ИКОНОМИЧЕСКИ ИЗСЛЕДВАНИЯ
ECONOMIC STUDIES

Volume 31 (2), 2022

Anton A. Gerunov[1]

PERFORMANCE OF 109 MACHINE LEARNING ALGORITHMS ACROSS FIVE FORECASTING TASKS: EMPLOYEE BEHAVIOR MODELING, ONLINE COMMUNICATION, HOUSE PRICING, IT SUPPORT AND DEMAND PLANNING[2]

This article puts the problem of forecasting in economic and business situations under scrutiny. Starting from the premise that accurate forecasting is now a key capability for analyzing problems of business operations and public policy, we investigate the performance of alternative prediction methods that include both traditional econometric approaches as well as novel algorithms from the field of machine learning. The article tests a total of 109 different regression-type algorithms across five pertinent business domains – employee absenteeism, success of online communication, real estate asset pricing, support ticket processing, and demand forecasting. The results indicate that forecasting algorithms tend to produce a set of widely dispersed outcome, with some methods such as random forecast and neural network implementations being able to consistently generate superior performance. We further argue that forecast accuracy is not necessarily predicated upon computational complexity and thus, an optimization decision between the costs and benefits of using a certain algorithm can feasibly be made.

Keywords: forecasting; algorithms; random forest; neural network; regression; machine learning
JEL: C44; C45; C52; D81

I. Introduction

Forecasting economic and business variables of interest has always been a central problem for econometrics. Taking its origin from early attempts in demand planning, this task has now expanded to any conceivable field of application from macroeconomic and financial forecasting, through consumer choice modelling, and into operations research. Yet, the standard econometric toolkit has expanded only at a relatively slow pace, displaying a preoccupation with forecasting problems over time series. While this has produced many

---

[1] *Anton A. Gerunov, Ph.D., Dr.Econ.Sc., Associate Professor at the Faculty of Economics and Business Administration, Sofia University "St. Kliment Ohridski", e-mail: a.gerunov@feb.uni-sofia.bg.*

[2] This paper should be cited as: *Gerunov, A. A. (2022). Performance of 109 Machine Learning Algorithms across Five Forecasting Tasks: Employee Behavior Modeling, Online Communication, House Pricing, IT Support and Demand Planning. – Economic Studies (Ikonomicheski Izsledvania), 31 (2), pp. 15-43.*

meaningful results and applications, particularly in the field of macroeconomics, growth theory, and financial economics, the non-time series forecasting problems have largely had to contend with leveraging a limited set of standard instruments such as the linear regression and its variations within the general linear model (GLM). New developments in the field of machine learning promise to expand the variety of available forecasting methods that can be fruitfully applied to different business domains. This article thus attempts to test a wide set of 109 alternative methods (or algorithms) by applying them to five different economic problems and compare their forecast performance. A short overview of the forecasting literature follows in the second section, while the third one presents the application of the algorithms under study to five very different problems that call for a regression approach. The fourth section discusses the results, and the fifth one concludes.

II. Literature Review

The issues of accurately forecasting variables of economic interest are hardly novel. Numerous approaches have been proposed, ranging from very simple ones such as exponential smoothing to rather complex such as neural networks (Hyndman, Athanasopoulos, 2014; Friedman, et al., 2001). Due to the specific of each business situation, forecasting research has often focused on a particular application area such as energy (Hong et al., 2016; Debnath, Mourshed, 2018), customer demand (Ferreira et al., 2016, Schaer, et al., 2019), operations research (van der Laan et al., 2016; Whitt, Zhang, 2019), as well as the classic strand of forecasting macroeconomic variables such as the interest rate (Kunze et al., 2017; Hassani et al., 2021), growth (Christensen et al. 2018; Cronin, McQuinn, 2020), and financial market performance (Mei et al., 2017; Chiu et al., 2018).

While enlightening, research that leverages a single or a few forecasting algorithms can hardly serve as guidance to identifying the best available method in terms of forecasting accuracy. To overcome this problem, the discipline has moved towards either comparative research designs, or so-called forecasting competitions such as the M-competitions where different methods compete against each other. Among the former, we should note the large proliferation of comparative studies in diverse domains. For example, Lago et al. (2018) present a comparative study of 27 state-of-the-art methods for forecasting electricity prices, that include both traditional and more advanced ones. The authors (ibid.) find that machine learning methods show overall better results than traditional statistical ones, with the deep learning methods being the best performers. Alawadi et al. (2020) compare 36 alternative machine learning algorithms that can be used to forecast temperature in a smart building. Their (ibid.) research finds that the ExtraTree algorithm (member of the Classification and Regression Trees, CART, family) shows the best performance in terms of forecasting accuracy. Tyralis et al. (2020) investigate 10 machine learning algorithms for streamflow forecasting, finding that neural networks show the best individual performance. Overall, an ensemble algorithm of the methods under investigation (so-called super learner) outperforms individual methods. There are also some smaller scale comparisons such as those by Koller et al. (2019) with seven algorithms, Salotti et al. (2018) with ten algorithms, Shih and Rajendran (2019) with eight algorithms, and a few others.

Forecasting competitions are similar in philosophy but employ a different research strategy and design. While comparative papers select a single or several sources of data and a single team or researcher tests a limited selection of algorithms, the forecasting competitions provide a large but fixed number of databases and challenge participants to find the best forecasting algorithms, crowdsourcing solutions from a potentially very large number of researchers. One of the first major undertakings in this direction is work by Makridakis et al. (1979) in the late nineteen-seventies which seeks to compare the performance of different forecasting methods on 111 datasets (the M-competition).

This approach also underpins later editions of the M-competitions. In the M-2 Competition, Makrdidakis et al. (1982) present results for 1001 time series that are forecasting using a diverse number of predominantly statistical algorithms. The M-3 Competition significantly expanded the sample of time series, bringing the total number of datasets to 3003 (Makridakis et al., 2000). The M-3 competition finds that while most complex methods do not necessarily produce the best forecasting performance, they still present a significant performance improvement over naïve. Those results validate results from other empirical work as well.

Most recently, Makridakis et al. (2020) published the results from the fourth wave of the M-competitions. This is the largest one to date, testing 61 alternative forecasting methods on over 100,000 different time series. Among the most important findings from M-4 are that the best six methods are significantly better than the others and that more complex methods (including ensembles of methods) have the potential to achieve higher predictive accuracy. The series of M-competitions provides a rigorous and robust evaluation on time series approaches but gives only limited insight about what methods to use in other situations where projections are potentially needed.

III. Methodology and Data

1. Business Domains under Study

The endeavour to pick out the most accurate algorithms for forecasting tasks is a multi-faceted problem. On the one hand, it must ensure that a relatively large number of algorithms are rigorously tested to give sufficient representativeness to the results, and on the other, it must not be constrained to a specific type of problem. In particular, time series analysis is amply researched, with the M-4 competition probably being the most recent comprehensive overview of time series forecasting methods (Makridakis et al., 2020). A novel contribution complementing the conclusions therein must focus on pertinent problems of forecasting of non-time series nature and apply a wide range of advanced methods to solve them.

Thus, this article selects five different forecasting tasks across diverse business domains that can be used as a testing ground. These situations are captured in five specific datasets and some initial work is done by the data creators and curators. This research aims to deepen and expand it to reach a set of novel conclusions with both scientific and practical value. Those five decision domains are as follows. First, we take up a task from operational research and model excessive workplace absenteeism using data from Martiniano et al. (2012), trying to project the hours that a given employee will be absent from his or her job leveraging a set of individual and firm-level explanatory variables. The second situation is a typical problem from digital marketing whereby the analyst must determine what types of online news and communication will be of most interest to a target audience. The challenge here is thus to forecast the number of shares of specific pieces of content, leveraging data from Fernandes et al. (2015).

The third problem is about the valuation of assets with a particular focus on real estate. Given the characteristics of a given unit of real estate, organizations are often interested to evaluate what its actual market price can be in order to facilitate their asset management processes. Again, the challenge here is to forecast the exact price given a dataset of property deals and a number of individual-level estate characteristics (Yeh, Hsu, 2018). The fourth problem under study here is a classic demand forecasting problem. Sharp rises and falls in market demand pose problems for any organization to assign resources and cope adequately. The focus in this situation lies on a logistics company and the volatility of its orders, using data provided by Ferreira et al. (2016). The final business problem under study is how to improve customer support by forecasting the length of the support operation (or open ticket), and use this forecast to minimize it. This is investigated via the dataset provided by Amaral et al. (2018). Table 1 summarizes all situations and sources of data.

Table 1

Types of Forecasting Problems and Data Sources

#	Business Situation	Data Source
1	Excessive absences from work	Martiniano, et al., 2012
2	Online communication	Fernandes, et al., 2015
3	Valuation of asset prices (real estate)	Yeh, Hsu, 2018
4	Sharp changes in market demand	Ferreira, et al., 2016
5	Support ticket processing	Amaral, et al., 2018

The diverse sets of business situations and the associated datasets provide for a wide scope of testing and ensure that the results obtained are sufficiently generalizable to be of use for both academics and practitioners. All the data sets are treated as individual-level data, thus abstracting from their time-series dimensions. This is done as the observations are independent of each other and we only expect a very weak correlation among those. Whenever a time dimensions may be relevant to the forecasting problem (e.g. day of the week for absences or holidays for demand), this is included as a separate explanatory variable.

2. Statistical and Machine Learning Algorithms

Ensuring results generalizability also entails the testing of a wide range of alternative statistical algorithms, some of them hailing from traditional econometrics, and some – from adjacent fields such as machine learning. There is a wide variety of such methods, most of which have not been formally tested on economic problems but for the most common ones such as variants of the linear models family (e.g. the multiple linear regression), support vector machines (SVM), classification and regression trees (CART) and random ensemble forest (RF) models, clustering (e.g. the k-Nearest Neighbors), as well as the ascendant class of neural network models. This research selects practically all regression methods applicable for the aforementioned types of problems, reaching a total of 109 algorithms, and tests them formally. The environment of choice is the R programming language and its packages (see Kuhn, 2008), and thus the R implementations of the selected methods. Table 2 summarizes the algorithms under study.

Table 2

Regression Algorithms Used for Forecasting

#	Method Name	R Implementation	#	Method Name	R Implementation
1	Model Averaged Neural Network	avNNet	56	Multi-Layer Perceptron, multiple layers	mlpWeight DecayML
2	Bagged MARS	bagEarth	57	Monotone Multi-Layer Perceptron Neural Network	monmlp
3	Bagged MARS using gCV Pruning	bagEarthGCV	58	Multi-Step Adaptive MCP-Net	msaenet
4	Bayesian Additive Regression Trees	bartMachine	59	Neural Network	neuralnet
5	Bayesian Generalized Linear Model	bayesglm	60	Neural Network	nnet
6	Boosted Tree	blackboost	61	Non-Negative Least Squares	nnls
7	The Bayesian lasso	blasso	62	Tree-Based Ensembles	nodeHarvest
8	Bayesian Ridge Regression (Averaged)	blassoAveraged	63	Non-Informative Model	null
9	Bayesian Ridge Regression	bridge	64	Parallel Random Forest	parRF
10	Bayesian Regularized Neural Networks	brnn	65	Neural Networks with Feature Extraction	pcaNNet
11	Boosted Linear Model	BstLm	66	Principal Component Analysis	pcr
12	Boosted Tree	bstTree	67	Penalized Linear Regression	penalized
13	Conditional Inference Random Forest	cforest	68	Partial Least Squares	pls
14	Conditional Inference Tree	ctree	69	Partial Least Squares Generalized Linear Models	plsRglm
15	Conditional Inference Tree	ctree2	70	Projection Pursuit Regression	ppr
16	Cubist	cubist	71	Quantile Random Forest	qrf
17	Stacked AutoEncoder Deep Neural Network	dnn	72	Quantile Regression Neural Network	qrnn
18	Multivariate Adaptive Regression Spline	earth	73	Ensembles of Generalized Linear Models	randomGLM
19	Elasticnet	enet	74	Random Forest	ranger
20	Tree Models from Genetic Algorithms	evtree	75	Radial Basis Function Network	rbfDDA
21	Random Forest by Randomization	extraTrees	76	Relaxed Lasso	relaxo
22	Ridge Regression with Variable Selection	foba	77	Random Forest	rf
23	Generalized Additive Model using LOESS	gamLoess	78	Random Forest Rule-Based Model	rfRules
24	Generalized Additive Model using Splines	gamSpline	79	Ridge Regression	ridge
25	Gaussian Process	gaussprLinear	80	Robust Linear Model	rlm
26	Gaussian Process with Polynomial Kernel	gaussprPoly	81	CART	rpart

#	Method Name	R Implementation	#	Method Name	R Implementation
27	Gaussian Process with Radial Basis Kernel	gaussprRadial	82	CART	rpart1SE
28	Stochastic Gradient Boosting	gbm	83	CART	rpart2
29	Multivariate Adaptive Regression Splines	gcvEarth	84	Quantile Regression with LASSO penalty	rqlasso
30	Fuzzy Rules via MOGUL	GFS.FR.MOGUL	85	Non-Convex Penalized Quantile Regression	rqnc
31	Generalized Linear Model	glm	86	Regularized Random Forest	RRF
32	Negative Binomial Generalized Linear Model	glm.nb	87	Regularized Random Forest	RRFglobal
33	Boosted Generalized Linear Model	glmboost	88	Relevance Vector Machines with Linear Kernel	rvmLinear
34	Elastic Net	glmnet	89	Relevance Vector Machines with Polynomial Kernel	rvmPoly
35	Generalized Linear Model with Stepwise Selection	glmStepAIC	90	Relevance Vector Machines with Radial Basis Function Kernel	rvmRadial
36	Hybrid Neural Fuzzy Inference System	HYFIS	91	Subtractive Clustering and Fuzzy c-Means Rules	SBC
37	Independent Component Regression	icr	92	Partial Least Squares	simpls
38	Partial Least Squares	kernelpls	93	Spike and Slab Regression	spikeslab
39	k-Nearest Neighbors	kknn	94	Sparse Partial Least Squares	spls
40	k-Nearest Neighbors	knn	95	Supervised Principal Component Analysis	superpc
41	Polynomial Kernel Regularized Least Squares	krlsPoly	96	Support Vector Machines with Linear Kernel	svmLinear
42	Radial Basis Function Kernel Regularized Least Squares	krlsRadial	97	Support Vector Machines with Linear Kernel	svmLinear2
43	Least Angle Regression	lars	98	L2 Regularized Support Vector Machine (dual) with Linear Kernel	svmLinear3
44	Least Angle Regression	lars2	99	Support Vector Machines with Polynomial Kernel	svmPoly
45	The lasso	lasso	100	Support Vector Machines with Radial Basis Function Kernel	svmRadial
46	Linear Regression with Backwards Selection	leapBackward	101	Support Vector Machines with Radial Basis Function Kernel	svmRadialCost
47	Linear Regression with Forward Selection	leapForward	102	Support Vector Machines with Radial Basis Function Kernel	svmRadialSigma
48	Linear Regression with Stepwise Selection	leapSeq	103	Bagged CART	treebag
49	Linear Regression	lm	104	Partial Least Squares	widekernelpls
50	Linear Regression with Stepwise Selection	lmStepAIC	105	Wang and Mendel Fuzzy Rules	WM
51	Model Tree	M5	106	eXtreme Gradient Boosting	xgbDART
52	Model Rules	M5Rules	107	eXtreme Gradient Boosting	xgbLinear
53	Multi-Layer Perceptron	mlp	108	eXtreme Gradient Boosting	xgbTree
54	Multi-Layer Perceptron, with multiple layers	mlpML	109	Self-Organizing Maps	xyf
55	Multi-Layer Perceptron	mlpWeightDecay			

It should be kept in mind that some implementations under R may be less efficient than implementations under other languages. To ensure comparability of calculation times, we opt for testing only methods implemented under this statistical programming language that can be accessed via a unified interface package (ibid.).

3. Forecasting Accuracy Metrics

A key strand of research has also focused on what criteria should be used to evaluate forecasting results. There is a wide range of indicators and metrics in the scientific literature for the accuracy of a prediction, and we can summarize four main groups of predictive error measures (Hyndman, 2006; Shcherbakov et al., 2013):

- **Measures that depend on the scale of measurement** – they take into account the difference between the estimated and the realized value reported against the scale of measurement. Examples of such measures are average absolute error or mean absolute deviation;
- **Measures taking into account the percentage deviation** – they do not factor in the scale of measurement, as they take into account the difference between the forecast and the realized value in percentages. For some of them, this can be in absolute value. An example is the average absolute percentage error;
- **Relative error measures** – they compare the average errors of a given test method with the errors of a naive forecasting approach. An example of such a measure is the mean scaled error;
- **Measures that are independent of the scale of measurement** – they express each forecast error in relation to the average error of a basic (naive) approach. Similar is the mean absolute scaled error.

Among the more popular forecast error measures, we will consider the mean error, root mean squared error, mean absolute error, mean absolute error, mean percentage error, mean absolute percentage error, and mean absolute scaled error. With ε_i we denote the error of the i-th observation, this error being equal to the difference between the forecast f_i^m realized observation y_i, or:

$$\varepsilon_i = y_i - f_i^m \tag{1}$$

Mean Error (ME) is thus a measure of the average difference between the forecasted and the actual values. It is defined as follows:

$$ME = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \tag{2}$$

This measure is not particularly appropriate as it averages both positive and negative deviations, thus cancelling out some of the variability. Therefore, the average error usually overestimates the predictive accuracy. To address this problem, either the squared error or the absolute error may be considered. Following this logic, we can calculate the following two measures. The root mean squared error (RMSE) looks at the averaged squares of the errors generated and allows one to obtain a measure on the same scale as the original variables. It is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i)^2} \tag{3}$$

Among the measures that use absolute values, it is worth mentioning the mean absolute error (MAE), defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| \tag{4}$$

The error rate can also be presented as proportion or percentage, as it is in the mean percentage error (MPE) and the mean absolute percentage error (MAPE). To calculate them, we define the percentage error of a forecast, p_i as:

$$p_i = \frac{(y_i - f_i^m)}{y_i} \tag{5}$$

The mean percentage error takes into account what is the mathematical expectation of the forecast error, expressed as a percentage, and is defined as follows:

$$MPE = \frac{1}{n}\sum_{i=1}^{n} 100 * p_i \tag{6}$$

Similarly, we also define the average absolute percentage error, using the absolute value of the percentage deviations:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} 100 * |p_i| \tag{7}$$

The latter two metrics may be problematic in cases when the realized value y_i is equal to zero. On those occasions, the fraction is not defined and the metric cannot be calculated. Apart from these presented measures, there are many others that are popular in the scientific literature and provide different perspectives on the predictive accuracy of a model, such as the mean scaled error, MSE, or the mean absolute scaled error, MASE (Shcherbakov et al., 2013; Prestwich et al., 2014; Mathai et al., 2016). The great challenge in assessing the forecasting accuracy of a forecasting model lies in judging which is the optimal measure for a given type of data in a given situation and for specific research goals. To facilitate comparison of the results presented here, we report the values of three separate forecasting accuracy metrics – the mean error (ME), the root mean squared error (RMSE) and the mean absolute error (MAE). The MPE and the MAPE are omitted as they are undefined in some cases.

4. Research Design

The five data sets are tested sequentially with each of the 109 alternative algorithms applied to them. Each of the sets is divided into two samples – a training set and a test set. The training set consists of 80% of the full dataset and is used for the initial estimation of the model. The test set comprises the remaining 20% and is used for comparing different algorithms. The idea behind this split is to minimize the risk of overfitting to the initial training data and only compare algorithm results based on unfamiliar data. To achieve this, once the model parameters are estimated on the training data, this model is used to forecast a target-dependent variable in the test data, and accuracy metrics are calculated. We report the Mean Error (ME), Root Mean Squared Error (RMSE), and the Mean Absolute Error (MAE). The Mean Percentage Error (MPE) and the Mean Absolute Percentage Error (MAPE) are also appropriate metrics, but in some of the situations, these cannot be calculated as the forecast value goes to zero, thus making the fraction undefined. Therefore, MPE and MAPE are not reported. Usually, accuracy metrics tend to be higher in the training set due to overfitting, and somewhat lower on the test set, which is why only accuracy metrics on the test set are reported.

In addition to accuracy metrics, the program for comparison measures the time needed for model training as a proxy for the computational resources required. The time is standardized with the most resource-intensive algorithm given a value of 100% and the rest – presented as fractions of this. This is how the complexity metric is defined – if a given algorithm has a complexity metric of 5%, this means that it is 20 times faster than the slowest algorithm under study. This additional metric allows one to quantitatively measure the tradeoff between costs of forecast (i.e. the time or resources needed to make it) and its benefits (as measured by accuracy). The introduction of some measure of the cost of forecasting allows the researcher to make an informed optimization decision of what method to use and the practitioner – to appropriately size the necessary infrastructure during operations. These considerations are relatively unimportant when dealing with a small dataset of fewer than 10,000 observations but become increasingly prominent as sizeable datasets (so-called “big data”) are analyzed to generate real-time forecasts on the order of millions or billions of observations. The general trend of economics to leverage ever-larger datasets, particularly for microeconomic problems, necessitates the introduction of additional considerations about the methods used and their computational efficiency.

IV. Results

This section provides a brief overview of the five economic situations under study and the application of all 109 algorithms to each of them. It should be noted that in each one, the value of forecasting lies in the organization’s ability to successfully embed it within its decision-making processes and its operational procedures. Furthermore, the accuracy of the forecast generates more value at large scales – even a small benefit applied over millions of transactions can significantly impact business operations. Thus, the proposed more complicated algorithms will be particularly useful in problems where a large number of transactions are being processed and are thus of greater interest to problems in microeconomics rather than to macroeconomic issues.

1. Workplace Absenteeism

Excessive absence from the workplace of key employees is one of the classic operational risks faced by the modern organization. This can be either a physical absence at a standard localized office or a refusal to perform work duties with virtual teams. The absence or incapacity of employees can lead to delays in processes, inability to complete operations in a timely manner, problems with management and transfer of organizational knowledge and reduced motivation of other employees in the team. All these effects are adverse and can lead to a general deterioration of the internal environment, lost profits and realized costs and losses for the organization. In this sense, it is essential to model the risk of absence from work and to take action to minimize it by improving motivation, facilitating employees and assisting them in their family and social duties.

Adequately modelling this phenomenon enables adequate measures that can address and minimize its adverse consequences. Absence from work is an indicator that is strongly correlated with a number of behavioural and situational characteristics of employees and is often a leading indicator of change in employee attitudes, including loyalty to the organization (Hassink, 2018). In this sense, management faces two main tasks in managing the risk of absence. The first is to provide the necessary staff to maintain the duration of the processes in the short term, for which the absent hours should be forecasted and managed. The second is to maintain high motivation in the medium term, and here the absence is rather an indicator of the attitude of employees. Thus, predicting absences at the individual level helps to identify potential intervention points.

To address this business problem, we use data from Martiniano et al. (2012), which describe the hours of absence of employees of a Brazilian courier company for the period from July 2007 to July 2010, as well as a wide range of other characteristics. The total number of observations in this database is 740. It contains data on the reasons for absence, time, situational characteristics (distance from work to workplace, workload), performance of work (fulfilment of objectives, disciplinary proceedings), as well as individual characteristics (age, education, height, weight, number of children, etc.) of the employees. The target variable of the task is the number of hours of absence of each employee, recorded as y. Martiniano et al. (2012) used these data to demonstrate the capabilities of a fuzzy neural network model to predict hours of absence, obtaining satisfactory results.

Building upon this, we evaluate the 106 alternative algorithms on the data under consideration, presenting their predictive accuracy, measured by the root of the root mean square error, in Figure 1. We note a significant group of algorithms with predictive accuracy in the vicinity of RMSE = 13, with almost all the alternatives considered having a root mean square errors for this task in the range from RMSE = 12 to RMSE = 14. The best algorithms register RMSE around 12, but there are a few with a particularly weak representation where RMSE > 15. While almost all algorithms produce satisfactory results, the difference between best and average ones is relatively small. The key consideration, in this case, is to avoid using any of the particularly weak methods, since their results register significantly higher error rates than the average ones.

Figure 1

Histogram of forecast accuracy for workplace absenteeism data

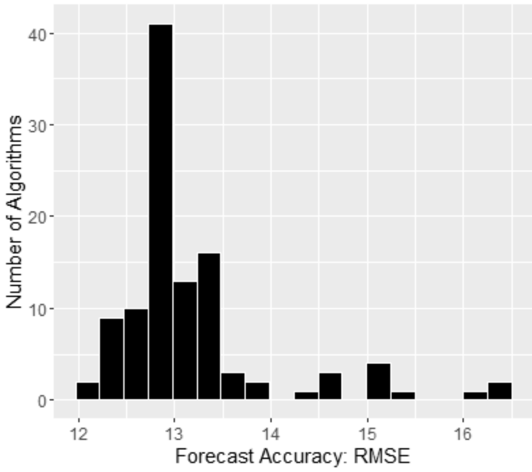


Table 1

Forecast accuracy of the top ten algorithms for workplace absenteeism data

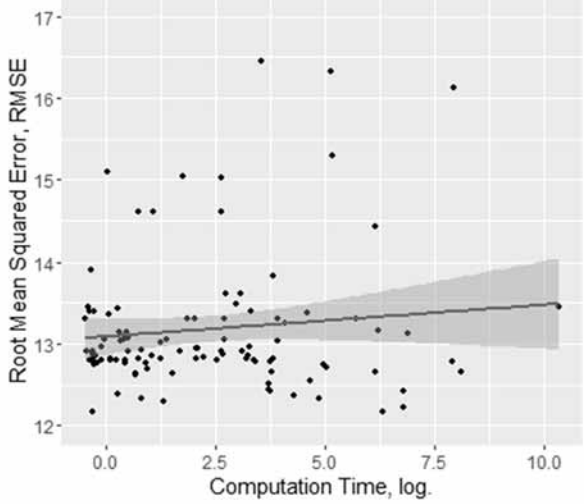
Algorithm	Method	Mean Error, ME	Root Mean Squared Error, RMSE	Mean Absolute Error, MAE	Complexity Measure
CART	rpart2	0.621	12.172	5.376	0.1%
eXtreme Gradient Boosting	xgbDART	0.043	12.174	5.459	1.8%
Random Forest by Randomization	extraTrees	-0.314	12.229	5.418	2.9%
Gaussian Process with Radial Basis Function Kernel	gausspr Radial	0.054	12.287	5.573	0.7%
Bagged CART	treebag	0.205	12.325	5.459	0.4%
Regularized Random Forest	RRFglobal	-0.603	12.330	5.724	0.4%
Conditional Inference Random Forest	cforest	-0.141	12.371	5.581	14.5%
Conditional Inference Tree	ctree	0.798	12.381	5.387	0.3%
Regularized Random Forest	RRF	-0.587	12.424	5.792	2.9%
Random Forest	rf	-0.500	12.430	5.688	8.5%

The methods with the lowest forecast errors are presented in Table 1. We immediately notice that six of the top ten methods are different implementations of the random forest family. The best predictive accuracy is found in a version of the classification and regression trees methods (rpart) with a root mean square error of 12.17 and an average error of 0.62, followed by an algorithm for extreme gradient boosting (xgbDART) with RMSE=12.17 and ME=0.04, respectively.

We note the extremely small difference in predictive accuracy between the top ten methods – it ranges from RMSE=12.17 to RMSE=12.43, which highlights the possibility of choosing a method among this already narrow set based on other considerations – e.g. the required computational resources, organizational experience and other business considerations, interpretability of results and integrate with other systems, etc. From the point of view of the resources used, we recognize that none of the top ten algorithms is particularly demanding from a computational perspective.

Figure 2

Link between forecast accuracy and computation time for workplace absenteeism data



The best method is 1,000 times faster than the slowest alternative and has a computation time commensurate with the multidimensional linear regression. These results are partly due to the popularity of the random forests, which leads to a large number of implementations that are highly optimal in terms of computation time and resources required. Those results can be leveraged by the modern organization to improve staff selection processes and ensure that unexpected drops in productivity due to absenteeism do not threaten business continuity.

2. Online News Sharing

The communication of an organization with its employees, customers, shareholders, regulators and other stakeholders has always been a key management process and an important component of operations to achieve the desired strategic goals. Historically, this communication has been subject to increasing automation, with technological innovation being an important driver of improving the flow of information and knowledge within and outside the structure. Digital transformation brings new channels of communication, and in some of them technological developments fundamentally change the needs of the management process. An example of this is social networks, where there is not a unilateral dissemination of information, but joint creation and active sharing, as organizations are exposed to a number of network and external effects of their actions. A major issue of communication operations is the extent to which a given information reaches the widest possible range of its identified recipients, and in the context of digital online communication, this is mainly a function of its sharing between different users. In this sense, an important part of the communication operations on the Internet of the modern organization consists in identifying the engines of distribution among users and the drivers of information utilization.

Given the specifics of communication operations, we define the main business problem as the need to identify the factors that make it lead to a high level of popularity of an online material and their use in organizational activities. To this end, it is important to model the engines of high distribution and to use them to the maximum extent to optimize publicity activities on the Internet. The purpose of this management process is to minimize the likelihood of failed communication, thereby reducing the unnecessary waste of resources in this activity. In an online environment, a natural and widely available measure of the reach of reached users is the number of shares of a material. Modern social networks allow users to easily share text, sound, picture or video, which not only evaluates the material, but also becomes channels for the distribution of shared messages. In this sense, the business goal is to ensure the highest possible number of shares of the organization's digital communication.

For this purpose, we use the data provided by Fernandes et al. (2015), which presents structured data on 39,797 news items published on the popular online portal Mashable. The authors (ibid.) have applied methods for analysis in natural language on the text of the provided news and have extracted basic quantitative characteristics of each of the texts, including variables such as word count, hyperlinks to different sources, emotional charge, mood polarization, publication category, day of the week, etc. The total number of provided characteristics is 60. In addition, a variable is available, which takes into account the number of shares of the given article on social networks, which we determine for the target variable of the present task. Fernandes et al. (2015) use this data to demonstrate the operation of a highly automated decision support system (expert system). They also test five different forecasting algorithms, finding that a random forest model led to the best results. Building on their work, we expand the scope of forecasting methods here (adding an additional 104) and show how this approach can be effectively used to identify and assess operational risks.

The target variable is forecasted using over 100 alternative algorithms, with the distribution of their predictive accuracy (RMSE) plotted in Figure 2. The vast majority of the tested algorithms have very good predictive accuracy from RMSE <11,000. We observe a notably shifted distribution, with a large number of approaches generating close to optimal results and a smaller number of algorithms with much lower performance.

The ten approaches with the highest predictive accuracy are presented in Table 2. It is noteworthy that six of them represent different implementations of the random forest method. The best algorithm, in this case, is the *ranger*, which is a highly optimized application of random forest that achieves comparable predictive accuracy at very low computational resource consumption. Its root mean squared error is only RMSE = 10,514, and the other ten algorithms score in close vicinity to this result.

Figure 3

Histogram of forecast accuracy for online news sharing data

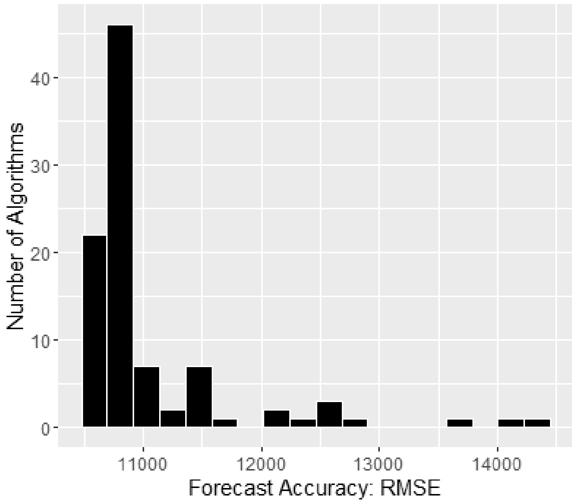


Table 2

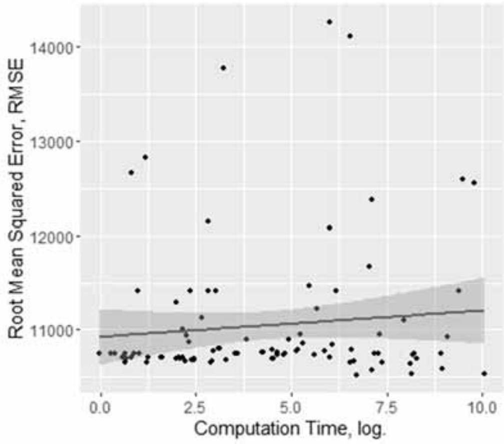
Forecast accuracy of the top ten algorithms for online news sharing data

Algorithm	Method	Mean Error, ME	Root Mean Squared Error, RMSE	Mean Absolute Error, MAE	Complexity Measure
Random Forest	ranger	384.52	10514.60	3566.87	0.35%
Regularized Random Forest	RRF	247.73	10532.25	3630.94	9.92%
Regularized Random Forest	RRFglobal	231.90	10533.87	3633.60	1.46%
Random Forest	rf	213.69	10573.80	3662.55	0.52%
Random Forest by Randomization	extraTrees	209.17	10584.73	3639.87	3.33%
Gaussian Process with Radial Basis Function Kernel	gaussprPoly	311.21	10636.78	3703.16	1.42%
Bayesian Ridge Regression	bridge	349.77	10651.67	3637.69	0.30%
Conditional Inference Tree	cforest	366.84	10656.54	3646.23	0.68%
glmnet	glmnet	370.03	10657.97	3607.08	0.00%
Lasso Regression	lasso	369.31	10659.49	3608.71	0.01%

In addition to the random forest calculation methods, we also observe the Gaussian process, Bayesian and lasso regression algorithms, as well as the generalized linear model. From the point of view of all calculated measures of predictive accuracy, the presented top ten algorithms are difficult to distinguish and the choice of the optimal one can be made on the basis of a suitable resource or organizational criterion.

For this purpose, the measure of complexity can be used as an approximation of the resource needs of the given algorithm. With the exception of the regularized random forest, all other algorithms are relatively fast, with *ranger* being 300 times faster than the slowest alternative considered. In this sense, it is an appropriate choice for a regression algorithm to solve similar forecasting problems. Since ensuring effective communication is now a key capability for businesses, public and non-governmental organizations, the ability to model and predict outreach is an important skill. Leveragin an advanced algorithm such as those presented here enable its users to improve both reach and engagement and thus ensure better results in their public relations and internal communications efforts.

Figure 4
Link between forecast accuracy and computation time for online news sharing data



3. Housing Prices

In addition to the usual business processes of production, supply and maintenance, a number of organizations also have activities for the acquisition and management of tangible fixed assets, including real estate. A key problem in the acquisition and management of the real estate is the determination of the correct price of the property. In standard practice, this often happens with the help of a dedicated expert appraiser, who combines objective market data with subjective judgment and adjustments so as to arrive at a definitive assessment. The main problem with this approach is that it largely depends on human judgment, which makes it relatively expensive, slow and more difficult to scale. These factors require relatively rare or even one-off property valuations, although a dynamic market environment often suggests significant dynamics.

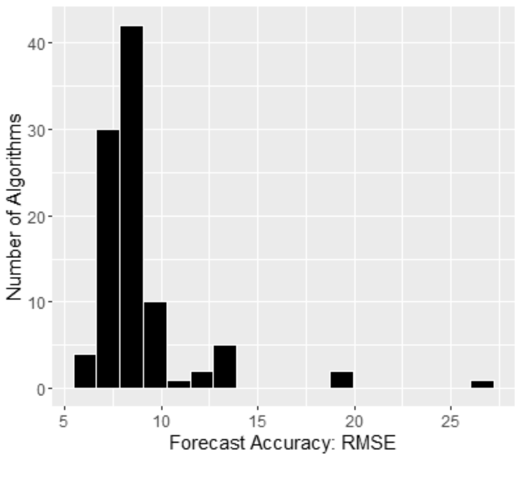
In this sense, it is important to improve the process of real-time assessment and forecasting of the value of a property, which will allow its rational management and the unlocking of maximum business values in operations with it. Such a task is of interest both to organizations involved in the purchase and sale of real estate (construction companies, brokers, etc.) and to organizations involved in the financing and securitization of transactions (banks, non-bank credit institutions, funds, etc.).

The main business problem is the need to determine and regularly update the correct price of a property, so that the organization can assess the effectiveness of potential disposal transactions with it (purchase, sale, rental, etc.), as well as to predict future price dynamics in order to minimize the risk of unexpected losses due to unfavourable trends in the real estate market. To tackle this task, we use data provided by Yeh & Hsu (2018). They (ibid.) compared a proposed new approach which they refer to as the comparative quantitative approach with four other alternatives – two hedonic pricing approaches, a multidimensional linear regression algorithm and a neural network model. Yeh and Hsu (2018) show the superiority of this comparative quantitative approach over the investigated alternatives. This study is a useful first step, but we recognize the need to expand the set of potential predictive algorithms and investigate their accuracy.

The database itself consists of 414 observations of real estate transactions in Taipei (Taiwan) against seven different characteristics - date, years of construction of the building, distance to the subway station, number of nearby shops, geographical coordinates (latitude and longitude), unit price area. For convenience in modelling, we divide the date into two components – year and sequence of the transaction within the calendar year (combination of day and month). The target variable is the unit price, and by its nature, it is a continuous numerical variable, whose prediction is of significant business interest. On the basis of real estate data, 106 alternative models from the field of machine learning are evaluated, and their predictive accuracy is investigated through their forecast errors. A histogram of the predictive accuracy as measured by the root mean square errors is presented in Figure 3. It is striking that the vast majority of methods have a realization of RMSE in the range of 7 to about 9. The best algorithms among the subjects have a predictive accuracy of RMSE <6.5. and those with the worst results can reach RMSE values of over 25. It is noteworthy that we notice a significant grouping around the average accuracy, but a few algorithms show extremely weak performance. Few algorithms have a slightly better performance than the average one. In this sense, there is potential business value in testing and selecting the best ones, even if this only allows one to avoid the very worst performers.

Figure 5

Histogram of forecast accuracy for housing prices data



The ten approaches with the lowest root mean squared error are presented in Table 3. It is immediately apparent that seven of them are different applications of the random forest family. They all report extremely good forecasting accuracy, with the root of the squared error of the predictions being in the range of RMSE = 6.46 to RMSE = 7.10. The other three non-random forest algorithms are 2 based on a kernel function and one based on a Gaussian process, whose predictive accuracy is about RMSE = 7.10.

Table 3

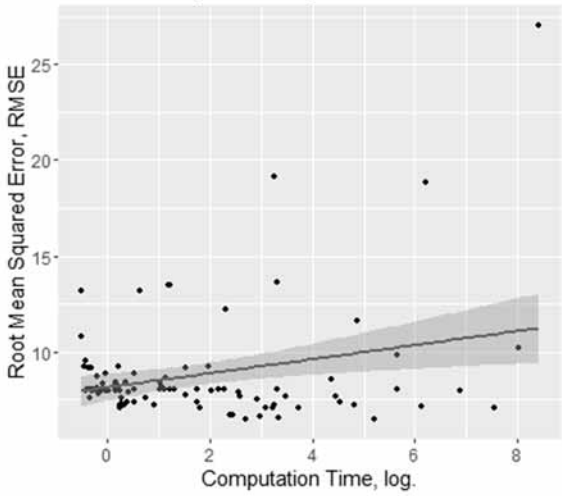
Forecast accuracy of the top ten algorithms for housing prices data

Algorithm	Method	Mean Error, ME	Root Mean Squared Error, RMSE	Mean Absolute Error, MAE	Complexity Measure
Regularized Random Forest	RRF	-0.750	6.459	4.831	64.8%
Quantile Random Forest	qrf	0.001	6.470	4.695	5.2%
Regularized Random Forest	RRFglobal	-0.832	6.568	4.890	9.9%
Random Forest	ranger	-0.878	6.600	4.884	7.0%
Parallel Random Forest	parRF	-0.936	6.689	4.965	3.8%
Random Forest	ranger	-0.943	6.689	4.908	4.0%
Radial Basis Function Kernel Regularized Least Squares	krlsRadial	-0.435	7.068	5.388	14.8%
Bayesian Additive Regression Trees	bartMachine	-0.795	7.076	5.353	11.1%
Random Forest by Randomization	extraTrees	-0.948	7.081	5.137	7.8%
Gaussian Process with Polynomial Kernel	gaussprPoly	-0.501	7.082	5.443	2.1%

The measure of complexity, accounting for the proportional computation time with respect to the most resource-intensive algorithm, also varies within very wide limits. The best method – that of the regularized random forest is only 35% faster than the slowest in the sample.

Figure 6

Link between forecast accuracy and computation time for housing prices data



On the other hand, the second best – the quantile random forest – is nearly 20 times faster than the most resource-intensive, and the difference in predictive accuracy between the two is almost imperceptible. This underlines that even with this type of task, it is possible to find the optimal point between the benefits and the costs of calculating given algorithms. Moreover, the calculation speed of the algorithm underlines the possibility of switching from asynchronous to synchronous operations, i.e. from model calculations and their subsequent use and future updates to real-time analytics that are used and trained simultaneously. On the more practical side, this level of accuracy provides encouraging implications for asset management. It enables organizations to generate up-to-date estimates of the actual value of immobile assets and thus make business decisions on their maintenance and disposal. On the flip-side, it enables individual owners and buyers to get a close approximation of the actual housing prices, thus decreasing the information asymmetry in the housing market.

4. Support Ticket Processing

Within the normal activities of the organization, problems and disturbances sometimes arise, which affect the end customers of its services - both internal and external. For their most effective removal as a good business practice, it is necessary to introduce a system for registration and processing of complaints by users of the service. The effectiveness of this system is essential both in terms of financial resources and as a source of knowledge about the constant evolutionary learning of modern organizations. The incident processing system is a direct link between the organization and the market and the various stakeholders using its products and services. In addition to being a source of information, effective incident processing is important in terms of providing a good level of service, maintaining the organizational reputation and caring for customer loyalty, all of which have potentially direct effects on the financial results of the structure. In this sense, the fast and efficient processing of signals from customers is a key competence for the modern organization and an important part of its usual operations.

The key business task in the processing of incident tickets by customers is ensuring a quick and efficient processing that ultimately leads to a cost-effective solution of the consumer problem. So one of the main indicators to be monitored is the time required from registration to closing of the given incident ticket, avoiding excessively long signal processing before reaching a satisfactory resolution. To this end, it is necessary to identify the main drivers of the potential delay, allowing it to be minimized. In addition, it is important to be able to automatically assess whether a signal is at risk of excessive delay, so as to direct organizational resources and attention to it. Therefore, the target variable for the forecast is the time required to process a given incident ticket.

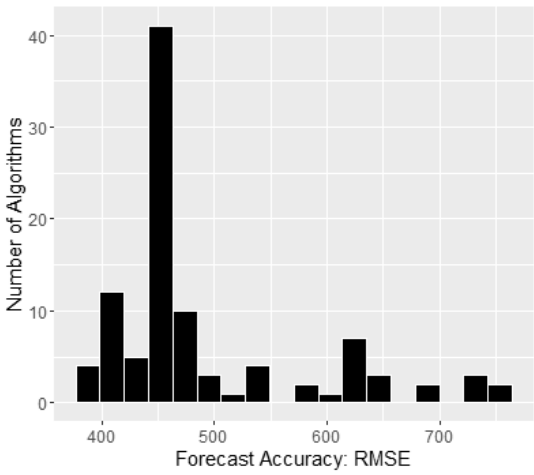
To investigate this task, we use data from an incident processing information system provided by Amaral et al. (2018). The data contains standard information, including identification numbers of incidents, responsible employees, processors, the request itself, and more. In essence, the information also includes a wide range of characteristics of the incidents themselves – the status of the incident, activity, number of assignments and reopens, reporting channel, symptoms and logical location, the presence of an error message and so on. Estimates of the severity of the problem are also available – those include effect strength, urgency, priority, double priority check, as well as completion characteristics – date and time, termination code, identifier of the terminating employee. The number of characteristics in the original data set is 36, and the observations are 141,712 in total.

Amaral et al. (2018) focus their work on selecting the appropriate characteristics to include in the forecasting model, leveraging annotated transition systems. They employ three experiments with different selection approaches, finding that peer reviews are better than some, but not all, automated approaches. The authors derive this conclusion using the mean absolute percentage error (MAPE) to measure forecast accuracy, but do not report other metrics for forecast accuracy. We build on the results in this article (ibid.) regarding the choice of characteristics to be included in the model and present a wide range of alternative forecasting approaches, some of which lead to higher prognostic accuracy compared to the approach proposed in the original article.

During information processing, we removed data that simply contains identifiers of various circumstances (agents, randomly generated request numbers, etc.), as they follow either a clearly predetermined sequence or are the result of chance and are thus unlikely to aid forecasting. Additionally, we remove all variables that have a large number of missing observations (over 30%), ending up with removing seven variables. Five of them have more than 98% missing, so their removal does not lead to significant loss of information. Based on the time stamps for the moment of signal registration and the moment of its termination, we construct the target variable – the necessary processing hours. After this, we select the complete set of observations that do not have missing data, which leads to a set of data with 108,247 observations that is used for subsequent analysis. This set contains 1 target and 18 explanatory variables, which is slightly more than the fifteen used by Amaral et al. (2018). Again, we estimate the 109 different algorithms under study on the incident management data. Of them, 102 reach convergence and can be used to generate forecasts.

Figure 7

Histogram of forecast accuracy for incident processing data



The distribution of forecast accuracy measured by the root mean squared error (RMSE) is visually displayed by the histogram in Figure 5. The vast majority of algorithms register an RMSE of about 450, but we also observe the long tail of the distribution. It shows the presence of a small number of algorithms with particularly poor performance. Under specific circumstances, it is possible to observe the tendency of certain algorithms to do particularly well with some tasks, which is offset by a particularly poor performance of others, which is known as the "No Free Lunch Theorem" of optimization, and we probably observe exactly those effects here.

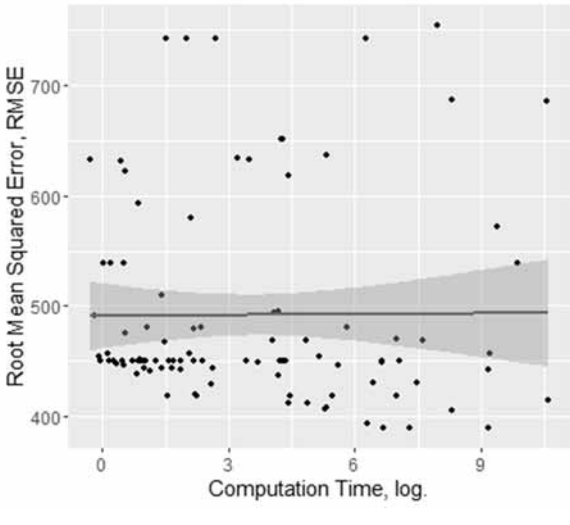
Table 4

Forecast accuracy of the top ten algorithms for incident processing data

Algorithm	Method	Mean Error, ME	Root Mean Squared Error, RMSE	Mean Absolute Error, MAE	Complexity Measure
Regularized Random Forest	RRFglobal	-23.95	388.82	183.81	2.0%
Regularized Random Forest	RRF	-24.22	389.64	184.59	13.2%
Random Forest	rf	-24.30	389.75	184.61	1.1%
Parallel Random Forest	parRF	-28.04	393.42	188.02	0.7%
Random Forest by Randomization	extraTrees	-23.79	405.33	187.67	5.6%
Random Forest	ranger	-24.27	406.17	188.66	0.3%
eXtreme Gradient Boosting	xgbLinear	-17.56	407.67	188.13	0.3%
eXtreme Gradient Boosting	xgbTree	-10.43	411.34	191.64	0.2%
Boosted Tree	bstTree	-8.33	412.05	185.00	0.1%
Bayesian Additive Regression Trees	bartMachine	-7.16	414.76	196.90	55.5%

Figure 8

Link between forecast accuracy and computation time for incident processing data



The ten methods with the highest predictive accuracy are summarized in Table 5. Eight of the ten methods are varieties of the random forest algorithm. The *RRFglobal* method achieves the highest forecast accuracy, with its RMSE = 388.8, followed by five other random forest implementations with comparable error rates.

The list of the ten most accurate algorithms also includes two methods for extreme gradient boosting. In terms of the complexity measure, the most accurate method has relatively good levels of resource intensity, being 50 times faster than the slowest alternative. If computational optimization is required, it is also possible to choose a method with comparatively high accuracy and lower computational needs.

This application of machine learning algorithms leverages a new type of data coming from the so-called process-aware information systems. This data enables an organization to fully map its business process – in this case, IT support – using information from its transactional data systems. The combination of process-aware data and advanced algorithmic modelling allows for a very precise business process management and control. This, in turn, enables gains in productivity and profitability.

5. Logistics Demand

One of the main operational concerns for the modern organization are the unexpected changes in the market environment, which can lead to excess or insufficient resources to meet demand. Many business activities are characterized by significant demand dynamics, and the main task consists of being able to meet the peak moments, minimize costs at low points and smooth the internal work process. Thus, it is especially important to be able to forecast demand and in the presence of significant deviations from its usual level – to take appropriate management action. The risk, in this case, is twofold. On the one hand, an excessively high level of demand can lead to an inability to serve potential customers and hence – lost revenue. On the other hand, excessively low levels of demand lead to unused resources such as employees and equipment and thus generate costs for the organization. Effective demand forecasting and management is also key to applying a flexible approach to resource allocation and helps to implement good management practices such as timely delivery (JIT), minimum product development and the application of the principles of flexible management (Agile).

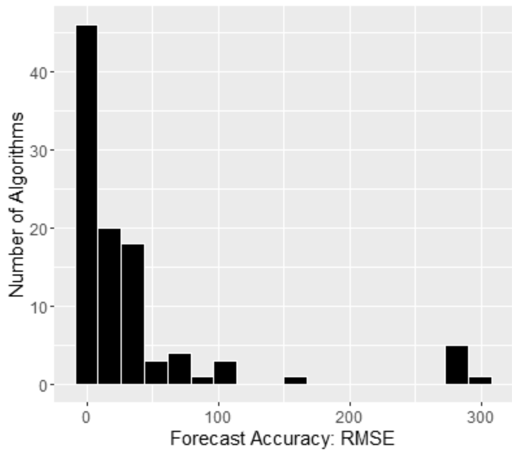
The business problem in demand analysis is the accurate forecasting of expected demand levels, which allows for taking the relevant measures should excessive deviations from the usual or expected load occur. The key goal is to correctly predict the peaks and troughs at which the organization should take appropriate action. To study this task, we use data from a large Brazilian logistics company provided by Ferreira et al. (2016). The database contains information for the orders received from the given company for a period of 60 days, and some of the variables are not described in detail in order to preserve sensitive business information. Data are available for the week and day of the order, for what part of the orders are defined as priority ones, how many of them are of a certain type (A, B, C), what part comes from the public sector, from the banking sector or from a traffic control centre. The total number of explanatory variables is 12. The array also contains information on the total daily number of orders, which is an indicator of the levels of market demand and so is the target variable of the current task.

We emphasize that the database provided by Ferreira et al. (2016) is extremely small, containing observations for only 60 days. Thus, the analysis of the database with alternative algorithms will show the possibilities for applications of machine learning methods over a very limited sample of data. This allows us to trace whether it is appropriate to apply big data algorithms instead of traditional statistical methods in cases of small sample analysis. Ferreira et al. (2016) use this data to demonstrate the application of an artificial neural network of the multilayer perceptron type. They found that this model was suitable for modelling demand even for a small number of observations. Based on their results, here we look at the applications of a much wider range of algorithms and further investigate how those numerous alternatives compare against each other.

Using the data under investigation, we estimate 109 alternative forecast models with different forecasting algorithms. Their predictive accuracy, measured by the root mean squared error, is presented in Figure 4. The histogram clearly shows that the small number of observations leads to unstable predictions. There are a large number of algorithms that reach extremely low RMSE values close to zero, which is likely symptomatic of overfitting. We also notice a small number of algorithms with extremely poor performance - RMSE in the neighbourhood of 300. This result is significantly worse than the naive forecast for each value (e.g. the average for the sample).

Figure 9

Histogram of forecast accuracy for logistics demand data



As a potential way to avoid relying on models with uncertain performance outside the training sample, we elect to not consider models with RMSE <1, emphasizing that in the vast majority of such cases, RMSE in fact approaches zero. Among the other algorithms, we can highlight the ones with the highest predictive accuracy, the top ten presented in Table 4. The most optimal models are the multinomial adaptive regression splines (MARS) with RMSE = 2.51 and RMSE = 2.61 for the implementation with pruning. A number of representatives of the family of generalized linear models as well as support vector machine implementations perform well in this task.

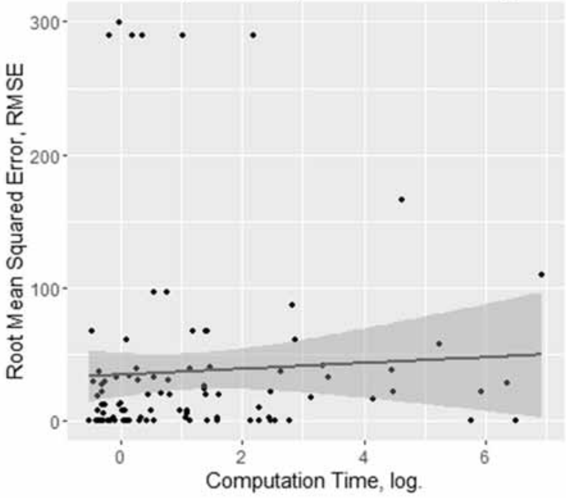
Table 5

Forecast accuracy of the top ten algorithms for logistics demand data

Algorithm	Method	Mean Error, ME	Root Mean Squared Error, RMSE	Mean Absolute Error, MAE	Complexity Measure
Bagged MARS	bagEarth	-0.056	2.509	2.076	1.12%
Bagged MARS using gCV Pruning	bagEarthGCV	0.107	2.605	2.185	0.48%
glmnet	glmnet	-1.188	2.780	2.384	0.14%
Gaussian Process with Linear Kernel	gaussprLinear	-0.966	2.843	2.357	0.28%
Bayesian Regularized Neural Networks	brnn	-0.910	2.853	2.374	0.08%
Support Vector Machines with Linear Kernel	svmLinear2	-3.153	5.865	5.129	0.07%
Support Vector Machines with Linear Kernel	svmLinear	-3.153	5.865	5.129	0.28%
Gaussian Process with Polynomial Kernel	gaussprPoly	-3.527	7.345	5.660	0.16%
Least Squares Support Vector Machine with Polynomial Kernel	svmPoly	-3.146	7.660	6.936	0.29%
Boosted Generalized Linear Model	glmboost	-2.669	7.863	6.994	0.10%

Figure 10

Link between forecast accuracy and computation time for logistics demand data



Again, we recognize that the best performing algorithms are not the most resource-intensive ones. This allows one to make an optimization decision for the optimal model for solving similar forecasting problems based on predictive accuracy and the measure of complexity. Given the accuracy and the complexity measure of the multinomial adaptive regression splines (MARS) without pruning, we conjecture it to be the algorithm with the best overall performance for this data. We emphasize that with such a small sample size, all results should be interpreted with caution and that practical applications should seek to avoid such limited samples.

The problem of demand forecasting is a classical microeconomic task that hails back to the roots of the discipline. At least since the 1950s, there have been rigorous applications of econometric methods for demand planning. However, the proliferation of machine learning approaches, and the explosion of data has spelt a new era for this common problem. The results here indicate that advanced modelling yields satisfactory forecasts even under significant data restrictions. The results are of obvious value – the ability to forecast demand enables the firm to more efficiently plan the resources to meet them, thus minimizing costs during downturns and maximizing profit during spikes. This has both bottom-line implications as well as reputational ones.

V. Discussion

1. Statistical and Modeling Insights

The article reviews the application of forecasting algorithms in five typical business situations – to model workplace absenteeism, corporate online communication, housing prices, logistics demand, and incident processing. The main task here is to reach an optimal tradeoff between potentially negative and potentially positive consequences through forecasting, planning and control of potential deviations. 109 different forecasting methods are tested in this paper in order to aid the analytic part of this task.

The first major insight of the study is to underline the importance of choosing the appropriate forecasting algorithm for modelling individual business domains. This is clearly visible both in the results above and in the rank distribution of predictive accuracy (see Figure 6). All the algorithms under investigation are ranked in accordance with their position in terms of accuracy. Thus, the one with the highest accuracy, measured by the root mean squared error, gets a rank of 1. The algorithm with the second highest precision is ranked as 2, and so on to the most inaccurate algorithm, which gets a rank of 109. Figure 6 presents a histogram of the averaged ranks of the studied methods. There are a small number of algorithms that consistently generate forecasts with a high level of accuracy in all the tasks studied. On the other hand, this is not the case with most algorithms, most of them having an average rank between 25 and 75. One also sees a small number of particularly unsuccessful algorithms with an average rank of over 80. This underlines the utility and benefits of applying statistical algorithm selection before application to specific problems.

Second, we recognize that there are groups of algorithms that generate consistently good results. These are often different applications of the random forest method, with different implementations of the random forest approach being very close to each other. This justifies the use of highly optimized implementations of the method (such as the *ranger* implementation in the R language) as they allow significant savings of computational resources virtually without loss of predictive accuracy. Classical statistical methods such as linear regression generally register much lower forecast accuracy than machine-learning methods. Thus, it may be appropriate to replace or at least supplement them with more complex algorithms from the field of machine learning. A notable exception to this is situations where extremely small samples are analyzed. We do not observe a difference between the linear regression model and alternative algorithms in such cases, as have seen in the logistics demand task. On the other hand, the accuracy measures in such a sample are unstable and less reliable, and this result should be interpreted with caution. In any case, all the methods under investigation perform better when trained with more data.

Figure 11

Histogram of average forecast accuracy across all datasets

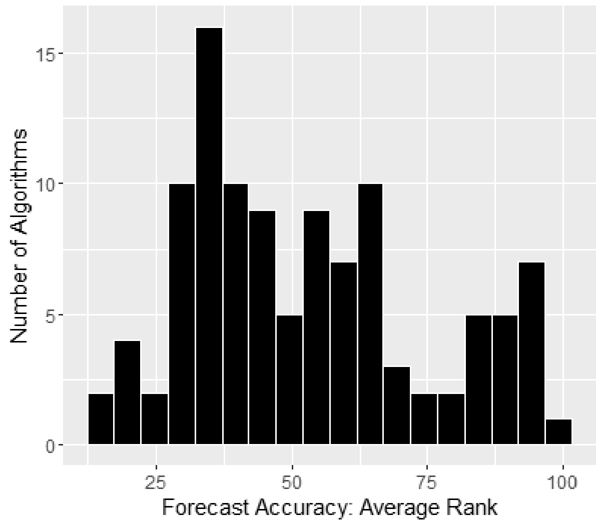
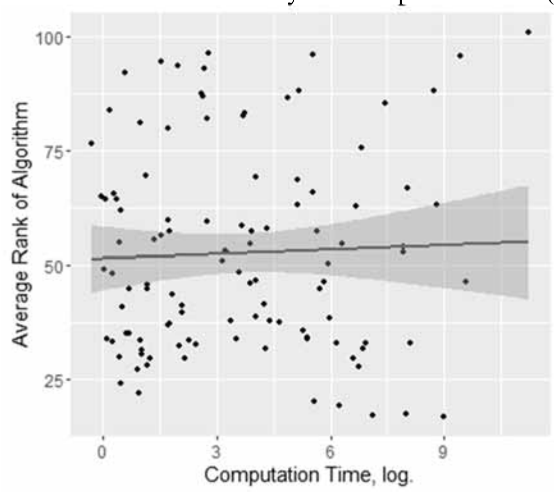


Figure 12

Link between forecast accuracy and computation time (log)



Third, we note that there is virtually no relationship between forecast accuracy and the required computational resources for the algorithms under study. Figure 7 presents graphically the relationship between the average rank of the algorithm and the log of the time required to calculate it. The visual inspection fails to reveal a clear link between these two characteristics of the considered algorithms. We formally test the relationship by regressing the mean rank on the required calculation time. Although the coefficient reaches statistical significance at levels of 5%, its magnitude is extremely small: $\beta = 0.0006$ and the explained variance is only 4%, indicating that it lacks practical significance. The almost imperceptible link between calculation time and predictive accuracy yet again underlines the possibility of choosing an algorithm with high accuracy and relatively low computational needs.

2. Economic Implications and Limitations of the Results

Each of the five business domains under investigation, in essence, presents different problems of information. Ranging from workplace absenteeism, online new sharing, modeling housing prices all the way into support tickets processing and demand planning, business activity is plagued by asymmetric information. In the domains under study, modern organization tend to only collect a limited set of data regarding those and are unable to fully predict an outcome under interest. For example, workplace data must be at an individual level and key employee characteristics such as a propensity for excessive absences could be difficult or even unethical to obtain. On the other hand, data on assets (such as housing) or processes (such as communication, IT support or demand planning) could be impossible or impractical to gather. In the end, the organization is left with only incomplete data that must be used to alleviate the economic problem of asymmetric information. Turning data into actionable knowledge thus requires algorithmic modelling to compensate for what is missing, and to quantify the impact and uncertainty of the decisions at hand. The article has thus reviewed leading algorithms to perform these tasks in a forecasting setting. Results are encouraging – businesses may reliably use a few out-of-the-box forecasting approaches such as random forests to generate useful projections. Those algorithms seem to work very well in medium-sized datasets where tens to hundreds of thousands of observations are described by tens of different predictors – a common situation in actual practice. Those results are in consonance with other research in the field (see, e.g., Fernandez-Delgado et al., 2014) and give businesses the confidence to base their analytic architecture on leveraging those algorithms. Moreover, the approach presented here gives a clear indication of the tradeoff between computational needs and forecasting accuracy, providing a primitive estimate of the cost-benefit tradeoff. This can be further elaborated and adapted for practical applications. The breadth of domains covered also hints at the generalizability of the results.

The limitations of the current study should also be noted and taken into account when interpreting its results. First, we are looking at five specific business situations that are described with a relatively sparse set of features. The specifics of the domain may have implications for the resulting forecast accuracy values. However, the robustness of top performers hints that these effects may be relatively small. More importantly, a sparse dataset will benefit more from sophisticated modelling that compensates for the missing data. Thus, the performance of forecasters may differ in more information-rich environments, with some studies indicated that deep neural networks are preferable in such settings (Chiroma et al., 2018). Second, this article has looked at five different business domains and tried to outline the best performers across the board. However, the well-known no free lunch theorem states that there is no single best tool for every optimization problem (Cao et al., 2019). Thus, it may very well be the case for a specific business need is best met by a custom approach that is less suitable for other domains. Finally, results have tended to focus on traditional business activities such as forecasting employee behaviour, success of communication, asset pricing, process efficiency and demand planning. This study has not looked at novel forecasting and machine learning tasks such as image or voice recognition that are increasingly important for modern data-driven organizations. The investigation of algorithm effectiveness in such tasks will have to be the subject of further research.

VI. Conclusion

This study is in partial response to the concern posed by Fernandez-Delgado et al. (2014) whether we really need hundreds of classification algorithms. The same question, probably even more poignant, can be asked if we need hundreds of alternative regression algorithms for forecasting. Taking 109 statistical and machine learning algorithms, this research aims to investigate their applicability, speed, and accuracy to five typical microeconomic problems: excessive workplace absenteeism, online communication success, valuation of real estate asset prices, forecasting sharp changes in market demand, and improving customer support. The results are stark – a small number of algorithms tend to outperform the rest, with the random forest family being a particularly strong candidate.

Moreover, results are dispersed – while there are some clear winners, there is a large number of algorithms with mediocre or sub-optimal performance. Therefore, it does make sense for the analyst or researcher to spend time and resources for optimal algorithm selection. Of particular note is the fact that traditional econometric methods such as the multiple linear regressions are rarely among the top performers. Finally, the tradeoff between computational complexity and forecast accuracy is not clear-cut. There are some methods that are not among the most computationally intensive ones but still reliably produce highly accurate results. This clearly shows the possibility to make an optimal economic decision about the type of algorithms and methods used.

In short, this article aims at exploring a large set of primarily machine learning algorithms and testing their applicability to typical tasks in economics and business. Those methods, and particularly members of the CART family, have proven to deliver accurate forecasts suitable for both academics and practitioners in the fields of economics and business. With the advent of big data, the econometric toolbox will have to expand so that its methods scale well to large volumes, are able to produce meaningful conclusions and results, and improve our understanding of underlying process drivers. The introduction of more advanced machine learning methods is a good first step in this direction. This necessitates more research in both methodology and substantive modelling questions but, in the end, holds the promise to expand the scope and depth of problems economists can handle with confidence.

References

Alawadi, S., Mera, D., Fernández-Delgado, M., Alkhabbas, F., Olsson, C. M., & Davidsson, P. (2020). A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. – Energy Systems, pp. 1-17.

Amaral, C. A., Fantinato, M., Reijers, H. A., Peres, S. M. (2018). Enhancing Completion Time Prediction Through Attribute Selection. – In: Information Technology for Management: Emerging Research and Applications, Springer, Cham, pp. 3-23.

Cao, Y., Chen, T., Wang, Z., Shen, Y. (2019). Learning to optimize in swarms. arXiv preprint arXiv:1911.03787.

Chiroma, H., Abdullahi, U. A., Alarood, A. A., Abdulhamid, S. M, Gabralla, L. A., Rana, N., Shuib, L., Hashem, I., Gbenga, D., Abubakar, A., Zeki, A., Herawan, T. (2018). Progress on artificial neural networks for big data analytics: a survey. – IEEE Access, 7, pp. 70535-70551.

Chiu, C. W. J., Harris, R. D., Stoja, E., Chin, M. (2018). Financial market volatility, macroeconomic fundamentals and investor sentiment. – Journal of Banking & Finance, 92, pp. 130-145.

Christensen, P., Gillingham, K., Nordhaus, W. (2018). Uncertainty in forecasts of long-run economic growth. – Proceedings of the National Academy of Sciences, 115(21), pp. 5409-5414.

Cronin, D., McQuinn, K. (2020). Are official forecasts of output growth in the EU still biased?. – Journal of Policy Modeling.

Debnath, K. B., Mourshed, M. (2018). Forecasting methods in energy planning models. – Renewable and Sustainable Energy Reviews, 88, pp. 297-325.

Fernandes, K., Vinagre, P., Cortez, P. (2015). A proactive intelligent decision support system for predicting the popularity of online news. – In: Portuguese Conference on Artificial Intelligence, Springer, Cham, pp. 535-546.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. – The journal of machine learning research, 15(1), pp. 3133-3181.

Ferreira, K. J., Lee, B. H. A., Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. – Manufacturing & Service Operations Management, 18(1), pp. 69-88.

Ferreira, R. P., Martiniano, A., Ferreira, A., Ferreira, A., Sassi, R. J. (2016). Study on daily demand forecasting orders using artificial neural network. – IEEE Latin America Transactions, 14(3), pp. 1519-1525.

Friedman, J., Hastie, T., Tibshirani, R. (2001). The elements of statistical learning (Vol. 1). Springer, Berlin: Springer series in statistics. doi. 10.1007/978-0-387-84858-7.

Hassani, H., Yeganegi, M. R., Gupta, R. (2021). Historical Forecasting of Interest Rate Mean and Volatility of the United States: Is there a Role of Uncertainty?. – Annals of Financial Economics, 2050018.

Hassink, W. (2018). How to reduce workplace absenteeism. IZA World of Labor.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond.

Hyndman, R. J. (2006). Another look at forecast-accuracy metrics for intermittent demand. – Foresight: The International Journal of Applied Forecasting, 4(4), pp. 43-46.

Hyndman, R. J., Athanasopoulos, G. (2014). Forecasting: principles and practice. Texts. doi. 10.1111/j.1467-856X.2005.00178.x.

Koller, M., Feldmaier, J., Diepold, K. (2019). A Comparison of Prediction Algorithms and Nexting for Short Term Weather Forecasts. arXiv preprint arXiv:1903.07512.

Kuhn, M. (2008). Building predictive models in R using the caret package. – Journal of Statistical Software, 28(5), pp. 1-26.

Kunze, F., Wegener, C., Bizer, K., Spiwoks, M. (2017). Forecasting European interest rates in times of financial crisis–What insights do we get from international survey forecasts?. – Journal of International Financial Markets, Institutions and Money, 48, pp. 192-205.

Lago, J., De Ridder, F., De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. – Applied Energy, 221, pp. 386-405.

Makridakis, S., Hibon, M. (2000). The M3-Competition: results, conclusions and implications. – International journal of forecasting, 16(4), pp. 451-476. doi. 10.1016/S0169-2070(00)00057-1.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. – Journal of forecasting,1(2), pp. 111-153. doi. 10.1002/for.3980010202.

Makridakis, S., Hibon, M., Moser, C. (1979). Accuracy of forecasting: An empirical investigation. – Journal of the Royal Statistical Society. Series A (General), pp. 97-145. doi. 10.2307/2345077.

Makridakis, S., Spiliotis, E., Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. – International Journal of Forecasting, 36(1), pp. 54-74.

Martiniano, A., Ferreira, R. P., Sassi, R. J., Affonso, C. (2012, June). Application of a neuro fuzzy network in prediction of absenteeism at work. – In: 7th Iberian Conference on Information Systems and Technologies (CISTI 2012), pp. 1-4. IEEE.

Mathai, A. V., Agarwal, A., Angampalli, V., Narayanan, S., Dhakshayani, E. (2016). Development of new methods for measuring forecast error. – International Journal of Logistics Systems and Management, 24(2), pp. 213-225.

Mei, D., Liu, J., Ma, F., Chen, W. (2017). Forecasting stock market volatility: Do realized skewness and kurtosis help?. – Physica A: Statistical Mechanics and its Applications, 481, pp. 153-159.

Prestwich, S., Rossi, R., Armagan Tarim, S., Hnich, B. (2014). Mean-based error measures for intermittent demand forecasting. – International Journal of Production Research, 52(22), pp. 6782-6791.

Salotti, J., Fenet, S., Billot, R., El Faouzi, N. E., Solnon, C. (2018). Comparison of traffic forecasting methods in urban and suburban context. – In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 846-853. IEEE.

Schaer, O., Kourentzes, N., Fildes, R. (2019). Estimating the market potential with pre-release buzz. – Lancaster University Management School, Management Science Paper Series, 1.

Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., Kamaev, V. A. E. (2013). A survey of forecast error measures. – World Applied Sciences Journal, 24(24), pp. 171-176.

Shih, H., Rajendran, S. (2019). Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation's blood supply. – Journal of healthcare engineering, 2019.

Tyralis, H., Papacharalampous, G., Langousis, A. (2020). Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. – Neural Computing and Applications, pp. 1-16.

Van der Laan, E., van Dalen, J., Rohrmoser, M., Simpson, R. (2016). Demand forecasting and order planning for humanitarian logistics: An empirical assessment. – Journal of Operations Management, 45, pp. 114-122.

Whitt, W., Zhang, X. (2019). Forecasting arrivals and occupancy levels in an emergency department. – Operations Research for Health Care, 21, pp. 1-18.

Yeh, I. C., Hsu, T. K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. – Applied Soft Computing, 65, pp. 260-271.