



Universidad del
Rosario

Facultad de Economía

Maestría en Economía de las Políticas Públicas

MÉTODOS COMPUTACIONALES DE POLÍTICAS PÚBLICAS 2020-II

Aproximación a dinámica del sector arrocerero, según Fedearroz

Ivonne Paola Ubaque Galán

5 de diciembre de 2020

APROXIMACIÓN A DINÁMICA DEL SECTOR ARROCERO, SEGÚN FEDEARROZ

AGENDA

1. Propuesta.
2. Métodos y herramientas utilizadas.
3. Resultados.
Periodos analizados.
Aspectos económicos.
Problemáticas.
4. Conclusiones.

1. PROPUESTA

- Se buscaba conocer los principales temas del gremio arrocero desde el primero de enero de enero de 2011.
- TLC con EEUU: mayo de 2012. Colombia otorgó un contingente de importación de 98 mil ton a partir del 6 año de entrada en vigor.
- Consumo per cápita anual: 41 kilos.
- Cerca de 500 mil familias siembran el arroz en 215 municipios ubicados en los Llanos Orientales, Tolima, Huila, Santanderes y algo en departamentos de la Costa Norte como Magdalena, Córdoba, Sucre y César.
- 500.000 hectáreas que están destinadas a la siembra de arroz y arrojan una producción aproximada de 1,7 millones de toneladas anuales.
- Los Llanos tienen mayor participación en la siembra con aproximadamente el 44,4%, la zona centro siembra el 29,5%, en el bajo Cauca el 13,6%, en la zona de los Santanderes el 7,8 % y el restante 4,5 % se cultiva en la Costa Norte.

2. MÉTODOS Y HERRAMIENTAS UTILIZADAS

seguro | fedearroz.com.co/new/index.php

Scraping a página
www.fedearroz.com

- Se extraen datos de 2942 noticias, desde el 8 de septiembre de 2008 a la fecha presente (30 de octubre de 2020).
- Se usaron librerías e instrumentos como: expresiones regulares, request, BeautifulSoup, numpy y matplotlib.

```
In [19]: #En este paso se identifica similitud en las URL de noticias para realizar el respectivo llamado de informaci
#para lo cual se encuentra similitud en la consecuencia de numero de noticia, de esta manera se genera un vec
#los link de noticias
```

```
noticias = re.findall('id=[0-9]+',str(links))
enlace = 'http://www.fedearroz.com.co/noticias/noticiasd2.php?'
datos = []
for i in range(len(noticias)):
    datos.append(enlace + noticias[i])
    print(datos[i])
print(len(datos))
```

```
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3223
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3224
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3221
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3207
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3206
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3200
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3199
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3198
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3220
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3197
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3196
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3202
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3203
http://www.fedearroz.com.co/noticias/noticiasd2.php?id=3204
```

Con expresiones regulares *findall* se extrae el *id* de cada una de las noticias para evadir la ventana emergente e ir directamente a la noticia.

Con *request* se realiza la extracción de la noticia. Luego se elimina texto basura, etiquetas y demás datos irrelevantes.

```
In [30]: #En este modulo se consulta todas las noticias y se limpia de texto basura que pudieron
#ser llamados en la extracción de datos del sitio web
```

```
texto =[]
textosunidos =''
for i in range(2924):
    html = requests.get(datos[i]).text
    soup = BeautifulSoup(html)
    type(soup)
    body.append(soup.find_all('p',{'class': "body"}))

    temporal = str(body[i])
    temporal = re.sub('\xa0\xa0|<p|class="body">|<n|</p>|<br/>|align="right"|vspace="5"|>|align="default"|</u>','',temporal)
    temporal = temporal.replace('[',',')
    temporal = temporal.replace(']',',')
    textosunidos = textosunidos+ temporal
    texto.append(temporal)
```

```
Out[31]: textosunidos
```

```
Out[31]: ' <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3223.jpg" <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3224.jpg" <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3221.jpg" <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3207.jpg" La Federación Nacional de Arroceros, Fedearroz - Fondo Nacional del Arroz realizará el próximo 26 de noviembre la Reunión Técnica Nacional y Feria Tecnológica del Sector Arrocero 2020. Entrada libre previa inscripción <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3206.jpg" Para volver a ver el Webinar: Manejo eficiente de la fertilización para optimizar la nutrición en el cultivo del arroz <img border="1" hspace="5" src="../../noticias-fotos/reductor.php?tamano=350&archivo=3200.jpg" Socialización de herramientas y procesos tecnológicos de eficiencia y nutrición de la planta de Arroz o del cultivo del arroz Bogotá D.C. 5 de octubre de 2020.La Federación Nacional de Arroceros - Fedearroz, rechaza enérgicamente los hechos ocurridos en la madrugada de hoy contra las instalaciones de la Planta de Secamie
```



```

n= len(fecha)
fechas_unidas=''
años =[]
for i in range(n):

    temporal = str(fecha[i])
    temporal = re.sub('<span|class="fecha">|</span>', '',temporal)
    año = re.findall('[0-9]+', temporal)
    años.append(año)
    fechas_unidas = fechas_unidas+temporal
    fechas.append(temporal)

notiXaño = []
añoLabel = []
actual = 2020
contador = 0

n= len(años)
#print (n)
for i in range(n):
    if(actual!=2007):
        if años[i][1] == str(actual):
            contador = contador+1
        else:
            notiXaño.append(contador)
            añoLabel.append(actual)

            actual = actual-1
            contador = 0
    else:
        i=i+n

```

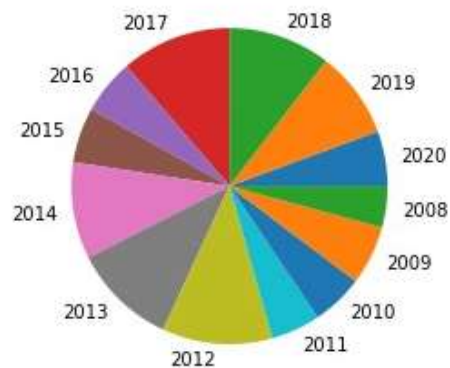
Con *expresiones regulares* para identificar los años más relevantes en gremio arrocero.

3. RESULTADOS

Periodos analizados

```
print(tab.tabulate(tabla, headers='firstrow',tablefmt='fancy_grid'))  
plt.pie(notiXaño, labels=añoIlabel)  
plt.show()
```

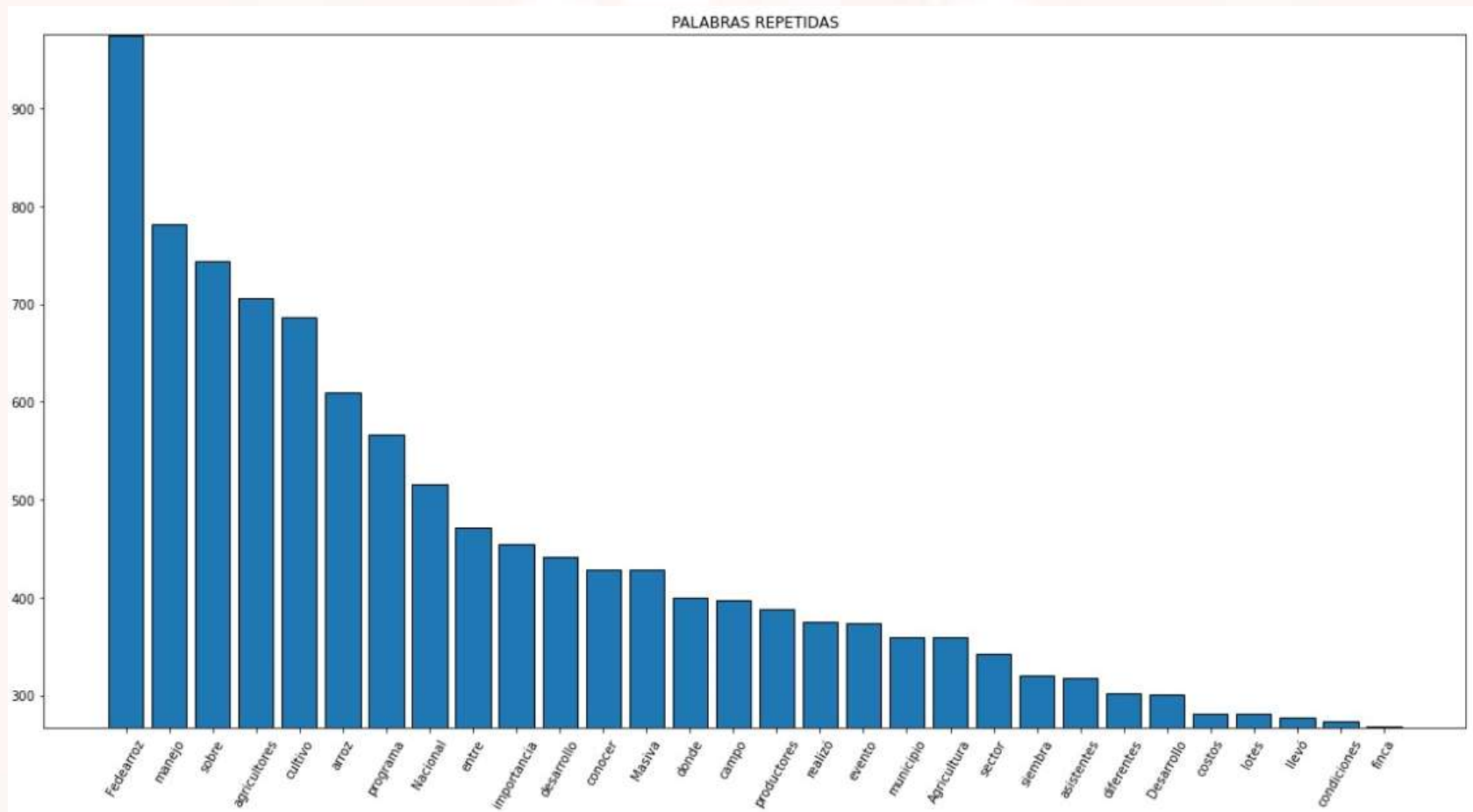
2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
164	259	305	330	163	165	288	310	328	147	155	176	123



2017: año de mayor producción arrocerá en el país.

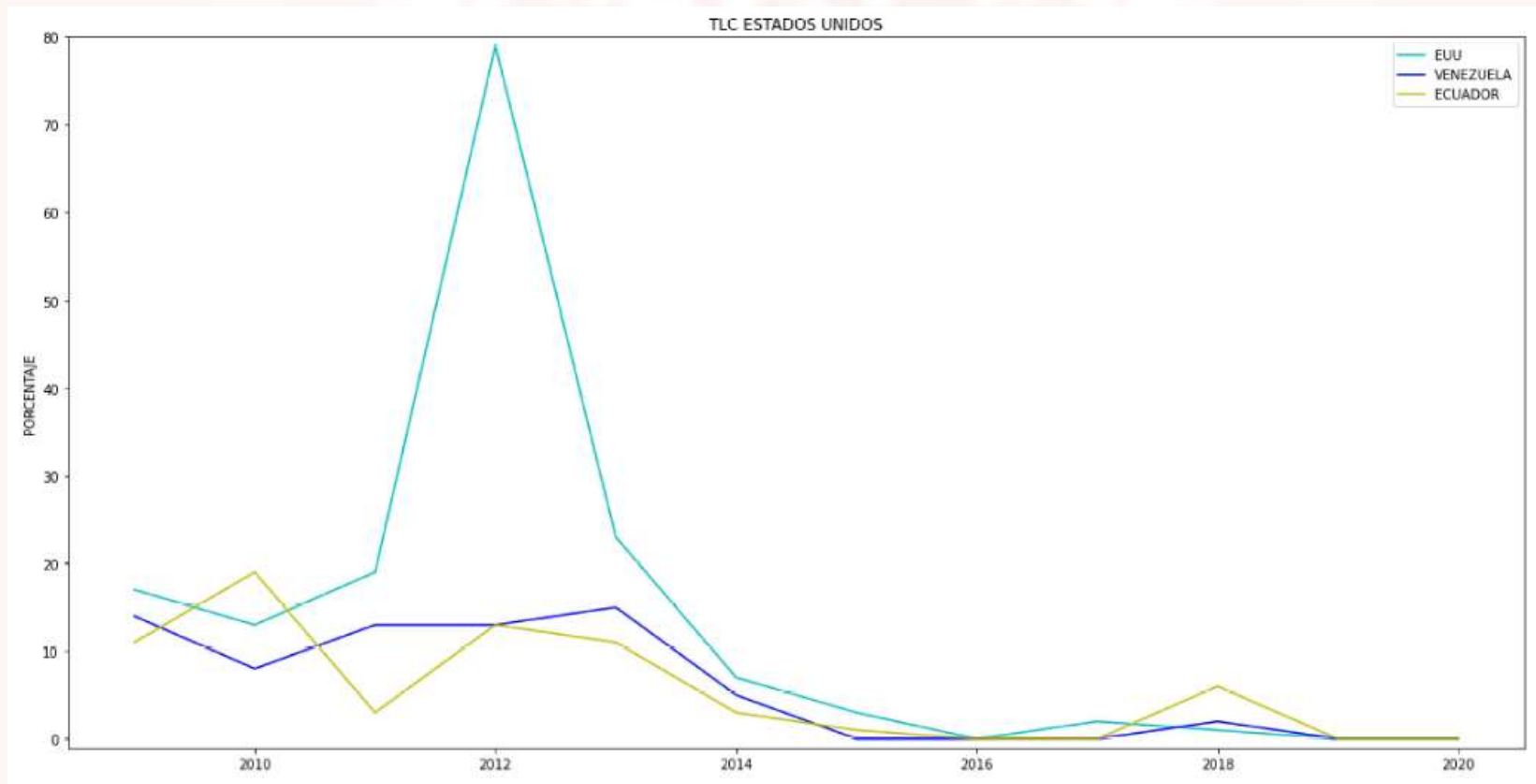
3. RESULTADOS

Aspectos económicos y problemáticas.



3. RESULTADOS

Aspectos económicos y problemáticas.



3. RESULTADOS

Aspectos económicos y problemáticas.

```
print('Arroz Roa',textosunidostxt.count(' roa '))  
print('Cartel',textosunidostxt.count('Cartel'))  
print('Sanción',textosunidostxt.count('sanción'))  
print('Flor huila',textosunidostxt.count('Flor huila'))  
print('Consejo de estado',textosunidostxt.count('Consejo de estado'))  
print('UNIÓN DE ARROCEROS',textosunidostxt.count('UNIÓN DE ARROCEROS'))
```

```
Arroz Roa 0  
Cartel 0  
Sanción 0  
Flor huila 0  
Consejo de estado 0  
UNIÓN DE ARROCEROS 0
```

4. CONCLUSIÓN

- En términos técnicos se observa que no hay parametrización en la publicación de la información.
- Se encontraron muchas etiquetas en la estructura del sitio web.
- En cuanto al alcance del contenido de las noticias, se evidencia un manejo modulado de la información.
- Siendo el gremio oficial de los productores arroceros, registra escasa información relevante y actualizada de la realidad nacional de un producto básico en la dieta de los colombianos.



Universidad del
Rosario

Facultad de Economía

Maestría en Economía de las Políticas Públicas

GRACIAS POR SU ATENCIÓN