologies application in credit problems remains open.

III. DISCUSSION OF RESULTS

While applying model ensembles, several simpler tasks are being solved simultaneously with one task. The investigated problem is either gradually or immediately divided into segments, for each of which a separate model is being developed, and the results of model calculations are being combined in a certain way.

However, it is not obvious that the result of combining several classifiers will give better quality than a separate model. To study this issue, let's consider the simplified casework of a committee that consists of three models (which will be called further the experts) that solve the problem of binary classification [5]. We denote the probability of a correct classification for each of them through p_1, p_2, p_3 . Probabilities are independent. Let the result of the committee work be determined by a simple vote, that is a certain example will be assigned to the class which is set by three or two experts. The probability of holding the correct classification by the committee is denoted by P_K .

As a result of the work of such committee, the following options are possible: all three experts conducted the classification correctly, one of three made a mistake, two of three or all three experts made mistakes.

The correct decision will be taken by the committee in the first two cases. By describing them we get four possible situations for the correct solution:

- all three experts didn't make mistakes – probability of the situation is calculated as $\theta_1 = p_1 \cdot p_2 \cdot p_3$;
- the first expert made a mistake and the other two conducted the classification correctly – the probability of the situation is calculated as $\theta_2 = (1 - p_1) \cdot p_2 \cdot p_3$;
- the second expert made a mistake and the other two conducted the classification correctly – the probability of the situation is calculated as $\theta_3 = p_1 \cdot (1 - p_2) \cdot p_3$;
- the third expert made a mistake and the other two correctly conducted the classification – the probability of the situation is calculated as $\theta_4 = p_1 \cdot p_2 \cdot (1 - p_3)$.

The probability of a successful committee's work will consider the probability of four situations mentioned above:

$$P_K = \theta_1 + \theta_2 + \theta_3 + \theta_4 = p_1 \cdot p_2 \cdot p_3 + (1 - p_1) \cdot p_2 \cdot p_3 + p_1 \cdot (1 - p_2) \cdot p_3 + p_1 \cdot p_2 \cdot (1 - p_3) \quad (1)$$

If one of the probabilities is fixed at a certain level, for example $p_3 = \text{const} = C$, then function (1) becomes a function of two variables. By changing the value of the constant it is possible to clearly reflect the behavior of the

function P_K , thus that means examining the result of the model committee work for different classification quality values given by individual experts.

Figure 1 shows a surface view corresponding to the form of the function P_K for the case when $p_3 = C = 0.3$ and $0 \leq p_1, p_2 \leq 0.3$. Under these conditions all three experts are characterized by low accuracy of classification, which is worse than could be obtained after random selection (for example, while throwing a coin).

Fig. 1 shows the zones of gradual growth of P_K function which are marked on the surface, ranging from 0 to the maximum value of 0.216. Such maximum value of the function corresponds to the probability of correct classification by the model committee, when the probabilities of all three experts are individually equal to 0.3 (but, apparently, overall accuracy is lower for each individual expert).

We investigate the dependence of the probability of correct classification by the committee on the accuracy of two models' classification in a full set of probability values from 0 to 1 at a fixed low accuracy of the third model at the level of 0.3. So, Table 1 shows the value that takes the function P_K at points with discrete coordinates with changing of p_1 and p_2 from 0 to 1 in step 0.1 while $p_3 = 0.3$.

Under these conditions, the surface of the function P_K is convex upward, which can be seen as the lines of color transition in Figure 1, as well as due to the calculations given in Table 1. The surface acquires the largest values along the line $p_1 = p_2$ (corresponding values are shown in bold in the table).

Any cross section of the surface P_K and the plane, which is perpendicular to this line, will correspond to the function of the form $p_1 = -p_2 + \text{const}$ ($\text{const} = p_2$ for $p_1 = 0$). For example, the cross section $p_1 = -p_2 + 0.2$, according to Table 1, corresponds to the following values: for $p_1 = 0$ and $p_2 = 0.2$, the value $P_K = 0.06$; another extreme point $p_1 = 0.2$ and $p_2 = 0$ also gives the value $P_K = 0.06$; for an average point $p_1 = 0.1$ and $p_2 = 0.1$, we have $P_K = 0.064$. As a result, from these calculations and the lines of transition of colors on Figure 1 it is clear that the surface has a slight bulge up. A similar situation is observed from the data of Table 1 for any other cross section of this surface that is perpendicular to the axis $p_1 = p_2$.

It is essential that for this convex surface at its highest level in all cases when $p_1 = p_2$ the value of the function $P_K < p_1 = p_2$. Consequently, in this case the committee gives a lower estimation accuracy than a separate expert, which can be seen from the Table 1. This situation is always approximate to the work of the committee, in which at least one of the models has a poor classification accuracy.

Such committee achieves exceeding the accuracy of the assessment above $P_K = 0.5$ only if the probability of a correct classification by two committee experts is not less than 0.6 (when the accuracy of the third expert's assessment is 0.3). And such accuracy of the classification that the first expert has, for example, at the level of 0.7, the model committee can

give only with the accuracy of the other expert at the level of 0.9.

Figures 2, 3 show the cases where the probability of correct classification by the third expert is higher ($p_3 = 0.5$ and 0.7, respectively).

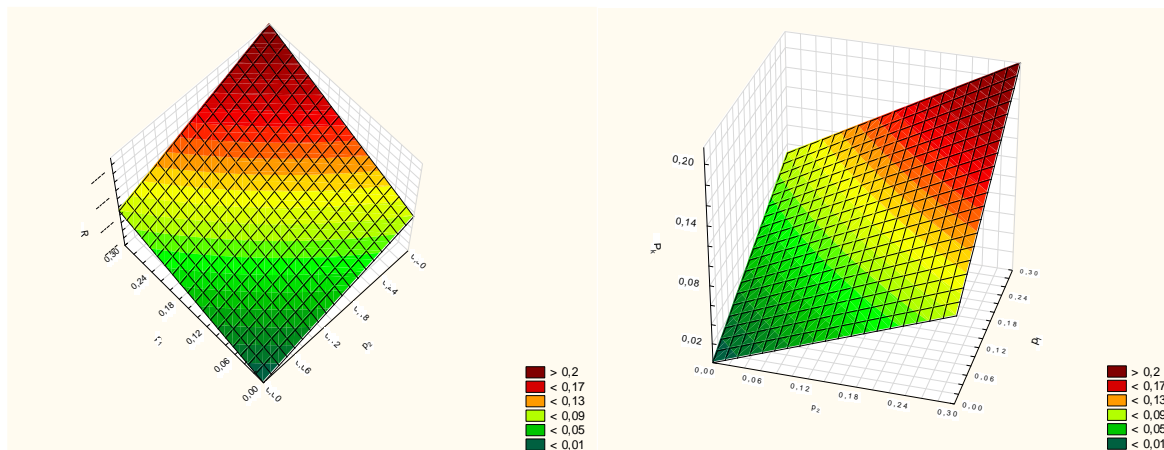


Fig. 1. The image of P_K surface in two views with $0 \leq p_1, p_2 \leq 0.3$, $p_3 = 0.3$

TABLE I. THE VALUE OF A CORRECT CLASSIFICATION PROBABILITY FOR THE MODEL COMMITTEE IN CASE OF A FIXED PROBABILITY OF A CORRECT CLASSIFICATION FOR ONE OF THE EXPERTS $p_3 = 0.3$

Probability of the correct classification of the second expert, p_2												
Probability of the correct classifications of the first expert, p_1		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	0	0	0.03	0.06	0.09	0.12	0.15	0.18	0.21	0.24	0.27	0.3
	0.1	0.03	0.064	0.098	0.132	0.166	0.2	0.234	0.268	0.302	0.336	0.37
	0.2	0.06	0.098	0.136	0.174	0.212	0.25	0.288	0.326	0.364	0.402	0.44
	0.3	0.09	0.132	0.174	0.216	0.258	0.3	0.342	0.384	0.426	0.468	0.51
	0.4	0.12	0.166	0.212	0.258	0.304	0.35	0.396	0.442	0.488	0.534	0.58
	0.5	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65
	0.6	0.18	0.234	0.288	0.342	0.396	0.45	0.504	0.558	0.612	0.666	0.72
	0.7	0.21	0.268	0.326	0.384	0.442	0.5	0.558	0.616	0.674	0.732	0.79
	0.8	0.24	0.302	0.364	0.426	0.488	0.55	0.612	0.674	0.736	0.798	0.86
	0.9	0.27	0.336	0.402	0.468	0.534	0.6	0.666	0.732	0.798	0.864	0.93
	1	0.3	0.37	0.44	0.51	0.58	0.65	0.72	0.79	0.86	0.93	1

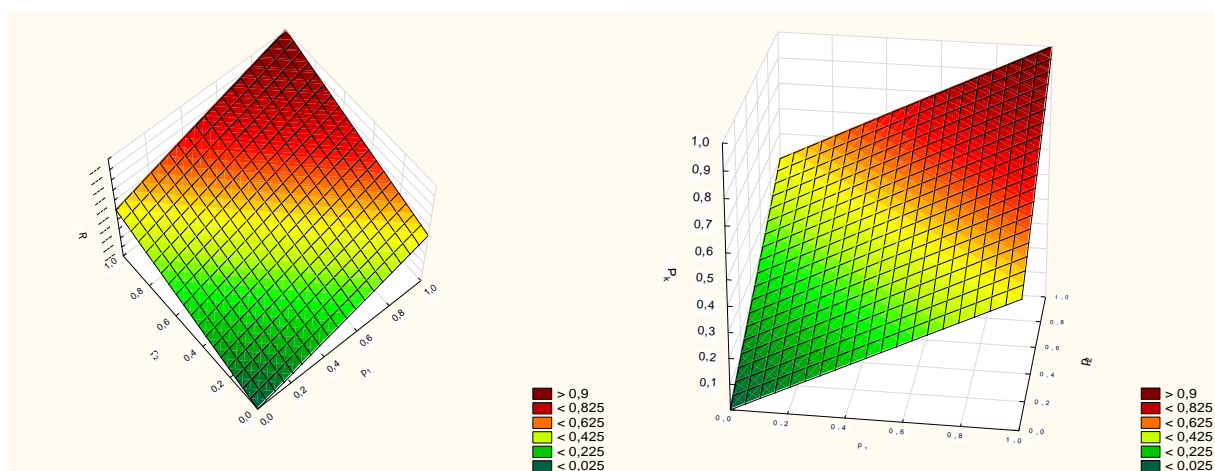


Fig. 2. The image of P_K surface in two views with $0 \leq p_1, p_2 \leq 1$; $p_3 = 0.5$

In Figure 2 $p_3 = 0.5$ while $0 \leq p_1, p_2 \leq 1$. Along the line $p_1 = -p_2 + 0.2$ while $p_3 = 0.5$, the values of P_K are the same, as can be seen from Table 2.

Function values P_K along the line $p_1 = p_2$ are also shown in bold in the Table 2. In addition, surface levels of the functions P_K grow linearly, which can be seen by the gradient of color transition in Figure 2, that is, it is an equal plane. In this case, the committee does not detract from the work of individual experts.

Fig. 3, where $p_3 = 0.7$ while $0.5 \leq p_1, p_2 \leq 1$, reflects the case when the surface P_K is concave.

In this case, for $p_1 = p_2$ we have $P_K > p_1 = p_2$. Consequently, the accuracy of such committee gives a better result than its individual models.

Figures 2 and 3 illustrate the main prerequisite for the creation of a model committee: if the accuracy of at least one individual expert has a probability of less than 0.5, then the committee with such model only loses accuracy. Increasing the accuracy of the committee's assessment of models should be expected only if the probability of correct classification is higher than 0.5 for all committee experts [5].

The logic of presenting the surfaces (Figures 1-3) in the three-dimensional space can be represented in one drawing on the plane (see Figure 4).

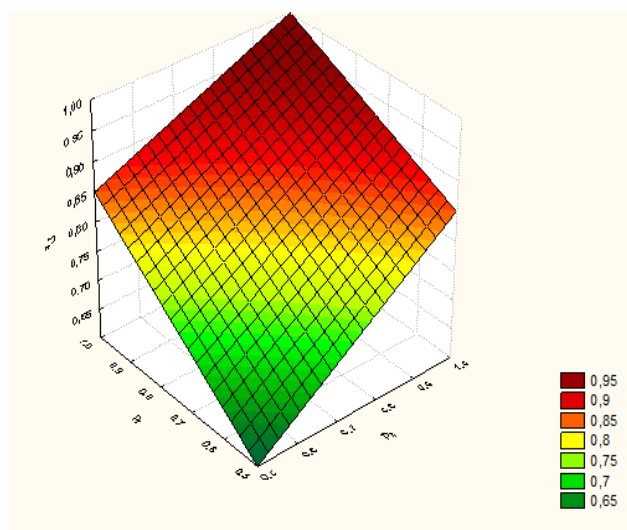


Fig. 3. The image of P_K surface with $0.5 \leq p_1, p_2 \leq 1$; $p_3 = 0.7$

TABLE II. THE VALUE OF THE CORRECT CLASSIFICATION PROBABILITY FOR THE MODEL COMMITTEE IN THE CASE OF A CORRECT CLASSIFICATION PROBABILITY FOR THE THIRD EXPERT $p_3 = 0.5$

		Probability of the correct classification of the second expert, p_2					
		0	0.1	0.2	0.3	0.4	0.5
Probability of the correct classifications of the first expert, p_1	0	0	0.05	0.1	0.15	0.2	0.25
	0.1	0.05	0.1	0.15	0.2	0.25	0.3
	0.2	0.1	0.15	0.2	0.25	0.3	0.35
	0.3	0.15	0.2	0.25	0.3	0.35	0.4
	0.4	0.2	0.25	0.3	0.35	0.4	0.45
	0.5	0.25	0.3	0.35	0.4	0.45	0.5

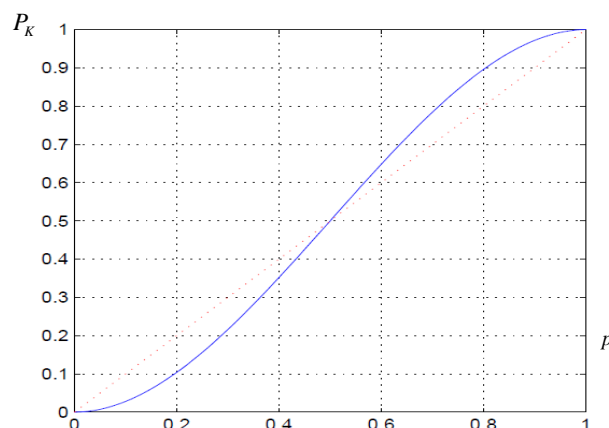


Fig. 4. The dependence of the accurate classification probability by the committee of models P_K from the exact classification of a separate expert p [5]

Figure 4 shows that the curve reflects the probability of accurate classification by the entire committee. On the horizontal axis, the precision of classification is postponed by one expert (for simplification it is assumed that all experts have the same probability p of guessing the borrowers' loan class). From Figure 4 it is seen that for the probability value $p < 0.5$ the accuracy of the committee is lower than the bisector (it is less than the accuracy of the individual expert). At point 0.5, the line crosses the bisector, indicating that the probability of accurate classification by the committee and all three experts is equal. Then the curve goes above the bisector – the committee shows higher precision of guessing than a separate expert.

The logical development of the idea of creating committees to improve the accuracy of classification is the committees' formation of committees, but in [5] this does not lead to a significant increase in the accuracy of such systems. Committees' systems only adapt to informational noise in the data of the study sample, but the accuracy of the new examples is decreasing (the effect of retraining appears).

The growth of the classification accuracy for the models in ensemble can be achieved through the implementation of these two conditions [5]:

- it is necessary to increase the evaluation accuracy of each individual model in the ensemble;
- it is necessary to achieve statistical independence of the errors in the models of the ensemble.

IV. CONCLUSION

The article explores the problem of finding methods to increase modeling efficiency by combining the results of calculations of several models. The expediency of using ensemble technology is substantiated. It is shown that ensemble calculations are more stable in the use of poorly structured information or its partial absence.

It is considered the simplified casework of a committee that consists of three models (experts) that solve the problem of binary classification. The result of the model committee work for different classification quality values given by individual experts is analyzed. It is proven that

increasing the accuracy of the committee's assessment should be expected only if the probability of correct classification is higher than 0.5 for all committee experts. The growth of the classification accuracy for the models in ensemble can be achieved through increasing the evaluation accuracy of each individual model and providing statistical independence of the errors in the models of the ensemble.

REFERENCES

- [1] Big Data at Raiffeisenbank. [Online]. Available: <https://habrahabr.ru/company/raiffeisenbank/blog/332496/>
 - [2] Big data: analytics and solutions – StatSoft Russia. [Online]. Available: <http://statsoft.ru/products/Enterprise/big-data.php>
 - [3] M. A. Beyer and D. Laney, The Importance of "Big Data": A Definition, Gartner, 2012. [Online]. Available: <http://www.gartner.com/id=2057415/> (Accessed 11 Oct 2017).
 - [4] Viktor Mayer-Schönberger and Kenneth Cukier, "Big Data: A Revolution That Will Transform How We Live, Work, and Think", *Paperback*, March 4, 2014.
 - [5] Scientific session of MIFI-2007, in *IX All-Russian Scientific and Technical Conf. "Neuroinformatics-2007" (Part 2)*, Moscow, Russia: MIFI, 2007.
 - [6] R. E. Schapire, "The Strength of Weak Learning. Machine Learning", No 5, pp. 197-227. 1990, Available at: <http://rob.schapire.net/papers/strengthofweak.pdf>
 - [7] Y. Freund and R. E. Schapire, "A decision-theoretical generalization of on-line learning and application to boosting". *Journal of Computer and System Sciences*, No. 55 (1), pp. 119-139, 1997, Available at: http://www.face-rec.org/algorithms/Boosting-Ensemble/decision-theoretic_generalization.pdf
 - [8] L. Breiman, *Random Forest*. Statistics Department at the University of California, Berkeley, CA, USA. [Online]. Available: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>
 - [9] A. Chopra and P. Bhilare, "Application of Ensemble Models in Credit Scoring Models". *Business Perspectives and Research*, vol. 6, issue 2, pp. 129-141, 2018.
- M. Zięba, and J. Świątek, "Ensemble Classifier for Solving Credit Scoring Problems", in *Technological Innovation for Value Creation. DoCEIS 2012. IFIP Advances in Information and Communication Technology*, vol. 372, Berlin, Heidelberg, Germany: Springer, 2012, p. 59-66.