

Inteligencia Artificial en la Gestión Agrícola: Uso de Modelos de Bosque Aleatorio para la Predicción de Producción y Reserva de Semillas en Brasil

 <https://doi.org/10.56238/sevned2024.023-006>

Eduardo Silva Vasconcelos

Doctor en Ciencias - Sistemas de Información
Instituto Federal de Goiás

Adriano Dawison de Lima

Doctorado en Agronomía
Universidad de Uberaba

Leandro Aureliano da Silva

Doctor en Ciencias - Sistemas de Información
Universidad de Uberaba

Luiz Fernando Ribeiro de Paiva

Doctor en Educación
Universidad de Uberaba

Débora Vasconcelos Melo

Máster en Gestión de las Organizaciones
Universidad Federal de Cataluña

Cleiton Silvano Goulart

Máster en Física
Universidad de Uberaba

RESUMEN

Este estudio aborda la aplicación de modelos de Inteligencia Artificial (IA), más específicamente de bosque aleatorio, para la predicción de la producción y reserva de semillas en la agricultura brasileña. El objetivo principal es contribuir al avance en la gestión y planificación de recursos, una acción crítica para aumentar la eficiencia y sostenibilidad del sector. El trabajo se destaca por la importancia de entender el papel de la IA en la optimización de prácticas agrícolas, proporcionando un marco de referencia para investigaciones futuras en la intersección entre tecnologías de IA y agricultura. Metodológicamente, el estudio implementó un riguroso proceso de recolección y procesamiento de datos proporcionados por el Ministerio de Agricultura y Ganadería de Brasil, cubriendo las cosechas de 2016/2017 a 2023/2024. La limpieza de datos precedió a la transformación de las variables categóricas a través de la codificación one-hot y la subsiguiente división del conjunto de datos en un 80% para entrenamiento y un 20% para pruebas. Utilizando la biblioteca scikit-learn, se configuró y evaluó un modelo de bosque aleatorio, empleando técnicas de validación como la división entrenamiento/prueba y la validación cruzada, además de las métricas de error cuadrático medio (MSE) y coeficiente de determinación (R^2) para medir la precisión y eficacia del modelo. Los resultados indican una correlación positiva moderada a fuerte entre las variables de tiempo y cantidad de semillas reservadas para ambos períodos de cultivo, Safra y Safrinha. Sin embargo, los análisis señalaron una variabilidad anual y una confianza diferenciada en las predicciones entre los períodos, sugiriendo la influencia de factores adicionales y la necesidad de modelos adaptativos. La concentración de la producción en pocas culturas fue identificada como un riesgo potencial, sugiriendo que la diversificación es clave para la resiliencia del sector. La capacidad de generalización del modelo fue evaluada, y el fenómeno de sobreajuste fue considerado una posibilidad ante las variaciones de precisión entre los datos de entrenamiento y prueba. Este estudio refuerza el potencial transformador que los modelos de IA, como el bosque aleatorio, poseen para la predicción y gestión agrícola, abriendo puertas para mejoras futuras y proporcionando subsidios valiosos para decisiones estratégicas basadas en datos en el sector agrícola.

Palabras clave: Inteligencia Artificial en Agricultura, Modelos de Bosque Aleatorio, Predicción de Producción de Semillas, Gestión Agrícola Sostenible, Análisis de Datos Agrícolas.



1 INTRODUCCIÓN

La agricultura enfrenta desafíos crecientes debido a las demandas de eficiencia y sostenibilidad. Este estudio aborda la importancia de la Inteligencia Artificial (IA) en la predicción de la producción de semillas, crucial para la planificación y la gestión de recursos en el sector agrícola. Nos enfocamos en el desarrollo y evaluación de un modelo de IA basado en bosque aleatorio, con el objetivo de mejorar las prácticas de toma de decisiones y la eficiencia operacional en la agricultura.

La creciente necesidad de optimización en la producción agrícola y la gestión de recursos ha impulsado la adopción de tecnologías avanzadas, como la Inteligencia Artificial (IA). En este contexto, la predicción de la producción de semillas emerge como un campo de estudio vital, ofreciendo perspectivas valiosas para el planeamiento agrícola y la sostenibilidad del sector. El presente trabajo explora la aplicación de modelos de IA, específicamente algoritmos de bosque aleatorio, para prever la producción de semillas, con el fin de mejorar el control y la gestión de recursos en la agricultura.

La importancia científica de este estudio reside en su contribución al entendimiento de cómo los modelos de IA pueden ser aplicados para mejorar la eficiencia y productividad agrícolas, además de proporcionar un marco para futuras investigaciones en la intersección entre IA y agricultura. La integración de la IA en la agricultura, especialmente para la predicción de la producción de semillas, tiene el potencial de transformar la gestión de recursos y la toma de decisiones en el sector. Este estudio demuestra la viabilidad y el valor de utilizar modelos de bosque aleatorio para este fin, destacando la importancia de tales enfoques para avances futuros en la agricultura sostenible y eficiente.

Este estudio tiene el objetivo general de desarrollar y evaluar un modelo de Inteligencia Artificial basado en técnicas de bosque aleatorio, destinado a la predicción de la producción y reserva de semillas en el contexto agrícola. Aspiramos a contribuir significativamente a la mejora de la gestión y la planificación de recursos en el sector, una necesidad imperante para el aumento de la eficiencia y sostenibilidad agrícolas. El estudio pretende no solo mejorar la planificación estratégica y la eficacia en la gestión de recursos, sino también impulsar la precisión y sostenibilidad en la agricultura, proporcionando directrices fundamentales para el desarrollo de estrategias de gestión y planificación avanzadas en el sector agrícola.

En términos específicos, el estudio busca: (1) emplear un análisis multifacético de los patrones de producción y reserva de semillas en la agricultura brasileña mediante técnicas estadísticas, incluyendo gráficos de líneas temporales, gráfico de Pareto, gráfico de dispersión, regresión lineal y tablas analíticas; (2) investigar la aplicabilidad y eficacia del algoritmo de bosque aleatorio, utilizado en este trabajo, para la modelación y predicción de la producción de semillas, evaluar el desempeño del modelo en términos de precisión, eficiencia y capacidad de generalización, y examinar la medida en que las predicciones basadas en Inteligencia Artificial contribuyen a optimizar la planificación y el control de recursos en la agricultura.

Estos objetivos buscan sintetizar la relevancia práctica y teórica de los modelos de Inteligencia Artificial, con énfasis en bosques aleatorios, en el análisis e interpretación de patrones de producción y reserva de semillas, destacando su aplicabilidad para elucidar tendencias, identificar variaciones y discernir factores críticos que influyen en estas dinámicas.

2 REFERENCIAL TEÓRICO

La revisión de la literatura contempla una serie de investigaciones previas que exploran la aplicación de la Inteligencia Artificial (IA) en la agricultura, enfatizando particularmente el uso de métodos de aprendizaje automático, como los bosques aleatorios, en el contexto de la predicción de producción. Esta revisión sintetiza tanto las contribuciones significativas para el avance tecnológico en el sector como estudios específicos que demuestran la eficacia de estas técnicas.

En el ámbito de las contribuciones y avances tecnológicos, Vieira Filho y Silveira (2012) discuten sobre la innovación en la agricultura, subrayando la complejidad del sector y la crucialidad del conocimiento y aprendizaje tecnológico para su desarrollo. Ellos argumentan que la innovación es un catalizador esencial para el mejoramiento de las prácticas agrícolas y para la sostenibilidad de la producción. Adama Brasil (2024) y Climate FieldView (2023) expanden esta discusión al abordar la importancia fundamental de la IA en la modernización de la agricultura. Estas fuentes iluminan aplicaciones específicas de la IA, tales como el sensoramiento remoto y el análisis de grandes volúmenes de datos en tiempo real, que han transformado la capacidad de monitoreo y gestión agrícola. Rehagro (2022) complementa esta visión, destacando las virtudes de la IA en la monitorización y optimización de la agricultura, enfatizando su papel en promover prácticas sostenibles y en elevar la eficiencia operacional en el campo.

En cuanto a los estudios específicos, las investigaciones llevadas a cabo por Mourtzinis et al. (2021), Saleem, Potgieter y Arif (2021), y Gadotti et al. (2022) evidencian la aplicabilidad y la eficiencia de técnicas de aprendizaje automático en el sector agrícola. Estos estudios proporcionan evidencias robustas de la superioridad de métodos como los bosques aleatorios y el aprendizaje profundo en realizar tareas predictivas específicas, apuntando a un desempeño notablemente superior en comparación con técnicas más tradicionales de modelización estadística. Ellos enfatizan la relevancia de los enfoques basados en IA como herramientas predictivas altamente competentes, capaces de manejar la complejidad de los datos agrícolas y de producir predicciones precisas, que son fundamentales para la planificación y la toma de decisiones informadas en el sector.

Por lo tanto, la literatura revisada subraya la progresión constante y el impacto potencial de la IA en la agricultura, indicando un horizonte prometedor para la integración continua de estas tecnologías avanzadas en las prácticas agrícolas contemporáneas. Estos avances no solo corroboran la posición de la IA como un pilar esencial en la modernización agrícola, sino que también endosan la



necesidad de investigación continua para explorar plenamente las capacidades y beneficios de los bosques aleatorios y del aprendizaje automático en la predicción agrícola.

3 METODOLOGIA

La metodología empleada para crear un modelo de "bosque aleatorio" tiene como finalidad estimar la cantidad de semillas que serán producidas y reservadas en la agricultura. Se trata de un método que utiliza la computación y la estadística para analizar datos sobre la producción agrícola pasada, con el fin de hacer predicciones confiables sobre la cantidad de reserva de semillas que se generará en el futuro. Este proceso ayuda a agricultores y gestores a planificar mejor sus actividades, anticipando la necesidad de semillas y optimizando recursos para una producción más eficaz y sostenible. Utilizando datos proporcionados por el Ministerio de Agricultura y Ganadería de Brasil, este estudio implementa un enfoque riguroso, desde la recolección hasta el procesamiento de los datos, culminando en el desarrollo de un modelo robusto para el análisis predictivo.

Este estudio adopta un enfoque cuantitativo, utilizando datos históricos de producción de semillas para desarrollar un modelo de bosque aleatorio. La eficacia del modelo se evalúa mediante métricas estadísticas, incluyendo el error cuadrático medio (MSE) y el coeficiente de determinación (R^2).

3.1 DISEÑO DEL ESTUDIO

El estudio fue diseñado para aplicar el modelo de bosque aleatorio, conocido por su capacidad de procesar grandes conjuntos de datos y evitar el sobreajuste. El objetivo fue evaluar la eficacia de este modelo en la predicción de la producción de semillas, incorporando variables como el período, la región, la especie de semilla y el área plantada.

3.2 RECOLECCIÓN DE DATOS

Los datos fueron obtenidos del Sistema de Gestión de Fiscalización del Ministerio de Agricultura y Ganadería de Brasil, abarcando información detallada desde la cosecha de 2016/2017 hasta 2023/2024. La exclusión de datos anteriores a 2016/2017 fue necesaria debido a inconsistencias identificadas.

3.3 PREPARACIÓN DE LOS DATOS

La preparación incluyó la limpieza y transformación de los datos, adoptando técnicas como la codificación one-hot para variables categóricas. La división de los datos en conjuntos de entrenamiento y prueba siguió la proporción estándar de 80/20.

3.4 DESARROLLO DEL MODELO

El modelo de bosque aleatorio se implementó en Python, utilizando la biblioteca scikit-learn. La selección de hiperparámetros se basó en prácticas estándar, considerando el equilibrio entre desempeño y tiempo de cómputo.

4 IMPLEMENTACIÓN Y EVALUACIÓN DEL MODELO DE BOSQUE ALEATORIO

La metodología empleada en la implementación y evaluación del modelo de bosque aleatorio es detallada, reflejando las diversas etapas necesarias para el análisis de predicción de la producción de semillas en la agricultura. Esta metodología está estructurada para permitir un análisis futuro detallado y criterioso de los resultados del modelo.

En la fase de entrenamiento, se realizaron procedimientos críticos, comenzando con la preparación inicial de los datos, donde se aplicaron técnicas de codificación para adecuar las variables categóricas al formato numérico exigido por el modelo:

```
from sklearn.preprocessing import OneHotEncoder
# Codificación One-Hot de las variables categóricas
encoder = OneHotEncoder(sparse=False)
X_encoded = encoder.fit_transform(X[['UF', 'Especie de Semente']])
```

A continuación, los datos se dividieron en conjuntos de entrenamiento y prueba, con el 80% destinado al entrenamiento y el 20% a la prueba, esencial para la posterior evaluación de la capacidad del modelo para generalizar a nuevos datos:

```
from sklearn.model_selection import train_test_split
# División de los datos en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Con los datos preparados, la construcción del modelo de bosque aleatorio se llevó a cabo utilizando la biblioteca scikit-learn, configurando hiperparámetros como la cantidad de árboles en el modelo:

```
from sklearn.ensemble import RandomForestRegressor
# Construcción del modelo de bosque aleatorio
modelo = RandomForestRegressor(n_estimators=100, random_state=42)
modelo.fit(X_train, y_train)
```

Para asegurar la fiabilidad de las predicciones del modelo, se implementó una estrategia de validación rigurosa, utilizando no solo la división entrenamiento/prueba sino también la validación cruzada para proporcionar una evaluación más robusta:

```
from sklearn.model_selection import cross_val_score
# Aplicación de validación cruzada
scores = cross_val_score(model, X, y, cv=5)
```

Los criterios de evaluación incluyeron métricas como el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2), fundamentales para medir la eficacia y la precisión del modelo:

```
from sklearn.metrics import mean_squared_error, r2_score
# Evaluación del modelo en los datos de prueba
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

Estas etapas metodológicas garanten que el modelo está bien ajustado a los datos, posibilitando la realización de predicciones confiables y precisas. La capacidad de generalización del modelo fue evaluada a través de la comparación de desempeño entre los conjuntos de entrenamiento y prueba, crucial para detectar sobreajuste o subajuste.

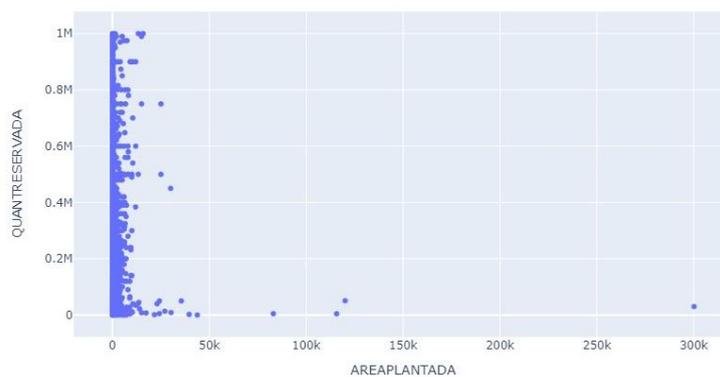
5 RESULTADOS E DISCUSIONES

5.1 ANÁLISIS E INTERPRETACIÓN DE PATRONES DE PRODUCCIÓN Y RESERVA DE SEMILLAS

Examinaremos los patrones de producción y reserva de semillas en la agricultura brasileña, utilizando un modelo de bosque aleatorio para interpretar datos complejos y multifacéticos. Se destaca el análisis de dos fases distintas de cultivo - la Safra y la Safrinha - y la comparación del desempeño del modelo en cada una. La interpretación de gráficos y tablas proporciona percepciones profundas sobre las dinámicas del sector, crucial para el desarrollo de estrategias eficaces en la gestión agrícola.

El Gráfico 1, a continuación, evidencia la compleja relación entre el área plantada y la cantidad de semillas reservadas, no demostrando una correlación directa clara. La Tabla 1 complementa este análisis, detallando la distribución de semillas reservadas entre los principales cultivos agrícolas, indicando una concentración significativa en pocas especies.

Gráfico 1 - Relación entre Área Sembrada y Cantidad Reservada



Fuente: Elaboración propia

El Gráfico 1 proporciona un análisis de la relación entre ‘Área Sembrada’ y ‘Cantidad Reservada’ en el contexto de la producción de semillas reservadas en Brasil, basado en los datos del SIGEF. La investigación se centra en discernir la tendencia general, identificar patrones y destacar valores atípicos. Se observa que el gráfico no revela una correlación lineal obvia entre ‘Área Sembrada’ y ‘Cantidad Reservada’, con la mayoría de los datos agrupados en el extremo menor de ‘Área Sembrada’, sin demostrar un incremento proporcional en ‘Cantidad Reservada’ a medida que aumenta ‘Área Sembrada’. Tal disposición sugiere la ausencia de una relación directa y robusta entre estas variables.

Además, el análisis resalta la presencia de valores atípicos, particularmente donde la ‘Cantidad Reservada’ permanece elevada independientemente del ‘Área Sembrada’, señalando posibles excepciones que podrían atribuirse a factores como alta productividad por área o variables externas no capturadas por el gráfico. La correlación de Pearson, cercana a 0.0973, indica una asociación muy débil, casi despreciable, corroborando la falta de una conexión fuerte entre las variables analizadas.

Este estudio concluye que la ‘Área Sembrada’ no sirve como predictor confiable de la ‘Cantidad Reservada’, inferiendo que otras variables no consideradas aquí, como la eficiencia en la producción de semillas, condiciones del suelo, prácticas agronómicas y climáticas, pueden ejercer influencias más significativas. La baja correlación refuerza la necesidad de un enfoque más integrado y multifactorial para entender la dinámica de la producción de semillas, implicando que análisis futuros deben incorporar una gama más amplia de datos y metodologías analíticas para capturar la complejidad del sector de semillas en Brasil.

Tabla 1: Análisis Porcentual de Semillas Reservadas para Principales Especies Agrícolas en Brasil

ESPECIE	QUANTRESERVADA	%	% Acumulado
<i>Glycine max</i> (L.) Merr. (SOJA)	3749435500,4430	66,2562	66,2562
<i>Triticum aestivum</i> L. (TRIGO)	877409998,2770	15,5047	81,7608
<i>Solanum tuberosum</i> L. (BATATA)	779113570,3830	13,7677	95,5285

Fonte: Elaboração Própria

La Tabla 1 ofrece un análisis cuantitativo de la reserva de semillas de tres cultivos primordiales en el agronegocio brasileño, resaltando la predominancia de la soja, el trigo y la papa en el espectro agrícola nacional. Específicamente, la soja, *Glycine max* (L.) Merr., se destaca significativamente, comprendiendo más de dos tercios del total de las reservas, enfatizando su posición central en la estructura agrícola y económica de Brasil. La relevancia del trigo, *Triticum aestivum* L., y de la papa, *Solanum tuberosum* L., también es notable, subrayando sus contribuciones esenciales a la seguridad alimentaria y a la diversificación agrícola.

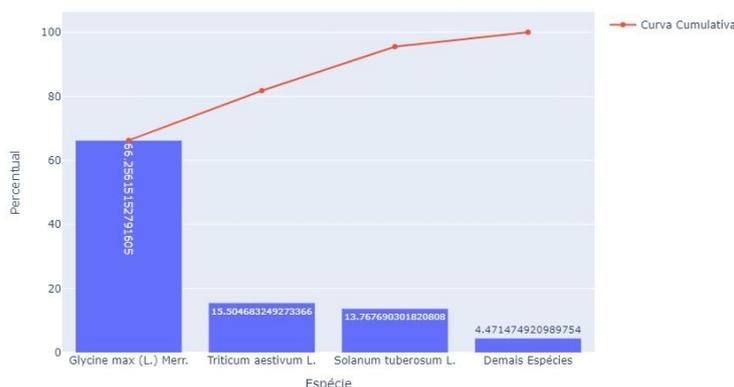
Estos datos apuntan a una concentración de la actividad agrícola en pocas culturas, que, aunque benefician económicamente, introducen riesgos asociados a la dependencia y a la falta de variedad, enfatizando la importancia de estrategias de diversificación. La expresiva concentración productiva en estas especies, responsable de más del 95% del total de las semillas reservadas, evidencia una dualidad entre fuerza económica y potencial vulnerabilidad ante adversidades, tales como fluctuaciones de mercado o eventos climáticos adversos.

Desde esta perspectiva, la Tabla 1 trasciende la mera exposición de datos, funcionando como un instrumento crítico para reflexión sobre políticas agrícolas. El análisis subraya la necesidad urgente de incentivar la diversificación de cultivos y adoptar prácticas agrícolas sostenibles, apuntando no solo a la robustez económica sino también a la resiliencia y la seguridad alimentaria de Brasil. Así, el estudio de estos datos se vuelve fundamental para la elaboración de políticas que equilibren producción, sostenibilidad y seguridad alimentaria en el contexto agrícola brasileño.

5.2 ANÁLISIS COMPARATIVO: SAFRA VS. SAFRINHA

La comparación entre "Data_Safra" y "Data_Safrinha" revela diferencias marcadas en el desempeño del modelo. El "Gráfico 2: Gráfico de Pareto para Especies" y el "Gráfico 3: Suma de la Cantidad Reservada de Semillas por Período" ilustran estas variaciones, mostrando cómo la producción de semillas responde a diferentes factores y condiciones en cada ciclo de cultivo.

Gráfico 2: Gráfico de Pareto para Especies



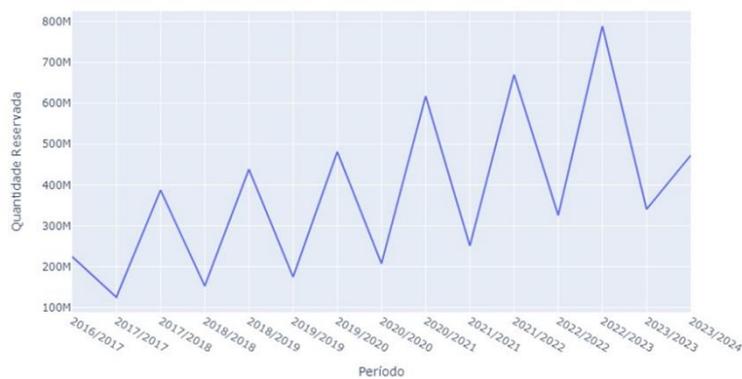
Fuente: Elaboración propia

En la presentación de los resultados expuestos en la Tabla 1, la utilización del gráfico de Pareto surge como una estrategia analítica destacada para elucidar la distribución de semillas entre los principales cultivos agrícolas en Brasil. Este gráfico es reconocido por su habilidad para representar visualmente la predominancia de ciertos factores, en este contexto, variedades específicas de cultivos, en contribuir significativamente a un resultado comprensivo, aquí siendo el agregado de las semillas reservadas. El empleo de este gráfico se fundamenta en su aptitud para demostrar el Principio de Pareto, o regla 80/20, elucidando que un segmento menor de causas tiende a ser responsable por la mayor parte de los resultados observados.

El Gráfico de Pareto, identificado como Gráfico 2, se establece como un mecanismo visual esencial para analizar y discutir la asignación y repercusiones de la reserva de semillas en el sector agrícola brasileño. Este instrumento facilita el reconocimiento de áreas que demandan atención prioritaria para futuras intervenciones e investigaciones, buscando una comprensión detallada sobre la configuración actual de la reserva de semillas en la agricultura del país, además de esclarecer sobre las fuerzas de mercado y potenciales para avances en el dominio.

Específicamente, el Gráfico de Pareto resalta el papel preeminente de cultivos como la soja, el trigo y la papa en el panorama de la producción de semillas brasileña, trayendo implicaciones sustanciales para la formulación de políticas agrícolas, delineamiento de estrategias y emprendimientos orientados a la sostenibilidad. Esta análisis indica la importancia de adoptar una táctica balanceada que no solo realce la producción eficaz y ecológica de estas culturas clave sino también incentive la diversificación agronómica para fortalecer la resiliencia y adaptabilidad del sector. La enfatización en la soja, que representa una significativa porción de la producción conforme los datos, sugiere que medidas políticas podrían ser orientadas para ampliar esta cultura, invirtiendo en áreas como investigación, mejoramiento genético e infraestructura logística, repercutiendo positivamente en el desarrollo agrícola nacional.

Gráfico 3: Suma de la Cantidad Reservada de Semillas por Período

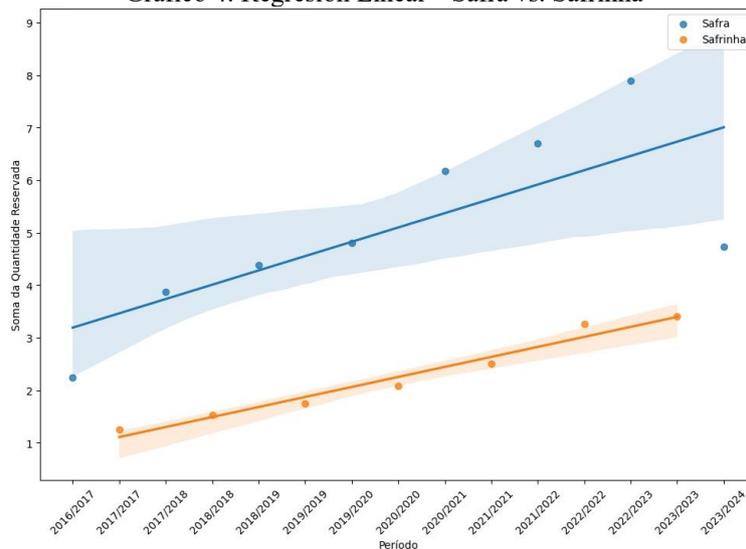


Fonte: Elaboração Própria

El Gráfico 3 proporciona un análisis longitudinal de la reserva de semillas en la agricultura brasileña de 2016 a 2024, revelando un patrón de fluctuaciones significativas que reflejan la dinámica del sector agrícola. No se observa una tendencia uniforme de crecimiento o declive en las reservas a lo largo de los años estudiados; en cambio, las variaciones anuales apuntan a una respuesta del sector a un conjunto diversificado de influencias, incluyendo factores ambientales, económicos y de mercado. Períodos como 2019/2020 y 2022/2023 se distinguen por picos en la reserva de semillas, posiblemente indicando años de abundancia o respuesta a aumentos en la demanda. Por contraste, el declive observado en 2023/2024 puede denotar adversidades enfrentadas por el sector.

El estudio también diferencia los ciclos de cultivo "Safrá" y "Safrinha", destacando la influencia de las condiciones climáticas y estrategias de plantación sobre la producción. La Safrá, período principal de cultivo, se beneficia de condiciones climáticas ideales y presenta alta productividad, mientras que la Safrinha, a pesar de las condiciones menos favorables y tradicionalmente menor productividad, representa un estadio crucial para la sostenibilidad y la continuidad de la producción agrícola. Estas observaciones subrayan la capacidad de adaptación y resiliencia del sector agrícola brasileño ante variables intermitentes, proporcionando una base para la comprensión profunda de las estrategias de manejo y de toma de decisiones en el contexto de la producción nacional de semillas.

Gráfico 4: Regresión Lineal – Safra vs. Safrinha



Fonte: Elaboração Própria

El Gráfico 4 ilustra la relación temporal entre los períodos de cultivo, Safra y Safrinha, y la suma de las cantidades reservadas de semillas en el contexto de la agricultura brasileña. La línea azul, representando la Safra, y la línea naranja, representando la Safrinha, son resultados de una regresión lineal que busca estimar la tendencia central de los datos de cada período. Las áreas sombreadas corresponden a los intervalos de confianza para cada regresión, que indican el grado de incertidumbre asociado a las predicciones del modelo.

El área sombreada más amplia alrededor de la línea de regresión de la Safra sugiere mayor variabilidad y menos confianza en las predicciones de reserva de semillas para este período. Esto implica que factores adicionales, potencialmente no capturados por el modelo, pueden estar influyendo en las reservas de semillas más allá del tiempo, como variaciones climáticas, políticas agrícolas, o cambios económicos. Por su parte, el área sombreada más estrecha alrededor de la línea de regresión de la Safrinha implica una mayor precisión en las predicciones, sugiriendo que la cantidad de semillas reservadas puede ser más consistentemente predicha con base en el tiempo para ese período.

La presencia de puntos por encima o por debajo de las líneas de regresión indica la variabilidad de los datos en relación con la tendencia modelada, y esos puntos pueden ser considerados outliers si están fuera del intervalo de confianza. Por ejemplo, si un punto de datos está por encima del área sombreada de la Safra, esto puede indicar un año en que la reserva de semillas fue excepcionalmente alta, posiblemente debido a una cosecha particularmente abundante o a políticas de estímulo a la reserva de semillas.

El coeficiente de Pearson para cada conjunto de datos cuantifica la fuerza de la relación lineal entre el tiempo y las reservas de semillas. Un coeficiente de Pearson de aproximadamente 0.755 para la Safra y de cerca de 0.983 para la Safrinha indica una fuerte correlación positiva para ambos períodos, con la Safrinha mostrando una relación casi perfecta. Sin embargo, la interpretación del coeficiente

debe considerar el contexto específico y los posibles factores confusos que pueden afectar las reservas de semillas.

Estos análisis estadísticos proporcionan insights fundamentales sobre la dinámica de producción y reserva de semillas, evidenciando un crecimiento continuo en ambas fases, pero con mayor previsibilidad y estabilidad durante la Safrinha. Se presta atención particularmente al comportamiento anómalo de los datos en 2023/2024, que destaca la necesidad de investigaciones adicionales para comprender las variaciones inesperadas y sus implicaciones para la gestión agrícola. La metodología empleada revela patrones críticos y ofrece una base sólida para la planificación estratégica y la toma de decisiones informada en el sector agrícola.

Por tanto, el análisis estadístico del Gráfico 4 demuestra la utilidad de métodos cuantitativos como la regresión lineal y el cálculo de intervalos de confianza para entender la dinámica de la producción agrícola y para informar decisiones estratégicas en el sector. El análisis sugiere que, mientras hay una tendencia positiva en el volumen de semillas reservadas a lo largo del tiempo para ambos períodos de cultivo, la variabilidad anual y la confianza en las predicciones pueden diferir sustancialmente entre la Safra y la Safrinha.

5.3 ANÁLISIS DE RENDIMIENTO DEL MODELO

Este capítulo se dedica a evaluar la eficacia del modelo de bosque aleatorio desarrollado para prever la producción de semillas en la agricultura brasileña. Este análisis está estructurado en tres subcapítulos distintos, cada uno enfocándose en diferentes conjuntos de datos y contextos para proporcionar una visión comprensiva del desempeño del modelo.

En el primer subcapítulo, "Predicción de la Producción de Semillas: Un Estudio sobre 'Data_Safra'", examinamos el rendimiento del modelo utilizando los datos específicos de la cosecha principal, enfocándonos en la precisión de las predicciones y en la capacidad del modelo de ajustarse a los datos históricos de este período crítico.

El segundo subcapítulo, "Predicción de la Producción de Semillas: Un Estudio sobre 'Data_Safrinha'", analiza las predicciones del modelo para la safrinha, un período de cultivo secundario, pero igualmente significativo, destacando las peculiaridades y los desafíos específicos de modelado para este ciclo de cultivo.

Finalmente, en el tercer subcapítulo, "Predicciones de Producción de Semillas: Un Análisis entre 'Data_Safra' y 'Data_Safrinha'", realizamos una comparación directa del rendimiento del modelo entre los dos períodos de cultivo. Este comparativo busca entender las variaciones y la consistencia del modelo en diferentes contextos agrícolas, contribuyendo a una comprensión más profunda de la eficacia de las técnicas de Inteligencia Artificial en prever dinámicas variables en el sector agrícola.

Cada subcapítulo contribuye a un entendimiento detallado de cómo el modelo de bosque aleatorio se adapta y responde a las complejidades de la producción de semillas, proporcionando insights valiosos para futuras aplicaciones y mejoras en el campo de la agricultura de precisión.

5.3.1 Predicción de la producción de semillas: Un estudio sobre 'data_safrá'

El análisis del modelo de regresión utilizando el subconjunto de datos 'data_safrá', extraído de un conjunto mayor, ofrece una visión cuantitativa y cualitativa sobre la eficacia de las predicciones relativas a la producción de semillas. Las métricas obtenidas — Error Cuadrático Medio (MSE) y Coeficiente de Determinación (R^2) — son esenciales para evaluar el modelo de bosque aleatorio adoptado en esta investigación.

La métrica de Error Cuadrático Medio (MSE) para los conjuntos general y de prueba es significativa, indicando una discrepancia notable entre los valores predichos y los reales. La precisión del modelo para nuevos datos es, por lo tanto, menor, como lo ilustra el MSE de aproximadamente 2,833,804,168.5. Este número elevado evidencia las limitaciones del modelo en la predicción de nuevos eventos, un aspecto crítico en la modelación predictiva.

Por otro lado, el MSE para el conjunto de entrenamiento muestra un rendimiento más prometedor, con un valor sustancialmente menor de 1,653,458,753.5. Esta diferencia destaca una mayor precisión del modelo en las condiciones bajo las cuales fue entrenado, un resultado esperado e indicativo de un ajuste apropiado a los datos de entrenamiento.

En cuanto al Coeficiente de Determinación (R^2), los valores para los conjuntos de prueba y entrenamiento revelan una disparidad importante. Para el conjunto de prueba, el R^2 de 0.4149566274957277 implica que solo aproximadamente el 41.5% de la varianza es explicada por el modelo. Aunque indica una capacidad de predicción, sugiere también que factores no contemplados en el modelo tienen un papel significativo. Para el conjunto de entrenamiento, el R^2 es de 0.6544389740961896, revelando que el modelo logra explicar aproximadamente el 65.44% de la varianza, demostrando un ajuste razonable a los datos utilizados para el entrenamiento.

Esta variación entre los conjuntos de entrenamiento y prueba, particularmente un porcentaje de acierto del test en relación al entrenamiento de 63.41%, sugiere un fenómeno conocido como sobreajuste (overfitting). Este ocurre cuando el modelo aprende las particularidades de los datos de entrenamiento a tal punto que falla en generalizar sus predicciones para nuevos datos, limitando su utilidad práctica. Esta constatación sugiere la necesidad de un equilibrio más refinado entre el ajuste del modelo y su capacidad de generalización.

5.3.2 Predicción de la producción de semillas: Un estudio sobre 'data_safrinha'

El estudio en cuestión utiliza el subconjunto 'data_safrinha' de un DataFrame más abarcador para modelar y evaluar las predicciones sobre la producción de semillas, utilizando un modelo de bosque aleatorio. Las métricas resultantes de la análisis, Error Cuadrático Medio (MSE) y Coeficiente de Determinación (R^2), constituyen herramientas vitales para la aferición del rendimiento del modelo predictivo.

El Error Cuadrático Medio (MSE) surge como un indicador de la media de los errores al cuadrado, es decir, las diferencias entre las predicciones del modelo y los valores reales observados. Un MSE elevado, como el obtenido tanto para el conjunto general como para el de prueba, con un valor alrededor de 3,113,846,766.768054, señala una magnitud considerable de error en las predicciones para los datos nuevos, apuntando a un área de atención y potencial mejora en la precisión del modelo. No obstante, el MSE obtenido para el conjunto de entrenamiento es sustancialmente menor, evidenciando un rendimiento más preciso en las predicciones con un valor de 1,386,346,989.5445054.

Al analizar el Coeficiente de Determinación (R^2 Score), que refleja la proporción de varianza en la variable dependiente que es explicada por las variables independientes en el modelo, notamos una distinción marcada entre los conjuntos de prueba y entrenamiento. Un R^2 de 0.7183833224143992 para el conjunto de prueba indica que aproximadamente el 71.84% de la varianza es capturada por el modelo, un valor considerable que atestigua la eficacia del modelo en la explicación de los datos. Este porcentaje aumenta para el conjunto de entrenamiento, donde el R^2 alcanza un robusto 0.8676384019256064, sugiriendo que el modelo se ajusta bien a los datos de entrenamiento.

Sin embargo, la diferencia entre las métricas de entrenamiento y prueba revela que, aunque el modelo muestra eficiencia en el aprendizaje de los patrones en los datos de entrenamiento, como lo indica un R^2 de entrenamiento relativamente alto, no mantiene el mismo nivel de precisión al ser aplicado a nuevos datos, una discrepancia que señala la presencia de sobreajuste.

Estos resultados, junto con el Porcentaje de Acierto del Test en Relación al Entrenamiento de 82.80%, destacan la necesidad de mayor investigación y ajuste fino del modelo para mejorar su generalización y capacidad predictiva. A pesar de las áreas para mejora, el modelo de bosque aleatorio se presenta como una aproximación robusta y prometedora para enfrentar la complejidad de los datos en la agricultura, con potencial para mejorar significativamente la precisión y eficacia del planeamiento y gestión de los recursos agrícolas.



5.3.3 Predicciones de producción de semillas: Un análisis entre 'data_safra' y 'data_safrinha'

Los datos revelan que el rendimiento del modelo varía entre los dos períodos, con un Coeficiente de Determinación (R^2) de 0.4149566274957277 para el "data_safra" y un R^2 considerablemente más alto de 0.7183833224143992 para el "data_safrinha". Este último sugiere que el modelo de bosque aleatorio logra explicar cerca del 71.84% de la variabilidad de los datos para el período de la safrinha, un contraste notable con el menor porcentaje de aproximadamente el 41.5% para el período de la safra.

Al analizar el Error Cuadrático Medio (MSE), notamos que, aunque los valores son altos para ambos conjuntos de datos, el MSE del "data_safrinha" es ligeramente superior, lo que podría sugerir una mayor discrepancia en las predicciones para este período. No obstante, es esencial considerar este dato junto con el R^2 , que nos cuenta una historia diferente, indicando una mejor capacidad de predicción para la safrinha.

Más notable es la discrepancia entre los resultados de entrenamiento y prueba para cada conjunto de datos. El "data_safra" muestra una diferencia más significativa, un indicador que puede señalar un sobreajuste del modelo a los datos de entrenamiento, reduciendo su eficacia en generalizar para datos no vistos. Por otro lado, el "data_safrinha" muestra una menor discrepancia entre los resultados de entrenamiento y prueba, reflejando una mejor generalización y una fuerte correlación lineal entre las variables, como lo indica el elevado coeficiente de Pearson de 0.9829378120964237.

Esta fuerte correlación para el "data_safrinha" reafirma la habilidad del modelo de bosque aleatorio en capturar las tendencias de los datos, permitiendo predicciones precisas y confiables para este período. El modelo no solo alcanza un desempeño superior en la predicción de los datos de safrinha, sino también revela la capacidad de los algoritmos de Inteligencia Artificial para adaptarse y aprender eficientemente de patrones consistentes y fuertemente lineales.

Los datos comparativos de los conjuntos "data_safra" y "data_safrinha" ilustran que el éxito en la modelización predictiva depende significativamente de la naturaleza y calidad de los datos de entrada. Las diferencias entre los dos períodos refuerzan la importancia de ajustar los modelos predictivos conforme a las características específicas de cada ciclo de producción de semillas. La análisis subraya la necesidad de una aproximación cuidadosa y contextualizada en la aplicación de técnicas de modelización, con el objetivo de maximizar la eficacia de las predicciones, esenciales para el planeamiento y la gestión agrícola.

6 CONCLUSIONES DEL ESTUDIO

Este estudio detallado, articulando el uso de modelos de bosque aleatorio, delinea un cuadro detallado de las dinámicas de producción y reserva de semillas en el contexto agrícola brasileño, ofreciendo una base sólida para futuras estrategias decisionales en el ámbito del sector. La



investigación subraya la complejidad intrínseca de las prácticas agrícolas en el país, iluminando las variaciones entre ciclos de cultivo distintos y regiones diversificadas. Con una mirada analítica y estadística, el estudio mapea la interacción entre una miríada de factores determinantes, enfatizando su interconectividad e influencia mutua.

La implementación del modelo de bosque aleatorio probó su capacidad analítica, administrando con eficiencia la extensa gama de datos del sector y extrayendo patrones que conducen a predicciones confiables. Aunque las variaciones en la precisión de las predicciones entre los ciclos de Safra y Safrinha hayan sido significativas, la aplicación del modelo resaltó la adaptabilidad necesaria para acomodar estas fluctuaciones, un aspecto clave para la modelización predictiva en la agricultura.

La concentración de la producción en un número limitado de cultivos emerge como una doble faceta, representando tanto una fortaleza económica como un punto de vulnerabilidad ante fluctuaciones del mercado y cambios climáticos. El estudio resalta, por tanto, la importancia crítica de promover la diversificación como un pilar para reforzar la resiliencia y la sostenibilidad del sector. Estrategias de gestión agrícola basadas en datos robustos son señaladas como fundamentales para responder con eficacia a las dinámicas descritas, implicando la necesidad de ajustes criteriosos en las prácticas de cultivo y en la planificación de la cosecha.

El trabajo señala igualmente la necesidad de investigaciones futuras más incisivas, que puedan desvelar aún más profundamente las relaciones entre las variables que afectan la producción agrícola y perfeccionar los modelos predictivos utilizados. La investigación abre perspectivas para la integración de datos climáticos, económicos y agronómicos en un enfoque multidisciplinar, con el objetivo de mejorar la capacidad predictiva y la gestión del sector.

Al proporcionar contribuciones valiosas tanto para la literatura científica como para la práctica agrícola, el estudio apoya a agricultores, gestores y formuladores de políticas en la toma de decisiones informadas y respaldadas por evidencias. En suma, el presente estudio no solo ha aclarado las complejas configuraciones productivas de semillas en Brasil, sino también ha ratificado la relevancia y la eficacia de los modelos de bosque aleatorio, representando un avance notable en la intersección entre ciencia de datos y agricultura.



REFERENCIAS

ADAMA BRASIL. Inteligência artificial na agricultura: quais são as principais funcionalidades? Disponível em: <https://www.adama.com/brasil/pt/inovacao/inteligencia-artificial-na-agricultura-quais-sao-principais-funcionalidades>. Acesso em: 17 jan. 2024.

CLIMATE FIELDVIEW. Quais as aplicações da inteligência artificial na agricultura? Disponível em: <https://blog.climatefieldview.com.br/inteligencia-artificial-agricultura>. Acesso em: 17 jan. 2024.

GADOTTI, Gizele I.; ASCOLI, Carla A.; BERNARDY, Ruan; MONTEIRO, Rita de C. M.; PINHEIRO, Romário de M. Machine Learning For Soybean Seeds Lots Classification. Scientific Paper, Special Issue: Artificial Intelligence. Eng. agríc. (Online), v. 42, spe. 2022. DOI: [org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022](https://doi.org/10.1590/1809-4430-Eng.Agric.v42nepe20210101/2022). Disponível em: <https://www.scielo.br/j/eagri/a/LtTLRpzgNQPWp5mw3qRMdtM/?format=pdf&lang=en>. Acesso em: 15 jan. 2024.

MOURTZINIS, S.; ESKER, P.D.; SPECHT, J.E.; et al. Advancing agricultural research using machine learning algorithms. Sci Rep, v. 11, 17879, 2021. DOI: [org/10.1038/s41598-021-97380-7](https://doi.org/10.1038/s41598-021-97380-7). Disponível em: <file:///C:/Users/User/Downloads/s41598-021-97380-7.pdf>. Acesso em: 14 jan. 2024.

REHAGRO. Inteligência artificial na agricultura: benefícios e aplicações. Disponível em: <https://rehagro.com.br/blog/inteligencia-artificial-na-agricultura/>. Acesso em: 17 jan. 2024.

SALEEM, M.H.; POTGIETER, J.; ARIF, K.M. Automation in Agriculture by Machine and Deep Learning Techniques: A Review of Recent Developments. Precision Agric, v. 22, p. 2053-2091, 2021. DOI: [org/10.1007/s11119-021-09806-x](https://doi.org/10.1007/s11119-021-09806-x). Disponível em: <https://link.springer.com/article/10.1007/s11119-021-09806-x>. Acesso em: 15 jan. 2024.

VIEIRA FILHO, José Eustáquio Ribeiro; SILVEIRA, José Maria Ferreira Jardim da. Mudança tecnológica na agricultura: uma revisão crítica da literatura e o papel das economias de aprendizado. Revista de Economia e Sociologia Rural, v. 50, n. 4, p. 651-666, dez. 2012. Disponível em: <https://www.scielo.br/j/resr/a/Pjz4mbbbKwDz8Vm4sbDY7mR/?format=pdf&lang=pt>. Acesso em: 18 jan. 2024.