

GRUPO 23

PROYECTO APRENDIZAJE NO SUPERVISADO

Desarrollo de Tipologías de Agricultores de Papa en Colombia
mediante Clustering No Supervisado

Integrantes:

- JAIME ANDRES UNRIZA VARGAS
- JHON FARLEY ADARVE DIAZ
- JHON NICOLAS GARZON RODRIGUEZ
- JINNETH CAROLINA PARADA AMAYA
- VICTOR ERNESTO VELANDIA SUAREZ



1. Resumen

El proyecto tiene como objetivo generar tipologías para los agricultores de papa en Colombia utilizando técnicas de aprendizaje no supervisado, específicamente métodos de clustering como K-medias y K-medoides. Utilizaremos información recolectada del Censo Nacional Agropecuario (CNA) y la Unidad de Planificación Rural Agropecuaria (UPRA). A través de la segmentación de agricultores según sus rendimientos y características socioeconómicas, esperamos proporcionar una herramienta útil para la toma de decisiones en políticas públicas y en la planificación del sector agrícola. Este proyecto está orientado principalmente a instituciones gubernamentales y organizaciones que trabajan en la mejora de la productividad agrícola y la sostenibilidad.

2. Introducción

El sector agrícola en Colombia enfrenta múltiples desafíos relacionados con la variabilidad de los rendimientos agrícolas y las condiciones socioeconómicas de los productores. En particular, los agricultores de papa, un cultivo clave en la economía agrícola del país, presentan una amplia diversidad en términos de prácticas agrícolas, acceso a recursos y niveles de producción. Esta diversidad requiere una caracterización precisa para la implementación de políticas y programas más eficaces.

El problema que abordamos en este proyecto es la falta de una caracterización clara y precisa de los agricultores de papa en Colombia, lo cual limita la eficacia de las políticas públicas y los programas de apoyo agrícola. Nuestra pregunta de investigación es: ¿Cómo podemos segmentar a los agricultores de papa en grupos homogéneos basados en sus rendimientos y características socioeconómicas, utilizando técnicas de clustering no supervisado?

La papa es un cultivo estratégico para la seguridad alimentaria y la economía agrícola de Colombia. Sin embargo, las disparidades en los rendimientos y el acceso a recursos entre los agricultores de papa impiden que las intervenciones sean efectivas. Una caracterización precisa permitirá identificar necesidades específicas y diseñar estrategias focalizadas que mejoren la productividad y la sostenibilidad del cultivo en el país.

El cliente potencial de este proyecto son las entidades gubernamentales, como el Ministerio de Agricultura y la UPRA (Unidad de Planificación Rural Agropecuaria), así como organizaciones agrícolas y cooperativas de productores. En el contexto organizacional, la falta de segmentación adecuada dificulta la implementación de políticas agrícolas eficientes. Este proyecto proporcionará una herramienta para mejorar la planificación y ejecución de programas de apoyo al agricultor.

Utilizando técnicas de clustering no supervisado, esperamos identificar patrones ocultos en los datos que no serían evidentes con métodos tradicionales. Esto permitirá clasificar a los agricultores en grupos homogéneos, basados en similitudes en sus prácticas y resultados, lo que a su vez facilitará la creación de políticas y programas más efectivos para mejorar la productividad y la equidad en el sector agrícola. Este problema pertenece al área de clustering dentro del aprendizaje no supervisado, donde buscamos agrupar datos multidimensionales en clusters significativos que representen diferentes tipologías de agricultores.

3. Revisión Preliminar de la Literatura

Revisamos literatura sobre la segmentación de agricultores y la aplicación de técnicas de clustering en el sector agrícola para establecer las similitudes y diferencias de los enfoques presentados en la literatura con el enfoque propuesto en nuestro proyecto:

- **Estudios sobre la segmentación de agricultores basados en características socioeconómicas y rendimientos.**
[Análisis y segmentación de productores agrarios para la transferencia del riesgo](#)
- **Aplicaciones de K-medias y K-medoides en la agricultura para la optimización de recursos.**
[Minería de datos espacial en la agricultura en Latinoamérica - Una aproximación conceptual](#)
- **Proyectos previos de tipologías en otros cultivos o sectores que puedan ofrecer un enfoque metodológico aplicable a nuestro caso.**
[Importancia de los objetivos sociales, ambientales y económicos de los agricultores en la adopción de maíz mejorado en Chiapas, México](#)

A continuación presentamos las similitudes y diferencias de los enfoques presentados en los artículos revisados frente al enfoque propuesto en nuestro proyecto:

Aspecto	Proyecto Actual (Tipologías en el Sector Agrícola)	Artículo 1: Análisis y Segmentación de Productores Agrarios para la Transferencia del Riesgo	Artículo 2: Minería de Datos Espacial en la Agricultura en Latinoamérica	Artículo 3: Importancia de los Objetivos Sociales, Ambientales y Económicos en la Adopción de Maíz Mejorado
Objetivo Principal	Desarrollar una tipología para agricultores de papa utilizando datos del CNA y UPRA.	Segmentar a los productores agrarios para mejorar la transferencia del riesgo mediante seguros agrícolas.	Revisar técnicas de minería de datos espacial en agricultura, con enfoque en Latinoamérica.	Analizar los objetivos de los agricultores de Chiapas en la adopción de maíz mejorado utilizando AHP y segmentarlos en base a opiniones y actitudes.
Técnicas Utilizadas	Clustering, análisis de datos agrarios y socioeconómicos para identificar grupos.	Análisis estadístico para la segmentación de productores.	Minería de datos espacial: clasificación, co-localización, clustering espacial, geovisualización.	Análisis de Componentes Principales (ACP), Análisis de Conglomerado (AC), Proceso Analítico Jerárquico (AHP).
Enfoque Metodológico	Segmentación basada en características socioeconómicas y de producción.	Segmentación basada en riesgos climáticos y de mercado.	Uso de algoritmos de minería de datos espacial para descubrir patrones no evidentes y útiles.	Segmentación de agricultores según opiniones, actitudes, aversión al riesgo, y priorización de objetivos económicos, socioculturales y ambientales.
Aplicación Geográfica	Específicamente a agricultores de papa en Colombia.	Análisis a nivel regional en el contexto del Perú.	Revisiones de casos en Asia y Latinoamérica, con énfasis en la optimización agrícola en países en desarrollo.	Agricultores de maíz en Chiapas, México.
Resultados Esperados	Identificación de grupos homogéneos de agricultores para mejorar políticas y apoyo sectorial.	Mejora en la segmentación para transferir riesgos y mejorar el acceso a seguros agrícolas.	Identificación de patrones espaciales en datos agrícolas, optimización del uso de recursos agrícolas.	Segmentación de agricultores en función de sus actitudes hacia la tecnología agrícola, con recomendaciones de estrategias diferenciadas para la adopción.
Contribución del Trabajo	Mejora en la caracterización de productores, contribuyendo a políticas agrícolas más precisas.	Desarrollo de herramientas para la transferencia del riesgo en el sector agrícola.	Base conceptual para la aplicación de técnicas de minería de datos espacial en la agricultura.	Propuesta de estrategias diferenciadas para mejorar la adopción de tecnología agrícola en diferentes segmentos de agricultores.

4. Descripción de los Datos

Los datos provienen del Censo Nacional Agropecuario (CNA) y la UPRA, y contienen información detallada sobre las Unidades Productoras Agropecuarias (UPA) de papa en Colombia. Las variables incluyen rendimiento por hectárea, características socioeconómicas de los agricultores, y prácticas agrícolas, entre otros. Presentaremos estadísticas descriptivas y visualizaciones iniciales para explorar la distribución de las variables y su relevancia para el proyecto.

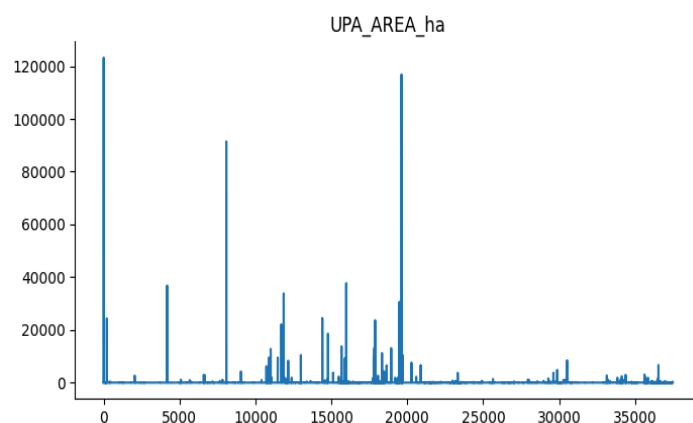
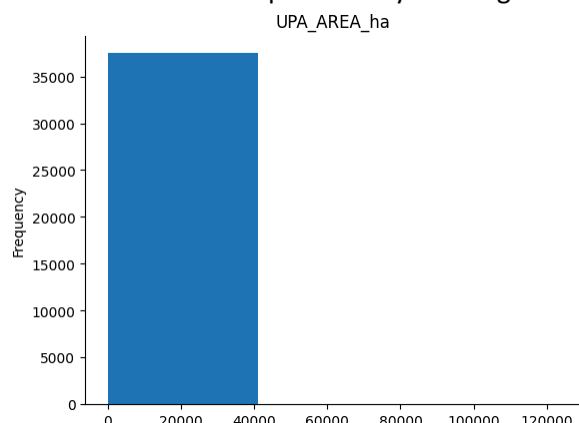
Una primera aproximación nos muestra que la base de datos disponible cuenta con 37,484 observaciones cada una asociada a una unidad productiva (UPA), y con 24 columnas o variables distribuidas de la siguiente forma (13 variables numéricas - 11 variables categóricas).

Variables numéricas: Son características que describen a cada UPA en términos de capacidad productiva y permiten entender un poco mejor su comportamiento y sus especialidades. A continuación se presentan algunas estadísticas descriptivas:

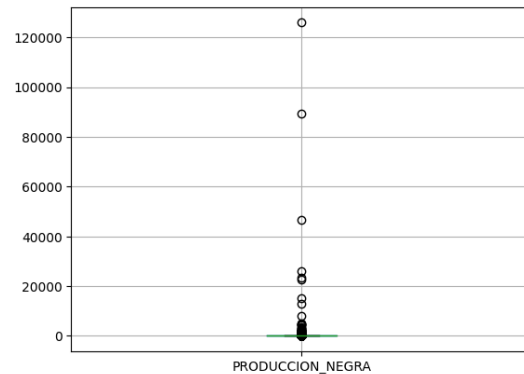
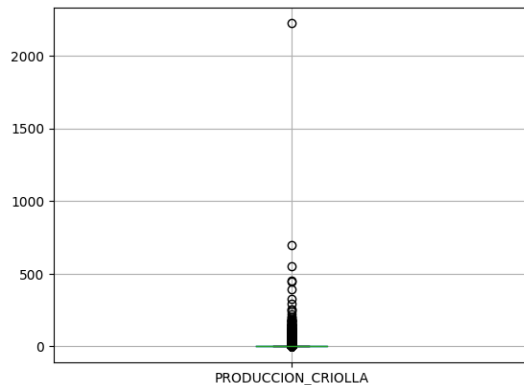
	UPA_AREA_ha	AREA_SEMBRADA_CRIOLLA	AREA_COSECHADA_CRIOLLA	PRODUCCION_CRIOLLA	RENDIMIENTO_CRIOLLA	AREA_SEMBRADA_NEGRA	AREA_COSECHADA_NEGRA
count	37484.000000	6221.000000	6221.000000	6221.000000	1258.000000	32869.000000	32869.000000
mean	30.816426	1.531621	0.276534	4.947853	19.086428	3.190960	2.371406
std	1118.108164	12.054708	2.044221	36.231268	3.510013	71.816320	63.757661
min	0.007254	0.010017	0.000000	0.000000	0.000273	0.010026	0.010026
25%	0.530939	0.105607	0.000000	0.000000	17.831738	0.127145	0.095182
50%	1.680011	0.321600	0.000000	0.000000	18.676614	0.539060	0.393605
75%	5.190562	1.000000	0.000000	0.000000	20.348524	1.905341	1.372049
max	123198.560100	572.782411	123.417969	2222.049240	112.083752	8192.000000	8192.000000

A continuación se presentan algunas características destacables de este análisis exploratorio:

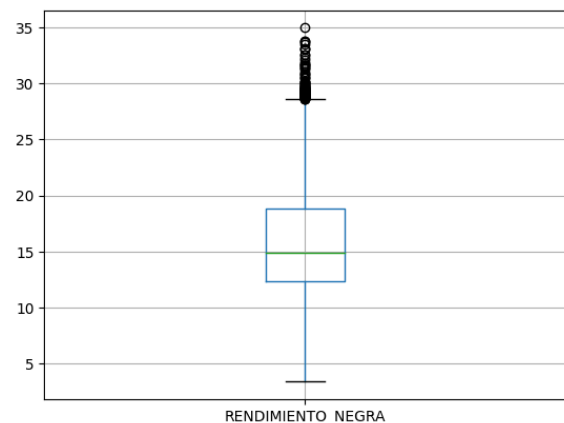
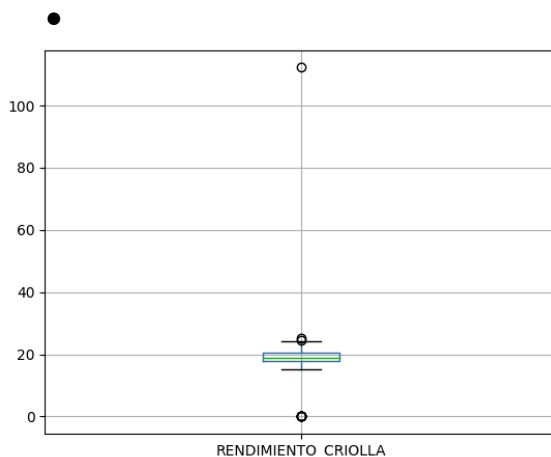
- La unidad productiva promedio cuenta con 30.8 hectáreas disponibles, con valores extremos (mínimo = 0.007, máximo = 123198.5). Un análisis gráfico muestra que la distribución de los datos se concentra alrededor del rango 0 - 40,000 hectáreas, con muy pocas observaciones saliéndose de ese intervalo, lo que sugiere en primera instancia que se cuenta con un grupo de UPAs con una diversidad de capacidades y estrategias de producción interesante.



- Se cuenta con datos de producción para dos variedades de papa (Criolla - Negra) y se encuentra una diferencia marcada entre unidades que se dedican a la producción de cada variable (6221 - Criolla, 32869 - Negra), lo que genera diferencias interesantes en los números de producción por toneladas, cómo se observa en los siguientes gráficos (una producción media de Criolla de 4.947 toneladas, comparada con las 38.094 para la variedad Negra).



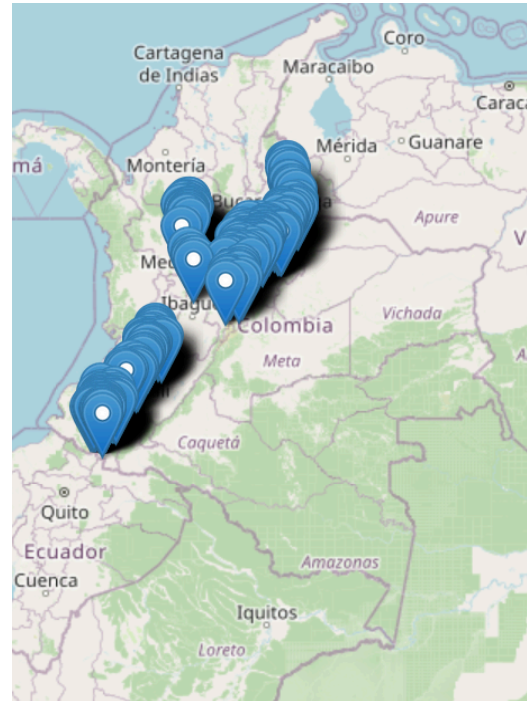
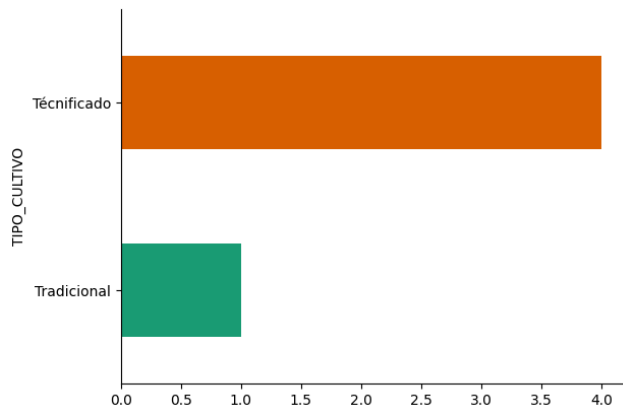
- También se tienen estadísticas de los rendimientos (cantidad producida por unidad de área de cultivo) de los cultivos por unidad productiva, con resultados que reflejan que para los datos disponibles, en promedio la variedad Criolla presenta mejores rendimientos para las UPAs que se dedicaron a su producción (19.086428 toneladas por hectárea para la variedad criolla, en comparación con las 15.589 de la variedad negra).



Variables categóricas - Geográficas: Son características más asociadas a los aspectos socio-económicos de cada unidad productiva, por ejemplo en este grupo de variables es posible encontrar la clasificación a la que pertenece cada unidad de acuerdo a su extensión territorial, o su estatus económico (informal, etc.). Estas dimensiones enmarcan a cada productor dentro de clasificaciones más claras y dan un contexto social al análisis.

A continuación se presentan algunas conclusiones:

- Las técnicas de cultivo modernizadas prevalecen dentro del grupo de estudio, lo que se podría interpretar como evidencia de los avances en la política agropecuaria que apoya a los diversos productores del país (facilidades de acceso a nuevas tecnologías, facilidades crediticias, etc.)



- Las unidades productivas estudiadas se ubican dentro de la zona agrícola más prolífica del país, la región andina que abarca todo el centro de Colombia siguiendo la ruta de la cordillera de los Andes; se caracteriza por su abundancia de suelo fértil, lo que sumado a sus características climáticas la convierten en la zona ideal para el desarrollo de diversos tipos de cultivos.

5. Metodología

Se implementarán técnicas de K-means y K-medoides para el clustering de los agricultores. Dependiendo de los resultados iniciales, se podrían considerar métodos adicionales como clustering jerárquico o DBSCAN. La elección final del método se basará en la capacidad de los clusters resultantes para proporcionar una tipología útil y accionable.

6. Bibliografía

1. Artículo 1:

- Salazar, A. & Espinosa, R. (2016). *Análisis y segmentación de productores agrarios para la transferencia del riesgo*. Revista de Estudios Agrarios, 24(2), 34-48.

2. Artículo 2:

- Gil-Torres, A. F., Monroy-García, A. L., & González-Sanabria, J. S. (2019). *Minería de datos espacial en la agricultura en Latinoamérica: Una aproximación conceptual*. Revista de Tecnología Agrícola, 35(3), 123-137.

3. Artículo 3:

- Sánchez-Toledano, B. I., Kallas, Z., & Gil, J. M. (2017). *Importancia de los objetivos sociales, ambientales y económicos de los agricultores en la adopción de maíz mejorado en Chiapas, México*. Revista Mexicana de Ciencias Agrícolas, 8(3), 269-287.