PREDICTIVE MODELS FOR INFANT MORTALITY IN THE STATE OF PARANÁ

Arnon Bruno Ventrilho dos Santos, Deborah Ribeiro Carvalho
Pontifícia Universidade Católica do Paraná (PUC-PR)
E-mails: asantos.quantum@gmail.com, drdrcarvalho@gmail.com

Abstract: The prediction of infant mortality guides preventive measures regarding deaths. The objective of this article is to evaluate models that could predict the early, late and post neonatal childhood mortality rate in the state of Paraná. Public available data were adopted for the period between 1996 and 2014. Predictive models for the period from 2015 to 2018 were compared, taking into account the mean absolute percentage error rate (MAPE). We also evaluated models for the period 2011 - 2014, given the possibility of comparing the expected results with the real value. We used three algorithms, namely: Linear Regression, Support Vector Machine Algorithm (SVM) and MultiLayer Perceptron (MLP). From the results, it was possible to identify greater precision in the models generated by MLPs, with a MAPE of 1% in 2015, 6% in 2016, 6% in 2017 and 8% in 2018, this model had an average accuracy of 30% higher if compared to the model that makes use of SVM (SMOreg), and in average it is 60% more accurate than the linear regression model. It is believed that increasingly refined adjustments in the MLP algorithm may further increase its accuracy and that the usage of other non-absolute quality measurements could demonstrate to which direction the prediction error is moving to.

Keywords: Linear Models, child mortality, predictions, time-series studies.

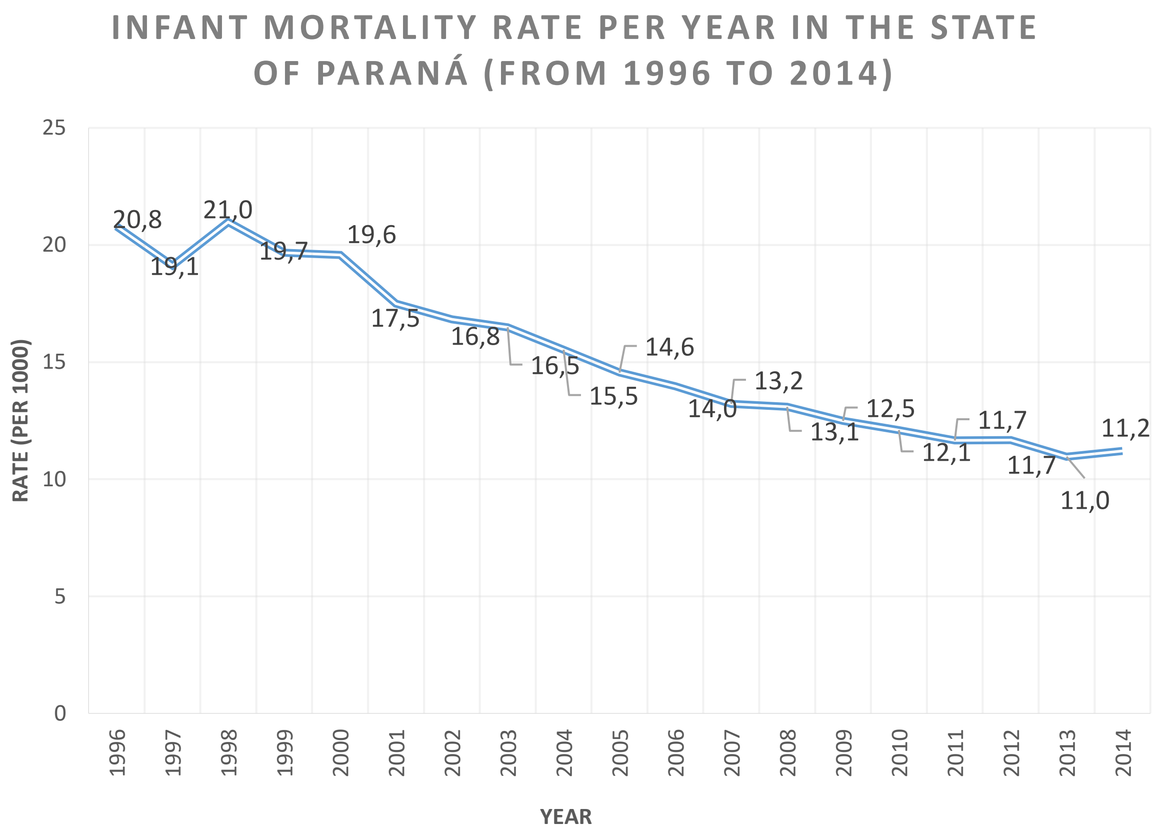
1. INTRODUCTION

Infant mortality represents a sensitive indicator of the health and living conditions of a population (United Nations Children's Fund, 1989). The infant mortality coefficient has the property to inform about the health of a population levels and summarizes the conditions of social, political and ethical welfare of a given social conformation (LOYAL & SZWARCWALD, 1996).

The ability to understand, predict and prevent these deaths is considered to be an important strategy for improving the quality of life and health. Researches that aims to understand and describe the possible causes of infant mortality in Brazil support the identification of the most relevant features thus enabling decision-making for preventive actions. This is the case for (Carvalho et. al, 2010), that through data mining, conducted an analysis of the characteristics that are more strongly linked to infant mortality in the State of Paraná between the years of 2000 and 2004. They found with this job relevant rules about data in Information System of live births (SINASC)-BIRTH CERTIFICATES, the mortality system – SIM, and the infant mortality research – SIMI (Carvalho et. al, 2010).

The State of Paraná, has shown a drop in infant mortality rate (Graph 1), although in 2014 that rate was just above the national average.

Graph 1: infant mortality rate in the State of Paraná between the years of 1996 and 2014



Fonte: TABNET - SIM/SINASC[1], adapted by the author.

While this downward trend is positive, the infant brazilian mortality rate remains high (United Nations Children's Fund, 2014), about 14.4 per 1000 live births.

In 2013, at an event on the evolution of policies against child mortality in the world, Brazil was presented as one of the countries that have managed to reduce significantly the rate of mortality in childhood through effective strategies, which, according to United Nations Children's Fund (UNICEF), mainly the social integration policies, public health policies and to better watch the awareness of society about the subject (United Nations Children's Fund, 2014).

This fall of approximately 70% in 25 years, should be seen as a direct consequence of actions exposed by the UNICEF in all units of the Federation. These actions result in fast and relevant impact on mortality rates.

However, it is not the objective of this study to assess actions that impact in reducing these rates, but rather assess which predictive computational model best identifies the trend in mortality rates in the State of Paraná in order to provide a model able to facilitate preventive decision making and thus better orient actions that accelerate the fall of these figures the levels each time closer to those defined by UNICEF as ideals..

This work aims therefore to identify evaluate predictive methods for infant mortality in the State of Paraná between the years of 1996 and 2014

[1] Available at: http://tabnet.datasus.gov.br Accessed in February 2016.

2. METHOD

For this work, initially, were selected through TABNET system all events of infant death in the State of Paraná between the years of 1996 and 2014 for the ICD-10, without any additional filters, organizing the data in table format, where the years are distributed in rows and the discrete variable (number of deaths) in a column. Subsequently, new data were generated following the same guidelines, differing, however, the subdivision of the Neonatal deaths deaths between early and late neonatal in order to verify if the predictive models used in this work would perform similarly to the timeseries when this series was divided into two new groups, as mentioned previously. Both files were then exported in standard .CSV (acronym for "comma seperated values"-comma-separated values). In order to obtain the infant mortality rate (which represents the number of infant deaths for every 1000 live births), it was also extracted from the TABNET database the number of live births for the period of interest to the study, also in ".csv" format. This was aimed at the extraction numbers and its later combination with the mortality data in order to obtain the infant mortality rate per 1000 births.
With the data of mortality and of live births, the infant mortality rate was calculated using Microsoft Excel.
To obtain this number, it was divisided the amount of children's deaths in the year by the number of live births, then the result was multiplied by 1000.
With these rates calculated, it was then possible to organize the results in a proper file for the algorithms and the tool used in this work, (". arff" format). 70% of the records of the datasets were used for training of these algorithms, while the remaining 30% were used to test the model learned from the training (Dobbin & Simon, 2011).

Extraction and data items

The data generated from TABNET, refer to 18 consecutive years (from 1996 to 2014), leading to the first data a file extraction with 32 rows of records. These data, as well as the next extraction, were imported into the local computer in the format ".csv".
It was extracted from the DATASUS the number of live births in the period of 1996 to 2014 in order to be able to calculate the rate of infant mortality (represented by 1000 live births) for future extractions, which refer to morbidity data.
The first imported file has several information about the query, in addition to the data of interest. For this reason, the software "Microsoft Excel 2013" was used to exclude information that are not interesting for the target analysis of this work, keeping, however, the column labels (years and number of deaths), in addition to the number of deaths spread throughout the time series (1996-2014). Finally, there is a file in the format ".csv" with 19 lines, distributed among quantitivas continuous variables (years) and discrete quantitative (number of deaths). This data series is also demonstrated by the Graph 1.
For the second extraction and dataset, which also refers to the period of 1996 to 2014, it was extracted the number of infant deaths, also the age group of individuals who came to death, in order to verify that the prediction can be more accurate when the data set is with stratified data between early neonatal deaths, late neonatal and post neonatal ones.
The file obtained from the criteria defined in the second extraction, was also imported in the ".csv" format. After all, the file generated registrates 138 rows, where 12 of these rows are information on the query held at TABNET system database. The data of interest for this work is located on the other 126 lines, organized between age groups ranging from "1 minute old," up to "1 year old" and distributed among quantitative continuous variables (years), discrete quantitative variables (number of deaths) and categorical variables ordinals (age group). These files have been compared with the number of live births to achieve the child mortality rate.

Summary measures

TABNET system describes in detail the time of death in the "detailed age" attribute. When the amount of deaths over the years by detailed age is extracted, however, there is a huge amount of columns (125 in all), which describe age groups ranging from 1 minute to 11 months, 364 days, 23 hours and 59 minutes of life. In order to improve the analysis, we conducted a dimensionality reduction process of the attribute"detailed age". The many dimensions of data sets increases the complexity of the techniques of manipulation and degrades the performance of data mining algorithms. To decrease these effects, dimensionality reduction techniques are intended to represent a set of dimension data in another space of smaller dimension, looking to maintain the characteristics of the set (Miyashiro, Camila, & Barioni, 2016).
With this process, the 125 columns were turned into 3 dimensions: early neonatal deaths, where deaths take place in live births between 0 and 6 days (6 days, 23 hours and 59 minutes), late neonatal, where deaths take place in live births between 7 and 27 days of life (27 days, 23 hours and 59 seconds) and post neonatal where the deaths take place in live births between 28 and 364 days (364 days , 23 hours and 59 seconds of life). This classification also obeys what is established in the International statistical classification of diseases (CID) and Health-related issues in your tenth review – ICD 10 (who, 2004, 2004a and 2004b), which is subdivided into the children's deaths in these 3 classes, as described by VIANNA (2007): infant mortality Early Neonatal, late Neonatal and infant mortality infant mortality post Neonatal.
The ".csv" file contains in its first three lines information on the query held at TABNET, these information were excluded from the document to ensure that only the necessary information for this work were used.

Bias risk-assessment

No method of assessing the risk of bias was used in this work.

3. Data mining and its Algorithms

Data mining

According to (FAYYAD, WEIR and DJORGOVSKI, 1993) data mining is a part of the process of KDD (acronym for Knowledge Discovery in Databases) which consists in the application of data analysis and discovery of algorithms that produce rules or models about the data collected. For this experiment, the data collected through DATASUS were organized by using the tool "Microsoft Excel 2013", keeping, however, data integrity, preserving the values, as described earlier, adapting these data for the interpretation of machine learning tool that will reap the standards and will show the results from the application of the algorithms mentioned below.

The Algorithms

Since this work considers the historical series of infant deaths between 1996 and 2014, representing all the time series database in the source for the situation in question, the regression in data mining task will be used to assess some of the time series prediction algorithms with the help of "WEKA" tool (Frank, Hall & Witten, 2011), aiming to identify which one offers better performance in terms of accuracy (identified through the Mean Absolute Percentage Error-MAPE)(Adhikari & Agrawal, 2013) when trying to predict values for the next periods of the

series, as well as the ones not yet seem by the algorithms (test-set). According to Fayyad (1996) "regression is a task to learn a function that maps a given item for a real estimated prediction variable". That is, what is desired is to use this strategy through a discrete variable (in this case, the total number of neonatal and perinatal deaths confirmed), to be able to describe the future behavior of the data series.
According to (Adhikari & Agrawal, 2013), prediction is possible through various time series models, such as stochastic models, multilayer perceptrons (MLP) and support vector machines (SVM).
For the prediction of the two datasets used in this study we tested some of the algorithms mentioned by (Adhikari & Agrawal, 2013), namely: Linear Regression algorithm, the MLP algorithm (Multilayer Perceptron) and a variation of the SVM (Support Vector Machines) algorithm, SMOreg algorithm (Sequential minimal optimization). These three were chosen, as being the ones available for testing in Weka time-series forecast extension.
The table below performs the comparison between the main features of the algorithms studied:

Table 1: Mapping of algorithms and its main features

Algorithm	# of Parameters	Adaptability	Generalization	Custo computacional
Linear Regression	Low	Low	Medium	Low
SVM	Low	High	High	High
MLP	High	High	High	HIgh

Number of parameters : This characteristic is important because it represents the complexity of the algorithm being tested. In order to put the tested algorithms in apparent equality conditions, the values used for parameters were those defined as standard in the machine learning tool we used (Frank, Hall & Witten, 2011). One of the three algorithms, the linear regression algorithm, is the one with the least amount of parameters to adjust is, therefore the less complex algorithm of three checked.
In accordance with the principle of parsimony (Adhikari & Agrawal, 2013) always the model with the smallest possible number of parameters must be chosen as the model that provides the best representation of the number parsed. That's because, according to this same principle, a smaller amount of parameters requires a smaller chance of errors and interferences in real results – principle similar to that of Occam's Razor in logic. We shall test this principle, for this specific problem, on the results section.
The SVM algorithm demanded adjustments to "optimizer" and "kernel" parameters. The "optimizer" is the learning algorithm and the "kernel" is the function that indicates how the algorithm will sort the results obtained in the hyperplane. These settings are very important to maximizing the ability of generalization of the algorithm.
For the SVM, the optimizer used was the "RegSmoImproved" (Bhattacharyya et al., 1999) and the “polynomial” kernel.
For the MLP algorithm, there is a lot of tweaking of parameters. There is no common heuristics to achieve the best settings other than the process of successive attempt to fit a model that has minimized the most common problems (such as the minimum-location, overfitting, underfitting), which indicate that the ability of generalization of the algorithm has been compromised. For this algorithm, the settings of parameters were as follows: batchSize = 100, hiddenLayers = (the sum of the number of neurons in the layer of input and output, divided by 2), learningRate = 0.3, momentum = 0.2.

Adaptability: One of the great advantages of using MLPs applied to time series is its ability to non-linear modelling, without any prior knowledge about the nature and distribution of the

statistical problem. The model, although demands a lot of tweaking of parameters, it adapts itself based on data learned during training, for this reason (Bhattacharyya et al., 1999) mentions that models based on artificial neural networks are data-driven and content self-adaptive by nature.
A similar conclusion applies to SVM. An important feature of this algorithm is that the training is equivalent to solving a linear quadratic optimization problem, so that the solution applying SVM is always unique and overall great, (Bhattacharyya et al., 1999). This algorithm has been used much for temporal problems since its inception in 1995.
The linear regression algorithm does not have great ability to adapt to the nature proposed by the problem because it's a linear analysis of a problem that is in its core non-deterministic, i.e. may or may not follow a linear trend. As statistical model, does not identify implicit characteristics identified during training in the other two methods, which can reduce its predictive capacity in comparison to other more robust methods.

Generalization: The linear regression model lacks large capacity of generalization because of its linear nature. Regression models are mathematical models that relate the behavior of a variable Y with another X when the function f that relates two variables is of type:

$$f\ (X)\ =\ a\ +\ b\ X$$

In this way, new values set for the series are found based on the appropriateness of the line which best describe the behavior of the dependent variable with respect to time.
On the other hand, the objective of the SVM algorithm is to use the principle of structural risk minimization (SRM) to find a decision rule with good generalization ability (Bhattacharyya et al., 1999) in this way, we can affirm that SVM can solve the problem of overfitting effectively – maximizing its generalization ability (Lin et al., 2007).
In contrast with its large amount of parameters, the MLP algorithm has excellent generalization capabilities since its parameters are properly adjusted to the problem in question and have minimized the risk of overfitting/underfitting and local minimum acceptable levels. Generally, this is a task that demands a large number of trials. (Lin et al., 2007) seem to indicate otherwise, mentioning that MLPs have good capacity of learning, but that usually cause overfitting/underfitting, impoverishing the ability of generalization.
In accordance with (Adhikari & Agrawal, 2013), we didn't visualize this kind of problem in the final results of our experiment, so that the model proposed here seems to be well adjusted to the problem.

Computational cost: This property of machine learning algorithms is something of great importance when selecting a model that can provide accurate results. In other words, it doesn't matter if the capacity of a given predictive model is little more than another if your computational cost is much higher.
This measure of computational cost is based on what was displayed throughout this experiment and what was presented by the literature.
By far, the fastest processing algorithm is the "linear regression". Because it is a large volume of calculations but relatively simple (resolution of linear equations for each point in the series), processors of today can easily find solutions quickly.
The MLP algorithm has a higher computational cost given the structure of the algorithm (Benhra, Benkachcha, & Hassani, 2013) and the amount of parameters that pass over the network until the optimal value have been found.
The algorithm of greater computational cost is the SVM. (Adhikari & Agrawal, 2013) mentions that this is the main drawback of this approach, because for increasing training, more computer resources are necessary to find a solution, which increases the temporal complexity of the

problem. Seeking to solve this problem, (Bhattacharyya et al., 1999) have developed the SmoReg algorithm, variation of the SVM which massively reduces the computational cost by dividing the problem to be calculated into subproblems that are if complementing over processing.

4. Results

Results of the summaries

After summarization measures described in the methodology section, this is the results obtained from the data extracted from TABNET.

It is also important to mention that a total of 665 deaths could not be classified according to the methodology previously described, because its registry does not have valid information about the age of death attribute, which caused invalidated the classification. Given the low percentage (1% of total registries), we believe that these did not interefered on the prediction algorithms.

Table 2 – Relation between the total infant deaths and births, obtaining infant mortality rate from it

Year	Births	Deaths	Rate (%)
1996	196.429	4083	20,79
1997	192.757	3688	19,13
1998	185.378	3889	20,98
1999	186.675	3671	19,67
2000	179.462	3514	19,58
2001	167.270	2925	17,49
2002	165.125	2779	16,83
2003	157.333	2595	16,49
2004	159.636	2479	15,53
2005	160.324	2332	14,55
2006	153.598	2146	13,97
2007	147.554	1950	13,22
2008	151.092	1978	13,09
2009	149.217	1864	12,49
2010	152.051	1840	12,1
2011	152.902	1781	11,65
2012	153.945	1796	11,67
2013	155.758	1707	10,96
2014	159.915	1791	11,2

Table 3-Summary of the second extraction Results

Age / Year	Early Neo Natal	Rate (%)	Late Neo Natal	Rate (%)	Post Neo Natal	Rate (%)	Unclassified Deaths
1996	1.973	10,04	480	2,44	1.668	8,49	38
1997	1.872	9,71	449	2,33	1.391	7,22	24
1998	1.871	10,09	457	2,47	1.589	8,57	28
1999	1.975	10,58	453	2,43	1.270	6,80	27
2000	1.873	10,44	432	2,41	1.209	6,74	0
2001	1.547	9,25	394	2,36	1.029	6,15	45
2002	1.485	8,99	441	2,67	880	5,33	27
2003	1.364	8,67	399	2,54	895	5,69	63
2004	1.287	8,06	408	2,56	808	5,06	24
2005	1.222	7,62	401	2,50	727	4,53	18
2006	1.137	7,40	354	2,30	676	4,40	21
2007	1.032	6,99	332	2,25	634	4,30	48
2008	1.093	7,23	312	2,06	602	3,98	29
2009	994	6,66	329	2,20	581	3,89	40
2010	1.036	6,81	314	2,07	540	3,55	50
2011	951	6,22	333	2,18	547	3,58	50
2012	956	6,21	318	2,07	566	3,68	44
2013	898	5,77	302	1,94	547	3,51	40
2014	974	6,09	325	2,03	541	3,38	49
Total	**25.540**	**8,17**	**7.233**	**2,13**	**16.700**	**5,34**	**665**

The data obtained from the two datasets were ultimately transferred to plain text documents (notepads of Microsoft Windows 10) and saved in two new "format file. arff" (acronym for Attribute-Relation File Format), the "etaria.arff" (representing the second TABNET data extraction) and another called "anos.arff" (which represents the first data extraction) preserving its formatting with comma separation when possible (first extraction) and inserting commas manually between the datasets in the second extraction.
Both datasets, (both the first and second data extraction), already in the "arff" format, still received additional data that describe the types of variables contained in the dataset for the WEKA Machine learning tool to be able to read the files properly. The results of the two extractions (already preprocessed) will, from now on, be referred to as "First" and "Second" datasets.

First dataset

```
@relation Mortalidade_Infantil_Anos

@attribute Anos date 'yyyy'
@attribute nObitos numeric

@data

1996,20.79
1997,19.13
1998,20.98
1999,19.67
2000,19.58
2001,17.49
2002,16.83
2003,16.49
2004,15.53
2005,14.55
2006,13.97
2007,13.22
2008,13.09
2009,12.49
2010,12.1
2011,11.65
2012,11.67
2013,10.96
2014,11.2
```

Second dataset:

```
@relation Mortalidade_Infantil_Faixa_Etaria
```

```
@attribute Anos date 'yyyy'
@attribute NeoPrecoce numeric
@attribute NeoTardia numeric
@attribute PosNeo numeric

@data

1996,1959,467,1657
1997,1867,445,1376
1998,1867,452,1570
1999,1976,447,1248
2000,1879,433,1202
2001,1535,386,1004
2002,1469,438,872
2003,1342,386,867
2004,1282,401,796
2005,1221,397,714
2006,1128,355,663
2007,1017,325,608
2008,1080,310,588
2009,989,319,556
2010,1032,304,504
2011,932,329,520
2012,950,309,537
2013,883,301,523
2014,964,315,512
```

Validation of results

It is said that a measurement is carried out with high precision if the errors are small (when compared to the original values) (UNIVERSITY of COIMBRA, 2003), therefore to assess the accuracy of the algorithms tested here, it was used the mean absolute percentage error (MAPE) under the test dataset created to validate the results.
Below are the results of each of the experiments in an attempt to predict the mortality rate from 2015 to 2018, having the last 4 years of the series (2011-2014) as the test dataset. In order to facilitate the identification of major and minor error experiments, the smallest MAPE identified was listed in blue, and the MAPE in red:

Results of the models
Table 4 – MAPE of the Algorithms for the first dataset (based on total Infant mortality rate)

Comparison with the test dataset:

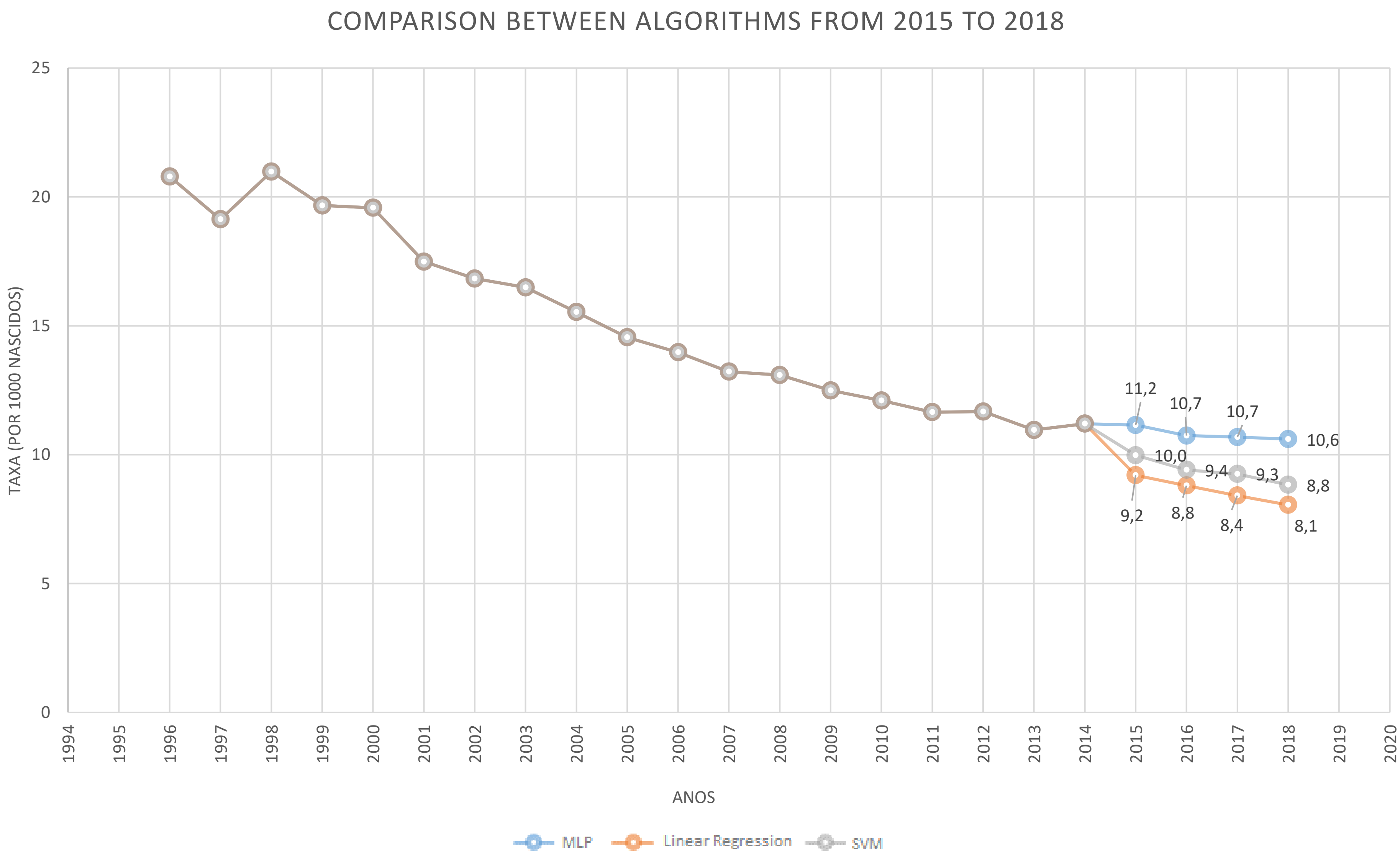
		Year			
	Algorithm	2011	2012	2013	2014
Infant Mortality	Linear Regression	6%	7%	8%	12%
	SMO	4%	4%	8%	8%
	MLP	2%	4%	7%	7%

Error prediction for the coming years:

		Year			
	Algorithm	2015	2016	2017	2018
Infant Mortality	Linear Regression	4%	6%	7%	12%
	SMO	7%	6%	7%	5%
	MLP	1%	6%	6%	8%

These results are also displayed graphically, as below:

Graph 2 – comparison between models for the prediction of infant mortality rate in the State of Paraná in the period between 2015 and 2018.


COMPARISON BETWEEN ALGORITHMS FROM 2015 TO 2018
TAXA (POR 1000 NASCIDOS)
ANOS
0
5
10
15
20
25
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
11,2
10,7
10,7
10,6
10,0
9,4
9,3
8,8
9,2
8,8
8,4
8,1
MLP
Linear Regression
SVM

Interesting to analyze the expected results for the total mortality rate with those results that have already been disclosed by the Ministry of health of the State of Paraná, which registered in 2015 a infant mortality rate of 10.9 per 1000, while the MLP model described in this work anticipated this result with 97.7% accuracy. Preliminary results for 2016 show that the model had similar accuracy (98,1%).
Another interesting result in the first dataset and partially confirms the conclusion of this study, is when it is attempted to forecast the total number of infant deaths in the State of Paraná from 2011 to 2014, using actual data for comparison between predicted x performed, according to the table below:

		Predictions					
Year	Deaths	Linear Reg.	% Abs. Error	MLP	% Abs. Error	SMO	% Abs. Error
2011	1781	1710,149	4%	1792,043	1%	1746,433	2%
2012	1796	1646,2085	9%	1761,791	2%	1699,61	6%
2013	1707	1598,6201	7%	1722,847	1%	1655,286	3%
2014	1791	1539,5769	16%	1696,261	6%	1617,845	11%

The MLP also has the smallest MAPE in predicting the number of infant deaths (without considering the rate per live births)

Table 4 – MAPE for the second dataset

		Year			
	Algorithm	2015	2016	2017	2018
Early NeoNatal Mortality	Regressão Linear	6%	6%	8%	11%
	SMO	15%	15%	17%	12%
	MLP	23%	25%	27%	22%
Late NeoNatal Mortality	Regressão Linear	9%	12%	12%	6%
	SMO	6%	5%	6%	8%
	MLP	7%	4%	4%	6%
Post NeoNatal Mortality	Regressão Linear	26%	28%	23%	27%
	SMO	17%	19%	18%	20%
	MLP	12%	13%	13%	14%

5. Conclusions

From the series of data harvested from the DATASUS through TABNET, after the due data preprocessing and predictive analytics with the WEKA tool, it was possible to verify that the algorithm that more accurately predicted the mortality rate was the MLP, which shown the smallest MAPE in 3 of the 4 analyses, even when the prediction was under the total number of deaths. The Linear regression was the less accurate in 3 of 4 reviews, but being the best in 1 of these 4, not confirming the principle of parsimony for the tests in this work. The SMO has remained balanced in all the series analysed. The patterns have repeated when we tried to

predict the total number of infant deaths in the 2011-2014 period, i.e. the MLP had the highest accuracy, followed by the SMO and Linear Regression.
Although MLP is the method that demand greater number of trials and tests (adjust the size of the network, number of neurons, learning rate, momentum) to perform with greater accuracy of prediction tasks (which demand the intervention of a specialist for these adjustments), this algorithm has been shown to have greater accuracy than the others here and certainly can serve as a good assistance to predict time series in different domains.
Linear regression proved to be ineffective because it is not the ideal algorithm for non-linear datasets and non-deterministic problems, as quoted by (Nau, 2014).
SMO also proved to be effective, though less than the MLP, which indicates that can it be used as additional task or replacing the MLP. Another interesting aspect of this approach is that it is not necessary to fit so many parameters to obtain accurate predictions.
This study could not accurately verify the amount of children's deaths in the State of Paraná in the year 2016 because this information is not yet available at TABNET. As an improvement, it is suggested that larger studies with predictive analysis to be carried out by testing other settings of parameters for the algorithms here tested in order to obtain maximum precision and contribute to decision-making in the area of health, particularly in the context of child mortality and other datasets with time series.

References

Agrawal, R. K, Adhikari, R; (2013). **An Introductory Study on Time Series Modeling and Forecasting**. Lambert Academic Publishing, Germany, p.1–p.68.

Andrade C, Szwarcwald C. **Análise espacial da mortalidade neonatal precoce no Município do Rio de Janeiro**, 1995-1996. Cad. Saúde Pública. 2001; 17(5): p.1199-p.1210

Balas, V.E; Mastorakis, N.; Popescu L.P; Popescu, M.L; **Multilayer Perceptron and Neural, WSEAS Transactions on Circuits and Systems**, Issue 7, Volume 8,University of Craiova, Romania, Bulgary, 2009, p.579 – p.588.

Benhra, J., Benkachcha, S., & H, E. H. (2013). **Causal Method and Time Series Forecasting model based on Artificial Neural Network**. International Journal of Computer Applications, 75(August). p.16 – p- 18

Bhattacharyya et al. (1999**). Improvements to Platt's SMO Algorithm for SVM Classifier Design, Department of Mechanical and Production Engineering, University of Singapore**, 1999, Singapore, p.117 – p-21

Cardon, A.; Müller, D.N; **Introdução às Redes Neurais Artificiais**, Universidade Federal do Rio Grande do Sul, Instituto de Informática, Curso de Pós-Graduação em Ciência da Computação, Porto Alegre, Brasil, 1994. p.4-p.9.

Dobbin, K. K; Simon, R. M; **Optimally splitting cases for training and testing high dimensional classifiers**, Univeristy of Georgia, BioMed Central, USA, 2011. P.2-p.8

Dos Santos, E.M.; **Teoria e Aplicação de Support Vector Machines à Aprendizagem e Reconhecimento de Objetos Baseado na Aparência**, Universidade Federal da Paraíba, Paraíba, Brasil, 2002. p.6 – p12.

Garner, S.R; Weka: **The Waikato Environment for Knowledge Analysis, Department of Computer Science**, University of Waikato, Hamilton, New Zealand, 1995. p.2-p.7.

Frank, E.; Hall M.A; Witten, I. H, **Data Mining, Practical Machine Learning Tools and Techniques**, Third Edition, Elsevier Inc., 2011. p.247, p.455.

Leite, A.; Silva, A. **Mortalidade Infantil: indicador das condições de vida das populações**, Brasil, p.1

Lin, K., Lin, Q., Zhou, C., & Yao, J. (2007). **Time Series Prediction Based on Linear Regression and SVR**, 2006, p.7–p.10.

Miyashiro, D. U., Camila, M., & Barioni, N. (2016). **Estudo de Técnicas de Redução de Dimensionalidade**, p.1–p.3.

Nau, R. (2014). **Notes on Linear Regression**., Fuqua School of Business, Duke University, EUA, p.1–p.18.

Vianna, R. C. X. F., Moro, C. M. C. de B., Moysés, S. J., Carvalho, D., & Nievola, J. C. (2010). **Mineração de dados e características da mortalidade infantil**. Cadernos de Saúde Pública, 26(3), p.535–p.542.