

**MÁQUINA DE VETORES DE SUPORTE PARA ESTIMAR A PRODUTIVIDADE DA SOJA**

Gabriela Karoline Michelin¹, Paulo Lopes de Menezes², Arnaldo Candido Junior³, Claudio Leones Bazzi⁴ & Marcela Marques Barbosa⁵

1 - Cientista da Computação, Mestranda na UTFPR/Medianeira-PR, gabrielamichelon@alunos.utfpr.edu.br

2 - Doutor em Engenharia Agrícola, Professor e Coordenador do Programa de Pós-Graduação PPGTA da UTFPR, plopes@utfpr.edu.br

3 - Doutor em Ciências da Computação e Matemática Computacional, Professor e Coordenador do Curso e Ciência da Computação da UTFPR, arnaldoc@utfpr.edu.br

4 - Doutor em Engenharia Agrícola, Professor e Diretor de Graduação e Educação Profissional da UTFPR/Medianeira-PR, bazzi@utfpr.edu.br

5 - Cientista da Computação, Mestranda na UTFPR/Medianeira-PR, marcelamarques.br@gmail.com

Palavras-chave:

inteligência artificial
nutrientes foliares
regressão

RESUMO

A soja é uma das principais oleaginosas da produção agrícola mundial, sendo muito utilizada para alimentação humana e animal, para produtos industrializados e ainda como fonte alternativa de biocombustível. É por suas grandes utilizações que é muito valorizada e cultivada no mundo. Portanto, este trabalho buscou aplicar uma técnica de inteligência artificial para prever a produtividade da soja e, conseqüentemente, maximizar a produção de uma área cultivável, aumentar o lucro do produtor e diminuir impactos ambientais. Utilizou-se, então, a técnica da máquina de suporte de vetores para buscar um modelo de predição da produtividade da soja por meio de dados de macro nutrientes presentes na folha da soja, permitindo assim, que sejam realizadas adubações somente nos locais necessários, preditos como pontos de baixa produtividade pelo melhor modelo de máquina de suporte de vetores obtido. Dentre todos os modelos criados, o melhor modelo de predição da produtividade conseguiu explicar 58% dos dados observados com as variáveis de nitrogênio, fósforo, potássio, cálcio e magnésio. Buscando utilizar menos variáveis para tornar a prática mais acessível ao produtor, aplicou-se um algoritmo que faz a seleção de variáveis de entrada para a máquina de suporte de vetores. Como resultado da utilização de menos variáveis de entrada, obteve-se um modelo que utilizou somente os dados de nitrogênio, fósforo e cálcio, sendo um modelo de predição menos custoso, pois com apenas três variáveis, explicou 58% dos dados observados, assim como o modelo que utilizou todas as variáveis.

Keywords:

artificial intelligence
foliar nutrients
regression

SUPPORT VECTOR MACHINE TO ESTIMATE THE SOYBEAN YIELD**ABSTRACT**

Soybean is one of the most highly valued and cultivated oleaginous in worldwide agriculture, due to its important use as a source of food, a resource to generate industrialized products, and an alternative source of biofuel. Therefore, this study sought to apply an artificial intelligence technique to predict soybean yield, and, therefore, maximize production from farmlands, increase the profit for the producer and reduce environmental impacts. Support vector machine technique was used to find a prediction model for soybean yield from macro nutrient data present in the soybean leaf, allowing fertilization to be carried out only in necessary locations predicted by best support vector machine model obtained to have low productivity. Among all the models created, the best prediction model of productivity was able to explain 58% of the data observed with nitrogen, phosphorus, potassium, calcium and magnesium variables. Seeking to use fewer variables and make the practice more accessible to the producer, an algorithm was applied in order to select input variables for the support vector machine. As result of the use of fewer input variables, a model that used only the data of nitrogen, phosphorus and calcium was obtained; this prediction model was found to be less costly because it explained with only three variables 58% of the data observed as well as the model that used all variables.

INTRODUÇÃO

A soja (*Glycine max*) é uma das oleaginosas mais importantes na produção agrícola do mundo devido à sua alta qualidade nutricional, como em proteínas e aminoácidos e por possuir um maior rendimento em relação a outras culturas (GANDHI, 2009). O uso de produtos criados a partir da soja para alimentação humana e animal e produtos industriais vêm crescendo a cada ano e, para isso, torna-se necessário buscar técnicas que auxiliem na maximização da produção de uma área agrícola. Para se ter uma ideia, a produção mundial de soja na safra 2015/2016 foi de aproximadamente 313,01 milhões de toneladas em uma área de 119,94 milhões de hectares (USDA, 2016).

Há técnicas de agricultura de precisão (AP) e técnicas computacionais que podem auxiliar no aumento da produção da soja e na exploração das características do solo e da planta. Lacerda et al. (2015) apontam para a importância de entender quais os nutrientes mais relevantes no desenvolvimento da soja, pois seus teores influenciam no crescimento e podem servir como auxiliares da previsão da produtividade.

Entre os nutrientes mais absorvidos pela planta da soja estão os macronutrientes, como: Nitrogênio (N), Fósforo (P), Potássio (K), Cálcio (Ca), Magnésio (Mg) e Enxofre (S). Assim, esse trabalho buscou obter um modelo computacional de previsão da produtividade da soja explorando as variáveis N, P, K, Ca e Mg. Para isso, utilizou-se máquina de suporte de vetores (SVM), que dentre as várias técnicas computacionais que podem ser utilizadas para estimar valores ou aprender padrões, é uma técnica conhecida como uma das melhores para resolver problemas de classificação e regressão (BOUBOULIS et al., 2015; CHANG & LIN, 2011; KARAMIZADEH et al., 2014).

A máquina de vetores de suporte é um algoritmo de aprendizado que se baseia na teoria de aprendizagem, o qual é um processo de indução de hipóteses a partir de uma experiência. A capacidade de aprendizado é considerada essencial para um comportamento inteligente. O aprendizado pode ser relacionado com a capacidade de memorizar, observar e explorar situações para aprender fatos, melhorar a capacidade cognitiva/motora e

organizar o conhecimento em representações úteis (FACELI et al., 2011).

O treinamento de uma máquina de vetores de suporte consiste em encontrar o hiperplano ótimo, isto é, aquele com a distância máxima dos padrões de treinamento mais próximos. Os vetores de suporte são os padrões (mais próximos) a uma distância determinada do hiperplano.

Alguns trabalhos utilizando APE Inteligência Artificial (IA) já foram desenvolvidos em várias culturas, como os trabalhos de Bocca (2014) e Rub (2009), os quais declaram ter tido melhores desempenhos utilizando SVMs do que as demais técnicas de estudo, para prever a produtividade da cana-de-açúcar e trigo, respectivamente.

Portanto, o objetivo desse trabalho foi aplicar a técnica de IA – máquina de suporte de vetores, para obter um modelo de previsão da produtividade da soja por meio dos nutrientes presentes na folha, buscando contribuir com a antecipação dos resultados da produtividade dos grãos. Dessa forma, possíveis deficiências de nutrientes na planta poderão ser corrigidas com adubação de cobertura na planta em um estágio inicial do desenvolvimento da cultura que, segundo Sfredo (2008), quando houver necessidade de aplicação de nutrientes para as plantas, deve-se efetuar-la antes do início da floração. Caso o modelo de SVM seja utilizado para estágios finais do desenvolvimento deve se realizar a adubação e calagem no solo após colheita para melhorar o solo para o próximo plantio, ou até mesmo, a adubação de base no momento do plantio. Dessa forma, pode se ter um aumento da produtividade e melhores rendimentos ao produtor e também auxiliar na diminuição da degradação ambiental.

MATERIAIS E MÉTODOS

Para atingir o objetivo de prever a produtividade da soja com uma SVM, foi necessário ter os dados dos nutrientes da folha da soja e os respectivos dados de produtividade. Com os dados em mãos, criou-se um programa computacional que fizesse o treinamento da SVM inserindo os dados desejados.

Os nutrientes analisados nesse trabalho foram coletados em 3 estágios do desenvolvimento da soja em duas áreas experimentais (Talhão 1, com

10 ha e Talhão 2, com 23,8 ha, Figura 1), com solos classificados como latossolo vermelho distroférico típico (FIDALSKI *et al.*, 2013), ambas cultivadas sob sistema de plantio direto por pelo menos 10 anos com sucessão de cultura soja/milho para fins comerciais.



Figura 1. Áreas experimentais do estudo.

As áreas estão localizadas no município de Serranópolis do Iguaçu/PR, sob coordenadas geográficas 25°24'28" S e 54°00'17" O com elevação média de 355 m. Foram determinados, por meio do software Pathfinder, duas grades amostrais irregulares, contendo 42 pontos amostrais no talhão 1 e 73 no talhão 2, com densidade de aproximadamente 3 pontos por ha.

A coleta dos dados foi realizada de forma manual no ano de 2012, em 3 estádios da safra da soja 2012/2013, V3, V6 e R1, onde o estádio V3 é o estádio vegetativo da planta com 3 folhas, o V6 é o estádio da planta com 6 folhas e o estádio reprodutivo da planta já com flores, chamado de R1 (WOLLENHAUPT *et al.*, 1997). Os pontos foram localizados por meio de um receptor GPS (GPS Trimble Pathfinder Pro XT). Para cada ponto amostral em um raio de 3 metros, coletou-se 12 folhas (WOLLENHAUPT *et al.*, 1997).

As folhas foram acondicionadas em sacolas de papel e posteriormente foram secas em estufa a 105°C. Após a secagem, as folhas foram debulhadas por um moinho de facas modelo SL 31 da marca SOLAB. Ao final de todo este processo as folhas foram encaminhadas para análise laboratorial dos

macronutrientes Ca, Mg, P, K, N, no Laboratório de Solos da Universidade Tecnológica Federal do Paraná (UTFPR) - Câmpus Pato Branco em convênio com o Instituto Agrônomo do Paraná (IAPAR), os quais seguem a metodologia de Tedesco *et al.* (1995) para a determinação de macronutrientes em tecido vegetal.

Para determinar a produtividade da soja 2012/2013 também foi realizada coleta de forma manual, nos meses de janeiro/fevereiro, coletando-se para cada ponto amostral, duas linhas de plantio, por um percurso de 1 m, representando uma porção de cerca de 0,9 m². Após a coleta, foi realizada a debulha manual e realizou-se o procedimento de secagem em estufa a 105° para posteriormente ser calculada a produtividade em toneladas por hectare (ton ha⁻¹), considerando uma umidade de grãos de 12%.

Para aplicar a técnica de Inteligência Artificial foi necessário realizar a normalização dos dados e a modelagem de acordo com o formato requerido por cada ferramenta a ser utilizada. Portanto, normalizou-se os dados entre 0 e 1, de acordo com a Equação 1 (COUTINHO *et al.*, 2016) antes de aplicá-los às ferramentas computacionais.

$$f(x) = y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Onde y é o valor normalizado correspondente ao valor original x , e x_{\min} e x_{\max} são os valores de mínimo e de máximo entre todos os valores da base de dados e y_{\min} corresponde a 0 e y_{\max} corresponde a 1.

Após a normalização, os dados foram formatados de acordo com a ferramenta utilizada, Waikato Environment for Knowledge Analysis (WEKA). Para utilizar a ferramenta de desenvolvimento NetBeans com a Application Programming Interface (API) da WEKA, os dados foram formatados como arquivos ARFF. Para obter o modelo de SVM, utilizou-se os parâmetros utilizados como padrão no treinamento do algoritmo SMOReg da API do software WEKA. O algoritmo de treinamento utilizou em todos os testes realizados o kernel polinomial.

Os dados utilizados como entrada no programa computacional de treinamento da SVM, foram

dos nutrientes N, P, K, Ca e Mg e o valor de saída foi o da produtividade da soja. O conjunto de treinamento representou uma porção de 60% do conjunto total, enquanto o conjunto de validação representou 30% dos dados e o conjunto de teste, os outros 10% restantes (SILVA et al., 2010).

O conjunto de validação foi usado na SVM logo após o treinamento, para verificar se a técnica aprendeu como obter a saída (produtividade da soja) por meio das entradas. O conjunto de teste só foi utilizado após escolher a rede com menor erro para somente aplicar as entradas deste conjunto no algoritmo e após verificar se as saídas estimadas são significativamente semelhantes ou não.

A estratégia utilizada para determinar o momento de parar o treinamento das técnicas de inteligência artificial foi a do treinamento com parada antecipada (HAYKIN, 2001), a qual permite comparar a cada ciclo de treinamento o valor estimado pela SVM com o valor observado da produtividade da soja. A partir disso, obtém-se um erro. Se o erro for tão baixo quanto aceitável, a SVM treinada foi satisfatória e encerram-se os treinamentos, caso contrário, deve ser continuado o treinamento até que se obtenha o erro esperado, ou até que o erro de validação comece a aumentar. O erro de validação analisado nas SVMs foi o Erro Quadrático Médio (EQM).

Utilizou-se o algoritmo CfsSubsetEval, técnica de seleção de variáveis disponibilizada pelo software WEKA, em que CFS remete à seleção de atributos baseada na correlação. Este algoritmo acopla uma fórmula de avaliação com uma medida de correlação e uma estratégia de busca heurística (SUBRAHMANYAM et al., 2013).

Segundo Subrahmanyam et al. (2013), a Correlation based Feature Selection (CFS) consegue identificar facilmente atributos redundantes e quais os atributos mais relevantes de uma base de dados.

Para avaliar o modelo baseado nas técnicas de inteligência artificial resultante deste trabalho utilizou-se: teste de normalidade do grupo de dados de teste para confirmar se os testes estatísticos paramétricos eram aplicáveis neste estudo; teste de hipóteses H_0 e H_1 para testar se os dados observados e os estimados poderiam ser considerados provenientes de uma mesma população (o que é esperado); coeficiente de determinação R^2 , o qual

reflete a qualidade do ajuste do modelo sendo possível verificar o quanto o modelo estimado se ajustou aos dados observados; o coeficiente de correlação de Pearson (R^P) para quantificar a correlação dos dados observados e estimados (dados de teste).

RESULTADOS E DISCUSSÃO

Realizou-se os testes de normalidade dos dados dos nutrientes e das produtividades pelos testes de Anderson - Darling, Kolmogorov - Smirnov, Shapiro - Wilk, Ryan - Joiner. Para os atributos, nitrogênio, potássio e cálcio, não foi possível verificar normalidade ao nível de 5% ($\alpha = 0.05$) de significância. Portanto, buscou-se transformar os dados não normais, segundo os testes de normalidade aplicados, para aplicar a estatística paramétrica na avaliação dos resultados. Os tipos de transformações realizadas foram a logarítmica, de raiz quadrada dos dados e também da divisão de $1/x$ (chamada de transformação do quociente).

Porém, como resultado, os testes de normalidade dos dados transformados pelas três formas de transformação, citadas acima, apresentaram valores de p mais distantes do nível de 5% de significância do que os valores de p obtidos antes da transformação dos dados.

Portanto, rejeita-se a hipótese de que os dados dos nutrientes seguem uma distribuição normal. A não normalidade dos dados deve-se à deficiência ou excesso de nutrientes ou de elementos químicos nocivos à planta. Desse modo, considerando que a normalidade dos dados não pode ser assumida, os testes de avaliação das máquinas de suporte de vetores a serem apresentados foram baseados em métodos estatísticos não paramétricos.

A Tabela 1 possui os resultados obtidos do modelo de SVM implementado com os parâmetros padrão de acordo com cada coleta realizada nos estádios V3 (estado vegetativo da planta da soja com três folhas), V6 (estado vegetativo da planta da soja com seis folhas), R1 (planta da soja no estado reprodutivo com flor) denominados neste trabalho como coleta 1, coleta 2 e coleta 3, respectivamente, dos nutrientes da soja.

Tabela 1. Resultado dos erros EQM, coeficiente de determinação e de correlação de Pearson para as melhores SVMs de cada coleta dos nutrientes da folha da soja (pontos amostrais aleatórios para conjunto de treinamento, validação e teste em cada coleta)

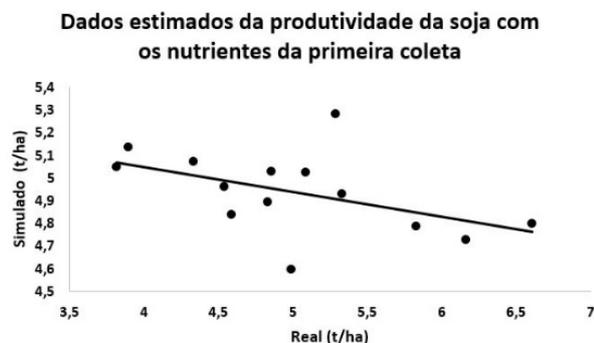
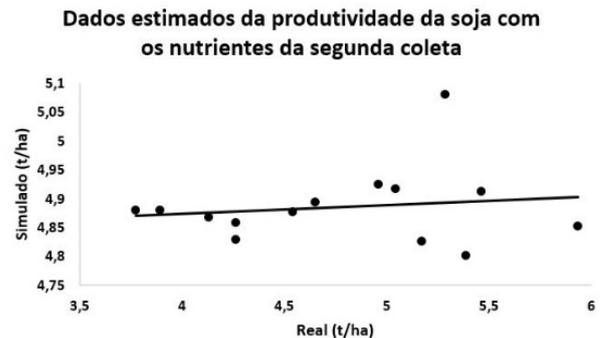
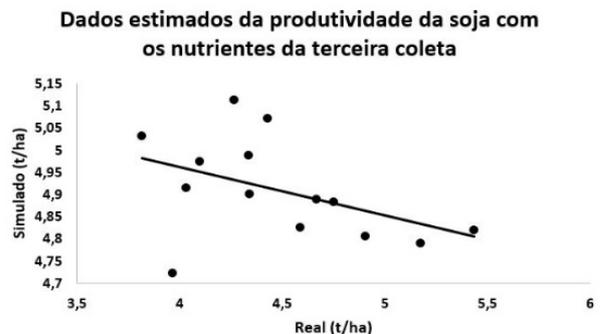
Coleta	Kernel	EQM	R ²	R ^P
1	Polinomial	0,1552	0,2342	-0,4839
2	Polinomial	0,0164	0,0227	0,1506
3	Polinomial	0,0253	0,1969	-0,4437

EQM-Mean Squared Error; R²-Coeficiente de Determinação; R^P-Coeficiente de Correlação de Pearson.

O menor EQM entre os três menores erros encontrados na fase de validação da SVM para cada coleta foi o da SVM da segunda coleta dos nutrientes da soja. No entanto, ter o menor EQM não significa ter o melhor coeficiente de determinação, pois o melhor, ou seja, maior coeficiente de determinação foi o da SVM da primeira coleta, que também teve o melhor índice de correlação dos dados observados e estimados, apesar de ser um modelo que explica pouco os dados observados, pois somente 23% dos dados previstos se ajustaram aos observados, e a correlação ser uma correlação negativa menor que 50%.

No entanto, a SVM da primeira coleta foi o modelo de predição dos dados dentre as três SVMs das três coletas, que conseguiu explicar os dados em maior percentual da variação dos dados, pois a terceira coleta obteve um modelo capaz de explicar aproximadamente 20% da variação dos dados originais, e a segunda coleta explicou apenas 2% dos dados observados.

Para melhor visualização dos resultados, as Figuras 2, 3 e 4 mostram a inclinação da reta da regressão linear dos dados observados e estimados da produtividade da soja em toneladas por hectare.

**Figura 2.** Resultado dos valores estimados da produtividade da soja com a SVM obtida com os dados da primeira coleta de nutrientes da folha.**Figura 3.** Resultado dos valores estimados da produtividade da soja com a SVM obtida com os dados da segunda coleta de nutrientes da folha.**Figura 4.** Resultado dos valores estimados da produtividade da soja com a SVM obtida com os dados da terceira coleta de nutrientes da folha.

Analisando as Figuras 2, 3 e 4, percebe-se que a Figura 2 tem a melhor relação de inclinação e proximidade dos pontos na reta de regressão, a qual representa o modelo de SVM, que explicou em maior percentual os dados observados e teve maior correlação dos dados estimados com os observados. A segunda coleta obteve uma inclinação muito pequena, enquanto a terceira coleta, apesar de uma inclinação mediana, teve pontos muito dispersos

da reta, mostrando a grande diferença dos valores estimados dos valores observados.

Na primeira coleta, quando a planta estava no estágio inicial, os dados das folhas possuíam maiores informações nutricionais, pois segundo Sfredo (2008), quanto mais velha a planta menor a absorção de nutrientes da folha, pois os nutrientes são transmitidos para as flores e vagens da planta. Portanto, o modelo pode ter maior relação dos macronutrientes com a produtividade em estágios iniciais, o que pode ter sido um dos motivos pelos quais a SVM da primeira coleta tenha explicado em maior percentual os dados observados.

No entanto, ter modelos que melhor se ajustam aos dados observados em estágios iniciais é mais adequado para realizar correção do solo na cobertura da soja já plantada, pois é em estágios anteriores ao de florescimento da planta que se deve realizar a adubação a tempo de corrigir deficiências ou excessos de nutrientes e melhorar o rendimento da safra momentânea (SFREDO, 2008).

Na Tabela 2, verifica-se que as SVMs das três coletas dos nutrientes da soja obtiveram p-valor > 0.05 , o que significa que o teste de Kruskal-Wallis aceita a hipótese de homogeneidade entre os dados observados e estimados ao nível de 5% de significância.

Pela classificação de Chowdhury et al.(2015), a correlação dos dados observados e estimados da SVM implementada com os dados da primeira e terceira coleta dos nutrientes da folha da soja é uma correlação linear moderada. Enquanto a SVM da segunda coleta possui correlação fraca.

O modelo de SVM da terceira coleta dos nutrientes da folha da soja por pouco não rejeita a hipótese de que os dados observados e estimados vêm de populações semelhantes. O valor crítico foi analisado na Tabela do qui-quadrado utilizada para testes de Kruskal-Wallis, e corresponde a 3,565, deste modo como valor de H é menor que o valor crítico nas coletas 1 e 2, não há diferenças significativas entre os grupos (dados observados e estimados), e para coleta 3, apesar de os dados poderem ser considerados de uma mesma população, as diferenças entre eles são significativas.

Realizou-se testes para as coletas 1, 2 e 3 separando os pontos amostrais das áreas A e B, devido as variabilidades espaciais que uma área pode ter e a outra não, causando influência nos testes. Portanto, realizou-se testes com SVMs para as coletas 1, 2 e 3 para a área A e B separadas. Verifica-se os resultados dos dados das áreas A e B separados na Tabela 3.

Tabela 2. Testes de homocedasticidade e semelhança de população entre os dados observados e estimados pela melhor SVM de cada coleta dos nutrientes da folha da soja

Kruskal-Wallis		
	H	P-valor
C1	0,0021	0,9633
C2	0,0084	0,9267
C3	7,8557	0,0051

C1 – Coleta 1; C2 – Coleta 2; C3 – Coleta 3; H – valor da estatística de Kruskal-Wallis; p-valor- é a probabilidade de significância.

Tabela 3. Resultado dos erros EQM, coeficiente de determinação e de correlação de Pearson para as melhores SVMs de cada coleta dos nutrientes da folha da soja somente na área A(mesmos pontos amostrais para conjunto de treinamento, validação e teste em cada coleta)

Área	Coleta	Kernel	EQM	R ²	R ^P
A	1	Polinomial	0,0262	0,3759	-0,6131
	2	Polinomial	0,0293	0,5868	-0,7660
	3	Polinomial	0,0247	0,2359	0,4856
B	1	Polinomial	0,0211	0,2290	0,4785
	2	Polinomial	0,0487	0,3799	0,6163
	3	Polinomial	0,0545	0,0824	0,2870

EQM-Mean Squared Error; R² - Coeficiente de Determinação; R^P-Coeficiente de Correlação de Pearson.

Os resultados foram satisfatórios para a área A (Tabela 3), pois obteve-se modelos de SVMs que explicaram de 23% a 58% da variação total dos dados observados da produtividade da soja, com destaque para a segunda coleta, a qual obteve maior coeficiente de correlação, mostrando uma correlação negativa mas classificada como forte de acordo com classificação de Chowdhury *et al.* (2015).

Para a área B, obteve-se coeficientes de determinação mais baixos, variando de 0,08 a 0,3799 e coeficientes de correlação de fraca a forte correlação dos dados observados e estimados. Assim como na primeira coleta da área A, a área B obteve na segunda coleta uma semelhança dos dados observados e estimados de aproximadamente 38% e uma forte correlação entre eles.

Os modelos da área A rejeitaram a hipótese de os dados observados e estimados serem homogêneos ao nível de 5% de significância pelo teste de Kruskal-Wallis e analisando o valor de H que é maior que o valor crítico do qui-quadrado (3,565), as diferenças entre os grupos (dados observados e estimados) são significativas. Portanto, os dados observados e estimados não podem ser considerados provenientes de uma mesma população.

Os resultados dos testes de Kruskal-Wallis para a área B apresentaram, em todas as coletas, homogeneidade dos dados observados e estimados ao nível de 5% de significância e verifica-se, por meio do valor de H, que as diferenças entre os conjuntos de dados observados e estimados não são significativas.

Analisando os resultados obtidos das SVMs

que utilizaram as variáveis selecionadas pela técnica CfsSubsetEval na área A, foram um pouco inferiores aos resultados das SVMs que utilizaram todos os atributos coletados da folha da soja, porém a diferença foi mínima.

Na Tabela 4 verifica-se os resultados referentes a área A e área B separadamente. Na área A, a segunda coleta dos nutrientes, obteve melhor desempenho utilizando apenas o Nitrogênio para a predição da produtividade, com 52% de semelhança dos dados estimados e observados, apresentando uma forte correlação negativa entre os dados observados e estimados.

A área B obteve modelos que explicaram os dados observados em 33% a 58% e os testes de Kruskal-Wallis mostraram que os dados observados e estimados eram provenientes de uma mesma população em ambas as coletas da área B. Assim como na área A os melhores resultados na área B se apresentaram na segunda coleta dos nutrientes da soja utilizando apenas Nitrogênio, Potássio e Cálcio. As correlações dos dados observados e estimados variaram de moderada à forte correlação.

Segundo Filho *et al.* (2013), o número de vagens por planta, o número de sementes por vagem e peso específico das sementes, são muito influenciados pelo manejo químico do solo, principalmente pelas adubações fosfatada e potássica. Os autores citam que a deficiência de P diminui o potencial de rendimento pela menor produção de flores e de vagens e quando há deficiência de K a planta pode produzir sementes mais leves, o que afeta o rendimento final da cultura.

No trabalho de Silva *et al.* (2013), utilizou-se

Tabela 4. Resultado dos erros EQM, coeficiente de determinação e de correlação de Pearson para as melhores SVMs treinadas somente com os atributos indicados pela técnica de seleção de atributos CfsSubsetEval de cada coleta dos nutrientes da folha da soja nas áreas A e B separadamente (mesmos pontos amostrais para conjunto de treinamento, validação e teste em cada coleta)

Área	Coleta	Kernel	Atributos	EQM	R ²	R ^P
A	1	Polinomial	K_Ca_Mg	0,026	0,3488	-0,5905
	2	Polinomial	N	0,0261	0,5201	-0,7211
	3	Polinomial	P_Ca_Mg	0,0246	0,1435	0,3788
B	1	Polinomial	N_Ca_Mg	0,0501	0,3358	0,5795
	2	Polinomial	N_P_Ca	0,0488	0,5866	0,7659
	3	Polinomial	K_Ca_Mg	0,0551	0,3900	0,6245

EQM-Mean Squared Error; R² - Coeficiente de Determinação; R^P - Coeficiente de Correlação de Pearson.

adubação pré e pós-plantio da soja e concluiu-se que com os dados utilizados no seu estudo que independentemente da forma de aplicação (pré ou pós-plantio), acrescentou ganhos significativos em produtividade por hectare. A importância de se realizar a adubação de forma localizada é devido à grande variabilidade espacial dos componentes químicos do solo, e os macro nutrientes possuem grande mobilidade no solo, ou seja, variam muito de ponto a ponto. Dessa forma, não se pode utilizar teores médios de referência dos nutrientes para o cálculo da dose de fertilizante a ser aplicada, pois podem ser doses superestimadas em determinadas áreas e insuficientes em outras (VIEIRA et al., 2010).

Como explica Sfredo (2008), para se corrigir problemas de nutrição das plantas já cultivadas há tempo da colheita, deve se realizar a adubação antes do florescimento da planta e, como observado neste trabalho, os modelos que explicaram mais os dados observados e tiveram maior correlação dos dados estimados e observados, foram os modelos de SVMs que utilizaram dados da segunda coleta da soja que, neste trabalho, correspondem a dados da planta no estágio V6, ou seja, no estágio vegetativo sem flores. Dessa forma, há uma grande possibilidade de o modelo ser útil aos produtores e servir como fonte de recomendação.

Portanto, é possível que a adubação correta, de acordo com a necessidade de cada local da área de plantio, realizada tanto no pré quanto no pós-plantio, pode auxiliar no aumento da produtividade da soja. Espera-se que com os resultados obtidos nesse trabalho, o modelo de SVM que foi capaz de explicar 58% dos dados preditos apenas com as variáveis de N, P e Ca auxilie na tomada de decisão do produtor, para saber onde é mais necessário realizar a adubação para se obter maiores ganhos e menos desperdícios.

CONCLUSÕES

- As SVMs obtidas com os dados das duas áreas juntas não tiveram bons modelos de predição, pois explicaram, no máximo, 23% dos dados observados (analisando o coeficiente de determinação). Já a separação das áreas obteve melhores desempenhos, com modelos capazes

de explicar até 58% dos dados observados (obtido na segunda coleta dos nutrientes da folha na área A).

- Em ambas as áreas, os melhores modelos de SVMs obtidos foram na segunda coleta da soja, os quais apresentaram forte correlação dos dados observados e preditos.
- A área A, que obteve maior coeficiente de determinação e correlação de Pearson, não apresentou dados estimados e observados provenientes de uma mesma população, segundo o teste de Kruskal-Wallis. A área B obteve um resultado de correlação e de coeficiente de determinação pouco menores, comparando-se aos dados resultantes do modelo da área A, que apresentou homogeneidade dos dados observados e estimados, ou seja, considerados provenientes de uma mesma população.
- A seleção de atributos mostrou bons desempenhos, pois tanto na segunda coleta da área A quanto na segunda coleta da área B, os modelos explicaram mais de 50% dos dados observados, obtendo-se menos variáveis de entrada, sendo assim, mais acessível ao produtor de forma que a análise foliar se torna menos custosa para aplicar a técnica de predição da produtividade da soja.

REFERÊNCIAS BIBLIOGRÁFICAS

BOCCA, F.F. **Produtividade de cana-de-açúcar: caracterização dos contextos de decisão e utilização de técnicas de mineração de dados para modelagem**. 2014. 98f. Dissertação (Mestrado em Engenharia Agrícola) — Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2014.

BOUBOULIS, P. et al. Complex Support Vector Machines for Regression and Quaternary Classification. **IEEE Transactions on Neural Networks and Learning Systems**, United States, v.26, n. 6, p.1260-1274, 2015.

CHANG, C.C.; LIN, C.J. LIBSVM: A Library for Support Vector Machines. **Association**

for **Computing Machinery Transactions on Intelligent Systems and Technology**, New York, v.2, n.3, p.1-27, 2011.

CHOWDHURY, A.K.; DEBSARKAR, A.; CHAKRABARTY, S. Novel Methods for Assessing Urban Air Quality: Combined Air and Noise Pollution Approach. **Journal of Atmospheric Pollution**, Turquia, v.3, n.1, p.1-8, 2015.

COUTINHO, E.R.; SILVA, R.M.; DELGADO, A.R.S. Utilização de Técnicas de Inteligência Computacional na Predição de Dados Meteorológicos. **Revista Brasileira de Meteorologia**, São José dos Campos, v.31, n.1, p.24-36, 2016.

FACELI, K. et al. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. 1. ed. Rio de Janeiro: LTC, 2011. 378p.

FIDALSKI, J.; TORMENA, C.A.; ALVES, S.J. Intervalo hídrico ótimo de um latossolo vermelho distrófico, após o primeiro período de pastejo contínuo de *brachiaria ruziziensis*, em sistema integração lavoura-pecuária. **Revista Brasileira de Ciência do Solo**. Viçosa, v.37, n.3, p.775-783, 2013.

FILHO, F.B. et al. Adubação com fósforo e potássio para produção e qualidade de sementes de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.48, n.7, p.783-790, 2013.

GANDHI, A.P. Quality of soybean and its food products. **International Food Research Journal**, Malaysia, v.16, n.1, p.11-19, 2009.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900p.

KARAMIZADEH, S. et al. In: **COMPUTER, COMMUNICATIONS, AND CONTROL TECHNOLOGY**, 4., 2014, Malaysia: International Conference, 2014. p.63-65.

LACERDA, J.J.de J. et al. Adubação, produtividade e rentabilidade da rotação entre soja e milho em solo

com fertilidade construída. **Pesquisa Agropecuária Brasileira**, Brasília, v.50, n.9, p.769-778, 2015.

MARTINS, I.S. et al. Produtividade de Soja sob Aplicação de Cloreto de Potássio em pré-plantio e pós-plantio. **Nucleus**, Ituverava, v.10, n.2, 2013.

RUB, G. Data mining of agricultural yield data: A comparison of regression models. In: **PROCEEDINGS OF THE 9TH INDUSTRIAL CONFERENCE ON ADVANCES IN DATA MINING**, 9., 2009, Heidelberg: Applications and Theoretical Aspects, Springer-Verlag, 2009. p.24-37.

SFREDO, G.J. **Soja no Brasil: calagem, adubação e nutrição mineral**. Londrina: Embrapa Soja, 2008. 148p.

SILVA, I.N.D.; SPATTI, D.H.; FLAUZINO, R.A. **Redes Neurais Artificiais para engenharia e ciências aplicadas: curso prático**. 1. ed. São Paulo: Artliber, 2010. 399p.

SUBRAHMANYAM, K. et al. Modified KS-test for Feature Selection. **IOSR Journal of Computer Engineering**, United States, v.13, n.3, p.73-79, 2013.

TEDESCO, M.J. et al. **Análise de solo, plantas e outros materiais**. 2. ed. Porto Alegre, 1995. 174p.

VIEIRA, S.R. et al. Variabilidade Espacial dos teores foliares de nutrientes e da produtividade da Soja em dois anos de cultivo em um Latossolo Vermelho. **Revista Brasileira de Ciência do Solo**, Viçosa, v.34, p.1503-1514, 2010.

USDA - United States Department of Agriculture. **World Agricultural Production**. United States: Office of Global Analysis, 2016. 27p.

WOLLENHAUPT, N.C.; MULLA, D.J.; CROWFORD, G.C. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In: PIERCE, J. F.; SADLER, E. J. (Eds.). **The Site-Specific Management for Agricultural Systems**. Madison, WI: American Society of Agronomy, 1997. p.19-53.