



**TRABAJO DE GRADO
Seminario**

**ALGORITMO COMPUTACIONAL PARA EL ANÁLISIS Y TOMA DE DECISIONES
EN DATOS DE VENTAS EN LA LICORERA JC GRANIZADOS, UTILIZANDO
ESTRATEGIAS DE MACHINE LEARNING**

Corporación Universitaria Remington.
Facultad de ingeniería
Tecnología en desarrollo de software

JUAN SEBASTIAN BUSTAMANTE BEDOYA
Tutor: Juan Carlos Briñez de León
Seminario
2024.

TABLA DE CONTENIDOS

RESUMEN	3
PALABRAS CLAVE.....	4
MARCO CONCEPTUAL Y CONTEXTUAL.....	4
PREGUNTA PROBLEMA	7
ACERCAMIENTO DE DATOS	8
DESCRIPCIÓN DE VARIABLES.....	8
POSIBLES APLICACIONES	9
APROXIMACIONES CON GRAFICOS	9
OBJETIVOS	12
OBJETIVO GENERAL:	12
OBJETIVOS ESPECIFICOS:	12
DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE	13
PROCESAMIENTO DE DATOS	17
RESULTADOS ADICIONALES.....	18
IMPLEMENTACIÓN EN CONTEXTOS REALES.....	19
CONCLUSIONES	19
REFERENCIAS	20

RESUMEN

La presente investigación se realizó con el fin de tener una predicción de ventas a partir de un modelo de regresión.

En este se utilizó la herramienta de machine learning con el modelo KNNR, utilizando los datos recolectados en la licorería JC granizados ubicada en el municipio de Sabaneta, Antioquia.

Se tuvieron en cuenta variables como el mes, la fecha, que día del fin de semana es, la base y el total de ingresos por día, con los datos recolectados a partir del software del establecimiento se realizó un procesamiento donde se analizaron las fechas exactas de cada día registrado, para así verificar si pertenecían al viernes, sábado, domingo o festivo, posterior a esto fueron presentados al algoritmo para tener los resultados solicitados.

Gracias a este modelo se puede predecir el valor de las ventas del fin de semana próximo y así el establecimiento realice la toma de decisiones, ya sea una estrategia de marketing para aumentar las ventas o también saber cuánta cantidad de pedido debe de realizar y si necesita apoyo de más personal.

Este modelo puede ser utilizado por los establecimientos encargados de la comercialización de productos ya que les permitiría conocer cómo van sus ventas y como mejorarlas.

PALABRAS CLAVE

Machine Learning, Algoritmo, Datos, Licorería, Licores, Marketing, Google Colab, Facturación, Ventas, Proveedor, Regresión, Predicción, Proyección.

MARCO CONCEPTUAL Y CONTEXTUAL

En Colombia se pueden encontrar diversos establecimientos donde se ofrecen diferentes productos relacionados con el licor, en muchos lugares son llamados licoreras o licorerías. También son consideradas en muchas ocasiones micro empresas ya que son unidades económicas que desarrollan comercialización de bebidas con o sin alcohol. (Del Pilar, 2019)

Las licorerías son de diferentes tipos, entre ellos están aquellos donde se puede comprar y beber en el establecimiento, allí adquieren un servicio en el punto. Otro tipo es donde solo ofrecen la venta, entrada y salida de licores y otras bebidas al por mayor. El mercado de bebidas alcohólicas es uno de los rubros que más utilidades y beneficios presenta, siendo América Latina uno de los continentes que más consumo de bebidas alcohólicas presenta, como tal el rubro es una buena oportunidad de negocio, por lo mismo es uno de los que más competencia tiene. (Guevara & Gordillo, 2023)

Siendo Colombia el país ubicado en la posición 12 en consumo de licor a nivel mundial según un estudio realizado en el año 2019, por esta razón una licorería es un buen emprendimiento en nuestro país.

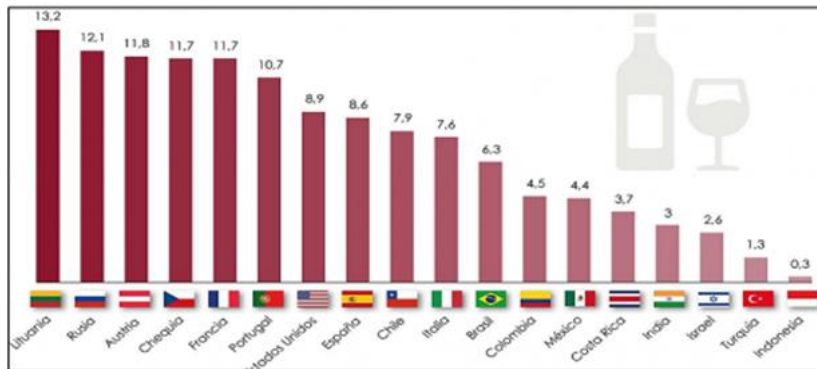


Figura 1. Países más bebedores en el año 2019

En Colombia no solo se distribuye en puntos como licorerías, sino también en tiendas de barrio o almacenes de cadena, además existe venta ilegal de bebidas alcohólicas en estado de adulteración lo que constituye una competencia desleal. (Zuleta & Jaramillo, 2000)

Todo esto hace que las licorerías tengan una gran desventaja, por ello debe de dar un valor agregado como lo es un buen servicio y música para disfrutar.

Para determinar las ganancias del establecimiento se tienen en cuenta variables como fechas y horarios, también el estrato socioeconómico donde esté ubicado, ya que este determina las preferencias por tipo de licor, en casos tales como lo son el Whisky (estratos de ingresos altos), el aguardiente (estratos de ingresos medios y bajos), el ron y el brandy (estratos de ingresos medios). (Zuleta & Jaramillo, 2000)

Para el comercio de licor también es muy importante sus proveedores ya que son los encargados de distribuir las bebidas con o sin alcohol, con la garantía que están en un

estado óptimo para el consumo de las personas, esto hace que los compradores tengan mayor confianza en el establecimiento comercial.

En la actualidad existen diversos software para organizar las finanzas de algún establecimiento, estos facilitan la automatización de procesos, pero en ocasiones es necesario conocer modelos de predicción para ver qué tan viable es la empresa, uno de los algoritmos más utilizados es el machine learning el cual automatiza la construcción de modelos de análisis de datos, este es una rama de la inteligencia artificial que basa su desarrollo en sistemas que aprenden de datos, identifican patrones y toman decisiones. (Manrique, 2020)

El machine learning se encarga de resolver situaciones a partir de los datos aportados, entre más datos se proporcionan se tendrán mejores resultados de predicción de eventos. Se deben de seguir una serie de pasos para realizar un modelo con base a este algoritmo, estos son:

1. Colección de datos
2. Procesamiento de datos
3. Exploración de datos
4. Se entrena el algoritmo
5. Evaluación de los algoritmos
6. Uso del modelo. (Manrique, 2020)



Figura 2. Pasos para modelo

También se utiliza un entorno de desarrollo que puede ser Google colab, esta es una herramienta de un servidor, adaptada para su integración con Google drive y GitHub, en esta se puede ejecutar notebooks tanto en la nube como en el entorno local. (Gavilanez & Saragozin, 2024)

PREGUNTA PROBLEMA

¿Cómo desarrollar una estrategia de predicción y marketing para próximos fines de semana en la licorera JC Granizados?

ACERCAMIENTO DE DATOS

Los datos fueron obtenidos de la licorería JC Granizados y se usaron Historiales de ventas por día (fin de semana), recibos de pagos y proveedores. Todo esto proviene un software interno y lo ofrece el establecimiento donde se aplicará la estrategia. Se realizó una tabla de Excel con cinco columnas, las cuales son mes, día, fin de semana, base y total de ventas. Con estos datos lograr implementar varias estrategias de marketing para mejorar las ventas y el servicio ofrecido en el establecimiento.

DESCRIPCIÓN DE VARIABLES

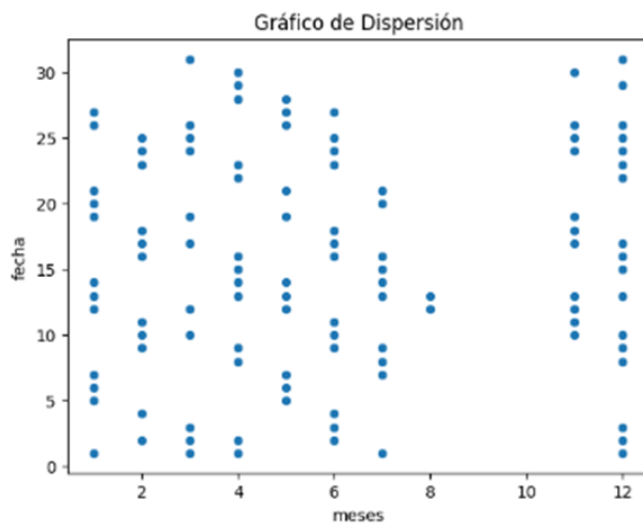
En la base de datos obtenida se encuentran datos tanto cualitativos como cuantitativos, entre las variables tenemos (mes, fecha) con esto nos podemos ubicar en el tiempo, (base) es con el capital con el que se inicia en caja las horas laborales en el establecimiento, (total de ventas) es el total de ganancias generado en un día de trabajo. Con esto se da una contextualización de las ganancias generadas enfocado en los fines de semana (viernes, sábado, domingo, festivos), teniendo en cuenta que la base de datos fue proporcionada por la licorería JC Granizados.

POSIBLES APLICACIONES

Con el modelo y la aplicación del algoritmo de machine learning se puede establecer una predicción de los días en fines de semana (viernes, sábado, domingo, Festivo) con más y menos ventas, así poder realizar estrategias de marketing como promociones en punto de venta en fechas especiales, eventos deportivos y fiestas a nivel local, además poder tener en cuenta las compras a proveedores. Este modelo de predicción puede ser aplicado a diferentes empresas y emprendimientos que comercializan productos.

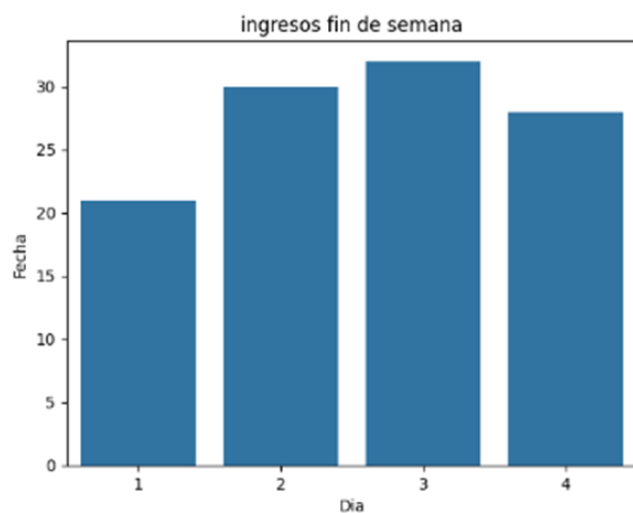
APROXIMACIONES CON GRAFICOS

Figura 3. Grafica de dispersión



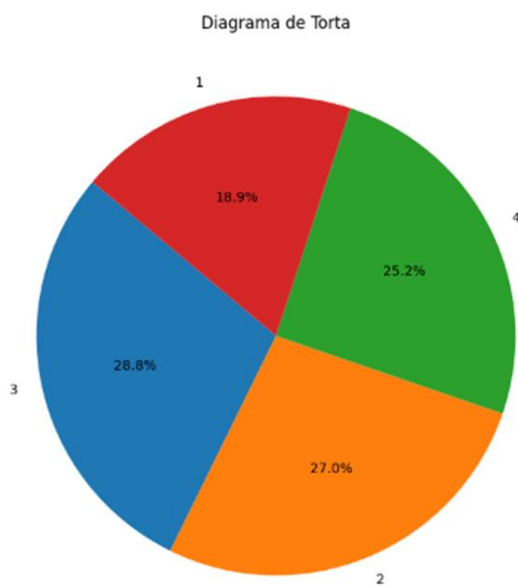
En la siguiente grafica se observa la dispersión de los datos, de los meses (1 al 12) con relación a las fechas registradas (1 al 31 de cada mes). Se puede observar que el mes que más tuvo movimientos y ventas es el mes 12 (diciembre).

Figura 4. Diagrama de barras



En este gráfico de barras, se evidencian los ingresos por fin de semana, aquí se realiza un análisis de las fechas (1 al 31 de cada mes) relacionando los días del fin de semana (Festivo: 1, viernes: 2, sábado: 3, domingo: 4). Se puede identificar que los días 2 – 3 son los días en los que más se ve movimiento y ventas.

Figura 5. Diagrama de torta



En el diagrama de torta se observa más detalladamente los porcentajes de cada uno de los días del fin de semana, identificando así, que el día que más movimiento y ventas presenta es el día 3 (sábado) con el 28.8%, seguido del día 2 (viernes) con el 27.0%.

OBJETIVOS

OBJETIVO GENERAL:

Implementar un algoritmo computacional para el análisis y toma de decisiones a partir de datos de JC GRANIZADOS, utilizando estrategias de machine learning.

OBJETIVOS ESPECIFICOS:

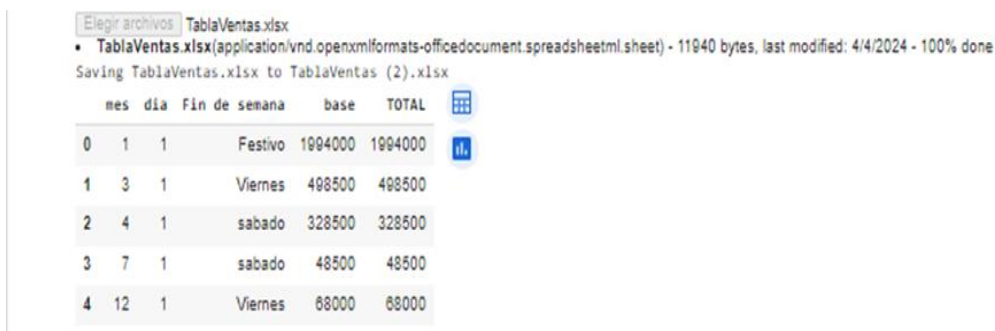
- Caracterizar y procesar los datos de interés, con miras a la toma de decisiones informadas.
- Implementar un algoritmo de Machine learning para la toma de decisiones en la licorera JC granizados a partir de los datos de interés.
- Evaluar y analizar el desempeño de los algoritmos implementados para la toma de decisiones.
- Validar el funcionamiento de toma de decisiones a partir de datos nuevos.

DESARROLLO E IMPLEMENTACIÓN DEL APRENDIZAJE

Se extrajeron los datos recolectados desde el 24 de marzo del 2023, hasta el 3 de marzo de 2024, estos fueron tomados desde el sistema de facturación de la licorería, haciendo énfasis en los fines de semana ya que son los días con mayor flujo de ventas, con esta información se evidenciaron valores fluctuantes, que tuvieron relación con fechas especiales, días de pago y/o eventos deportivos, estos serán la base para determinar las estrategias a realizar para la mejora de los ingresos.

Posterior a esto se realizó una tabla de Excel agrupando los datos, teniendo en cuenta 5 variables (mes, día, fin de semana, base y total de ventas). Usando el algoritmo se comenzó con la primera parte donde se validan y se dividen en entradas (Mes, Día, Fin de Semana, Base) y las salidas (Total de ventas). Luego se procede a tomar los modelos con los que se realizara el algoritmo para la predicción.

Se carga el archivo de Excel en el algoritmo de machine learning, donde se observan 5 variables (mes, día) que se organizan en orden ascendente, otra variable es (fin de semana) lo que indica que es fin de semana teniendo en cuenta los días (viernes, sábado, domingo, festivo), por último, el (total de ventas) con el cual se indica el total de ventas registradas en el día luego el cierre de caja.



mes	día	Fin de semana	base	TOTAL	
0	1	1	Festivo	1994000	1994000
1	3	1	Viernes	498500	498500
2	4	1	sabado	328500	328500
3	7	1	sabado	48500	48500
4	12	1	Viernes	68000	68000

Figura 6. Bustamante, J.

Se obtiene la información de los datos registrados, donde Mes, Día, Base y Total son variables Int (numéricas), por otro lado, Fin de Semana es Object (objeto).

```
[21] #Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111 entries, 0 to 110
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   mes              111 non-null    int64
1   día              111 non-null    int64
2   Fin de semana    111 non-null    object
3   base             111 non-null    int64
4   TOTAL           111 non-null    int64
dtypes: int64(4), object(1)
memory usage: 4.5+ KB
```

Figura 7. Bustamante, J.

Se hace análisis de datos.

```
[23] #Análisis de los datos
Conjunto_Datos.describe()
```

	mes	día	base	TOTAL
count	111.000000	111.000000	1.110000e+02	1.110000e+02
mean	5.972973	15.38038	3.188838e+05	3.188838e+05
std	3.745202	8.48253	3.500075e+05	3.500075e+05
min	1.000000	1.00000	4.500000e+03	4.500000e+03
25%	3.000000	9.00000	9.375000e+04	9.375000e+04
50%	5.000000	15.00000	2.119000e+05	2.119000e+05
75%	9.500000	23.00000	4.068000e+05	4.068000e+05
max	12.000000	31.00000	1.994000e+06	1.994000e+06

Figura 8. Bustamante, J.

Para un análisis más enfocado, se elimina la columna Total, ya que es el mismo valor de la base por tanto es innecesaria para la predicción.

	mes	día	Fin de semana	base
0	1	1	Festivo	1994000
1	3	1	Viernes	498500
2	4	1	sabado	328500
3	7	1	sabado	48500
4	12	1	Viernes	88000
5	2	2	Sabado	382500
6	3	2	sabado	207000
7	4	2	domingo	341200
8	6	2	Viernes	314000
9	12	2	Sabado	391000
10	3	3	Domingo	92000
11	6	3	sabado	813500
12	12	3	Viernes	803700
13	2	4	Sabado	88500
14	6	4	Domingo	533000
15	1	5	viernes	15000
16	5	5	Viernes	288000
17	1	6	Sabado	80000
18	5	6	Domingo	148000
19	1	7	viernes	52500

Figura 9. Bustamante, J

Se reemplazan las variables cualitativas a cuantitativas como:

Festivo = 1 viernes = 2 sábado = 3 domingo = 4

Así poder identificar más fácil los días y para que el algoritmo pueda cuantificar los resultados de manera gráfica.

```
# Cambiando rango de valores del dataframe
Estrategia = {'FESTIVO':1, 'VIERNES':2, 'SABADO': 3, 'DOMINGO': 4}
Conjunto_Datos['Fin de semana']=Conjunto_Datos['Fin de semana'].map(Estrategia)
Conjunto_Datos.head(20)
```

	mes	día	Fin de semana	base
0	1	1	1	1994000
1	3	1	2	498500
2	4	1	3	328500
3	7	1	3	48500
4	12	1	2	88000
5	2	2	3	382500
6	3	2	3	207000
7	4	2	4	341200
8	6	2	2	314000
9	12	2	3	391000
10	3	3	4	92000
11	6	3	3	613500
12	12	3	2	803700
13	2	4	3	88500
14	6	4	4	533000
15	1	5	2	15000
16	5	5	2	288000
17	1	6	3	80000
18	5	6	4	148000
19	1	7	2	52500

Figura 10. Bustamante, J.

PROCESAMIENTO DE DATOS

Procesando los datos, se ajustan como una matriz numérica donde se toman las 3 variables de entrada anteriores (Mes, Dia, Fin de semana) dando como salida la variable (Base).

```
0s [23] #Ajustando los datos a una matriz numérica
import numpy as np
Dataset_Matriz = np.array(Dataset)
#Dividiendo los datos en entrada y salida
Entradas = Dataset_Matriz[:,0:3]
Salida = Dataset_Matriz[:, -1]
```

Figura 11. Bustamante, J.

Luego se procede a elegir el modelo (KNNR) con el cual se hará la predicción de las ventas para el próximo fin de semana y así poder realizar estrategias de marketing en la licorería.

```
0s [30] #Modelo para predicciones de series de tiempo (KNN: Buen o si sabemos que los datos no se salen del rango)
from sklearn.neighbors import KNeighborsRegressor as KNNR
Modelo_1=KNNR()
Modelo_1.fit(Entradas, Salida)
```

+ KNeighborsRegressor
KNeighborsRegressor()

Figura 12. Bustamante, J.

Se identifican las 3 variables de entrada (Mes, Dia, Fin de semana), También se implementa el modelo de regresión (KNNR) el cual dará una proyección a futuro.

```

#Se ingresan los datos de entrada para generar una predicción
Nueva_entrada = np.zeros((1,3))
Nueva_entrada[0,0]=float(input('Ingrese el valor del mes(1-12): '))
Nueva_entrada[0,1]=float(input('Ingrese el valor del fecha(1-31): '))
Nueva_entrada[0,2]=float(input('Ingrese el valor del Fin de semana(1,2,3,4) : '))

Proyeccion_1 = Modelo_1.predict(Nueva_entrada)
#Proyeccion_2 = Modelo_2.predict(Nueva_entrada)

print('')
print('')
print('Según los datos ingresados, la proyección del siguiente fin de semana, usando KNN será: ',Proyeccion_1[0])
print('')
#print('Según los datos ingresados, la proyección para el siguiente mes usando ANN será: ',Proyeccion_2[0])

Ingrese el valor del mes(1-12): 

```

Figura 13. Bustamante, J.

RESULTADOS ADICIONALES

Finalmente ejecutamos el modelo, y se obtiene como resultado la predicción para el próximo fin de semana, teniendo en cuenta el mes, la fecha y que día del fin de semana se toma (festivo, viernes, sábado, domingo). Se ingresan los datos los cuales se usarán, arrojando como resultado la proyección de las ganancias de ventas para el próximo fin de semana.

```

#Se ingresan los datos de entrada para generar una predicción
Nueva_entrada = np.zeros((1,3))
Nueva_entrada[0,0]=float(input('Ingrese el valor del mes(1-12): '))
Nueva_entrada[0,1]=float(input('Ingrese el valor del fecha(1-31): '))
Nueva_entrada[0,2]=float(input('Ingrese el valor del Fin de semana(1,2,3,4) : '))

Proyeccion_1 = Modelo_1.predict(Nueva_entrada)
#Proyeccion_2 = Modelo_2.predict(Nueva_entrada)

print('')
print('')
print('Según los datos ingresados, la proyección del siguiente fin de semana, usando KNN será: ',Proyeccion_1[0])
print('')
#print('Según los datos ingresados, la proyección para el siguiente mes usando ANN será: ',Proyeccion_2[0])

Ingrese el valor del mes(1-12): 6
Ingrese el valor del fecha(1-31): 18
Ingrese el valor del Fin de semana(1,2,3,4) : 2

Según los datos ingresados, la proyección del siguiente fin de semana, usando KNN será: 379700.0

```

Figura 14. Bustamante, J.

IMPLEMENTACIÓN EN CONTEXTOS REALES

Machine learning es una inteligencia artificial que utilizada en el contexto de ventas genera predicciones a futuro con precisión y flexibilidad según los cambios del mercado. (Correa, 2023)

En la era digital esta herramienta es muy usada en diferentes empresas como modelo de predicción de demanda para así mejorar los procesos internos de esta, y dar mejores resultados. Muchas empresas han visto como procesos de venta, recolección de datos, compra de insumos, etc. se optimizan gracias al uso de machine learning.

CONCLUSIONES

- Maching learning es una herramienta de inteligencia artificial, que utiliza la recolección de datos para proyectar y mejorar la toma de decisiones.
- Este tipo de proyectos con base en machine learning pueden ser de mucha utilidad para empresas y establecimientos de comercio, ya que facilita la proyección de ventas y ganancias a futuro.
- Aplicando el modelo de machine learning se logro en la licorera JC Granizados pronosticar el valor aproximado de los ingresos para los próximos fines de semana, con esto realizar estrategias de marketing y pedidos a proveedores en caso de necesitarlo.
- Se busca a futuro que esta herramienta optimice los procesos de venta y así obtener un incremento en los ingresos de la licorera JC Granizados.

REFERENCIAS

- Correa, A. (01 de 02 de 2023). *bibliotecadigital.udea.edu.co*. Obtenido de bibliotecadigital.udea.edu.co:
https://bibliotecadigital.udea.edu.co/bitstream/10495/35745/3/CorreaAlejandro_2023_ModelosPrediccionDemanda.pdf
- Del Pilar, V. (01 de 01 de 2019). *repositorio.uladech.edu*. Obtenido de repositorio.uladech.edu:
https://repositorio.uladech.edu.pe/bitstream/handle/20.500.13032/13410/AUDITORIA_TRIBUTACION_HUANCAAYA_VELA_LEYDI_DEL_PILAR.pdf?sequence=4&isAllowed=y
- Gavilanez, A., & Saragozin, B. (01 de 03 de 2024). *repositorio.utc.edu.ec*. Obtenido de repositorio.utc.edu.ec: <https://repositorio.utc.edu.ec/bitstream/27000/12029/1/PI-002692.pdf>
- Guevara, L., & Gordillo, M. (02 de 02 de 2023). *repositorio.ujcm.edu.pe*. Obtenido de repositorio.ujcm.edu.pe:
https://repositorio.ujcm.edu.pe/bitstream/handle/20.500.12819/1812/Leidi-Maria_tesis_titulo_2023.pdf?sequence=1&isAllowed=y
- Manrique, E. (01 de 01 de 2020). *media.proquest.com*. Obtenido de media.proquest.com:
https://media.proquest.com/media/hms/PFT/1/0SI9G?_s=%2FDsLIWFwzl4jIAjKWZEDaaPRdU4%3D
- Zuleta, L., & Jaramillo, L. (01 de 10 de 2000). *www.repository.fedesarrollo.org*. Obtenido de www.repository.fedesarrollo.org:
https://www.repository.fedesarrollo.org.co/bitstream/handle/11445/986/Repor_Octubre_2000_Zuleta_y_Jaramillo.pdf?sequence=1&isAllowed=y