

Utilização de Redes Neurais e Regressão Linear na Predição de Resultados de Jogos do Novo Basquete Brasil (NBB)

Alex Michelon¹, Osmar Oliveira Braz Junior¹, Pablo Schoeffel¹

¹Departamento de Engenharia de Software – Universidade do Estado de Santa Catarina (UDESC) – Centro de Educação Superior do Alto Vale do Itajaí (CEAVI) – Ibirama – SC – Brasil

{alexmichelonn, osmar.oliveira.braz}@gmail.com,
pablo.schoeffel@udesc.br

Abstract. *The use of learning machine to sports analysis has become an important support to teams, coaches and athletes of several sports. The result's prediction, besides to be a tool to technical analysis, can be used to strategic goals by sports director and by people who bet. In this context, this paper compares the use of artificial neural networks and linear regression to identify the best technique to predict results of basketball matches. The study uses scouts (“numerical statistics in the game’s aspects”) from seasons of NBB (Novo Basquete Brasil) to verify the accuracy on the predictions comparing to real results. Tests performed identified that linear regression was the most efficient technique found, which had more than 75% of assertiveness. To validate the tests, two simulations were performed using the best technique and the data of a new NBB season. The results reached more than 68% of assertiveness.*

Keywords: *Prediction, artificial intelligence, basketball.*

Resumo. *A utilização de aprendizado de máquina para análises esportivas tem se tornado um importante aliado aos clubes, técnicos e atletas de diversas modalidades. A predição de resultados, além de ser um instrumento para análise técnica, pode ser utilizada com fins estratégicos por dirigentes e até apostadores. Nesse contexto, esse estudo faz comparações entre a utilização de redes neurais e regressão linear, a partir das estatísticas relacionadas aos jogos (scouts) de temporadas do NBB (Novo Basquete Brasil), a fim de identificar qual a técnica que permite maior acurácia sobre a assertividade de resultados em comparação a embates já ocorridos. A partir de testes gerados, foi identificado que a técnica regressão linear se mostrou a mais eficiente, permitindo índice de assertividade superior a 75% sobre os dados de testes. Para validação foram geradas duas simulações a partir deste método, compreendendo jogos de uma nova temporada da liga NBB, cujos resultados chegaram a mais de 68% de assertividade.*

Palavras-chave: *Predição, Inteligência Artificial, Basquetebol.*

1. Introdução

Ao longo do tempo, as atividades esportivas expandiram-se para além das áreas nas quais foram concebidas. O basquetebol, a exemplo de outros esportes, desenvolveu

ramos como a profissionalização, a gestão em administração esportiva e o conceito de esporte como negócio. Esta evolução também se estende à área tecnológica, onde o entretenimento aliado ao consumismo elevou a importância do esporte ao cotidiano das pessoas por todo o mundo [Azevêdo 2009].

Com o desenvolvimento da tecnologia, a área esportiva tornou-se foco de estudos e inovações. Uma delas, ligada ao esporte, foi à análise de estatísticas de jogo. Após a década de 80, especialistas iniciaram estudos na área de estatística, no intuito de fornecer informações antes apenas conhecidas através da experiência de indivíduos e, mesmo assim, de forma subdesenvolvida [Cao 2012].

Hoje, técnicos, assistentes, especialistas e simpatizantes dispõem de ferramentas que, baseados em *scout* (informações relacionadas a aspectos gerais envolvidos em um jogo), são capazes de efetuar predições de resultados. Na área de basquetebol, previsões são possíveis devido à enorme quantidade e detalhamento dos dados existentes, aliada à alta disponibilidade encontrada. Neste contexto, estão inseridas técnicas como redes neurais artificiais e regressão linear [Cao 2012].

Existem diversos estudos baseados em técnicas de estatísticas e algoritmos de aprendizagem para buscar previsões de resultados no basquetebol e outros esportes. Porém, não foram encontrados trabalhos relacionados às competições nacionais, tão pouco trabalhos que disponibilizem algum meio de geração de simulações de resultados.

Sendo assim, este estudo propõe a criação de um mecanismo para a geração de predições de jogos do NBB, o Novo Basquete Brasil. Serão usadas como base as informações obtidas a partir do processamento do *scout* dos jogos de temporadas anteriores, aplicadas a redes neurais artificiais e regressão linear, avaliando os índices de assertividade na predição de resultados de jogo.

Esse artigo está organizado da seguinte forma: na Seção 2 é apresentada uma breve contextualização sobre o NBB e predição esportiva; na Seção 3 são descritos os trabalhos mais similares encontrados na literatura. Na Seção 4 é descrita a metodologia utilizada e a Seção 5 apresenta os resultados. Por fim, a Seção 6 apresenta as considerações finais do trabalho.

2. Contextualização

O NBB (Novo Basquete Brasil) é o atual campeonato brasileiro masculino adulto de basquetebol organizado pela LNB (Liga Nacional de Basquetebol). O NBB foi criado em 2009 com o objetivo de recolocar o basquetebol entre os esportes mais praticados do país [LNB 2016].

Neste contexto, os clubes têm recorrido ao uso de tecnologia, a fim de lhes permitir vantagens competitivas. Uma das possibilidades é o uso da computação como forma de prever ou antecipar situações de jogo, tal como seus resultados. A predição, de acordo com Pereira (2011), “significa antecipação, previsão, adivinhação”. Esta é acompanhada de estratégias de avaliação, correção, inferência e processos cognitivos de compreensão, ou seja, está ligada intimamente com o aspecto humano [Pereira 2011].

Os especialistas colocam que a gestão dos dados estatísticos é de grande valia ao avanço da qualidade no esporte, principalmente para aproveitamento de técnicos e integrantes de comissões técnicas [Vendite 2005]. Para obter vantagens competitivas, vários esportes utilizam de técnicas de predição [Cao 2012].

Com a alta disponibilidade de informações, principalmente referente à NBA (liga norte-americana de basquetebol), muitos aspectos do esporte puderam ser analisados, corrigidos e otimizados, com o objetivo de produzir resultados significantes e permitir maior conhecimento das situações envolvidas no jogo [Cao 2012].

Para que a tarefa de predição se torne possível, torna-se necessário que os mecanismos de pesquisa descubram padrões, testem opções, reconheçam situações comuns. Neste âmbito encontram-se softwares como o Weka, o qual disponibiliza algoritmos de aprendizado e técnicas de comparação estatística para elaboração de análises em base de dados [Cao 2012].

3. Trabalhos Correlatos

Utilizando como base de pesquisa temas como: predições de resultados em basquetebol, *data mining* no esporte e comparação entre algoritmos de classificação de dados, foram encontrados quatro trabalhos similares à proposta deste artigo.

Cao (2012) propôs um estudo que baseia-se em técnicas de mineração de dados para efetuar a predição de resultados de basquetebol. Para isto, foram utilizados dados obtidos dos websites Databasebasketball.com, NBA.com e, principalmente, Basketball-reference.com. As informações são decorrentes de temporadas da NBA entre os anos de 2006 e 2011 e são constituídas de: i) estatísticas dos últimos 10 jogos dos times e de seus oponentes; ii) estatísticas de jogos recentes entre os dois times; iii) números de jogos nos últimos 5 dias; iv) dias de descanso antes do jogo; v) estatísticas gerais dos times na última temporada.

Como resultado destes testes, aplicadas para as técnicas de regressão linear, *naïve bayes*, máquina de vetores de suporte (método de aprendizado assistido através da análise de dados por regressão) e redes neurais; obtiveram-se níveis de acurácia das técnicas baseadas na divisão das amostras classificadas corretamente sobre a quantidade total de amostras, sendo 67.82%, 65.82%, 67.22% e 66,67%, respectivamente.

Fabian [Fabian 2008] propôs um estudo no qual foi concebida uma ferramenta para efetuar análise de embates de basquetebol, para que técnicos, assistentes e simpatizantes do esporte possam entender e conhecer seus oponentes. A aplicação, denominada WASP (*Weka for Advanced Statistical Prediction*), utiliza de *scouts* de jogadores de equipes da NBA para efetuar predições por atletas, no intuito de formar resultados da equipe em cada jogo.

A avaliação do estudo evidenciou que algumas características de jogo tiveram melhor nível de assertividade entre predição e dados reais. A análise das predições também apontou que tende haver um jogador em ambas as equipes que possui diferença substancial entre os resultados previstos e reais em certa característica de jogo. Este padrão foi classificado como “um desastre do jogador”, fato que implica na predição final do resultado do embate.

Poropudas (2011) propôs um estudo em que, através de um algoritmo matemático, pode-se classificar e avaliar equipes de basquete de acordo com o seu desempenho, utilizando a diferença de placares entre os times em sua sede e fora de sua sede. Estas classificações servem como base para o cálculo de quatro parâmetros no modelo: vitórias, percentuais de vitórias, diferença de pontos e diferença de pontos nos

últimos 60 dias, os quais são aliados ao fator da equipe jogar em sua sede ou na sede de seu adversário.

Os dados utilizados foram da NBA, dentre as temporadas de 2005-2006 e 2009-2010, sendo que cada uma destas compreende 1230 jogos. Este estudo baseou-se na utilização do algoritmo do filtro de Kalman, o qual se mostrou satisfatoriamente eficiente na classificação dos fatores espaço-tempo, obtendo percentuais de acurácia entre 65 e 70%, baseado nos resultados apresentados.

West e Lamsal (2008) disponibilizaram um estudo cujo objetivo é, por meio da análise de classificação e predição, prever quais times de futebol americano dentre as faculdades disputarão o campeonato nacional da modalidade e qual será o vencedor.

Como fonte de dados, a proposta utilizou informações dos anos de 2004 a 2007, de 240 subligas da *NCAA Football College Division* (FCS, primeira divisão do futebol americano colegial dos Estados Unidos), sendo que em cada uma destas ocorreram 120 jogos. Estes dados foram coletados de fontes on-line: Yahoo! Sports, ESPN.com, Jeff Sagarin do USA Today, NCAA.com e o website cfbstats.com [West and Lamsal 2008].

Na aplicação da predição nas equipes participantes do FCS, considerando as próximas fases e a grande final do torneio, obteve-se o índice de 59.4% de assertividade [14].

O principal diferencial do corrente projeto em relação aos trabalhos mencionados consiste em explorar o ambiente preditivo no que diz respeito a dados do NBB. Isso pela razão de não se ter conhecimento sobre alguma proposta parecida que objetiva a análise de dados de competições nacionais, os quais, por intermédio das características de seus embates, equipes e jogadores, podem demonstrar aspectos e padrões diferentes em relação a ligas internacionais.

4. Metodologia

O intuito é o de descobrir, dentre os métodos de redes neurais artificiais e regressão linear, qual apresenta maior acurácia na classificação dos jogos do NBB, permitindo que outros sistemas ou usuários possam utilizar a técnica escolhida para efetuar predições de basquetebol. Para isso, o trabalho foi realizado seguindo as seguintes etapas: i) coleta e manipulação dos dados; ii) treinamento; iii) testes; iv) análise e comparação dos resultados dos testes; v) análise dos resultados.

4.1. Coleta e Manipulação de Dados

Os dados utilizados como base para o trabalho foram extraídos do website do LNB (<http://www.lnb.com.br/estatisticas/>). Os *scouts* utilizados correspondem aos jogos realizados entre as temporadas 2008/2009 e 2012/2013, com exceção dos jogos das equipes Saldanha da Gama e Londrina, visto que as médias destas equipes não foram localizadas no website até o final da análise dos dados. Considerando este fato, o conjunto total de amostras somaram 1234 instâncias. A temporada 2014, utilizada para o conjunto de dados de validação, somam 229 amostras.

As características formadoras do *scout* utilizado neste estudo correspondem à pontuação, assistências, rebotes, percentual de acerto de lances livres, cestas de três e de dois pontos; recuperações de posse de bola, bloqueios, erros e eficiência de cada equipe; considerando as médias finais em cada temporada.

No intuito de testar a ação e acurácia dos métodos candidatos, foram criados três tipos de arquivos diferentes quanto à forma de apontar o vencedor de cada embate. O primeiro, denominado “A ou B”, consiste em atribuir ao vencedor a letra “A”, caso a equipe da sede fosse o vencedor e com a letra B, caso a equipe visitante fosse o vencedor. O segundo arquivo foi chamado de “1 ou 0”, possuindo comportamento semelhante ao “A ou B”, alterando apenas a letra “A” pelo número 1 e a letra “B” pelo número 0. O terceiro tipo de arquivo foi denominado de “diferença de pontos” e taxa o vencedor através da subtração entre a pontuação final da equipe da sede e a pontuação final da equipe visitante para o respectivo jogo. Caso o resultado seja positivo, a equipe da sede é a vencedora, caso o resultado seja negativo, o visitante é o vencedor.

4.2. Treinamento

Para o treinamento foi utilizada a ferramenta Weka (2016). Sobre as redes neurais artificiais, foram necessários alguns testes para a definição de parâmetros. A partir dos resultados destes testes realizados com 15 combinações dos parâmetros taxa de aprendizagem, a taxa de aprendizagem de 0,3 obteve o melhor índice dentre todos (93,5981% de assertividade) e, por este motivo, foi o valor padrão adotado nos treinamentos e testes para o método redes neurais artificiais.

Com relação ao número de neurônios da camada oculta da rede, os testes apontaram que o modelo com melhor percentual, atingindo nível de acurácia superior a 93%, utilizou 21 neurônios na camada oculta, parâmetro que foi escolhido para o modelo de treinamento e teste deste projeto, no âmbito de redes neurais artificiais.

4.3. Testes

Nesta proposta, os testes foram efetuados utilizando dois métodos: a opção *Supplied test set* (A) (conjunto de teste) e a opção *Cross-validation* (B) (validação cruzada). O arquivo aplicado a opção (A) foi formado pelos dados da temporada 2013/2014 do NBB, compreendendo os jogos até a data de 13 de março de 2014. Até este período, somaram-se 229 jogos, os quais foram utilizados para validar o arquivo de treinamento, quando utilizada a opção de teste por conjunto de teste.

Para a técnica de regressão linear, é importante frisar que o modelo utilizado para seleção de atributos foi o método M5. Como restrição nesta etapa, pode-se citar que o algoritmo de regressão linear não classifica atributos nominais. Desta forma não foi possível efetuar testes desta técnica com o tipo de arquivo “A ou B”. Para execução de testes com o arquivo “1 ou 0”, foi necessário designar o atributo “vencedor” como numérico.

4.4. Análise dos Dados de Testes

A Tabela 1 demonstra as configurações e resultados dos testes efetuados. Esta tabela contém os percentuais de instâncias corretamente e incorretamente classificadas, informação utilizada para escolha do método de implementação. Algumas informações tiveram legendas criadas tais como épocas (E), *folds* (F), taxa de aprendizagem (TA), neurônios na camada oculta (CO).

Tabela 1 – Resultado dos testes por técnica e parâmetros

Técnica Utilizada	Tipo de Arq.	Config. Utilizada WEKA	Total Instânc. Classif.	Instâncias Classificadas Corretamente (%)		Instâncias Classificadas Incorretamente (%)		Erro Absoluto Médio	
Redes neurais artificiais	1 ou 0	Validação cruzada	1463	F: 2	72,5906%	F: 2	27,4094%	F: 2	0.378
				F: 5	73,6159%	F: 5	26,3841%	F: 5	0.3398
				F: 10	73,2057%	F: 10	26,7943%	F: 10	0.3319
				F: 20	74,7779%	F: 20	25,2221%	F: 20	0.3349
				F: 40	74,2310%	F: 40	25,7690%	F: 40	0.3373
Redes neurais artificiais	1 ou 0	Conjunto de teste	229	E: 1000	65,0655%	E: 1000	34,9345%	E: 1000	0.3491
				E: 2000	66,3755%	E: 2000	33,6245%	E: 2000	0.3380
				E: 5000	64,6288%	E: 5000	35,3712%	E: 5000	0.3471
				E: 10000	65,5022%	E: 10000	34,4978%	E: 10000	0.3438
				E: 20000	64,6288%	E: 20000	35,3712%	E: 20000	0.3512
				E: 30000	64,1921%	E: 30000	35,8079%	E: 30000	0.3541
				E: 40000	63,3188%	E: 40000	36,6812%	E: 40000	0.3629
Regressão linear	1 ou 0	Conjunto de teste	229	68,9956%		31,0044%		0.3399	
Regressão linear	1 ou 0	Validação cruzada	1463	F: 2	73,8209%	F: 2	26,1791%	F: 2	0.3607
				F: 5	73,8893%	F: 5	26,1107%	F: 5	0.3575
				F: 10	74,2310%	F: 10	25,7690%	F: 10	0.3561
				F: 20	74,5044%	F: 20	25,4956%	F: 20	0.355
Redes neurais artificiais	A ou B	Validação cruzada	1463	F: 2	72,4539%	F: 2	27,5461%	F: 2	0.3792
				F: 5	73,7526%	F: 5	26,2474%	F: 5	0.3399
				F: 10	72,8640%	F: 10	27,1360%	F: 10	0.3294
				F: 20	74,7779%	F: 20	25,2221%	F: 20	0.3352
				F: 40	73,8209%	F: 40	26,1791%	F: 40	0.3412
Redes neurais artificiais	Diferença de Pontos	Validação cruzada	1463	F: 2	70,5400%	F: 2	29,4600%	F: 2	0.0976
				F: 5	70,6083%	F: 5	29,3917%	F: 5	0.0951
				F: 10	71,3602%	F: 10	28,6398%	F: 10	0.0972
				F: 20	72,0437%	F: 20	27,9563%	F: 20	0.0944
Redes neurais artificiais	Diferença de Pontos	Conjunto de teste	229	E: 1000	67,2489%	E: 1000	32,7511%	E: 1000	0.1303
				E: 2000	51,9651%	E: 2000	48,0349%	E: 2000	0.3302
				E: 5000	50,2183%	E: 5000	49,7816%	E: 5000	0.4515
				E: 10000	49,3449%	E: 10000	50,6550%	E: 10000	0.7546
				E: 20000	49,7816%	E: 20000	50,2183%	E: 20000	0.7798
				E: 30000	51,0917%	E: 30000	48,9082%	E: 30000	0.7262
				E: 40000	54,1484%	E: 40000	45,8515%	E: 40000	0.6951
Regressão linear	Diferença de Pontos	Conjunto de teste	229	68,5590%		31,4410%		0.0857	
Regressão linear	Diferença de Pontos	Validação cruzada	1463	F: 2	73,7526%	F: 2	26,2474%	F: 2	0.0884
				F: 5	74,2310%	F: 5	25,7690%	F: 5	0.0883
				F: 10	74,7095%	F: 10	25,2905%	F: 10	0.0879
				F: 20	75,1196%	F: 20	24,8804%	F: 20	0.0875
Redes neurais com Regressão linear	Diferença de Pontos	Conjunto de teste	229	E: 1000	68,5590%	E: 1000	31,4410%	E: 1000	0.1464
				E: 2000	63,7555%	E: 2000	36,2445%	E: 2000	0.2273
				E: 5000	55,8952%	E: 5000	44,1048%	E: 5000	0.5684
				E: 10000	55,0218%	E: 10000	44,9782%	E: 10000	0.6577
				E: 20000	59,3886%	E: 20000	40,6114%	E: 20000	1.769
Redes neurais com Regressão linear	Diferença de Pontos	Validação cruzada	1463	F: 2	69,1729%	F: 2	30,8271%	F: 2	0.1181
				F: 5	68,4894%	F: 5	31,5106%	F: 5	0.1204
				F: 10	69,4463%	F: 10	30,5537%	F: 10	0.1149
				F: 20	69,0362%	F: 20	30,9638%	F: 20	0.1151

Os resultados obtidos com redes neurais, apesar de não apresentarem a melhor acurácia dentre os modelos, tiveram níveis satisfatórios de predição, considerando o nível de maior assertividade encontrado (Figura 1a).

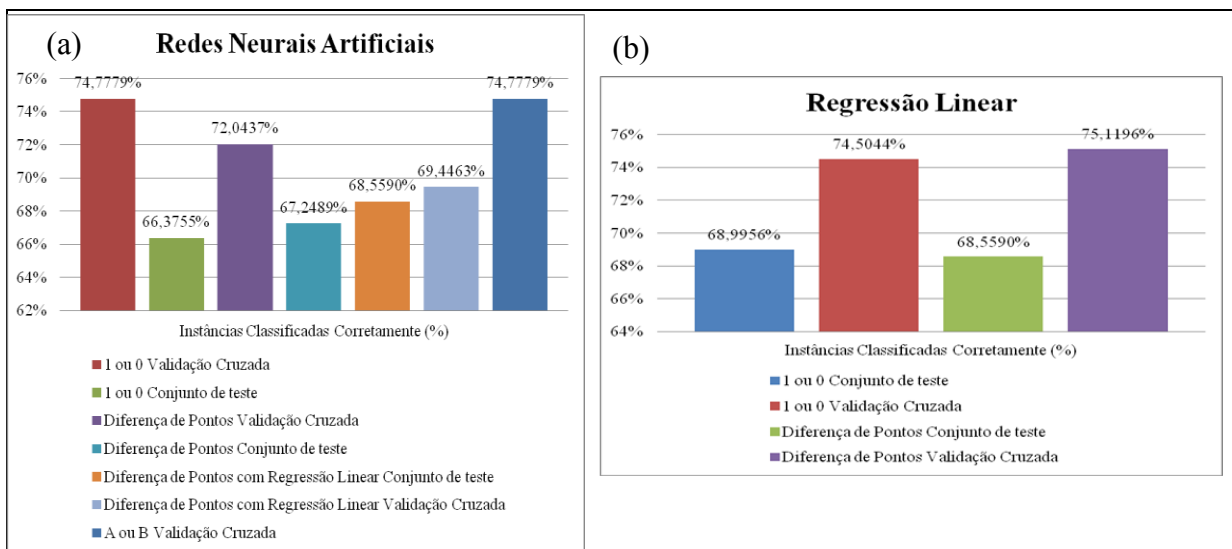


Figura 1 – (a) Resultados de testes com redes neurais artificiais (b) Resultados de testes com regressão linear

A influência do tipo de avaliação com validação cruzada se mostra fortemente presente na técnica de regressão linear. Nesta técnica, houve-se uma pequena disparidade entre os índices dos resultados dos métodos de avaliação conjunto de teste e validação cruzada. Enquanto a validação cruzada obteve índices de acerto de 74,5044% para o tipo de arquivo “1 ou 0” e de 75,1196% para o arquivo “Diferença de pontos”, o conjunto de teste obteve para os mesmos arquivos, respectivamente, 68,9956% e 68,5590%. A Figura 1b apresenta tais dados.

Os modelos que obtiveram maiores índices de acerto na classificação dos dados são demonstrados na Figura 2a.

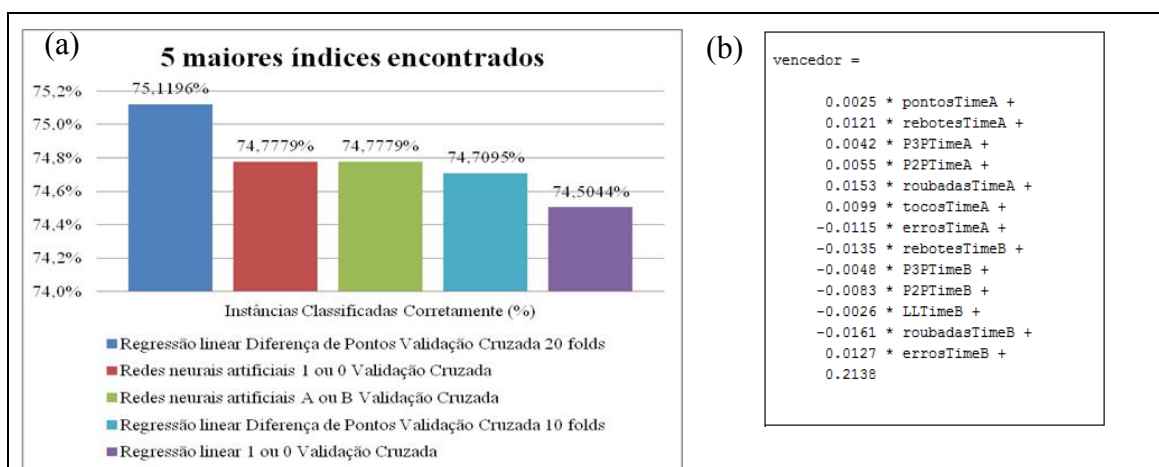


Figura 2 - (a) Melhores resultados dos testes (b) Modelo de regressão linear obtido

A partir dos resultados apresentados na Figura 2a, o método de regressão linear aplicado ao tipo de saída “diferença de pontos”, utilizando a configuração de validação cruzada com 20 *folds*, foi a técnica que mais conseguiu classificar mais instâncias corretamente, alcançando um nível de 75,1196% de assertividade. Esta técnica gerou um modelo de avaliação para geração das previsões decorrentes do conjunto de dados do NBB, conforme é demonstrado na Figura 2 (b).

Percebe-se na Figura 2 (b) que as entradas (fatores) que mais afetaram no resultado foram: rebotes, roubadas de bola e erros. Desta forma, conforme a proposta do estudo, regressão linear foi o método de classificação utilizado para aplicação e análise dos resultados.

5. Resultados

A fim de validar a técnica de previsão selecionada foi realizada uma simulação. Esta consiste na aplicação do modelo de regressão linear encontrado como mais adequado ao contexto sobre o conjunto de dados do NBB da temporada 2013/2014, compreendendo a execução de dois testes.

O primeiro consistiu em submeter os jogos do NBB diretamente à equação obtida através do resultado do teste com a técnica regressão linear, utilizando o arquivo “diferença de pontos”, aplicado à validação cruzada utilizando 20 *folds*. Por meio deste, intitulado “sem autoalimentação”, serão obtidos os resultados utilizando exatamente a mesma função de regressão e dados dos testes realizados. Já o segundo teste, intitulado “com autoalimentação”, consistiu na inserção dos resultados preditos dos jogos a cada rodada, isto para que a próxima rodada seja predita considerando todo o conjunto de dados somado aos resultados da rodada anterior.

Tais testes foram efetuados utilizando as médias finais das equipes na temporada 2013/2014, de forma a considerar que tais médias não foram alteradas conforme a execução das previsões. De certa forma, tal ação pode influenciar os resultados obtidos em relação aos ocorridos, sendo que comportamentos alheios às características consideradas, as quais se mostrem determinantes na melhora ou piora de determinada equipe, não foram considerados neste estudo de caso.

O teste “sem autoalimentação” foi elaborado a partir do conjunto de amostras existentes na fase classificatória da temporada 2013/2014 do NBB, totalizando 272 jogos. Como resultado da aplicação do modelo de previsão nessas partidas, obteve-se índice de assertividade de 68,01%, sendo que este modelo classificou 185 instâncias corretamente.

Já o teste “com autoalimentação” para a fase classificatória da temporada 2013/2014 do NBB, teve a primeira rodada de jogos da temporada, considerando o modelo utilizado no teste “sem autoalimentação”. Este retornou o percentual de 68,75% de assertividade. Dentre as 272 instâncias classificadas, este retornou o vencedor do jogo corretamente de acordo com o resultado real em 187 amostras.

6. Considerações Finais

O desenvolvimento deste projeto objetivou a previsão de partidas do NBB na temporada 2013/2014. Para isto, por intermédio dos resultados dos jogos entre os anos de 2008 e

2013, foram estudadas técnicas de aprendizado e de classificação de dados, no intuito de escolher a técnica que proporcionasse maior assertividade em testes.

Esta proposta foi reduzida em número de técnicas de classificação e predição em relação a outros trabalhos citados, como em Cao (2012). No entanto, a realização de testes em vários tipos de arquivos da mesma base, utilizando diferentes configurações assegurou que, para os métodos estudados, fosse escolhida a técnica que melhor se adaptou aos dados selecionados.

A escolha dos métodos de redes neurais artificiais e regressão linear também colaboraram para que os testes dos dados fossem inseridos em duas vertentes: aprendizagem de máquina e estatística aliada à mineração de dados. Desta forma, pode-se observar com clareza com qual das áreas conseguiria obter melhores resultados. Isto em contraposição aos trabalhos de Poropudas (2011) e West e Lamsal (2008), os quais utilizaram apenas técnicas de regressão.

O nível de acurácia do modelo deste projeto também alcançou um nível significativo de assertividade em relação aos demais. Com exceção do trabalho de Fabian (2008), os quais não apresentaram o número exato de amostras nem o percentual de acertos das classificações em relação ao número de instâncias, este projeto obteve em seu mais ajustado método, o maior nível de acertos dentre todos os trabalhos pesquisados. A proposta de Poropudas (2011), a qual apresentou níveis de 65 a 70% de acertos em uma população de 1230 amostras, serve como ponto de afirmação desta proposta que, para um total de 1463 instâncias, obteve níveis de assertividade acima de 75%.

Desta forma, este estudo consegue disponibilizar um modelo para a geração de predições, apresentando e utilizando a classificação de um método comprovado teoricamente, além de informação quantitativa referente aos processos de testes.

Apesar dos resultados positivos, algumas situações devem ser consideradas como limitações e que podem ter impactado nos resultados finais. O mais significativo fator não previsto é a mudança de jogadores entre clubes a cada temporada, o que provoca também novos índices de aproveitamento na média de cada equipe, em caso de considerar as médias das equipes como base de análise. Além disso, apesar de se tratar de conjunto de teste, o número de instâncias utilizadas nesta etapa pode ser considerado insuficiente.

Porém, essas limitações podem ser consideradas pontos de melhoria, pois todos eles podem ser corrigidos ou complementados no presente trabalho e, possivelmente, gerarão resultados ainda mais satisfatórios.

Sugere-se como extensões e complementos a esse trabalho a implementação de treinamentos e testes utilizando outros dados resultantes de cada jogo. Uma alternativa seria buscar o resultado da equipe no jogo, utilizando predições por embates e, a partir do comportamento da equipe em um referido jogo, poder prever resultados. Quanto a isto, poderiam ser utilizados também *scouts* dos jogadores que, somados, formariam o *scout* da equipe. Esta alternativa foi utilizada no trabalho de Fabian (2008). Também sobre isto, outra sugestão é permitir o cadastro de estatísticas por equipe a cada rodada, para que os resultados possam ser efetuados na medida em que os jogos aconteçam, com médias atuais das equipes.

Considerando as técnicas empregadas, apesar deste estudo propor a utilização da regra delta genérica, as redes neurais artificiais podem ainda ser exploradas quanto à adição de parâmetros que as tornem mais eficientes para classificação.

7. Referências

- Azevêdo, P. H. (2009). "O esporte como negócio: Uma visão sobre a gestão do esporte nos dias atuais". Revista Estudos, vol. 36, n. 9/10, p. 929-939. Disponível em <<http://revistas.ucg.br/index.php/estudos/article/viewFile/1167/810>>. Acessado em: 23 de abril de 2014.
- Cao, C. (2012). "Sports data mining technology used in basketball outcome prediction." A dissertation submitted in partial fulfilment of the requirements of Dublin Institute of Technology for the degree of M.Sc. in Computing (Data Analytics). 2012. 106p.
- Fabian, J. (2008). "Weka for advanced statistical prediction (WASP), A basketball match-up prediction application." CSULA (California State University) summer 2008 – fall 2008. 15p.
- LNB. (2016). "Sobre a LNB". Disponível em: < <http://www.lnb.com.br/lnb/>>. Acessado em: 23 de fevereiro de 2016.
- Pereira, V. W. (2011). "Predição, compreensão e situação de compreensão". Revista do Programa de Pós-Graduação em Letras da Universidade de Passo Fundo - v. 7 - n. 1 - p. 91-103 - jan./jun. 2011. 2011. 13p.
- Poropudas, J. (2011). "Kalman filter algorithm for rating and prediction in basketball". University of Helsinki - Faculty of Social Sciences Statistics - Master's Thesis, May 2011. 2011. 88p.
- Vendite, C. C.; Vendite, L. L.; Moraes, A. C. (2005). "Scout no futebol: Uma ferramenta para a imprensa esportiva". Rio de Janeiro: Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação. XXVIII Congresso Brasileiro de Ciências da Comunicação. 2005. 10p.
- WEKA, (2016). The University of Waikato. Weka 3: Data mining software in Java. Disponível em:<<http://www.cs.waikato.ac.nz/ml/weka/>>. Acessado em: 23 de fevereiro de 2016.
- West, B. T.; Lamsal, M. (2008). "A new application of linear modeling in the prediction of college football bowl outcomes and the development of team ratings". Journal of Quantitative Analysis in Sports, vol. 4, nro. 3, jul. 2008. University of Michigan. 21p.