

Previsão do Índice de Desenvolvimento Humano e da expectativa de vida nos países da América Latina por meio de técnicas de mineração de dados

Forecasting the human development index and life expectancy in Latin American countries using data mining techniques

Celso Bilynkiewicz dos Santos¹

Luiz Alberto Pilatti²

Bruno Pedroso¹

Deborah Ribeiro Carvalho³

Alaine Margarete Guimarães¹

Abstract *The predictability of epidemiological indicators can help estimate dependent variables, assist in decision-making to support public policies, and explain the scenarios experienced by different countries worldwide. This study aimed to forecast the Human Development Index (HDI) and life expectancy (LE) for Latin American countries for the period of 2015-2020 using data mining techniques. All stages of the process of knowledge discovery in databases were covered. The SMOReg data mining algorithm was used in the models with multivariate time series to make predictions; this algorithm performed the best in the tests developed during the evaluation period. The average HDI and LE for Latin American countries showed an increasing trend in the period evaluated, corresponding to $4.99 \pm 3.90\%$ and 2.65 ± 0.06 years, respectively. Multivariate models allow for a greater evaluation of algorithms, thus increasing their accuracy. Data mining techniques have a better predictive quality relative to the most popular technique, Autoregressive Integrated Moving Average (ARIMA). In addition, the predictions suggest that there will be a higher increase in the mean HDI and LE for Latin American countries compared to the mean values for the rest of the world.*

Key word *Forecasting, Data mining, HDI, Life expectancy, Latin America*

Resumo *A previsibilidade de indicadores epidemiológicos pode contribuir na projeção de variáveis dependentes, auxiliar em tomadas de decisões para sustentar ou não políticas públicas e justificar o cenário vivido pelos países e o mundo. O artigo tem por objetivo prever o Índice de Desenvolvimento Humano e a expectativa de vida nos países latino-americanos no período de 2015 a 2020, utilizando técnicas de mineração de dados. Foram percorridas as etapas do processo Descoberta de Conhecimento em Base Dados. Adotaram-se para previsões modelos desenvolvidos com séries multivariadas através do algoritmo de mineração de dados SMOReg, que apresentaram melhor desempenho em testes desenvolvidos durante o experimento. As médias do Índice de Desenvolvimento Humano e da expectativa de vida nos países latino-americanos tendem a aumentar no período analisado, respectivamente, $4,99 \pm 3,90\%$ e $2,65 \pm 0,06$ anos. Experiências multivariadas possibilitam maior aprendizagem dos algoritmos, aumentando sua precisão. As técnicas de mineração de dados apresentaram melhor qualidade nas previsões em relação à técnica mais popular, ARIMA. As previsões sugerem média de crescimento do IDH e EV dos países latino-americanos maiores que a mundial.*

Palavra-chave *Previsão, Mineração de dados, IDH, Expectativa de vida, América Latina*

¹ Setor de Ciências Biológicas e da Saúde, Universidade Estadual de Ponta Grossa. Av. General Carlos Cavalcanti 4748, Uvaranas. 84030-900 Ponta Grossa PR Brasil. bilynkiewicz@uepg.br

² Universidade Tecnológica Federal do Paraná. Curitiba PR Brasil.

³ Programa de Pós-Graduação em Tecnologia em Saúde, Pontifícia Universidade Católica do Paraná. Curitiba PR Brasil.

Introdução

A maioria dos países latino-americanos passa por um processo de desenvolvimento humano muito semelhante, possivelmente em função do contexto histórico de suas emancipações políticas e características sociais e culturais. Este processo de desenvolvimento pode ser avaliado a partir do Índice de Desenvolvimento Humano (IDH), adotado pelo *United Nations Development Programme* (UNDP) para medir os avanços na qualidade de vida de um país¹⁻³ a partir da média geométrica entre indicadores de educação, saúde e renda⁴, que classifica a maior parte dos países da América Latina como países em desenvolvimento com IDH alto⁵.

Por apresentar dependência da componente “saúde”, medida por uma vida longa e saudável, o índice é amplamente utilizado em pesquisas da área de saúde⁶⁻¹³, enquanto sua componente decomposta mais utilizada neste tipo de estudos é a expectativa de vida (EV)¹⁴⁻²⁷, que também se encontra entre os indicadores mais utilizados para avaliar o desenvolvimento socioeconômico de um país.

A literatura apresenta um quantitativo muito grande de pesquisas recentes⁶⁻²⁷ que colocam a EV ou o IDH como variáveis balizadoras de estudos de saúde.

A previsibilidade do IDH ou de suas componentes pode auxiliar em tomadas de decisões governamentais, apoiando ou não políticas públicas, caso os valores reais correspondam às expectativas das previsões. Suas projeções também podem ser utilizadas em estudos prospectivos de diferentes áreas, entre elas a da saúde, para explicar o comportamento futuro de variáveis dependentes.

A literatura oferece uma variedade de técnicas de previsão, dentre as quais se destacam as previsões a partir de técnicas de Mineração de Dados (MD) aplicadas em diferentes áreas²⁸⁻³⁴, entre elas a da saúde^{33,34}. No entanto, não foi possível identificar estudos predizendo o IDH ou a EV dos países da América Latina.

Diante desta abertura na literatura, o presente estudo tem como objetivo prever o IDH e a EV nos países latino-americanos para o período de 2015 a 2020, a partir de seus dados históricos, utilizando técnicas de MD.

Pretende-se com este estudo contribuir na projeção destes indicadores utilizados em pesquisas epidemiológicas e, também, contribuir na avaliação de algoritmos e modelos empregados, a partir das comparações entre as previsões e as

tendências divulgadas pelo UNDP para o IDH em períodos anteriores às projeções e entre suas medidas de qualidade de previsões.

Materiais e métodos

A partir dos dados históricos do IDH de 188 países filiados ao UNDP, referentes ao período de 1990 a 2014, foram desenvolvidas todas as etapas do processo de Descoberta de Conhecimento em Base Dados (DCBD)³⁵, apresentadas nas próximas subseções. Durante o processo, foi definido o algoritmo e modelo com melhor desempenho para efetivar previsão do IDH e da EV nos 22 países latino-americanos filiados ao UNDP no período de 2015 a 2020.

A maioria das etapas do processo de DCBD foi realizada no ambiente de MD WEKA³⁶, utilizando a técnica de *Forecast*, através de uma interface de programação de aplicativos disponível a partir da versão 3.7.

Durante o processo, foi avaliado o desempenho de diferentes algoritmos baseados em funções. A partir do algoritmo com melhor desempenho, foram desenvolvidos modelos de previsão, e seus resultados, comparados com os últimos relatórios do UNDP, para definição dos modelos mais eficientes.

Para avaliar os resultados, utilizaram-se as medidas de qualidade das previsões das séries temporais: erro absoluto médio (MAE), erro quadrático médio (MSE), raiz quadrada do erro quadrático médio (RMSE), erro percentual absoluto médio (MAPE), precisão direcional (DAC), erro absoluto relativo (RAE), e raiz quadrada do erro quadrático relativo (RRSE). Também, foram realizados testes estatísticos de análise de variância, *T* de *Student* pareado por países, em diferentes momentos do processo de DCBD, adotando com nível de significância, $\alpha = 0,05$.

Paralelo ao DCBD, foi desenvolvido, através do software SPSS, um modelo de previsão utilizando-se a técnica de previsão mais popular, *Autoregressive Integrated Moving Average* (ARIMA) para, ao final dos testes, compará-la às técnicas de MD.

Pré-Processamento de mineração de dados

O pré-processamento iniciou-se com a obtenção dos dados de IDH e EV da base do UNDP³⁷, atualizada em 24 de julho de 2014 e do seu relatório referente ao IDH 2013³. Esta fonte de dados pode sofrer atualizações pontuais a qualquer

momento, e à divulgação de um novo relatório de IDH anual, as séries temporais podem sofrer atualizações mais significativas.

A partir destas fontes, foi desenvolvida uma base de dados específica, através do Microsoft Access, com as séries temporais referentes ao período de 1980 a 2013. Após a implementação desta base de dados, foi realizada a etapa de DCBD de “exploração da base de dados”, por meio da Linguagem de Consultas Estruturadas (*Structured Query Language – SQL*), resultando na estatística descritiva das séries temporais, incorporadas às Tabelas 3 e 4, apresentada na seção de Resultados.

Ao final do Pré-Processamento de MD, foram selecionadas 90 séries temporais do IDH para testes, separadas em dois lotes de dados, sendo o primeiro para testar a previsão do IDH 2013 e o segundo para previsão do IDH 2014, com dados anteriores ao período de previsão. Cada lote de dados foi utilizado para o desenvolvimento de: i) um modelo multivariado global (MMG), treinado com séries multivariadas correspondentes aos 188 países filiados à UNDP; ii) 22 modelos multivariados específicos (MME), treinados com grupos que variavam de dois a 45 países com capacidade explicativa do índice de cada país latino-americano; iii) 22 modelos univariados (MU), treinados com séries correspondentes a cada país da América Latina, resultando no total de 45 modelos por lote de dados. O MMG foi treinado com os dados de 188 países para aumentar a experiência do algoritmo.

Para o desenvolvimento dos MMEs, foram selecionados conjuntos de dados de IDH de países candidatos a previsores de cada país da América Latina (atributo meta). Os conjuntos de dados foram escolhidos através do algoritmo de seleção baseada em correlação: *Correlation-based Feature Selection (CFS)*³⁸, por meio do método de

validação cruzada. Este algoritmo prioriza conjuntos de atributos (variáveis independentes) muito relacionados com o atributo meta (variável dependente) e pouco relacionados entre si.

Mineração de dados

Nesta etapa, foi selecionando o algoritmo mais adequado ao estudo, testando os algoritmos pertencentes ao grupo de aprendizagem baseada em funções: *Least Median Squared, Linear Regression, Multilayer Perceptron, RBF Network, SMOReg, Gaussian Processes*.

Para poupar custos operacionais, os testes preliminares foram realizados apenas para previsão do IDH 2013, em que foi selecionado o algoritmo *SMOReg*³⁹, por apresentar os melhores resultados para as diferentes categorias de modelos (Tabela 1).

Finalizando a etapa de MD, 90 modelos foram desenvolvidos para conclusão dos testes com o uso do algoritmo *SMOReg*: dois MMGs, 44 MMEs e 44 MUs. Estes modelos foram comparados na previsão do IDH de 22 países da América Latina nos períodos de 2013-2014 para escolha do modelo de melhor desempenho (Tabela 2).

Pós-Processamento de Mineração de Dados

Os resultados dos modelos alimentaram uma base de dados que permitiu comparações entre os valores reais e as previsões, assim como entre as medidas de qualidade dos modelos. Os valores reais referentes ao IDH 2013 foram consultados no UNDP³, em 24 de julho de 2014, enquanto os valores referentes ao IDH 2014 foram consultados após sua atualização e divulgação realizada pelo UNDP⁵, em 14 de dezembro de 2015.

Depois de percorridas todas as etapas de DCBD para testar algoritmos e modelos, retor-

Tabela 1. MAE dos modelos de previsão desenvolvidos com algoritmo de MD baseado em funções.

	Previsão		IDH 2013				
	Modelo	MMG	SMM		UM		
	Estatística	μ	\pm	μ	\pm	μ	\pm
Algoritmo	SMOReg	0,0002 ^a	0,00005	0,0008 ^{to}	0,0005	0,0014 ^a	0,0007
	Gaussian Processes	0,0011 ^b	0,0008	0,0117 ^c	0,0057	0,0174 ^f	0,0088
	RBF Network	0,0165 ^d	0,0079	0,0161 ^d	0,0062	0,0160 ^e	0,0070
	Multilayer Perceptron	**	**	0,0021 ^b	0,005	0,0020 ^b	0,0007
	Linear Regression	**	**	*	*	0,0025 ^c	0,0028
	Least Median Squared	**	**	*	*	0,0044 ^d	0,0046

p < 0,05 em a < b < c < d < e < f (comparados por coluna). * não permitiu testes com conjuntos de países. ** não permitiu testes com todos países.

Tabela 2. Medidas de qualidade dos modelos desenvolvidos para testar a previsão do IDH dos países da América Latina filiados ao UNDP.

Horizonte de Previsão		IDH 2013-2014					
		MMG ^a		MME ^b		MU ^c	
Medida de Qualidade	Modelo	μ	\pm	μ	\pm	μ	\pm
		Estatística					
	Precisão direcional - DAC**	98,61	4,23	96,46	5,54	96,11	5,72
	Erro absoluto médio - MAE*	0,0002	0,00005	0,0008	0,0005	0,0014	0,0007
	Erro percentual absoluto médio - MAPE*	0,026	0,006	0,12	0,07	0,21	0,08
	Erro quadrático médio - MSE	0,0	0,0	0,0	0,0	0,0	0,0
	Erro absoluto relativo - RAE*	3,53	0,57	17,49	11,28	29,58	15,11
	Raiz quadrada do erro quadrático médio - RMSE*	0,0002	0,00005	0,001	0,001	0,002	0,001
	Raiz quadrada do erro quadrático relativo - RRSE*	3,51	0,65	25,15	13,80	40,20	17,38

* $p < 0,01$ em $a < b < c$. ** $p < 0,05$ em $a > b > c$.

naram-se as etapas anteriores, repetindo-se o processo para previsão do IDH e EV no período de 2015 a 2020, aplicando-se apenas o algoritmo e o modelo com melhor desempenho, SMOReg e MMU, respectivamente. Antes da efetivação da previsão, as séries temporais de 1980 a 2014 foram atualizadas em 14 de dezembro de 2015, pois, a cada divulgação de um novo relatório, a base de dados da UNDP³⁷ pode sofrer atualizações mais significativas³.

Resultados

A Tabela 1 apresenta estatísticas de resumo do MAE dos testes realizados para escolha do algoritmo de melhor desempenho, aferindo ao SMOReg o melhor desempenho.

A Tabela 2 apresenta estatísticas de resumo das medidas de qualidade dos testes de previsão do IDH 2013-2014 dos países da América Latina, para seleção do melhor modelo de previsão.

Verificou-se, por meio da Tabela 2, que os MMUs apresentaram as melhores medidas de qualidade das previsões, correspondendo aos maiores valores de DAC e menores tipos de erros (MAE, RMSE, MAPE, RAE e RRSE), em relação aos MMEs e MUs.

A Figura 1 apresenta o MAE de testes realizados com modelos de previsões adotados neste estudo e comparados com o método ARIMA.

Observou-se que os modelos desenvolvidos através de técnicas de MD apresentaram os menores erros absolutos em relação ao modelo ARIMA.

A Tabela 3 apresenta os cinco últimos pontos de observações das séries temporais históricas do

IDH nos países latino-americanos³⁷, as previsões do índice (2015 a 2020) estimadas pelo algoritmo SMOReg em MMGs, o resumo estatístico do índice mundial e na América Latina e seu o percentual de crescimento para o período de projeção.

A Figura 2 apresenta direções e previsões dos modelos (linhas tracejadas), tendências (linhas contínuas) divulgadas pelo UNDP^{3,5} e previsões do MMUs do IDH 2015-2020 de alguns países da América Latina, que apresentaram as melhores e piores medidas de qualidade das previsões, apesar de não apresentarem diferenças estatísticas significativas entre os valores nominais das previsões e os valores das tendências já divulgadas (2014 e 2015).

A Figura 3 apresenta a curva de crescimento do IDH mundial, com os valores médio, máximo, mínimo e de variâncias, registrados ao longo do período, além da média da América Latina e seus países com o maior e menor IDH.

A Tabela 4 apresenta os cinco últimos pontos de observações das séries temporais históricas da EV nos países latino-americanos³⁷, as previsões para variável (2015 a 2020) estimadas pelo algoritmo SMOReg em MMGs, resumo estatístico da variável no mundo e na América Latina e o seu percentual de crescimento para o período de projeção.

A Figura 4 apresenta a curva de crescimento da EV mundial, com os valores médio, máximo, mínimo e de variâncias, registrados ao longo do período, além da média dos países desenvolvidos, da América Latina e seus países como a maior e menor EV.

Discussão

Previsões do IDH

Em relação às previsões do IDH, ressalta-se que atualizações significativas dos índices de alguns países podem limitar o estudo. Segundo o UNDP¹, as estimativas internacionais e nacionais de dados podem apresentar inconsistência, uma vez que as agências de dados internacionais consultam os dados nacionais e, eventualmente, estimam dados inexistentes para comparação entre países. Em relação a estas atualizações, observaram-se diferenças significativas entre os valores de IDH, divulgados em 24 de julho de 2014³ e 14 de dezembro de 2015⁵.

Algumas previsões deste estudo, decorrentes dos testes para seleção do algoritmo e modelo, de

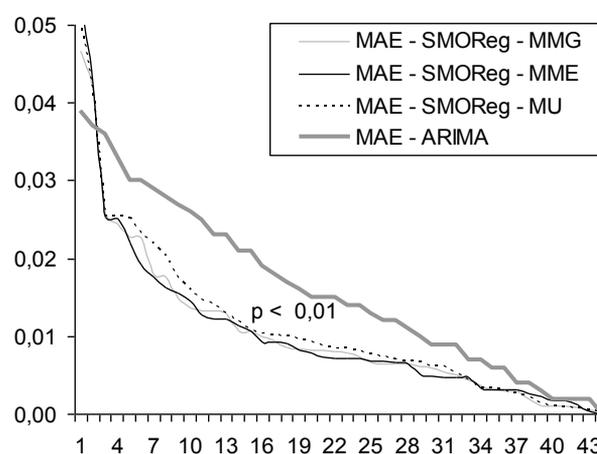


Figura 1. MAE acumulado por modelos, resultantes das previsões do IDH 2013-2014 dos países da América Latina.

Tabela 3. Últimos pontos de observações das séries temporais históricas do IDH dos países da América Latina, suas projeções para 2015 a 2020 e resumo estatístico do índice mundial e dos países latino-americanos.

País*	Últimos Pontos de Observação %					Horizonte de Previsão						Variação %	
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2015-2020	
ARG	0,811	0,818	0,831	0,833	0,836	0,843	0,85	0,858	0,865	0,874	0,882	5,50	
CHL	0,814	0,821	0,827	0,83	0,832	0,836	0,844	0,852	0,86	0,869	0,877	5,43	
URY	0,78	0,784	0,788	0,79	0,793	0,799	0,805	0,812	0,819	0,826	0,833	4,98	
PAN	0,761	0,759	0,772	0,777	0,78	0,786	0,793	0,8	0,807	0,815	0,823	5,49	
CUB	0,778	0,776	0,772	0,768	0,769	0,774	0,78	0,787	0,794	0,802	0,809	5,21	
CRI	0,75	0,756	0,761	0,764	0,766	0,77	0,776	0,783	0,789	0,796	0,803	4,78	
VEN	0,757	0,761	0,764	0,764	0,762	0,761	0,765	0,769	0,774	0,779	0,785	2,97	
MEX	0,746	0,748	0,754	0,755	0,756	0,76	0,766	0,772	0,778	0,784	0,791	4,63	
BRA	0,737	0,742	0,746	0,752	0,755	0,761	0,769	0,778	0,786	0,795	0,804	6,45	
PER	0,718	0,722	0,728	0,732	0,734	0,739	0,745	0,752	0,759	0,766	0,773	5,26	
ECU	0,717	0,723	0,727	0,73	0,732	0,734	0,739	0,743	0,748	0,754	0,759	3,68	
COL	0,706	0,713	0,715	0,718	0,72	0,724	0,73	0,736	0,743	0,75	0,757	5,19	
BLZ	0,709	0,711	0,716	0,715	0,715	0,715	0,717	0,72	0,723	0,726	0,728	1,86	
DOM	0,701	0,704	0,708	0,711	0,715	0,72	0,726	0,733	0,74	0,747	0,754	5,41	
PRY	0,668	0,671	0,669	0,677	0,679	0,681	0,686	0,691	0,696	0,701	0,707	4,05	
SLV	0,653	0,658	0,662	0,664	0,666	0,669	0,674	0,682	0,689	0,696	0,703	5,48	
BOL	0,641	0,647	0,654	0,658	0,662	0,667	0,671	0,677	0,683	0,689	0,695	4,96	
GUY	0,624	0,63	0,629	0,634	0,636	0,637	0,64	0,644	0,648	0,651	0,655	3,00	
NIC	0,619	0,623	0,625	0,628	0,631	0,636	0,643	0,651	0,66	0,668	0,676	7,13	
GTM	0,611	0,617	0,624	0,626	0,627	0,632	0,639	0,647	0,655	0,663	0,671	7,05	
HND	0,61	0,612	0,607	0,604	0,606	0,61	0,616	0,622	0,629	0,635	0,641	5,84	
HTI	0,471	0,475	0,479	0,481	0,483	0,487	0,492	0,497	0,502	0,507	0,513	6,13	
América Latina	μ	0,699	0,703	0,707	0,71	0,712	0,716	0,721	0,727	0,734	0,74	0,747	4,99
	±	0,081	0,081	0,083	0,083	0,083	0,083	0,083	0,084	0,084	0,085	0,086	3,90
Mundial	μ	0,679	0,683	0,687	0,689	0,691	0,695	0,7	0,705	0,711	0,717	0,723	4,63
	±	0,157	0,156	0,156	0,155	0,155	0,154	0,155	0,155	0,155	0,155	0,155	0,2

* código internacional de países com três letras.

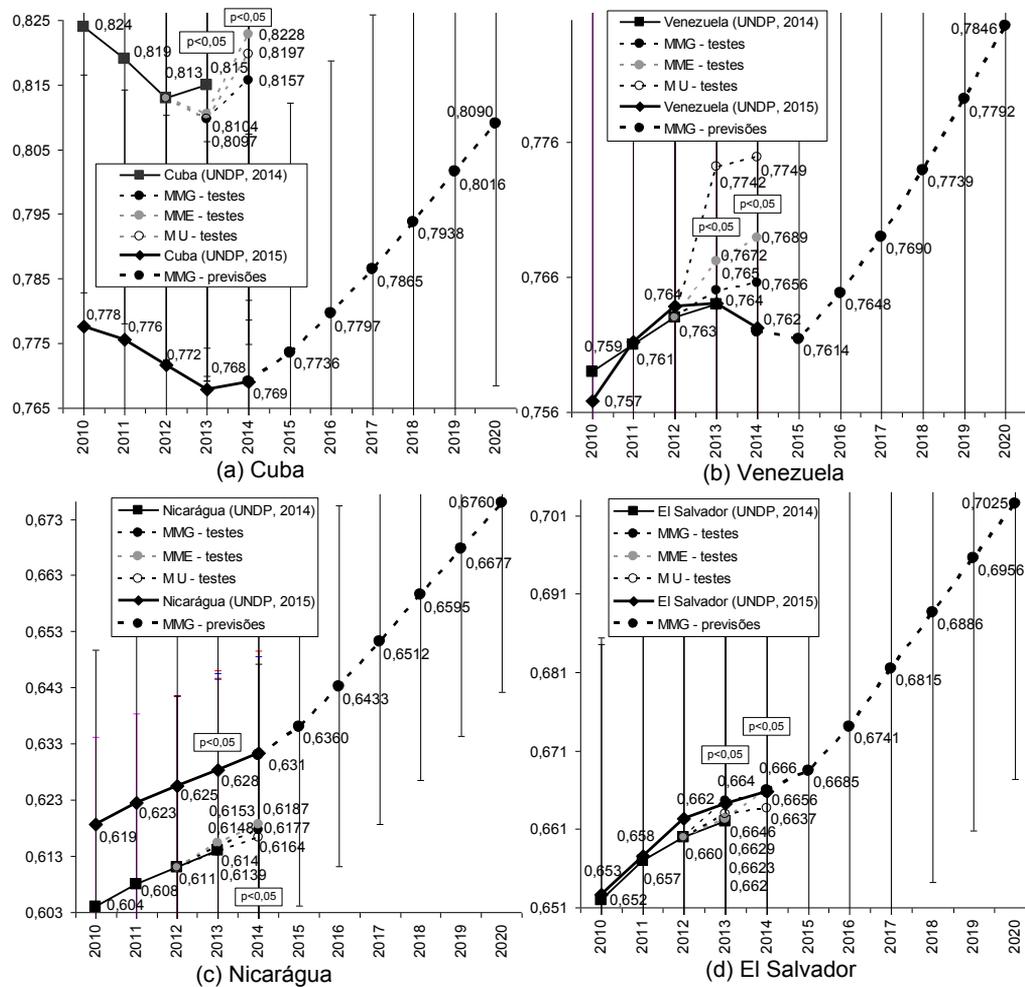


Figura 2. Previsões e tendências do IDH de Cuba(a), Venezuela(b), Nicarágua(c), El Salvador(d).

previsão, apontaram contradições com as direções das tendências do IDH de outros estudos^{3,5} em relação a todos os resultados dos modelos para Cuba, em 2013 (Figura 2a), e Venezuela, em 2014 (Figura 2b). Cuba também apresentou o maior MAE em relação à previsão do IDH 2014, além de apresentar as maiores diferenças entre relatórios da UNDP^{3,5}. Já as previsões para a Nicarágua, em 2013 (Figura 2c), e El Salvador, em 2014 (Figura 2d), apresentaram os menores MAE. Estas comparações, quando favoráveis as previsões, no caso da Nicarágua e El Salvador, podem sustentar políticas públicas ou econômicas adotadas por estes países para o desenvolvimento do índice e, quando desfavoráveis as estimativas, no caso da Venezuela, podem levantar questionamentos de suas políticas ou de seus da-

dos, que podem encontrar-se desatualizados ou inconsistentes, como foi caso de Cuba.

Bolívia (+0,61%) e Cuba (+0,13%) apresentaram, respectivamente, o maior e menor ganho percentual do IDH no último período (2013-2014) em relação aos países da América Latina, enquanto a Venezuela (-0,61%) apresentou perda no índice^{3,5}.

A partir das previsões apresentadas neste estudo (Tabela 3), observa-se que o Uruguai poderá atingir, até 2016, o mesmo nível de desenvolvimento da Argentina e Chile, atualmente classificados⁵ como países desenvolvidos com desenvolvimento humano muito elevado (IDH > 0,799). O mesmo poderá ocorrer com o Panamá em 2017, Cuba em 2019 e Costa Rica em 2020. Os países em desenvolvimento, Paraguai e El

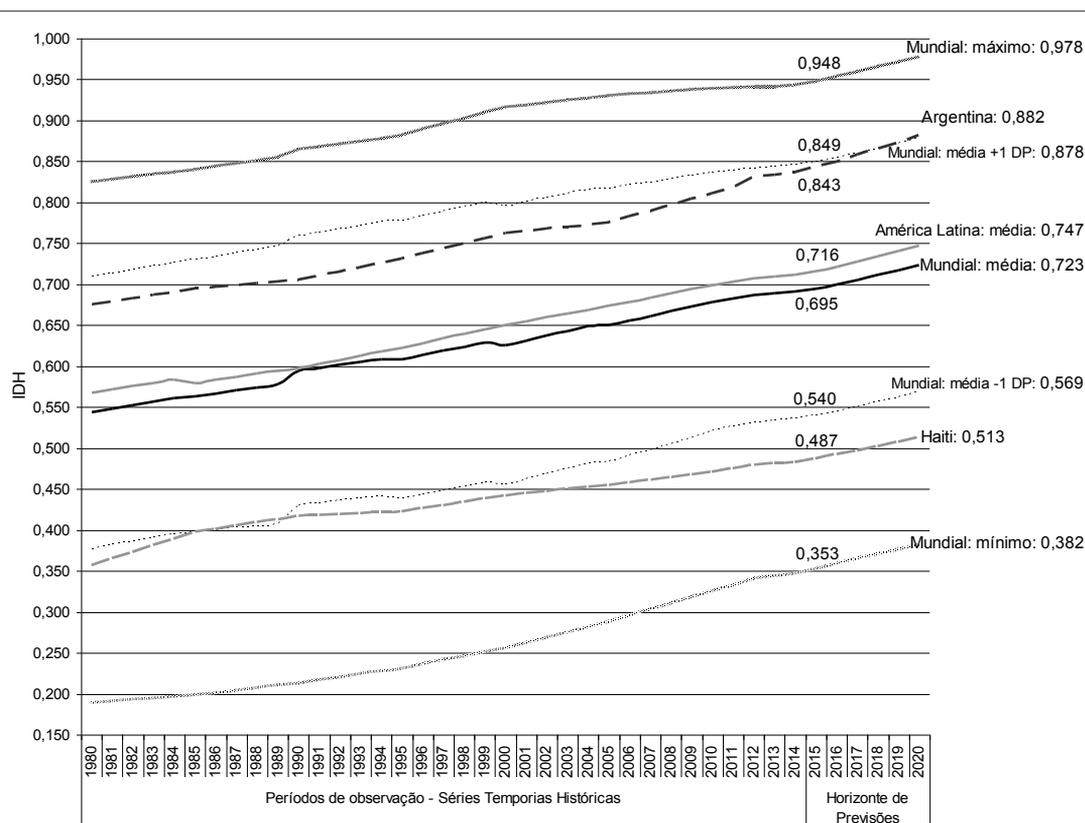


Figura 3. Séries Temporais Históricas (1980 -2014) e Horizonte de Previsões (2015 a 2020) do IDH Mundial, da América Latina e de Países Latino-americanos com maior (Argentina) e menor (Haiti) índice.

Salvador, poderão passar da classe de desenvolvimento humano médio ($0,599 < \text{IDH} < 0,7$) para desenvolvimento humano elevado ($0,699 < \text{IDH} < 0,8$), respectivamente, em 2019 e 2020.

O IDH médio dos países da América Latina ($4,99 \pm 3,90\%$) nas projeções deste estudo, referentes ao período de 2015 a 2020, apresentam uma expectativa de crescimento acima da média mundial ($4,63 \pm 0,20\%$), mantendo a mesma tendência⁵ que destaca a América Latina e o Caribe com os maiores IDH, classificados como alto e com índices superiores às regiões da Europa, Ásia, Pacífico, Estados Árabes e África subsaariana.

Nicarágua (7,13%) e Guatemala (7,02%) tendem a apresentar maior crescimento do índice para o mesmo período, enquanto Belize (1,86%) tende a apresentar o menor crescimento. O Haiti (6,13%), apesar de tender a um crescimento acima da média mundial, permanecerá como único país da América Latina classificado⁵ como sub-desenvolvido ($\text{IDH} < 0,55$). Os demais países

tendem a manter-se na mesma classe de desenvolvimento humano, mesmo apresentando crescimento do índice.

O Brasil, que atualmente passa por um momento de crise econômica⁴⁰, não perceberá mudanças tão significativas no IDH, no caso de avanços significativos na EV ou na educação como se prevê, pois com a nova metodologia de cálculo, estes avanços tendem a ser atenuados pela renda baixa, em função do déficit do PIB. A média geométrica empregada no cálculo do índice reduz o nível de substitutibilidade entre as dimensões, em virtude de que o baixo desempenho em uma determinada dimensão não pode ser mais compensado pelo maior desempenho em outra dimensão⁵. Apesar das críticas⁴¹ ao novo tipo de cálculo do IDH, observa-se que este privilegia países com menor desigualdade entre suas componentes⁴, pois a média geométrica tende a ser cada vez menor que a média aritmética à medida que se aumenta a variância entre as componentes.

Tabela 4. Últimos pontos de observações das séries temporais históricas da expectativa de vida nos países da América Latina e projeções para 2015 a 2020.

País*	Últimos Pontos de Observação %					Horizonte de Previsão						Variação %	
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020		2015-2020
CHL	80,4	80,7	81,1	81,4	81,7	82,14	82,63	83,19	83,81	84,50	85,25	4,35	
CRI	78,8	78,9	79,1	79,2	79,4	79,58	79,82	80,09	80,40	80,75	81,13	2,18	
CUB	79	79,1	79,2	79,3	79,4	79,53	79,64	79,78	79,95	80,17	80,41	1,27	
PAN	76,8	77	77,2	77,4	77,6	77,86	78,12	78,43	78,76	79,14	79,56	2,52	
URY	76,6	76,7	76,9	77	77,2	77,37	77,59	77,84	78,13	78,47	78,83	2,11	
MEX	76,1	76,2	76,4	76,6	76,8	77,04	77,33	77,70	78,13	78,63	79,18	3,10	
ARG	75,6	75,8	75,9	76,1	76,3	76,54	76,79	77,09	77,43	77,80	78,20	2,49	
ECU	75	75,2	75,4	75,7	75,9	76,25	76,65	77,15	77,72	78,35	79,06	4,16	
NIC	73,7	74	74,3	74,6	74,9	75,36	75,90	76,56	77,30	78,13	79,05	5,55	
PER	73,7	73,9	74,1	74,3	74,6	74,93	75,36	75,89	76,51	77,22	77,99	4,55	
BRA	73,3	73,6	73,9	74,2	74,5	74,91	75,35	75,86	76,43	77,08	77,79	4,41	
VEN	73,6	73,7	73,9	74	74,2	74,38	74,61	74,85	75,14	75,46	75,81	2,17	
COL	73,3	73,5	73,7	73,9	74	74,22	74,44	74,71	75,00	75,34	75,72	2,33	
DOM	72,7	72,9	73,1	73,3	73,5	73,79	74,11	74,51	74,96	75,47	76,05	3,47	
HND	72,4	72,6	72,8	72,9	73,1	73,37	73,69	74,09	74,58	75,16	75,81	3,71	
SLV	71,9	72,2	72,5	72,8	73	73,41	73,94	74,59	75,30	76,11	77,03	5,52	
PRY	72,3	72,5	72,6	72,8	72,9	73,09	73,25	73,47	73,69	73,97	74,27	1,88	
GTM	70,9	71,1	71,4	71,6	71,8	72,14	72,56	73,07	73,66	74,34	75,11	4,61	
BLZ	69,7	69,7	69,8	69,9	70	70,07	70,14	70,22	70,28	70,32	70,34	0,48	
BOL	66,4	66,9	67,5	67,9	68,3	68,87	69,44	70,09	70,80	71,63	72,55	6,23	
GUY	66	66,1	66,2	66,3	66,4	66,54	66,72	66,91	67,12	67,34	67,57	1,77	
HTI	61,3	61,7	62,1	62,4	62,8	63,23	63,66	64,13	64,65	65,24	65,89	4,91	
América Latina	μ	73,2	73,4	73,6	73,8	74	74,3	74,6	75	75,4	75,9	76,5	3,33
	±	4,47	4,43	4,39	4,38	4,36	4,34	4,33	4,33	4,35	4,37	4,42	2,05
Mundial	μ	69,9	70,2	70,5	70,8	71	71,4	71,7	72	72,4	72,8	73,3	3,18
	±	8,89	8,73	8,59	8,47	8,37	8,27	8,22	8,19	8,19	8,21	8,24	2,20

* código internacional de países.

Previsões da Expectativa de Vida

Foi possível confrontar as previsões com estudos recentes de outras agências internacionais^{42,43}, que já disponibilizam a EV de 2015 de seus países filiados. No entanto, observam-se diferenças em suas séries temporais em relação à fonte de dados^{3,5,37} utilizada no treinamento dos modelos desenvolvidos neste estudo, que o limita até que se divulguem novos valores de EV ou outros estudos para possibilitar comparações.

Este indicador também pode apresentar inconsistência, pois muitos óbitos não são registrados corretamente¹⁶.

Analisando as séries históricas mundiais de EV, antes das previsões para a América Latina, através da base de dados do UNDP³⁷, observa-se

que, nos últimos 34 anos, a média mundial foi de $67,84 \pm 2,89$ anos. A menor EV registrada no período foi a da população do Camboja, em 1980, com média de 27,5 anos, enquanto a da população mundial nesse período era de $61,62 \pm 10,5$ anos. Em 1995, Ruanda apresentou a menor EV (31,50 anos), bem abaixo dos padrões mundiais da época, de $65,44 \pm 10,18$ anos. No último relatório⁵, Hong Kong - China registrou a maior EV (84 anos em 2014) e a Suazilândia apresentou a menor EV (49 anos), enquanto os padrões mundiais apresentaram a EV média de $71,03 \pm 8,37$ anos. O Haiti é o país da América Latina que sempre apresentou a menor EV, a qual era, em 2014, de 62,8 anos, com perspectivas, segundo as projeções deste estudo, de chegar à EV de 65,89 anos até 2020, acima de -1DP (65,06 anos) da

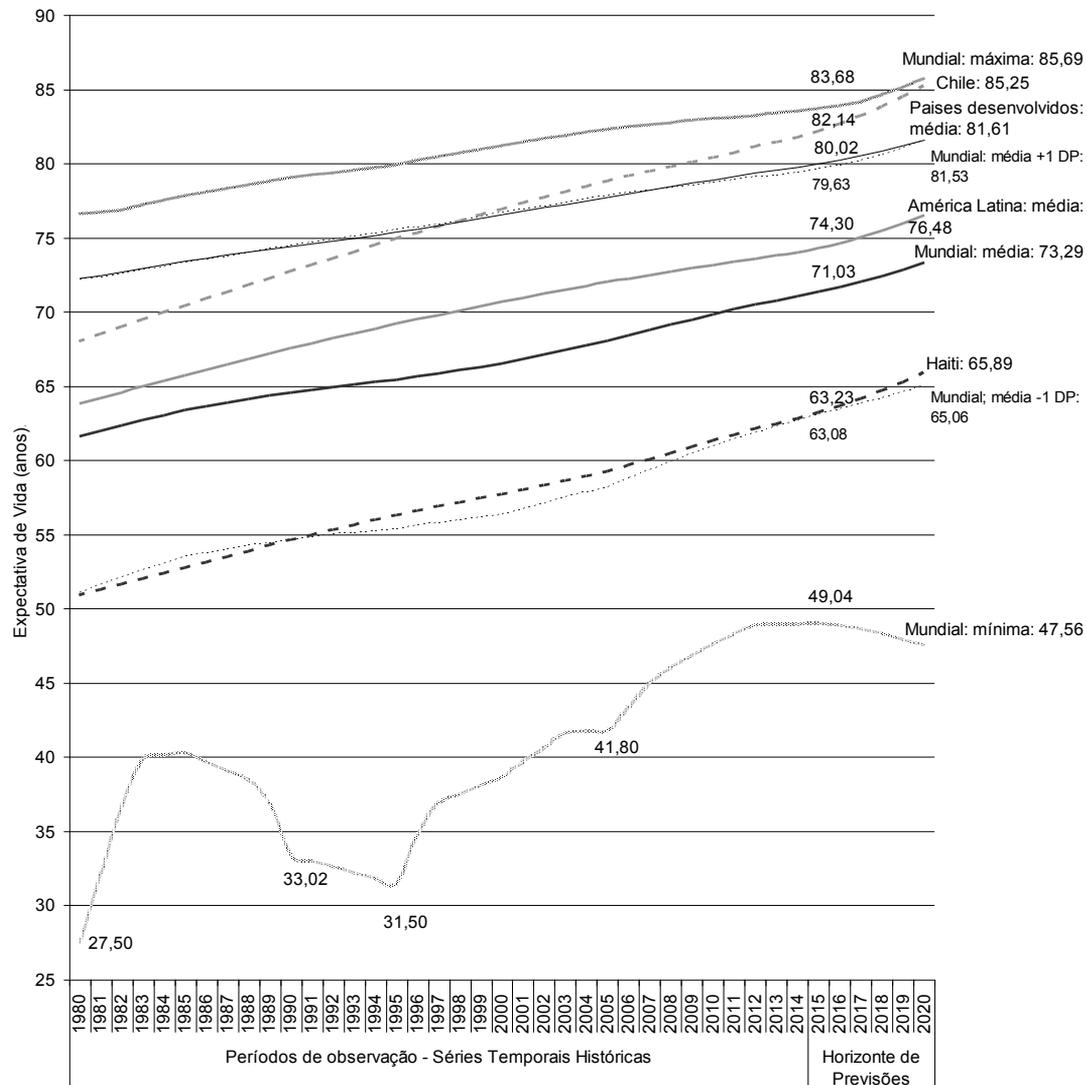


Figura 4. Séries temporais históricas (1980 -2014) e horizonte de previsões (2015 a 2020) da EV mundial, da América Latina e de países latino-americanos com a maior (Chile) e menor (Haiti) EV.

previsão da média mundial. Outros países, como Belize (70 anos), Bolívia (68,3 anos) e Guiana (66,4 anos), também apresentam EV abaixo da média mundial, enquanto a maioria, correspondente a 81,82% dos países da América Latina, apresentam EV acima da média mundial. A EV média da América Latina historicamente^{37,42,43} sempre se apresentou acima da média mundial.

As projeções deste estudo estimam que nos próximos seis anos a média da EV da América Latina aumente de 74 para $76,5 \pm 4,42$ anos, enquanto a mundial é estimada em $73,29 \pm 8,24$ anos, e que em 2015 já seria de $74,3 \pm 4,34$, o que se confirmou em outro estudo⁴².

Atualmente³⁷, o Chile (81,7 anos) apresenta a maior EV da América Latina, com média superior aos demais países desenvolvidos ($79,9 \pm 2,81$), e tende a chegar até 2020 à EV de 85,25 anos, também superior à prevista neste estudo para os países desenvolvidos ($81,61 \pm 3,12$ anos).

Apesar das EV se apresentarem crescentes, Kanso et al.¹⁴ apontam que a EV aos 60 anos aumentaria em 20% se esses óbitos causados por morte evitáveis não ocorressem e que a sobremortalidade masculina foi observada em quase todas as causas de morte evitáveis analisadas, o que pode estar relacionado à maior exposição a fatores de risco e à menor utilização dos serviços

de saúde pelos homens. Por outro lado, estudos¹⁹ apontam uma desvantagem notável do sexo feminino em relação à EV saudável.

Os dados de EV, principalmente analisados de maneiras decompostas, com ou sem saúde, por sexo e regiões, podem ser utilizados em políticas públicas, como referências para cálculo de planos de saúde e de contribuições previdenciárias⁴⁴, decursos para pagamentos de pensões^{45,46}, como justificativas para reformas previdenciárias⁴⁷, no planejamento do futuro da assistência à saúde¹⁵ e qualidade de vida dos idosos⁴⁸⁻⁵⁰, bem como para prever o aumento de doenças relacionadas ao envelhecimento⁵¹.

Método, Modelos e Algoritmo de Previsão

Expressiva parte da literatura sugere que previsões combinadas podem melhorar as previsões individuais³⁰. Isso foi visível nos modelos multivariados (MM), que apresentam melhores resultados que os MUs. Nos MMs, o algoritmo aprendeu com o comportamento de histórico das séries temporais de todos ou de grupos de países, enquanto nos MUs a aprendizagem se limitou às séries temporais do país-alvo.

Os MMGs apresentaram melhor desempenho que os demais. Mas esta vantagem relativa do preditor multivariado pode ser diferente para cada país. Outros estudos⁵² também destacam vantagens dos modelos multivariados, principalmente se existirem fortes relações entre as séries temporais, o que também ocorreu no presente estudo.

Os testes de análise de variância sugerem evidências da inexistência de diferenças significativas entre as previsões dos modelos e as tendências divulgadas pelo UNDP³ para o IDH 2013 e 2014. No entanto, os MMGs apresentaram as melhores medidas de qualidade acumuladas durante todo período de treinamento e previsão, com a maior DAC e os menores tipos de erro em relação aos demais modelos.

A eficiência dos MMGs pode ser implicitamente explicada pelas interdependências e vulnerabilidade dos países, mencionadas em outros estudos³.

Em relação ao método de avaliação dos modelos, observa-se que as medidas de qualidade MAE e DAC são suficientes para qualificar a previsão do IDH ou EV, dispensando-se a necessidade da análise de variância, pois apesar de não existirem diferenças estatísticas significativas

entre previsões dos modelos e os valores reais, o MAE permitiu identificar os melhores modelos, confirmando estudos⁵³ que discutem o uso das medidas de qualidade específicas para previsão.

O algoritmo SMOReg apresentou as melhores medidas de qualidade durante os testes de previsão em relação a outros do mesmo paradigma de aprendizagem (Tabela 1), confirmando estudos anteriores⁵⁴ e também reafirmado vantagens no uso das técnicas de MD (Figura 1) em relação a outras técnicas mais populares de previsão, como ARIMA, já apontadas em estudos anteriores⁵⁵.

As maiores dificuldades enfrentadas neste método de previsão se referem ao custo operacional. As etapas de pré e pós-processamento de MD consumiram aproximadamente 80% do custo operacional, como sugerem estudos anteriores⁵⁶. Falta de acesso a dados totalmente consistentes foi outro problema, frequente em grandes bases de dados⁵⁷, pois atualizações nos dados de observação já publicadas limitaram o estudo, diminuindo a previsibilidade.

Conclusão

Modelos desenvolvidos a partir de séries temporais multivariadas, apesar de mais complexos, apresentaram melhor precisão que os modelos desenvolvidos a partir de séries univariadas.

As séries temporais multivariadas possibilitam maior aprendizagem dos algoritmos com o aumento de diferentes experiências históricas univariadas.

As técnicas de mineração de dados apresentaram melhor qualidade nas previsões em relação à técnica mais popular, ARIMA.

O IDH é um índice robusto com grande previsibilidade e vulnerabilidade, utilizado em pesquisas epidemiológicas, principalmente como limitador demográfico ou parâmetro comparativo.

As previsões apontam que as médias de crescimento do IDH e da EV nos países da América Latina tendem a permanecer maiores que a média mundial.

As contradições entre a previsão e os valores reais do índice ou de suas componentes, se comparados, futuramente poderão desencadear reflexões e auxiliar em tomadas de decisões para sustentação ou não de políticas públicas voltadas ao planejamento e gestão de saúde e, também, justificar o cenário vivido pelos países e mundo.

Colaboradores

CB Santos, contribuição substancial para a concepção, planejamento, análise e interpretação dos dados e participação na aprovação da versão final do manuscrito. LA Pilatti, B Pedroso, DR

Carvalho e AM Guimarães, contribuição substancial para a concepção e planejamento, na revisão crítica do conteúdo e participação na aprovação da versão final do manuscrito.

Referências

1. United Nations Development Programme (UNDP). *Human Development Report (HDR) 1990: Concept and Measurement of human development*. New York: UNDP; 1990.
2. Alkire S. Human development: Definitions, critiques, and related concepts. *UNDP-HDRO Occasional Papers* 2010.
3. United Nations Development Programme (UNDP). *Human Development Report (HDR) 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*. New York: UNDP; 2014.
4. Kovacevic M. Review of HDI critiques and potential improvements, UNDP. *Human Development Reports* 2010; 33.
5. United Nations Development Programme (UNDP). *Human Development Report (HDR) 2015. Rethinking Work for Human Development*. New York: UNDP; 2015.
6. Percio J, Medina NH, Luna EA. Visual Impairment and Human Development in Brazil. *Int J Epidemiol* 2015; 44(Supl. 1):i157.
7. Sadosky ADI, Poton WL, Reis-Santos B, Barcelos MRB, Silva ICM. Índice de Desenvolvimento Humano e prevenção secundária de câncer de mama e colo do útero: um estudo ecológico. *Cad Saude Publica* 2015; 31(7):1539-1550.
8. Tavares LF, Castro IRR, Levy RB, Cardoso LO, Claro RM. Dietary patterns of Brazilian adolescents: results of the Brazilian National School-Based Health Survey (PeNSE). *Cad Saude Publica* 2014; 30(12):2679-2690.
9. Szuster DAC, Caiiffa WT, Andrade EIG, Acurcio FA, Cherchiglia ML. Sobrevida de pacientes em diálise no SUS no Brasil. *Cad Saude Publica* 2012; 28(3):415-424.
10. Castro JMD, Rodrigues-Júnior AL. A influência da mortalidade por causas externas no desenvolvimento humano na Faixa de Fronteira brasileira. *Cad Saude Publica* 2012; 28(1):195-200.
11. Kariminia A, Chokephaibulkit K, Pang J, Lumbiganon P, Hansudewechakul R, Amin J, Kumarasamy N, Puthanakit T, Kurniati N, Nik Yusoff NK, Saphonn V, Fong SM, Razali K, Nallusamy R, Sohn AH, Sirisanthana V. Cohort Profile: The TREAT Asia Pediatric HIV Observational Database. *Int J Epidemiol* 2011; 40(1):15-24.
12. Martinez EZ, Roza DL, Caccia-Bava MCGG, Achcar JA, Dal-Fabbro AL. Gravidez na adolescência e características socioeconômicas dos municípios do Estado de São Paulo, Brasil: análise espacial. *Cad Saude Publica* 2011; 27(5):855-867.
13. González-Zapata LI, Estrada-Restrepo A, Álvarez-Castaño LS, Álvarez-Dardet C, Serra-Majem L. Exceso de peso, aspectos económicos, políticos y sociales en el mundo: un análisis ecológico. *Cad Saude Publica* 2011; 27(9):1746-1756.
14. Kanso S, Romero DE, Leite IC, Marques A. A evitabilidade de óbitos entre idosos em São Paulo, Brasil: análise das principais causas de morte. *Cad Saude Publica* 2013; 29(4):735-748.
15. Mendes ACG, Sá DA, Miranda GMD, Lyra TM, Tavares RAW. Assistência pública de saúde no contexto da transição demográfica brasileira: exigências atuais e futuras. *Cad Saude Publica* 2012; 28(5):955-964.
16. Chiavegatto Filho ADP, Laurenti R. Decomposição da diferença da expectativa de vida de Minas Gerais em relação ao Rio de Janeiro e São Paulo, Brasil. *Cad Saude Publica* 2013; 29(6):1131-1140.
17. Cervantes CAD, Botero MA. Average years of life lost due to breast and cervical cancer and the association with the marginalization index in Mexico in 2000 and 2010. *Cad Saude Publica* 2014; 30(5):1093-1102.
18. Campolina AG, Adami F, Santos JLF, Lebrão ML. A transição de saúde e as mudanças na expectativa de vida saudável da população idosa: possíveis impactos da prevenção de doenças crônicas. *Cad Saude Publica* 2013; 29(6):1217-1229.
19. Camargos MCS, Gonzaga MR. Viver mais e melhor? Estimativas de expectativa de vida saudável para a população brasileira. *Cad Saude Publica* 2015; 31(7):1460-1472.
20. Stringhini S, Polidoro S, Sacerdote C, Kelly RS, van Veldhoven K, Agnoli C, Grijoni S, Tumino R, Giurandella MC, Panico S, Mattiello A, Palli D, Masala G, Gallo V, Castagné R, Paccaud F, Campanella G, Chadeau-Hyam M, Vineis P. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. *Int J Epidemiol* 2015; 44(4):1320-1330.
21. Li L, Hardy R, Kuh D, Power C. Life-course body mass index trajectories and blood pressure in mid life in two British birth cohorts: stronger associations in the later-born generation. *Int J Epidemiol* 2015; 44(3):1018-1026.
22. Lacey RE, Sacker A, Kumari M, Worts D, McDonough P, Booker C, McMunn A. Work-family life courses and markers of stress and inflammation in mid-life: evidence from the National Child Development Study. *Int J Epidemiol* 2015; 45(4):1247-1259.
23. Hendi AS. Trends in U.S. life expectancy gradients: the role of changing educational composition. *Int J Epidemiol* 2015; 44(3):946-955.
24. Morton SM, De Stavola BL, Leon DA. Intergenerational determinants of offspring size at birth: a life course and graphical analysis using the Aberdeen Children of the 1950s Study (ACONF). *Int J Epidemiol* 2014; 43(3):749-759.

25. Anstey KJ, Kingston A, Kiely KM, Luszcz MA, Mitchell P, Jagger C. The influence of smoking, sedentary lifestyle and obesity on cognitive impairment-free life expectancy. *Int J Epidemiol* 2014; 43(6):1874-1883.
26. Brunekreef B, Von Mutius E, Wong GK, Odhiambo JA, Clayton TO, Group tIPTS. Early life exposure to farm animals and symptoms of asthma, rhinoconjunctivitis and eczema: an ISAAC Phase Three Study. *Int J Epidemiol* 2012; 41(3):753-761.
27. Mackenbach JP, Looman CW. Life expectancy and national income in Europe, 1900-2008: an update of Preston's analysis. *Int J Epidemiol* 2013; 42(4):1100-1110.
28. Mangalova E, Agafonov E. Wind power forecasting using the k-nearest neighbors algorithm. *Int J Forecasting* 2014; 30(2):402-406.
29. Silva L. A feature engineering approach to wind power forecasting: GEFCom 2012. *Int J Epidemiol* 2014; 30(2):395-401.
30. Rodrigues BD, Stevenson MJ. Takeover prediction using forecast combinations. *Int J Forecasting* 2013; 29(4):628-641.
31. Correa FE, Gama J, Pizzigatti Correa PL, Alves LRA. Data mining frequent temporal events in agrieconomic time series. *IEEE Lat Am T* 2015; 13(7):2329-2334.
32. Sousa WRN, Couto MS, Castro AF, Silva MPS. Evaluation of desertification processes in ouricuri-pe through trend estimates of times series. *IEEE Lat Am T* 2013; 11(1):602-606.
33. Xie Y, Schreier G, Hoy M, Liu Y, Neubauer S, Chang DCW, Redmond SJ, Lovell NH. Analyzing health insurance claims on different timescales to predict days in hospital. *J Biomed Inform* 2016; 60:187-196.
34. Winters-Miner LA, Bolding PS, Hilbe JM, Goldstein M, Hill T, Nisbet R, Walton N, Miner GD. Biomedical Informatics. In: Winters-Miner LA, Bolding PS, Hilbe JM, Goldstein M, Hill T, Nisbet R, Walton N, Miner GD. *Practical Predictive Analytics and Decisioning Systems for Medicine*. Cambridge: Academic Press; 2015. p. 42-59.
35. Fayyad UM, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI magazine* 1996; 17(3):37.
36. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 2009; 11(1):10-18.
37. United Nations Development Programme (UNDATA). *Human Development Index trends, 1980-2013*. New York: UNDATA; 2014.
38. Hall MA. *Correlation-based feature selection for machine learning* [thesis]: The University of Waikato; 1999.
39. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to the SMO algorithm for SVM regression. *IEEE T Neur Net Lear* 2000; 11(5):1188-1193.
40. Watts J. Brazil's health system woes worsen in economic crisis. *Lancet* 2016; 387(10028):1603-1604.
41. Ravallion M. Troubling tradeoffs in the human development index. *J Dev Econ* 2012; 99(2):201-209.
42. Central Intelligence Agency (CIA). *The World Factbook*. Langley: CIA; 2016.
43. World Bank. *World Development Indicators*. Washington: The World Bank; 2016.
44. Inoue JT, Rodrigues CG, Afonso LE. Tábua de mortalidade e expectativa de vida saudável: uma aplicação à população beneficiária de planos de saúde privados no Brasil em 2008. In: *Anais do 12º Congresso USP de Controladoria e Contabilidade*; 2012; São Paulo. p. 1-15.
45. Brasil. Lei nº 13.183, de 4 de novembro de 2015. Altera as Leis nºs 8.212, de 24 de julho de 1991, e 8.213, de 24 de julho de 1991, para tratar da associação do segurado especial em cooperativa de crédito rural e, ainda essa última, para atualizar o rol de dependentes, estabelecer regra de não incidência do fator previdenciário, regras de pensão por morte e de empréstimo consignado, a Lei nº 10.779, de 25 de novembro de 2003, para assegurar pagamento do seguro-defeso para familiar que exerça atividade de apoio à pesca, a Lei nº 12.618, de 30 de abril de 2012, para estabelecer regra de inscrição no regime de previdência complementar dos servidores públicos federais titulares de cargo efetivo, a Lei nº 10.820, de 17 de dezembro de 2003, para dispor sobre o pagamento de empréstimos realizados por participantes e assistidos com entidades fechadas e abertas de previdência complementar e a Lei nº 7.998, de 11 de janeiro de 1990; e dá outras providências. *Diário Oficial da União* 2015; 5 nov.
46. Lu B, He W, Piggott J. Should China introduce a social pension? *The Journal of the Economics of Ageing* 2014; 4:76-87.
47. Rocha FRE. A previdência social no Brasil: uma política em reestruturação. *Temporalis* 2016; 2(30):453-473.
48. Rosa VD. Atividade física e a qualidade de vida de mulheres idosas. *FACES* 2016.
49. Vecchia RD, Ruiz T, Bocchi SCM, Corrente JE. Qualidade de vida na terceira idade: um conceito subjetivo. *Rev Bras Epidemiol* 2005; 8(3):246-252.
50. Minayo MCS, Hartz ZMA, Buss PM. Qualidade de vida e saúde: um debate necessário. *Cien Saude Colet* 2000; 5(1):7-18.
51. Salgado Filho N, Brito DJA. Doença renal crônica: a grande epidemia deste milênio. *J Bras Nefrol* 2006; 28(2):1-5.
52. Peña D, Sánchez I. Measuring the advantages of multivariate vs. univariate forecasts. *J Time Ser Anal* 2007; 28(6):886-909.
53. Armstrong JS. Evaluating forecasting methods. In: Armstrong JS, editor. *Principles of forecasting*: Springer; 2001. p. 443-472.
54. Li C, Jiang L. Using locally weighted learning to improve SMOreg for regression. *Pacific Rim International Conference on Artificial Intelligence*; 2006; Berlin Heidelberg: Springer; 2006. p. 375-384.
55. Hong T, Pinson P, Fan S. Global energy forecasting competition 2012. *Int J Forecasting* 2014; 30(2):357-363.
56. Mannila H. Data mining: machine learning, statistics, and databases. *Scientific and Statistical Database Management, International Conference on*; 1996: IEEE Computer Society; 1996. p. 2-2.
57. Witten IH, Frank E. *Practical machine learning tools and techniques*. 2nd ed. San Francisco: Morgan Kaufmann; 2005.

Artigo apresentado em 25/05/2016

Aprovado em 21/10/2016

Versão final apresentada em 23/10/2016