

DATOS DE PRODUCCION DE CULTIVOS INDIA

Leslie Johana Galviz

Cristian Machado

2024

****Docente:** Yennyfer Portilla Yela**

Universidad del Valle - Facultad de Ingeniería

Probabilidad y Estadística

Santiago de Cali

Tabla de Contenido

- Introducción
- Objetivos
- Informe sobre Variables de Datos de Cultivo en India
- Métodos
- Resultados De La Investigación
- Conclusiones
- Referencias

Introducción

En un mundo donde la seguridad alimentaria y la sostenibilidad agrícola son cada vez más cruciales, comprender las dinámicas de la producción de cultivos se convierte en un pilar fundamental para la formulación de políticas agrícolas y la toma de decisiones estratégicas en el sector agroalimentario.

India es uno de los grandes productores de alimentos a nivel mundial. Tiene cultivos clave, como arroz, trigo, caña de azúcar, algodón, frutas y verduras. Específicamente, India produce alrededor del 10.9% de las frutas y el 8.6% de las verduras a nivel mundial. También es el mayor productor de legumbres, cubriendo aproximadamente el 25% de la producción global y satisfaciendo el 27% de la demanda mundial de esta importante fuente de proteínas. Además, India lidera la producción de leche y contribuye significativamente a la producción de especias, té y ganado.

Este estudio se centra en un extenso conjunto de datos sobre la producción de cultivos en India, abarcando más de 246,000 registros a lo largo de varios años. La información incluye diversos tipos de cultivos, métricas de producción y datos regionales que reflejan la compleja realidad agrícola del país. Analizaremos las tendencias temporales y regionales en la producción de cultivos, identificando patrones significativos y variaciones regionales que influyen en la eficiencia y productividad agrícola.

Objetivos

Objetivo Principal

Desarrollar modelos predictivos que ayuden a anticipar futuros niveles de producción y detectar riesgos potenciales en la producción de alimentos en India, aplicando técnicas de series de tiempo y otras herramientas estadísticas, con el fin de predecir el comportamiento de la producción alimentaria en los próximos tres años y su impacto en la economía global.

Objetivos Específicos

- Recopilar y analizar datos históricos de producción agrícola en India para identificar tendencias, patrones estacionales y fluctuaciones en la producción de diversos alimentos a lo largo del tiempo.
- Identificar y evaluar los riesgos potenciales que podrían afectar la producción de alimentos, como cambios climáticos, plagas, y políticas gubernamentales, y desarrollar estrategias para mitigar estos riesgos en función de las proyecciones obtenidas de los modelos predictivos.

Informe sobre Variables de Datos de Cultivo en India

1. State_Name (Nombre del Estado)

- **Descripción:** Representa el nombre del estado de la India donde se registraron los datos del cultivo.
- **Importancia:** Permite analizar variaciones en la producción agrícola entre diferentes estados, ayudando a identificar áreas con mayor potencial de producción y aquellas que podrían requerir apoyo adicional.
- **Clasificación:** Cualitativa.

2. District_Name (Nombre del Distrito)

- **Descripción:** Especifica el distrito dentro del estado donde se recopilaron los datos del cultivo.
- **Importancia:** Facilita un análisis más detallado a nivel regional, permitiendo la identificación de patrones locales y el diseño de políticas específicas para mejorar la producción en distritos con bajo rendimiento.
- **Clasificación:** Cualitativa.

3. Crop_Year (Año de Cultivo)

- **Descripción:** Indica el año en el que se cosechó el cultivo.
- **Importancia:** Es esencial para el análisis de tendencias temporales en la producción agrícola, permitiendo comparar datos de diferentes años y evaluar el impacto de factores externos como cambios climáticos o políticas agrícolas.
- **Clasificación:** Cuantitativa.

4. Temporada (Temporada Agrícola)

- **Descripción:** Indica la temporada agrícola (por ejemplo, Kharif, Rabi) durante la cual se cultivó el cultivo.
- **Importancia:** Comprender las temporadas de cultivo ayuda a anticipar períodos de producción y escasez, así como a planificar mejor el uso de recursos hídricos y otros insumos agrícolas.
- **Clasificación:** Cualitativa.

5. Cultivo (Tipo de Cultivo)

- **Descripción:** Identifica el tipo de cultivo que se cultivó (por ejemplo, arroz, trigo, maíz).
- **Importancia:** Permite segmentar la producción agrícola y estudiar las preferencias del mercado, así como las necesidades de los agricultores en función de los cultivos específicos.
- **Clasificación:** Cualitativa.

6. Área (Superficie)

- **Descripción:** Representa la superficie total de tierra (en hectáreas o acres) utilizada para el cultivo.
- **Importancia:** Proporciona datos sobre la escala de producción y la disponibilidad de tierras para cultivos, lo que es crucial para estimar la capacidad de producción y la eficiencia del uso del suelo.
- **Clasificación:** Cuantitativa.

7. Producción (Cantidad Total)

- **Descripción:** Indica la cantidad total (en toneladas métricas) del cultivo producido en el área especificada.
- **Importancia:** Esta variable es clave para el análisis de la productividad agrícola y la planificación de la oferta en el mercado. Ayuda a prever el comportamiento futuro de la producción y su impacto en la economía global.
- **Clasificación:** Cuantitativa.

Métodos

La limpieza de datos es un paso esencial para garantizar la calidad y fiabilidad de los análisis realizados. En este estudio, se busca optimizar el conjunto de datos sobre la producción de cultivos en India para obtener conclusiones más precisas.

Proceso de Limpieza

1. **Carga de Datos:** Se cargaron múltiples archivos CSV desde un directorio y se combinaron en un único conjunto de datos.
2. **Separación de Columnas:** Inicialmente, todos los datos estaban en una sola columna. Se separaron en varias columnas representativas como: Estado, Distrito, Año de Cultivo, Temporada, Cultivo, Área y Producción.
3. **Detección de Datos Faltantes:** Se realizó un análisis de los datos faltantes, identificando registros incompletos mediante un mapa de datos faltantes.
4. **Conversión de Tipos de Datos:** Las columnas se convirtieron a sus tipos adecuados, asegurando que los valores numéricos se trataran correctamente para análisis futuros.
5. **Eliminación de Datos Faltantes y Duplicados:** Se eliminaron filas con datos faltantes y registros duplicados para mantener la integridad del conjunto de datos.
6. **Identificación y Eliminación de Valores Atípicos:** Se revisaron los datos en busca de valores atípicos utilizando gráficos, y se eliminaron aquellos que no cumplieran con criterios razonables.
7. **Normalización de Texto y Filtrado:** Se normalizó el texto de las columnas para estandarizar la nomenclatura y se filtraron registros no válidos, asegurando que solo se mantuvieran entradas con valores positivos y años dentro de un rango razonable.

Proceso de Procesamiento de Datos

Ahora contamos con un conjunto de datos más preciso, lo que nos permite incorporar nuevas variables para la formulación de conclusiones. Utilizando diversas técnicas estadísticas, podremos llegar a conclusiones de alta calidad, minimizando la dispersión de datos y garantizando una interpretación más clara y relevante para poder culminar el objetivo.