



REDE NEURAL APLICADA À ANÁLISE DE DADOS DE PERCURSOS DE CICLISTAS PARA PREVISÃO DE MÉTRICAS DE DESEMPENHO

Emanuela da Rosa Machado¹

Rodrigo Cesar Nunes Maciel²

Resumo: O ciclismo é um esporte que apresenta um crescimento exponencial. Pode-se demonstrar este fato pelo crescimento da produção e venda de bicicletas desenvolvidas para a prática deste esporte, conseqüentemente observa-se uma tendência na busca por ferramentas computacionais de auxílio a prática da atividade amadora e / ou profissional. Alguns aplicativos existentes no mercado permitem ao usuário fazer o registro das atividades realizadas, porém a previsão de métricas relevantes ao esporte não são abordadas. Sendo assim algumas características competitivas do ciclismo e a demonstração do desempenho do praticante não podem ser previstas. Desta forma o objetivo deste trabalho baseou-se em implementar uma Rede Neural Artificial tipo *Perceptron* de Múltiplas Camadas para previsão de métricas de desempenho de um grupo de ciclistas da cidade de Criciúma, visando o desenvolvimento de uma ferramenta computacional de auxílio à prática deste esporte. A avaliação da Rede Neural Artificial que foi realizada por meio de uma análise do coeficiente de determinação R^2 , da dispersão dos valores medidos em comparação aos valores previstos, obteve-se o valor percentual de 97%. Além de análises em relação ao Erro Quadrático Médio e do Erro Absoluto Médio. A validação da Rede Neural Artificial foi realizada por meio de um novo conjunto de dados de atividades já registradas, os valores reais e os valores previstos foram comparados, sendo os obtidos pelo modelo muito próximo dos valores reais. O artigo presente obteve como resultado o desenvolvimento de uma ferramenta computacional de auxílio a prática do ciclismo para a previsão da velocidade média de circuitos pré-determinados, considerando os fatores que apresentam uma correlação negativa com a variável em estudo. Os valores previstos pelo modelo, por meio de sua validação, apresentaram um valor de erro muito baixo em comparação aos valores reais.

Palavras-Chave: Rede Neural Artificial. Regressão Linear. Ciclismo. Python. Keras.

1 INTRODUÇÃO

O ciclismo sempre esteve presente na vida das pessoas, seja durante a infância com os pais ensinando a como andar de bicicleta ou na vida adulta como meio de transporte. De acordo com o Instituto de Pesquisa Econômica Aplicada (IPEA), no Brasil há mais bicicletas que carros, respectivamente 50 milhões contra 41

¹ Graduanda UniSATC. E-mail: emanuelarmachado@gmail.com

² Prof. UniSATC. E-mail: rodrigo.maciel@satc.edu.br



milhões. O total de viagens feitas por bicicletas é em torno de 7%, com potencial de atingir 40% (IPEA, 2017). Algo que tem crescido de forma acentuada nos últimos anos é o ciclismo como forma de esporte. O número de pessoas que veem aderindo à prática tem aumentado de forma exponencial.

Conforme os dados disponibilizados pela Abraciclo, a indústria brasileira superou até novembro de 2019 a projeção de produção total do ano. Fato que retrata claramente um crescimento consistente da demanda por bicicletas brasileiras de médio e alto valor agregado. No ranking geral a categoria mais produzida foi a *Mountain Bike* (MTB), ideais para o uso em trilhas e terrenos acidentados, com 425.372 unidades e 47,3% de participação. A categoria Urbana ficou em segundo lugar com 330.235 unidades e 36,7% de participação (Abraciclo, 2019). No ano de 2020, as medidas de isolamento social contra o Covid-19 impulsionaram as vendas do segmento, crescendo em 118% entre os meses de junho e julho, na comparação do mesmo período no ano de 2019 (Abraciclo, 2020).

O crescimento da produção de bicicletas desenvolvidas à prática do esporte, indica o aumento dos ciclistas adeptos e conseqüentemente observa-se uma tendência no aumento da demanda por aplicativos que registrem as atividades realizadas. Existem inúmeros aplicativos disponíveis para o registro de atividades físicas, incluindo o ciclismo. Dados sobre a prática desse esporte indicam que os ciclistas buscam superar seus próprios limites, diminuindo o tempo de trajeto, alcançando velocidades médias cada vez maiores, aumentando suas rotas e o nível de dificuldade envolvida nos percursos, tudo isso por um motivo, vencer a si mesmo e suas próprias limitações.

Os aplicativos disponíveis tem como característica comum fazer apenas o registro dos dados, sem apresentar previsões conclusivas em relação as próximas práticas, logo as características competitivas do praticante não são atendidas. Desta forma, não é possível gerar informações sobre as métricas de desempenho relacionadas ao seu processo evolutivo ou a previsão destas em relação as próximas atividades. Este fator é de extrema importância para aqueles que praticam e buscam uma constante evolução.

Existem novas tecnologias computacionais que podem auxiliar os atletas na prática de seus esportes. Uma tecnologia com grandes destaques atuais e inúmeras aplicações são as Redes Neurais Artificiais (RNA). Por meio de uma RNA, dados, como distância percorrida, elevação, velocidade, tempo de trajeto, entre



outros, poderão ser filtrados e analisados com o objetivo de gerar uma previsão de métricas relacionadas ao desempenho do atleta a curto prazo. As RNAs têm tido papel importante em diferentes segmentos, sendo amplamente utilizada por ser uma tecnologia atual e com diversos escopos. Na mecatrônica, o uso dessa tecnologia é visto em procedimentos de automação industrial, controle de próteses, leitura de dados, entre outros.

Neste contexto e com a necessidade de aplicativos que realizem a previsão de métricas relacionadas ao desempenho por meio dos trajetos selecionados, surgiu a ideia da utilização dessa tecnologia como meio de solucionar a dificuldade de acompanhar e prever a evolução do atleta a curto prazo. A aplicação em questão poderá realizar uma análise de dados já existentes, coletados a partir de aplicativos disponíveis no mercado. Informações essas que poderão ser disponibilizadas para que o usuário analise o seu desempenho e auxilie o seu treinamento, ajudando o ciclista a superar seus próprios limites. Contribuindo dessa forma com todos aqueles que utilizam o ciclismo como forma de esporte.

Com base neste contexto, este trabalho teve como objetivo implementar uma RNA tipo *Perceptron* de Múltiplas Camadas para previsão de métricas de desempenho de um grupo de ciclistas da cidade de Criciúma, visando o desenvolvimento de uma ferramenta computacional de auxílio à prática deste esporte.

2 REVISÃO BIBLIOGRÁFICA

Nesse capítulo serão abordados temas de suma importância para o desenvolvimento do projeto em questão. A pesquisa dos temas foi baseada em periódicos nacionais e internacionais, livros, teses e dissertações.

2.1 O CICLISMO

A prática de atividades esportivas traz benefícios a saúde física, a longo prazo temos o aumento das capacidades cardíacas e pulmonares, além de uma maior eficiência do sistema musculoesquelético, levando a um bom estado geral do organismo. As mudanças positivas em fatores comportamentais também devem ser ressaltadas, como a diminuição do nível de ansiedade e tensão (ALMEIDA, 2004).



O ciclismo vem conseguindo um espaço considerável em nosso país, de acordo com os dados do IBGE, dos esportes que apresentaram frequência semanal maior, ou seja, quatro vezes ou mais por semana, o ciclismo está em segundo lugar com 39,8% (IBGE, 2015).

2.1.1 Modalidades do ciclismo

Sua versatilidade é um dos pontos que merece destaque ao mencionar a difusão da prática, afinal essa atividade é indicada para todo tipo de pessoa que tenha condições físicas e motoras para pedalar. “Por sua prática ser ao ar livre tornou-se ainda mais atrativo ao público geral” (IBGE, 2015). Como demonstrado por meio dos dados da Abraciclo, a produção de bicicletas para o transporte e para diferentes modalidades do esporte tem expandido. Dentre as diversas modalidades do ciclismo, optou-se por dar destaque as duas mais comuns:

- **Ciclismo de Estrada:** Modalidade praticada com as chamadas *speeds*, são bicicletas caracterizadas por seus pneus bastante finos e por serem mais leves que as tradicionais. Sua aerodinâmica favorece o ganho de velocidade e estabilidade nas estradas, sendo essencialmente direcionadas para a prática da modalidade e são mais difíceis de pilotar (BITTENCOURT, 2019).

- **Mountain Bike:** É a modalidade mais popular no Brasil, devido a possibilidade de uso em diversos tipos de terrenos, sendo que as mesmas são desenvolvidas para o uso em áreas de terra, porém pode-se fazer uso em asfalto com uma velocidade reduzida. Elas contam com suspensões e com pneus largos com cravos que dão aderência ao pedalar na terra (BITTENCOURT, 2019).

A partir do crescimento do esporte, a procura por meios de registrar os trajetos executados e acompanhar os dados de evolução vem sendo cada vez maior. Com a utilização do sinal de GPS (*Global Positioning System*) presente nos aparelhos celulares, as informações dos sinais emitidos por satélites estão disponíveis a todos. Seu uso permite determinar as coordenadas do terreno e com a utilização de um aplicativo que faça o registro destas informações, transformou-se em um feito incrivelmente simples para cada ciclista obter registros de suas atividades. Além de informações de rota, é possível o registro de dados de elevação total, distância



percorrida, tempo de trajeto, entre outros (SEGADILHA, 2014). O registro desses dados é pertinente, pois por meio de uma informação, o desenvolvimento de várias aplicações e análises torna-se viável, incluindo a situação problema do artigo presente.

2.2 APRENDIZADO DE MÁQUINA

Data Mining é um processo de extração de padrões escondidos de grandes conjuntos de dados. Buscando relacionar esses conjuntos usando a estatística, matemática, inteligência artificial e técnicas de aprendizado de máquina (BERSON, 2000). Enquanto o *Data Mining* foca-se na exploração dos dados e descobertas de propriedades não conhecidas, o *Machine Learning* se concentra na previsão e classificação com base nas propriedades aprendidas a partir de um conjunto de dados de treinamento (SCHNEIDER, 2016). Através das informações apresentadas, podemos definir o *Machine Learning* como a automatização de uma interação a partir de uma inteligência artificial.

Logo é possível dizer que a procura por formas de agrupar e esquadrihar esses dados, com a finalidade de realizar buscas, solucionar problemas e realizar previsões, gerou o surgimento de várias técnicas baseadas nos conceitos explorados. Dentre essas técnicas podemos citar a criação das RNAs, pode-se classificar o processo citado como uma ferramenta baseada em *Machine Learning*.

2.3 REDES NEURAIAS ARTIFICIAIS

No cenário global atual, um fator que ganhou relevância é a análise de dados. Estamos soterrados em uma quantidade exorbitante de dados em todos os segmentos. “A própria internet representa um diagrama grande de conhecimento que contém uma enorme enciclopédia de referências cruzadas” (GRUS, 2018). Por si só, os dados, não apresentam valor agregado, são apenas informações soltas, todavia é possível transformá-los em respostas, sendo necessário realizar os questionamentos corretos. Os conjuntos de dados diferem de inúmeras formas. Os atributos usados para descrever os objetos de dados podem ser de diferentes tipos e os conjuntos de dados podem ter características especiais (TAN; STEINBACH; KUMAR, 2009). Com o uso da tecnologia correta a análise de dados torna-se uma ferramenta relevante no



desenvolvimento de várias aplicações. As redes neurais são uma das ferramentas na vanguarda para a criação de aplicações a partir de um conjunto de dados.

Fundamentado na composição dos neurônios, McCulloch e Pitts acrescentaram a ideia da resposta de um neurônio individualmente considerado pode ser descrita por uma sentença lógica proposicional, que descreve na presença de quais estímulos um neurônio é ativado. Por meio desse rudimento, esses neurônios são descritos como os elementos de uma rede neural. Esta rede poderá aprender com os sinais recebidos, por meio de mudanças na estrutura da rede motivada (MCCULLOCH; PITTS, 1943). Em concordância a Grus (2018), pode-se definir que a RNA é um modelo preditivo motivado pela forma como o cérebro funciona. Baseado no entendimento da atividade cerebral, tornou-se possível executar atividades complexas, as quais o cérebro é capaz de executar com elevada efetividade, tais como: o reconhecimento de padrões, a percepção e o controle. As RNAs consistem de neurônios artificiais que desenvolvem cálculos similares sobre suas entradas.

Citando Haykin (2001), “uma RNA (RNA) é um sistema de processamento massivamente paralelo, composto por unidades simples com capacidade natural de armazenar conhecimento e disponibilizá-lo para uso futuro”. Essa técnica implementa padrões de detecção e algoritmos de aprendizado de máquina para construir modelos de predição para base de dados históricos em larga escala (PINHEIRO, 2008). O campo da Inteligência Artificial não teria a desenvoltura apresentada atualmente sem o uso das RNAs.

Apesar da existência de outros algoritmos para a resolução de problemas de regressão, tais como: regressão linear, k Vizinhos mais Próximos, Árvore de Decisão e Máquina de Vetores de Suporte, a escolha pela utilização de RNAs deu-se por sua capacidade de aprendizado em qualquer segmento e por não haver limites para a quantidade de informação que ela possa processar, sendo este um projeto para o desenvolvimento de uma ferramenta computacional de auxílio ao esporte há a possibilidade de que novos dados possam ser insertos no modelo. A técnica de RNA tem sido aplicada com sucesso na área dos esportes, como pode ser verificado nos trabalhos publicados (TEIXEIRA, 2019.), (ALVES, 2010) e (JUNIOR; SANTIAGO; TINÓS, 2020).



2.4 ARQUITETURA REDE NEURAL

A arquitetura de uma RNA é definida pela forma com que os neurônios estão estruturados (interconectados). Esse fator é determinante por estar intimamente relacionado ao algoritmo de aprendizagem a qual será feito o uso para treiná-la. Dentre as redes mais comuns podemos citar as Redes Neurais *Perceptron* de Múltiplas Camadas, as Redes Neurais *Convolucionais* e as Rede Neurais Recorrentes. A Rede Neural *Perceptron* de Múltiplas Camadas, é considerada a arquitetura mais comum, sendo utilizada para resolução dos mais diversos tipos de problemas, este modelo foi proposto por Rosenblat em 1958, sendo considerado como uma extensão do *Perceptron* de camada única, como será visto na subseção 2.3.2.

A Rede Neural Convolutiva é uma variação da Rede Neural *Perceptron* de Múltiplas Camadas. Sua inspiração veio do processamento biológico de dados visuais e vem sendo aplicada com sucesso no processamento e análise de imagens digitais. Conhecida por sua capacidade de aplicar filtros em dados visuais, mantendo a relação de vizinhança entre os pixels da imagem ao longo do processamento da rede. A Rede Neural Convolutiva tem se tornado amplamente utilizada, principalmente nas aplicações de classificação, detecção e reconhecimento em imagens e vídeos (VARGAS, 2016).

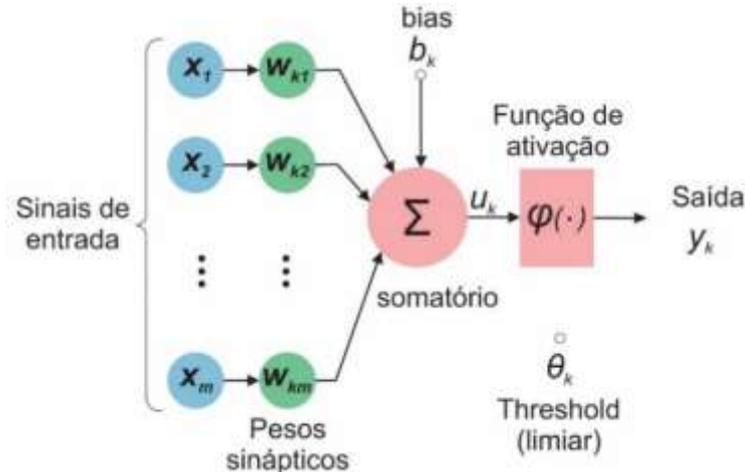
Da necessidade de realizar previsões sobre os dados sequenciais e aproveitar da estrutura de RNAs, surgiu a arquitetura das redes neurais recorrentes. Onde as RNAs tradicionais apresentam limitações, a Rede Neural Recorrente é utilizada para a solução de problemas com eventos sequenciais, quando um determinado dado ou valor não pode ser avaliado de maneira isolada. São redes com laço de recorrência (*loops*), permitindo que as informações persistam (GOODFELLOW, 2016). Exemplos do uso dessa rede podem ser visto na previsão da taxa de câmbio (COSTA, 2020), na previsão de ataques epiléticos (FREITAS, 2019), na utilização para análise de sentimentos em textos (KUNAS, 2020), entre diversos outros exemplos que fazem uso de dados sequenciais.

2.4.1 Redes Perceptron

Haykin (2001) define um neurônio como uma unidade de processamento de informações, sendo fundamental para a operação de uma rede neural. A rede neural mais simples conhecida é a *Perceptron*, que se aproxima de um único neurônio com “m” entradas binárias. Ela calcula a soma ponderada de suas entradas e “dispara” se essa soma for zero ou maior (GRUS, 2018).

O diagrama da Fig. 1 ilustra o modelo de um neurônio, que seria a base para o projeto de RNAs. Os sinais de entrada, chamados de *inputs*, são representados pelo x . A letra w indica os respectivos pesos de cada entrada. O somatório é uma combinação linear dessas entradas ponderada pelos pesos, gerando dessa forma a saída (*output*). Por fim, ele é avaliado por uma função de perda. Em resumo as entradas são multiplicadas pelos pesos, em seguida o resultado das multiplicações é somado e após passarem pela função de ativação, a saída final é obtida.

Figura 1: Modelo não-linear de um neurônio



Fonte: Adaptado de Haykin (2001)

O *Perceptron* consiste em dois tipos de nodos: “nodos de entrada que são usados para representar os atributos de entrada e um nodo de saída, que é usado para representar a saída do modelo” (TAN; STEINBACH; KUMAR, 2009). Os nodos de uma arquitetura de rede neural são conhecidos comumente como neurônios ou unidades. Sob a percepção de RNAs um neurônio é uma unidade computacional que recebe uma entrada, realiza uma operação matemática e entrega uma saída.



A descoberta dos limites da capacidade de representação do perceptron é usualmente apontada como um dos principais fatores para a redução na pesquisa em redes neurais nas décadas de 1970 e boa parte de 1980 (RUSSELL; NORVIG, 2010). A partir da segunda metade da década de 1980, os limites da capacidade de representação do *Perceptron*, foram superados devido à adoção de diferentes modelos do neurônio e de estruturas complexas formadas por múltiplas camadas de neurônios.

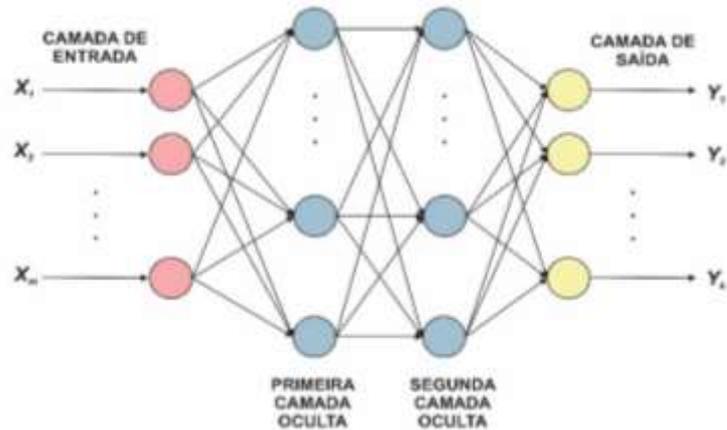
Dada a diversidade de modelos existentes, não será viável apresentar aqui uma análise detalhada das principais variantes e alternativas de modelagem introduzidas desde o fundamento da teoria em 1943.

2.4.2 Rede Neural *Perceptron* de Múltiplas Camadas

A Rede Neural *Perceptron* de Múltiplas Camadas possui uma complexidade maior do que a de um modelo *Perceptron*. Das muitas topologias, esta é rede comumente utilizada para modelagem de processos. Sua distinção é dada pela presença de camadas intermediárias, denominadas também de camadas ocultas. São nessas camadas que os neurônios efetivamente atuam como unidades processadoras. Essa arquitetura torna a rede capaz de extrair estatísticas de ordem mais alta, não correspondendo à camada de saída. O uso de uma ou mais camadas intermediárias permite o aumento do poder computacional de processamento não-linear e de armazenamento da rede (LIU, 2008).

Nesse contexto, podemos afirmar que o conjunto de saídas dos neurônios de cada camada é utilizado como entrada para a camada seguinte. Cada unidade executa uma soma ponderada de suas entradas e passa este nível de ativação por meio de uma função de transferência para produzir sua saída.

Os sinais de saída dos neurônios da última camada, constituem-se na resposta global da rede para o padrão de entrada fornecido. O fluxo de sinal da rede progride para frente. As entradas surgem da esquerda e vão para a direita passando por cada uma das camadas existentes (HAYKIN, 2001). Um requisito importante para o uso de uma RNA é que deve existir uma relação entre as entradas e saídas propostas. A Fig. 2 representa o modelo de uma rede neural múltiplas camadas.

Figura 2: Modelo Rede Neural *Perceptron* de Múltiplas Camadas

Fonte: Adaptado de Kolisny (2018)

No momento da criação de uma rede neural, uma função de ativação é definida, vindo a ser fixa. Essa função é responsável pela alteração do sinal recebido em cada neurônio. Os neurônios ocultos ou os de saída, conhecidos como nós computacionais, são desenvolvidos com função de realizar cálculos.

Primeiramente faz-se o cálculo do sinal funcional, sendo esse expresso por uma função não-linear do sinal de entrada e dos pesos sinápticos relacionados ao neurônio em questão. Os pesos sinápticos são atribuídos a cada um dos neurônios durante o treinamento da RNA, podendo ser fixos ou variáveis. Com base no sinal de erro encontrados entre as respostas desejadas e aquelas obtidas pelos neurônios de saída, de acordo com o método de treinamento utilizado, os pesos e limiares dos neurônios são ajustados (HAYKIN, 2001).

2.5 MÉTODOS DE APRENDIZAGEM

As RNAs possuem inúmeras características, sendo a sua capacidade de aprendizado uma das mais marcantes. Seu aprendizado é realizado por meio de um processo iterativo de ajustes de pesos e limiares. Existem processos que são capazes de ajustar não apenas os pesos da rede, mas também sua arquitetura e as funções de ativação dos neurônios (VON ZUBEN, 1996).

A aprendizagem é o processo pelo qual os parâmetros livres de uma rede neural são adaptados por um mecanismo de apresentação de estímulos fornecidos pelo ambiente no qual a rede está inserida, citando Haykin (2001). O tipo de



treinamento é definido pela forma na qual os parâmetros são modificados. A definição do autor citado ao conceito de aprendizagem implica na seguinte sequência de eventos:

- Apresentação de estímulos a rede neural;
- Alteração dos parâmetros livres da rede;
- Novo padrão de resposta ao ambiente.

Pode-se definir os paradigmas de aprendizagem em três principais: supervisionada, não-supervisionada e por esforço. Na aprendizagem supervisionada há um conjunto de amostras de entradas-saídas, onde o ambiente é desconhecido, a rede se ajusta pela combinação do sinal de entrada com o sinal de erro sendo essa a diferença entre a saída desejada e a que a rede fornece.

Na aprendizagem não-supervisionada, faz-se uso da aprendizagem competitiva, na qual os neurônios competem entre si a fim de serem mais ativos, por meio de algoritmos auto-organizados. A rede se adapta a regularidades estatísticas dos dados de entrada, desenvolvendo dessa forma a habilidade de criar representações internas para codificar características das entradas e gerar novas classes de maneira automática (RAUBER, 2005).

A abordagem de aprendizagem por esforço enfatiza a aprendizagem mediante a interação com o ambiente, sem ter o embasamento de uma supervisão ou um modelo completo do ambiente em questão, visa minimizar um índice escalar de desempenho (HAYKIN, 2001).

2.5.1 Aprendizagem Supervisionada

Segundo Fabris (2005), a aprendizagem supervisionada é o método de treinamento mais comumente empregado para a aprendizagem das RNAs, neste paradigma de aprendizagem há um supervisor com conhecimento sobre o ambiente no qual a rede está inserida, ou seja, o aprendizado se dá por meio de intervenção externa. O objetivo é encontrar a ligação mais adequada entre pares de estímulos de entrada e saída conhecidos. Para chegar a este ponto com esse método de aprendizagem, a rede ajusta os seus parâmetros (os pesos sinápticos e as tendenciosidades). De acordo com Lorenzi (2009), “para cada padrão de entrada



submetido, será comparada a resposta desejada e a calculada pela RNA, sendo ajustados os pesos das conexões, de forma a minimizar o erro”.

O algoritmo de retropropagação foi proposto por Rumelhart, Hinton e Willian (1986) para treinamento supervisionado. Essa aprendizagem consiste em dois passos, por meio das diferentes camadas da rede: a propagação e a retropropagação. Na propagação um padrão é apresentado à camada de entrada da rede, o resultado é propagado por intermédio desta, camada por camada, até que a resposta seja comparada à saída produzida pela última camada, a camada de saída, os pesos sinápticos são mantidos fixos neste passo. Na retropropagação, a saída obtida é comparada à saída desejada para o padrão apresentado. Caso a saída obtida for diferente da desejada, o erro é calculado sendo propagado a partir da camada de saída até a camada de entrada. Os pesos das conexões das unidades de camadas internas vão sendo modificadas conforme o erro é retropropagado (SILVA, 2017).

De acordo com De Moraes (2016), a Eq. (1) é a representação matemática para o cálculo do erro de cada neurônio da rede no instante de tempo discreto n , onde e_j representa o sinal de erro, d_j indica o valor desejado na saída do neurônio e y_j o valor encontrado.

$$e_j(n) = d_j(n) - y_j(n) \quad (1)$$

O princípio de aprendizagem deste algoritmo é realizar uma estimativa do erro das camadas intermediárias por meio de uma estimativa do efeito que estas causam no erro da camada de saída (HAYKIN, 2001). Alves (2019) descreve na Eq. (2) a forma como o erro é utilizado para ajustar os pesos das conexões de cada camada:

$$\Delta w_{ij} = \eta e_j x_i \quad (2)$$

Na Eq. (2), η representa a taxa de aprendizagem, e_j é o erro de saída do neurônio j e x_i é a entrada deste neurônio i . O produto destes valores, Δw_{ij} , é definido como a variação dos pesos entre os neurônios j e i . Alves (2019) afirma que “o aprendizado da rede acontece a cada interação onde os valores são acrescidos

desta variação. A minimização do erro indica que a rede otimizou os valores dos pesos e a rede alcançou o aprendizado”.

3 PROCEDIMENTOS EXPERIMENTAIS

Com o objetivo de dividir as etapas de desenvolvimento desse trabalho, segue abaixo os tópicos das principais atividades executadas durante a realização deste projeto. Em cada etapa será discutido os principais assuntos relacionados à mesma. A Fig. 3 representa o fluxograma interativo do procedimento experimental adotado.

Figura 3: Fluxograma interativo do procedimento experimental



Fonte: A autora (2020)

3.1 BASE DE DADOS

A primeira etapa deste projeto baseou-se na escolha da base de dados para ser utilizada pela RNA. Para a seleção destes dados fez-se o uso da plataforma *Kaggle*, onde foi possível localizar três bases de dados com as palavras-chaves: “Métricas de Ciclismo ao ar livre”, “Métricas de Ciclismo” e “Dados Strava”. A pesquisa



por estas palavras-chaves foi realizada na língua inglesa. O *Kaggle* é uma comunidade *on-line* de cientistas de dados pertencente ao Google. Por meio desta plataforma é possível localizar dados necessários para diversos tipos de aplicações.

Para o uso de RNAs a quantidade de dados é de grande importância. Por meio da base de dados formada pelos três conjuntos de dados citados, foi possível obter uma quantidade de amostras satisfatória e dados diversificados, possibilitando o treinamento e os testes da RNA. Os dados utilizados possuem informações do ano de 2014 até o ano de 2019. A base de dados foi filtrada para conter apenas informações de atividades de ciclismo ao ar livre, as métricas utilizadas foram a distância percorrida, a elevação máxima no percurso, o ganho total de elevação ao final da atividade, a velocidade média do atleta e tempo que a atividade física durou.

O conjunto de dados continha 11100 amostras. Todos os valores apresentados nessas métricas são quantitativos, logo não foi necessário realizar a conversão de variáveis qualitativas por um tratamento de dados. Fez-se apenas a conversão dos valores de metros para quilômetros e de segundos para horas, sendo estas convertidas para números decimais. A finalidade desta conversão foi manter o padrão dos dados gerados pelo aplicativo *Strava*, que foram utilizados posteriormente para a validação.

3.2 SOFTWARES E BIBLIOTECAS

Para elaboração do programa, foi utilizado o software livre *Jupyter*, sendo este um ambiente de desenvolvimento integrado utilizado para a linguagem *Python*. Para o desenvolvimento do programa em questão fez-se uso de algumas bibliotecas. Uma biblioteca é uma coleção de subprogramas utilizados no desenvolvimento de software, dentre as principais podemos citar:

- **Pandas:** é uma ferramenta de manipulação de dados de alto nível. Sendo construído no pacote Numpy e sua estrutura de dados chave é chamada de *DataFrame*. Este permite o armazenamento e manipulação de dados tabulares em linhas de observações e colunas de variáveis.
- **NumPy:** é um pacote para linguagem Python capaz de suportar *arrays* e matrizes multidimensionais.



- Sklearn: é um conjunto de ferramentas simples e eficientes para análise preditiva de dados. Construído em NumPy, SciPy e matplotlib.
- Matplotlib: de acordo com a documentação geral da biblioteca, a mesma fornece uma forma de plotagem semelhante ao MATLAB. Destina-se principalmente a plotagens interativas e casos simples de geração de plotagem programática.
- TensorFlow: é uma biblioteca de aplicações de aprendizado de máquina e pesquisa de redes neurais profundas desenvolvida pelo Google, que dá a possibilidade de serem utilizados diversos algoritmos de aprendizado e otimizações para o treinamento.
- Keras: é uma API (Interface de programação de aplicativo) de aprendizado profundo escrita em Python, executada na plataforma de aprendizado de máquina TensorFlow. O uso desta API facilita a programação de RNAs, automatizando diversas rotinas e suportando *frameworks* de baixo nível, como o TensorFlow.

3.3 PRÉ-PROCESSAMENTO DOS DADOS

A criação do *DataFrame* deu-se pelo uso da biblioteca Pandas, mediante a importação do arquivo do tipo valores separados por vírgula (CSV). Em seguida realizou-se a limpeza dos dados para a remoção de valores nulos presentes no conjunto de dados. A remoção dos *outliers*, que são definidos como valores atípicos, ou seja, que possuem grande afastamentos das demais séries, foi realizada por meio da avaliação do *zscore*. A biblioteca utilizada neste caso foi a *scipy.stats.zscore*, capaz de calcular a pontuação z de cada valor da amostra, em relação à média e ao desvio padrão da amostra. A média, os desvios e a amplitude são especialmente sensíveis aos *outliers*, por este fato a remoção deles torna-se importante para o problema em questão. A normalização dos dados não foi realizada, foram realizados testes com a normalização, porém os resultados entrados foram mais satisfatórios sem o seu uso.

Após a aplicação de todas as etapas de pré-processamento, obteve-se um conjunto final formado por 11011 linhas de amostras, trazendo informações sobre distância percorrida, elevação máxima do percurso, ganho total de elevação, tempo de atividade e velocidade média. Percebe-se que em 10% dos dados do conjunto inicial foram removidos por apresentarem grande distanciamento das séries. A Tab.1 mostra informações de cada parâmetro deste conjunto final, mostrando a quantidade de dados, a média dos valores, o desvio padrão médio, os valores mínimos e máximos



do conjunto.

Tabela 1: Informações sobre o conjunto de dados

	Distância	Elevação Máxima	Ganho Total de Elevação	Tempo de Atividade	Velocidade Média
Quantidade	11011	11011	11011	11011	11011
Média	25,258114	0,265217	0,270261	1,212447	20,594603
Desvio padrão	27,951061	0,200631	0,420713	1,33264	4,503447
Mín.	0,000200	0,000000	0,000000	0,000100	0,018000
Máx.	250,534000	2,583800	3,896800	11,846400	54,205200

Fonte: A autora (2020)

3.4 ANÁLISE ESTÁTISTICA

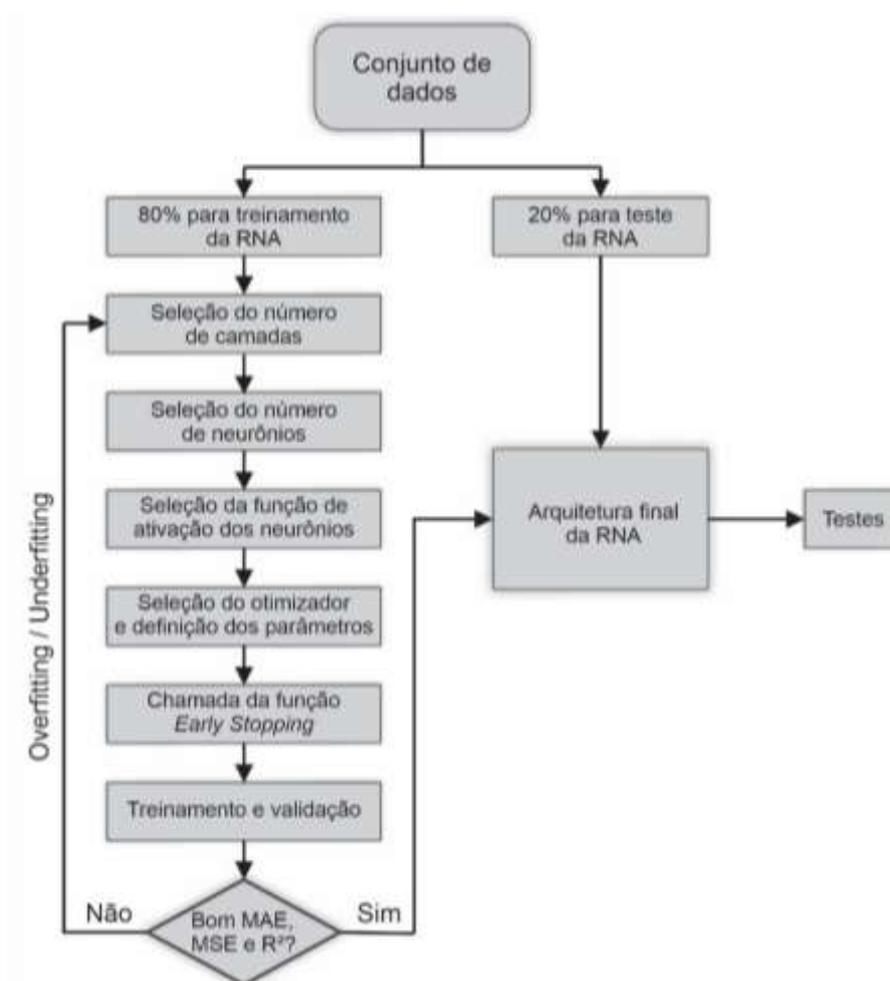
A fim de dar continuidade ao desenvolvimento, foi feita uma análise da correlação entre as variáveis, a correlação procura entender como uma variável se comporta em um cenário onde a outra apresenta variação. Este método busca identificar a existência de alguma correlação entre a variabilidade de ambas. O *DataFrame* de saída pode ser interpretado para qualquer célula, a correlação da variável de linha com a variável de coluna é o valor da célula. A correlação de uma variável consigo mesma tem o valor de 1, logo todos os valores diagonais apresentaram o valor de 1.

Nesta etapa do procedimento realizou-se uma análise da correlação existente entre cada variável por meio da correlação de Person. Este método busca avaliar a relação linear entre duas variáveis contínuas. Podemos afirmar que uma relação linear positiva existe quando o coeficiente de correlação se aproxima de 1, nota-se um aumento no valor de uma variável quando a outra também aumenta. Já a correlação negativa ou inversa, é definida quando o coeficiente se aproxima de -1, sendo possível dizer que as variáveis são correlacionadas, mas nesse caso quando o valor de uma variável aumenta o da outra diminui.

3.5 REDE NEURAL ARTIFICIAL

Os passos para construção da RNA e definição de sua arquitetura final foi esquematizado no fluxograma da Fig. 4.

Figura 4: Fluxograma da escolha da arquitetura da Rede Neural



Fonte: A autora (2020)

Os dados das amostras foram aleatoriamente separados em conjunto de treinamento e conjunto de teste. As proporções geralmente usadas são 80/20 ou 70/30, neste caso definiu-se 80% do total para o conjunto de treinamento e 20% do total para o conjunto de teste. Para a condução da pesquisa, foi realizada uma análise exploratória no conjunto de dados de treinamento, variando-se os parâmetros das RNAs como o número de camadas, a quantidade de neurônios em cada camada, tamanho do lote (batch) e o número de épocas.

Realizou-se a chamada da função *Early Stopping* para a parada do treinamento da rede para evitar problemas relacionados a *overfitting*, quando o modelo acaba por decorar o que deve ser feito e *underfitting*, quando o modelo não consegue encontrar relações entre as variáveis



3.6 ANÁLISE DE DESEMPENHO DA REDE E VALIDAÇÃO

A etapa de análise de desempenho do modelo foi realizada por meio de uma comparação entre os valores previstos e os valores medidos da parcela de dados separados para teste, comparados com métricas estatísticas de Erro Médio Quadrático (*MSE*), Erro Absoluto Médio (*MAE*), Raiz do Erro Quadrático Médio (*RMSE*) e do Coeficiente de Determinação (R^2).

Alguns dados de atividades de ciclistas da região foram coletados por meio do uso do aplicativo *Strava*. Sendo este um aplicativo desenvolvido para o registro de atividades físicas que incorpora recursos de uma rede social. Por meio da coleta destes dados uma nova base de amostras foi gerada. Nesta etapa, verificou-se os resultados obtidos com o uso das RNAs em relação aos resultados reais obtidos de atividades já realizadas, para a avaliação do modelo na previsão da saída definida, realizando a validação do projeto desenvolvido com uma nova base de dados reais.

4 RESULTADOS E DISCUSSÕES

4.1 ANÁLISE ESTATÍSTICA DOS DADOS

O cálculo das correlações entre todas as variáveis selecionadas gerou os coeficientes apresentados na Tab. 2. Nos dados presentes esperava-se obter uma maior correlação da velocidade média em relação a distância e ao tempo de atividade. A baixa correlação encontrada pode ser causada por alterações existentes no conjunto de dados existente.

Tabela 2: Correlação entre as variáveis

	Distância	Elevação Máxima	Ganho Total de Elevação	Tempo de Atividade	Velocidade Média
Distância	1,000000	0,512226	0,770288	0,952377	0,232983
Elevação Máxima	0,512226	1,000000	0,783975	0,556238	-0,062382
Ganho Total de Elevação	0,770288	0,783975	1,000000	0,807576	-0,009518
Tempo de Atividade	0,952377	0,556238	0,807576	1,000000	0,023009
Velocidade Média	0,232983	-0,062380	-0,009518	0,023009	1,000000

Fonte: A autora (2020)



No estudo em questão percebeu-se pela observação dos valores apresentados na Tab.01 a existência de uma correlação positiva de todas as variáveis em relação a distância percorrida e o tempo de atividade. Notou-se também que a variável que apresentou uma correlação mais baixa foi a velocidade média. Além disso foi possível observar uma correlação negativa da velocidade média em relação a elevação máxima e o ganho total de elevação, indicando que o aumento desses fatores pode gerar uma diminuição no valor da velocidade média.

Após a análise dos fatores de correlação apresentados, definiu-se as variáveis de entrada e de saída da RNA. As entradas utilizadas foram a distância percorrida, elevação máxima, ganho total de elevação e o tempo de atividade. Optou-se por apenas uma variável de saída, selecionada por uma questão competitiva do esporte, sendo está a velocidade média.

O acompanhamento da velocidade média no ciclismo é um modo importante de medir o desenvolvimento, habilidade e preparo do atleta. Apesar da velocidade média apresentar uma correlação positiva com o tempo e distância, percebeu-se a influência de outros fatores em sua determinação. O nível de dificuldade de um percurso é definido pelo tipo de terreno, quantidade de subidas e descidas e inclinação das subidas, fatores que podem ser analisados por meio das variáveis de entrada de elevação máxima do terreno e do ganho total de elevação.

4.2 REDE NEURAL ARTIFICIAL

A arquitetura implementada classifica-se como sendo do tipo *Perceptron* de Múltiplas Camadas, com aprendizagem supervisionada por meio do algoritmo de retropropagação e otimizador Adam, sendo este um algoritmo de otimização com método simples de implementar e é computacionalmente eficiente, tem poucos requisitos de memória, é invariante ao reescalonamento diagonal dos gradientes, sendo bem adequado para problemas que são grandes em termos de dados e / ou parâmetros.

Fez-se uso do método *Sequential* para a criação do modelo, de acordo com a documentação oficial do Keras, o método *Sequential* cria uma estrutura linear e em pilha das camadas da rede neural. Isso definirá a forma da rede neural. No restante da rede neural, são criadas duas camadas ocultas e mais uma de saída, todas por intermédio do método *keras.layers.Dense*. Uma camada densa, é simplesmente uma



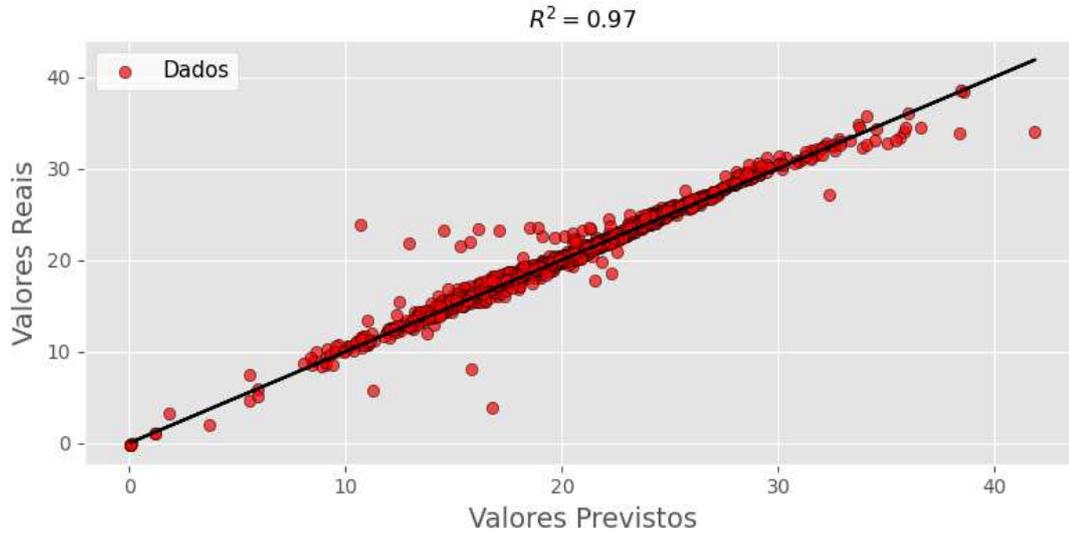
camada densamente conectada com as outras, significando que cada neurônio da camada terá uma ligação com os neurônios da camada anterior e da camada subsequente, além de implementar a operação de ativação e também o parâmetro bias.

A função de ativação utilizada nas duas primeiras camadas foi a *ReLU*, sendo essa é uma função não linear, o que torna fácil a propagação dos erros para trás e a ativação de várias camadas de neurônios por meio dessa função. A função de perda *MSE* (Erro Quadrático Médio na tradução) foi aplicada para indicar o quanto as previsões desviaram dos valores de destino. No treinamento da rede os pesos são atualizados com base nesse valor. A RNA que apresentou os menores valores de MAE, MSE e RMSE, com o maior valor do R^2 , foi construída com três camadas com respectivamente 16-32-1 neurônios. Foram usadas 250 épocas, sendo este o parâmetro para determinar o número de vezes que um algoritmo de aprendizado vê o conjunto de dados completo para a análise. O tamanho do lote (*batch*) foi definido em 64.

4.3 ANÁLISE DE DESEMPENHO DA REDE E VALIDAÇÃO

A avaliação da performance da RNA foi feita por meio da comparação entre os valores medidos do conjunto de teste e os valores previstos pela própria RNA. A Fig. 7 apresenta o diagrama de dispersão da previsão da velocidade média pela RNA em relação aos valores reais obtidos na base de dados e separados para testes. Observa-se o valor de R^2 de cerca de 97%.

Figura 7: Gráfico de dispersão entre os valores de velocidade média previstos e reais



Fonte: A autora (2020)

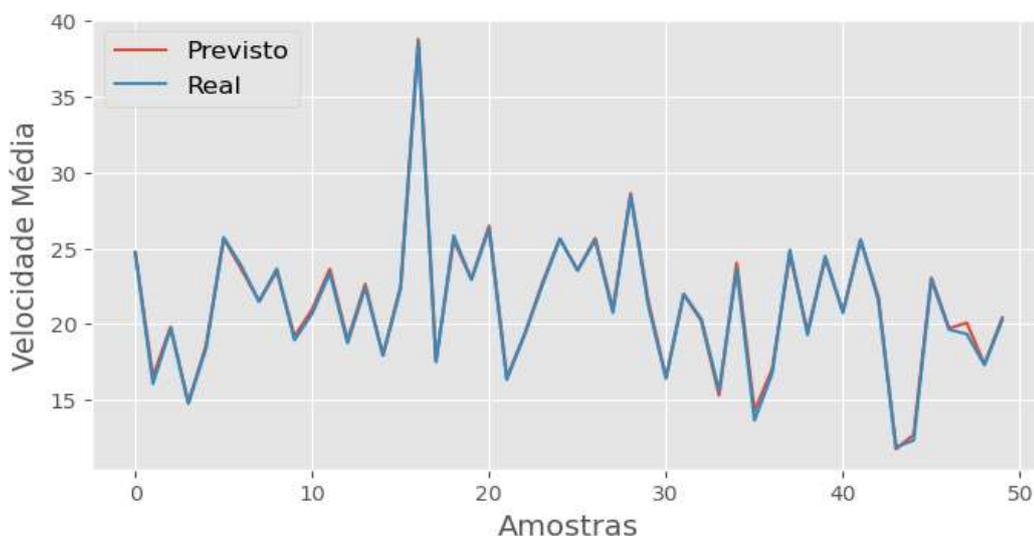
A obtenção do valor do coeficiente de determinação pode ser observada na Eq.(3):

$$R^2 = 1 - \frac{SQR}{SQT} \quad (3)$$

O coeficiente de determinação é obtido pela comparação da regressão da soma dos quadrados à soma total dos quadrados. O valor de SQR oferece a variação de Y considerando as variáveis X utilizadas no modelo. O valor de SQT É o somatório das diferenças entre o valor alvo real e seu valor médio elevado ao quadrado. O R^2 é uma forma utilizada para a avaliação de modelos de RNAs para problemas de regressão, sendo uma maneira de determinar o quanto o modelo em questão está prevendo corretamente.

O gráfico de linhas apresentado pela Fig. 8 mostra mais um comparativo entre os valores previstos e os valores reais, foram usadas 50 amostras para a criação do gráficos, para a melhor visualização do leitor. É possível notar que a RNA é capaz de reproduzir em sua resposta as tendências de variação da saída ao longo do tempo.

Figura 8: Gráfico de linhas entre os valores previstos e reais

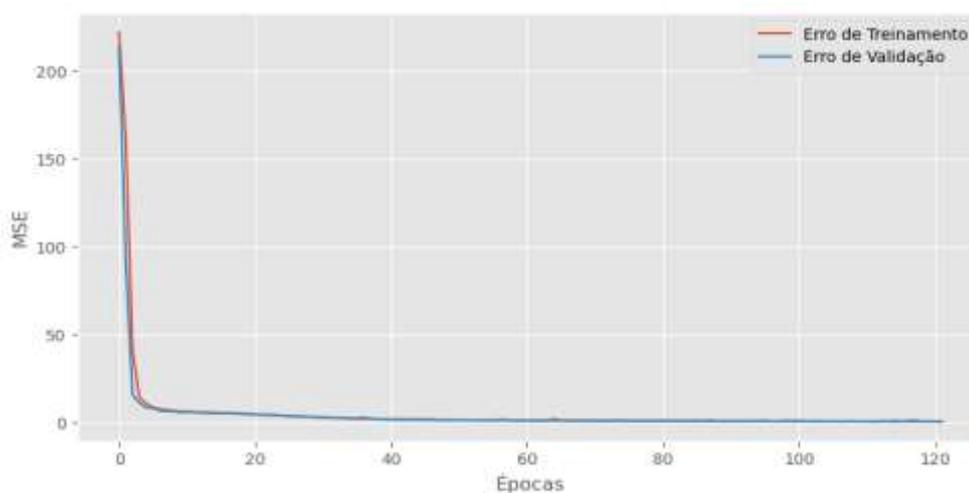


Fonte: A autora (2020)

Para os valores de MSE, MAE e RMSE encontrou-se respectivamente 0,654, 0,391 e 0,809. A Fig. 9 apresenta um gráfico com o MSE e a Fig. 10 traz os valores do MAE, determinados no treinamento e na validação após a chamada da função *Early Stopping*, indicando que a rede convergiu aproximadamente na época 120, ou seja, o *score* de validação não obteve mais aumento.

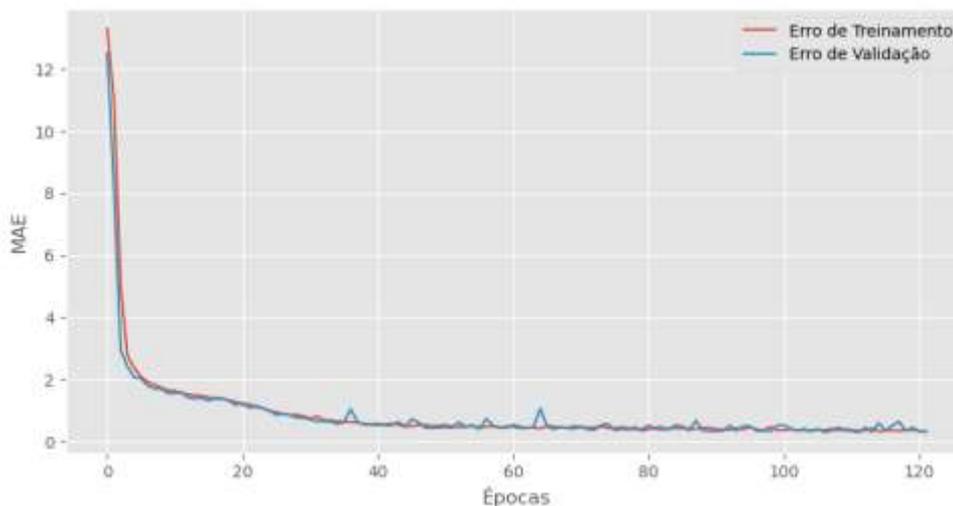
Com a análise deste gráfico é possível perceber que o erro fica estável, sendo que o erro da validação acompanha o erro de treinando, não tornando a ser maior que este.

Figura 9: Gráfico de linhas do Erro Médio Quadrático (MSE)



Fonte: A autora (2020)

Figura 10: Gráfico de linhas do Erro Médio Absoluto (MAE)



Fonte: A autora (2020)

Para finalizar o projeto e validar a performance da rede foram coletados alguns dados de atividades de um grupo de ciclistas da cidade de Criciúma e montada uma pequena base de dados, no exemplo em questão foram utilizadas cinco amostras de atividades do ciclismo já realizados, onde os dados da velocidade média eram de conhecimento. A Tab. 3 apresenta os dados, na entrada da rede nas quatro primeiras colunas, a quinta coluna representa a velocidade registrada e a sexta coluna traz a velocidade média prevista pela rede.

Tabela 3 – Utilização de dados reais do Strava para previsão pela RNA

Distância (km)	Elevação Máxima (km)	Ganho Total de Elevação (km)	Tempo de Atividade (h)	Velocidade Média Registrada (km/h)	Velocidade Média Prevista (km/h)
10,62	0,058	0,064	0,54	19,8	19,93
32,43	0,091	0,266	2,14	15,2	15,19
21,1	0,132	0,3	1,53	13,8	13,42
63,29	0,121	0,732	4,51	14	14,58
40,62	0,268	0,968	2,34	17,4	17,45
51,36	0,069	0,468	2,65	19,4	19,79

Fonte: A autora (2020)

Verificou-se que os resultados previstos pela RNA foram muito próximo dos valores reais, determinando desta forma que existe um bom funcionamento da Rede Neural, resultando em uma ferramenta computacional de auxílio a prática do ciclismo,



a velocidade média foi prevista pela RNA considerando as variáveis que afetam a variável de saída por meio de uma correlação negativa.

5 CONCLUSÕES

O presente trabalho buscou desenvolver e verificar a eficácia de uma RNA construída em Python, por meio de suas bibliotecas, para a previsão de métricas de desempenho de um grupo de ciclistas da região de Criciúma. Desenvolveu-se desta forma uma ferramenta computacional de auxílio para análise e previsão de dados relacionados a este esporte. O conjunto de dados utilizado para o treinamento e validação da rede foi proveniente de amostras de dados localizadas na plataforma *Kaggle*, por meio da união de três conjuntos obteve-se a base de dados utilizada neste trabalho.

A Rede Neural gerada foi uma *Perceptron* de Múltiplas Camadas, a com a arquitetura proposta, obtive erro quadrático médio baixo e um coeficiente de determinação de 97%, indicando a eficácia da rede. Demonstrando por meio deste estudo que, a partir de uma seleção e tratamentos de dados relacionados ao esporte, que apresentam influência nos dados selecionados como saída da RNA, e por meio do ajuste de seus parâmetros, é possível fazer o uso desta tecnologia como forma de previsão de métricas do esporte.

A previsão dos dados de velocidade média por intermédio dos dados do trajeto busca auxiliar os ciclistas no treinamento para competições, onde os dados do percurso possam ser inseridos e considerando a dificuldade do terreno, a velocidade média ideal seja prevista. Além de colaborar com a evolução de ciclistas amadores com pouco tempo para a prática do esporte.

Para trabalhos futuros, sugere-se o desenvolvimento de uma aplicação visual. Por meio da qual o usuário possa inserir os dados de trajeto em seu *smartphone* e o modelo apresente os dados da velocidade média para a realização daquela atividade. Tornando desta forma a ferramenta computacional proposta viável para o uso prático. Acredita-se que a inclusão de dados sobre as características das bicicletas utilizadas e as condições meteorológicas seja relevante, pois podem ter influência na atividade, infelizmente não foi possível adicioná-los a este trabalho.



REFERÊNCIAS

- ALMEIDA, Eduardo Barbosa et al. Gasto calórico nas atividades de trabalho e cotidianas, dos carteiros que utilizam bicicleta. **Rev. Bras. Cine. Des. Hum. ISSN**, v. 1415, p. 8426, 2004.
- ALVES, Paulo Venâncio Rodrigues. **Fusão sensorial de um sonar e de um laser scanner utilizando redes neurais artificiais**. 2019.
- Associação Brasileira dos Fabricantes de Motocicletas, Ciclomotores, Motonetas, Bicicletas e Similares. **Indústria de Bicicletas supera, até Novembro, projeção de produção total de 2019**. Disponível em: <<http://www.abraciclo.com.br/2019/1443-industria-de-bicicletas-supera-ate-novembro-projecao-de-producao-total-de-2019>>. Acesso em: 31 de março de 2020.
- BERSON, Alex; SMITH, Stephen; THEARLING, Kurt. **Building data mining applications for CRM**. McGraw-Hill Professional, 1999.
- BITTENCOURT, Gabriela. **Bicicletas para todos os gostos e modalidades de ciclismo: saiba qual é o modelo ideal para você**. Disponível em: <<https://globoesporte.globo.com/eu-atleta/noticia/bicicletas-para-todos-os-gostos-e-modalidades-de-ciclismo-saiba-qual-e-o-modelo-ideal-para-voce.ghtml>>. Acesso em: 08 de abril de 2020.
- COELHO FILHO, Osmar; SACCARO JUNIOR, Nilo Luiz. **Cidades Cicláveis: avanços e desafios das políticas cicloviárias no Brasil**. IPEA, 2017.
- DE MORAIS, Leonardo Cesar et al. **Previsão do preço da energia elétrica no mercado ibérico utilizando uma rede neural artificial *multi-layer perceptron* com *backpropagation***. *Etic-encontro de iniciação científica-issn 21-76-8498*, v. 12, n. 12, 2016.
- GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep learning**. MIT press, 2016.
- GRUS, Joel. **Data Science do zero: Primeiras regras com o Python**. Alta Books Editora, 2018.
- HAYKIN, Simon. **Redes neurais: princípios e prática**. Bookman Editora, 2007.
- Kaggle. **Outdoor Cycling Metrics**. Disponível em: <<https://www.kaggle.com/dorinaferencsik/outdoor-cycling-metrics>>. Acesso em: 20 de agosto de 2020.
- KOLISNYK, T. Ye et al. **Application of an Artificial Neural Network for Design of Sustained-Release Matrix Tablets Containing Vaccinium Myrtillus Leaf Powder Extract**. 2018.
- KÜNAS, Cristiano Alex; HECK, Leandro Perius; PADOIN, Edson Luiz. **Implementação de Rede Neural Artificial em Plataforma GPU Aplicada na**



Análise de Sentimentos em Textos. In: Anais da XX Escola Regional de Alto Desempenho da Região Sul. SBC, 2020. p. 41-44.

LIU, Y. L. A. et al. **Uso de rede neural percéptron multi-camadas na classificação de patologias cardíacas.** Trends in Applied and Computational Mathematics, v. 9, n. 2, p. 255-264, 2008.

LORENZI, Alexandre. **Aplicação de redes neurais artificiais para estimativa da resistência à compressão do concreto a partir da velocidade de propagação do pulso ultra-sônico.** 2009.

MCCULLOCH, Warren S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, n. 4, p. 115-133, 1943.

PINHEIRO, Carlos Andre Reis. **Inteligência Analítica: Mineração de dados e descoberta de conhecimento.** Editora de Ciência Moderna Ltda., 2008.

PNAD: 2015: manuais, guias, etc. **Pesquisa nacional por amostra de domicílios: 2015: manual de entrevista suplemento: prática de esporte e atividade física: algumas características das relações de trabalho primeira infância.** IBGE, Coordenação de Trabalho e Rendimento, 2015.

RAUBER, Thomas Walter. **Redes neurais artificiais.** Universidade Federal do Espírito Santo, 2005.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**, 3ª edição. Upper Saddle River: Prentice Hall, 2010.

SCHNEIDER, Pedro Henrique. **Análise preditiva de Churn com ênfase em técnicas de Machine Learning: uma revisão.** 2016. Tese de Doutorado.

SEGADILHA, Ana Beatriz Pereira et al. **Identificação dos fatores que influenciam na escolha da rota pelos ciclistas: estudo de caso da cidade de São Carlos.** 2014.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao datamining: mineração de dados.** Ciência Moderna, 2009.

TEIXEIRA, Felipe Guimarães. **Classificação do desempenho do atleta de judô utilizando regressão logística e rede neural.** 2019.

VARGAS, Ana Caroline Gomes; PAES, Aline; VASCONCELOS, Cristina Nader. **Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres.** In: Proceedings of the xxix conference on graphics, patterns and images. 2016.

VON ZUBEN, Fernando José et al. **Modelos paramétricos e não-paramétricos de redes neurais artificiais e aplicações.** 1996.