The Impact of Data Science, Big Data, Forecasting, and Predictive Analytics on the Efficiency of Business System

Băcescu-Cărbunaru Angelica and Popovici Mariluzia

Abstract Facing the actual context and nowadays' challenges related to the new actions generated by pandemic factors, we will center the paper on the following areas: Data Science, Big Data, Big Data analytics, which are impacted by forecasting and predictive analytics thus generating improved results for organizations and business units. We will focus on practical implementations in banking and other technical-economical industries. The actual challenges, related to global problems, enforce companies to reshape their business model to current events. Only companies managing businesses optimally adjusted will resist. Organizations should synergistically correlate all their business components, internal units, and processes with the environment. The paper will present the latest methods and technologies used in Big Data analytics for an organization's efficiency and productivity increase. We will focus on the forecasting part, by making a correspondence with predictive analytics. Through forecasting, estimating future values of indicators, managerial Key Performance Indicators, based on specified forecasting algorithms, will show the value of this field, in parallel with predictive analytics for business strategies. Between these two areas, there are interdependencies, each of them having different approaches, providing methods and techniques for the goal of increasing outputs, measured by indicators, usually revealed in views, reports, dashboards, and other system outputs and layers. Pointing on these sections of predictions will reveal the impact of these fields on decision improvement and return enhancement for organizations. The current pandemic circumstances, correlated with afferent effects: teleworking / working from home, improved communications, increased medical actions and treatments etc., reveal the information and predictive' actions values for mankind future evolution.

Keywords Data science · Big data · Statistics · Economic intelligence · Forecasting · Support vector machine

B.-C. Angelica (✉)
Bucharest University of Economic Studies, Academy of Romanian Scientists, Bucharest, Romania
e-mail: mihaela.carbunaru@csie.ase.ro

P. Mariluzia
Vodafone Technologies, Bucharest, Romania

85

1 Introduction

In this paper we will present new domains: data-technologies—Data Science, Big Data, forecasting, predictive analytics, but also new concepts like: Continuous Intelligence, SMART Data, SMART Analytics, etc., and their impact on businesses efficiency.

The aim is to show how the novelty and actuality of informatics systems (targeting financial-banking systems) storing and processing data with Big Data platforms, using forecasting methods, particularly Machine Learning algorithms, from the predictive-analytics sub-domain, can increase efficiency in the financial-banking industry, achieving future benefit-growth.

One of the recent areas of knowledge is *Data Science*—exploring different concepts associated with data—methods, techniques, algorithms, processes, and systems, in order to extract information and mainly knowledge and insights from different data types, structured and unstructured data, in order to apply results and findings in a large number of industries and areas from real life, so as to increase benefits. As an interdisciplinary field, this domain is related to data analysis, modeling, informatics, statistics, machine learning—interconnecting theories and knowledge from many areas—like mathematics, information technology, predictive analysis, etc. The most actual concepts and technologies used to work with data are as follows: Databases, Economic Intelligence, Data Warehouse, Data Mining, Big Data, predictive analytics, forecasting, etc. Some of the concepts we will discuss and detail in the next paragraphs are Big Data and its techniques; predictive analytics, forecasting, but also M.L. and Artificial Intelligence (we will also mention and frame modern concepts like SMART Data, SMART Analytics, Continuous Intelligence, Strategic decisions, etc.). Referring to other technologies interconnected with Data Science—like Data Bases, Economic Intelligence, most recently studied: Continuous Intelligence, data processing, data analytics—we will mention the *Big Data* area, which deals with data collection mechanisms—analyzing, extracting information from datasets with high volumes or increased complexity, so huge that they cannot be operated with traditional methods and techniques through classic software platforms for data processing. So, the concept of Big Data points out the large data volumes, which hamper classical applications to achieve efficient time and cost results, by significant sizes. Some of the most important Big Data companies in 2021 are Google, VMware, Splunk, Amazon, Oracle, IBM, HP Enterprise, Teradata, Mu Sigma, SAP, EMC, etc. (Analytics Insight 2021). The most important advantages of Big Data are *Analytics* and *predictive analysis*. "Predictive analysis should be based on taking into account in the analytical methodology possible the most *modern prognostic models* and a large amount of data necessary to perform the most accurate predictive analysis. In this way, the result of the prediction analysis performed will be the least subject to the risk of analytical error, i.e., an incorrectly designed forecast" (Gwoździewicz et al. 2020).

Big Data analytics tools can predict outgrowth accurately; in this way, driving businesses and organizations to optimal decisions and strategies, while optimizing their processes and operations in an efficient manner, while also reducing risks. *Big Data Analytics* is the process of huge dataset investigations, structured or unstructured data, of various types, collected in different ways, aligned per variated models, the main objective being detecting hidden *patterns*, correlations—predicting customers behavior and preferences, trends in market movements. All of these elements are helpful aspects for decisional processes and organizational strategies. (Improving customers' experience, developing new products, enhancing innovation area, avoiding risks by identifying fraudulent activities, etc.) There are many actual platforms developed for the Big Data Analytics area. Several of them—strongly outlined on the market—are Storm, Cloudera, Apache Hadoop, Grid grain, Space Curve, SAS, Microsoft Azure, etc. "It is a further progression of Business Intelligence (BI) and data mining combined with statistical techniques. Business Intelligence processes help analysis of internal and external data to enable business executives to make intelligent decisions" (Jeble et al. 2016). Economic Intelligence is closely linked to BI.

The process of making predictions, anticipating future trends, based on historical and present data, usually by analyzing actual tendencies, is represented by ***forecasting.*** Bajari et al. have recently shown that "Clearly, data is used in many aspects of company decision-making, and our focus is on the application to demand forecasting. We believe that this is a particularly good area to study, as the success of forecasting models is relatively straightforward to assess" (Bajari et al. 2019).

Data analytics represents an area of raw data analysis, extracting information, knowledge, and conclusions from these data. Generally, associated techniques and processes are automated; revealing trends and metrics resulting from data investigated. Outputs and conclusions are used in processes optimization, efficiency enhancement for systems, or organizations. Strongly related to forecasting, there is the predictive analytics area, nowadays growing conjunctively with ITC and Data Science fields. ***Predictive analytics*** represents operations over data, using mathematical-statistical algorithms, lately machine learning techniques, in order to identify the expectations/probability for future outputs, taking into consideration influencing factors. Research on this area is based on historical events and data, providing the best estimation of what will happen in the future. "This analytics is concerned with forecasting and statistical modelling to determine the future possibilities based on supervised, unsupervised, and semi-supervised learning model" (Sivarajah et al. 2017).

We will focus our research on the forecasting area, taking into consideration techniques of forecasting. One of the forecasting areas, with the parallel growing domain, *Machine Learning*, is widely used together with Mathematics and computing, Probability and Statistics, and other connected domains.

2   Literature Review

The aim of this study is to present the area of modern data-technologies: Data Science, pointing to Big Data platforms newly utilized in companies, with forecasting algorithms, from the Machine Learning field (and other new technologies like digitalization, distant-communication, etc.). Many authors have addressed issues related to new data technologies, from different points of view, depending on the specificities of the activity.

In this regard, we will mention some of the most recent research papers. Jeble et al. (2016) published a study exploring the use of Big Data and predictive analytics, and Gwoździewicz et al. (2020) developed a study on the use of Data processing tools in a period of fast digitalization. Also, Bajari et al. (2015) studied and published an article on the evolution of Big Data in the modern economy and Big data impact on firm performance, and Sivarajah et al. (2017) elaborated a study that deals with critical analysis based on Big Data and analytical methods. Mikalef et al. (2019) examined also the relations between Big data and firm performance and the possibilities to meet future challenges. Bajari et al. (2015) studied how to use Machine Learning methods for demand estimation. Carbo-Valverde et al. (2020) published an extensive paper on the ML approach to the digitalization process inside the bank system, and Ampountolas et al. (2021) deal with the study of the ML approach for micro-credit scoring and have published a paper on this topic. He et al. (2014) wrote and presented a material at an international conference paper about the importance of the loss of customers' prediction of commercial banks, based on the Support Vector Machine model. Sabbeh (2018) was concerned with how the ML techniques are used for customer retention and publishes an article with a focus on SVM techniques, and Aydiner et al. (2019) conducted a study on what role plays business process performance in the relationship between the adoption of business analytics and firm performance.

The data used in the analysis within the article are taken from the statistical yearbook 2018 and other works of the National Institute of Statistics, the Romanian Association of Banks (ARB), and the official site of the National Bank of Romania. The final aim is to show how the novelty and actuality of ITC systems (we will focus on informatics financial-banking systems) and Big Data platforms can increase efficiency in the financial-banking industry, realizing future benefit-growth—we will also use forecasting methods and algorithms for this. This endeavor is realized in order to identify the impact of these new modern technologies on decisions and strategies influencing the efficiency of business systems. Using Machine Learning algorithm (in particular, SVM) we will point out the increased efficiency in business units after using modern data technologies (financial-banking units being one of the impacted areas).

Tudorel Andrei, the president of INS, recalled that "Big Data information is essential for the modernization of statistics in the globalized world." (CECAR Business Magazine 2021). Bajari et al. claim that there has been a high level of interest in modeling consumer behavior in the fields of computer science and statistics (Bajari

et al. 2015). In a 2018 comparative study, Sabbeh says that customer retention has been increasingly investigated in many business domains including banking. In May 2019, Mikalef et al., who published in the Journal of Business Research, showed that big data analytics, when applied to problems of specific domains such as healthcare, service provision, supply chain management, and marketing, can offer substantial value.

3 Research Questions/Aims of the Research

3.1 Data

The current research uses annual time series data for the period 2018–2020 gathered from several sources, namely the Romanian Statistical Yearbook (2018), the Romanian Association of Banks (ARB)/the Romanian banking system (2021); Financial Magazine, 2021, the National Bank of Romania (2021) official sites, etc. Apart from the technological development (in particular digitalization using Big Data and other Data platforms), business sector development growth is also included in the model as a control variable.

3.2 Technological Development Impacted the Financial-Banking Sector

Several figures of the Romanian *banking systems* show that this sector holds *76.1%* of the assets of the Romanian financial system (Q1 2020). Credit institutions have invested substantial amounts in the development of *modern technologies and digital solutions* during the last years. In the COVID-19 pandemic context, the number of various banking operations realized remotely increased, the digitalization of this industry was promoted, so it improved and grew. The value and number of payment transactions with cards issued by resident payment service providers increased by almost 10% since the beginning of the COVID-19 pandemic, thus accelerating the digitization process in the banking system. With this process of digitization increase (using modern storage and operating systems) and optimization of operational expenses, in 2019, the network of banking units registered a *reduction of 6.4%*, up to 4,*758* banking branches, while the number of employees reached *53,106*. For the financial-banking sector, using modern systems, the above figures mean cost reductions so, *efficiency increase*." Understanding the digital jump of bank customers is key to design strategies to bring on board and keep online users, as well as to explain the increasing competition from new providers of financial services" (Carbo-Valverde et al. 2020).
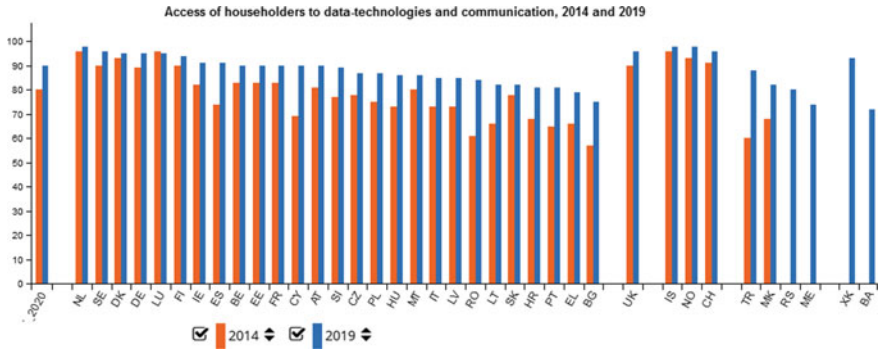
Fig. 1 Access of householders to data-technologies and communication, 2014 and 2019. *Source* https://www.bancherul.ro/stire.php?id_stire=20262&titlu=[big-data-technologies-can-lead-to-anti-competitive-behaviors-warns-the-competition-council]

From a global perspective, the Romanian data market, including Big Data technologies, recorded an increase in the period 2013–2019, the average annual compound growth rate (CAGR) being 26.4%, well above the European average (CAGR 7.4%) (on data market, digital data is marketed as products or services). The above figures indicate an orientation of the economy toward the use of *Big Data technologies*, the introduction of *artificial intelligence,* and a continuous process of digital transformation. Thus, the number of data providers has increased significantly, from approximately 2,400 in 2013 to 5,750 in 2019 (Bancherul Homepage 2020).

Referring to technologies used by customers, especially those related to data and communication, according to the information from the official site of the European Union [https://ec.europa.eu/], and by comparing data from 2014 and 2019, we can see in the following figure (Fig. 1), for Romania, in 2014, that the percentage of ITC technologies utilization by householders was 61%; in 2019, the percentage was 84%, which appears as an average value comparing with the other countries studied.

According to the data provided by the Institute National of Statistics, in the second figure below (Fig. 2), we can see that for the period 2014–2019, the subjects using data technologies and communication for banking operations increased from 17.2% in 2014 to 29.4% in 2019. A significant increase: 5.1%, in 2017, compared to this previous year (only 18% from the total of users).

These figures indicate growth in using data and communication technologies and also banking services for transactions (based on new ITC technologies implementations on a business level in this case, the banking field).

Also, the figure below (Fig. 2), another source for our data input, shows the transactions in Romania, 2014–2019 (this means ITC technologies utilization increased last years due to the implementation of new communication and data technologies in business and organizations) (EUROSTAT Homepage 2021).

Based on the above data collected from different sources and processed, it was realized a first set of data sources for the application of the research methods and associated algorithm.
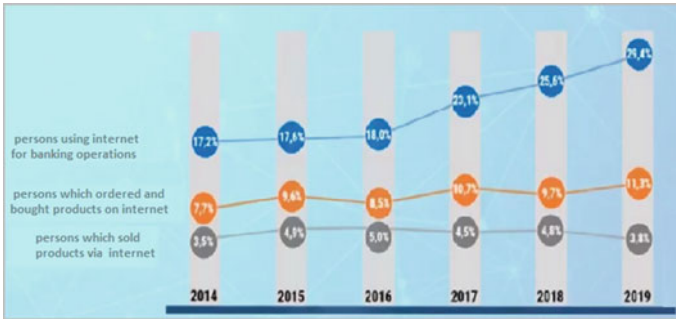
Fig. 2 Internet services utilization—online transactions, 2014–2019. *Source:* https://www.fac ebook.com/INSTATISTICA/videos/2927143920718411/

3.3 Applying Modern Forecasting with Machine Learning Algorithm—SVM

The research aims to apply Machine Learning forecasting algorithm (in particular—Support Vector Machine) in order to identify the impact of new data-technologies implemented (Big Data and digitalization) on online operations, increased transaction volumes in the banking sector, leading to business efficiency in units within the financial-banking area (this endeavor focuses on the identification of increased growth in efficiency for the next period for those units using new data-technologies (above-mentioned). "Increasing financial consumption demand of customers further intensifies the competition among commercial banks. To increase their profits for continuing operations and enhance the core competitiveness, commercial banks must avoid the loss of customers while acquiring new customers. We study commercial bank customer churn prediction based on the SVM model and use a random sampling method to improve the SVM model, considering the imbalanced characteristics of customer data sets. The results show that this method can effectively enhance the prediction accuracy of the selected model" (He et al. 2014). Using the Support Vector Machine algorithm, the purpose is to find the hidden impact of increased ITC technologies (Data based in particular), on predicted increased efficiency and better decisions and strategies on organization's management.

4 Research Methods

One of the data sets used in the study is described in the table below (Table 1). Data refers to financial-banking indicators from Romania, Dec. 2019–Dec. 2020.

The paragraphs below present the model specification and methodology applied in order to obtain the estimated results and prove efficiency increase.

Table 1 Aggregate indicators for credit institutions (Banks, branches of foreign banks, Credit coop)[a]

	Dec. 2019	Mar. 2020	Jun. 2020	Sep. 2020	Dec. 2020
Number of credit institution	34	34	33	34	34
of which foreign banks	7	7	7	8	8
Total net assets (billion lei)	495.2	518.4	518.5	533.1	560.2
ROA[b] (%)	1.34	1.27	1.08	1.17	0.98
ROE[c] (%)	12.21	11.70	9.81	10.48	8.86
Rate of return for base activity (%)	184.09	168.08	180.02	185.58	185.51

Source Adapted from BNR
[a]The indicators include only banks and Credit coop
[b]Return on assets (Annualized net profit/Total average assets)
[c]Return on equity (Annualized net profit/Average own capital)

4.1 Methodology and Specification of the Model

The current study, by applying Machine Learning algorithm: Support Vector Machine, in particular, will assess the impact of ITC development (particularly data technologies, Big Data, and digitalization) on financial-banking sector efficiency increase. Referring to the Support Vector Machine algorithm, the model applied is presented below. This is a supervised learning algorithm, generally used for classification tasks, but it is also suitable for regression tasks. SVM separates classes/clusters by drawing a decision boundary (therefore, to find the maximum margin hyper-planes it offers the best solution for generalization). For the present case, linearly separable, SVM aims to find a function, a hyper plane, which maximizes the margin between two opposite classes, and therefore, it optimally separates the data into two categories. An SVM classifier tries to maximize the next function (Lp), with respect to the vectors $\vec{w}$ and constant b, to ensure that it is the function with the maximum margin that will separate the two classes,

$$L_P = 1/2\|\vec{w}\|^2 - \sum_{i=1}^{t} \alpha_i y_i(\vec{w}\vec{x}_i + b) + \sum_{i=1}^{t} \alpha_i, \tag{1}$$

where t is the number of training examples;

- αi, (i = 1, …, t), are non-negative numbers such that the derivatives of Lp with respect to αi are zero and αi are the Lagrange multipliers;
- $\vec{w} = \sum_{i=1}^{t} \alpha_i y_i x_i$,
- Lp is called the Lagrangian.

"In this equation, the vectors w and constant b define the hyper plane (Popovici and Bacescu 2021)."

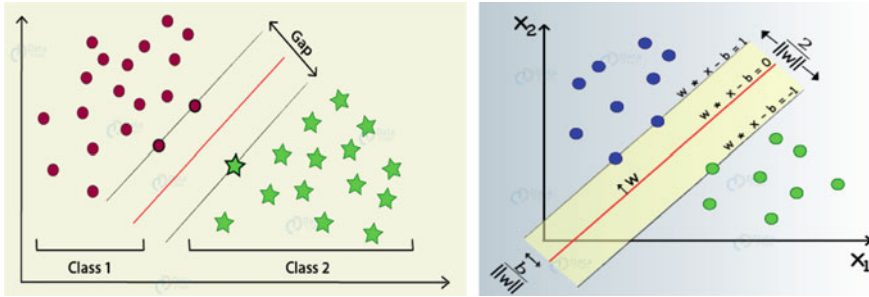A graphical view of the meaning of the SVM algorithm applied is shown below (see Fig. 3).

Fig. 3 SVM algorithm and limits graphical view. *Source* Top 10 algorithms in data mining—https://www.researchgate.net/publication/

Finally, the main goal is to classify units based on the IT data technologies applied (for financial-banking sector units) and to identify outputs resulting in efficiency growth, due to classification realized. Below, the SVM applied algorithm will identify these classes per labels applied ('Significant Efficiency Increase'—0, 'Not Significant Efficiency Increase'—1). Based on this information we will formulate conclusions at the end of this study.

4.2 Applied Algorithm and Results

Regarding the methods used, we will investigate classification for efficiency by SVM algorithm using Python programming language structures. In testing and Python coding writing, we use *Scikit-learn (Sklearn)*—one of the most useful libraries for machine learning from Python. The code below shows some of the pre-processing data as well as the matching data, results for the SVM classifier.

```
import matplotlib.pyplot as plt
import numpy as nm
import pandas as pd
data_set=pd.read_csv('test_data.csv')
x=data_set.iloc[:,[2,3]].values
y=data_set.iloc[:, 4].values
from sklearn.model_selection import train_test_data
x_train, x_test, y_train, y_test - train_test_data (x,y, test_size=0.25, random_state=0)
from sklearn.processing import StandardScaler
st_x=StandardScaler()
x_train=st_x.fit_transform(x_train)
x_test=st_x.transform(x_test)
from sklearn.svm import SVC # "Support vector classifier"
classifier = SVC (kernel='linear', random_state=0)
classifier.fit(x_train, y_train)
```

Fig. 4 Output of SVM classifier, dividing units into two regions. *Source* Studied information from https://www.analyticssteps.com
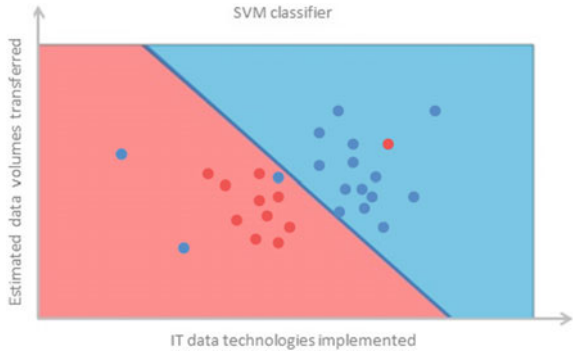
In order to predict output for tested data, we will use the vector: Using the previous

```
output_pred = classifier.predict (x_test)
```

code, we will get the output from the image below (Fig. 4), which will classify business units based on the IT Data technologies applied identifying efficiency growth, due to the classification realized. Classes could be ('Significant Efficiency Increase', 'Not Significant Efficiency Increase').

As we will see in the output image below, the SVM classifier divides the units into two regions with *increased efficiency* and slight *increase efficiency*. Units with *increased efficiency* are in the red area, with red scatter points and those with *slight increase efficiency* are in the blue area with blue points. As expected from our research, we estimated increased efficiency for business units applying modern data techniques (Big Data in particular).

The hyper-plane divided the above-mentioned classes into significant and not-significant efficiency increase variables, which is in line with our estimation.

5   Analysis/Results Interpretation

This paper evaluated the usefulness of machine learning models in assessing defaulting in a financial-banking businesses environment.

We measured increases provided by data-associated informatics systems (Big Data and Smart data platforms) for financial-banking units—referring to volumes of customer transactions, online transactions, digitalization, data platforms, etc. The present study is based on data for the period 2019–2020. The study provides the technique and analysis for the construction of a technological development index in order to trace its impact of data technologies over financial-banking sector growth.

The analysis is based on the Machine Learning model—Support Vector Machine—supervised learning algorithm.

Access to real data is usually not easy. Most articles currently use online datasets, already prepared in a certain format for the model to work well with most machine learning algorithms. In this paper, we confirmed that machine learning algorithms could implicitly predict a real data set in the financial-banking field. The available literature on the identification of outputs resulting in increasing the efficiency of the financial-banking sector did not pay special attention to the impact of ITC development by applying the ML algorithm and in particular SVM, which we did. These elements make this work have its own identity. Based on the findings of this paper, future studies will focus on how to achieve increases in the efficiency of the financial-banking system by implementing new data-related technologies and mass digitization. Big Data Analysis is the often difficult process of examining Big Data to uncover information—such as hidden patterns, correlations, market trends, and customer preferences—that can help financial-banking entities make informed business decisions. On the other hand, the algorithm used in this paper is static in nature and does not take into account the temporal aspects of increasing efficiency. So, we did not estimate how long the efficiency of the organizations will increase. If we can estimate the average time for its growth, the financial-banking unit can consider what other steps need to be taken in the implementation of new technologies and digitalization. Future studies will take macroeconomic variables, such as inflation and unemployment rates, into account in our models to estimate their impact.

These results are relevant to perceive the real transformations behind the digital evolution of consumers. While previous theories and studies have emphasized the importance of the technological components of service and consumer perceptions to explain the evolution of digitalization, other machine learning models demonstrate the phenomenon that customers become digital primarily for the needs of obtaining banking information and, later, for performing transactional services.

Although the models reported in several studies (Apostoulos et al., Sabeh et al.), have the three most accurate results, and the best performances (random forest, XGBoost, and Adaboost), respectively, these are all classifier sets and tree-based algorithms. We are sure that the performance metrics should not be limited to these tree-based algorithms, and the main argument is that SVM allows not only the best accuracy on the training data but leaves, at the same time, a lot of space for the correct classification of the future data. While such models are commonly used to obtain estimates with reasonable accuracy, these techniques were the prelude to the success of ML-SVM models that have been widely applied in many fields.

Numerous studies on the applicability of machine learning techniques have been implemented in various financial-banking fields due to their ability to recognize trends in data evolution (Carbo-Valverde et al. 2020). Other studies indicated that a combination of machine learning methods could provide high accuracy of estimates (Bajari et al. 2015). Our beliefs are that the model applied in the present study offers an appropriate solution in order to generalize the increase of efficiency for the business units that apply advanced data techniques. In ML applications, vector

support machines, SVM, are considered one of the safest and most accurate methods of all study algorithms.

Overall, our study confirms the need to conduct research that covers the entire process of implementing new technologies on Big Data and digitization, rather than focusing on a single dimension. In addition, our research finds that applying machine learning techniques to consumer research provides more accurate results that increase business efficiency.

6   Conclusions

In our study, we intended to emphasize the impact of ITC data technologies (particularly Big Data and digitalization) on the efficiency of the business units (financial-banking businesses in particular).

Findings indicate that ITC data-technologies development (Data platforms like Big Data with specific techniques, digitalization, distant communication, and SMART data analytics), increase the efficiency of the financial-banking system (first of all by cost-reduction and also by implementing and increasing innovation in the IT area, applied on financial-banking products and services—like transactions, loans, etc.). Technological evolution in the IT area—databases, tools, and software applications and analysis and forecasting applications—had a major impact at the micro, but also macroeconomic level—driving the evolution and direction of each unit and organization from the financial-banking area.

Databases and DBMSs, Statistics, forecasting, predictive analytics represent important areas of study and practice of data, helping a lot to make business decisions, based on the information extracted from the stored data. Like any field, these domains are constantly evolving, with new systems, technologies, methods, and techniques developing constantly, systems that help to optimize and streamline business decisions.

The analytical results showed that machine learning algorithms are able to be used to model the efficiency of banking units. Like any other research activity, our study has certain limitations. Therefore, it would be interesting to examine whether emerging economies can face the same process of bank digitization as developed economies. Despite these limitations, we believe that the results of this study are valuable for other researchers and practitioners interested in understanding how digitalization is making its mark in people's lives and implicitly in the work of financial-banking units. Another limitation emerged from the sample selection (credit institution from Romania). Although our effort focused on using real data from profile institutions in Romania, we will base the experimental analysis on a more extensive data set in future works. While some broad qualitative conclusions can be drawn from our results about the importance of different SVM implementation techniques, it is possible that the particular choice of features, etc. may not be universally applicable in other countries and other institutions. Using an extensive data set could not only increase the performance of the model, but could also provide more accurate estimates. Similarly, we

could better control the number of estimated values while understanding the limits of machine learning algorithms. The inclusion of temporal issues at the macroeconomic level is another area of interest for future research. The findings of the study suggest that financial-banking units could benefit from new Big Data technologies and the digitization process by providing services that better meet customer needs and respond more quickly. So, the classification of customers using techniques and data similar to SVM models would make it possible to offer some more personalized digital services. The method is not yet commonly used in the financial-banking field. We believe that practitioners will find arguments in favor of their immediate application (flexibility, ease of use, and scalability).

Our study has a modest contribution to the analytical analysis at the level of financial-banking institutions of the role that the SVM algorithm can play, proposing an efficiency estimation model that takes into account the identification of the results that led to its increase. Therefore, it is recommended that authorities in the business sector (particularly in the banking area) should implement innovative ITC and Data technologies in order to increase the business area and also the efficiency of organizations.

Based on the results of this paper, in future studies we will address ways to increase efficiency for units in the financial-banking sector, proposing the application of modern data techniques in identifying patterns on the one hand complex, and on the other hand non-obvious, hidden in a database with millions of data points. At the same time, we will consider the influences of economic shocks caused by extreme phenomena, such as global pandemics (for example COVID-19), on the growth rate of efficiency.

References

Ampountolas, A., Nde, T.N., Date, P., Constantinescu, C.: A machine learning approach for micro-credit scoring. Risks **9**(3), 50 (2021). https://doi.org/10.3390/risks9030050. https://www.mdpi.com/2227-9091/9/3/50. Accessed 10 Mar 2021

Aydiner, A.S., Tatoglu, E., Bayraktar, E., Zaim, S., Delen, D.: Business analytics and firm performance: The mediating role of business process performance. J. Bus. Res. **96** 228–237 (2019)

Bajari, P., Nekipelov, D., Ryan, S., P., Yang, M.: Machine learning methods for demand estimation. Am. Econ. Rev. **105**(5), 481–485 (2015). https://doi.org/10.1257/aer.p20151021. Accessed 15 Apr 2021

Bajari, P., Chernozhukov, V., Hortaçsu, A., Suzuki, J.: The impact of big data on firm performance: an empirical investigation. AEA Papers and Proceedings, vol. 109, pp. 1–5 (2019). https://doi.org/10.1257/pandp.20191000. Accessed 10 Mar 2021

Bancherul Homepage (2020). https://www.bancherul.ro/stire.php?id_stire=20262&titlu=[big-data-technologies-can-lead-to-anti-competitive-behaviors-warns-the-competition-council. Accessed 20 Dec 2020

Carbo-Valverde, S., Cuadros-Solas, P., Rodríguez-Fernández, F.: A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. PLoS ONE **15**(10), e0240362 (2020). https://doi.org/10.1371/journal.pone.0240362. Accessed 17 May 2021

CECAR Business Magazine (2021). https://www.ceccarbusinessmagazine.ro/[president-ins-most-of-data-sources-are-owned-by-profit-oriented-private-companies]-a4883/

Gwoździewicz, S., Prokopowicz, D., Grzegorek, J.: The use of advanced data processing tools in the era of digitization; Chapter 4, pp. 93–128 (2020). https://www.researchgate.net/publication/344467694.Aaccessed 10 Apr 2021

He, B., Shi, Y., Wan, Q., Zhao, X.: Prediction of customer attrition of commercial banks based on SVM model. In: 2nd International Conference on Information Technology and Quantitative Management, ITQM, Procedia Computer Science, vol. 31, pp. 423–430 (2014). https://www.sciencedirect.com/science/article/pii/S1877050914004633. Accessed 12 May 2021

Jeble, S., Kumari, S., Patil, Y.: Role of big data and predictive analytics (2016). https://www.researchgate.net/profile/Yogesh-Patil-30/publication/309809606. Accessed 15 Apr 2021

Mikalef, P., Boura, M., Lekakos, G., Krogstie, J.: Big data analytics and firm performance: findings from a mixed-method approach. J. Bus. Res. **98**, 261–276 (2019)

National Bank of Romania (2021)

ORACLE Homepage (2021). https://www.oracle.com/ro/data-science/machine-learning/what-is-machine-learning/. Accessed 03 Mar 2021

Popovici, M., Bacescu-Carbunaru, A.: Data management and data governance. Correlations with the area of databases and statistics. Romanian Statistical Review–Supplement no. 2, pp. 121–138 (2021)

RG Homepage (2021). https://www.researchgate.net/publication/29467751_Top_10_algorithms_in_data_mining. Accessed 18 Feb 2021

Sabbeh, S.F.: Machine Learning techniques for customer retention–a comparative study, (IJACSA). Int. J. Adv. Comput. Sci. Appl. **9**(2), 273–281 (2018). https://thesai.org/Downloads/Volume9No2/Paper_38-Machine_Learning_Techniques_for_Customer_Retention.pdf

Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of Big Data challenges and analytical methods. J. Bus. Res. **70**, 263–286 (2017). https://www.sciencedirect.com/journal/journal-of-business-research. Accessed 12 Apr 2021

Romanian Association of Banks (ARB)/Romanian banking system (2021). https://www.arb.ro/sistemul-bancar-din-romania. Accessed 03 Mar 2021

The Romanian Statistical Yearbook (2018)