

Aprendizagem de Máquina para Avaliação da Eficácia de Seguro Climático Multiameaça

Marcos Roberto Benso¹

1 - Departamento de Engenharia de Biosistemas da Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo)

Neste trabalho uma proposta de esquemas de seguro paramétrico multiameaça com base em algoritmos de aprendizagem de máquina é apresentada. Tendo a produção de soja no Brasil como estudo de caso, dados da Pesquisa Agrícola Municipal (1988-2023) e observação meteorológicas do *Brazilian daily weather gridded data* (1961–2020) foram combinadas. Os dados meteorológicos foram utilizados como insumo para o cálculo de índices climáticos mensais, incluindo médias (precipitação e temperaturas médias) e índices extremos para produção de modelos preditivos com base em aprendizagem de máquina. Com base nos indicadores, foram construídos quatro cenários de políticas de seguro: S1: precipitações médias, S2: temperaturas médias, S3: precipitações e S4: temperaturas médias e complexo, que inclui as médias mensais e extremos climáticos. Foram testados como modelos candidatos, regressão linear, Random Forest (RF), XGBoost e Gradient Boosting Machine (GBM), apresentando, respectivamente, os valores de R² para teste de 0,51, 0,61, 0,60 e 0,65. A avaliação das políticas de seguro foi feita com base na comparação com produtividades estimadas pelos modelos de aprendizagem de máquina (\hat{y}) comparadas com a produtividade municipal observada pela PAM/IBGE (y). O modelo de seguro segue um modelo de indenização (p) em nível municipal, cuja produtividade média é dada por $E(y)$ e

$$p = \max\left(\frac{\hat{y} - E(y)}{E(y)}, 0\right) \cdot E(y)$$

. Com base nos quatro cenários avaliados com o modelo GBM, foram calculados indicadores de eficácia de seguros como potencial de redução de risco (RRP, sacas/ha) e redução de risco relativa (RRR, %) que comparam o balanço de caixa de uma fazenda situada em um dado município com e sem adoção de apólice de seguros indexados multiameaça. Uma diferença estatisticamente significativa com p -valor $< 0,05$ de RRP foi encontrada entre os quatro cenários. O cenário S2 apresentou a pior performance com RRP de -5,4 sacas/ha, o valor negativo indica a tendência da adoção de uma política que leva apenas em consideração a temperatura média, é pior do que o cenário sem adoção de nenhuma política de redução de risco. Os cenários S1: precipitação apenas, S3: precipitação e temperatura e S4: complexo apresentaram valores positivos de NRR, indicando que os cenários com redução de risco são superiores ao cenário sem redução de risco. Considerando os dados dos modelos avaliados em nível dos seis biomas brasileiros, a precisão dos modelos permanece alta em todos os biomas e esquemas, variando de 0,74 (Amazônia sob precipitação) a 0,89 (Caatinga sob precipitação e temperatura combinadas). A maioria dos biomas apresenta ligeiros ganhos de precisão ao adicionar temperatura aos gatilhos de precipitação ou ao migrar para o modelo complexo. A redução do risco líquido destaca onde os esquemas de índice evitam perdas de culturas de forma mais eficaz. A Caatinga se destaca sob o esquema de precipitação com uma RRP de 19,61 sacas/ha/ano e uma RRR de 0,40, o que significa que 40% das perdas esperadas são evitadas. A Amazônia também apresenta desempenho moderado, em torno de 8 sc/ha/ano (RRR 0,55).

Palavras-chave: Ciência de Dados; Redução de Risco de Desastre; Soja.