

19-21 de Octubre 2022 | Granada

INTERNATIONAL CONFERENCE ON REGIONAL SCIENCE

Challenges, policies and governance of the territories in the post-covid era

Desafíos, políticas y gobernanza de los territorios en la era post-covid

XLVII REUNIÓN DE ESTUDIOS REGIONALES

XIV CONGRESO AACR



COMUNICACIÓN

Título: Datos abiertos para la prospectiva y dinámica del mercado agrícola. El caso del café

Autores y e-mails de todos:

Luiza Petrosyan
(luipet@inf.upv.es)

Instituto Universitario de Matemática Pura y Aplicada
Universitat Politècnica de València

Alicia García García
(agarciaga@unav.es)
Universidad de Navarra

Fernanda Peset
(mpesetm@upv.es)
Instituto Universitario de Matemática Pura y Aplicada
Universitat Politècnica de València

A. Ferrer Sapena
(anfersa@upv.es)
Instituto Universitario de Matemática Pura y Aplicada
Universitat Politècnica de València

Andrea Sixto Costoya
(andrea.sixto@uv.es)
Unidad de Información e Investigación Social y Sanitaria – UISYS
Universitat de València

Área Temática: *(S14 Generación de escenarios prospectivos en la Nueva Ruralidad)*

Resumen

El objetivo de esta investigación es realizar un análisis prospectivo para analizar la dinámica del mercado agrícola y su oferta comercial a partir de recoger datos abiertos de distintos recursos. Se ha establecido para ello un protocolo sistemático de descarga de datos de diferentes fuentes para generar una masa de información suficiente que permita aplicar procedimientos de análisis utilizando la inteligencia artificial y estableciendo de manera automática los parámetros de análisis. Se ha elegido como caso de estudio el café

y utilizado como fuentes de descarga automática las recogidas en Agrelma (<https://www.agrelma.com/>), Alibaba (<https://www.alibaba.com/>) y Twitter. Además, se han escogido como fuentes de información tres portales gubernamentales de datos abiertos de los países que son los mayores productores de café actualmente: Brasil, Colombia y Vietnam. Los resultados obtenidos han permitido desarrollar una herramienta de análisis capaz de buscar y recopilar toda la información pública posible sobre la dinámica del mercado del café.

Introducción

En la actualidad los datos abiertos son uno de los elementos clave para el ámbito de la agricultura. A partir de su uso se está desarrollando la agricultura 4.0 y numerosas aplicaciones que permiten incrementar el rendimiento de las producciones agrícolas, siendo los datos abiertos un input fundamental para la agricultura de precisión (Top et al., 2022). Estos datos se encuentran aportados fundamentalmente por la administración, que a su vez deben contener los principios reconocidos por la Open Knowledge Foundation, donde reúnan las siguientes características: disponibles y accesibles (que los datos estén abiertos a la ciudadanía), reutilizables y redistribuibles (posibilidad de reutilizar y compartir los datos) y de participación universal (que cualquiera pueda utilizar los datos) (Abella et al., 2019; Open Knowledge Foundation, 2013; Royo-Montañés & Benítez-Gómez, 2019). Los datos que nos aporta la administración y las distintas instituciones gubernamentales aportan información de aspectos como es la producción, el clima o las parcelas, entre otros. Además de estas fuentes de datos, existen otras que proporcionan empresas privadas como Agrelma o Twitter y otras que provienen de sensores instalados en las parcelas de manera individual o colectiva.

De acuerdo a las categorías establecidas por el proyecto Global Open Data For Agriculture and Nutrition (GODAN) para el sector agrícola, se precisan 14 tipologías de datos, entre los que destacan los datos geográficos, los datos meteorológicos y los datos de mercado (Global Open Data for Agriculture and Nutrition, 2019). Estas 14 tipologías se agrupan en los siguientes apartados: 1) datos de la legislación y de la administración; 2) datos sobre recursos naturales, de la tierra y del medio ambiente; 3) datos agronómicos y de tecnologías agrícolas; y 4) datos socioeconómicos. Aunque muchos de estos datos los proporcionan las administraciones y, en función del grado de apertura de las mismas, se pueden localizar en mayor o menor medida, es también probable que se deba recurrir a otro tipo de datos estadísticos de diversas fuentes que pueden complementar los datos que se encuentran en los portales de datos abiertos.

En este contexto, el objetivo del presente estudio, que se engloba dentro del proyecto DATAUSE (<https://datause.es/>), es analizar la dinámica del mercado agrícola, concretamente del café como producto, y su oferta comercial a partir de la recogida de datos abiertos de distintas fuentes.

Metodología

- Fuentes de obtención de datos

Se ha establecido un protocolo sistemático de descarga de datos de diferentes fuentes para generar una masa de información suficiente para poder aplicar procedimientos de análisis utilizando la inteligencia artificial y estableciendo de manera automática los parámetros de análisis. Para ello, se han escogido como fuentes de información los datos de los portales de venta online Agrelma y Alibaba y la red social Twitter. En el caso de Agrelma, se ha elegido como fuente por su perfil especializado en la venta de productos agrícolas, ya que permite el contacto directo con los productores, importadores y centros de compra, y está dirigido fundamentalmente a grandes productores y compradores. Por su parte, Alibaba ha sido seleccionado por ser portal de venta directa a consumidores de gran impacto a nivel internacional. Finalmente, se ha elegido la red social Twitter por tener una mejor extracción de los contenidos que circulan por la misma, pudiendo realizar búsquedas más precisas que en otras redes. Además, se considera la elección de varios portales de datos abiertos gubernamentales que, en sus secciones dedicadas a los datos sobre el sector agrícola, se ajusten lo más posible a las tipologías estipuladas por GODAN.

- Producto escogido

Una vez determinadas las fuentes de datos, se procede a escoger el producto a analizar, que en este caso ha sido el café. La elección del mismo ha venido motivada porque es la tercera bebida más consumida en el mundo, la segunda materia prima más comercializada después del petróleo y uno de los productos a los que afecta el cambio climático de manera más directa, lo que supone dificultades añadidas para su producción (Global Open Data for Agriculture and Nutrition, 2019).

- Análisis de datos

Con respecto al análisis de los datos abiertos proporcionados por los gobiernos, se han analizado los portales del gobierno de Colombia (gov.co), de Brasil (dados.gov.br) y de Vietnam (data.gov.vn). La elección de los mismos es debido a que son los tres principales productores de café en la actualidad (Foreign Agricultural Service, 2019; International Coffee Organization, 2020).



Figura 1. Los 3 principales portales de datos abiertos de los países productores como fuente de datos

Por otro lado, se analizan los datos provenientes de los mercados Agrelma y Alibaba para identificar aspectos como quiénes son los productores de café, los comerciantes, los países implicados, los precios, y las condiciones de venta. Por lo que respecta a los datos

estructurados y textuales de Twitter, el análisis se ha enfocado en averiguar la siguiente información: las cuentas más activas, los temas relevantes y las relaciones entre los actores del mercado.

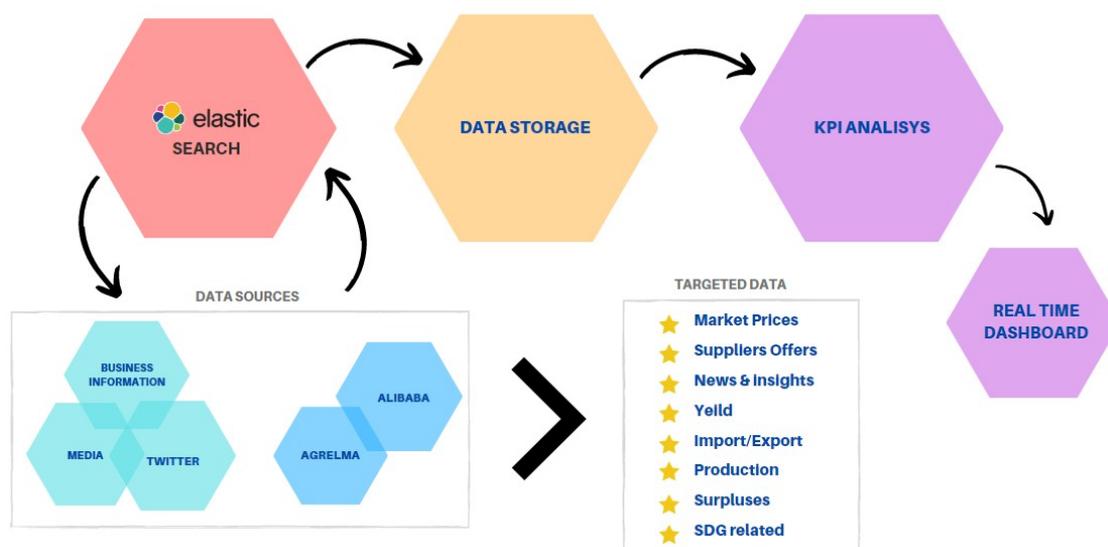


Figura 2. Diagrama del análisis y procesamiento de datos

Resultados

La relación entre ambas fuentes complementarias (Agrelma y Alibaba y la red social Twitter), junto con los portales de datos abiertos, ha permitido caracterizar la dinámica del mercado del café y construir los perfiles reales de los actores.

Tras analizar los portales de datos abiertos de los países seleccionados, se ha observado que el de Colombia ofrece gran cantidad de datos organizados de acuerdo a las categorías planteadas por GODAN, seguido del portal de datos de Brasil y, por último, el de Vietnam. Esto corrobora lo ya reflejado en el Global Open Data Index realizado por la Open Knowledge Foundation en 2017 (Global Open Data Index, 2017).

Finalmente, los resultados de este estudio, insertado dentro en el marco del proyecto DATAUSE, han permitido la construcción de una herramienta (Digiproof AI) de análisis de los mercados por parte de nuestros colaboradores KPI RISK ETHICS & COMPLIANCE capaz buscar y recopilar toda la información pública posible que responda a los objetivos planteados. La herramienta se basa en una metodología OSINT (Open-Source Intelligence) que es inteligencia de fuentes abiertas. Consiste en una serie de procesos que tienen como misión hacer uso de fuentes de carácter público para poder buscar y recopilar toda la información pública posible sobre el objetivo en concreto con el fin de poder interpretar esa información y darle una utilidad.

Conclusiones

La modernización y digitalización de los sistemas agrarios ha ido evolucionando en las últimas décadas lidiando y superando sus propias barreras y dificultades. Entre estas dificultades, estaría el hecho de que se trata de un sector a menudo disperso, donde a

veces es complicado acercar tecnologías ligadas al uso de los datos, o incluso acercar el propio acceso a un Internet de calidad, a las personas que están trabajando en el día a día en el campo. Por ello, todas las contribuciones que persigan acortar esta distancia entre el trabajo a pie de campo y el acceso a los datos son un plus para la mejora del I+D en el sector agrícola. Proyectos como DATAUSE y propuestas como la que se presenta en este estudio, que buscan diseñar y optimizar herramientas para acercar las fuentes de información que ofrecen los datos abiertos al trabajo en el ámbito de la agricultura, son un paso más para seguir avanzando en un contexto donde cada vez es más necesario tener en cuenta el uso y explotación de los datos para conseguir la máxima eficiencia.

Palabras Clave: *Café, Datos agrícolas, Datos abiertos, Reutilización.*

Clasificación JEL: O1, O13, Q13, Q17,

Referencias bibliográficas

Abella, A., Ortiz-de-Urbina-Criado, M. & De-Pablos-Herederó, C. (2019). The process of open data publication and reuse. *Journal of the Association for Information Science and Technology*, 70(3), 296–300. <https://doi.org/10.1002/asi.24116>

Foreign Agricultural Service. (2019). *Coffee: World Markets and Trade*. <https://downloads.usda.library.cornell.edu/usda-esmis/files/m900nt40f/sq87c919h/8w32rm91m/coffee.pdf>

Global Open Data for Agriculture and Nutrition. (2019). *Agriculture*. <https://www.godan.info/pages/agriculture>

Global Open Data Index. (2017). *Place overview*. <https://index.okfn.org/place/>

International Coffee Organization. (2020). *Coffee Production Report*. <https://www.ico.org/prices/po-production.pdf>

Open Knowledge Foundation. (2013). *Defining Open Data*. <https://blog.okfn.org/2013/10/03/defining-open-data/>

Royo-Montañés, S. & Benítez-Gómez, A. (2019). *Portales de datos abiertos. Metodología de análisis y aplicación a municipios españoles*. 05, 1–13.

Top, J., Janssen, S., Boogaard, H., Knapen, R. & Şimşek-Şenel, G. (2022). Cultivating FAIR principles for agri-food data. *Computers and Electronics in Agriculture*, 196(February). <https://doi.org/10.1016/j.compag.2022.106909>