

NORMALIZAÇÃO DE UM BANCO DE DADOS PARA TREINAMENTO DE APRENDIZADO DE MÁQUINA: APLICAÇÃO SOBRE OS EFEITOS DE VARIÁVEIS NA PRECISÃO DA SEMEADURA

RAFAEL DOS SANTOS ESTECHE¹; JEAN PIETRO COLET DE CARLI²;
MATEUS AUGUSTO THEODORO RODRIGUES³; MARLON SOARES
SIGALES⁴; ÂNGELO VIEIRA DOS REIS⁵; MATEUS BECK FONSECA⁶

¹UNIVERSIDADE FEDERAL DE PELOTAS – rafael.esteche@ufpel.edu.br

²UNIVERSIDADE FEDERAL DE PELOTAS – jean.carli@ufpel.edu.br

³UNIVERSIDADE FEDERAL DE PELOTAS – a.t.r_mateus@outlook.com

⁴UNIVERSIDADE FEDERAL DE PELOTAS – marlon.sigales@ufpel.edu.br

⁵UNIVERSIDADE FEDERAL DE PELOTAS – areis@ufpel.edu.br

⁶UNIVERSIDADE FEDERAL DE PELOTAS – mateus.foseca@ufpel.edu.br

1. INTRODUÇÃO

Existe a estimativa que em 2050 a população mundial irá chegar a 10 bilhões de pessoas. Como decorrência deste crescimento exponencial, acarretará vários desafios à sociedade, um deles será o fornecimento de alimentos para segurança alimentar mundial, para isso a produção de alimentos deve aumentar em 50% até essa data (FAO, 2021).

Neste contexto, o aumento da produção de alimentos, de forma sustentável, torna-se um desafio para a agricultura no que se refere a otimização de práticas de manejos de cultura. Com o objetivo de aproveitar ao máximo a área disponível para obter maior produção, a semeadura de precisão é essencial para garantir que as sementes sejam distribuídas uniformemente, de acordo com as recomendações agrônômicas. Esse implemento agrícola possui um mecanismo especializado para esta função, chamado dosador de sementes (Andrade, 2022).

Dosadores de diferentes tecnologias e metodologias são empregados no processo, cada qual possui precisão de semeadura e coeficientes de variação próprios. Métricas para comparação e averiguação da qualidade destes dispositivos sobre os fatores que afetam sua precisão são utilizadas baseadas em técnicas estatísticas.

O estudo de Andrade (2022), gerou um extenso banco de dados de intervalos temporais entre ejeções de sementes efetuados por três modelos diferentes de dosadores de sementes, obtidas variando em diversas velocidades de semeadura, inclinações longitudinais e presença ou ausência de tubo condutor de sementes, com parcelas experimentais com 5 repetições de 250 amostras cada.

O presente trabalho está sendo desenvolvido como parte inicial do projeto Instrumentação eletrônica aplicada à agricultura 4.0, desenvolvido pelo Grupo de Instrumentação Eletrônica do CENg-UFPEL e tem o objetivo de realizar a normalização desse banco de dados, antes organizado para ser analisado por meios de comparações e testes estatísticos, para uma forma que seja possível compreender os fatores de influência sobre a precisão dos dosadores de sementes, por meio de aprendizado de máquina.

2. METODOLOGIA

Em posse dos dados de Andrade (2022), que estavam organizados por nome, em que cada arquivo apresentava os fatores da parcela experimental, sendo eles a presença ou não de tubo condutor, o ângulo de inclinação longitudinal do dosador e a velocidade simulada do implemento, e internamente constavam cinco colunas representando as repetições aleatorizadas com 250 amostras cada. Estes arquivos foram coletados de 3 dosadores de tecnologia pneumática presentes no mercado, com variação no fator inclinação presentes no conjunto [-11, -5, 0, 5, 11] graus, velocidades simuladas compreendidas no intervalo de 5 a 20 km/h com variação de 1km/h entre cada parcela experimental, e a presença ou não de tubo (ANDRADE et al., 2019) e coletados através do software (ANDRADE et al., 2021).

O processo de normalização dos dados se iniciou com a criação de um arquivo contendo a informação de todos as parcelas experimentais, separando em colunas os dados dos fatores de velocidade, ângulo e presença de tubo condutor, estas foram agregadas de maneira que cada linha correspondesse a coleta de todas as amostras para aqueles fatores, além disto foram feitas as médias da taxa de sementes aceitáveis, de falhas e múltiplas.

De acordo com as metodologias de ensaio de dosadores de semente de Coelho (1996), apresentado na Tabela 1, definida através da distância os espaçamentos aceitáveis, falhos e múltiplos, adaptou essa metodologia para trabalhar com o tempo, conforme com o tempo de referência (T_{ref}), encontrado a partir da distância de referência (X_{ref}) agrônômica para aquela velocidade (v) em questão, através da equação $\Delta T_{ref} = X_{ref} \cdot 3,6 \cdot (v)^{-1}$. Os dados foram agrupados em razões percentuais médias de aceitáveis quando ($0,5 \cdot T_{ref} \leq T_{amostra} \leq 1,5 \cdot T_{ref}$), de múltiplas ($0,5 \cdot T_{ref} > T_{amostra}$) e de falhas ($1,5 \cdot T_{ref} < T_{amostra}$), além disso foram apresentados os coeficientes de variação destas médias, para que se possam conferir os desvios das médias.

Tabela 1: Classificação dos tipos de espaçamentos entre sementes de acordo com os intervalos de variação de distâncias.

Tipo de espaçamento	Intervalo de tolerância para variação de X_i
Múltiplos	$X_i < 0,5 \cdot X_{ref}$
Aceitáveis	$0,5 \cdot X_{ref} < X_i < 1,5 \cdot X_{ref}$
Falhos	$X_i > 1,5 \cdot X_{ref}$

3. RESULTADOS E DISCUSSÃO

O cabeçalho da tabela normalizada é demonstrado na Figura 2. Nela são apresentados os fatores analisados em conjunto com os resultados influenciados por eles nas colunas nomeadas “somatório Aceitável”, “somatório Falho”, “somatório múltiplo” e “coeficiente de variação”.

tubo	angulo(graus)	velocidade(km.h-1)	somatorioAceitavel(%)	somatorioMultiplo(%)	somatorioFalho(%)	CV_Aceitavel(%)	CV_Multiplo(%)	CV_Falho(%)	
127	s	0	20	96.64	2.40	0.96	0.110871	0.536433	1.128274
31	c	-5	20	82.88	8.56	8.56	0.110871	0.536433	1.128274
29	c	-5	18	87.12	6.32	6.56	0.110871	0.536433	1.128274
12	c	-11	17	88.56	5.44	6.00	0.110871	0.536433	1.128274
27	c	-5	16	86.00	8.24	5.76	0.110871	0.536433	1.128274
84	s	-11	9	86.64	2.96	10.40	0.110871	0.536433	1.128274
124	s	0	17	96.96	2.08	0.96	0.110871	0.536433	1.128274
23	c	-5	12	86.80	6.72	6.48	0.110871	0.536433	1.128274
142	s	11	19	95.60	2.80	1.60	0.110871	0.536433	1.128274
72	c	5	13	85.28	4.56	10.16	0.110871	0.536433	1.128274
134	s	11	11	96.96	2.40	0.64	0.110871	0.536433	1.128274
5	c	-11	10	84.88	7.44	7.68	0.110871	0.536433	1.128274
42	c	0	15	96.24	3.12	0.64	0.110871	0.536433	1.128274

Figura 2- Tabela Normalizada

Os resultados das análises estatísticas obtidos por Andrade (2021) serão adicionados à tabela para fim de treinamento futuro de um aprendizado de máquinas que aprenderá a analisar os dados através dos fatores de entrada com os fatores de saída e o coeficiente de variação, tentando identificar uma lei que esboce os efeitos de cada fator sobre as métricas de desempenho.

Para isso estes dados serão expostos a diversos modelos de aprendizado de máquinas, como Random Forest, Árvore de decisão, Gradient Boosting. Verificando sua aproximação aos testes estatísticos iniciais, e desta forma apresentando um padrão de resposta modelo, ao qual será submetido um banco de dados com respostas conhecidas não submetido ao treinamento, para verificação.

4. CONCLUSÕES

A tabela normalizada foi gerada com base nos dados do experimento de Andrade et al. (2022), o que facilita a compreensão das influências dos fatores no padrão de deposição de sementes dos dosadores analisados, mesmo que visualmente. Espera-se que o aprendizado de máquina consiga identificar os padrões que influenciam nos diferentes dosadores, e desta forma os fabricantes possam aprimorar seu equipamento, tornando-os mais competitivos e entregando melhores equipamentos aos agricultores de um modo geral, elevando produtividades e reduzindo desperdícios que podem ser evitados através de demonstrações de estudos como esse e os que serviram de base para este conduzidos pelo Núcleo de Inovação em Maquinas e Equipamentos Agrícolas da Universidade Federal de Pelotas (NIMEq-UFPEL).

5. REFERÊNCIAS BIBLIOGRÁFICAS

FAO. Desenvolvimento sustentável nas indústrias de semestres é vital. FAO Food and Agricultural Organization of the United Nations. Roma, 2021. Acessado em 16 julh. 2023. Disponível em: <https://news.un.org/pt/story/2021/11/1769342>

ANDRADE, Henrique Gonçalves. Efeito de velocidade, inclinação longitudinal e sistema de condução de sementes em dosadores pneumáticos. Faculdade de Agronomia Eliseu Maciel, Universidade Federal de Pelotas, Pelotas, 2022. Acessado em 21 agos 2023.

Disponível

em:https://guaiaca.ufpel.edu.br/bitstream/handle/prefix/8870/dissertacao_henrique_andrade.pdf?sequence=1&isAllowed=y

Coelho, J.L.D. Ensaio e certificação das máquinas para semeadura. In: MIALHE, L. G. Máquinas agrícolas: ensaios e certificação. Piracicaba: FEALQ, 1996. Cap. 11, p. 551-570.

ANDRADE, H. G.; SIGALES, M. S.; ARAUJO, A. S.; WALKER, E.; REIS, A. V. Desenvolvimento de um sensor para contar o número e o intervalo de tempo de sementes ejetadas por dosadores de semeadoras. In: **XXVIII CONGRESSO DE INICIAÇÃO CIENTÍFICA E X SEMANA INTEGRADA UFPEL**, 2019, Pelotas. Anais. . . [S.l.: s.n.], 2019.

ANDRADE, H. G.; SIGALES, M. S.; CASELATTO M. R. A.; HIDACA G.; WALKER Eduardo; REIS, A. V. Feramenta Computacional para Auxílio na coleta e tratamento de dados de distribuição espacial de sementes dosadas em bancada. In: **XXV ENPÓS Encontro de Pós-Graduação UFPEL**, 2021, Pelotas. Anais