

# Utilização de Redes Neurais Artificiais para previsão de campeão da *UEFA Champions League* baseada em estatísticas de temporadas anteriores

Lauro Tremea Culau<sup>1</sup>, Mateus Felipe Eisenkramer<sup>1</sup>

<sup>1</sup>Departamento de Computação  
Universidade de Santa Cruz do Sul (UNISC)  
Santa Cruz do Sul – RS – Brazil

{laurotc,mfelipe}@mx2.unisc.br

**Abstract.** *Statistics have been always part of the sports, and with the grown of technologies, it is becoming easier to get information to compare, use them to predict how good a player will be, or which team will has more chances to become champion. From this analisys, this article is using the more important statistics in soccer, and with neural networks, trying to predict the champion of the UEFA Champions League. The API provided by ADReNA system will be the tool used, together with back-propagation as the model for the RNA. As the League has not finished yet it is not possible to say if the result is correct, but for to have an acceptable prediction the team predicted to be the champion, has to be one of the teams which is playing in the League this year.*

**Resumo.** *As estatísticas sempre estiveram presentes nos esportes, e com o crescimento da tecnologia isso está ficando muito mais fácil de se obter, comparar e usá-las para prever o quanto um jogador pode render numa temporada ou qual o time que tem maiores chances de ser campeão. A partir desta premissa, o trabalho pretende utilizar algumas das principais estatísticas do futebol, juntamente com redes neurais para buscar prever o campeão da UEFA Champions League. A ferramenta utilizada será a API do sistema ADReNA usando como modelo de RNA o back-propagation. Como a Liga dos Campeões do ano atual ainda não está finalizada, não é possível de se dizer se o resultado é correto ou não. Porém, para ter uma previsão aceitável, o time que for o escolhido como possível campeão deve estar participando da competição atual.*

## 1. Introdução

O mundo esportivo é muito baseado em estatísticas, principalmente em esportes coletivos onde há vários fatores envolvidos, como por exemplo, tempo de jogo, ataques, o quanto a defesa está funcionando, acertos de passes, entre muitos outros, e muitos times se baseiam nas estatísticas de jogadores para decidir se o contratam ou não ou mesmo treinadores usam isso para avaliar o desempenho do time [Albert and Koning 2007] [Vaz de Melo et al. 2008].

Um grande exemplo disso são as ligas de esportes dos Estados Unidos, como por exemplo a Liga de Futebol Americano (NFL) e a Liga de Basquete Americana (NBA). Técnicos e auxiliares analisam estatísticas de jogadores que atuam para ver se eles tem

condições, vão agregar valor e trazer bons resultados para o time, e inclusive para determinar seu salário [Vaz de Melo et al. 2008]. Além disso estudam como os outros times jogam para prever o que podem fazer para neutralizar as jogadas do adversário, ou seja, analisam o máximo possível de dados para melhorar os resultados do time [Albert et al. 2005]. A mídia é muito interessada às estatísticas também, em que analistas esportivos baseiam seus comentários não apenas em opinião própria, mas sim com um embasamento de duas, três e até quatro temporadas passadas para analisar o jogador, ou o time em pauta.

Além da análise de estatísticas para a melhora do desempenho do time, elas são muito usadas no meio de apostas, onde especialistas analisam e buscam identificar quem é o time que tem mais chances de se tornar o candidato ao título e que movimenta uma grande quantia de dinheiro em todos os países e na maioria dos esportes, mesmo não sendo coletivos [Vaz de Melo et al. 2008]. Para se ter uma noção, no último campeonato de basquete universitário dos Estados Unidos, ocorrido em Março, com a precoce eliminação de times candidatos ao título e times coadjuvantes [ESPN 2016a], chamados de *underdogs*, avançando no campeonato, o mercado de apostas teve muitas variações, ganhos e perdas de grandes montantes de dinheiro [ESPN 2016b].

O presente trabalho busca de alguma forma utilizar Redes Neurais Artificiais (RNAs) para prever campeões de ligas baseado apenas em estatísticas. Visto que elas são largamente usadas na cobertura esportiva pela mídia, na contratação de jogadores e nas estratégias de jogo, pode-se buscar usá-las também para prever resultados de campeonatos.

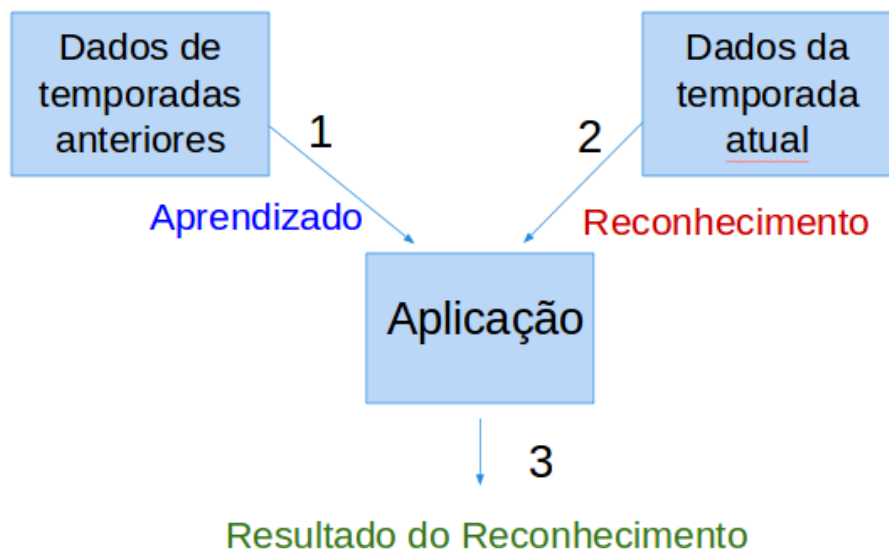
O restante do trabalho está dividido da seguinte maneira: Na Seção 2 está descrito o Domínio da Aplicação, com a explicação detalhada do objetivo do trabalho. Na Seção 3 é explicado o modelo de Rede Neural Artificial (RNA) a ser utilizado. A Seção 4 apresenta a ferramenta ADReNA, cuja API será utilizada no trabalho. A Implementação do trabalho é detalhada na Seção 5 e os Resultados obtidos pelo trabalho são descrito na Seção 6. Por fim é na Seção 7 é apresentada a Conclusão e as Referências utilizadas no trabalho.

## **2. Domínio da Aplicação**

A aplicação a ser desenvolvida tem um domínio focado na área de previsão esportiva e apostas. Usando premissas as premissas citadas anteriormente na Introdução do trabalho e definindo o esporte a ser usado para os testes, que será o futebol, a proposta é de utilizar uma Rede Neural Artificial (RNA) para treinar o programa com as estatísticas dos times de campeonatos passados, e a partir dos dados do campeonato deste ano, determinar o seu futuro campeão.

Utilizando como entrada de dados as estatísticas com os dados médios de cada time da fase de grupos de temporadas passadas, como por exemplo, posse de bola, chutes a gol, faltas cometidas, vitórias, derrotas, entre outros, e com seus respectivos campeões ensinar a rede. E a partir das estatísticas médias da temporada atual, buscar prever o campeão, que ainda não está definido. A Figura 1 apresenta a ideia básica do funcionamento da aplicação.

Para a realização do trabalho, os dados utilizados foram os da *UEFA Champions League*, mais conhecida no Brasil como Liga dos Campeões da Europa. O funcionamento



**Figura 1. Diagrama de funcionamento da aplicação. Fonte: Autores**

da Liga é baseado em fase de grupos e fase final. A fase de grupos são oito grupos de quatro times cada, que fazem dois jogos. Os primeiros times de cada grupo se classificam para a fase final, que após sorteio se enfrentam e jogam jogos eliminatórios até a final, que é realizada em jogo único em uma cidade pré-definida antes do torneio começar [UEFA 2016].

Os dados estatísticos foram obtidos através do site [WhoScored 2016], que mantém dados atualizados sobre diversas ligas e torneios de futebol no mundo inteiro. A partir deste site, foram coletados os dados pertencentes às temporadas de 2010 à 2016 da fase de grupos da Liga dos Campeões. Com os dados obtidos, foram criados arquivos *.json* para cada temporada, com os seus respectivos valores, divididos por time, juntamente com o campeão de cada temporada.

O site apenas dispunha dos dados separados por um apanhado geral de dados, os dados defensivos e ofensivos. Para encontrar os dados necessários, foi necessário fazer o *merge* dos arquivos para as informações estatísticas do time conter todas as necessárias para os testes. Os dados obtidos e ajustados que serão utilizados pela aplicação para o aprendizado e reconhecimento são mostrados na Figura 2 em um arquivo *.json* referente a um time, e listados mais detalhadamente abaixo:

- Id único do time;
- Nota geral, corresponde a média das notas da atuação do time em questão durante os jogos em que participou;
- Média de chutes por jogo;
- Média de chutes no gol por jogo;
- Total de cartões vermelhos;
- Total de cartões amarelos;
- Média de posse de bola por jogo;
- Média de jogadas aéreas ganhas por jogo;
- Média de roubadas de bola por jogo;

- Média de interceptações por jogo;
- Média de faltas por jogo;
- Média de escanteios por jogo;
- Média de chutes concedidos por jogo;
- Média de dribles ganhos por jogo;
- Média de faltas cometidas por jogo;
- Porcentagem de passes realizados com sucesso por jogo.

```
{
  "ranking": 1,
  "teamId": 52,
  "teamName": "Real Madrid",
  "rating": 7.2926874,
  "redCard": 0,
  "yellowCard": 17,
  "possession": 0.54896915,
  "aerialWonPerGame": 7.1666665,
  "shotsPerGame": 21.333334,
  "shotOnTargetPerGame": 7.8333335,
  "dribbleWonPerGame": 12.833333,
  "foulGivenPerGame": 12.666667,
  "shotsConcededPerGame": 9.666667,
  "tacklePerGame": 28,
  "interceptionPerGame": 26.833334,
  "foulsPerGame": 17.666666,
  "offsideGivenPerGame": 3.3333333,
  "passSuccess": 0.825821
},
```

**Figura 2. Exemplo dos dados de um time dentro um arquivo .json. Fonte: Autores**

O único ano no qual o campeão não está escrito no arquivo é o ano de 2016, que é o ano cujo campeão o trabalho busca prever, visto que ainda não há um campeão decidido. O restante do funcionamento da implementação da aplicação está descrito na Seção 5.

### **3. Modelo *back-propagation* RNA**

Redes Neurais Artificiais (RNA) são técnicas aplicadas na computação que se baseiam no funcionamento do cérebro. Inspiradas pela hipótese de usar a ideia de neurônios, foram iniciadas pesquisas nessa área, usando o neurônio como modelo para a criação do modelo matemático em 1943 por McCulloch e Pitts. A rede neural é composta por unidades, chamadas de neurônios, conectadas por *links* e estes se comunicam diferentes unidades. Cada um desses *links* possuem um peso associado que determina a força da conexão, e servem de dados de entrada para os neurônio, que aplicam cálculos sobre eles, e partir

destes cálculos, resultados e propagação dos resultados que há o aprendizado da rede neural [Russell and Norvig 2009].

Na maioria das RNAs as entrada seguem em apenas um sentido, as entradas são processadas e geram saídas, elas são conhecidas como *feed-forward network*. Elas são normalmente organizadas em camadas em que cada camada recebe o valor de saída da camada anterior. Nas redes de múltiplas camadas existem as camadas internas que não se conectam diretamente com a saída, são apenas usadas para realizar os cálculos de aprendizagem. Existem tanto as RNAs com apenas uma camada, em que a entrada é diretamente conectada com a saída, ou então as multicamadas onde há mais de uma camada interna e a camada de entrada não se conecta com a de saída [Russell and Norvig 2009].

No trabalho proposto, será utilizada uma RNA com *back-propagation*. Para entender, é necessário explicar o conceito de redes neurais com multi camadas (*multilayer networks*). O princípio de uma rede multi camadas é de que a saída de uma é a entrada de outra até que as saídas da última camada cheguem até camada de saída e esta tenha como saída o resultado final. Porém, com as camadas internas não é possível medir o erro pois não há como comparar com os dados de treinamento. Felizmente, é possível fazer o *back-propagate* do erro que é medido na camada de saída e enviá-lo para as camadas internas. Com esse envio para trás, os pesos são atualizados e a saída pode se tornar diferente ocasionando um erro menor e que seja aceitável [Russell and Norvig 2009].

O treinamento da rede é feito a partir de um *training set* onde os dados são enviados para a camada de entrada, calculados pelas internas e enviados para a de saída, até que a rede aprenda o padrão, ou obtenha uma taxa de erros bem baixa. A partir deste aprendizado, pode-se testar novas entrada, e testar se elas serão detectadas corretamente com o padrão aprendido pela rede.

#### **4. Ferramenta ADReNa**

O Ambiente de Desenvolvimento de Redes Neurais Artificiais (ADReNA) é uma ferramenta desenvolvida por Matheus Henrique Kist, ex-aluno do curso de Ciência da Computação da Unisc, no ano de 2013. O principal objetivo desta aplicação é de fornecer uma ferramenta para desenvolvimento e testes de RNAs. A ferramenta desenvolvida permite a modelagem de redes *Backpropagation* e *Kohonen*, e conter todos os métodos referentes às RNAs, desde a sua criação, o treinamento e a execução para o reconhecimento de padrões [Kist and Frozza 2013].

Utilizando a API em linguagem Java, disponibilizada pelo autor, a proposta do trabalho, é de treinar a rede com as estatísticas médias de cada time durante a fase de grupos da Liga dos Campeões de temporadas passadas, e a partir dos dados da temporada atual, testar a RNA para verificar se a rede conseguiu reconhecer os padrões, e desta forma buscar prever o campeão da temporada atual, que ainda não está definido.

#### **5. Implementação**

A implementação do trabalho foi realizada utilizando a linguagem Java, desta forma, foi necessário o uso da API do ADReNA na linguagem Java. A aplicação recebe arquivos do tipo *.json* com os dados de cada temporada do campeonato, e considerando os dados estatísticos descritos na Seção 2 foram aplicadas uma série de transformações para formatar

os dados da maneira adequada para o uso de como o ADReNA recebe aceita através da sua API.

Como é possível de se visualizar na Figura 2 e também por serem usados dados médios de toda a fase de grupos da Liga, há muitos valores decimais. Para o melhor funcionamento, todos os valores decimais foram arredondados para inteiro e a partir deles feitos os tratamentos de intervalos, nos casos de porcentagem, como posse de bola e a quantidade de acerto de passe. Nesses casos, foram definidos intervalos de valores e estes salvos para serem os valores a serem usados para estes dados estatísticos e para o envio dos dados para o ADReNA da forma que são aceitos pela API, os dados estatísticos foram convertidas para binário.

Após a conversão para *bits*, os dados das temporadas anteriores são enviados para a API do ADReNA, esta que realiza o aprendizado e exporta de volta para a aplicação os arquivos de treinamento, da estrutura da rede e de como é feito o reconhecimento da rede em arquivos *.json*.

Para o treinamento, foram utilizadas as competições entre os anos de 2010 e 2015. Os dados estatísticos de cada time contém 74 bits, que são gerados a partir da conversão que foi explicada acima. Como cada temporada possui 32 times, a entrada que é enviada para o ADReNA é de 2368 bits, que contém todos os dados estatísticos, e a saída retornada por ele é de 12 bits que é o ID do time campeão, que posteriormente é convertido para ser apresentado ao usuário. Os 12 bits foram definidos visto que o ID dos times varia de números pequenos até números maiores que precisam de maior número de bits para serem representados.

Feito o treinamento, os arquivos são salvos, e ao chamar a função de reconhecimento utilizando o arquivo da temporada atual convertido para *bits*, a aplicação novamente faz o chamado à API do ADReNA para enviar os dados, e para isso a API importa a estrutura de rede e a estrutura de como é feito o reconhecimento. Com esses dados, ela realiza o reconhecimento e a API retorna uma *string* com o valor dos *bits* da resposta, estes que são convertidos para um número inteiro. O valor obtido nessa conversão corresponde ao ID do time que a RNA preveu para ser o campeão da temporada atual do campeonato, a partir disso é necessário buscar qual o nome do time o qual possui o ID obtido, e este será o campeão, baseado nas estatísticas.

## 6. Resultados

Para obtermos uma visão mais abrangente de qual seria o comportamento da rede neural tendo como entrada os dados estatísticos de todos os times presentes na fase de grupos da Liga dos Campeões da Europa entre os anos de 2010 e 2015, testamos diversas arquiteturas. Tendo em vista que para o teste de reconhecimento do campeão foi utilizado os dados do ano de 2016, abaixo consta a tabela com os resultados obtido na Tabela 1.

Mesmo variando pouco o número de neurônios nas camadas intermediárias o resultado sempre foi o mesmo, a rede retornava o campeão para o ano de 2016 sendo o time com ID 5. Infelizmente, dentre os times que foram utilizados como entrada para a rede, nenhum tinha o id mencionado, o que significa que a rede não pode apreender com sucesso a prever o campeão da liga com os dados estatísticos providos a ela.

Porém, quando a rede neural foi treinada utilizando somente os dados de um ano

**Tabela 1. Resultados dos Testes com diferentes camadas de neurônios**

Camada 1	Camada 2	Camada 3	Id Campeão obtido
4	6	5	5
4	6	5	5
6	5	4	5
6	5	5	5

da competição utilizada no estudo, o retorno do reconhecimento do campeão para o ano de 2016 sempre era o campeão dos dados utilizados como entrada. O que significa que a rede conseguiu apreender a reconhecer o campeão com sucesso somente quando era treinada com os dados da competição em um ano específico. Não sendo capaz de inferir corretamente o campeão quando recebia dados de competições de vários anos.

Não foi possível variar muito o número de camadas intermediárias nem o de neurônios por camada por falta de recursos computacionais. Quando a estrutura da rede era configurada com mais camadas ou neurônios, a API do ADReNA acusava falta de memória para prosseguir com o treinamento da rede neural.

## 7. Conclusão

A previsão do campeão com base nos dados estatísticos da fase de grupos da Liga dos Campeões da Europa prove um desafio grande. Tanto pela falta de relação entre os dados estatísticos utilizados com o campeão em si quanto pela impossibilidade de se utilizar uma estrutura de rede neural mais complexa.

Acredita-se que os dados esportivos como número de cartões amarelos, número de chutes a gol, posse de bola, pouco refletem em quem será o campeão de uma competição tão disputada como a Liga dos Campeões. Uma alternativa talvez seja utilizar outras bases de dados, baseadas em conhecimento humano sobre os times. Como número de apostas realizadas a favor do time em questão, média de torcedores do time que compareceram aos jogos ou até a avaliação de notícias em jornais ou revistas sobre os jogos do time.

## Referências

- Albert, J., Bennett, J., and Cochran, J. (2005). *Anthology of Statistics in Sports*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics.
- Albert, J. and Koning, R. (2007). *Statistical Thinking in Sports*. CRC Press.
- ESPN (2016a). Loss by bettors' choice spartans good news for vegas sportsbooks. Disponível em: <[http://espn.go.com/chalk/story/\\_/id/15009976](http://espn.go.com/chalk/story/_/id/15009976)>. Acesso em: 25 de abril de 2016.
- ESPN (2016b). Michigan state sees most vegas action ahead of ncaa tournament. Disponível em: <[http://espn.go.com/chalk/story/\\_/id/14990303](http://espn.go.com/chalk/story/_/id/14990303)>. Acesso em: 25 de abril de 2016.
- Kist, M. and Frozza, R. (2013). *Ambiente de modelagem e teste de redes neurais artificiais*. Universidade de Santa Cruz do Sul (UNISC).

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice Hall Press, Upper Saddle River, NJ, USA, 3rd edition.

UEFA (2016). Champions league. Disponível em: <<http://www.uefa.com/uefachampionsleague/>>. Acesso em: 25 de abril de 2016.

Vaz de Melo, P. O., Almeida, V. A., and Loureiro, A. A. (2008). Can complex network metrics predict the behavior of NBA teams? In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 695–703, New York, NY, USA. ACM.

WhoScored (2016). Football statistics. Disponível em: <<http://whoscored.com>>. Acesso em: 25 de abril de 2016.