



UNIVERSIDADE FEDERAL DO MARANHÃO

Coordenação de Ciência da Computação

Cicero-Joe Rafael Lima de Oliveira

**Modelos de Regressão Aplicados ao Aprendizado  
Motor**

São Luís - MA

2023

Cicero-Joe Rafael Lima de Oliveira

## **Modelos de Regressão Aplicados ao Aprendizado Motor**

Monografia apresentada como requisito parcial para obtenção do título de bacharel em Ciência da Computação, à Coordenação do Curso de Ciência da Computação, da Universidade Federal do Maranhão.

Coordenação de Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Paulo Rogério de Almeida Ribeiro

São Luís - MA

2023

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Oliveira, Cicero-Joe Rafael Lima de.

Modelos de Regressão aplicados ao Aprendizado Motor /  
Cicero-Joe Rafael Lima de Oliveira. - 2023.

44 f.

Orientador(a): Paulo Rogério de Almeida Ribeiro.

Monografia (Graduação) - Curso de Ciência da  
Computação, Universidade Federal do Maranhão, São Luís,  
2023.

1. Análise de Dados. 2. Aprendizado Motor. 3.  
Métricas de Avaliação. 4. Modelos de Regressão. 5.  
Previsão de Erro. I. Ribeiro, Paulo Rogério de Almeida.  
II. Título.

Cicero-Joe Rafael Lima de Oliveira

## **Modelos de Regressão Aplicados ao Aprendizado Motor**

Monografia apresentada como requisito parcial para obtenção do título de bacharel em Ciência da Computação, à Coordenação do Curso de Ciência da Computação, da Universidade Federal do Maranhão.

Trabalho de Graduação. São Luís - MA, 26 de Julho de 2023

---

**Prof. Dr. Paulo Rogério de Almeida  
Ribeiro**  
Orientador  
Universidade Federal do Maranhão

---

**Prof. Dr. Geraldo Braz Junior**  
Examinador  
Universidade Federal do Maranhão

---

**Prof. Dr. Ivo José da Cunha Serra**  
Examinador  
Universidade Federal do Maranhão

São Luís - MA  
2023

*À memória de meus avôs Cicero Oliveira e Sebastião Azevêdo*

# Agradecimentos

Agradeço a Deus, sem o qual eu não possuiria saúde, sabedoria e paciência para realizar qualquer esforço.

Agradeço aos meus pais, Cicero e Joelma, por sempre estarem dispostos a se doarem por meus sonhos e objetivo, por estarem continuamente me nutrindo de maneira física, intelectual e emocional. Acima de tudo sendo exemplo de perseverança, melhoria contínua e integridade . Agradeço a minha irmã Raabe, pelo constante cuidado e amizade, mostrando um apoio incondicional na conquista de meus objetivos. Também agradeço ao meu cunhado e amigo, Frank, por ser compreensivo e solícito.

Agradeço ao meus amigos do *Grupo*, que tive prazer de conhecer durante o bacharelado e que sempre que necessário deram suporte, ainda aos meus queridos que tive o privilegio de ter a amizade fora do ambiente acadêmico.

Agradeço ao *LACMOR*, personificado pelo professor Alexandre Muniz que contribuiu para minha melhoria como aluno e profissional. Adicionalmente estou grato ao meu orientador Prof. Paulo Rogério, por ser paciente comigo e estar dedicado com o projeto. Também, aos professores de ciência da computação que estão sempre empenhados em repassar o conhecimento que possuem.

No mais, agradeço a todos que me ajudaram e aconselharam durante minha vida acadêmica. Pois cada caractere presente neste projeto representa, também, uma parte de vocês. Muito obrigado!

*"Não importa o quão fraco ou indigno você se sinta, mantenha seu coração aquecido, cerre os dentes e siga em frente."*

(Kimetsu no Yaiba, Kyojuro Rengoku)

# Resumo

Este trabalho tem como objetivo utilizar algoritmos de regressão para prever o erro de um bloco em uma tarefa de aprendizado motor. Existe uma lacuna na literatura em relação aos estudos de regressão aplicados ao aprendizado motor. O uso de modelos preditivos fornece informações valiosas sobre atividades motoras, estimando o potencial de melhoria com base em dados iniciais e avaliando o desempenho, facilitando assim o processo de aprendizado de habilidades motoras. O estudo explora conceitos de regressão aplicados ao aprendizado motor, incluindo Regressão por Perceptron de Múltiplas Camadas (Regressão MLP), Máquina de Vetores de Suporte (SVR) e Regressão Linear (LR). A metodologia descreve o experimento de aprendizado motor, aquisição de dados e análise de resultados. Para avaliar a precisão dos modelos preditivos utilizou-se Erro Absoluto Médio (MAE), Erro Médio Quadrático (MSE), Coeficiente de Determinação ( $R^2$ ) e Erro Percentual foram empregadas como métricas. Os resultados demonstram valores satisfatórios para as métricas usadas (MAE, MSE,  $R^2$ , Diferença Percentual Máxima e Diferença Percentual Mínima) para todos os modelos (SVR, LR e Regressão MLP). Em conclusão, os modelos de regressão se mostraram precisos para prever o desempenho do aprendizado motor, contribuindo para uma melhor compreensão do processo de aprendizagem e auxiliando no desenvolvimento de estratégias de treinamento personalizadas. Os resultados sugerem que os modelos de regressão podem ser ferramentas poderosas para estimar e otimizar o desempenho de aprendizagem em aquisição de habilidades motoras.

**Palavras-chave:** Aprendizado Motor, Modelos de Regressão, Previsão de Erro, Algoritmos Preditivos, Métricas de Avaliação, Análise de Dados.



# Abstract

This study aims to use regression algorithms to predict the error of a block on a motor learning task. There is a gap in the literature regarding regression studies applied to motor learning. The use of predictive models provides valuable information about motor activities, estimating the potential for improvement based on initial data and evaluating performance, thus facilitating the process of learning motor skills. The study explores regression concepts applied to motor learning, including Multilayer Perceptron Regression (MLP Regression), Support Vector Machine (SVR), and Linear Regression (LR). The methodology involves understanding the motor learning experiment, data acquisition and analysis of the results. Performance evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination ( $R^2$ ), and Percentage Error were employed to assess the accuracy of the predictive models. The results show that all models (SVR, LR, and MLP Regression) have a good performance for all metrics (MAE, MSE,  $R^2$ , Maximum Percentage Difference, and Minimum Percentage Difference). In conclusion, regression models proved accurate in predicting motor learning performance, contributing to a better understanding of the learning process and assisting in the development of personalized training strategies. The results suggest that regression models can be powerful tools for estimating and optimizing learning performance in motor skill acquisition.

**Keywords:** Motor Learning, Regression Models, Error Prediction, Predictive Algorithms, Evaluation Metrics, Data Analysis.

# Lista de ilustrações

Figura 1 – Desenho parcial do experimento . . . . .	22
Figura 2 – Conjunto de figuras do erro por bloco para cada indivíduo no experimento.	28
Figura 3 – Conjunto de figuras do erro por bloco para cada indivíduo no SVRLinear.	31
Figura 4 – Conjunto de figuras do erro por bloco para cada indivíduo no modelo de Regressão Linear. . . . .	34
Figura 5 – Conjunto de figuras do erro por bloco para cada indivíduo no modelo MLP de Regressão. . . . .	37
Figura 6 – Conjunto de figuras traçadas em Regressão Linear para o erros preditos	40

# Lista de tabelas

Tabela 1 – Exemplo da organização atual do <i>dataset</i> . . . . .	23
Tabela 2 – Resultado de Métricas SVRLinear . . . . .	29
Tabela 3 – Conjunto de dados preditos - SVRLinear . . . . .	30
Tabela 4 – Resultado de Métricas Regressão Linear . . . . .	33
Tabela 5 – Conjunto de dados preditos - Regressão Linear . . . . .	33
Tabela 6 – Resultado de Métricas MLP de Regressão . . . . .	36
Tabela 7 – Conjunto de dados preditos - MLP de Regressão . . . . .	36
Tabela 8 – Resultados das Métricas Gerais . . . . .	39

# Lista de abreviaturas e siglas

SVR	<i>Support Vector Machines Regression</i>
LR	<i>Linear Regression</i>
MLP	<i>Multilayer Perceptron</i>
$R^2$	Coefficiente de Determinação
RMSE	Erro quadrático médio
MAE	Erro absoluto médio
tDCS	Estimulação Transcraniana por Corrente Direta
tRNS	Estimulação Transcraniana por Ruído Aleatório

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>13</b>
<b>1.1</b>	<b>Justificativa</b>	<b>14</b>
<b>1.2</b>	<b>Objetivos</b>	<b>14</b>
1.2.1	Objetivos Específicos	14
<b>1.3</b>	<b>Organização do Trabalho</b>	<b>15</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
<b>2.1</b>	<b>Regressão com Máquina de vetores de suporte</b>	<b>16</b>
<b>2.2</b>	<b>Regressão Linear</b>	<b>17</b>
<b>2.3</b>	<b>Regressão por Perceptron de Multicamadas</b>	<b>18</b>
2.3.1	Solucionador LBFGS	18
<b>3</b>	<b>METODOLOGIA</b>	<b>20</b>
<b>3.1</b>	<b>Dataset Utilizado</b>	<b>20</b>
3.1.1	Obtenção dos Dados	21
3.1.2	Remoção de <i>Outliers</i>	22
3.1.3	Projeto Atual	23
<b>3.2</b>	<b>Materiais</b>	<b>23</b>
<b>3.3</b>	<b>Modelos de Regressão</b>	<b>24</b>
3.3.1	Regressão por vetores suporte	24
3.3.2	Regressão Linear	24
3.3.3	Regressão por Perceptron Multicamadas	25
<b>3.4</b>	<b>Métricas de Avaliação</b>	<b>25</b>
3.4.1	Erro quadrático médio (MSE)	25
3.4.2	Erro absoluto médio (MAE)	25
3.4.3	Coefficiente de Determinação ( $R^2$ )	26
3.4.4	Erro Percentual	26
<b>4</b>	<b>RESULTADOS</b>	<b>27</b>
<b>4.1</b>	<b>Resultados do Aprendizado Motor</b>	<b>27</b>
<b>4.2</b>	<b>Resultados dos Regressores</b>	<b>29</b>
4.2.1	Resultados SVRLinear	29
4.2.2	Resultados Regressão Linear	33
4.2.3	Resultados MLP de Regressão	36
4.2.4	Resultados Comparativos	39

<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>42</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>43</b>

# 1 Introdução

O pleno desempenho de atividades motoras influenciam em muitas áreas da vida de um ser humano, desde o ato de escovar os dentes à prática de um trabalho (e.g., jornalismo, construção civil, gastronomia etc). O sistema motor humano é altamente adaptativo e capaz, ajustando-se com base nas experiências e em práticas cíclicas. À medida que uma determinada atividade é realizada, repetidamente, podem haver modificações no sistema motor, por exemplo aprimoramentos no desempenho e na eficiência do movimento.

Segundo [Krakauer et al. \(2019\)](#), há dois aspectos do aprendizado: Obtenção de Habilidades - o processo pelo qual o indivíduo adquire a habilidade de movimento apropriada para a execução de uma tarefa específica com exatidão e precisão; Manutenção de Habilidades - a disposição interna do indivíduo em assegurar os níveis de rendimento das habilidades adquiridas em situações de adversidade.

O processo de aprendizado motor requer regularidade na prática da atividade. A prática constante permite que o sistema motor se beneficie das repetições e consolidação dos padrões motores adequados, levando a um ganho contínuo no desempenho da tarefa, ([KRAKAUER et al., 2019](#)).

No trabalho de [Dayan e Cohen \(2011\)](#) descreve-se aprendizado *online*, como a melhoria obtida durante a atividade para aquisição de habilidade. Os ganhos de habilidade podem ser obtidos, também, entre as sessões sem qualquer exercício prático, sendo isto o aprendizado *offline*.

O estudo de [Prichard et al. \(2014\)](#) investigou os efeitos de estimulações cerebrais elétricas não invasivas - Estimulação Transcraniana por Corrente Direta (tDCS) e Estimulação Transcraniana por Ruído Aleatório (tRNS) - no aprendizado motor (*online* e *offline*). Os participantes saudáveis foram submetidos a diferentes protocolos de estímulo enquanto realizavam uma tarefa motora de decalque de palavras. Além disso, o estudo introduziu uma metodologia para monitoramento do erro durante a tarefa, possibilitando uma análise mais detalhada do desempenho dos participantes.

O projeto de [CARVALHO \(2018\)](#), baseado no experimento proposto por [Prichard et al. \(2014\)](#), consistiu na elaboração e construção de um software automatizado e padronizado para a tarefa de decalque de palavras. Essa ferramenta auxiliou na realização e monitoramento da tarefa de aprendizado motor, permitindo a observação da precisão de execução do decalque.

Posteriormente, [SANTOS \(2019\)](#) utilizou o software desenvolvido por [CARVALHO \(2018\)](#) para coletar dados de treinamento motor de voluntários. O objetivo foi estimar

a quantidade de sessões necessárias para aprender uma tarefa motora. Algoritmos de classificação como *K-Nearest Neighbors*, Árvore de Decisão e Aprendizagem de Máquina Automática (AutoML) com "Auto Weka" entre outros foram aplicados.

Adicionalmente, o estudo realizado por [Silva \(2022\)](#), utilizando os dados de [SANTOS \(2019\)](#), teve como objetivo utilizar uma Rede Neural Recorrente para estimar o momento em que um indivíduo irá atingir um determinado nível de desempenho em uma tarefa de aprendizado motor.

Essas pesquisas contribuem para o avanço do conhecimento na área de aprendizado motor, no entanto, as literaturas vigentes, pouco relatam o uso de algoritmos de regressão para previsão de erro durante o aprendizado motor. Faz-se necessário a utilização de soluções de predição para compreender a situação do desempenho do aprendiz e o que se espera para sessões de aprendizado futuras.

## 1.1 Justificativa

No entanto, ainda há uma lacuna na literatura (vide: [Prichard et al. \(2014\)](#), [SANTOS \(2019\)](#) e [Silva \(2022\)](#)) em relação aos estudos de regressão aplicados ao aprendizado motor. A utilização de modelos preditivos nessa área pode trazer benefícios significativos, pois permite entender e estimar a progressão do desempenho motor com base nos dados iniciais. Isso possibilita uma visão mais clara do potencial de melhoria de um indivíduo ao longo do tempo, bem como uma avaliação do seu desempenho atual. As previsões podem auxiliar na identificação de pontos fortes e fracos no aprendizado motor de cada indivíduo, permitindo o desenvolvimento de estratégias personalizadas de treinamento. Assim sendo, essas informações podem auxiliar no planejamento dos treinamentos, por exemplo mais direcionado a cada paciente, assim como na definição de metas realistas para cada indivíduo.

## 1.2 Objetivos

O objetivo geral deste trabalho consiste em utilizar um algoritmo de regressão para predição de um erro de um bloco comparando o que foi predito com o erro real, utilizando o erro referente a blocos de treinamento realizados.

### 1.2.1 Objetivos Específicos

- Estudar a aplicação de regressão no *datasets* de aprendizado motor.
- Comparar o erro de um bloco futuro (estimado) com o bloco real.
- Avaliar diferentes regressores.



## 1.3 Organização do Trabalho

A estrutura deste trabalho apresenta-se como:

- O Capítulo 2 trata da fundamentação teórica das técnicas utilizadas. Descreve-se os algoritmos de regressão e suas especificações aplicadas no projeto atual.
- O Capítulo 3 descreve a metodologia por etapas, nomeadamente: obtenção e estruturação do *dataset*, materiais utilizado, modelos de regressão e métricas de avaliação aplicadas.
- O Capítulo 4 apresenta os resultados obtidos do experimento motor e dos regressores, discutindo as relações entre eles, bem como suas implicações.
- O Capítulo 5 descreve as conclusões do trabalho, assim como propostas para trabalhos futuros.

## 2 Fundamentação Teórica

Este capítulo apresenta os conceitos explorados para o desenvolvimento do estudo de regressão aplicados ao aprendizado motor.

### 2.1 Regressão com Máquina de vetores de suporte

A *Support Vector Machine Regression (SVR)* é uma técnica de aprendizado de máquina que pode ser aplicada a problemas de regressão. É uma extensão da classificação *SVM*, que foi adaptada para lidar com problemas de previsão de valores contínuos ao invés de classes discretas.

A abordagem do *SVM* para regressão envolve a construção de um hiperplano de separação ótimo no espaço de entrada, maximizando-se a margem entre os dados de treinamento e o hiperplano criado por meio de um conjunto de vetores de suporte.

O objetivo é encontrar uma função de regressão que se aproxime dos dados de treinamento, mas que contenha uma margem de tolerância em torno dos pontos de dados. A previsão de novos valores é feita calculando a distância entre o ponto de dados a ser previsto e os vetores de suporte, e usando essa informação para estimar o valor correspondente, (SMOLA; SCHÖLKOPF, 2004).

Uma das características distintivas do *SVR* é o uso de uma função de *kernel*, que permite mapear os dados de entrada para um espaço de maior dimensionalidade. A utilização do *kernel* linear permite que o *SVR* modele relacionamentos não lineares entre as variáveis independentes e a variável dependente.

No caso específico do *kernel* linear, a função de regressão é linear no espaço de características original. Portanto o hiperplano, criado foi obtido por uma função linear, a qual separa os pontos de características( entradas) com retas, (HUANG; KECCMAN; KOPRIVA, 2006).

A equação geral de um modelo *SVR* com *kernel* linear é representada da seguinte forma:

$$f(x) = \beta_0 + \sum_{i=1}^n \alpha_i \langle x, x_i \rangle \quad (2.1)$$

Onde:

1.  $f(x)$  é o valor predito pela função de regressão
2.  $\beta_0$  é o termo de bias<sup>1</sup>

---

<sup>1</sup> Análogo ao coeficiente angular em uma função afim

3.  $\alpha_i$  são os coeficientes de suporte que determinam a influência de cada vetor de suporte  $x_i$
4.  $\langle x, x_i \rangle$  é o produto escalar entre os vetores  $x$  e  $x_i$

Destaca-se que os vetores  $x$  e  $x_i$  são relativos a duas características (entradas) do modelo.

## 2.2 Regressão Linear

A regressão linear é uma técnica estatística amplamente utilizada para modelar o relacionamento entre uma variável dependente contínua e uma ou mais variáveis independentes. É um dos métodos mais simples e interpretáveis de modelagem preditiva e possui uma ampla gama de aplicações em diversos campos, incluindo ciências sociais, economia, engenharia e ciências da saúde.

A regressão linear busca encontrar a melhor linha reta que representa a relação linear entre a variável independente e a variável dependente. Essa reta é determinada pela minimização dos erros entre os valores observados e os valores preditos pelo modelo.

A equação geral de um modelo de regressão linear simples é representada da seguinte forma:

$$Y = \beta_0 + X\beta_1 + \varepsilon \quad (2.2)$$

Onde têm-se que:

1.  $Y$  é a variável dependente
2.  $\beta_0$  é o intercepto da linha de regressão
3.  $X$  é a variável independente
4.  $\beta_1$  é o coeficiente de regressão que indica a inclinação da linha de regressão
5.  $\varepsilon$  é o termo de erro que representa a variação não explicada pelo modelo

Os coeficientes de regressão estimados fornecem informações sobre a magnitude e a direção do efeito da variável independente na variável dependente. Além disso, a regressão linear permite a realização de inferências estatísticas sobre os coeficientes, como testes de hipóteses e intervalos de confiança.

Uma extensão da regressão linear é a regressão linear múltipla, ou seja envolve mais de uma variável independente. A regressão linear múltipla permite modelar relações mais complexas entre a variável dependente e as variáveis independentes, levando em consideração a influência de múltiplos fatores simultaneamente (JAMES et al., 2023).

Menciona-se também a regressão linear ponderada que atribui pesos diferentes aos pontos de dados com base em suas importâncias.

## 2.3 Regressão por Perceptron de Multicamadas

Redes Neurais *Multilayer Perceptron* (*MLP*) são uma classe popular de modelos de aprendizado de máquina que consistem em uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída. Cada camada é composta por um conjunto de neurônios artificiais (também chamados de nós) que estão interconectados por sinapses. Cada sinapse é ponderada por um valor que representa a importância dessa conexão.

O funcionamento do *MLP* é baseado em uma abordagem conhecida como *feedforward*, em que as informações fluem através da rede em uma única direção, neste caso da camada de entrada para a camada de saída. Cada neurônio nas camadas (ocultas e de saída) calcula uma combinação linear das entradas ponderadas pelos pesos sinápticos e aplica uma função de ativação para produzir o valor de saída (LECUN; BENGIO; HINTON, 2015).

Uma das vantagens do *MLP* é a sua capacidade de aprender e modelar relações não-lineares complexas entre as variáveis de entrada e a variável de saída. Além disso, os *MLP's* podem ser usados para lidar com conjuntos de dados de alta dimensionalidade, são altamente flexíveis e adaptáveis a diferentes tipos de problemas.

Ao longo dos anos, os MLPs têm sido aprimorados com várias técnicas e variações, como regularização, *dropout*, normalização de lotes e a utilização de diferentes funções de ativação. Essas melhorias têm contribuído para o sucesso e a ampla aplicabilidade dos MLPs em uma ampla gama de problemas de aprendizado de máquina.

Durante o treinamento do *MLP*, os pesos sinápticos são ajustados iterativamente usando algoritmos de otimização, por exemplo o *backpropagation*. O objetivo é minimizar uma função de perda que mede a discrepância entre as saídas do modelo e os valores alvo. O processo de treinamento envolve a propagação do erro calculado na camada de saída de volta para as camadas ocultas, ajustando gradualmente os pesos sinápticos para reduzir o erro (LECUN; BENGIO; HINTON, 2015).

### 2.3.1 Solucionador LBFGS

Uma das implementações mais comuns de *MLP* é a utilização do *solver* *lbfgs*, que é um algoritmo de otimização baseado no método de Quasi-Newton. Esse *solver* é especialmente adequado para problemas de regressão, utilizando aproximações simples de matrizes Hessianas para otimizar as sinapses, sendo assim, ao invés de armazenar plenamente essa matriz de aproximações densas, são armazenados apenas poucos vetores

que representam as aproximações de forma implícita, (LIU; NOCEDAL, 1989).

A utilização do solucionado lbfgs em *MLP* oferece várias vantagens, por exemplo: convergência rápida, eficiência computacional e capacidade de lidar com conjuntos de dados de tamanho menor. Além disso, é adequado para problemas com alta dimensionalidade e não requer a definição de uma taxa de aprendizado manualmente.

## 3 Metodologia

Este capítulo descreve a metodologia proposta para o presente estudo, iniciando pela compreensão do experimento de aprendizado motor (software de análise, aquisição de dados e análise de resultados) que produziu a primeira versão do *dataset* utilizado neste projeto.

Os dados mais aderentes ao critérios da atividade de aprendizado foram utilizados conforme (Subseção 3.1.1). Em sequencia alterou-se a composição do conjunto de dados, filtrando conteúdo que não estava consistente com o grupo (Subseção 3.1.2). Em continuidade organizou-se este conjunto, utilizando médias entre valores, bem como, divisão de grupo de teste e treino, retratado no Projeto Atual (Subseção 3.1.3).

Os algoritmos de predição (*SVRLinear*, Regressão Linear e MLP com solucionador *lbfgs*) foram escritos para a previsão de erros relacionados às atividades de aprendizado motor. Escolhidos devido à características particulares como: capacidade computacional (simples/robusto), tolerância a erro e adaptação à pequenos *datasets*.

Os modelos *SVRLinear* (Subseção 3.3.1), Regressão Linear (Subseção 3.3.2) e MLP (Subseção 3.3.3) foram treinados utilizando um conjunto de treinamento e seus parâmetros foram ajustados com base nesses dados.

Visando aferir os regressores, utilizou-se métricas clássicas para algoritmos de regressão, tais como: *Mean Absolute Error* - MAE (Subseção 3.4.2), *Mean Squared Error* - MSE (Subseção 3.4.1), Coeficiente de Determinação -  $R^2$  (Subseção 3.4.3) e Erro Percentual. Elas foram utilizadas para monitorar e avaliar a capacidade dos modelos na adaptação aos dados e de realizar previsões precisas dos erros na tarefa de aprendizado motor.

Em resumo, a metodologia proposta neste trabalho se fundamenta na aquisição e modelagem dos dados, e no uso de modelos de regressão para prever os erros associados aos experimentos reais. A aplicação desses modelos demanda uma etapa de treinamento e avaliação, além da utilização de métricas de desempenho apropriadas. As representações visuais dos resultados são úteis para facilitar a compreensão e interpretação dos erros estimados pelos modelos, assim como, para comparar os erros reais e os preditos.

### 3.1 Dataset Utilizado

O conjunto de dados utilizado neste trabalho foi obtido a partir de trabalhos prévios realizados por SANTOS (2019) e CARVALHO (2018). Esses trabalhos estruturaram a aquisição dos dados, desde plataforma de decalque das palavras até a execução do experimento prático. A compreensão dos erros entre as palavras reais e dos decalques foi

um dos principais focos destes estudos.

No contexto deste projeto, o conjunto de dados passou por modificações em sua estrutura, visando a aplicação de técnicas de regressão. Essas alterações incluíram: o uso dos erros relacionados aos blocos e indivíduos de cada sessão, a separação das sessões em partes distintas. Além disso, foi aplicado o conceito de divisão treino-teste no experimento. Essas modificações foram necessária para adequar o conjunto de dados às necessidades dos algoritmos, com a finalidade de obter resultados significativos.

### 3.1.1 Obtenção dos Dados

Os primeiros dados foram obtidos por meio de um conjunto de experimentos realizados utilizando um *software* desenvolvido por CARVALHO (2018) para avaliar o aprendizado motor. Esse *software* foi projetado para fornecer imagens de palavras em fonte cursiva na língua portuguesa, utilizando a fonte *League Scripts*<sup>1</sup>, com o intuito dos participantes realizarem o fac-símile dessas palavras.

O programa foi desenvolvido em *Python* utilizando as biblioteca *PyGame*, que permitiu a criação de uma interface gráfica para a exibição das palavras e a interação do participante. Além disso, o processamento das imagens foi realizado utilizando a biblioteca *OpenCV*, possibilitando a análise automática dos traçados feitos por cada participante para cada imagem apresentada.

Dessa forma, o programa ofereceu uma ferramenta automatizada para a aplicação do experimento, por meio da análise dos delineamentos realizados para cada imagem apresentada. Isso permitiu a coleta sistemática dos dados e facilitou a análise posterior dos resultados obtidos no experimento (CARVALHO, 2018).

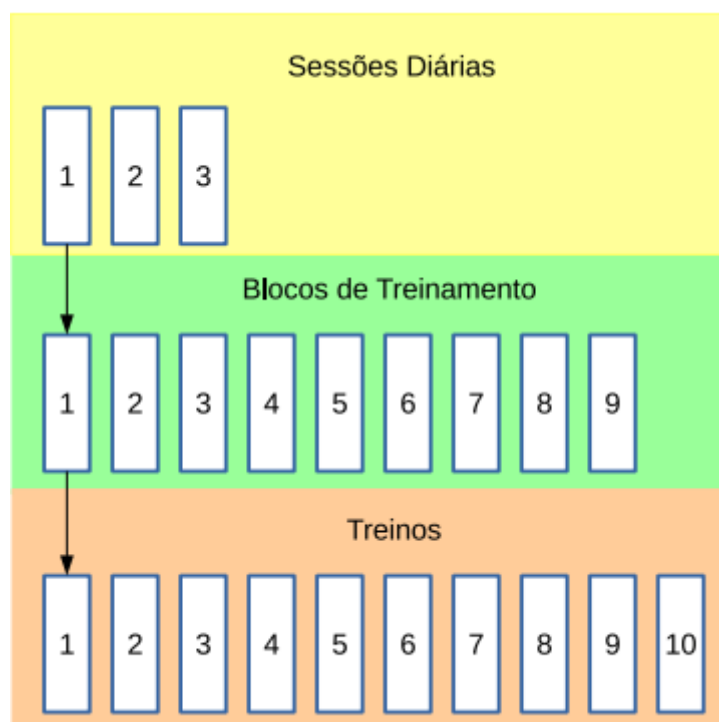
Utilizando o *software* foi possível realizar o experimento de aprendizado, no qual foram projetados testes para medir a diferença entre a imagem real e o erro de desenho produzido por cada participante.

Inicialmente, o experimento contou com a participação de vinte pessoas que possuíam bom estado de saúde e eram destros dominantes. Cada participante passou por dois questionários. O primeiro (*Edinburgh Handedness Inventory*) tinha como objetivo avaliar a destreza com a mão direita ou esquerda. Caso o participante utilizasse ambas as mãos no cotidiano ou fosse canhoto, ele seria excluído do experimento. Em seguida, foi aplicado o segundo questionário (*Beck Depression Inventory*) para determinar a saúde mental dos candidatos. Aqueles que apresentaram resultados que comprometiam a execução do experimento foram desconsiderados (SANTOS, 2019). Após as análises dos questionários restaram 12 indivíduos para a prosseguimento do experimento.

<sup>1</sup> <https://www.theleagueofmoveabletype.com/>

Para o presente projeto, foram aplicadas uma série de 10 treinos agrupados em 10 blocos, realizados ao longo de três sessões consecutivas. Ou seja, similar a Figura 1, sendo que nessa tem-se apenas 9 blocos ao invés de 10 como no presente projeto.

Figura 1 – Desenho parcial do experimento



Fonte: (SANTOS, 2019)

A aquisição de dados foi organizada em treinos (*trials*), blocos e sessões. Durante as tentativas (*trials*), o participante possui um período de 10 segundos multiplicado pelo número de letras da palavra, durante esse tempo o voluntário deveria utilizar a mão esquerda para desenhar a palavra apresentada. A informação extraída desse desenho é obtido pela a diferença entre a palavra mostrada e o que foi escrito pelo avaliado, sendo denotada como métrica de erro do aprendizado motor (SANTOS, 2019).

Após a conclusão dos experimentos, o conjunto de dados foi formatado em formato CSV, para análise posterior. O CSV é utilizado para armazenar dados tabulares, permitindo a organização dos dados em linhas e colunas de forma clara e estruturada. O CSV facilita a manipulação e o processamento dos dados durante a análise.

### 3.1.2 Remoção de *Outliers*

O projeto executado por Silva (2022), que teve como objetivo a remoção de *outliers*, também faz parte da composição do *dataset* (Subseção 3.1.3), utilizado nas regressões. Esse projeto foi fundamental para compreender a composição e a integridade das informações presentes no conjunto de dados. Desta forma, por meio da identificação e remoção de



anomalias obteve-se um conjunto de dados mais confiável e adequado para a aplicação de técnicas de modelagem.

Com o objetivo de eliminar quaisquer dados assimétricos, foram aplicados no conjunto de dados (experimento - 3.1.1) o Intervalo Interquartil (IQR). Essas técnicas visam dar consistência aos dados, permitindo uma análise mais precisa e confiável, (SILVA, 2022).

Após a aplicação dessas técnicas de pré-processamento, restaram apenas 10 participantes com dados válidos no *dataset*, que foram utilizados para a análise e modelagem subsequentes.

### 3.1.3 Projeto Atual

Os dados do 10 participantes estavam organizados nas colunas: *volunteer*, erro sessão 1, erro sessão 2, erro sessão 3, bloco, *trial*. As colunas de erro foram renomeadas para "session\_1\_error", "session\_2\_error" e "session\_3\_error". Para cada indivíduo aplicou-se a média das *trials* em cada bloco, ou seja, a média das *trials* do bloco 1 para o indivíduo 1, média das *trials* do bloco 2 para o indivíduo 1, e assim sucessivamente até o bloco 10 para o indivíduo 1, assim como para os demais indivíduos. Essa organização permitiu obter um total de 100 casos para realização das regressões.

O *dataset* foi dividido em conjuntos de treinamento e teste, utilizando uma proporção de 70% para treinamento e 30% para teste. Essa divisão possibilitou avaliar o desempenho dos modelos em dados não vistos durante o treinamento, fornecendo uma medida objetiva da capacidade de generalização dos modelos. Mostra-se na 1 um exemplo do formato final, m

Tabela 1 – Exemplo da organização atual do *dataset*

index	volunteer	block	session_1_error	session_2_error	session_3_error
0	id01	1	0.1750	0,1657	0,1565

## 3.2 Materiais

Para a análise dos dados, foi utilizado um computador de mesa equipado com as seguintes especificações: processador AMD Ryzen 5 4600G, 16GB de memória RAM, placa de vídeo GTX 1050 Ti e sistema operacional Windows 10.

A implementação do projeto foi feita em *Python* 3.10 utilizando as bibliotecas *Pandas*, *NumPy*, *Matplotlib* e *Seaborn*. O *Pandas* foi utilizado para a manipulação e análise dos dados, enquanto o *NumPy* forneceu suporte para cálculos numéricos. As bibliotecas *Matplotlib* e *Seaborn* foram utilizadas para a visualização dos resultados e criação de gráficos.

A seleção de ferramentas foi feita com base na sua ampla aceitação e uso na comunidade científica e de análise de dados, devido ao oferecimento de recursos poderosos. O ambiente de desenvolvimento fornecido pelo *Visual Studio* em conjunto com *Jupyter Notebook*, sendo que este pode ser utilizado dentro *Visual Studio*.

### 3.3 Modelos de Regressão

Todos os métodos de regressão utilizados foram implementados com o auxílio da biblioteca *SciKit Learn*. Esta biblioteca é amplamente reconhecida e utilizada na comunidade acadêmica devido à sua versatilidade e facilidade de uso na criação de modelos de aprendizado de máquina. O *SciKit* oferece diversas funções que facilitam a construção e avaliação de algoritmos de regressão, proporcionando uma ampla gama de recursos para análise de dados. O uso dessa biblioteca contribuiu para a eficiência e precisão dos modelos desenvolvidos neste trabalho de pesquisa (PEDREGOSA et al., 2011).

#### 3.3.1 Regressão por vetores suporte

A configuração dos parâmetros do modelo SVR inclui os seguintes valores dentre os padrões de entrada, utilizou-se:

1. Variáveis de características: "session\_1\_error", "session\_2\_error"
2. Variável de rótulo: "session\_3\_error"
3.  $C = 1000$
4. Epsilon = 0,01

Os parâmetros personalizados foram definidos para ajustar o desempenho do modelo SVR de acordo com as necessidades e características específicas do conjunto de dados. O valor  $C$  representa o parâmetro de penalidade do erro, controlando a suavidade da função de decisão. O valor Epsilon determina a largura da margem de erro. As variáveis de controle de erro são particularidades deste modelo (Seção 2.1).

#### 3.3.2 Regressão Linear

As entradas utilizadas foram as seguintes:

1. Variáveis de características: "session\_1\_error", "session\_2\_error"
2. Variável de rótulo: "session\_3\_error"

Utilizou-se a representação simples do algoritmo de regressão linear (Seção 2.2).

### 3.3.3 Regressão por Perceptron Multicamadas

Os parâmetros de execução do MLP Regressor que foram alterados em relação aos padrões são os seguintes:

1. Variáveis de características: "session\_1\_error", "session\_2\_error"
2. Variável de rótulo: "session\_3\_error"
3. Solver: lbfgs
4. Iterações Máximas: 1000
5. Tamanho das Camadas Ocultas: 3 camadas com 100 unidades cada

O uso do solver lbfgs foi escolhido devido ao tamanho limitado do *dataset* utilizado, tornando-o uma opção ideal para o problema em questão. A utilização do solucionador está relacionado com sua capacidade de otimizar o regressor baseando-se em aproximações simples de matrizes hessianas (Subseção 2.3.1).

## 3.4 Métricas de Avaliação

Quatro métricas foram utilizadas para avaliação do desempenho, são estas: MSE, MAE,  $R^2$  e Erro Percentual.

### 3.4.1 Erro quadrático médio (MSE)

O *MSE* é calculado somando o quadrado das diferenças entre os valores reais ( $y$ ) e os valores previstos ( $\hat{y}$ ), e em seguida, dividindo essa soma pelo número de amostras no conjunto de dados.

$$\text{MSE}(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N} \quad (3.1)$$

Quanto menor o valor do *MSE*, melhor é o desempenho do modelo, indicando uma maior precisão nas previsões. Fornecendo mais peso aos erros maiores, penalizando mais modelos imprecisos. O *MSE* é uma métrica útil para identificar a dispersão dos erros.

### 3.4.2 Erro absoluto médio (MAE)

A métrica mede a média dos resíduos, ou seja, das diferenças entre os valores observados ( $y$ ) e os valores estimados ( $\hat{y}$ ) pelo modelo. O MAE fornece uma medida direta da magnitude média dos erros, sem levar em consideração sua direção.

$$\text{MAE}(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N} \quad (3.2)$$

Assim como o MSE, para o MAE, quanto menor o valor obtido, melhor é o desempenho do modelo, expondo-se os erros próximos da mediana são mais importantes do que os erros extremos (*outliers*).

### 3.4.3 Coeficiente de Determinação ( $R^2$ )

O coeficiente de determinação ( $R^2$ ) é uma métrica fundamental para a avaliação da adaptação de um modelo de regressão. Ele mede a proporção da variabilidade dos dados observados que é explicada pelo modelo. Essa métrica varia de 0 a 1, e seu valor indica a quantidade de variabilidade dos dados de resposta (variável dependente -  $y$ ) que pode ser atribuída às variáveis preditoras incluídas no modelo.

$$R^2(y, \hat{y}) = 1 - \frac{SQ_{res}}{SQ_{tot}} \quad (3.3)$$

Quanto mais perto do 1, maior a adaptação do modelo. Contudo não necessariamente representa a precisão das previsões.

### 3.4.4 Erro Percentual

A métrica avalia o desempenho dos modelos de previsão, comparando os erros preditos ( $\hat{y}$ ) com os erros observados (real -  $y$ ) e quantificando a precisão das previsões em termos percentuais. Quanto menor o erro percentual, mais próximo o valor predito está do valor real.

$$\text{Erro Percentual}(\%)(y, \hat{y}) = \frac{y - \hat{y}}{y} * 100 \quad (3.4)$$

O erro percentual foi aplicado para cada dado presente no *dataset* de resultados. Quanto menor o erro percentual, mais próximo o valor predito está do valor real. Fornece uma medida relativa do erro do modelo, independente da escala dos dados.

## 4 Resultados

Nesta seção serão apresentados os resultados obtidos com os algoritmos de predição (*SVRLinear*, Regressão Linear e MLP com solucionador *lbfgs*), do erro em uma sessão, usando o *dataset* de aprendizado motor.

Inicialmente verifica-se o aprendizado *online* e *offline* dos participantes no experimento (Seção 4.1). Posteriormente, mostra-se os valores preditos e compara-os com os valores reais para *SVRLinear* (Subseção 4.2.1), Regressão Linear (Subseção 4.2.2) e MLP (Subseção 4.2.3). Por último compara-se os resultados dos regressores (Subseção 4.2.4).

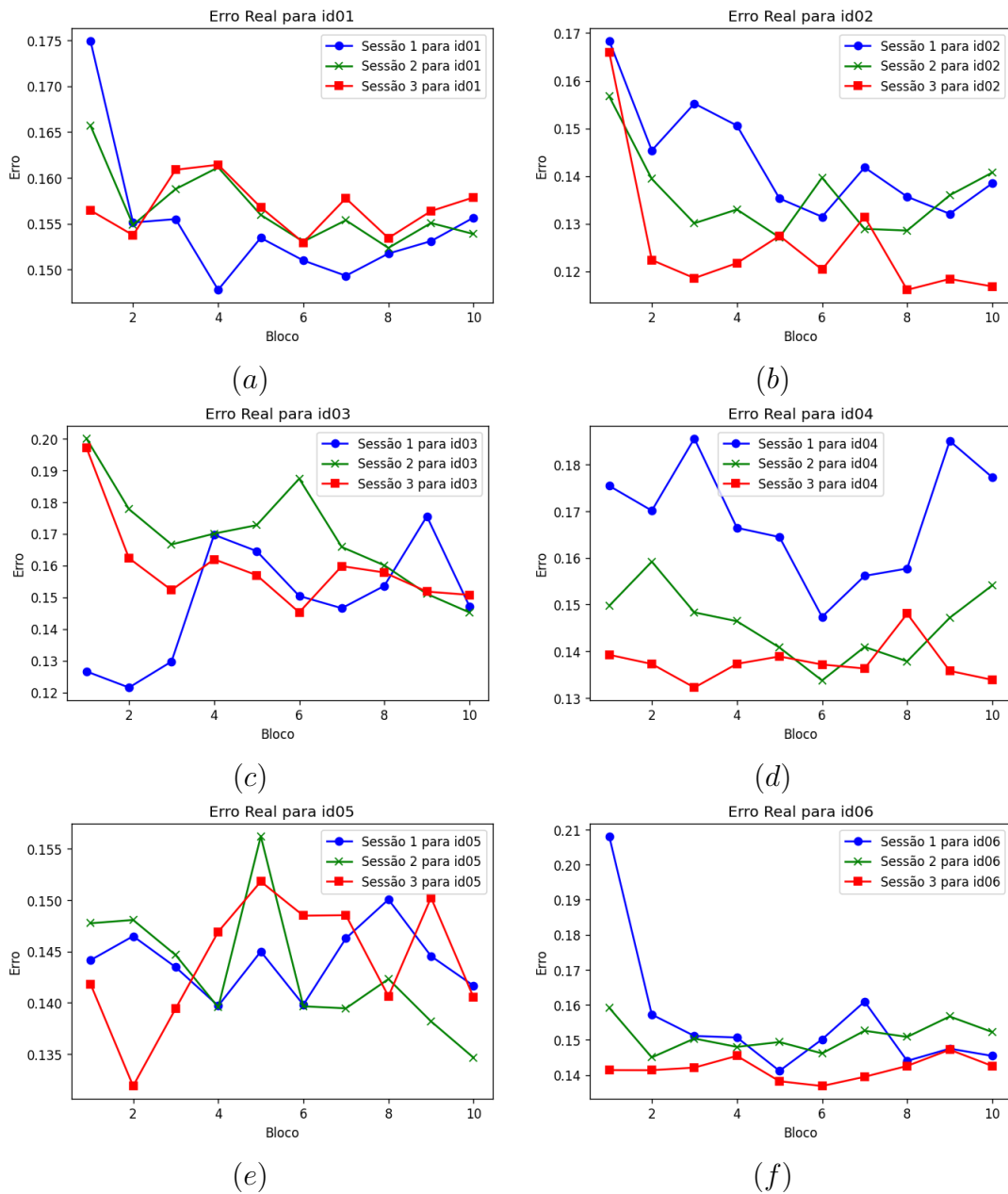
### 4.1 Resultados do Aprendizado Motor

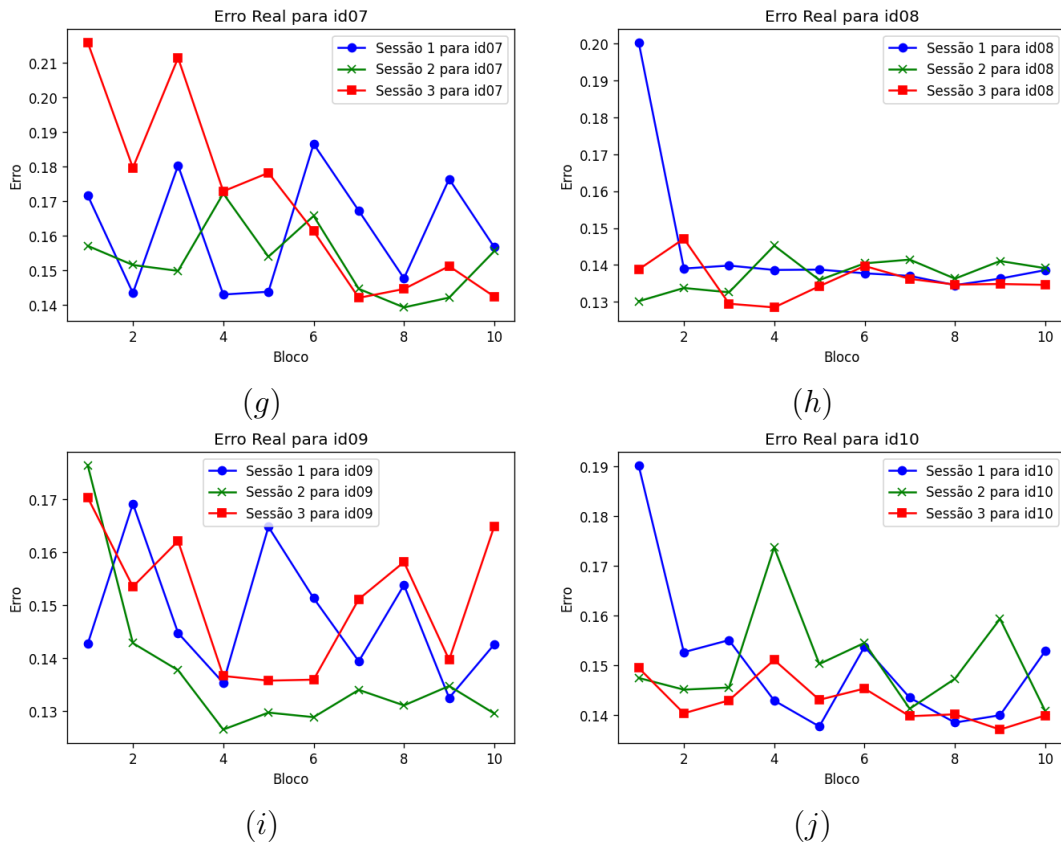
A Figura 2 mostra o aprendizado motor da tarefa, ou seja, uma diminuição do erro (do aprendizado motor) ao longo do experimento, isto é, o primeiro bloco da primeira sessão (Bloco 1 da Sessão 1) apresenta um erro maior do que o último bloco da última sessão (Bloco 10 da Sessão 3), com exceção da Figura 2 (a e i).

Percebe-se pela Figura 2 uma diminuição do erro médio (do aprendizado motor) para cada bloco (eixo das abcissas), ou seja, aprendizado *online*. Percebe-se aprendizado *online* em várias sessões para vários participantes (Sessão 1 da Figura 2 (a), Sessão 3 da Figura 2 (b), Sessão 2 da Figura 2 (c), Sessão 1 da Figura 2 (f), etc). Para a terceira seção do experimento esse aprendizado *online* aconteceu em todos os casos, com exceção da Figura 2(a).

Adicionalmente, a Figura 2 demonstra uma diminuição do erro médio para cada sessão (1 - Azul, 2 - Verde e Vermelho - 3), ou seja, aprendizado *offline*. Foi identificado um processo de aprendizado *offline* em pelo menos uma das seções de teste para os voluntários (Nas Sessões 2 e 3 da Figura 2 (d), Sessão 3 da Figura 2 (f), Sessões 2 e 3 da Figura 2 (h)). As Figuras 2 (a), 2 (b), 2 (c), 2 (e) e 2 (g) não mostram claramente esse aprendizado *offline*.

Figura 2 – Conjunto de figuras do erro por bloco para cada indivíduo no experimento.





Fonte: acervo do autor.

## 4.2 Resultados dos Regressores

### 4.2.1 Resultados SVRLinear

A regressão *SVRLinear* foi aplicada ao *dataset* de treinamento utilizando a abordagem de *Support Vector Regression* com um *kernel* linear. A Tabela 2 mostra os resultados desse regressor.

Tabela 2 – Resultado de Métricas SVRLinear

	MAE	MSE	$R^2$	Maior Diferença %	Menor Diferença %
LinearSVR	0,009999	0,000205	0,153341	27,10%	0,14%

A Tabela 3 mostra o índice/linha, ID, bloco, valor real, valor predito, diferença (real e predito) e diferença percentual, sendo que os dados estão ordenados por menor diferença percentual (absoluta). A menor diferença nominal (*index*/linha 0) foi 0.0002 em relação ao valor real. A maior diferença nominal (*index*/linha 29) foi 0.0573. Destaca-se que dentre os cinco melhores resultados (menor diferença), quatro pertencem ao id01.

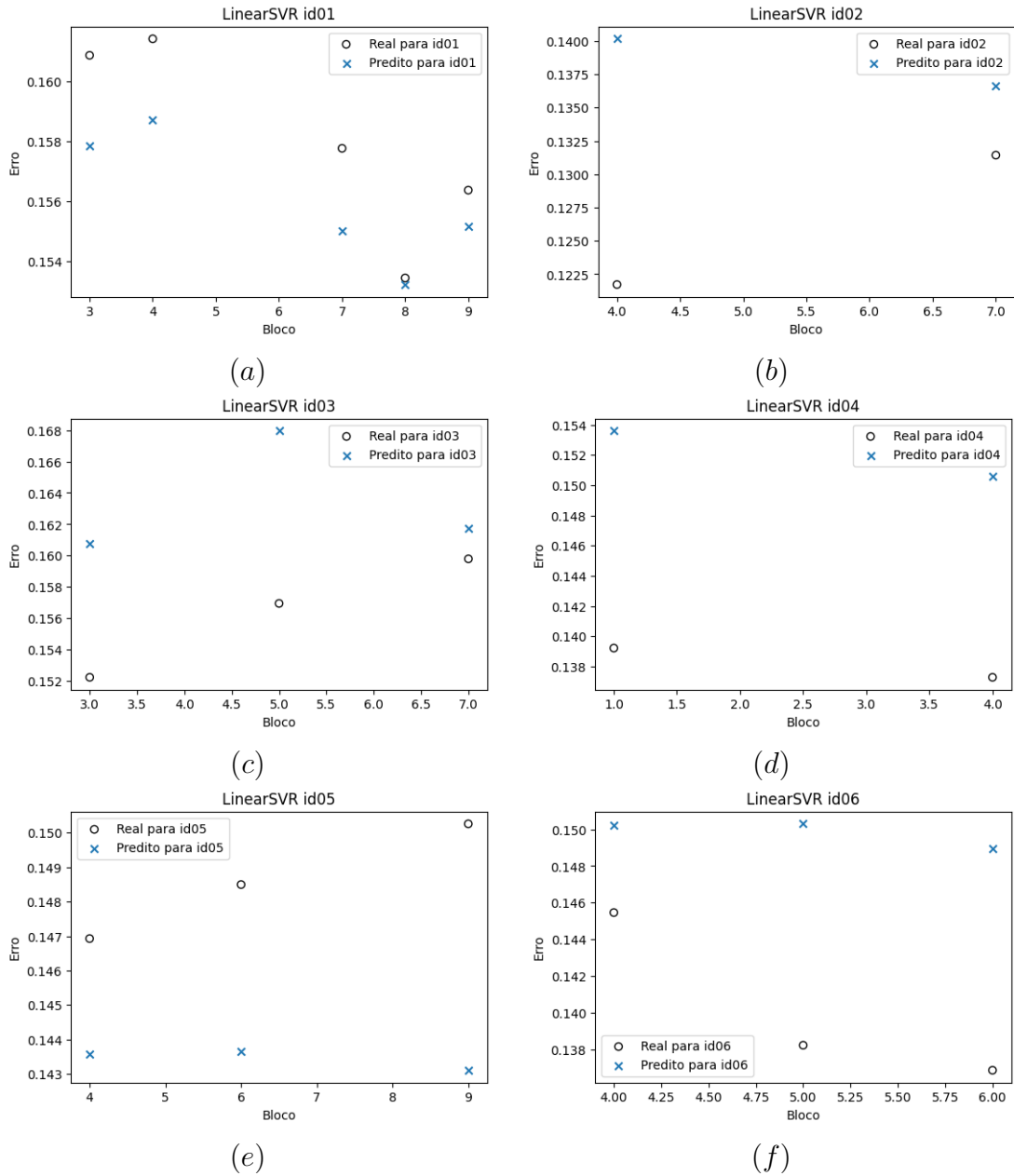
Tabela 3 – Conjunto de dados preditos - SVRLinear

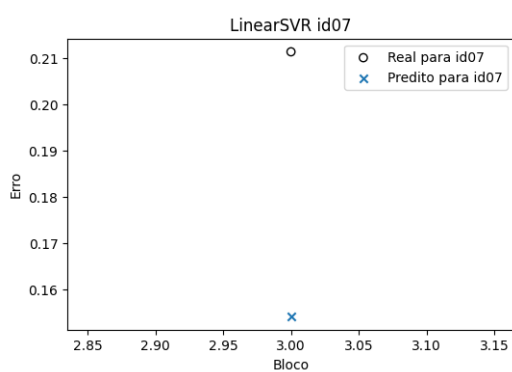
index	id	bloco	real	predito	diff	diff_percentual
0	id01	8	0,153433	0,153213	0,000220	-0,14%
1	id01	9	0,156370	0,155148	0,001222	-0,78%
2	id03	7	0,159790	0,161733	-0,001944	1,22%
3	id01	4	0,161429	0,158724	0,002706	-1,68%
4	id01	7	0,157768	0,155008	0,002760	-1,75%
5	id01	3	0,160878	0,157839	0,003039	-1,89%
6	id05	4	0,146928	0,143582	0,003346	-2,28%
7	id08	6	0,139726	0,144004	-0,004278	3,06%
8	id06	4	0,145464	0,150227	-0,004763	3,27%
9	id05	6	0,148495	0,143659	0,004836	-3,26%
10	id02	7	0,131434	0,136660	-0,005226	3,98%
11	id10	10	0,139972	0,145641	-0,005669	4,05%
12	id10	3	0,142994	0,148984	-0,005991	4,19%
13	id05	9	0,150258	0,143106	0,007152	-4,76%
14	id08	2	0,147134	0,139659	0,007475	-5,08%
15	id08	7	0,136246	0,144613	-0,008368	6,14%
16	id03	3	0,152226	0,160760	-0,008534	5,61%
17	id08	9	0,134850	0,144295	-0,009445	7,00%
18	id10	6	0,145368	0,154859	-0,009491	6,53%
19	id03	5	0,156944	0,167955	-0,011011	7,02%
20	id09	7	0,151167	0,139856	0,011312	-7,48%
21	id06	5	0,138235	0,150297	-0,012062	8,73%
22	id06	6	0,136873	0,148948	-0,012074	8,82%
23	id04	4	0,137290	0,150569	-0,013278	9,67%
24	id04	1	0,139224	0,153588	-0,014364	10,32%
25	id10	4	0,151113	0,166694	-0,015581	10,31%
26	id02	4	0,121721	0,140166	-0,018445	15,15%
27	id08	4	0,128506	0,147291	-0,018785	14,62%
28	id09	3	0,162151	0,142817	0,019334	-11,92%
29	id07	3	0,211355	0,154071	0,057283	-27,10%

A Figura 3 retrata alguns dos resultados previamente demonstrados na Tabela 3, nomeadamente coluna valor real e predito, no entanto de maneira gráfica. Observa-se o melhor resultado (*index*/linha 0 da Tabela 3) no bloco 8 da Figura 3(a). Similarmente, para o pior resultado (*index*/linha 29 da Tabela 3) no bloco 3 da Figura 3(g).

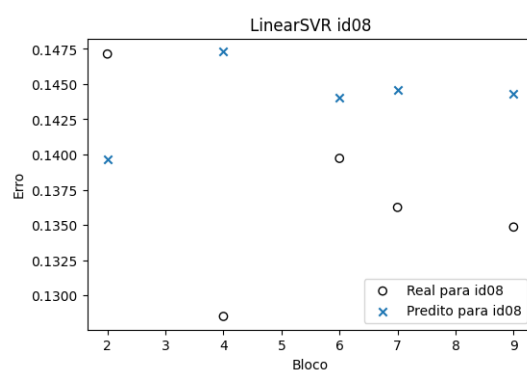


Figura 3 – Conjunto de figuras do erro por bloco para cada individuo no SVRLinear.

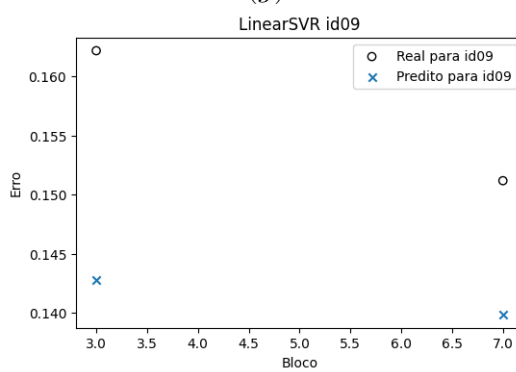




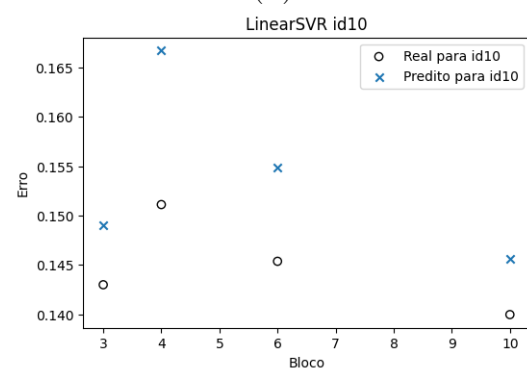
(g)



(h)



(i)



(j)

Fonte: acervo do autor.

## 4.2.2 Resultados Regressão Linear

Em seguida, foi realizada a regressão linear. O modelo recebeu os dados de treinamento e avaliado utilizando o conjunto de teste. A Tabela 4 apresenta os resultados métricos desse algoritmo de regressão.

Tabela 4 – Resultado de Métricas Regressão Linear

	MAE	MSE	$R^2$	Maior Diferença %	Menor Diferença %
LinearReg	0,009295	0,000213	0,121323	29,72%	0,09%

Tabela 5 – Conjunto de dados preditos - Regressão Linear

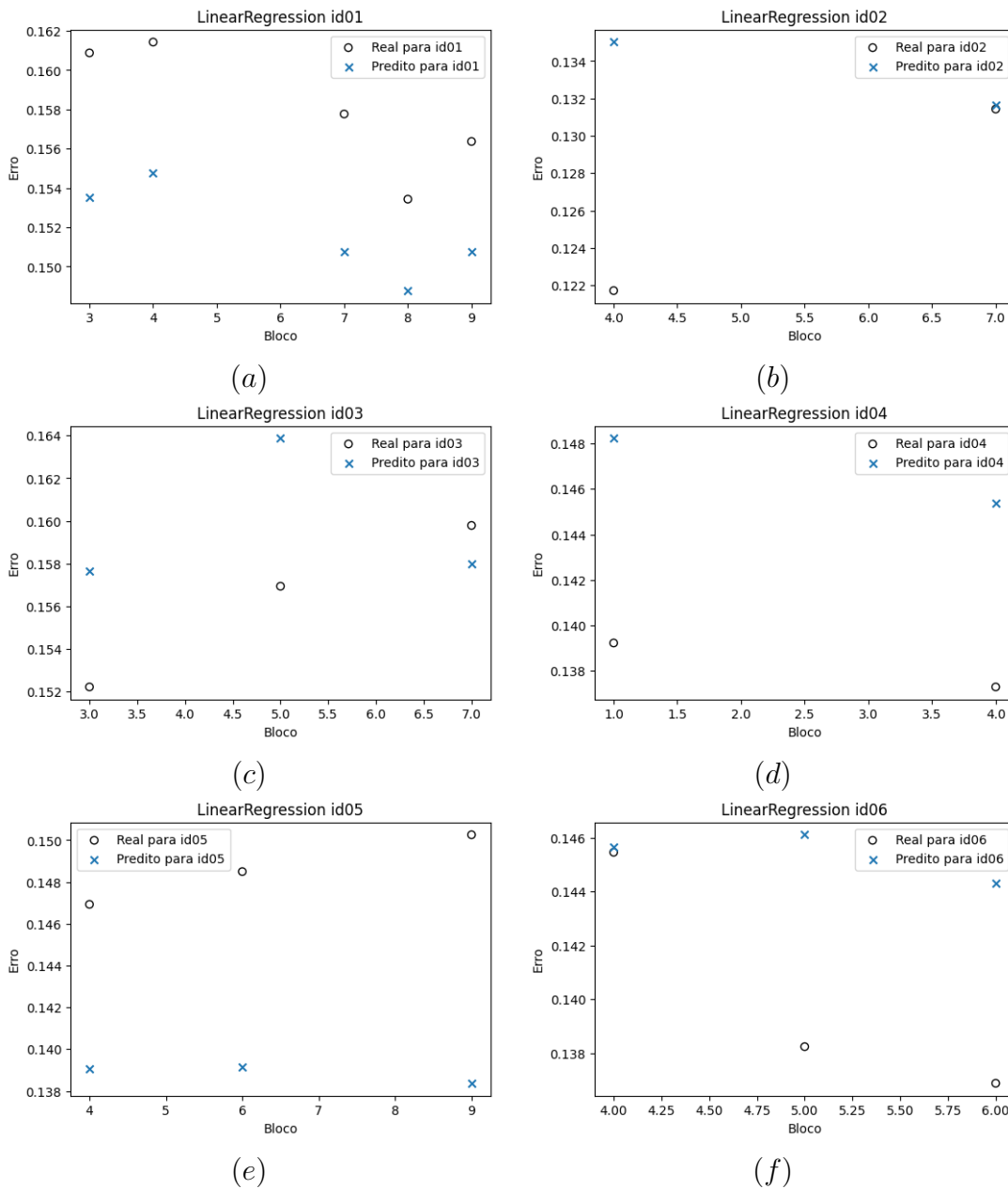
index	id	bloco	real	predito	diff	diff_percentual
0	id08	6	0,139726	0,139595	0,000132	-0,09%
1	id06	4	0,145464	0,145660	-0,000196	0,13%
2	id02	7	0,131434	0,131660	-0,000226	0,17%
3	id10	10	0,139972	0,140718	-0,000747	0,53%
4	id10	3	0,142994	0,144169	-0,001175	0,82%
5	id03	7	0,159790	0,158005	0,001785	-1,12%
6	id08	7	0,136246	0,140268	-0,004022	2,95%
7	id01	8	0,153433	0,148779	0,004654	-3,03%
8	id10	6	0,145368	0,150440	-0,005072	3,49%
9	id08	9	0,134850	0,139960	-0,005110	3,79%
10	id03	3	0,152226	0,157650	-0,005424	3,56%
11	id01	9	0,156370	0,150772	0,005598	-3,58%
12	id01	4	0,161429	0,154769	0,006660	-4,13%
13	id03	5	0,156944	0,163866	-0,006923	4,41%
14	id01	7	0,157768	0,150776	0,006993	-4,43%
15	id01	3	0,160878	0,153524	0,007355	-4,57%
16	id06	6	0,136873	0,144327	-0,007454	5,45%
17	id05	4	0,146928	0,139069	0,007859	-5,35%
18	id06	5	0,138235	0,146116	-0,007881	5,70%
19	id04	4	0,137290	0,145390	-0,008100	5,90%
20	id04	1	0,139224	0,148225	-0,009001	6,47%
21	id05	6	0,148495	0,139147	0,009348	-6,30%
22	id05	9	0,150258	0,138371	0,011886	-7,91%
23	id08	2	0,147134	0,134945	0,012189	-8,28%
24	id10	4	0,151113	0,163398	-0,012285	8,13%
25	id02	4	0,121721	0,135017	-0,013296	10,92%
26	id08	4	0,128506	0,143037	-0,014531	11,31%
27	id09	7	0,151167	0,135137	0,016030	-10,60%
28	id09	3	0,162151	0,138056	0,024094	-14,86%
29	id07	3	0,211355	0,148542	0,062813	-29,72%

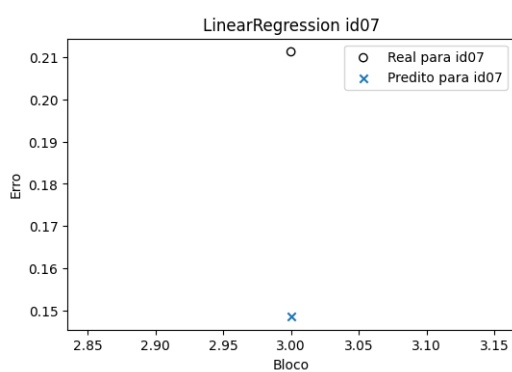
A Tabela 5 mostra o índice/linha, ID, bloco, valor real, valor predito, diferença (real e predito) e diferença percentual, sendo que os dados estão ordenados por menor

diferença percentual (absoluta). A menor diferença percentual (*index*/linha 0) foi 0,09% em relação ao valor real. Para maior diferença (*index*/linha 29) foi obtido 29,72%.

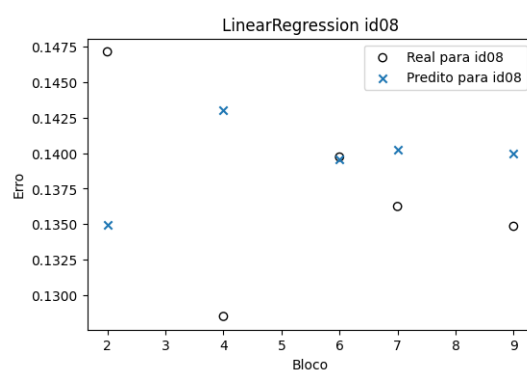
A Figura 4 retrata os resultados demonstrados na Tabela 5 visando uma análise visual da predição. Observa-se o melhor resultado (*index*/linha 0 da Tabela 5) no bloco 6 da Figura 4(h). Similarmente, para o pior resultado (*index*/linha 29 da Tabela 5) no bloco 3 da Figura 4(g).

Figura 4 – Conjunto de figuras do erro por bloco para cada indivíduo no modelo de Regressão Linear.

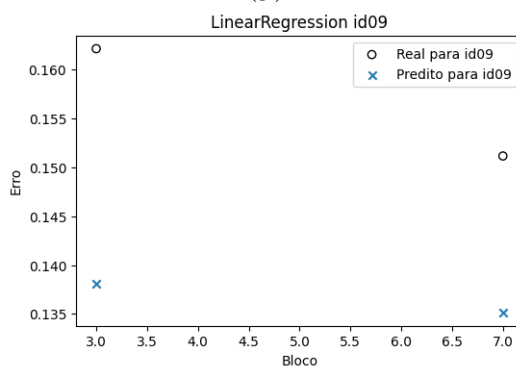




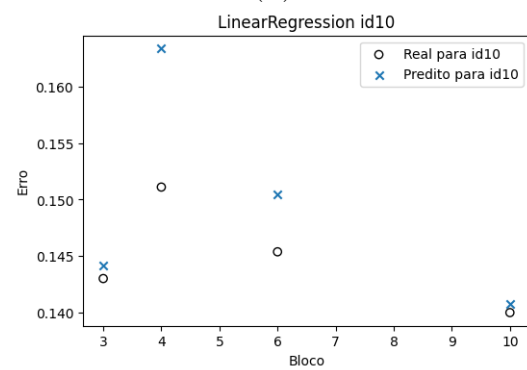
(g)



(h)



(i)



(j)

Fonte: acervo do autor.

### 4.2.3 Resultados MLP de Regressão

A regressão MLP (*Multilayer Perceptron*) foi aplicada ao *dataset* de treinamento utilizando a perspectiva de *Multilayer Perceptron* com um *solver* otimizado de Método Quasi-Newton. A Tabela 6 mostra os resultados desse regressor.

Tabela 6 – Resultado de Métricas MLP de Regressão

	MAE	MSE	$R^2$	Maior Diferença %	Menor Diferença %
MLPRegressor	0,009347	0,000283	0,162799	27,92%	0,13%

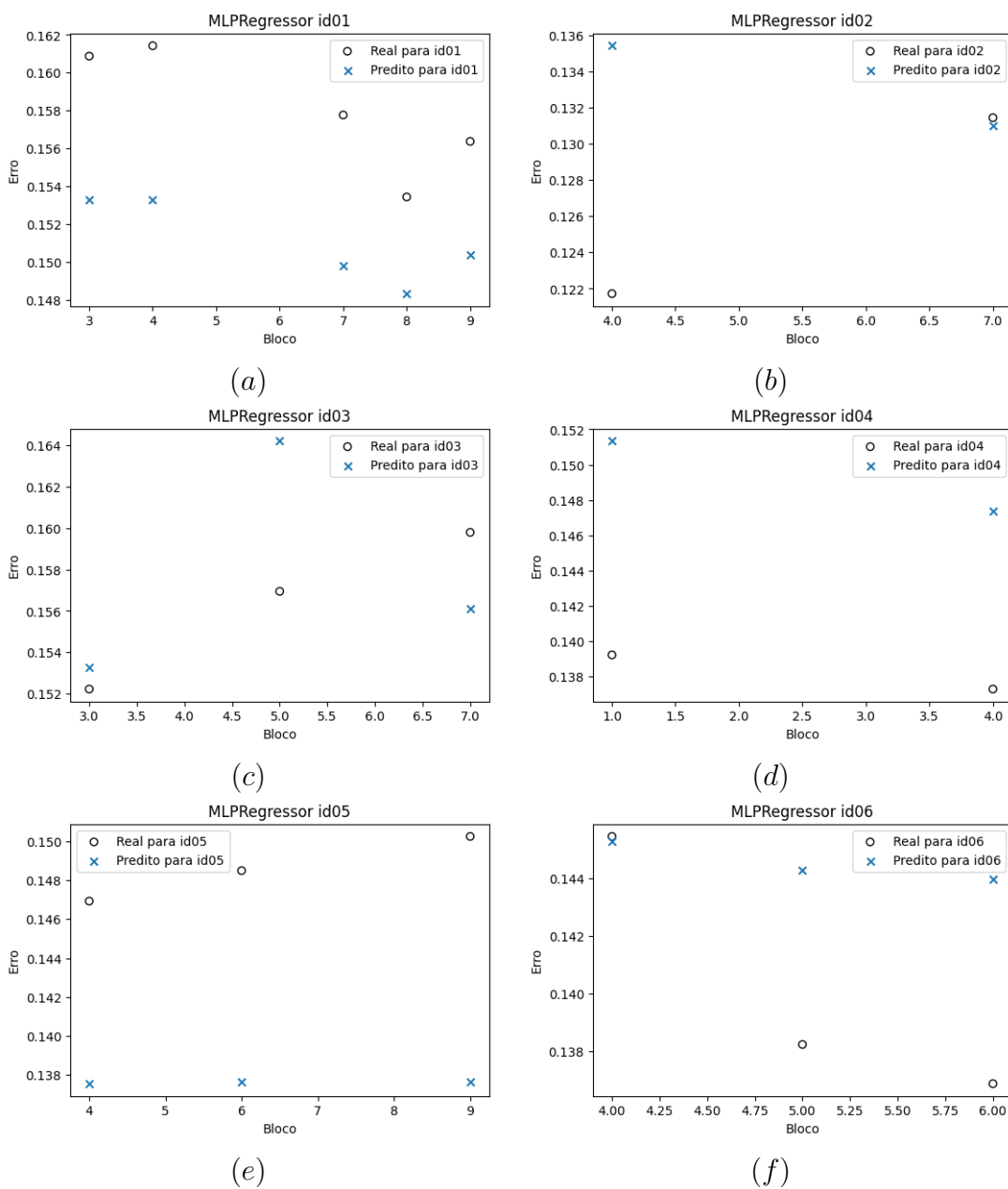
Tabela 7 – Conjunto de dados preditos - MLP de Regressão

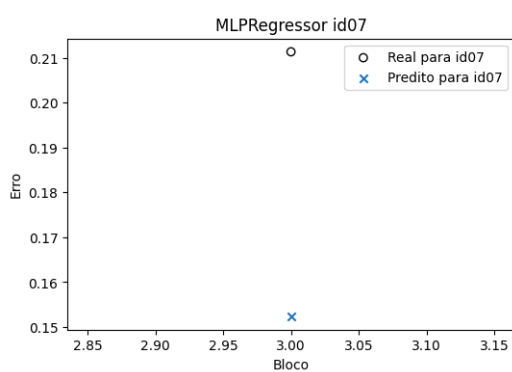
index	id	bloco	real	predito	diff	diff_percentual
0	id06	4	0,145464	0,145279	0,000185	-0,13%
1	id02	7	0,131434	0,130988	0,000445	-0,34%
2	id03	3	0,152226	0,153254	-0,001028	0,68%
3	id10	10	0,139972	0,141030	-0,001058	0,76%
4	id10	3	0,142994	0,144551	-0,001558	1,09%
5	id08	6	0,139726	0,137734	0,001992	-1,43%
6	id08	7	0,136246	0,138251	-0,002006	1,47%
7	id08	9	0,134850	0,137858	-0,003008	2,23%
8	id03	7	0,159790	0,156087	0,003703	-2,32%
9	id10	6	0,145368	0,150159	-0,004791	3,30%
10	id01	8	0,153433	0,148321	0,005113	-3,33%
11	id01	9	0,156370	0,150368	0,006002	-3,84%
12	id06	5	0,138235	0,144284	-0,006049	4,38%
13	id06	6	0,136873	0,143965	-0,007092	5,18%
14	id03	5	0,156944	0,164195	-0,007252	4,62%
15	id01	3	0,160878	0,153274	0,007604	-4,73%
16	id01	7	0,157768	0,149809	0,007959	-5,05%
17	id01	4	0,161429	0,153278	0,008151	-5,05%
18	id05	4	0,146928	0,137537	0,009391	-6,39%
19	id10	4	0,151113	0,160550	-0,009437	6,24%
20	id04	4	0,137290	0,147370	-0,010080	7,34%
21	id05	6	0,148495	0,137625	0,010871	-7,32%
22	id04	1	0,139224	0,151333	-0,012109	8,70%
23	id08	4	0,128506	0,141056	-0,012550	9,77%
24	id05	9	0,150258	0,137613	0,012645	-8,42%
25	id08	2	0,147134	0,133620	0,013514	-9,19%
26	id02	4	0,121721	0,135407	-0,013686	11,24%
27	id09	7	0,151167	0,133855	0,017313	-11,45%
28	id09	3	0,162151	0,137354	0,024797	-15,29%
29	id07	3	0,211355	0,152348	0,059006	-27,92%

A Tabela 7 mostra o índice/linha, ID, bloco, valor real, valor predito, diferença (real e predito) e diferença percentual, sendo que os dados estão organizados por menor diferença percentual (absoluta). A menor diferença nominal (*index*/linha 0) foi 0.0002 em relação ao valor real. A maior diferença percentual (*index*/linha 29) foi 0.0590.

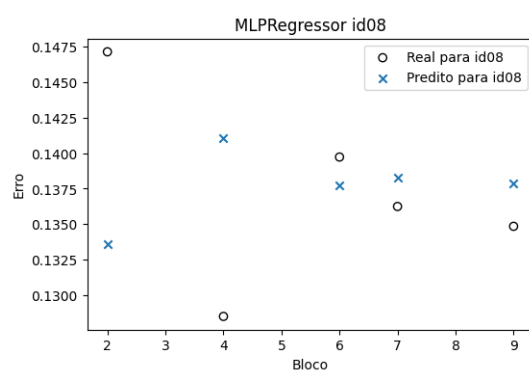
A Figura 5 retrata visualmente os resultados demonstrados na Tabela 7. Observa-se o melhor resultado (*index*/linha 0 da Tabela 7) no bloco 4 da Figura 5(f). Similarmente, para o pior resultado (*index*/linha 29 da Tabela 7) no bloco 3 da Figura 5(g).

Figura 5 – Conjunto de figuras do erro por bloco para cada indivíduo no modelo MLP de Regressão.

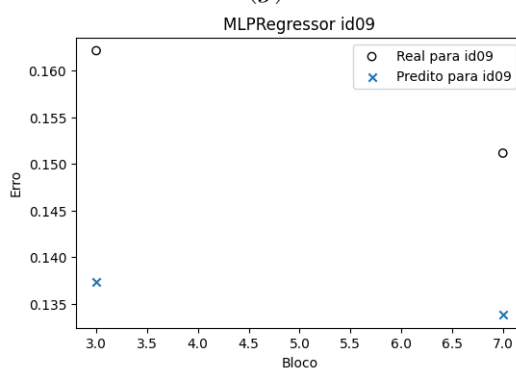




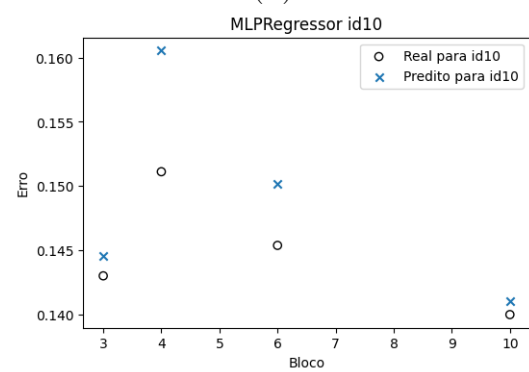
(g)



(h)



(i)



(j)

Fonte: acervo do autor.



#### 4.2.4 Resultados Comparativos

Os resultados gerais (regressores e métricas) são sumarizados na Tabela 8:

Tabela 8 – Resultados das Métricas Gerais

	MAE	MSE	$R^2$	Maior Diferença %	Menor Diferença %
LinearSVR	0,009999	0,000205	0,153341	27,10%	0,14%
LinearReg	0,009295	0,000213	0,121323	29,72%	0,09%
MLPRegressor	0,009347	0,000283	0,162799	27,92%	0,13%

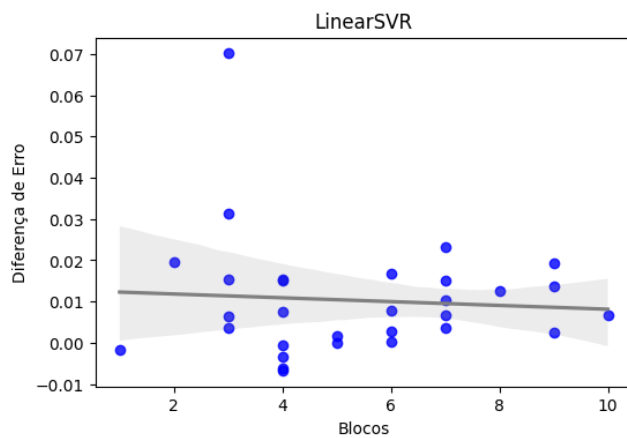
Na Tabela 8 identifica-se que todos os modelos (SVR, LR e Regressão MLP) apresentam métricas (MAE, MSE,  $R^2$ , Maior Diferença Percentual e Menor Diferença Percentual) com valores aceitáveis. Por exemplo, o modelo de *SVM* de Regressão possui o menor valor de erro quadrático médio (*MSE*), enquanto o *MLPRegressor* obteve o melhor coeficiente de determinação ( $R^2$ ). A Regressão Linear possui o melhor valor de erro absoluto médio (*MAE*), mesmo sendo o pior modelo em termos de coeficiente de determinação.

No contexto da regressão linear, destaca-se que esse modelo apresenta a maior e a menor diferença entre os erros reais e os erros preditos para a terceira seção do experimento.

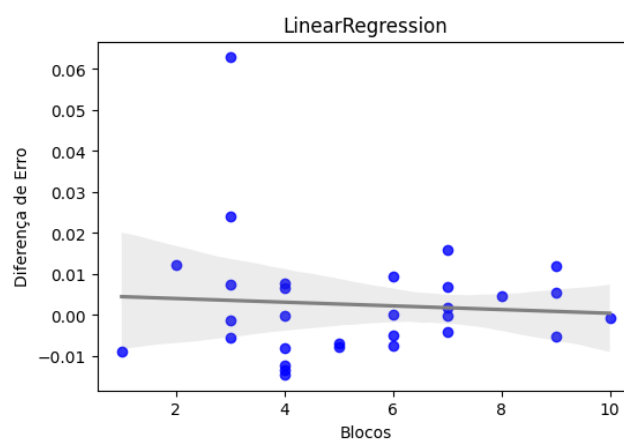
As Figuras 6(a), 6(b), 6(c) demonstram o impacto dos erros preditos que possuem os maiores percentuais. Utilizou-se uma regressão linear para traçar a tendência desse conjunto de erros, incluindo também uma curva de confiança caracterizada pelo sombreamento da reta na Figura 6.

Essa análise proporciona uma visão mais abrangente do desempenho dos modelos em relação aos erros preditos com diferenças percentuais significativas. As figuras fornecem *insights* valiosos sobre a qualidade das previsões e a confiabilidade dos resultados obtidos, o que permite uma avaliação mais precisa do desempenho dos modelos em situações em que ocorrem variações acentuadas nos erros.

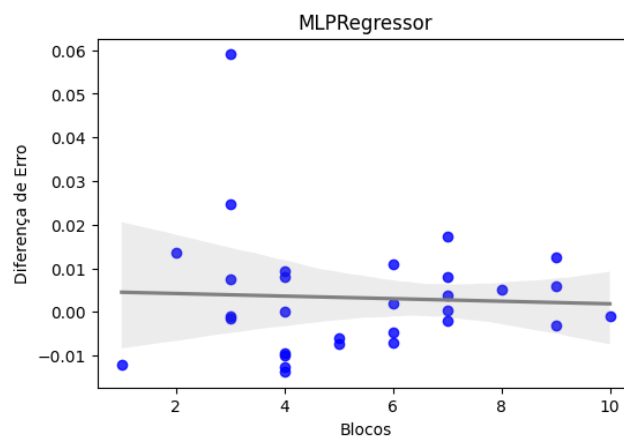
Figura 6 – Conjunto de figuras traçadas em Regressão Linear para o erros preditos



(a)



(b)



(c)

Fonte: acervo do autor.

Constata-se com base nos modelos apresentados nas Figuras (SVR - 6(a), LR - 6(b) e 6(c) - MLP Regressora), apesar de apresentarem diferenças superiores a 0,03, em termos absolutos, tais variações não afetam significativamente a tendência geral dos modelos.

Nota-se que todos os gráficos exibem um coeficiente angular que tende a zero. Ou seja, sugerem que os resultados finais (erro predito) dos modelos são aproximadamente equivalentes.

Evidencia-se que cada modelo possui sua especialidade, robustez e capacidade de adaptação. Conquanto, um fator determinante para o resultado final é o conjunto de dados utilizado, o qual influencia diretamente devido aos ajustes realizados e ao tamanho do grupo de informações em questão.

## 5 Conclusão

O objetivo do trabalho foi utilizar algoritmos de regressão para a predição do erro em um bloco para uma tarefa de aprendizado motor. A partir dos resultados obtidos, foi possível observar que os modelos de regressão utilizados apresentaram um bom desempenho na predição do erro, fornecendo estimativas consistentes e próximas aos valores reais.

Apesar dos resultados promissores, é importante ressaltar algumas limitações deste estudo. Em primeiro lugar, o tamanho do conjunto de dados utilizado foi relativamente pequeno, o que pode ter impactado a generalização dos modelos de regressão. Além disso, foram consideradas apenas variáveis específicas para a predição do erro, não explorando outros fatores que podem influenciar o desempenho motor. Outra limitação é o foco exclusivo na predição do erro em blocos e sessões de treinamento, não abrangendo outras fases, por exemplo *trials*, do experimento de aprendizado motor.

Evidencia-se a importância da qualidade dos dados para obter resultados confiáveis. No mais, evidenciou-se que modelos mais simples (Regressão Linear) podem ter resultados análogos aos mais complexos (MLP).

Contribuiu-se para o campo da tarefa de aprendizado motor ao propor uma abordagem baseada em algoritmos de regressão para a predição do erro em um bloco de treinamento. Os resultados obtidos podem auxiliar no desenvolvimento de estratégias mais eficientes de treinamento, bem como, tentar prever resultados futuros a partir dos dados atuais.

Como trabalhos futuros, sugere-se a ampliação do conjunto de dados, incluindo um maior número de participantes e a coleta de dados em diferentes contextos e tarefas motoras. Além disso, é importante explorar outras variáveis e técnicas de análise de dados para aprimorar a precisão dos modelos de regressão. Também é recomendado investigar a predição de novas sessões e blocos utilizando os erros referentes às *trials* e explorar outros algoritmos de aprendizado de máquina para comparação e aprimoramento dos resultados. Por fim, é fundamental realizar estudos que envolvam indivíduos com danos motores, a fim de aplicar e validar a abordagem proposta em um contexto clínico real, por exemplo para predição de reabilitação motora.

## Referências

- CARVALHO, M. B. F. d. *Plataforma de avaliação automatizada do aprendizado de habilidade motora*. Monografia (Bacharelado - BICT) — Universidade Federal do Maranhão, São Luís, MA, Brasil, 2018. Citado 3 vezes nas páginas 13, 20 e 21.
- DAYAN, E.; COHEN, L. G. Neuroplasticity subserving motor skill learning. *Neuron*, Elsevier BV, v. 72, n. 3, p. 443–454, nov. 2011. Disponível em: <<https://doi.org/10.1016/j.neuron.2011.10.008>>. Citado na página 13.
- HUANG, T.-M.; KECCMAN, V.; KOPRIVA, I. *Kernel Based Algorithms for Mining Huge Data Sets*. Springer-Verlag, 2006. 11–57 p. Disponível em: <<https://doi.org/10.1007/3-540-31689-2>>. Citado na página 16.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R.; TAYLOR, J. Linear regression. In: *Springer Texts in Statistics*. Springer International Publishing, 2023. p. 69–134. Disponível em: <[https://doi.org/10.1007/978-3-031-38747-0\\_3](https://doi.org/10.1007/978-3-031-38747-0_3)>. Citado na página 17.
- KRAKAUER, J. W.; HADJIOSIF, A. M.; XU, J.; WONG, A. L.; HAITH, A. M. *Motor Learning*. Wiley, 2019. 613–663 p. Disponível em: <<https://doi.org/10.1002/cphy.c170043>>. Citado na página 13.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, Springer Science and Business Media LLC, v. 521, n. 7553, p. 436–444, maio 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>. Citado na página 18.
- LIU, D. C.; NOCEDAL, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, Springer Science and Business Media LLC, v. 45, n. 1-3, p. 503–528, ago. 1989. Disponível em: <<https://doi.org/10.1007/bf01589116>>. Citado na página 19.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 24.
- PRICHARD, G.; WEILLER, C.; FRITSCH, B.; REIS, J. Effects of different electrical brain stimulation protocols on subcomponents of motor skill learning. *Brain Stimulation*, Elsevier BV, v. 7, n. 4, p. 532–540, jul. 2014. Disponível em: <<https://doi.org/10.1016/j.brs.2014.04.005>>. Citado 2 vezes nas páginas 13 e 14.
- SANTOS, M. R. d. *Modelos de predição aplicados ao aprendizado motor*. Dissertação (Mestrado) — PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO/CCET, 2019. Disponível em: <<https://tedebc.ufma.br/jspui/handle/tede/tede/2571>>. Citado 5 vezes nas páginas 13, 14, 20, 21 e 22.
- SILVA, L. C. C. da. *Aplicação de Redes Neurais Recorrentes para a Predição do Aprendizado Motor*. Monografia (Bacharelado) — Universidade Federal do Maranhão, São Luís, MA, Brasil, 2022. Citado 3 vezes nas páginas 14, 22 e 23.

---

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and Computing*, Springer Science and Business Media LLC, v. 14, n. 3, p. 199–222, ago. 2004. Disponível em: <<https://doi.org/10.1023/b:stco.0000035301.49549.88>>. Citado na página 16.