



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
CURSO DE ENGENHARIA MECÂNICA

Rafael Albuquerque de Lima Pacheco

Utilização de modelos computacionais de aprendizagem de máquinas para previsão de valores de temperatura do rolamento da multiplicadora de um aerogerador em operação comercial.

FORTALEZA
2019

Rafael Albuquerque de Lima Pacheco

Utilização de modelos computacionais de aprendizagem de máquinas para previsão de valores de temperatura do rolamento da multiplicadora de um aerogerador em operação comercial.

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Mecânica do Departamento de Engenharia Mecânica da Universidade Federal do Ceará, como parte dos requisitos para obtenção do Título de Engenheiro Mecânico.

Orientador: Prof. Dr. Paulo Alexandre Costa Rocha

FORTALEZA

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

P121u Pacheco, Rafael Albuquerque de Lima.

Utilização de modelos computacionais de aprendizagem de máquinas para previsão de valores de temperatura do rolamento da multiplicadora de um aerogerador em operação comercial. / Rafael Albuquerque de Lima Pacheco. – 2019.

68 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Mecânica, Fortaleza, 2019.

Orientação: Prof. Dr. Paulo Alexandre Costa Rocha.

1. Energia Eólica. 2. Aprendizagem de Máquinas. 3. Manutenção Preditiva. 4. Regressão. 5. Support Vector Machine. I. Título.

CDD 620.1

Rafael Albuquerque de Lima Pacheco

Utilização de modelos computacionais de aprendizagem de máquinas para previsão de valores de temperatura do rolamento da multiplicadora de um aerogerador em operação comercial.

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia Mecânica do Departamento de Engenharia Mecânica da Universidade Federal do Ceará, como parte dos requisitos para obtenção do Título de Engenheiro Mecânico.

Aprovada em ____/____/____.

BANCA EXAMINADORA

Prof. Dr. Paulo Alexandre Costa Rocha
Universidade Federal do Ceará (UFC)

Prof. Dr. Rômulo Nascimento Rodrigues
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Elicivaldo Lima
Universidade Federal do Ceará (UFC)

M. Sc. Adão Linhares Muniz
Secretaria de Infraestrutura do Estado do Ceará (SEINFRA)

A Deus, criador e Senhor de todas as coisas.

Aos meus pais, José Antônio Pacheco e Rosilene Albuquerque de Lima Pacheco, a minha irmã Débora Albuquerque de Lima Pacheco e a minha amada esposa Marília Brilhante Queiroz por toda a paciência e apoio incondicional.

AGRADECIMENTOS

Agradeço primeiramente a Deus por ter me feito capaz de chegar até aqui.

Aos meus pais por sempre me apoiarem e suportarem durante o percurso que me trouxe até aqui, por seus inúmeros conselhos e admoestações e por buscarem produzir em mim a maturidade necessária para que eu consiga alcançar lugares mais altos.

Ao Prof. Dr. Paulo Alexandre Costa, por ter me aceitado como seu orientando e por ter feito uma excelente orientação, disponibilizando seu tempo e atenção sempre que possível.

Aos professores que eu tive a honra de ser aluno ao longo da graduação, pois por meio deles adquiri muito conhecimento, em alguns deles pude sentir o carinho e cuidado paternos que para mim foram de grande valia.

Aos professores participantes da banca examinadora pelo tempo investido na leitura e avaliação deste trabalho.

A minha esposa Marília Brilhante Queiroz por ter sido paciente e por ter me apoiado em todos os momentos com carinho e dedicação, sempre me inspirando a vencer a inércia e utilizar mais do potencial que tenho.

A minha irmã por estar comigo e ser sempre solícita, você foi por muito tempo o meu referencial de amor, carinho e família.

A família da minha esposa por terem me abraçado, aconselhado e apoiado, vocês são como minha família.

Aos meus amigos por me apoiarem e por sempre acreditarem em mim e estarem comigo em todos os momentos independente da distância geográfica.

A Energo Engenharia e a todos os meus colegas de trabalho, especialmente ao Adão Linhares Muniz, ao Samuel Araújo Lima, ao Gustavo Carvalho e ao Augusto Fontenele Magalhães.

A minha família materna, especialmente às minhas tias Rosângela e Robersivânia e à minha avó Laís, por terem me recebido em Fortaleza e me ajudado na adaptação à outra realidade longe dos meus pais.

Aos meus avós paternos, Sebastião Cota Pacheco e Terezinha de Jesus Pacheco, em memória, por me inspirarem a ser uma pessoa melhor, as memórias que tenho de vocês estarão sempre comigo.

RESUMO

Com a popularização da manutenção preditiva a busca por ensaios não destrutivos e técnicas de inspeção não invasivas tem aumentado constantemente em vários setores da indústria e com isso o uso de tecnologias inovadoras tem se destacado. O uso de ultrassom, imagens termográficas, análise de vibração e videoscopia são exemplo dessas técnicas. Outro tema em evidência é o uso de métodos computacionais de aprendizagem de máquinas. Esses métodos foram desenvolvidos nos anos 70, mas devido às limitações de processamento e armazenamento de dados só começaram a ser mais explorados por volta dos anos 2000 e tem sido cada vez mais utilizados. A versatilidade dos métodos fez com que fossem implementados a diversas áreas de estudo como biotecnologia, engenharia, estatística e marketing. Na engenharia esses métodos têm sido aplicados com sucesso para os mais diversos objetivos e o custo de implementação desses modelos tem reduzido cada vez mais. O objetivo do presente trabalho foi programar os métodos computacionais de aprendizagem de máquinas, Regressão Linear, Regressão Polinomial e Máquina de Vetores Suporte, desenvolvidos na linguagem de programação R a um conjunto de dados coletados do sistema SCADA de um aerogerador em operação comercial com intuito de prever valores de temperatura para o rolamento do eixo principal da multiplicadora, cruzar os resultados obtidos com os diferentes modelos e apresentar a comparação entre os métodos. Concluiu-se que o método com maior assertividade, 10,3% a mais do que a Regressão Linear e 8,96% Regressão Polinomial, foi a Máquina de Vetores Suporte com erro médio de aproximadamente 1,25°C para o aerogerador 1 e aproximadamente 1,36°C para o aerogerador 2 e coeficiente de correlação de 85,6% por cento para o aerogerador 1 e de 88,3% para o aerogerador 2.

Palavras-chave: Energia Eólica. *Machine Learning*. Aprendizagem de Máquinas. Manutenção Preditiva. R. Regressão. *Support Vector Machine*.

Sumário

1. Introdução.....	7
1.1. Justificativa	8
1.2. Objetivos	8
1.2.1. Objetivo Geral	8
1.2.2. Objetivos Específicos	8
2. Fundamentação Teórica	9
2.1. Manutenção.....	9
2.2. Aprendizagem de Máquinas	11
2.2.1 Regressão Linear	13
2.2.2 Regressão Polinomial	13
2.2.3 Máquina de Vetores Suporte (SVM).....	14
2.3. Aerogeradores.....	16
3. Materiais e Métodos.....	18
4. Discussão e Resultados.....	19
4.1. Métodos aplicados.....	19
4.2. Gráficos dos dados – Aerogerador 1	19
4.3. Gráficos dos dados – Aerogerador 2	24
4.4. Gráficos de previsão – Regressão Polinomial – Aerogerador 1	28
4.5. Gráficos de previsão – Regressão Polinomial – Aerogerador 2	33
4.6. Gráficos de previsão – <i>Support Vector Machine</i> – Aerogerador 1	38
4.7. Gráficos de previsão – <i>Support Vector Machine</i> – Aerogerador 2	43
4.8. Gráficos de previsão – Regressão Linear – Aerogerador 1.....	48
4.9. Gráficos de previsão – Regressão Linear – Aerogerador 2.....	53
5. Conclusão	59
6. Trabalhos Futuros	60
Anexo I.....	62
Algoritmo Regressão Linear	62
Algoritmo Regressão Polinomial.....	63
Algoritmo Máquina de Vetores Suporte	64
Anexo II.....	65

1. Introdução

A manutenção surgiu com a necessidade de reparos em ferramentas e se popularizou com a criação de máquinas. Atualmente existem máquinas para praticamente qualquer atividade e estas precisam passar por processos de manutenção para manterem suas funções dentro dos padrões de aceitabilidade operacional.

A princípio o processo de manutenção utilizado era apenas o de manutenção corretiva o qual se dava como um reparo após uma falha, com intuito de manter as funções do objeto reparado. A manutenção evoluiu de forma a prevenir que as falhas acontecessem e recebeu o nome de manutenção preventiva. Atualmente a forma mais moderna de manutenção é a preditiva que consiste em estudar ou monitorar o comportamento de máquinas ou componentes com intuito de prever a ocorrência de falhas para que se possa agir de forma proativa realizando a manutenção ou substituição de componentes antes que estes falhem e sem causar grandes impactos na produtividade (MULLER, 2008).

Os métodos de manutenção preditiva têm se popularizado pela possibilidade de programar manutenções para períodos que impactam de forma diminuta nos resultados produtivos e financeiros. O uso de métodos computacionais vem aumentando por conta da possibilidade de analisar ou monitorar de forma remota utilizando sensoriamento e formas de aquisição de dados que quase sempre já estão presentes nas máquinas (VAPORTEC, 2016).

Nesse contexto de evolução das técnicas de manutenção e da análise de grandes volumes de dados os processos de aprendizagem de máquinas têm se desenvolvido com o objetivo de facilitar conclusões a partir do monitoramento e também realizar previsões quanto a falhas e com isso evitá-las.

Existem vários métodos de aprendizagem de máquinas que podem ser divididos em duas classes principais, métodos descritivos e métodos preditivos. Os métodos descritivos podem ser divididos em métodos de associação, de agrupamento, de detecção de desvios, de padrões sequenciais e de sumarização. Os métodos preditivos podem ser divididos em métodos de classificação e métodos de regressão (GRANATYR, 2017).

Essas técnicas são aplicadas a vários setores da economia, nesse trabalho o foco foi o setor de geração de energia eólica. No setor de energia eólica as paradas de manutenção precisam ser especialmente planejadas pois dependem das condições climáticas como vento, chuva e descargas atmosféricas, regulamentados por norma. Dessa forma o monitoramento do desgaste e a previsão de falhas dos componentes é de extrema importância para a otimização da operação e dos processos de manutenção.

Os métodos utilizados nesse trabalho são de regressão e implementados em R com objetivo de prever valores de temperatura para o rolamento DE (*drive end*, porção motora) do eixo rápido da multiplicadora de um aerogerador em operação.

1.1. Justificativa

A aplicação de métodos computacionais para previsão de comportamentos tem sido foco de várias empresas, universidades e pesquisadores. O mercado de monitoramento, previsão e predição tem crescido bastante e com isso a necessidade por soluções simples e eficientes.

Esse trabalho objetiva a aplicação de conhecimentos adquiridos ao longo do curso de graduação em engenharia mecânica, utilizando principalmente conhecimentos em estatística, programação e manutenção de equipamentos para produzir modelos de aprendizagem de máquinas implementados em R capazes de fornecer previsões com boa assertividade e baixo custo computacional.

1.2. Objetivos

1.2.1. Objetivo Geral

Programar e testar modelos de aprendizagem de máquinas visando a previsão de valores de temperatura de um componente específico de um aerogerador real e realizar a comparação entre os resultados de previsão obtidos por cada um dos modelos entre si e entre os valores reais adquiridos pelo sistema SCADA.

1.2.2. Objetivos Específicos

- Estudo de diferentes algoritmos de aprendizagem de máquinas para previsão de valores.
- Comparação dos resultados obtidos em cada um dos algoritmos estudados.

- Análise de erro absoluto e coeficiente de correlação para cada um dos métodos aplicados.

2. Fundamentação Teórica

2.1. Manutenção

A manutenção tem feito parte do cotidiano das pessoas desde o surgimento das primeiras ferramentas e sua importância e participação tem aumentado com o aumento da complexidade das máquinas e ferramentas utilizadas.

A evolução da manutenção pode ser dividida basicamente em seis fases, a saber: Antes da Revolução industrial, Era a vapor, Corretiva, Preventiva, Racionalização e produtiva total (PALMEIRA, 2001).

A princípio a manutenção tinha como característica apenas a restauração da funcionalidade de uma ferramenta ou máquina, ou seja, a princípio se utilizava determinada máquina ou ferramenta até que perdesse a capacidade de realizar a função para a qual foi projetada, no conceito anterior a revolução industrial quem realizava as manutenções das máquinas e ferramentas era o operador que em muitos casos era o próprio inventor da máquina ou ferramenta. A participação das máquinas na economia e a complexidade das máquinas eram tão pequenas que não havia necessidade de equipes, pessoal especializado ou plano de manutenção (PALMEIRA, 2001).

Com a revolução industrial as máquinas evoluíram e aumentaram sua participação na economia e com isso sua complexidade, com isso surgiu a necessidade de pessoal especializado e procedimentos de manutenção mais eficientes para não impactar socialmente a comunidade que estava diretamente ligada à produção ou consumo dos bens produzidos por essas máquinas (PALMEIRA, 2001).

Com a primeira guerra mundial, entre os anos 1914 e 1918, a importância das máquinas como indicativo de poder das nações foi consolidada e com isso equipes especializadas em máquinas ou componentes surgiram e se capacitaram a realizar manutenções de caráter corretivo de forma direta e eficiente (PALMEIRA, 2001).

Com a segunda guerra mundial, entre os anos 1939 e 1945, o advento dos aviões e a importância quase que determinista dos mesmos no resultado da guerra, o grau de confiabilidade exigido na fabricação de peças e componentes fez com que o

conceito de manutenção preventiva tivesse início e com ela o uso de tecnologias eletrônicas embarcadas capazes de auxiliar na operação, programação e manutenção das máquinas (PALMEIRA, 2001).

Entre 1950 e 1970 as indústrias já representavam a maior parte da economia dos países desenvolvidos e nos países subdesenvolvidos as principais atividades eram ligadas a produção de insumos para as indústrias. Com a escassez de uma das mais importantes matérias primas para os processos industriais durante a crise do petróleo surgiu a necessidade da gestão inteligente dos recursos e maquinário utilizando as primeiras técnicas de monitoramento com foco em economia, provocou o surgimento da engenharia de manutenção visando além da prevenção de falhas, a execução da manutenção com o menor custo total possível (PALMEIRA, 2001).

Devido à globalização houve um aumento da demanda por produtos que mantivessem a qualidade e consistência independentemente do aumento da produção, com isso o surgimento de novas técnicas de monitoramento e de controle da qualidade e desempenho, acompanhadas pelo desenvolvimento tecnológico buscando garantir a maior eficiência produtiva possível utilizando vários conceitos de administração, controle e gerenciamento da produção, como as teorias de Taylor e Ford, os modelos de produção baseados na indústria japonesa Toyota, sistemas de produção enxuta, minimização de estoques e organização dos processos produtivos e de manutenção, a aplicação desse conjunto de conceitos deu-se o nome de manutenção produtiva total.

A importância das máquinas e dos bens de consumo manufaturados cresceu e a evolução tecnológica em implementações para inspeção e monitoramento foi tamanha que a engenharia da manutenção procurou antever possíveis falhas e estimar a vida útil de máquinas e componentes, daí o conceito de manutenção preditiva. No contexto de manutenção preditiva existem várias técnicas aplicáveis além do monitoramento remoto, tais como, análise por ultrassom, ensaio por líquido penetrante, inspeção termográfica, análise de vibração e ruído, para tanto o grau de controle, automação e monitoramento nas máquinas aumentou gerando um volume de informação que tornou a análise manual ineficiente, com isso a implementação das ciências da computação, análises estatísticas e conhecimentos de engenharia aconteceu de forma inevitável e vem trazendo inúmeros benefícios para a indústria de forma geral.

Os métodos de análise de dados têm se desenvolvido e aumentado a capacidade de processamento dados e a assertividade das informações fornecidas, mas o custo computacional para obter esses resultados muitas vezes é proibitivo. A busca agora é por métodos com boa assertividade e confiabilidade e baixo custo computacional.

2.2. Aprendizagem de Máquinas

Os métodos de aprendizagem de máquinas consistem em algoritmos que processam dados de forma a encontrar relações existentes, ou comportamentos previsíveis sendo divididos em métodos descritivos e métodos preditivos. Os métodos descritivos têm por objetivo identificar relações entre variáveis como detecção de padrões, detecção de desvios, agrupamento, associação e sumarização. Já os métodos preditivos têm por objetivo prever algum valor ou atributo e são divididos em classificação e regressão. Tanto os métodos descritivos como os métodos preditivos podem ser aplicados em praticamente qualquer campo de estudo. Os métodos preditivos são o foco desse trabalho e podem ser amplamente utilizados para várias finalidades, desde tarefas simples como a recomendação de produtos similares em compras online até tarefas mais complexas como a previsão de resposta de uma determinada bactéria a um certo tipo de proteína, ou ainda, a previsão do ciclo de vida de um determinado componente em uma máquina (GRANATYR, 2017).

A subdivisão dos métodos de aprendizagem de máquinas do tipo preditivo se dá devido ao tipo das variáveis. Se as variáveis são qualitativas, por exemplo gênero, cor, raça ou marca então o problema é rotulado como classificação, ou seja o algoritmo visa interpretar as características de uma nova entrada e aloca-la no grupo correspondente a tais características. Um exemplo prático de um problema de classificação é o conjunto de carros em uma concessionária, desses carros têm-se os seguintes atributos, cor, tipo e motorização, sendo assim é possível dividir os carros em grupos pela cor, tipo, motorização, cor e tipo, cor e motorização ou motorização e tipo, nesse exemplo se a concessionária adquirisse um carro *hatch* com motor 1.0 e de cor branca ele poderia ser automaticamente alocado em um grupo segundo as características desejadas ou ainda, retirando um carro do conjunto e sabendo que é um *hatch* 1.0 pode-se dizer qual a provável cor do carro baseado nas informações adquiridas do banco de dados. Outro exemplo prático é, dado a idade, o salário, o

histórico de crédito e o nível de endividamento de uma pessoa é possível prever o pagamento ou não pagamento de uma dívida (JAMES, 2013).

Quando as variáveis do problema são do tipo quantitativas, por exemplo altura, peso, potência, torque ou idade o problema é de regressão, ou seja, o algoritmo busca, baseado na análise de um conjunto de dados, prever qual seria o valor de um determinado atributo para uma dada entrada. Um exemplo prático de um problema de regressão é, dado um conjunto de dados de casas com informações sobre área construída, número de quartos, número de pavimentos, número de banheiros, bairro e preço deseja-se prever o preço de um imóvel dado as características do mesmo, nesse exemplo se a casa a qual deseja-se prever o preço tiver dois quartos, dois banheiros, setenta metros quadrados, um pavimento e for localizada no centro é possível para o algoritmo após análise do banco de dados, prever o preço da casa baseado nas relações encontradas entre as variáveis.

Para a maioria dos problemas de aprendizagem de máquinas esses métodos preditivos podem ser divididos em duas grandes categorias, aprendizagem supervisionada e aprendizagem não supervisionada (SMOLA, 2010).

Métodos de aprendizagem supervisionada para previsão consistem em um processo de aprendizagem onde existe um objetivo conhecido e a partir de um conjunto de dados procura-se prever um determinado atributo, ou seja, para cada observação da variável a ser predita (X_i , onde $i = 1, \dots, n$), existe um valor de resposta associado (Y_i) e deseja-se encaixar um modelo que relacione a resposta aos atributos de previsão, doravante denominados preditores, com o objetivo de prever com precisão a resposta para futuras observações, por exemplo, dado um banco de dados de pessoas com informações de salário e anos de estudo é possível prever quanto uma pessoa ganharia se tivesse determinado tempo de estudo ou quantos anos uma pessoa deveria estudar para ter um certo salário. Nesse caso é possível aplicarmos um modelo de regressão linear.

Outra aplicação de aprendizagem supervisionada chamada de inferência, é utilizada para entender a relação entre a resposta e os preditores, por exemplo, coletado um conjunto de dados das 500 maiores empresas dos EUA. Para cada empresa registra-se o lucro, número de funcionários, setor e o salário do CEO. O interesse é entender quais fatores afetam o salário do CEO (JAMES, 2013).

Nos métodos de aprendizagem não supervisionada a situação é mais complexa, uma vez que para cada observação de um atributo $i = 1, \dots, n$, observamos um vetor de valores X_i mas nenhum valor de resposta associado Y_i . Sendo assim é impossível aplicar um modelo de regressão linear, uma vez que não existe uma variável de resposta a ser prevista e por isso chamamos de não supervisionada, por não termos uma variável de resposta para comparar os resultados das análises. Embora a regressão linear não possa ser aplicada existem ferramentas que podemos utilizar para analisar esse tipo de problema. Um exemplo de ferramenta é a análise de cluster ou *clustering* (JAMES, 2013).

2.2.1 Regressão Linear

A regressão linear é uma abordagem simples de aprendizagem supervisionada que busca prever uma resposta quantitativa baseada em um único atributo preditor assumindo que existe uma relação praticamente linear entre eles. Dessa forma, baseado na equação matemática de uma reta proposta pelo matemático francês Pierre de Fermat no livro de geometria analítica intitulado *Introdução aos lugares geométricos planos e sólidos* que foi publicado em 1629 (BALL, 1960) podemos escrever a seguinte equação: $Y \approx \beta_0 + \beta_1 \cdot X$. Onde Y é o valor a ser previsto, X é o atributo preditor e β_0 e β_1 são coeficientes de interceptação e inclinação da reta em relação aos eixos (JAMES, 2013).

Se o problema de regressão for com múltiplas variáveis, ou seja, os dados analisados são caracterizados por mais de uma variável que determina o comportamento e a distribuição dos valores, o lugar geométrico para a previsão deixa de ser uma reta e passa a ser um conjunto de retas contidas em um hiperplano.

2.2.2 Regressão Polinomial

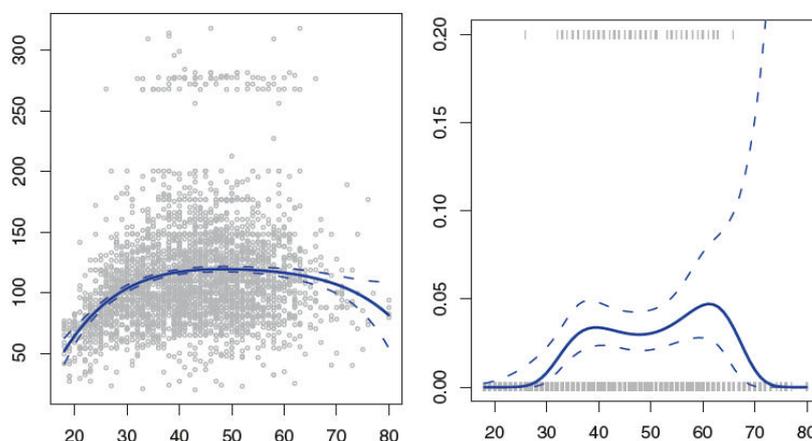
A regressão polinomial difere da regressão linear pela substituição de um modelo baseado numa função linear, do tipo $Y \approx \beta_0 + \beta_1 \cdot X$, por um modelo baseado numa função polinomial, do tipo $Y \approx \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X^2 + \beta_n \cdot X^n$.

O método de regressão polinomial apresenta como vantagem sobre o modelo de regressão linear a possibilidade de adaptação a perfis não lineares de distribuição de dados. A adaptação do modelo polinomial pode ser ajustada mudando o grau do

polinômio para que a curva gerada pela função se aproxime da curva de distribuição dos dados.

A Figura 1 apresenta a direita um gráfico de dispersão, pontos em cinza, com uma curva produzida por um polinômio de quarto grau, linha sólida, e o intervalo de 95% de confiança representado pelas linhas pontilhadas. O gráfico da esquerda representa um polinômio de quarta ordem aplicado aos mesmos dados do gráfico da direita porem com um filtro para valores acima de 250. Novamente a linha sólida é a curva gerada pela função polinomial e a linha pontilhada o intervalo de confiança de 95%.

Figura 1 – Gráficos de dispersão com uma curva polinomial



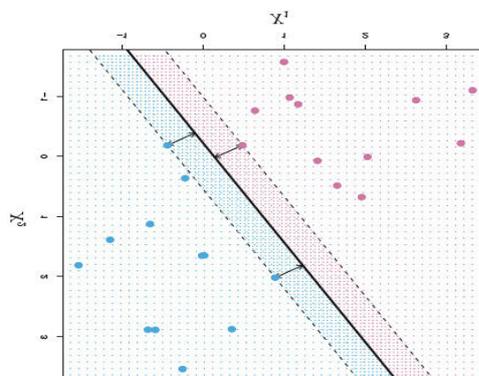
Fonte: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.

2.2.3 Máquina de Vetores Suporte (SVM)

O algoritmo SVM é um algoritmo de classificação e regressão derivado de um algoritmo chamado classificador de margem máxima que tem por objetivo traçar hiperplanos de separação binária de observações com a margem máxima em base de dados. O que diferencia o SVM do seu predecessor é o fato de o SVM ter a capacidade de criar variáveis de suporte aumentando a dimensão do problema e assim traçar hiperplanos não lineares em um espaço vetorial maior do que o espaço vetorial de origem do problema separando as várias classes ou determinando valores utilizando núcleos de classificação ou regressão, sendo capaz de resolver problemas não binários.

A Figura 2 mostra um conjunto de dados onde foi aplicado o classificador de margem máxima, a linha mais escura demarca o hiperplano de separação das duas classes, as linhas pontilhadas demarcam a área de decisão do classificador, os pontos sobre as linhas pontilhadas são vetores suporte e a distância entre eles e o hiperplano de separação é a margem máxima.

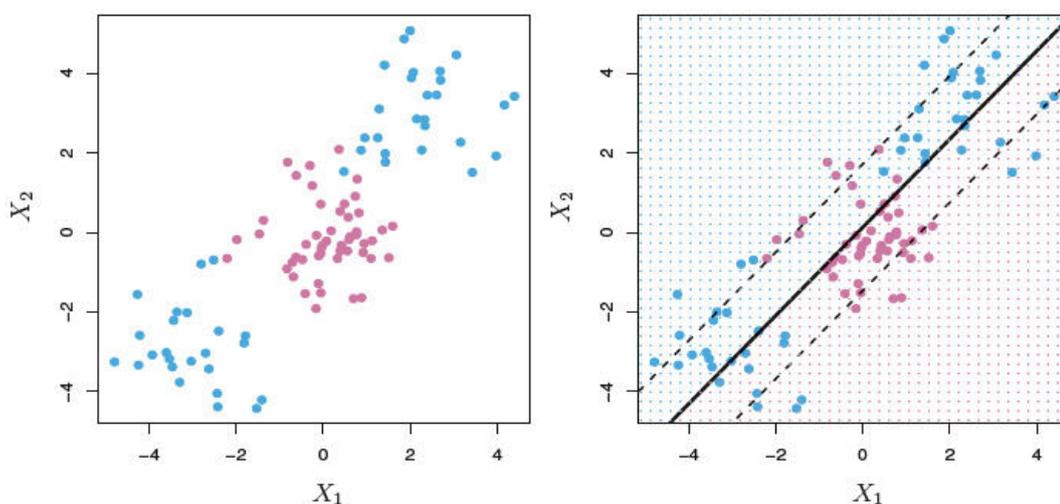
Figura 2 – Gráfico de dispersão (margem máxima)



Fonte: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.

A Figura 3 representa um problema onde a solução linear não pode ser aplicada pois não existe hiperplano de separação que divida a base de dados em duas classes distintas. O gráfico da esquerda ilustra a distribuição dos dados e o gráfico da direita ilustra o desempenho insuficiente do classificador de margem máxima nesse tipo de problema.

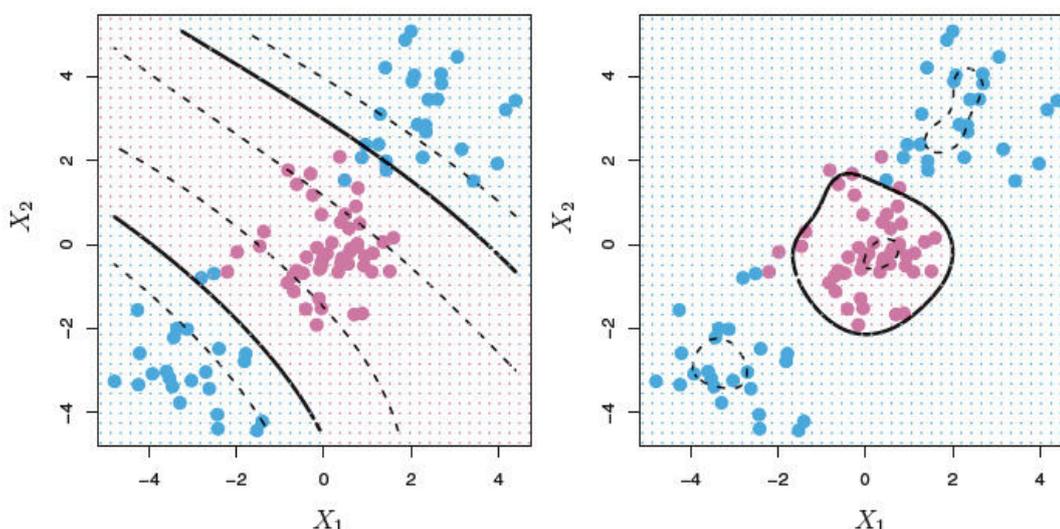
Figura 3 – Gráfico de dispersão (margem máxima problema não linear)



Fonte: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.

A Figura 4 mostra o desempenho do método SVM na classificação do mesmo problema não linear apresentado na Figura 3. No gráfico da esquerda o núcleo de separação utilizado é do tipo polinomial de grau 3, já no gráfico da direita, o núcleo de separação é do tipo radial. Nesse exemplo ambos os núcleos têm bons resultados na definição das fronteiras entre as classes comprovando a maior eficiência do SVM em relação ao seu predecessor.

Figura 4 – Gráfico de dispersão (SVM problema não linear)

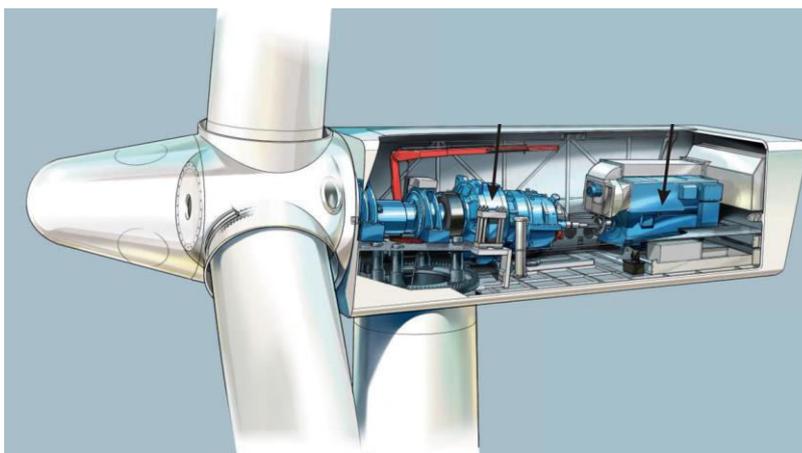


Fonte: Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013.

2.3. Aerogeradores

Aerogeradores são máquinas rotativas que utilizam o escoamento do ar (vento) por meio de hélices com perfis aerodinâmicos, para a geração de energia elétrica. A conversão de energia de rotação (cinética) em energia elétrica é feita por um gerador eletromagnético que possui um campo magnético, com a rotação do eixo do aerogerador as cargas são induzidas ao movimento e com isso geram corrente elétrica. (SAMPAIO, 2015). Como normalmente os aerogeradores tem dimensões grandes e giram a velocidades relativamente baixas, mas com torque elevado, é necessário que seja feito um aumento da rotação em detrimento do torque, usualmente essa elevação de rotação é feita por uma associação de engrenagens chamada de *gearbox* ou multiplicadora. A Figura 5 mostra a Nacele de um aerogerador e a disposição de alguns dos componentes principais dando destaque para a multiplicadora e gerador.

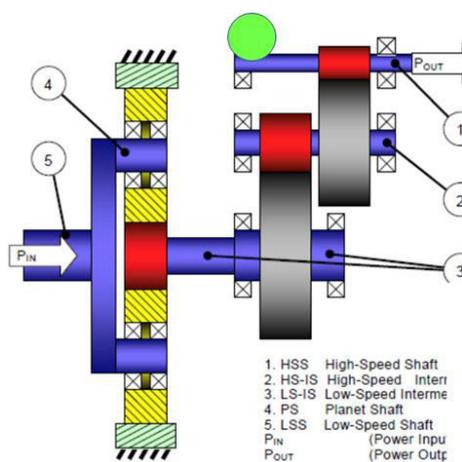
Figura 5 – Desenho representativo de uma Nacele



Fonte: Delfos Intelligent Maintenance, 2019.

A Figura 6 mostra a representação esquemática de uma multiplicadora e a posição do rolamento DE, objeto de estudo, marcado com um círculo verde.

Figura 6 – Desenho representativo de uma multiplicadora



Fonte: Delfos Intelligent Maintenance, 2019.

3. Materiais e Métodos

Existem várias linguagens de programação com bom desempenho estatístico, contudo para a implementação dos métodos analisados nesse trabalho foi escolhida a linguagem de programação R por ser uma linguagem de alto nível, com bom desempenho estatístico, possuir um repositório com várias aplicações, várias bibliotecas e funções que facilitam a programação, análise e comparação dos métodos estudados.

Como ambiente de programação e desenvolvimento foi escolhida a interface de desenvolvimento R Studio por ser de fonte aberta, robusta e simples de utilizar.

Os métodos escolhidos para estudo nesse trabalho foram a regressão linear, a regressão polinomial e a máquina de vetores suporte.

Os códigos dos algoritmos foram todos programados em R para esse trabalho e estão disponíveis no Anexo I.

O computador utilizado foi um laptop da marca Dell modelo 14N4050 com processador Intel core I5 2410M de 2.4 GHz e 12 GB de memória RAM. Dado o baixo custo computacional dos modelos implementados o processamento dos algoritmos levou apenas alguns segundos mesmo para um elevado número de valores em variáveis diversas.

Os dados utilizados foram cedidos por uma empresa de geração de energia que decidiu permanecer anônima, bem como optou por não divulgar a localização do complexo de geração nem a marca e modelo das turbinas. Todos os dados foram coletados pelo sistema SCADA das máquinas e gravados em arquivos de extensão CSV (*Comma Separated Values*) que é um formato de arquivo de texto usado para trocar dados de planilhas entre aplicativos. Os arquivos de dados foram abertos no R Studio utilizando a função de leitura de CSV do R.

Todos os valores faltantes foram removidos da base de dados utilizando lógica de programação para que não influenciassem na distribuição dos dados, outra forma de tratamento de valores faltantes seria a substituição dos mesmos pela média dos dados para cada variável, mas a utilização da média geraria pontos fora do comportamento natural da distribuição gráfica de valores e por isso não foi utilizada.

4. Discussão e Resultados

Os modelos implementados nesse trabalho têm como objetivo prever valores de temperatura do rolamento do eixo rápido da multiplicadora, ou rolamento DE do HSS, com base nos valores de temperatura ambiente, temperatura da multiplicadora e temperatura do eixo principal.

Os métodos de regressão foram escolhidos com base no bom desempenho, simplicidade estatística e de programação e baixo custo computacional.

4.1. Métodos aplicados

O método mais simples aplicado foi o de regressão linear. Esse método consiste em procurar uma equação para um hiperplano de separação que se aproxime do comportamento da distribuição dos pontos da base de dados.

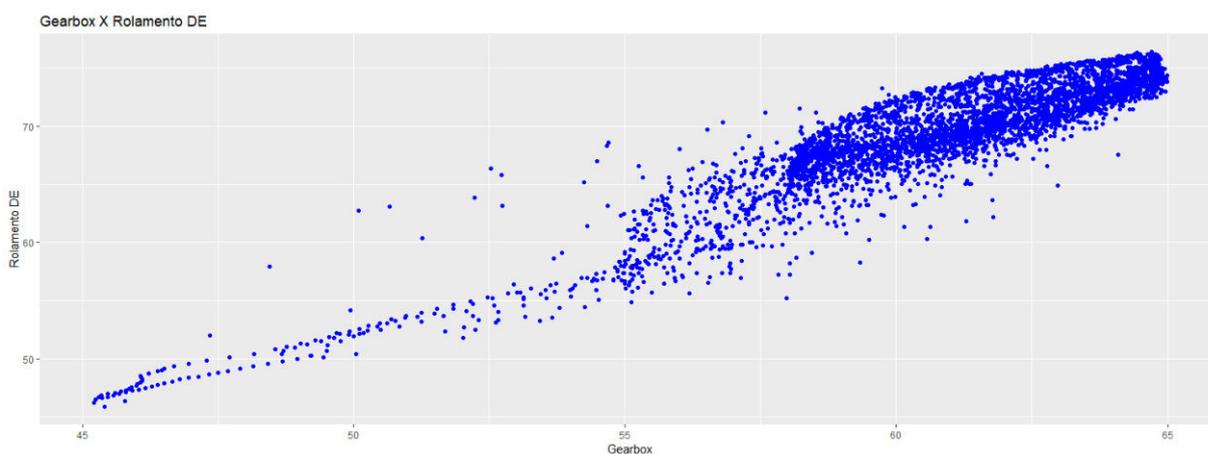
O método de regressão polinomial consiste em procurar uma curva gerada por uma função polinomial que se adapte a distribuição dos dados a serem previstos.

O método de máquinas de vetores de suporte consiste na determinação de hiperplanos de separação com a margem máxima otimizada.

4.2. Gráficos dos dados – Aerogerador 1

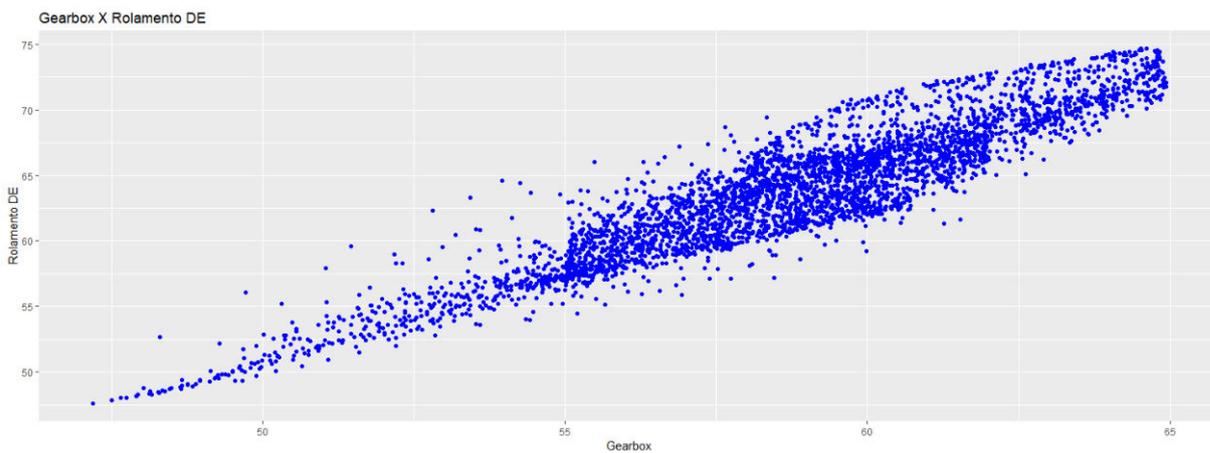
Os gráficos de distribuição apresentado nas Figura 7 a 17 mostram a distribuição dos valores de temperatura do rolamento da multiplicadora em função dos dados de temperatura da multiplicadora do aerogerador 1 para os meses de janeiro a novembro.

Figura 7 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Janeiro).



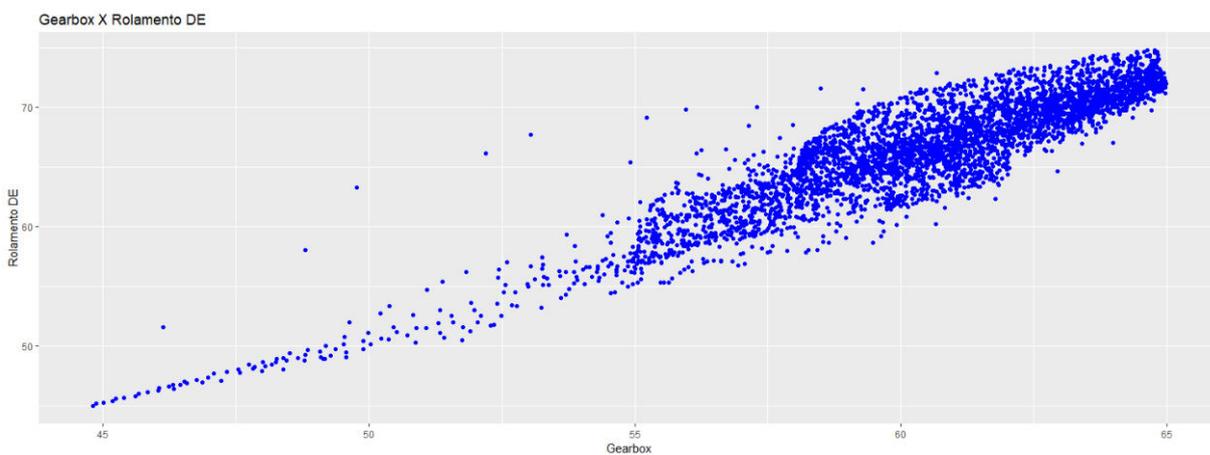
Fonte: Elaborada pelo autor.

Figura 8 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Fevereiro).



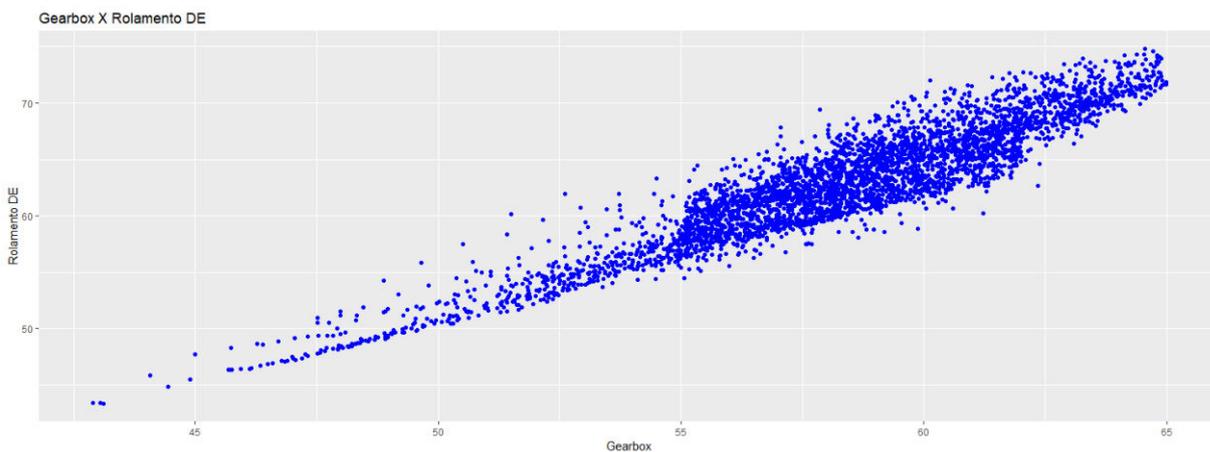
Fonte: Elaborada pelo autor

Figura 9 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Março).



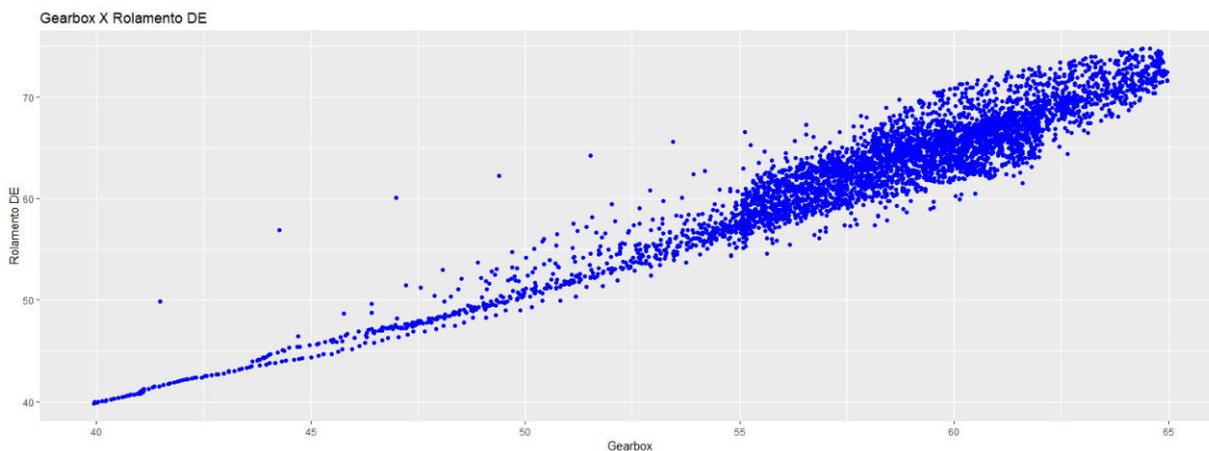
Fonte: Elaborada pelo autor

Figura 10 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Abril).



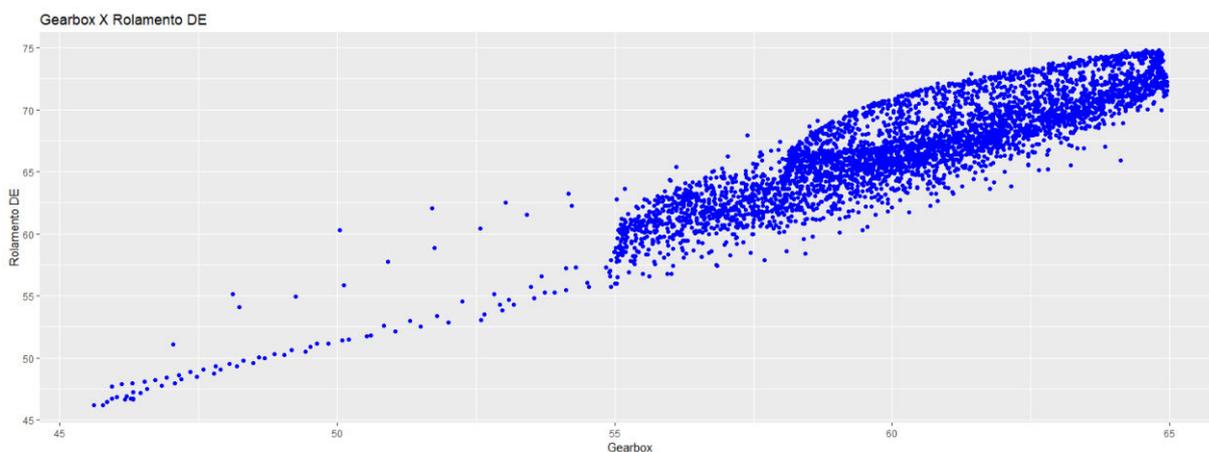
Fonte: Elaborada pelo autor

Figura 11 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Maio).



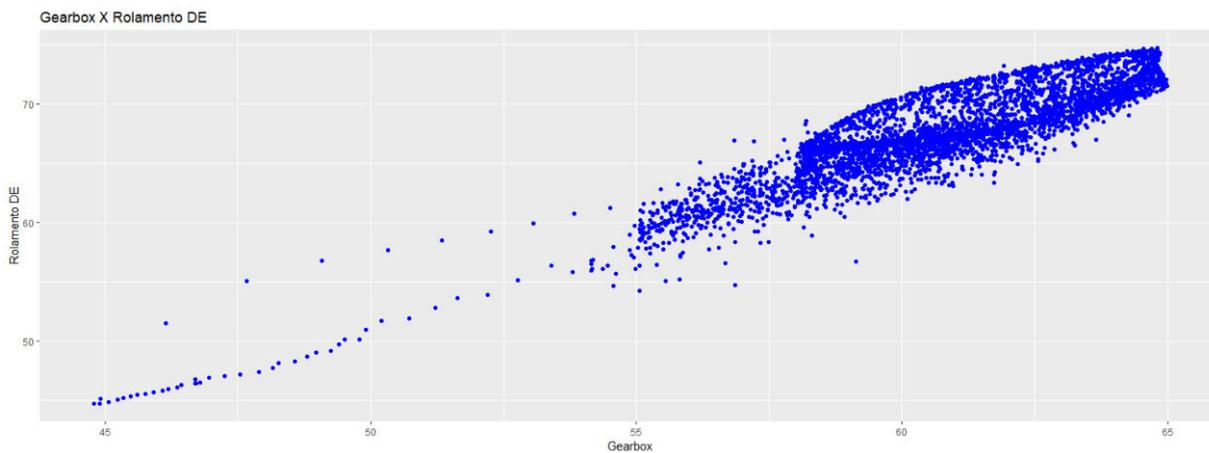
Fonte: Elaborada pelo autor

Figura 12 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Junho).



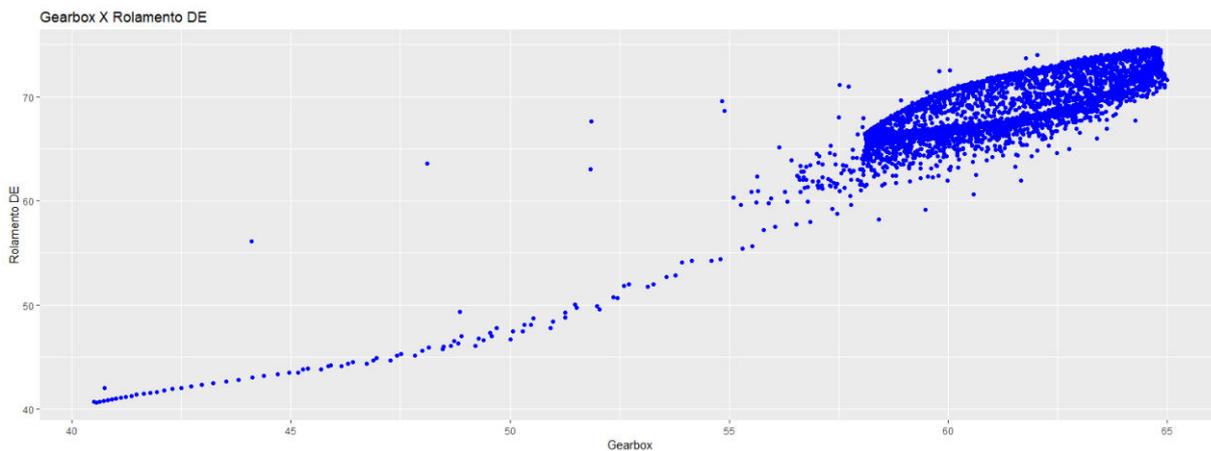
Fonte: Elaborada pelo autor

Figura 13 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Julho).



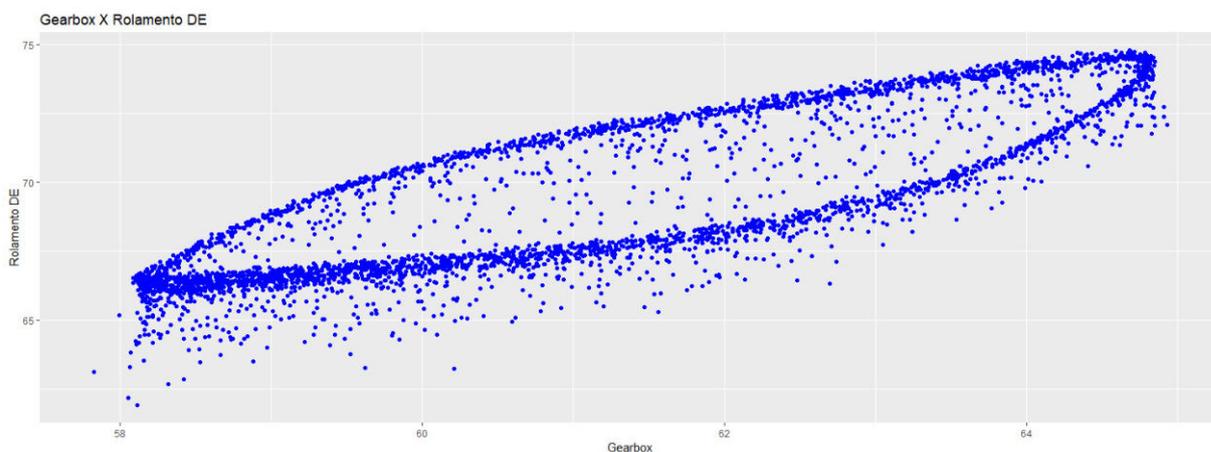
Fonte: Elaborada pelo autor

Figura 14 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Agosto).



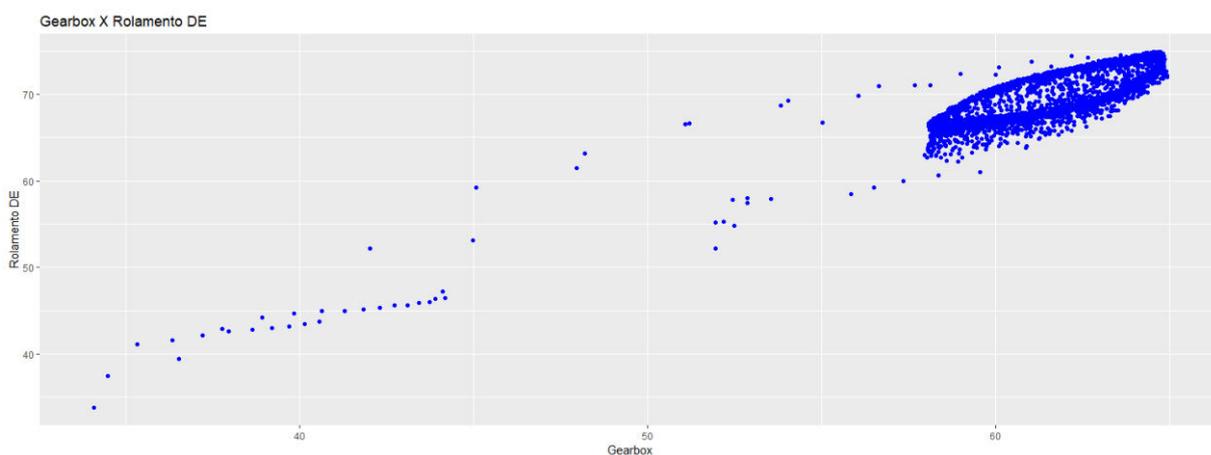
Fonte: Elaborada pelo autor

Figura 15 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Setembro).



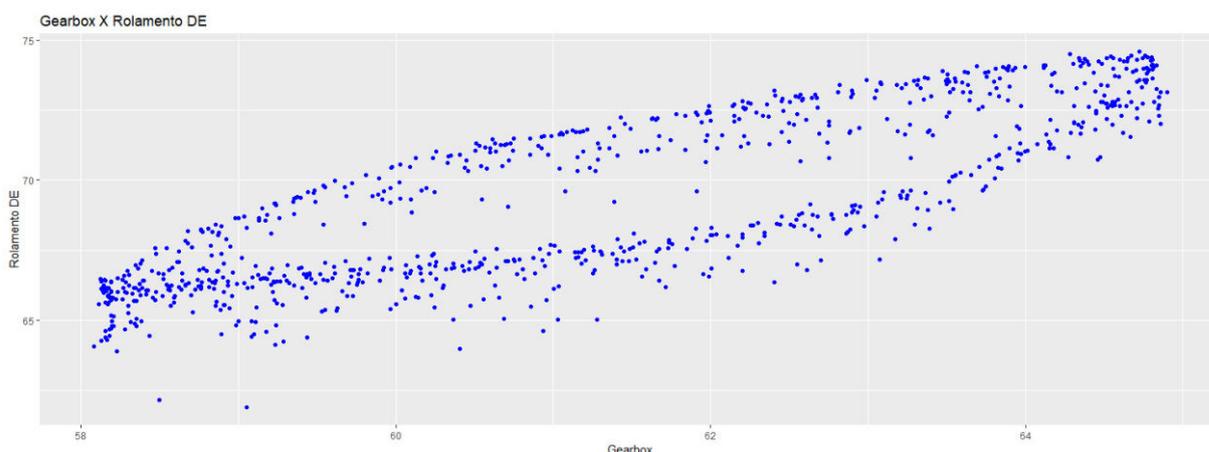
Fonte: Elaborada pelo autor

Figura 16 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Outubro).



Fonte: Elaborada pelo autor

Figura 17 – Gráfico de dispersão - temperatura rolamento1 x multiplicadora1 (Novembro).



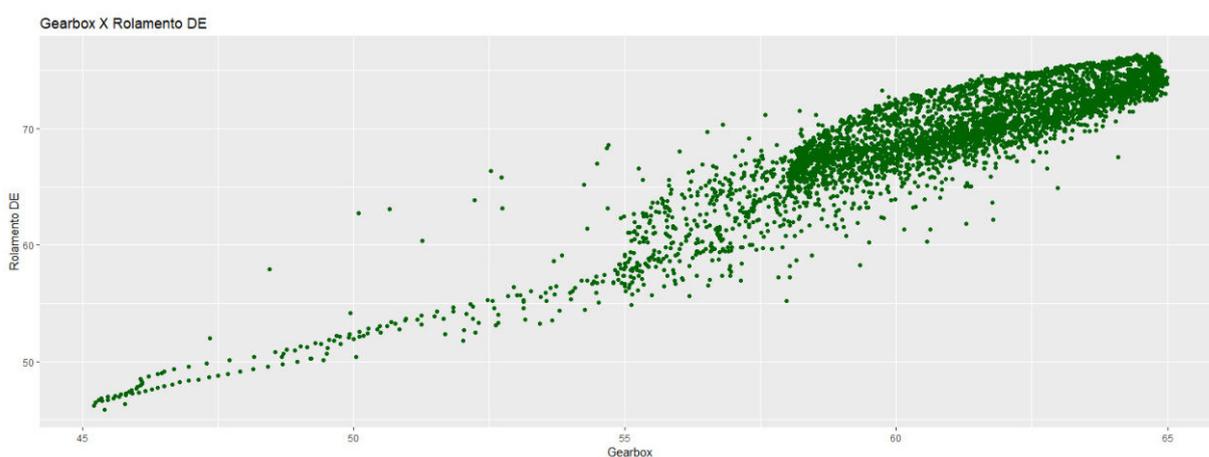
Fonte: Elaborada pelo autor

Os gráficos de janeiro a agosto tem um comportamento similar enquanto os dados de setembro e novembro apresentam um comportamento diferente pois o regime de vento nesses meses é diferente fazendo com que a máquina funcione em altas rotações o que mantém a temperatura dos componentes rotativos mais alta, dessa forma observamos apenas o comportamento para temperaturas na faixa de cinquenta e oito a sessenta e cinco graus célsius que se assemelham a uma curva de histerese o que faz sentido por se tratar da variação de temperatura de massas diferentes, cerca de vinte toneladas para a *multiplicadora* e duzentos quilos para o rolamento. Por se tratarem de dados de temperatura de duas massas tão diferentes a inércia térmica pode provocar esse tipo de comportamento.

4.3. Gráficos dos dados – Aerogerador 2

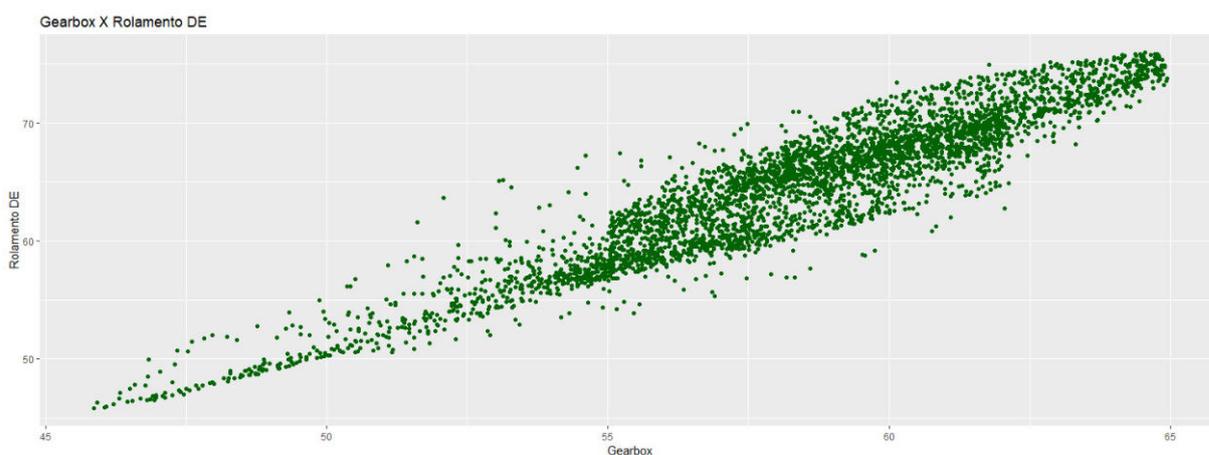
Os gráficos de distribuição apresentados nas Figura 18 a 28 mostram a distribuição dos valores de temperatura do rolamento da *multiplicadora* em função dos dados de temperatura da *multiplicadora* do aerogerador 2 para os meses de janeiro a novembro.

Figura 18 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Janeiro).



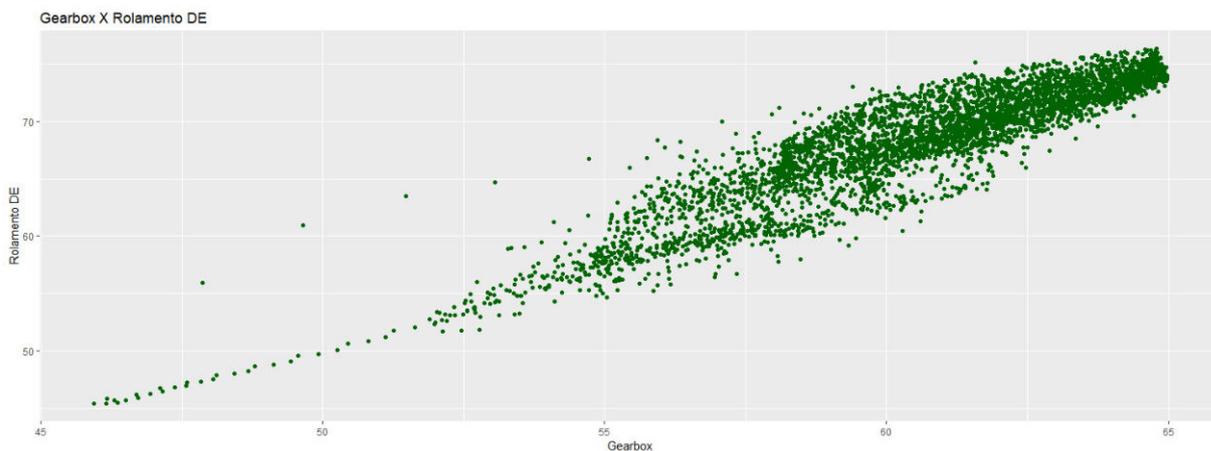
Fonte: Elaborada pelo autor.

Figura 19 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Fevereiro).



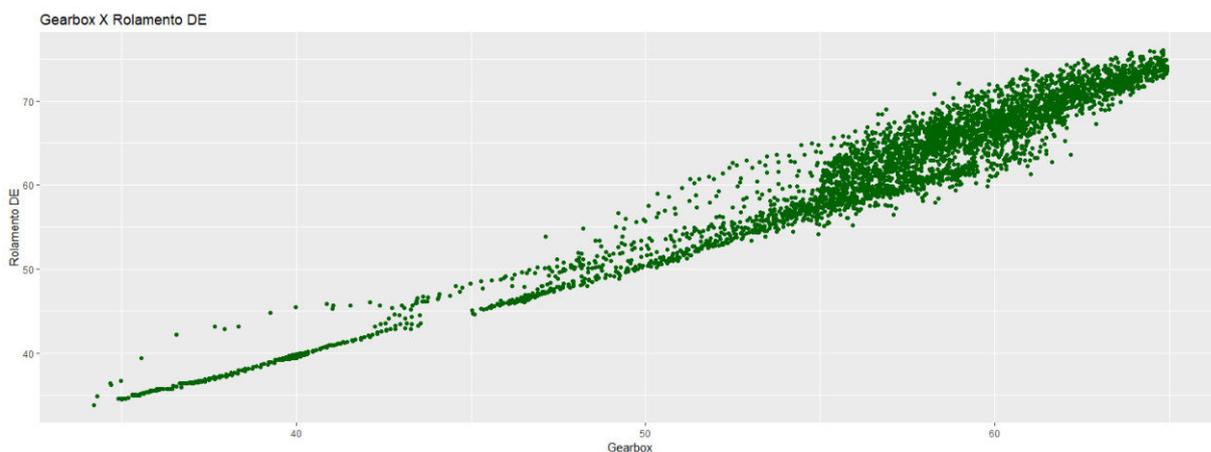
Fonte: Elaborada pelo autor

Figura 20 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Março).



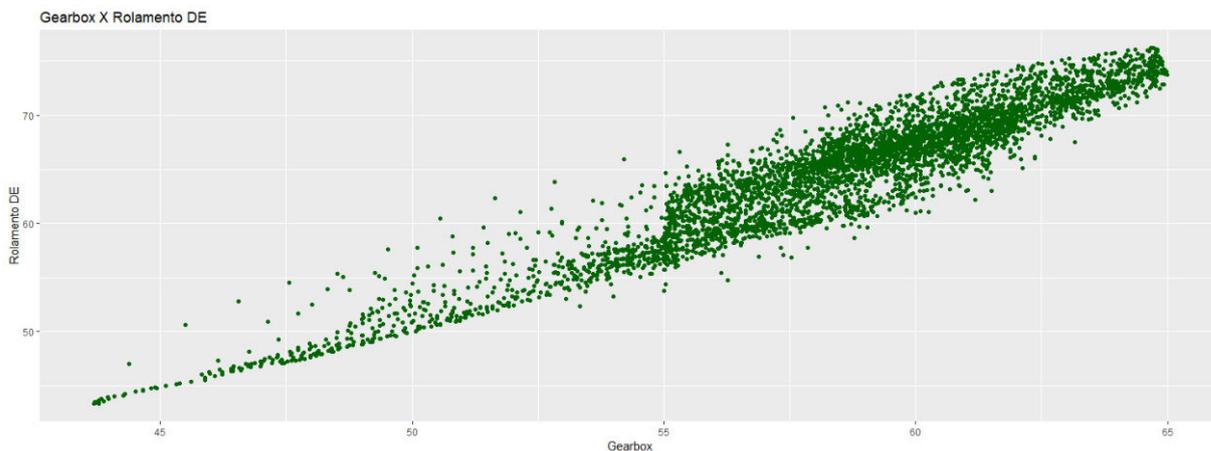
Fonte: Elaborada pelo autor

Figura 21 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Abril).



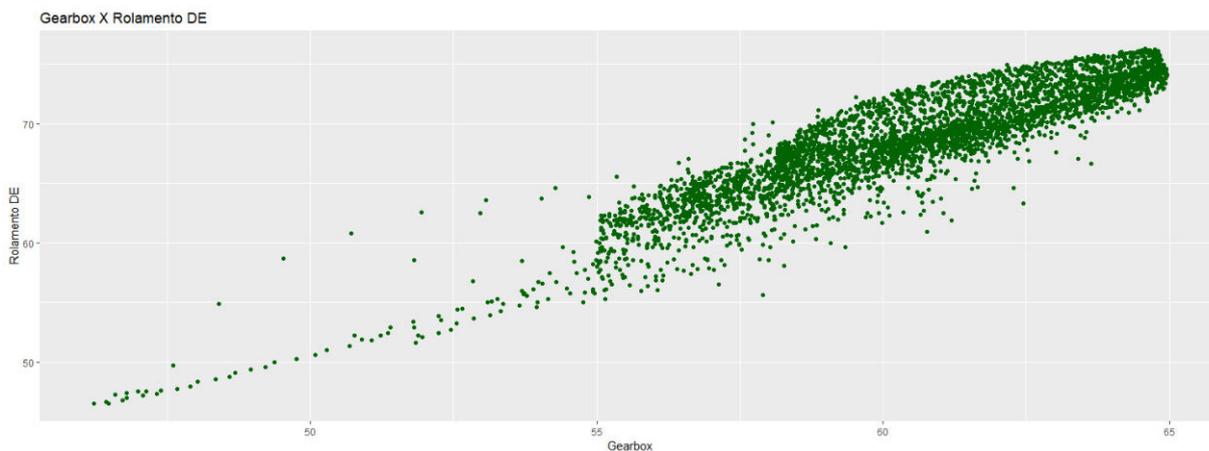
Fonte: Elaborada pelo autor

Figura 22 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Maio).



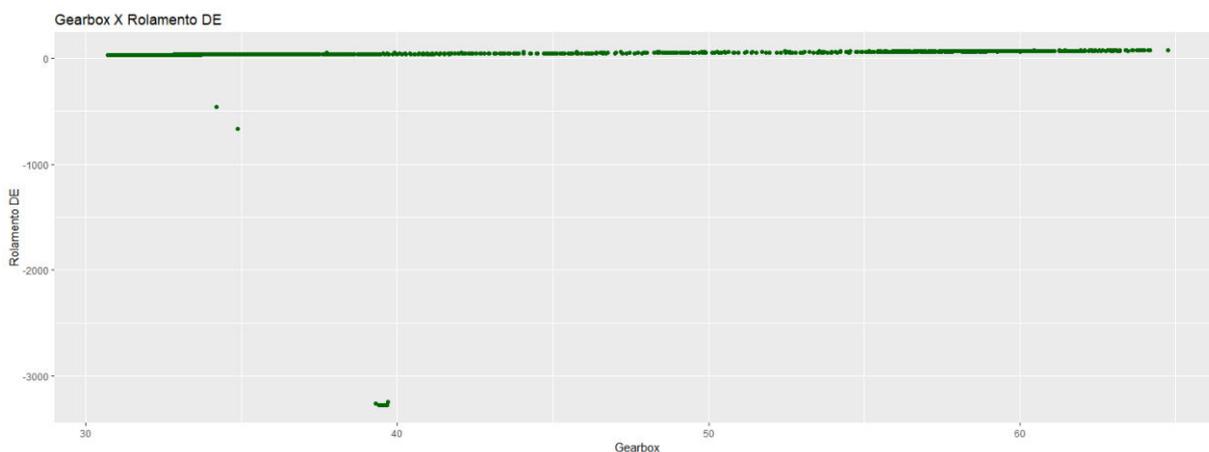
Fonte: Elaborada pelo autor

Figura 23 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Junho).



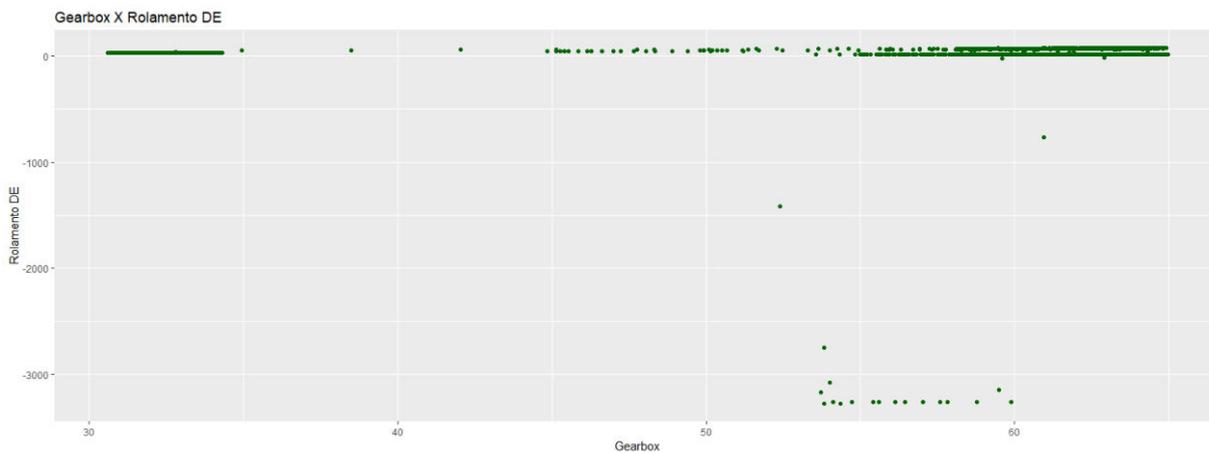
Fonte: Elaborada pelo autor

Figura 24 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Julho).



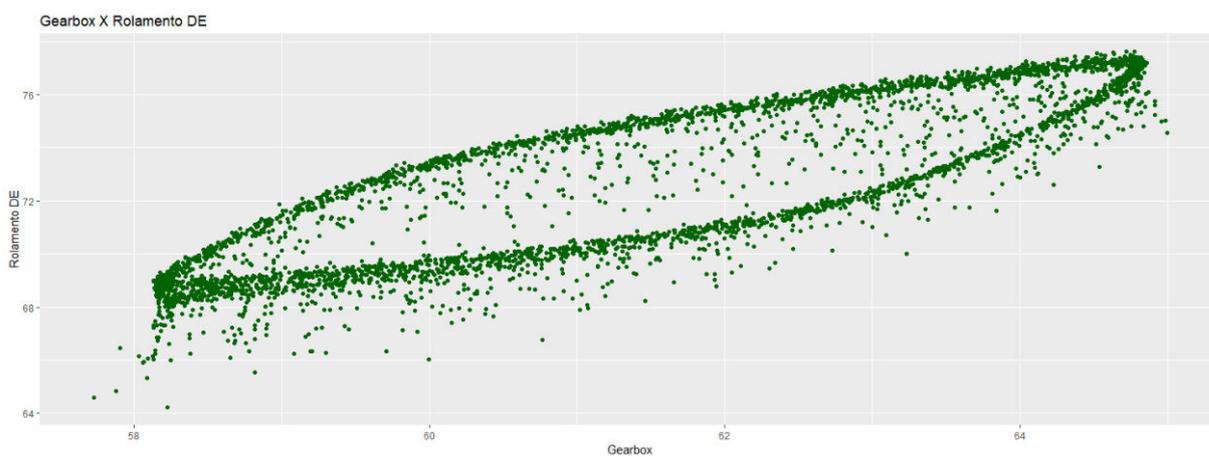
Fonte: Elaborada pelo autor

Figura 25 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Agosto).



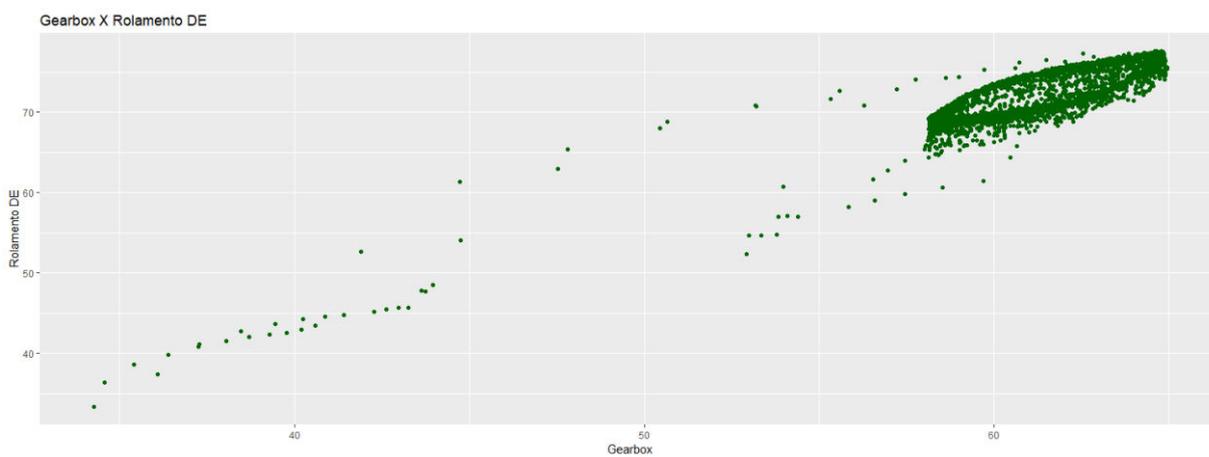
Fonte: Elaborada pelo autor

Figura 26 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Setembro).



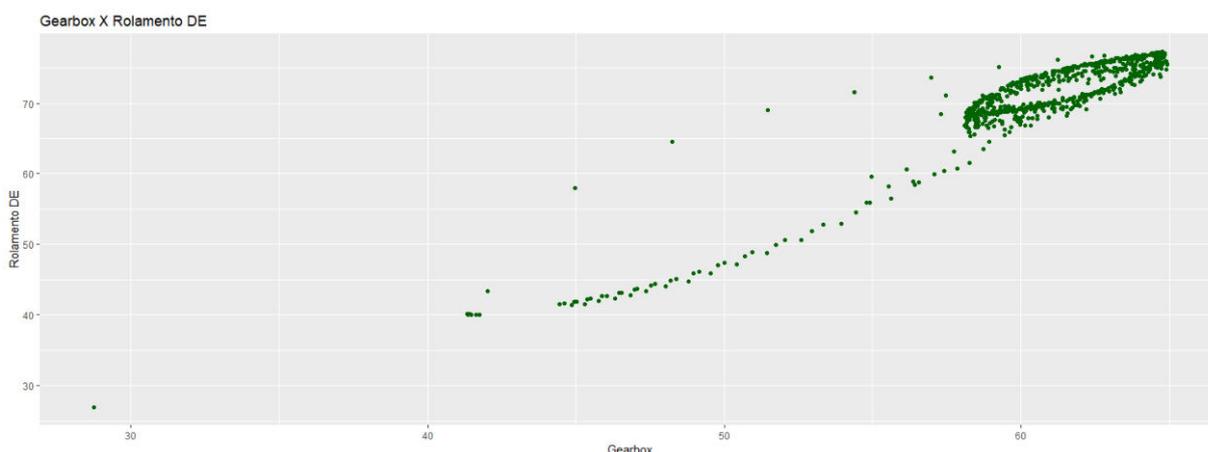
Fonte: Elaborada pelo autor

Figura 27 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Outubro).



Fonte: Elaborada pelo autor

Figura 28 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Novembro).



Fonte: Elaborada pelo autor

Para os meses de julho e agosto os dados de temperatura do rolamento da *multiplicadora* mostram que houve falha nos sensores registrando temperaturas da ordem de três mil graus célsius negativos. O mês de setembro não apresentou temperaturas abaixo dos cinquenta e sete graus célsius e isso pode ser justificado pelo regime de vento, alta velocidade do vento e a constância do vento nesse mês fez com que a temperatura dos componentes rotativos se mantivesse alta durante todo o mês. No mês de novembro observou-se um ponto de temperatura baixa para os componentes, abaixo dos trinta graus célsius, próximo da temperatura ambiente, esse ponto pode ser explicado como um provável desligamento da máquina com travamento do eixo rápido (HSS). Nos demais meses o comportamento é próximo do esperado para um ano típico.

4.4. Gráficos de previsão – Regressão Polinomial – Aerogerador 1

Os gráficos mostrados abaixo, Figura 29 a Figura 39, trazem os dados do sistema SCADA e os dados previstos pelo modelo de regressão polinomial utilizando um polinômio de sexta ordem. Em azul são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 1. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 1.

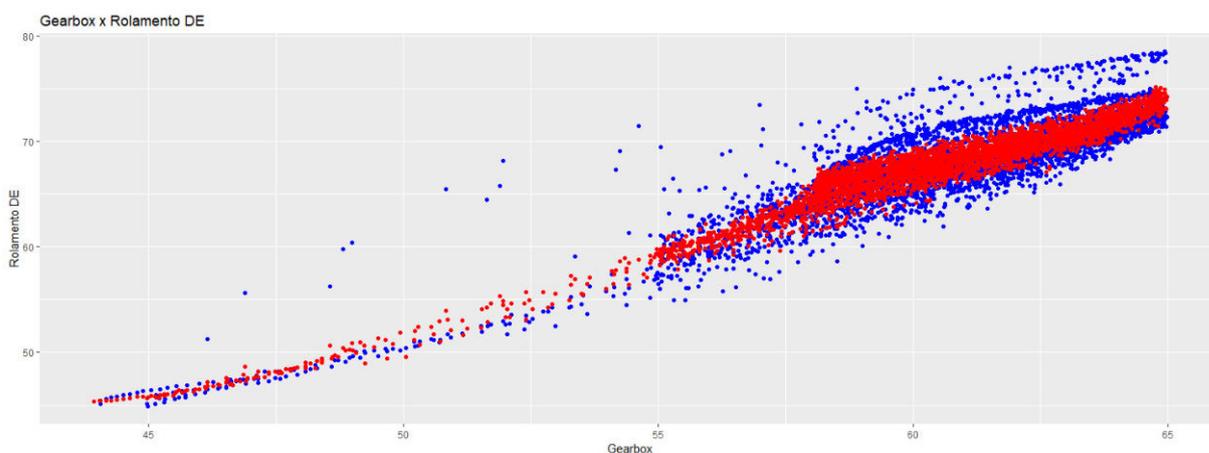
Tabela 1 – Dados de correlação e erro – Regressão Polinomial – Aerogerador 1

Regressão Polinomial - AEROGERADOR 1		
Mês	Coefficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,848	1,534
Fevereiro	0,899	1,271
Março	0,880	1,240
Abril	0,900	1,267
Maio	0,927	1,339
Junho	0,851	1,380
Julho	0,849	1,365
Agosto	0,849	1,496
Setembro	0,692	1,303
Outubro	0,788	1,423
Novembro	0,689	1,618
Média	0,849	1,365

Fonte: Elaborada pelo autor

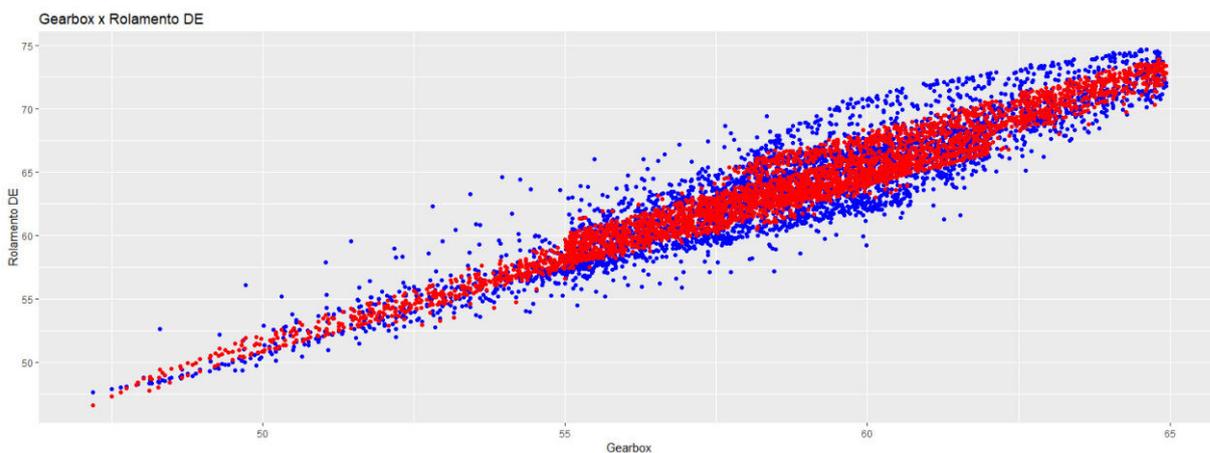
Pela Tabela 1 é possível observar que o coeficiente de correlação foi sempre superior a sessenta e oito por cento e o erro médio absoluto inferior a 1,62 graus célsius.

Figura 29 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Janeiro).



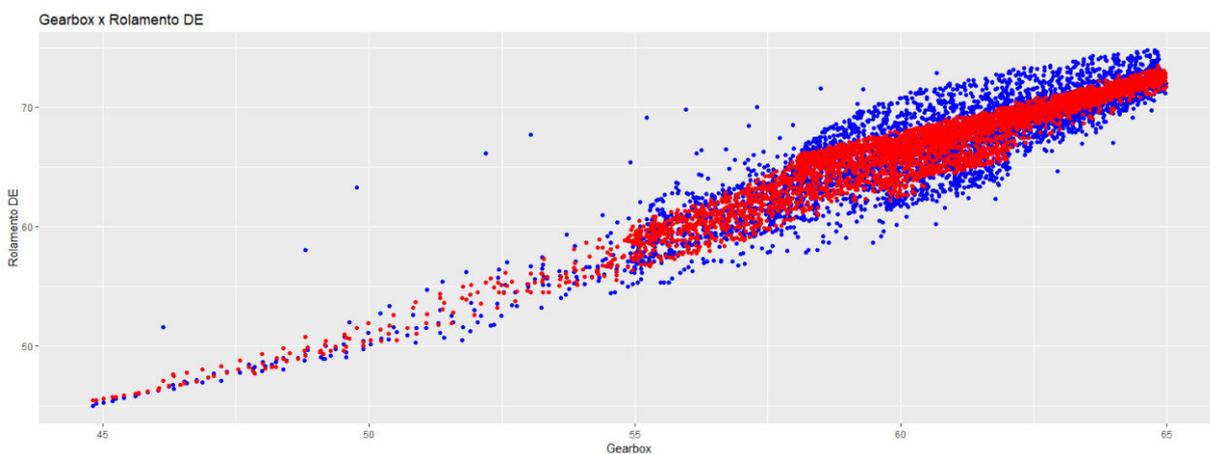
Fonte: Elaborada pelo autor

Figura 30 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Fevereiro).



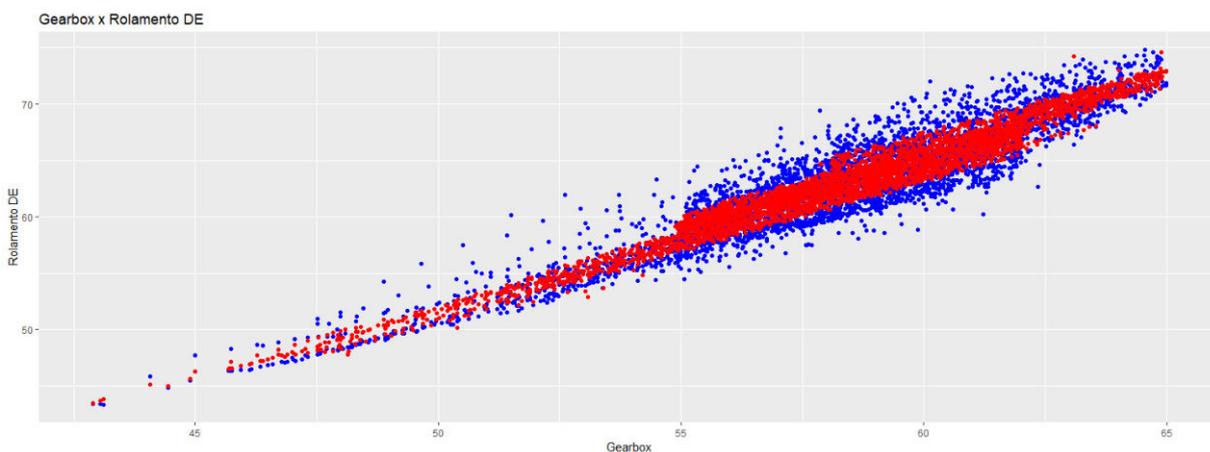
Fonte: Elaborada pelo autor

Figura 31 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Março).



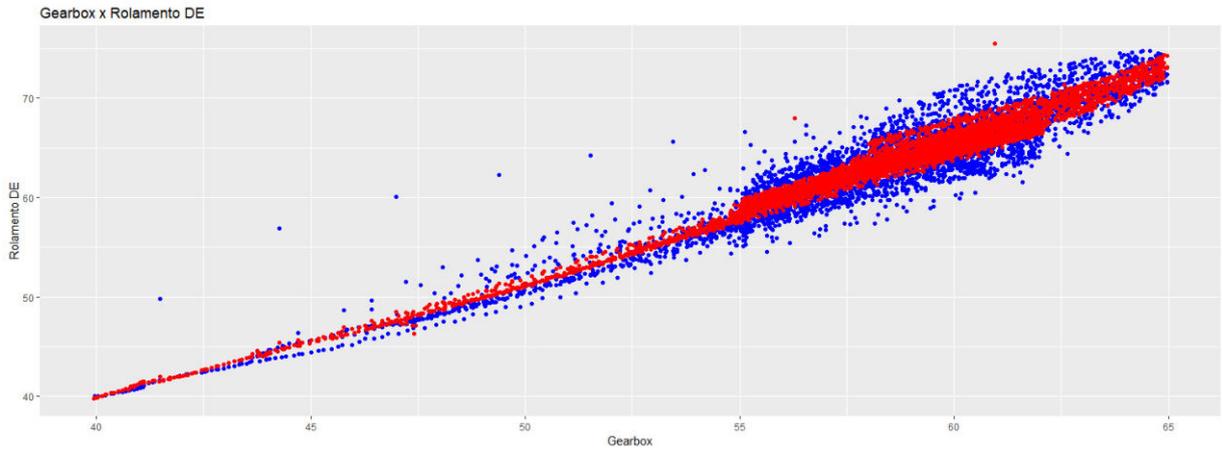
Fonte: Elaborada pelo autor

Figura 32 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Abril).



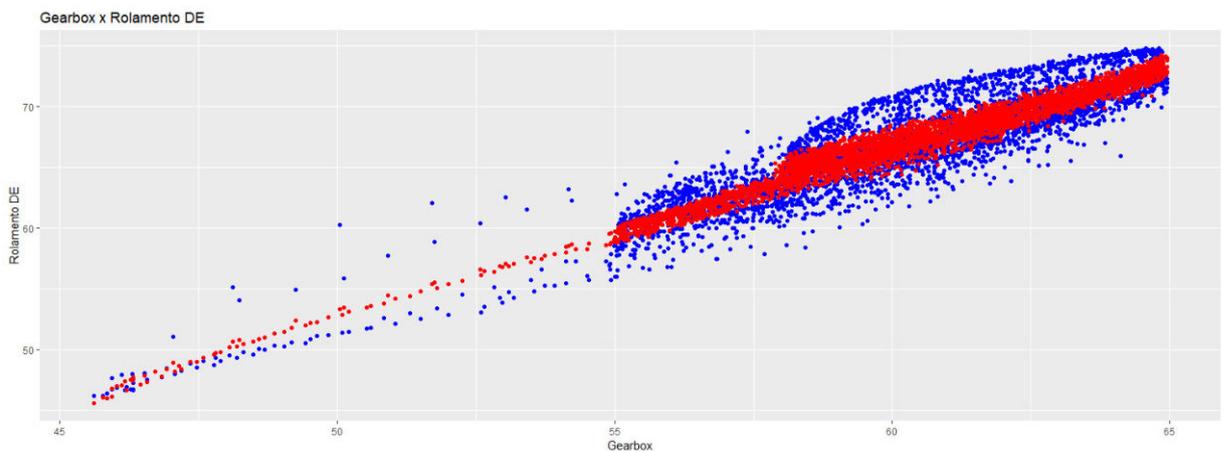
Fonte: Elaborada pelo autor

Figura 33 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Maio).



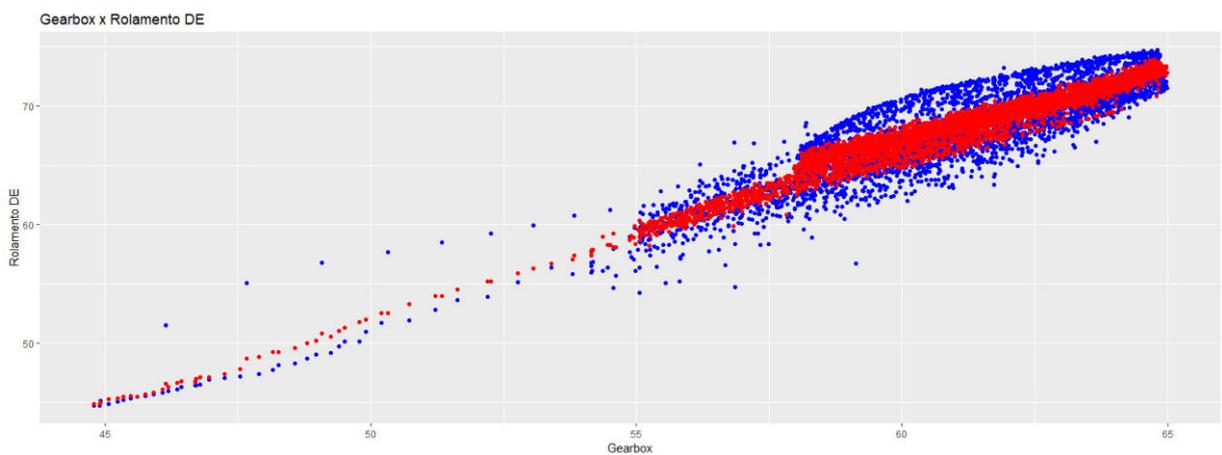
Fonte: Elaborada pelo autor

Figura 34 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Junho).



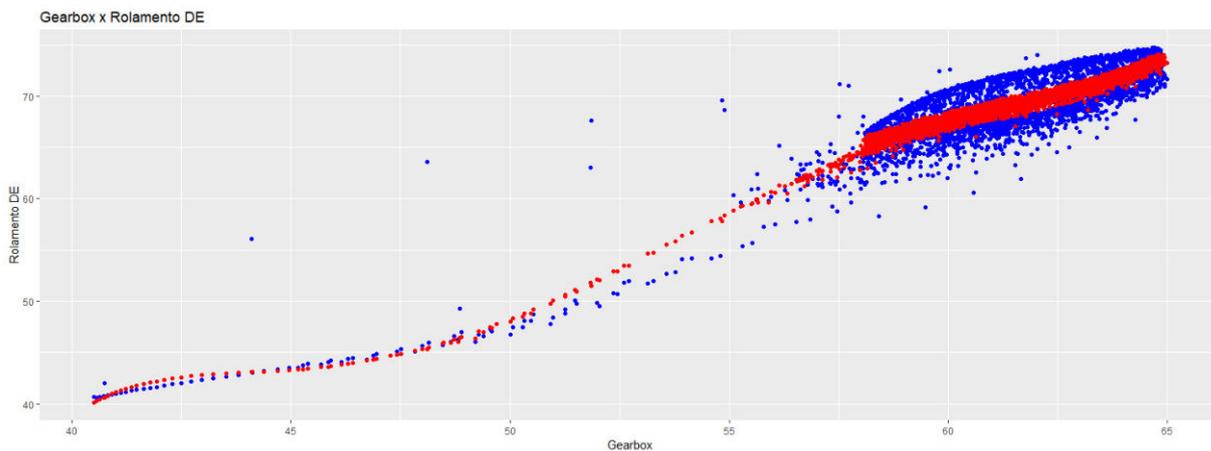
Fonte: Elaborada pelo autor

Figura 35 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Julho).



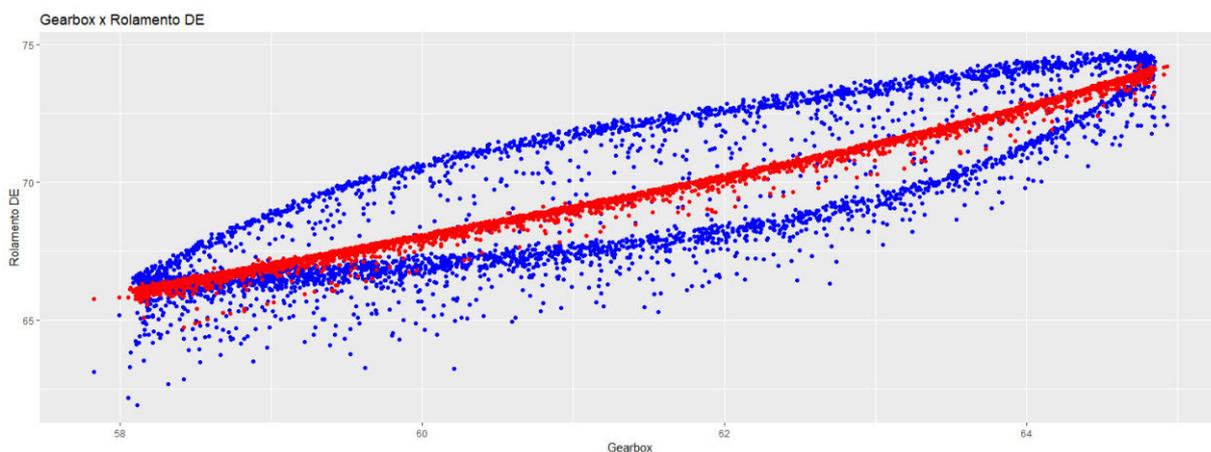
Fonte: Elaborada pelo autor

Figura 36 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Agosto).



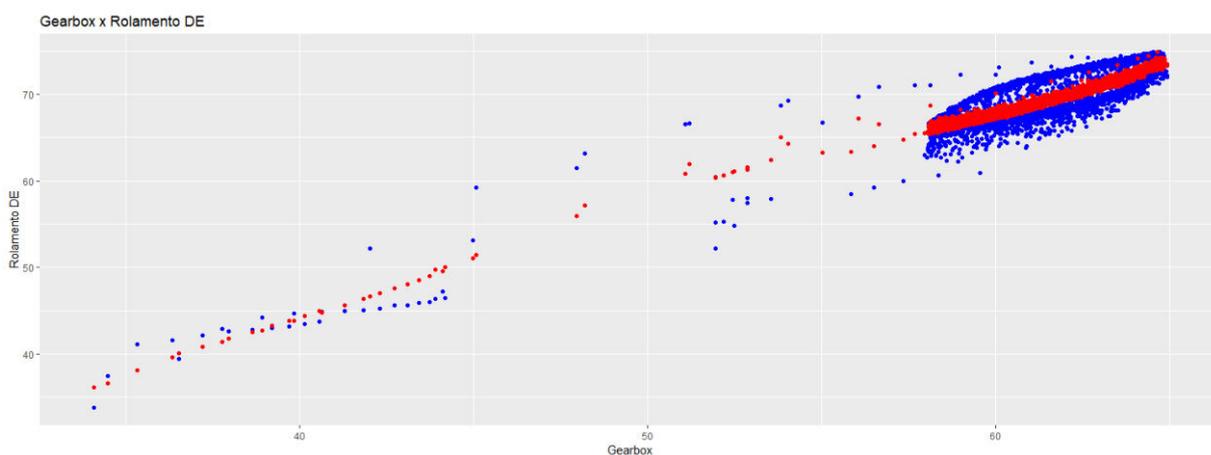
Fonte: Elaborada pelo autor

Figura 37 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Setembro).



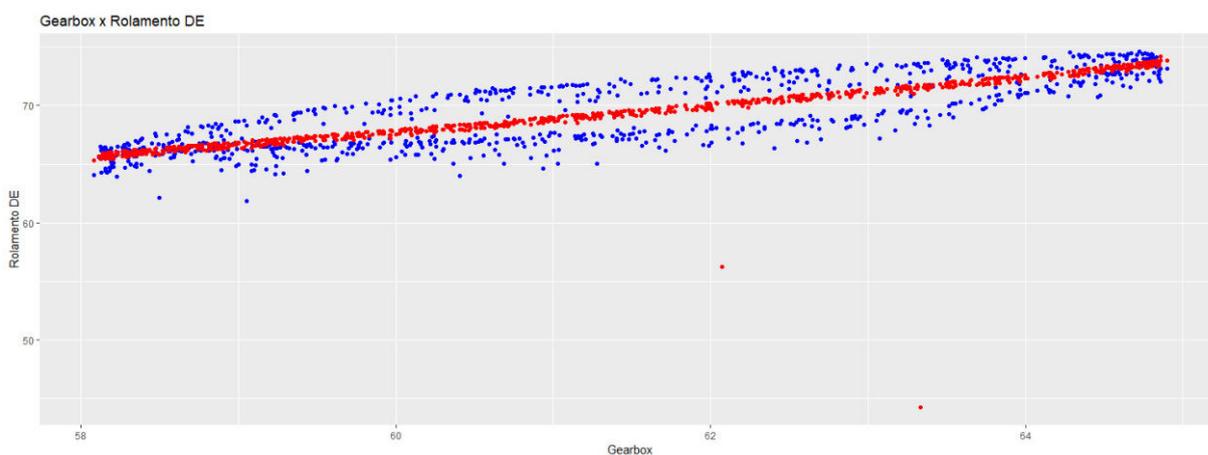
Fonte: Elaborada pelo autor

Figura 38 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Outubro).



Fonte: Elaborada pelo autor

Figura 39 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Novembro).



Fonte: Elaborada pelo autor

4.5. Gráficos de previsão – Regressão Polinomial – Aerogerador 2

Os gráficos mostrados abaixo trazem os dados do sistema SCADA e os dados previstos pelo modelo de regressão polinomial utilizando um polinômio de sexta ordem. Em verde são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 2. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 2.

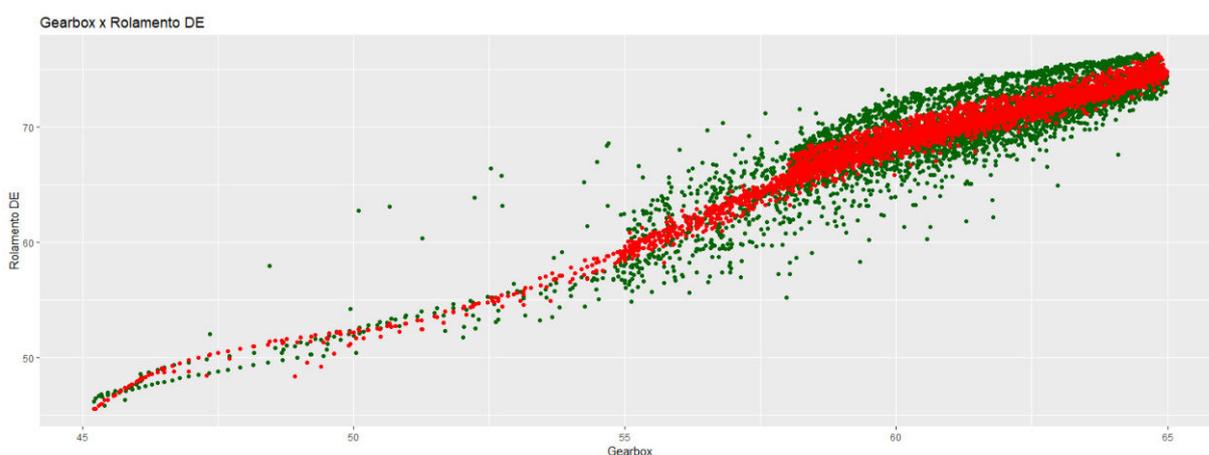
Tabela 2 – Dados de correlação e erro – Regressão Polinomial – Aerogerador 2

Regressão Polinomial - AEROGERADOR 2		
Mês	Coefficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,879	1,416
Fevereiro	0,895	1,531
Março	0,888	1,397
Abril	0,952	1,511
Mai	0,908	1,529
Junho	0,855	1,352
Julho	0,313	53,325
Agosto	0,118	54,908
Setembro	0,702	1,418
Outubro	0,810	1,459
Novembro	0,939	1,597
Média	0,879	1,511

Fonte: Elaborada pelo autor

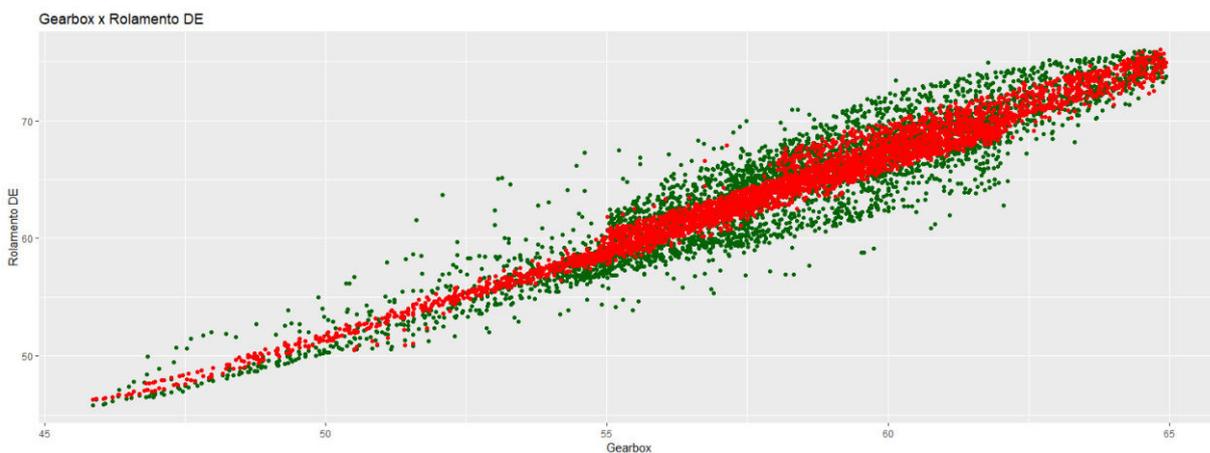
Da Tabela 2 pode-se observar que os meses onde a máquina apresentou defeito, julho e agosto, o coeficiente de correlação diminuiu bastante e o erro médio absoluto aumentou bastante. Nos demais meses o coeficiente de correlação foi superior a setenta por cento e o erro médio absoluto não foi superior a um virgula seis graus célsius.

Figura 40 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Janeiro).



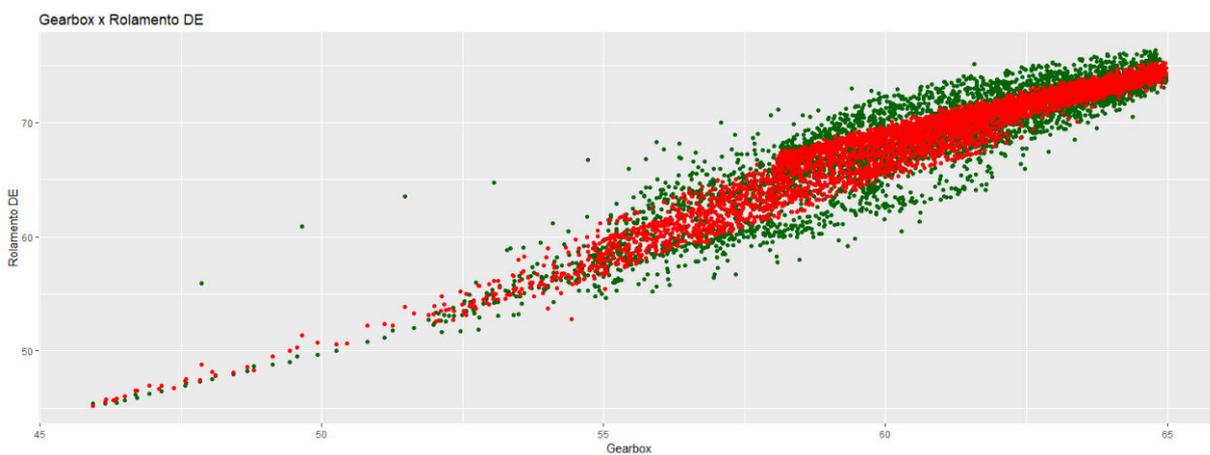
Fonte: Elaborada pelo autor

Figura 41 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Fevereiro).



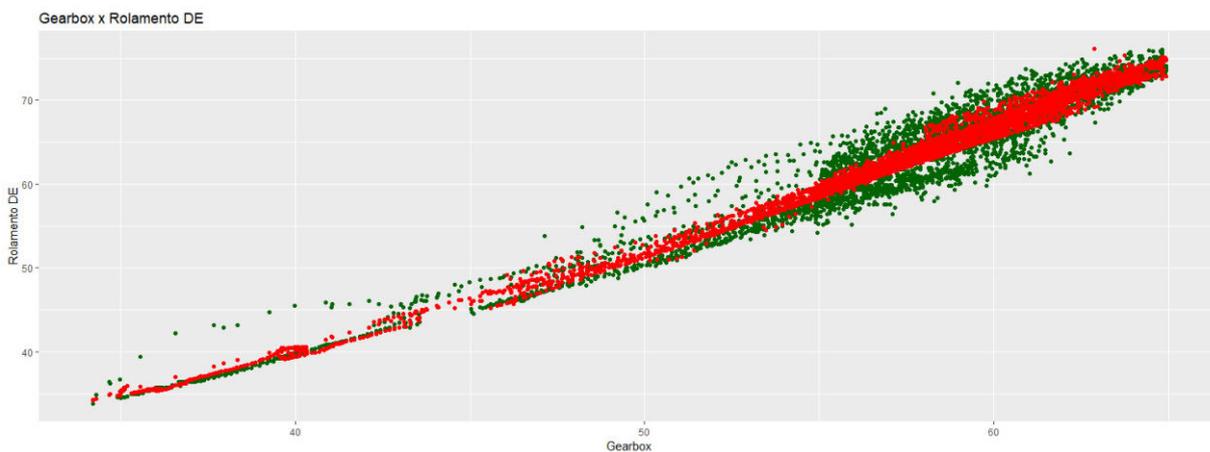
Fonte: Elaborada pelo autor

Figura 42 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Março).



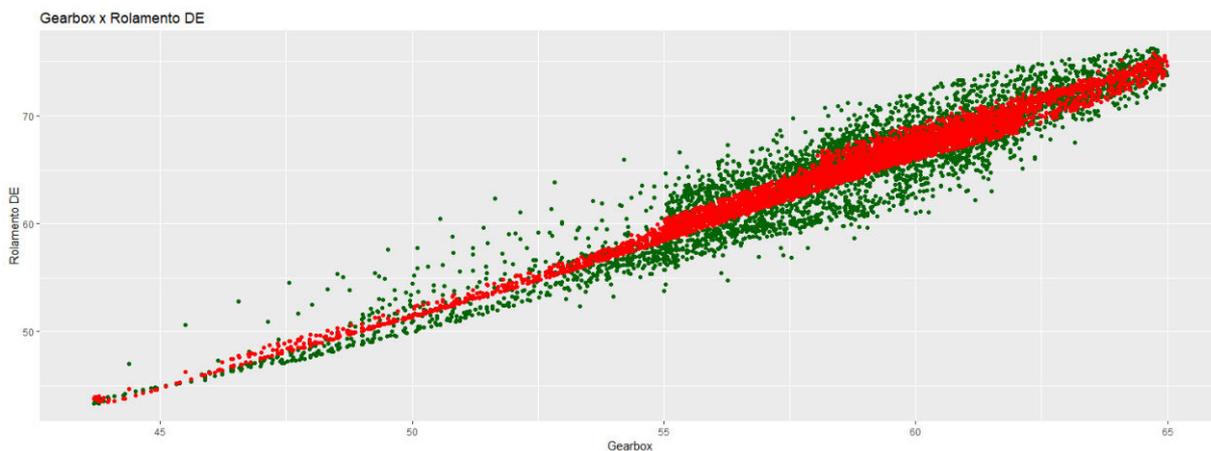
Fonte: Elaborada pelo autor

Figura 43 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Abril).



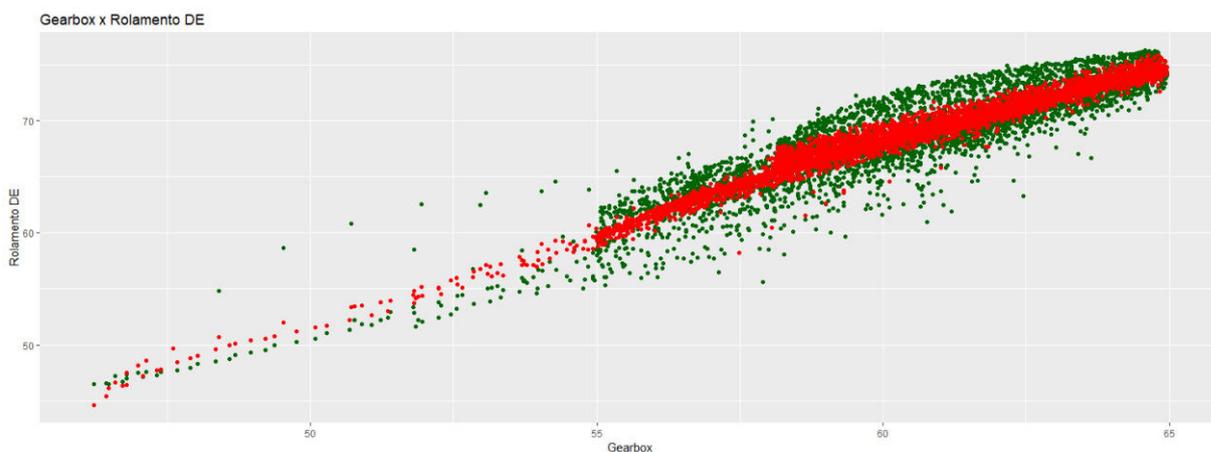
Fonte: Elaborada pelo autor

Figura 44 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Maio).



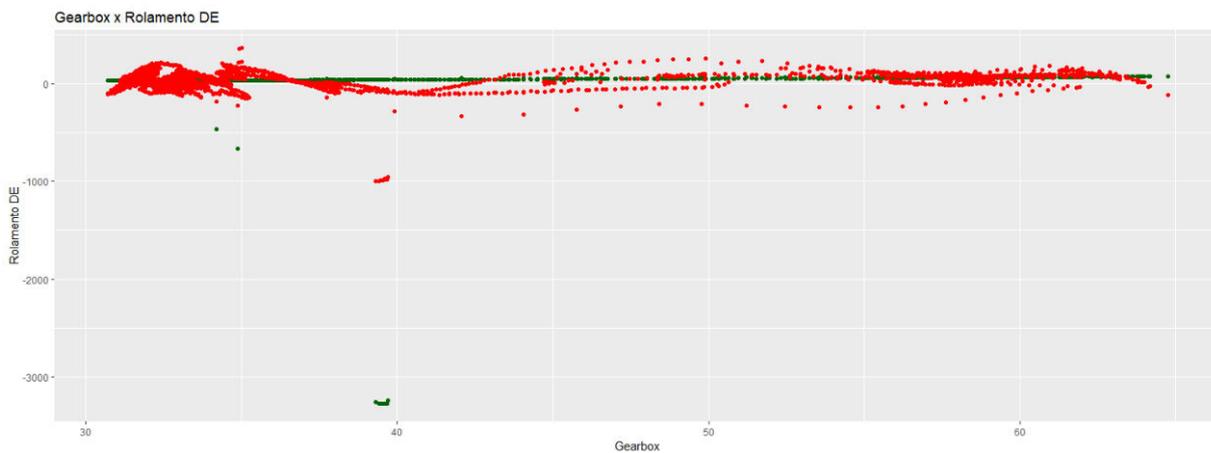
Fonte: Elaborada pelo autor

Figura 45 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Junho).



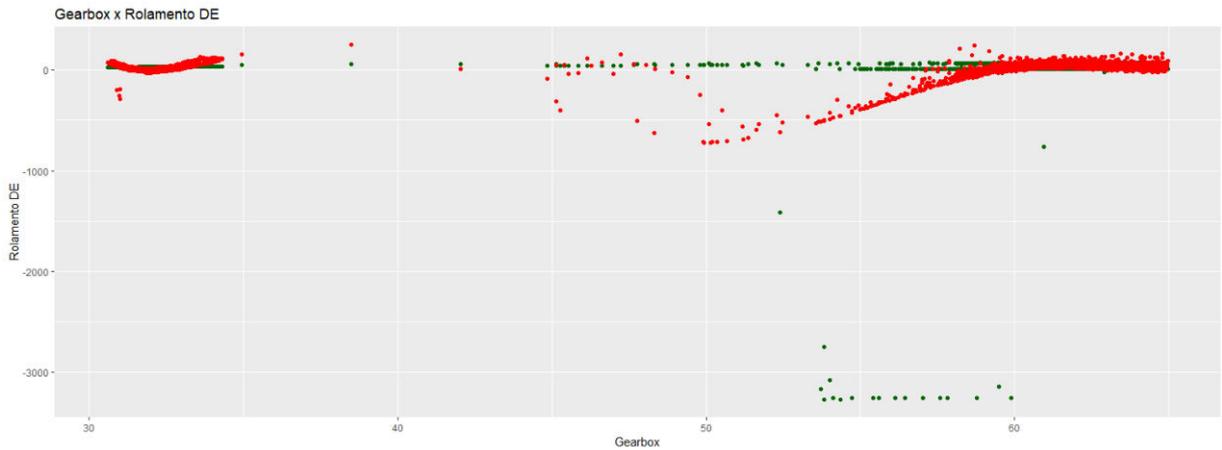
Fonte: Elaborada pelo autor

Figura 46 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Julho).



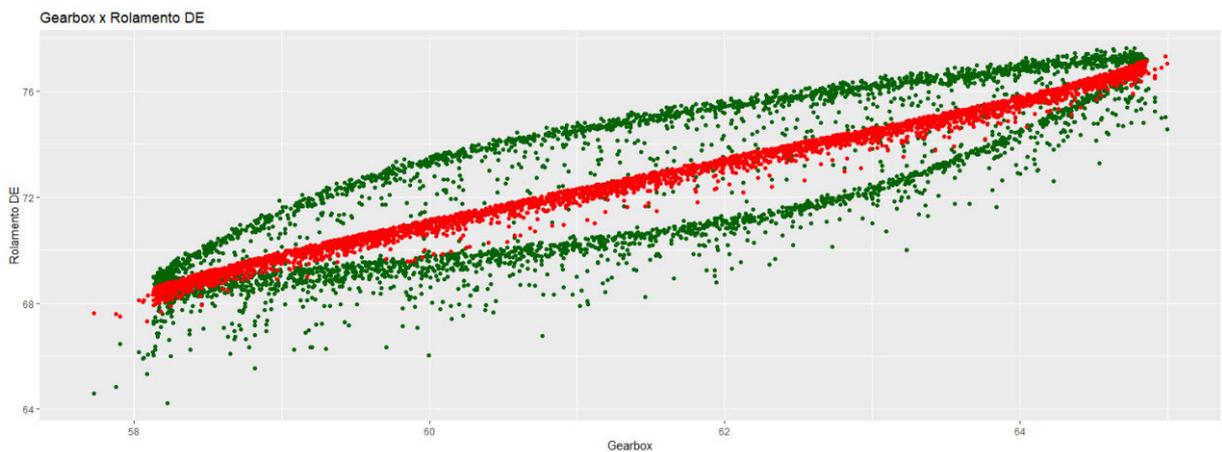
Fonte: Elaborada pelo autor

Figura 47 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Agosto).



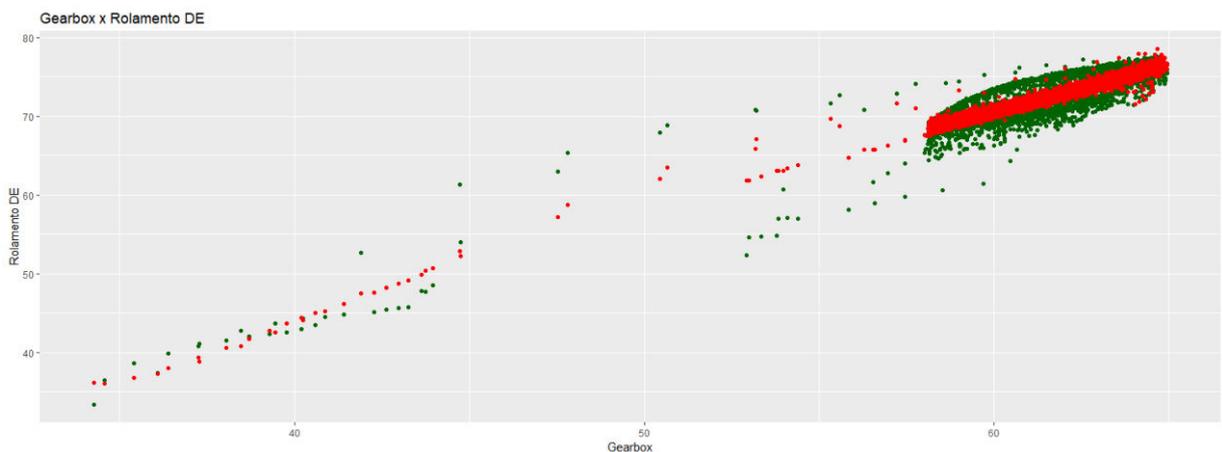
Fonte: Elaborada pelo autor

Figura 48 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Setembro).



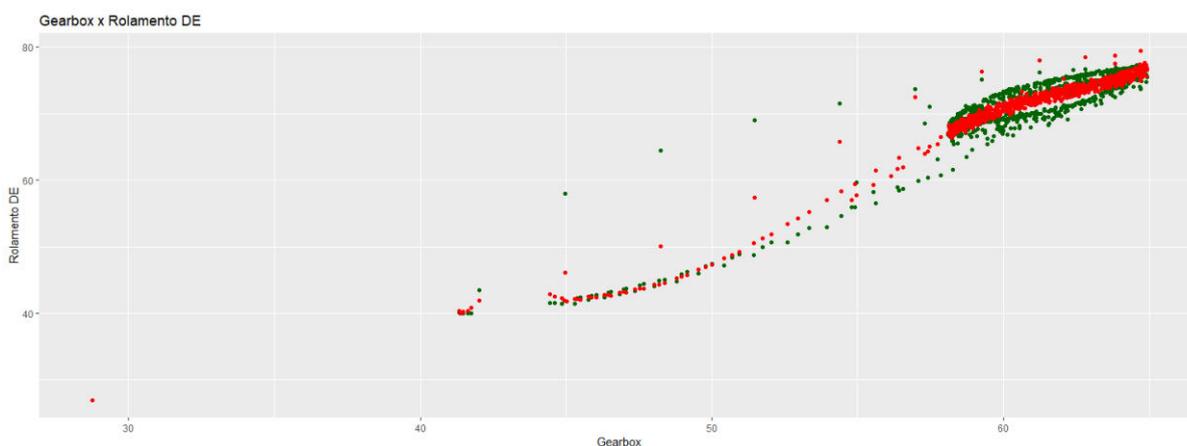
Fonte: Elaborada pelo autor

Figura 49 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Outubro).



Fonte: Elaborada pelo autor

Figura 50 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Novembro).



Fonte: Elaborada pelo autor

É possível observar que nos meses onde o sensor apresentou defeito, Figura 46 e Figura 47, o modelo computacional perdeu aderência e apresentou comportamento anômalo confirmando o que a Tabela 1 mostrou quanto aos coeficientes de correlação e valores de erro extremamente indesejados.

4.6. Gráficos de previsão – *Support Vector Machine* – Aerogerador 1

Os gráficos mostrados abaixo trazem os dados do sistema SCADA e os dados previstos pelo modelo de aprendizagem de máquinas *Support Vector Machine* utilizando o *kernel* radial uma vez que dentre as opções, linear, polinomial e sigmoide, foi o que obteve melhores resultados. O custo utilizado no modelo foi o custo padrão (um), a título de teste foi aplicado o modelo com custo de mil mas os resultados não apresentaram ganhos significativos em termos de assertividade e o tempo de cálculo aumentou substancialmente. Em azul são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

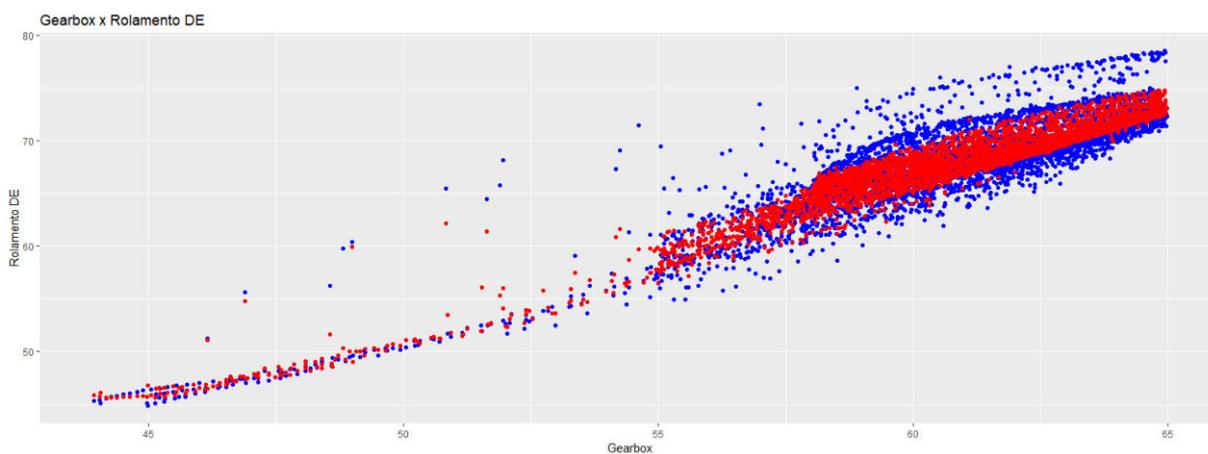
Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 3. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 3.

Tabela 3 – Dados de correlação e erro – SVM – Aerogerador 1

SVM - AEROGERADOR 1		
Mês	Coeficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,856	1,380
Fevereiro	0,907	1,156
Março	0,891	1,161
Abril	0,909	1,168
Maiο	0,933	1,218
Junho	0,859	1,273
Julho	0,855	1,207
Agosto	0,847	1,344
Setembro	0,670	1,253
Outubro	0,778	1,261
Novembro	0,667	1,327
Média	0,856	1,253

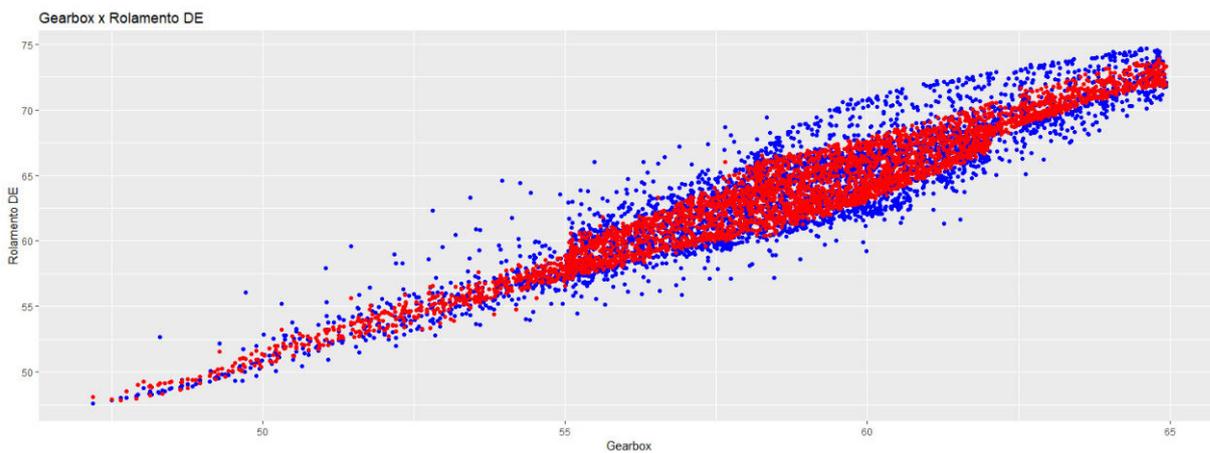
Fonte: Elaborada pelo autor

Figura 51 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Janeiro).



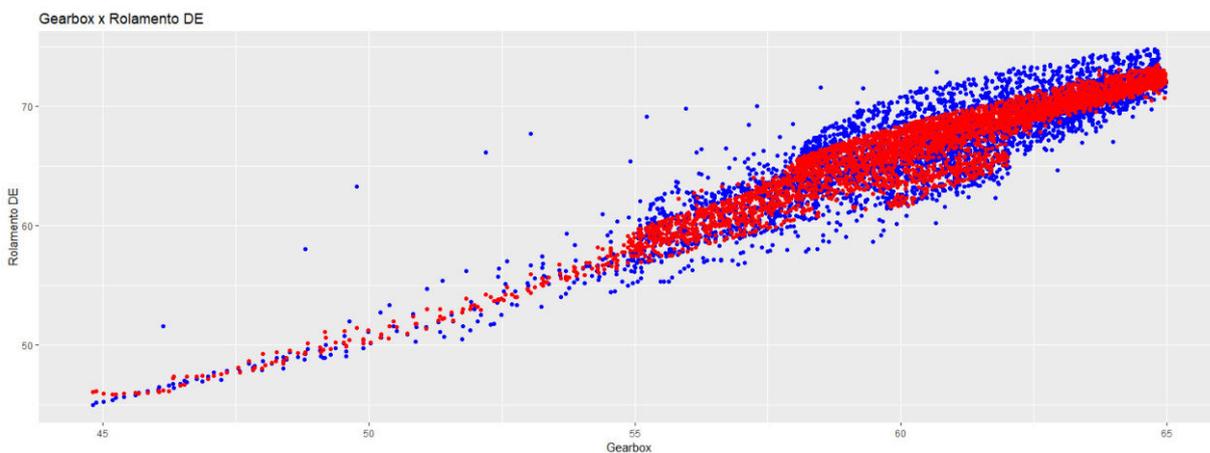
Fonte: Elaborada pelo autor

Figura 52 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Fevereiro).



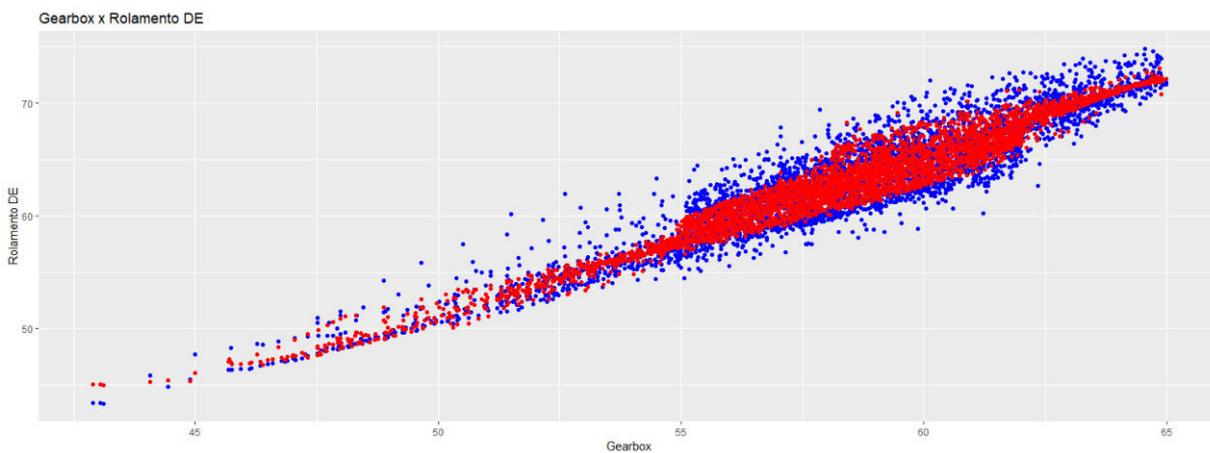
Fonte: Elaborada pelo autor

Figura 53 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Março).



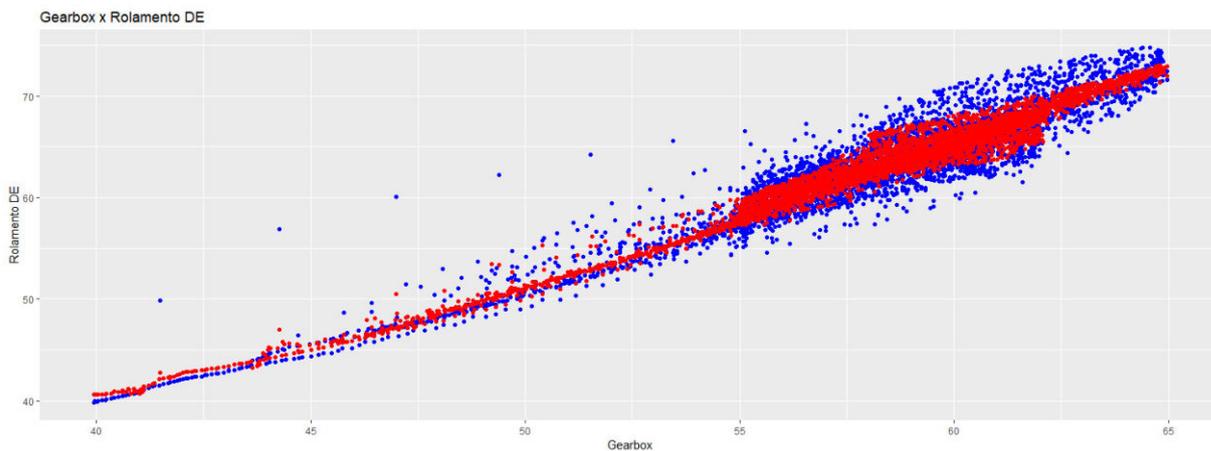
Fonte: Elaborada pelo autor

Figura 54 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Abril).



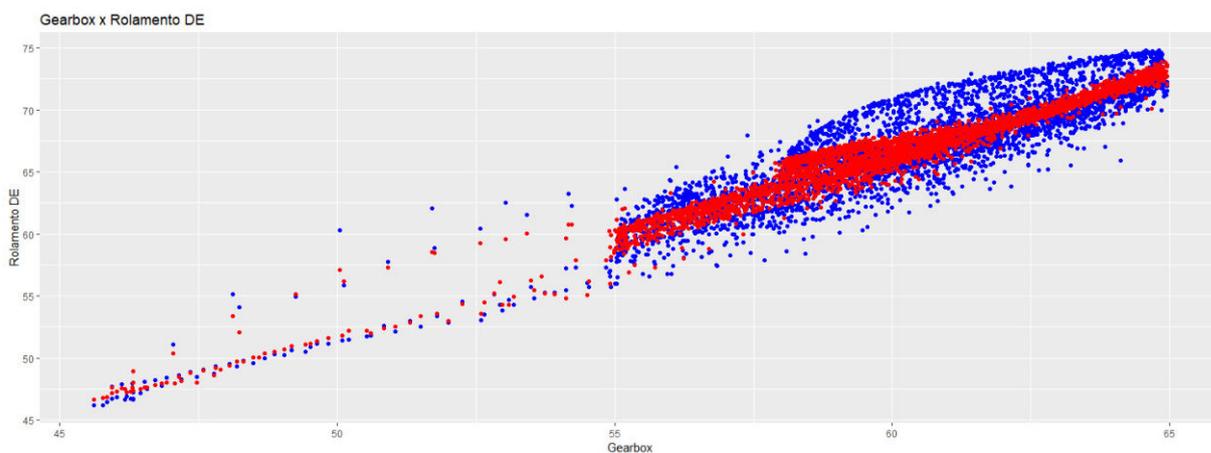
Fonte: Elaborada pelo autor

Figura 55 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Maio).



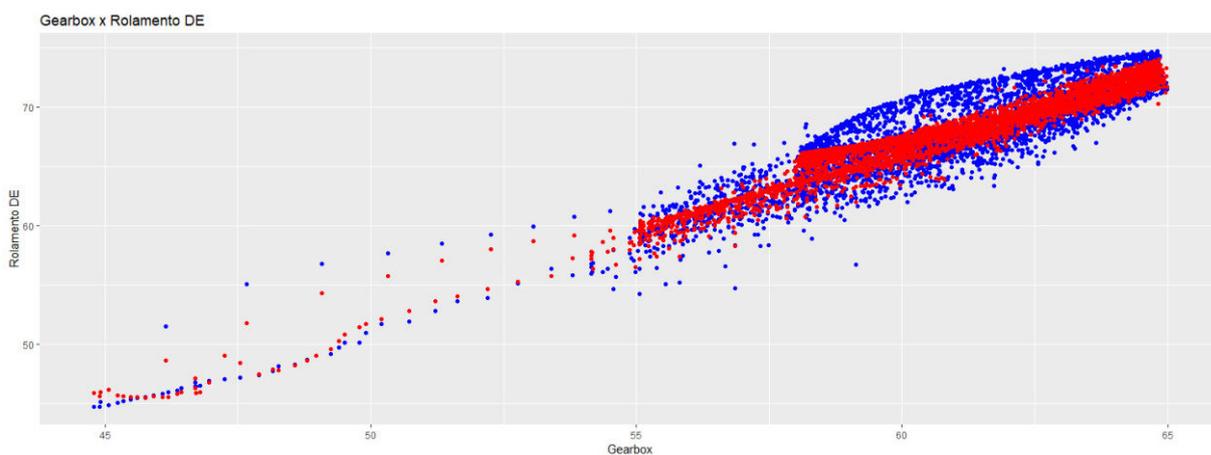
Fonte: Elaborada pelo autor

Figura 56 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Junho).



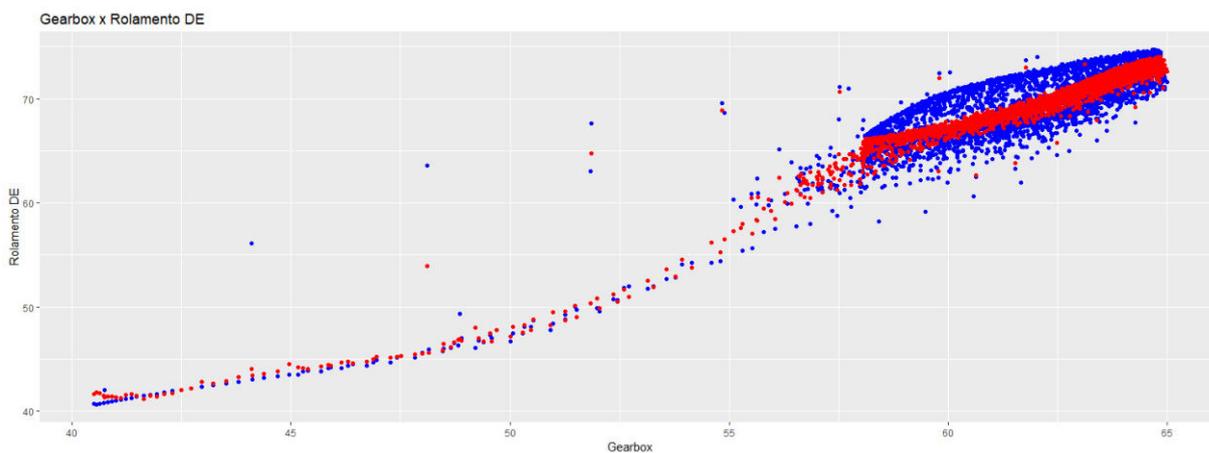
Fonte: Elaborada pelo autor

Figura 57 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Julho).



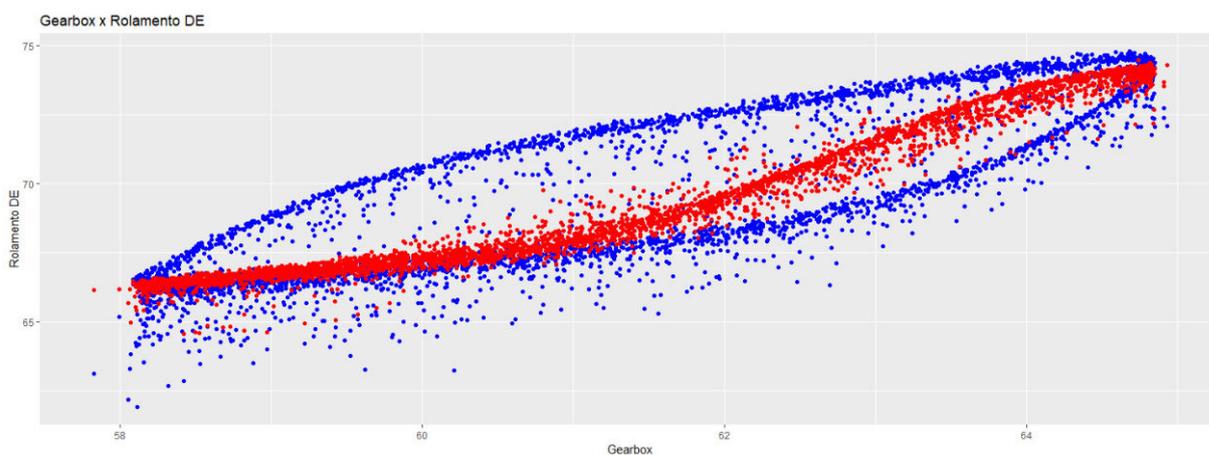
Fonte: Elaborada pelo autor

Figura 58 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Agosto).



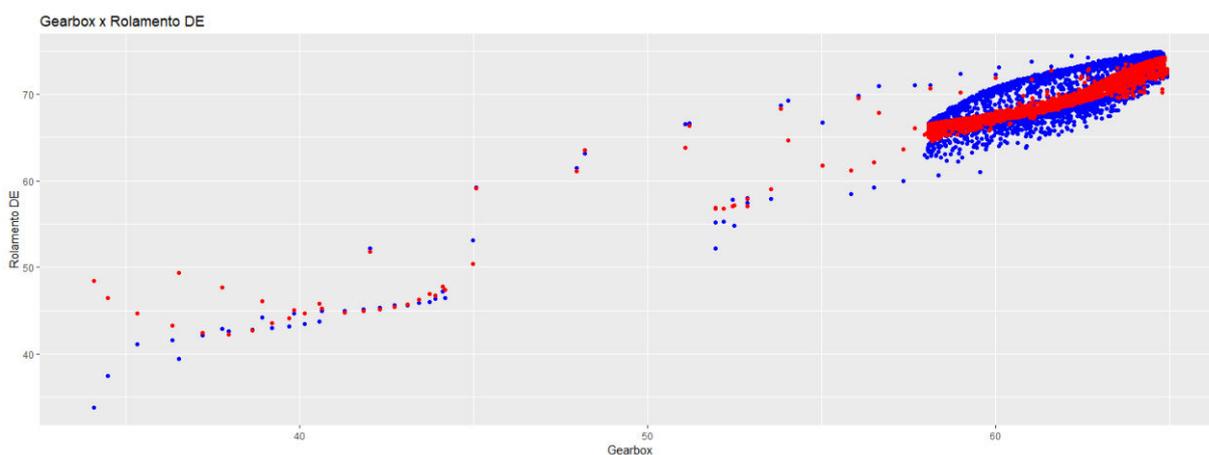
Fonte: Elaborada pelo autor

Figura 59 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Setembro).



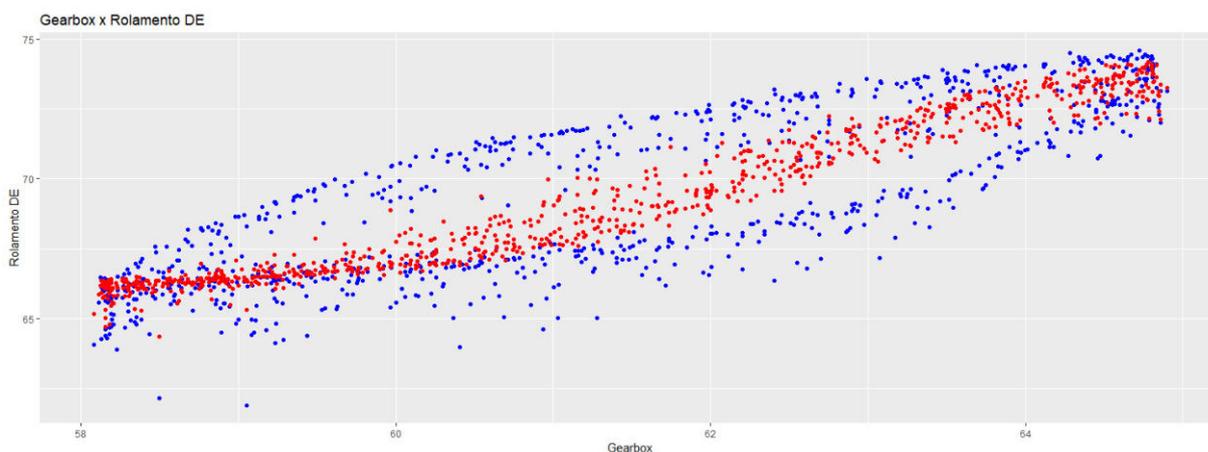
Fonte: Elaborada pelo autor

Figura 60 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Outubro).



Fonte: Elaborada pelo autor

Figura 61 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Novembro).



Fonte: Elaborada pelo autor

4.7. Gráficos de previsão – *Support Vector Machine* – Aerogerador 2

Os gráficos mostrados abaixo trazem os dados do sistema SCADA e os dados previstos pelo modelo de aprendizagem de máquinas *Support Vector Machine* utilizando o *kernel* radial uma vez que dentre as opções, linear, polinomial e sigmoide, foi o que obteve melhores resultados. O custo utilizado no modelo foi o custo padrão (um), a título de teste foi aplicado o modelo com custo de mil mas os resultados não apresentaram ganhos significativos em termos de assertividade e o tempo de cálculo aumentou substancialmente. Em verde são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 4. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 4.

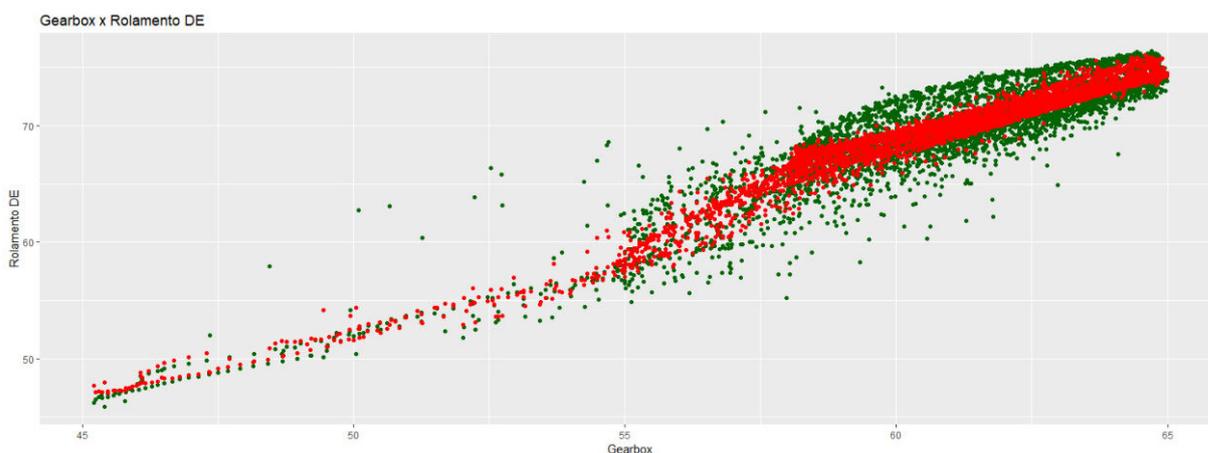
Tabela 4 – Dados de correlação e erro – SVM – Aerogerador 2

SVM - AEROGERADOR 2		
Mês	Coefficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,883	1,294
Fevereiro	0,908	1,355
Março	0,892	1,219
Abril	0,957	1,394
Maio	0,916	1,361
Junho	0,869	1,259
Julho	0,078	10,566
Agosto	-0,001	32,349
Setembro	0,687	1,358
Outubro	0,808	1,351
Novembro	0,929	1,455
Média	0,883	1,358

Fonte: Elaborada pelo autor

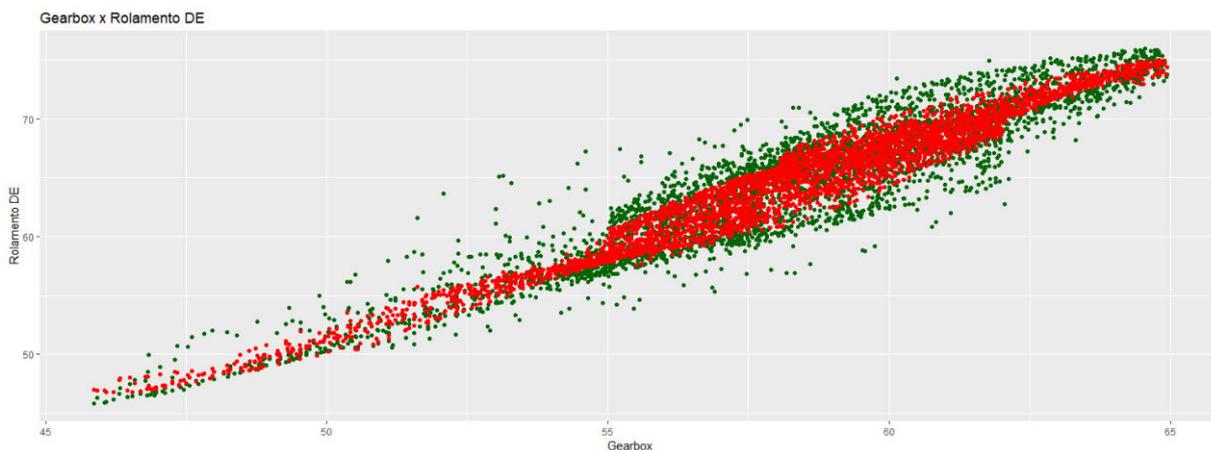
Conforme a Tabela 4 o valor do coeficiente de correlação no mês de agosto foi negativo e isso pode ser explicado pela fraca correlação entre os dados uma vez que o sensor de temperatura do rolamento da multiplicadora apresentou defeito, mas o sensor de temperatura da multiplicadora continuou funcionando. Esse fenômeno também ocorreu no mês de julho, mas o coeficiente de correlação, apesar de baixo, não chegou a ser negativo.

Figura 62 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Janeiro).



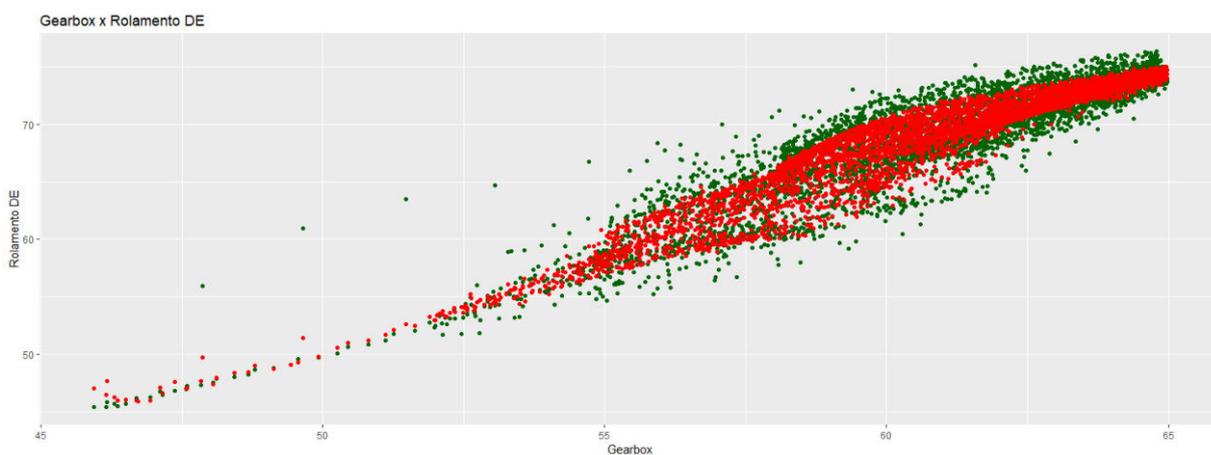
Fonte: Elaborada pelo autor

Figura 63 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Fevereiro).



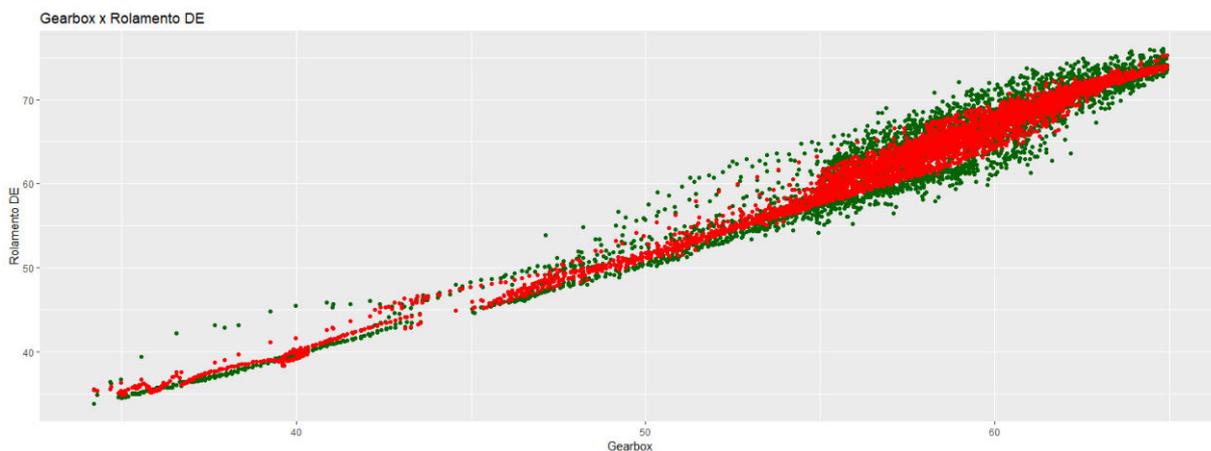
Fonte: Elaborada pelo autor

Figura 64 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Março).



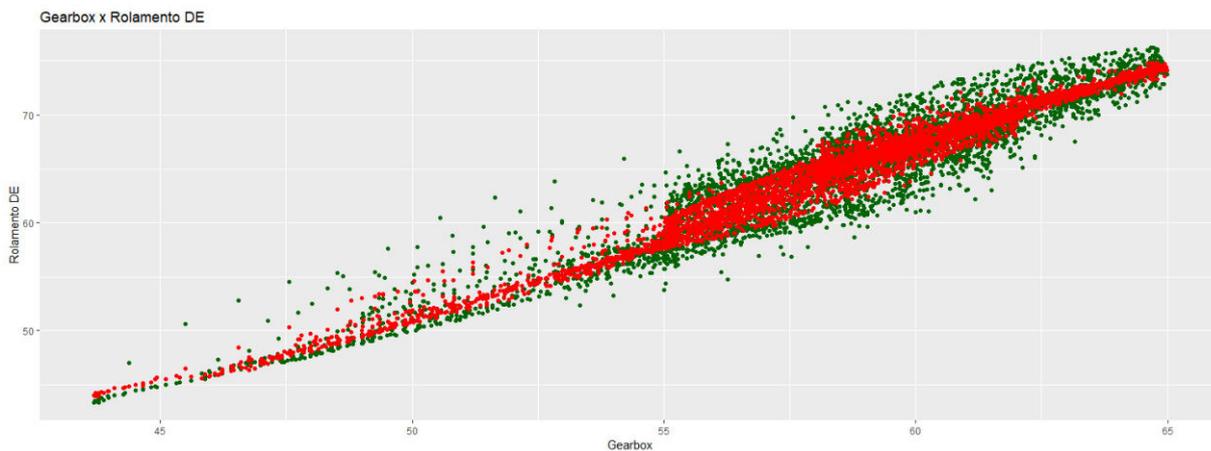
Fonte: Elaborada pelo autor

Figura 65 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Abril).



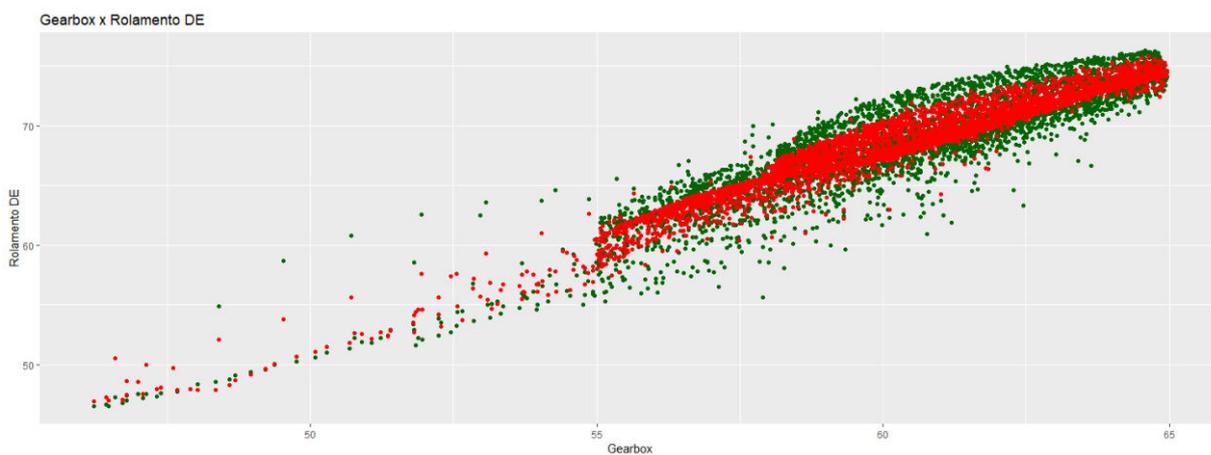
Fonte: Elaborada pelo autor

Figura 66 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Maio).



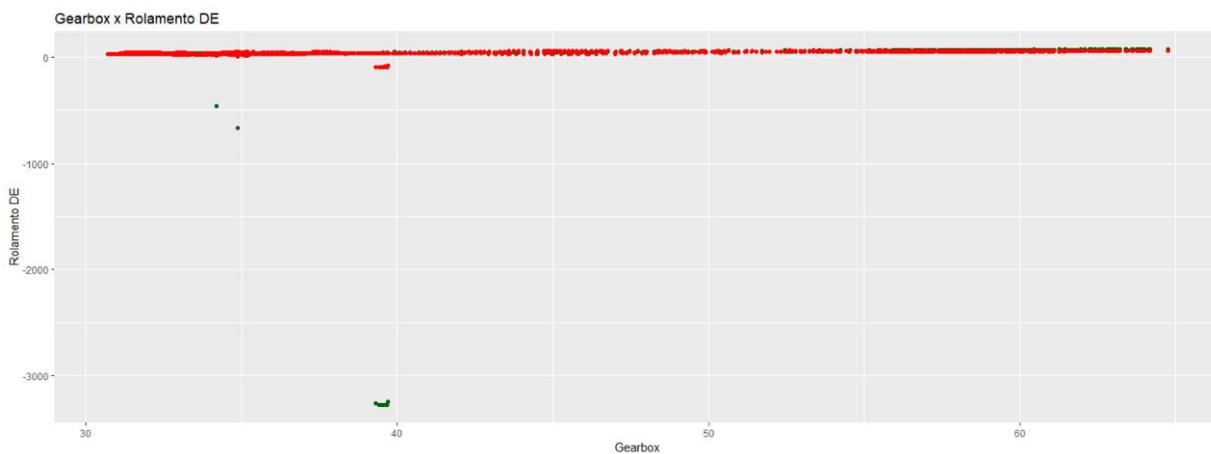
Fonte: Elaborada pelo autor

Figura 67 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Junho).



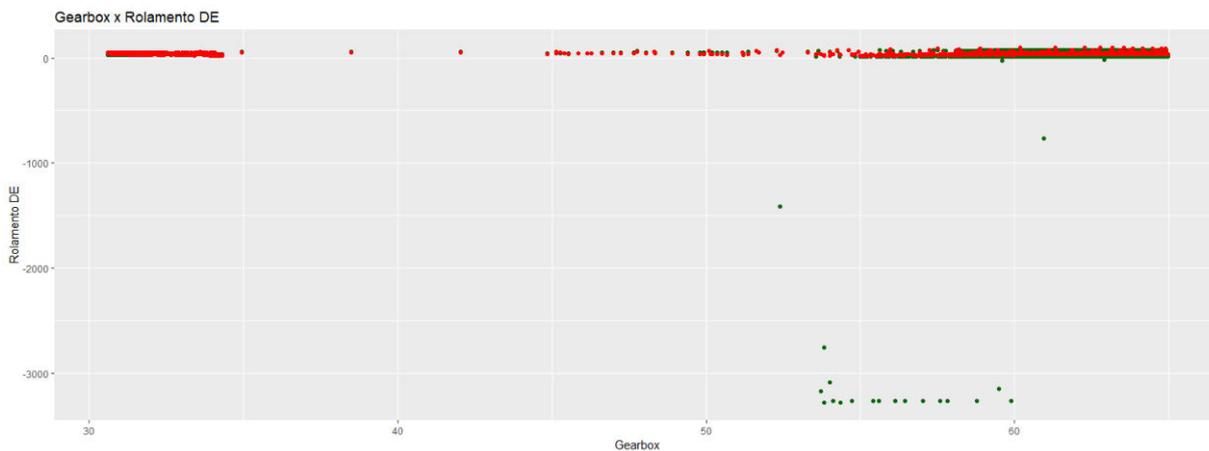
Fonte: Elaborada pelo autor

Figura 68 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Julho).



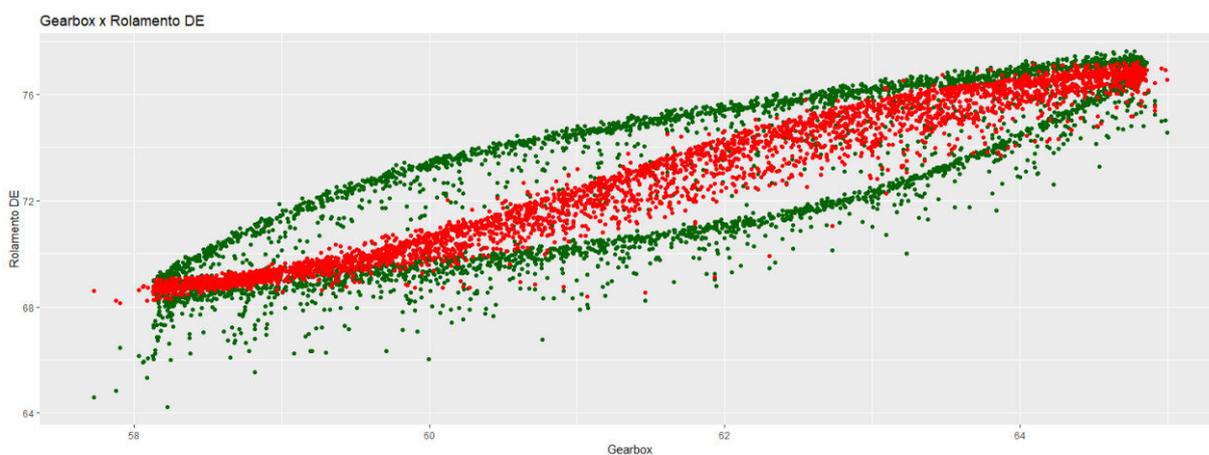
Fonte: Elaborada pelo autor

Figura 69 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Agosto).



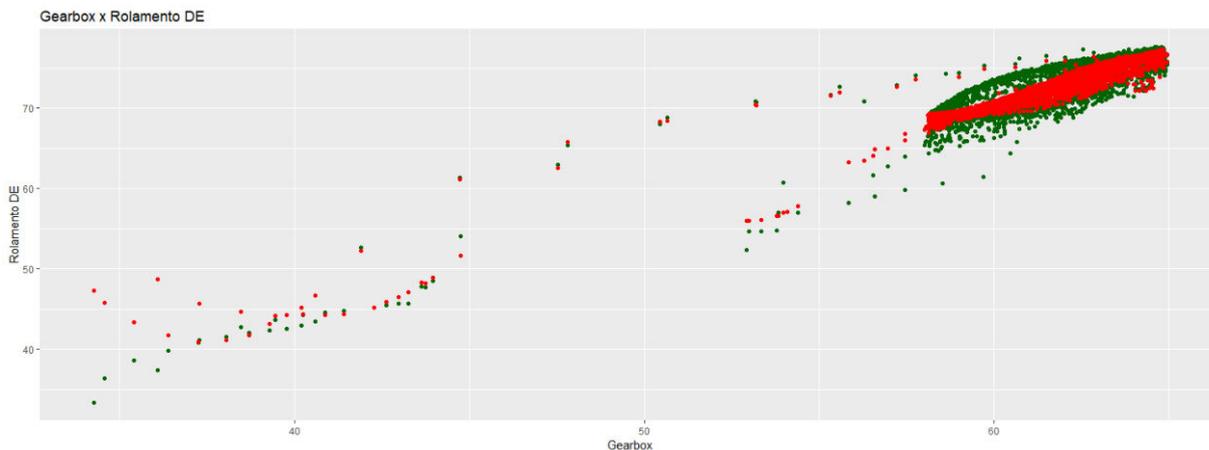
Fonte: Elaborada pelo autor

Figura 70 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Setembro).



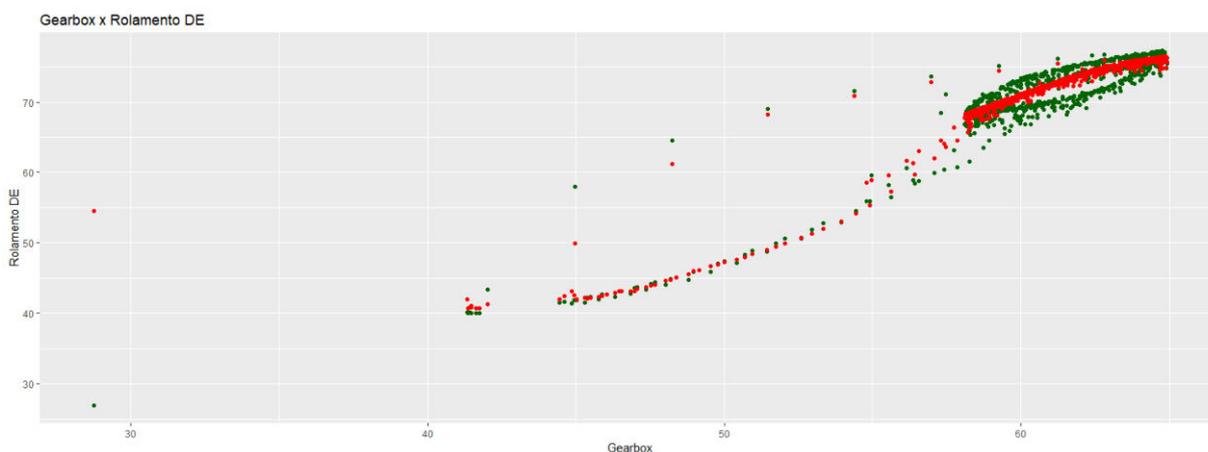
Fonte: Elaborada pelo autor

Figura 71 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Outubro).



Fonte: Elaborada pelo autor

Figura 72 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Novembro).



Fonte: Elaborada pelo autor

4.8. Gráficos de previsão – Regressão Linear – Aerogerador 1

Os gráficos mostrados abaixo trazem os dados do sistema SCADA e os dados previstos pelo modelo de aprendizagem de regressão linear simples. Em azul são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

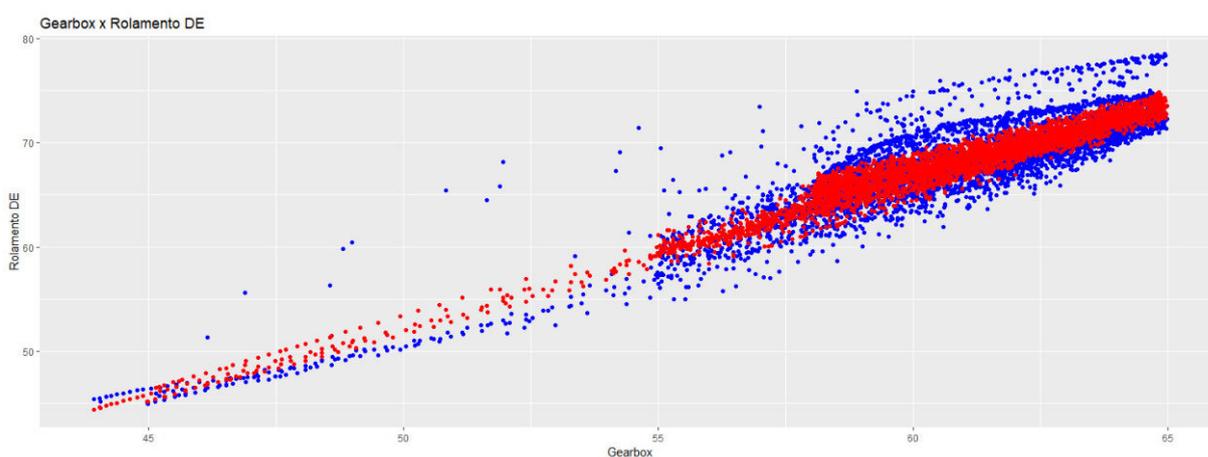
Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 5. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 5.

Tabela 5 – Dados de correlação e erro – Regressão Linear – Aerogerador 1

Regressão Linear - AEROGERADOR 1		
Mês	Coeficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,837	1,523
Fevereiro	0,899	1,252
Março	0,880	1,270
Abril	0,896	1,300
Maio	0,923	1,335
Junho	0,844	1,383
Julho	0,838	1,344
Agosto	0,827	1,480
Setembro	0,695	1,382
Outubro	0,773	1,425
Novembro	0,682	1,443
Média	0,838	1,382

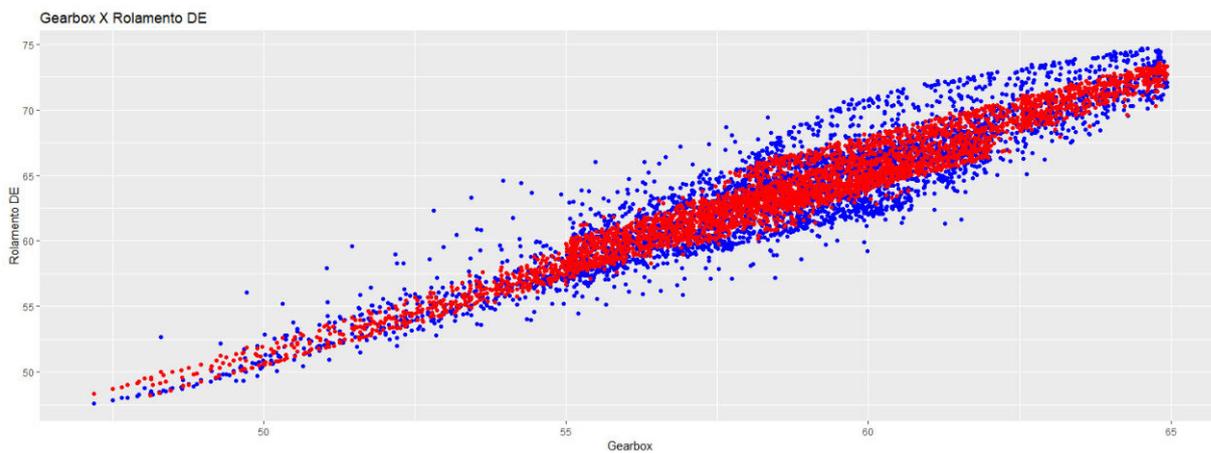
Fonte: Elaborada pelo autor

Figura 73 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Janeiro).



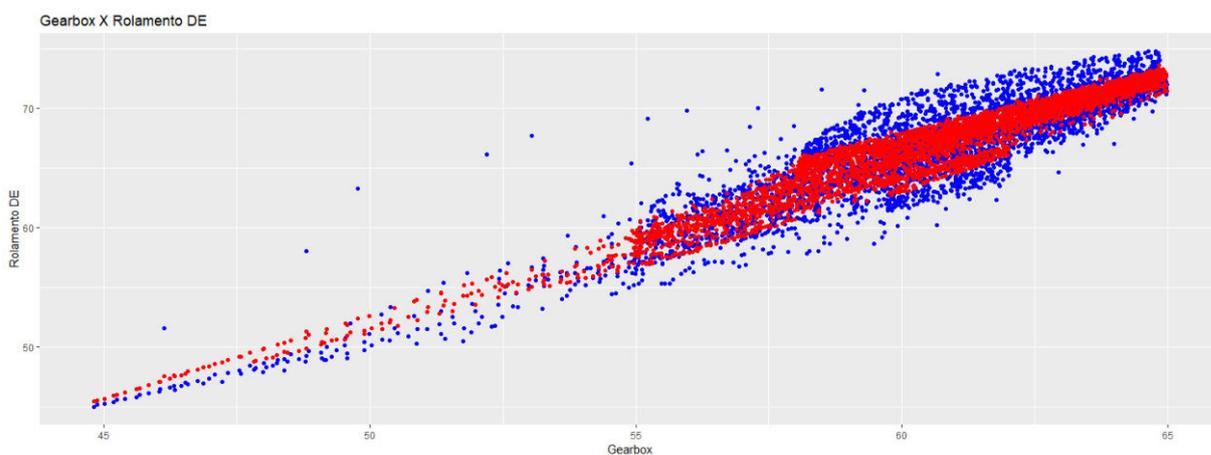
Fonte: Elaborada pelo autor

Figura 74 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Fevereiro).



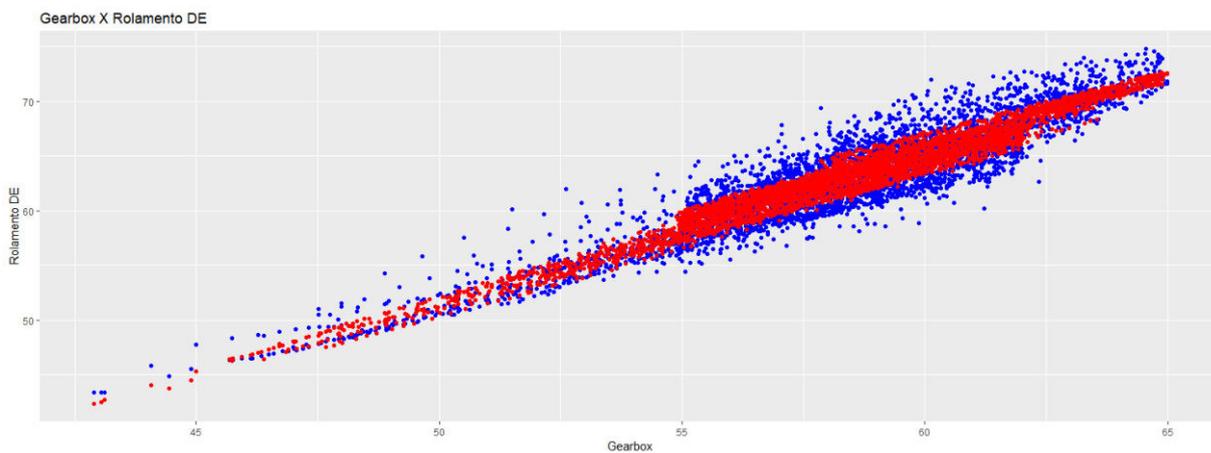
Fonte: Elaborada pelo autor

Figura 75 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Março).



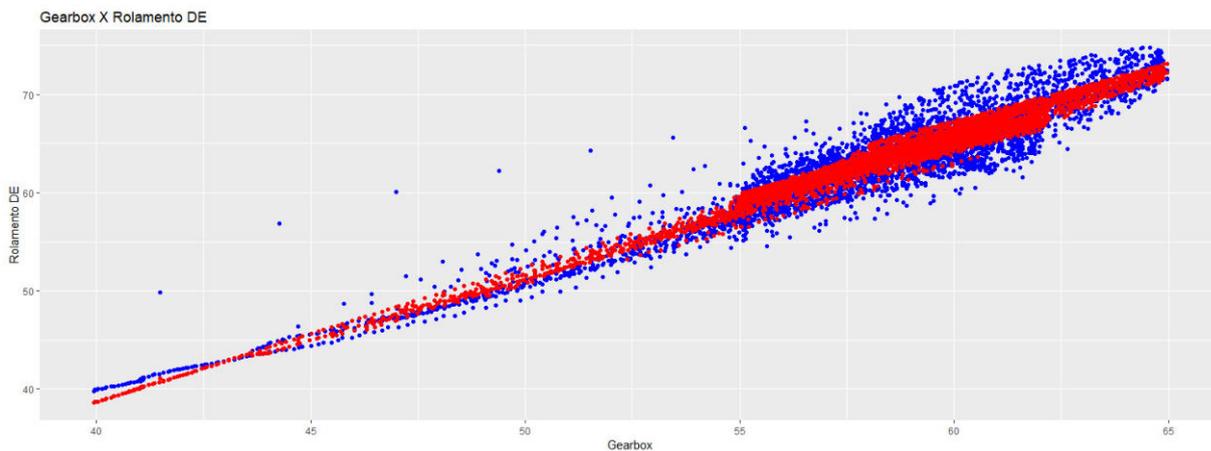
Fonte: Elaborada pelo autor

Figura 76 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Abril).



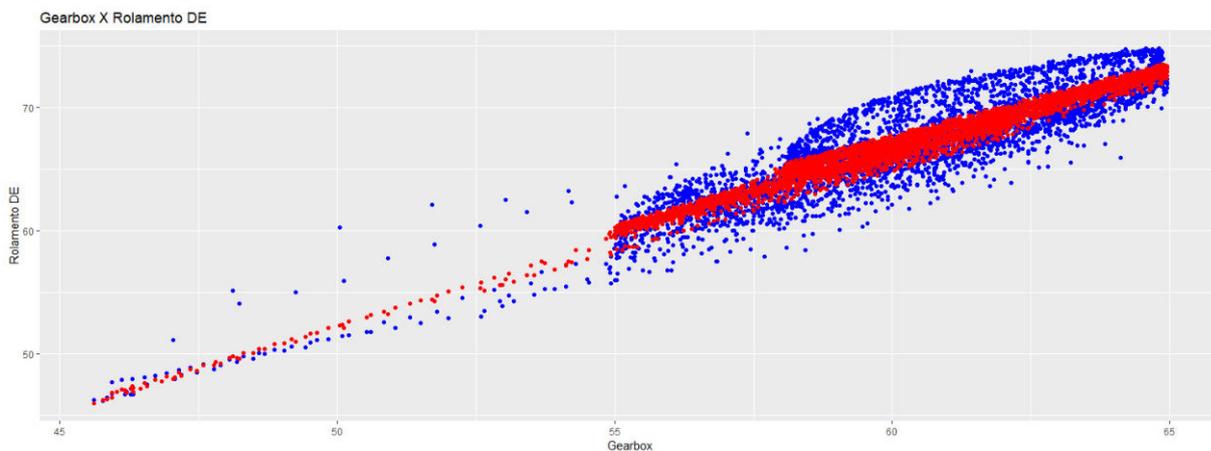
Fonte: Elaborada pelo autor

Figura 77 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Maio).



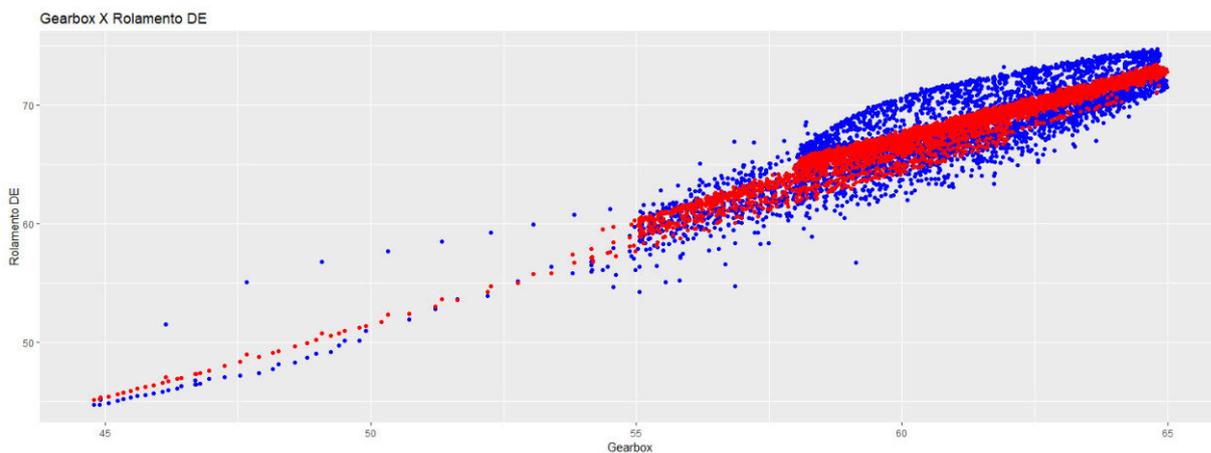
Fonte: Elaborada pelo autor

Figura 78 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Junho).



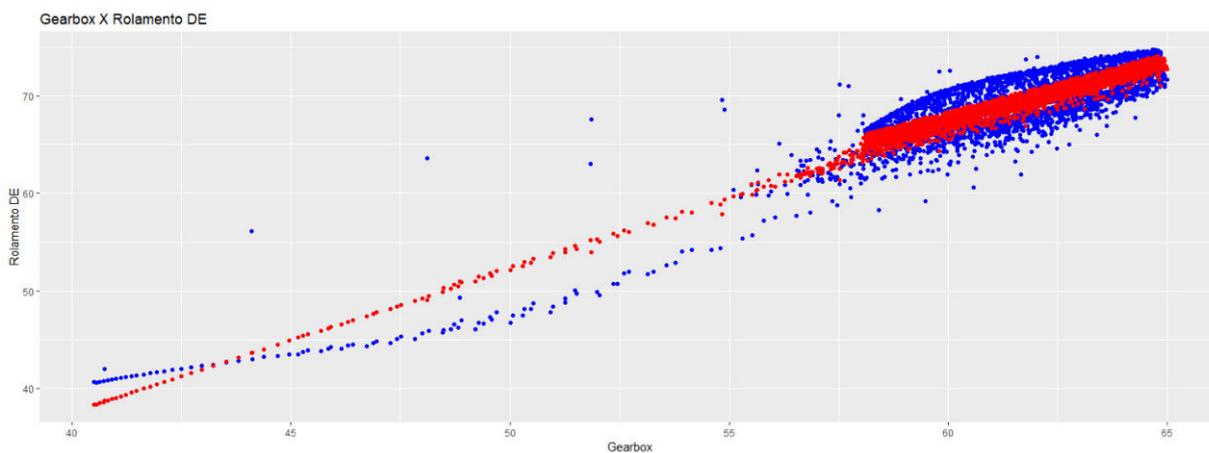
Fonte: Elaborada pelo autor

Figura 79 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Julho).



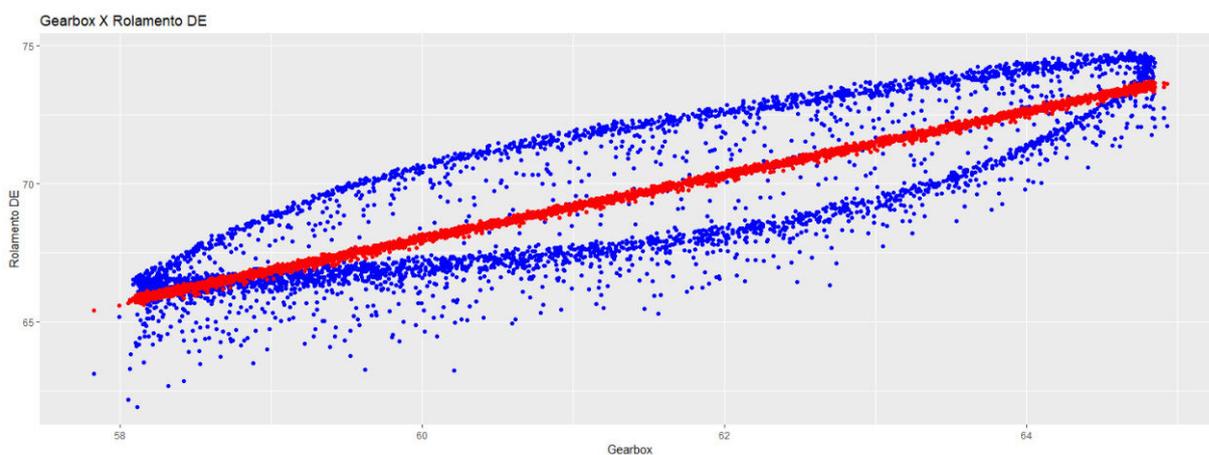
Fonte: Elaborada pelo autor

Figura 80 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Agosto).



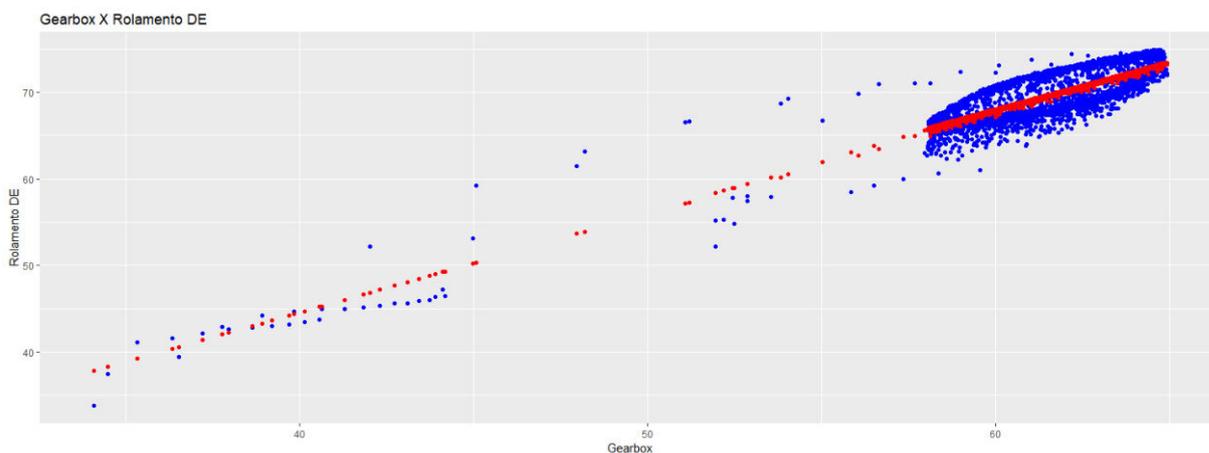
Fonte: Elaborada pelo autor

Figura 81 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Setembro).



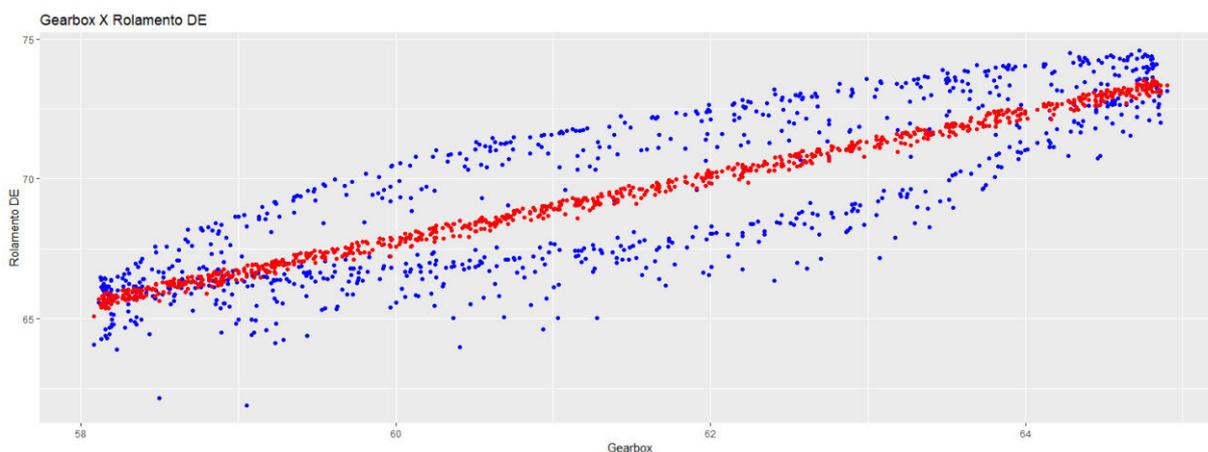
Fonte: Elaborada pelo autor

Figura 82 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Outubro).



Fonte: Elaborada pelo autor

Figura 83 – Gráfico de dispersão – temperatura rolamento1 x multiplicadora1 (Novembro).



Fonte: Elaborada pelo autor

4.9. Gráficos de previsão – Regressão Linear – Aerogerador 2

Os gráficos mostrados abaixo trazem os dados do sistema SCADA e os dados previstos pelo modelo de aprendizagem de regressão linear simples. Em verde são os dados adquiridos pelo sistema SCADA e em vermelho são os dados previstos pelo modelo computacional.

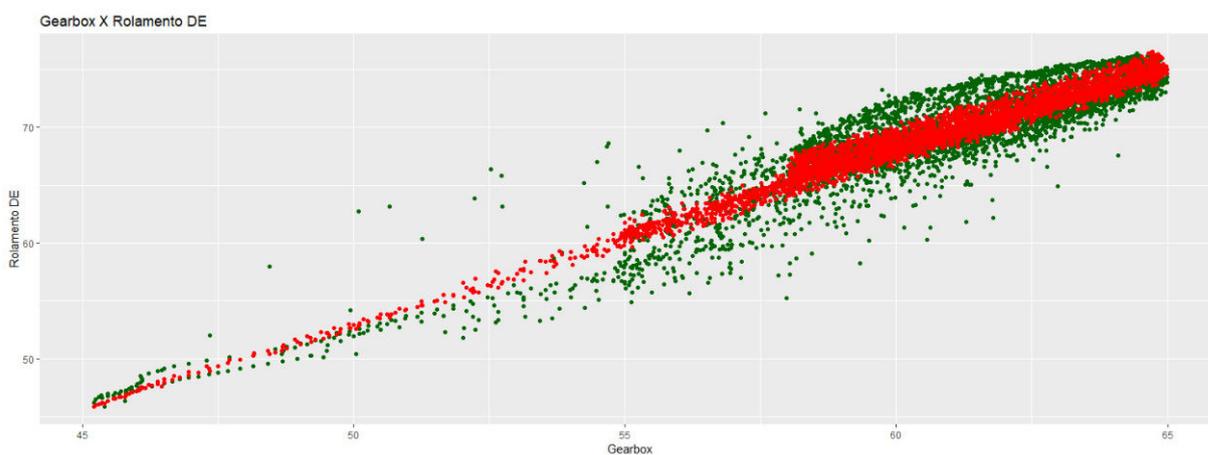
Para esse modelo e aerogerador o erro absoluto médio foi calculado pelo método dos mínimos quadrados e os valores estão apresentados na Tabela 6. O coeficiente de correlação entre os dados e os valores previsto foram calculados e também estão na Tabela 6.

Tabela 6 – Dados de correlação e erro – Regressão Linear – Aerogerador 2

Regressão Linear - AEROGERADOR 2		
Mês	Coefficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,864	1,429
Fevereiro	0,894	1,503
Março	0,879	1,389
Abril	0,944	1,707
Maio	0,904	1,525
Junho	0,845	1,412
Julho	0,033	23,066
Agosto	0,001	30,270
Setembro	0,704	1,451
Outubro	0,785	1,514
Novembro	0,894	1,852
Média	0,864	1,514

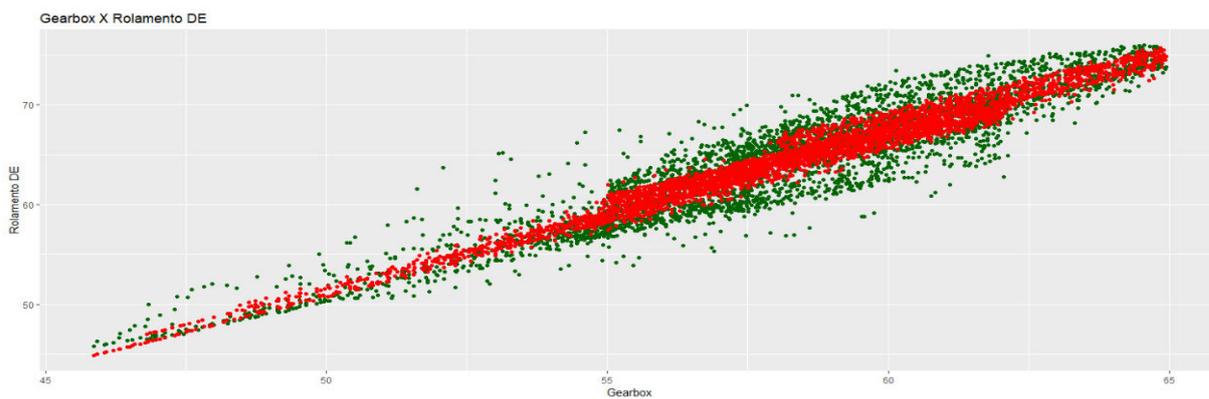
Fonte: Elaborada pelo autor

Figura 84 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Janeiro).



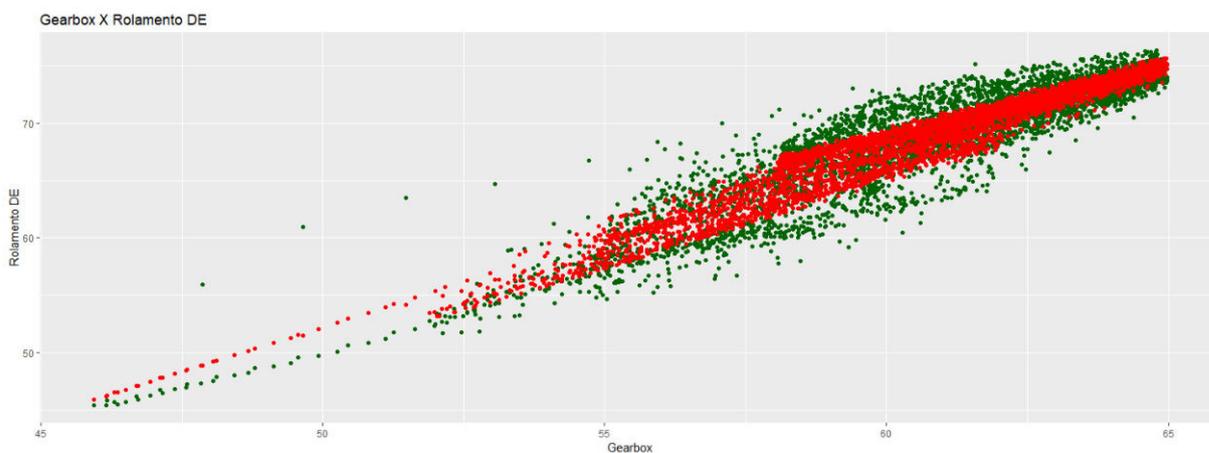
Fonte: Elaborada pelo autor

Figura 85 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Fevereiro).



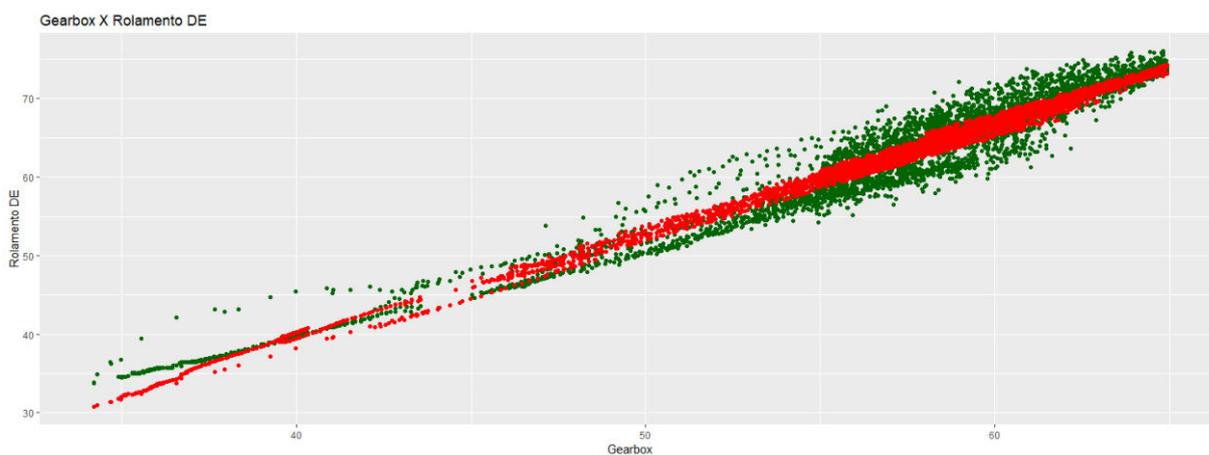
Fonte: Elaborada pelo autor

Figura 86 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Março).



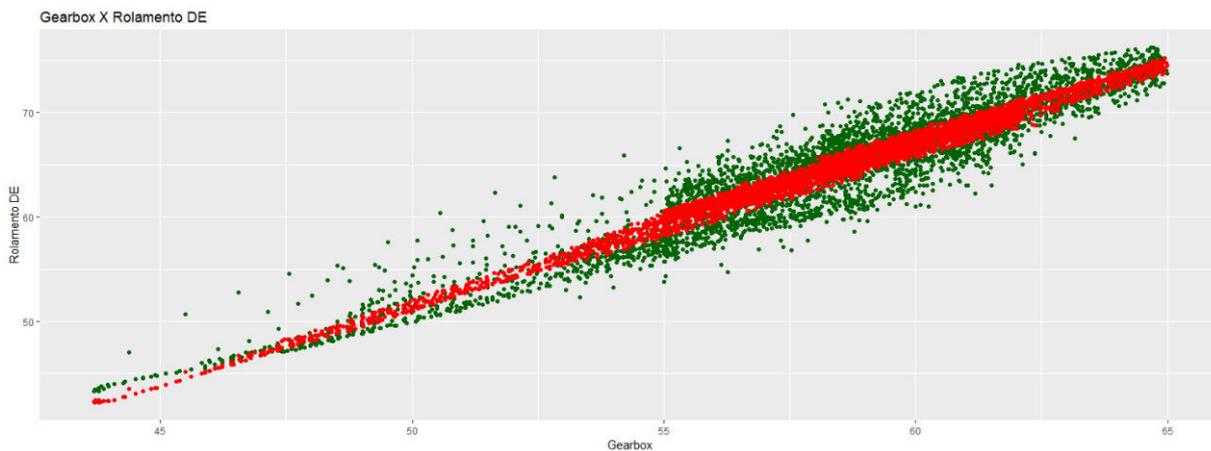
Fonte: Elaborada pelo autor

Figura 87 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Abril).



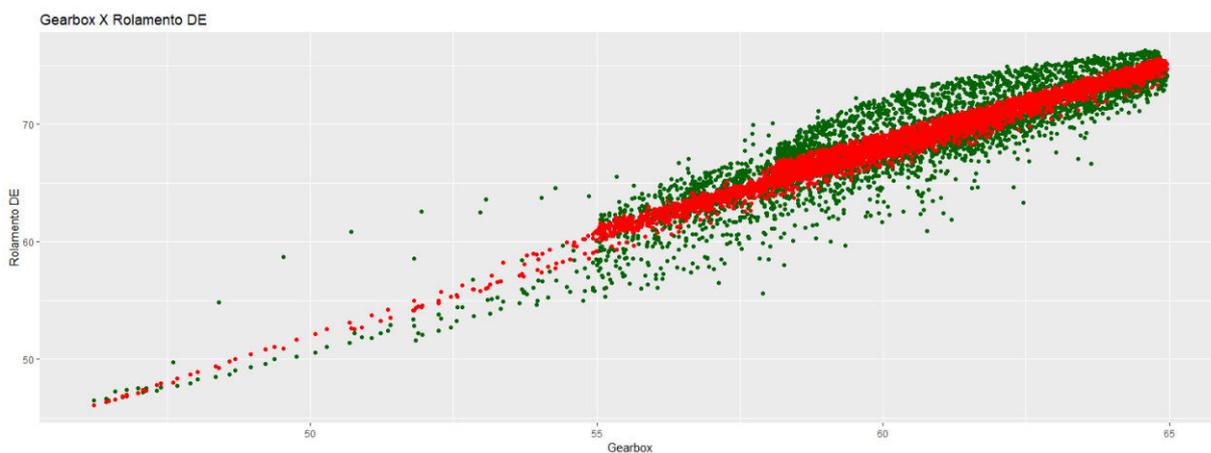
Fonte: Elaborada pelo autor

Figura 88 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Maio).



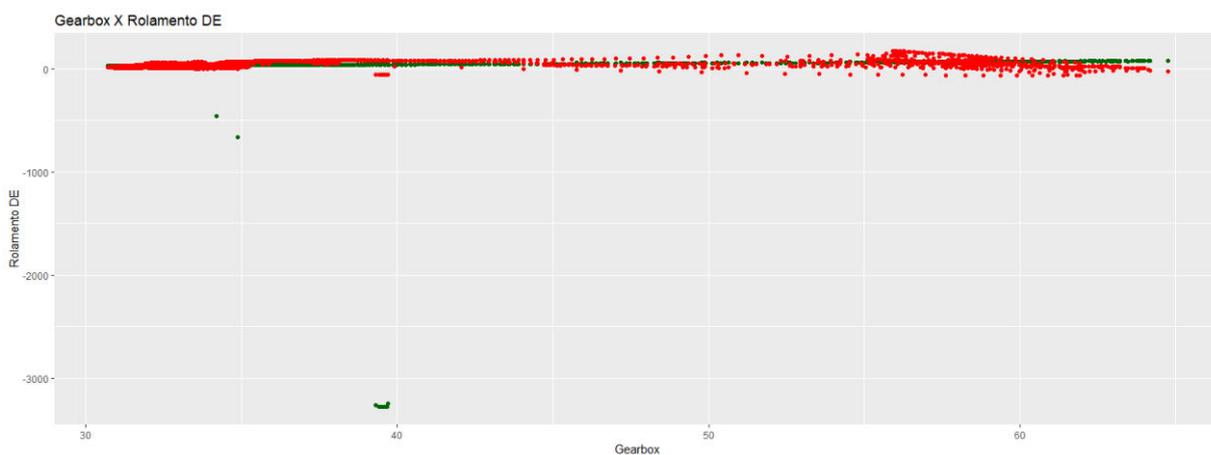
Fonte: Elaborada pelo autor

Figura 89 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Junho).



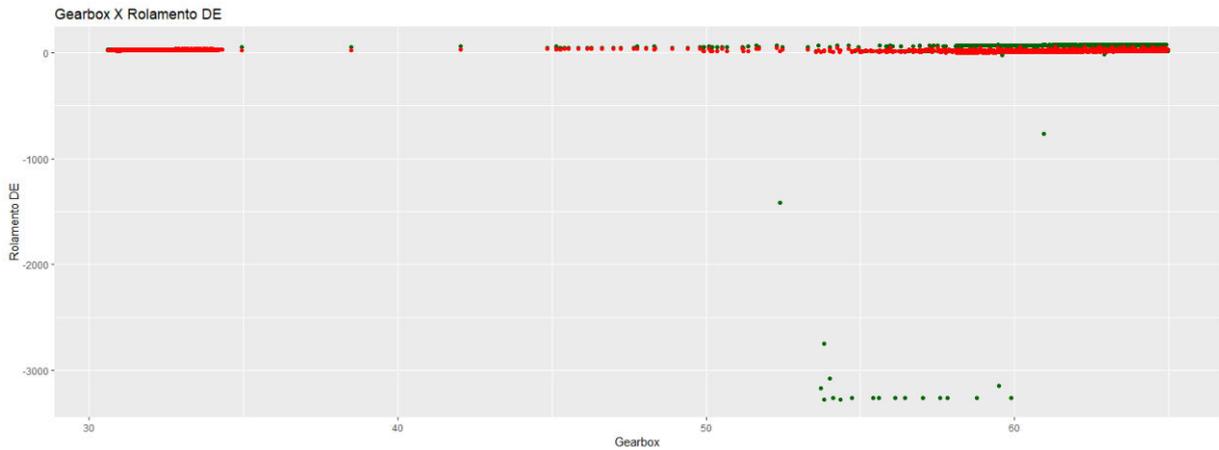
Fonte: Elaborada pelo autor

Figura 90 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Julho).



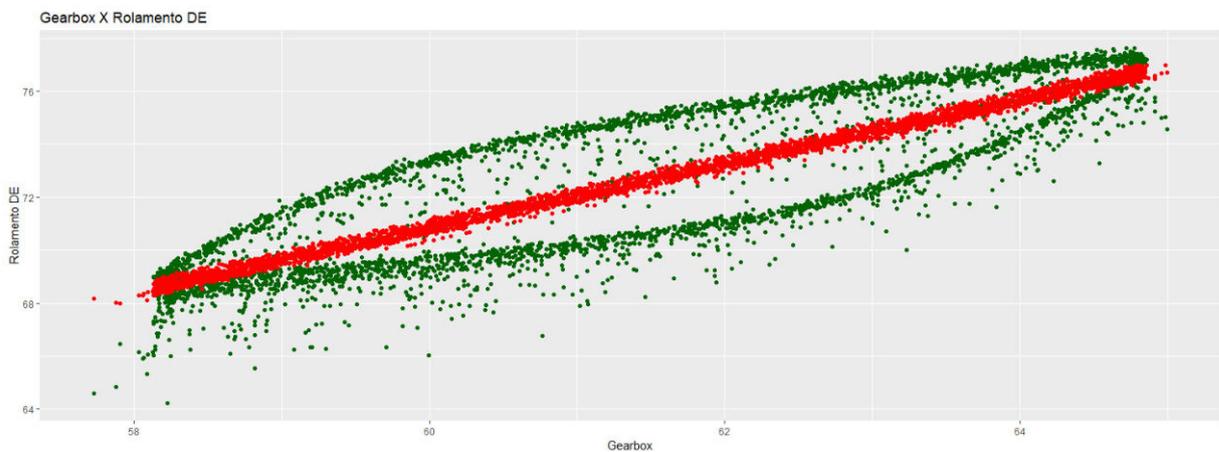
Fonte: Elaborada pelo autor

Figura 91 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Agosto).



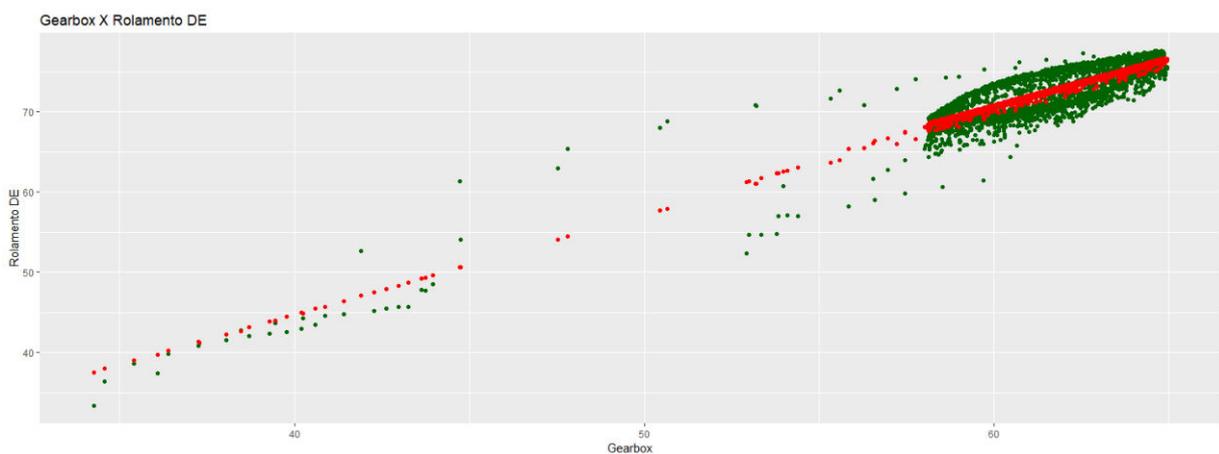
Fonte: Elaborada pelo autor

Figura 92 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Setembro).



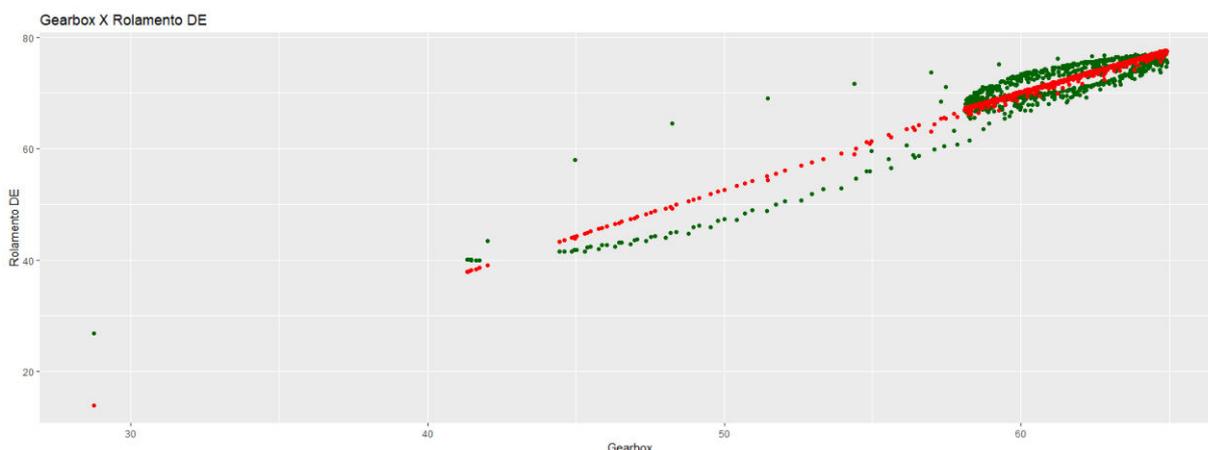
Fonte: Elaborada pelo autor

Figura 93 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Outubro).



Fonte: Elaborada pelo autor

Figura 94 – Gráfico de dispersão – temperatura rolamento2 x multiplicadora2 (Novembro).



Fonte: Elaborada pelo autor

A comparação entre os modelos em termos de coeficiente de correlação e erro médio absoluto estão apresentados na Tabela 7.

Tabela 7 – Média anual de coeficiente de correlação e erro médio absoluto.

Regressão Linear - AEROGERADOR 1		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,838	1,382
Polinomial - AEROGERADOR 1		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,849	1,365
SVM - AEROGERADOR 1		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,856	1,253
Regressão Linear - AEROGERADOR 2		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,864	1,514
Polinomial - AEROGERADOR 2		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,879	1,511
SVM - AEROGERADOR 2		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,883	1,358

Fonte: Elaborada pelo autor

Pela tabela pode-se observar que a correlação média anual entre os resultados previstos e os valores do SCADA foi sempre superior a oitenta e oito por cento e o erro médio absoluto foi sempre inferior a 1,52°C.

5. Conclusão

Dentre os métodos estudados o que apresentou melhor resultado em termos de acurácia foi o modelo de aprendizagem de máquinas utilizando a Máquina de Vetores Suporte (*Support Vector Machine*, SVM). Os outros métodos também apresentaram bons resultados uma vez que a média anual dos erros para cada máquina foi um valor baixo conforme mostrado na Tabela 7.

O coeficiente de correlação para o aerogerador 1 utilizando a Máquina de Vetores Suporte foi 1,8% maior do que a correlação utilizando a Regressão Linear e 0,7% maior do que a correlação utilizando a Regressão Polinomial. Para o aerogerador 2 o SVM obteve correlação 1,9% superior a Regressão Linear e 0,4% superior a Regressão polinomial.

O erro médio absoluto para o SVM foi 10,3% menor do que a Regressão Linear e 8,96% menor do que a Regressão Polinomial para o aerogerador 1. Para o aerogerador 2 o erro médio absoluto do SVM foi 11,49% menor do que a Regressão Linear e 11,28% menor do que a Regressão Polinomial.

É possível observar através da diminuição do coeficiente de correlação e das discrepâncias entre os pontos observados e previstos no gráfico que o comportamento dos dados é anômalo e imprevisível quando um sensor apresenta falha.

Para o presente trabalho foram utilizadas a temperatura ambiente, temperatura da multiplicadora, temperatura do rolamento do eixo principal e temperatura do rolamento DE da multiplicadora.

6. Trabalhos Futuros

Para trabalhos futuros recomenda-se que o parâmetro de potência ativa da máquina seja utilizado entre as variáveis de aprendizagem, pois essa variável carrega informações importantes e por ser função da rotação, quando utilizada para o treinamento dos modelos os resultados podem apresentar aumento da correlação e da acurácia.

Outra sugestão é a utilização de uma duplicata da base de dados avançada ou defasada em um intervalo de medição para treinamento e validação dos modelos facilitaria a adaptação dos mesmos à distribuição dos dados e possivelmente melhoraria seu coeficiente de correlação e sua acurácia por trazer informações sobre o comportamento da máquina quanto ao aumento ou diminuição da rotação.

A título de ilustração do aumento da correlação e acurácia dos modelos com a utilização da potência ativa do aerogerador no Anexo II está uma tabela de valores de correlação e erro médio absoluto das previsões utilizando o modelo de aprendizagem de máquinas, Máquina de Vetores Suporte (*Support Vector Machine*, SVM), uma vez que este obteve o melhor resultado nos testes, para comparação com os resultados obtidos nos testes utilizando o mesmo modelo sem essas variáveis.

Referências

BALL, W. W. R.A short account of the history of mathematics. Dover Publications. 1960.

Ethem Alpaydin. Introduction to machine learning — 2nd ed. The MIT press 2010.

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning: with Applications in R. New York: Springer, 2013. Print.

Introduction to Machine Learning Alex Smola and S.V.N. Vishwanathan – Disponível em: <<http://alex.smola.org/drafts/thebook.pdf>> Acesso em 05 de abril de 2019.

Machine Learning e Data Science com R de A à Z. Professor Doutor Jones Granatyr – Disponível em: <<https://www.udemy.com/machine-learning-e-data-science-com-r/>> Acesso em 25 de março de 2019.

MULLER, A.; MARQUEZ, A.C.; IUNGA, B. On the concept of e-maintenance: Review and current research. Reliability Engineering and System Safety, v. 93, p. 1165– 1187, 2008.

PALMEIRA, Jorge Nassar. Manutenção produtiva total: o caso eletronorte. 2001. 128 f. Dissertação (Mestrado Executivo) – EBAPE, Fundação Getúlio Vargas, Rio de Janeiro, 2001.

Repositório particular da Delfos Intelligent Maintenance, 2019.

Sampaio, Alan Santos Junior Estudo de um Gerador de Indução Assíncrono autoexcitado por capacitores/ Alan Sampaio Santos Junior. - Rio de Janeiro: UFRJ/ Escola Politécnica, 2015.

Anexo I

Algoritmo Regressão Linear

```

1 # Leitura do banco de dados
2 base = read.csv('data_4102_y2018m01.csv')
3
4 # Seleção das variáveis de interesse
5 base$device_id = NULL
6 base$sample_time = NULL
7 base$air_density = NULL
8 base$wind_spd = NULL
9 base$norm_wind_spd = NULL
10 base$incl_temp = NULL
11 base$gen_torque = NULL
12 base$gen_spd = NULL
13 base$bear_a_temp = NULL
14 base$bear_b_temp = NULL
15 base$temp_gen_1 = NULL
16 base$temp_gen_2 = NULL
17 base$temp_gen_cool_air = NULL
18 base$power = NULL
19
20 # Remoção de dados faltantes
21 base <- base[!(is.na(base$gbx_temp)) | !(is.na(base$gbx_bear_temp)),]
22
23 # Implementação do modelo de regressão linear
24 regressor = lm(formula = gbx_bear_temp ~ ., data = base)
25 summary(regressor)
26
27 # Exibição dos coeficientes da reta e correlação entre as variáveis
28 b0 = regressor$coefficients[1]
29 b1 = regressor$coefficients[2]
30
31 previsao = predict(regressor, newdata = base [-4])
32
33 # install.packages('miscTools')
34 library(miscTools)
35 cc = rsquared(base[['gbx_bear_temp']], resid = base[['gbx_bear_temp']] - previsao)
36
37 #install.packages('ggplot2')
38 library(ggplot2)
39 ggplot() +geom_point(aes(x = base$gbx_temp, y = base$gbx_bear_temp), colour = 'blue') +
40   geom_point(aes(x = base$gbx_temp, y = predict(regressor, newdata = base[-4])), colour = 'red') +
41   ggtitle('Gearbox X Rolamento DE') + xlab('Gearbox') + ylab('Rolamento DE')
42
43 # erro médio absoluto
44 |mean(abs(base[['gbx_bear_temp']] - previsao))
45

```

Algoritmo Regressão Polinomial

```

1 # leitura do banco de dados
2 base = read.csv('data_4101_y2018m11.csv')
3
4 # seleção de variáveis de interesse
5 base$device_id = NULL
6 base$air_density = NULL
7 base$wind_spd = NULL
8 base$norm_wind_spd = NULL
9 base$nc1_temp = NULL
10 base$gen_torque = NULL
11 base$gen_spd = NULL
12 base$bear_a_temp = NULL
13 base$bear_b_temp = NULL
14 base$temp_gen_1 = NULL
15 base$temp_gen_2 = NULL
16 base$temp_gen_cool_air = NULL
17 base$power = NULL
18 base$sample_time = NULL
19
20 scale(base)
21
22 # remoção de dados faltantes
23 base <- base[!(is.na(base$gbx_temp)) | !(is.na(base$shaft_bear_temp)),]
24
25 # instalação do pacote de machine learning
26 # install.packages('caTools')
27
28 library(caTools)
29 set.seed(1)
30
31 # regressao polinomial - aumentando a dimensão da base de dados
32 base2 = base
33 base2$sample_time = NULL
34
35 base2$amb_temp2 = base2$amb_temp^2
36 base2$amb_temp3 = base2$amb_temp^3
37 base2$amb_temp4 = base2$amb_temp^4
38 base2$amb_temp5 = base2$amb_temp^5
39 base2$amb_temp6 = base2$amb_temp^6
40
41 base2$gbx_temp2 = base2$gbx_temp^2
42 base2$gbx_temp3 = base2$gbx_temp^3
43 base2$gbx_temp4 = base2$gbx_temp^4
44 base2$gbx_temp5 = base2$gbx_temp^5
45 base2$gbx_temp6 = base2$gbx_temp^6
46
47 base2$shaft_bear_temp2 = base2$shaft_bear_temp^2
48 base2$shaft_bear_temp3 = base2$shaft_bear_temp^3
49 base2$shaft_bear_temp4 = base2$shaft_bear_temp^4
50 base2$shaft_bear_temp5 = base2$shaft_bear_temp^5
51 base2$shaft_bear_temp6 = base2$shaft_bear_temp^6
52
53
54 # divisão da base de dados
55 divisao2 = sample.split(base2$gbx_temp, SplitRatio = 0.75)
56 base_treinamento2 = subset(base2, divisao2 == TRUE)
57 base_teste2 = subset(base2, divisao2 == FALSE)
58
59 # previsão
60 regressor2 = lm(formula = gbx_bear_temp ~ ., data = base_treinamento2)
61 summary(regressor2)
62 cc2 = summary(regressor2)$r.squared
63
64 previsao2 = predict(regressor2, newdata = base_teste2[-4])
65
66 # analisando o coeficiente de correlação
67 #install.packages("miscTools")
68 library(miscTools)
69 cc = rSquared(base_teste2[['gbx_bear_temp']], resid = base_teste2[['gbx_bear_temp']] - previsao2)
70
71
72 # plotando gráficos
73 library(ggplot2)
74 ggplot() +geom_point(aes(x = base$gbx_temp, y = base$gbx_bear_temp), colour = 'blue') +
75   geom_point(aes(x = base2$gbx_temp, y = predict(regressor2, newdata = base2[-4])), colour = 'red') +
76   ggtitle('Gearbox x Rolamento DE') + xlab('Gearbox') + ylab('Rolamento DE')
77
78 # erro médio absoluto
79 mean(abs(base_teste2[['gbx_bear_temp']] - previsao2))
80

```

Algoritmo Máquina de Vetores Suporte

```

1 #1 leitura do banco de dados
2 base = read.csv('data_4102_y2018m11.csv')
3
4 #2 seleção de variáveis de interesse
5 base$device_id = NULL
6 base$sample_time = NULL
7 base$air_density = NULL
8 base$wind_spd = NULL
9 base$norm_wind_spd = NULL
10 base$incl_temp = NULL
11 base$gen_torque = NULL
12 base$gen_spd = NULL
13 base$bear_a_temp = NULL
14 base$bear_b_temp = NULL
15 base$temp_gen_1 = NULL
16 base$temp_gen_2 = NULL
17 base$temp_gen_cool_air = NULL
18 #base$power = NULL
19
20 #3 remoção de dados faltantes
21 base <- base[!(is.na(base$gbx_temp)) | !(is.na(base$gbx_bear_temp)),]
22
23
24 #4 instalação do pacote de machine learning
25 # install.packages('caTools')
26
27 library(caTools)
28 set.seed(1)
29
30 #5 divisão da base de dados em teste e treinamento
31 #divisao = sample.split(base$gbx_temp, SplitRatio = 0.75)
32 #base_treinamento = subset(base, divisao == TRUE)
33 #base_teste = subset(base, divisao == FALSE)
34
35 #6 implementação do SVM
36 # install.packages('e1071')
37 library(e1071)
38
39 regressor = svm(formula = gbx_bear_temp ~., data = base, type = 'eps-regression',
40                kernel = 'radial', cost = 1 )
41 summary(regressor)
42 previsao = predict(regressor, newdata = base [-4])
43
44 # install.packages('miscTools')
45 library(miscTools)
46 cc = rSquared(base[['gbx_bear_temp']], resid = base[['gbx_bear_temp']] - previsao)
47
48
49 # plotando gráficos
50 library(ggplot2)
51 ggplot() +geom_point(aes(x = base$gbx_temp, y = base$gbx_bear_temp), colour = 'darkgreen') +
52   geom_point(aes(x = base$gbx_temp, y = predict(regressor, newdata = base[-4])), colour = 'red') +
53   ggtitle('Gearbox x Rolamento DE') + xlab('Gearbox') + ylab('Rolamento DE')
54
55 # erro médio absoluto
56 mean(abs(base[['gbx_bear_temp']] - previsao))

```

Anexo II

SVM – Utilizando a Potência ativa – Aerogerador 1.

SVM (Power) - AEROGERADOR 1		
Mês	Coeficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,906	1,041
Fevereiro	0,956	0,769
Março	0,935	0,822
Abril	0,964	0,700
Maio	0,972	0,761
Junho	0,913	0,905
Julho	0,895	0,951
Agosto	0,875	1,170
Setembro	0,684	1,197
Outubro	0,802	1,170
Novembro	0,698	1,242
Média	0,906	0,951

Fonte: Elaborada pelo autor

SVM – Utilizando a Potência ativa – Aerogerador 2.

SVM (Power) - AEROGERADOR 2		
Mês	Coeficiente de correlação (%)	Erro médio absoluto (°C)
Janeiro	0,942	0,851
Fevereiro	0,966	0,807
Março	0,947	0,803
Abril	0,987	0,778
Maio	0,972	0,781
Junho	0,929	0,853
Julho	0,076	11,505
Agosto	0,017	33,722
Setembro	0,669	1,278
Outubro	0,829	1,250
Novembro	0,940	1,309
Média	0,940	0,853

Fonte: Elaborada pelo autor

SVM – Comparação da média utilizando a potência e sem utilizar a potência.

SVM(Power) - AEROGERADOR 1		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,906	0,951
SVM(Power) - AEROGERADOR 2		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,940	0,853
SVM - AEROGERADOR 1		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,856	1,253
Regressão Linear - AEROGERADOR 2		
Média anual	Coeficiente de correlação (%)	Erro médio absoluto (°C)
	0,883	1,358

Fonte: Elaborada pelo autor