



# Análisis de descripciones de partidas

Ing. Aguilera Márquez Genaro  
Act. Castro Rodríguez Jorge Octavio  
Lic. Gómez Álvarez Leopoldo Javier  
Lic. Pérez Vázquez Ernesto



# AGENDA



Objetivo del  
Proyecto

1

Metodología  
de trabajo

3

Caso de  
negocio

2

Conclusiones

4



2

# Objetivo

- ◆ Por medio de mecanismos de procesamiento de lenguaje natural se realizará reconocimiento de patrones estructurales sobre las descripciones capturadas en el proceso de empeño de artículos en los puntos de recepción, con la finalidad de generar una agrupación de objetos basado en sus características.
- ◆ Con base a los grupos creados se plantea proporcionar a negocio la opción de enfocarse en el análisis de aquellos que representen y proporcionen mayor beneficio a la institución.



1



3

# Caso de negocio

Actualmente la captura de las descripciones de las partidas realizadas por los mecanógrafos se realiza sin el seguimiento de políticas definidas respecto a la precisión de las palabras capturadas; de igual manera no existe lineamientos para estandarizar las frases vs los artículos, lo anterior conlleva a tener un equipo de 10 personas analizando y clasificando todos los días los artículos de manera batch, lo que implica lo siguiente:

- (-) Precisión de clasificación (-\$)
- (+) Tiempo de procesamiento (-\$)
- (+) Tiempo en cambio de precios (-\$)
- (-) Apego (-\$)
- (+) Riesgos (-\$)



# Metodología de trabajo

Entendimiento  
del caso de  
negocio



3.1

Limpieza y  
Clasificación  
de las etiquetas



3.2

Tokenización,  
Stemming,  
frecuencia y  
Clasificación



3.3

TF-IDF y  
segmentación  
de la bolsa  
de palabras



3.4

Modelado



3.5

Autoetiquetado



3.6

Monitoreo/Mejoras

5

3.1

# ENTENDIMIENTO DEL CASO DE NEGOCIO

# Justificación del caso de negocio

## Paso 1

### Recepción

Las prendas son revisadas, evaluadas y se documenta la descripción de los artículos

## Almacenamiento

Se resguardan para su futura operación CLASIFICACIÓN A ALTO NIVEL

## Paso 2

### Clasificación

Se analizan las descripciones DE CADA UNA DE LAS PARTIDAS DE MANERA MANUAL PARA UNA CLASIFICACIÓN DETALLADA Y MOVIMIENTO DE PASE



SNMP

## Paso 3

## MOVIMIENTO

Movimiento físico

## Paso 4

Segmentación por  
importancia de negocio  
NLP + Modelado

3.2

## LIMPIEZA y CLASIFICACIÓN

# Limpieza y clasificación de las etiquetas

Diccionario

Palabra Original	Palabra Corregida
CHISP	CHISPA, DIAMANTE
CHURUM	CHURUMBELA
DIVINO	MOTIVO_0
ESPIG	MOTIVO_0
FILIG	FILIGRANA
TRENZAD	TRENZADA
ZIRC	CIRCONITA

ENUNCIADO ORIGINAL:  
DOS.ANILLOS....2.PAR.ARRACADA  
S.UNA.DE.ELLA.TORZAL....PAR.AR  
ETES....PAR.BROQULES.CALIBRAD  
OLA.ORO.BAJO.GRS.8.6.....D(0.6)

Descripción

Limpieza

Sin  
Stopwords

Stemming

Clasificación

ENUNCIADO SIN STOPWORDS: dos  
anillo 2 par arracadas torzal par  
aretes par broqules calibradola oro  
bajo grs 8.6 d 0.6

CLASIFICACION: oro

ENUNCIADO LIMPIO: dos anillos 2  
par arracadas una de ella torzal par  
aretes par broqules calibradola oro  
bajo grs 8.6 d(0.6)

ENUNCIADO CON STEMMS: dos  
anill 2 par arrac torzal par aret par  
broql calibradol oro baj grs 8.6 d 0.6

--- Clasificación ---  
-----  
Registros: 349999  
-----  
Oro: 300423  
Relojes: 23136  
Metales P: 10644  
Otros: 15796  
-----

Base Descripciones: 349,999

3.3

## TOKENIZACIÓN, STEMMING y FRECUENCIAS

# Tokenización, Stemming, frecuencia y Clasificación

Aplicando la **Clasificación** tenemos la posibilidad de ver que hay dentro de cada uno de los grupos definidos con anterioridad

```
--- Clasificación ---  
-----  
Registros: 349999  
-----  
Oro: 300423  
Relojes: 23136  
Metales P: 10644  
Otros: 15796
```

Tokenización

WORDS - ORO: ['collar', 'torzal', 'tres', 'colores', 'oro', 'mediano',...]

WORDS - RELOJES: ['reloj', 'dama', 'tag', 'hever', 'caratula', 'negra',...]

WORDS - OTROS: ['kit', 'herramienta', 'proskit', 'mod', 'ser', 'nv',...]

WORDS - METALES\_P: ['cien', 'monedas', 'plata', 'ley', '0.720', 'gramos',...]

Stemming

STEMMS - ORO: ['coll', 'torzal', 'tres', 'color', 'oro', 'median',...]

STEMMS - RELOJES: ['reloj', 'dam', 'tag', 'hev', 'caratul', 'negr',...]

STEMMS - OTROS: ['kit', 'herramient', 'proskit', 'mod', 'ser', 'nv',...]

STEMMS - METALES\_P: ['cien', 'moned', 'plat', 'ley', '0.720', 'gram',...]

Frecuencias

FREQ WORDS - ORO: [('oro', 304853), ('gramos', 228347), ('mediano', 134001), ('bajo', 113862),...]

FREQ WORDS - RELOJES: [('reloj', 22879), ('pulso', 19664), ('caja', 19081), ('acero', 17144),...]

FREQ WORDS - OTROS: [('mod', 8765), ('ser', 8145), ('manch', 6155), ('ray', 4937),...]

FREQ WORDS - METALES\_P: [('gramos', 7800), ('plata', 6806), ('ley', 5276), ('anillo', 2796),...]

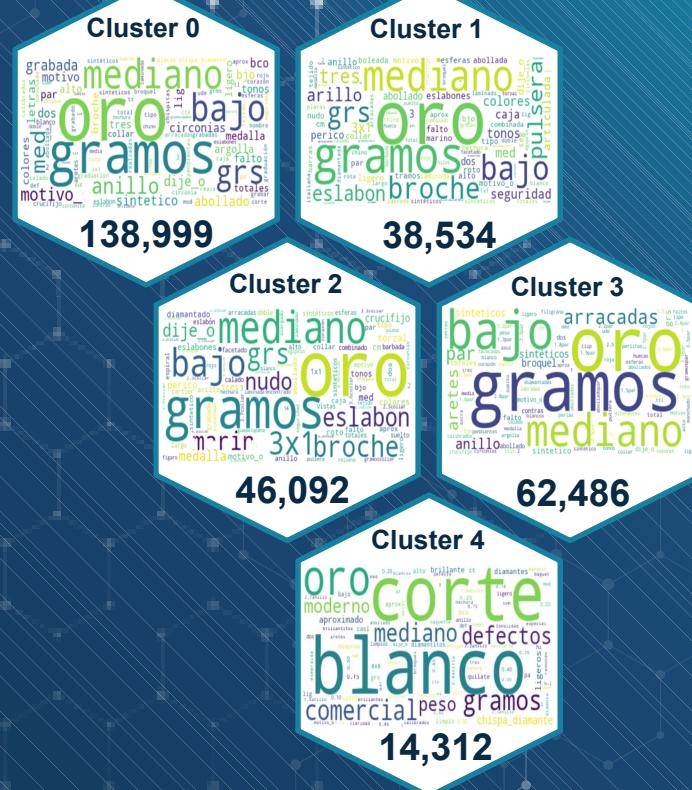
3.4

## TF-IDF y SEGMENTACIÓN

# TF-IDF y segmentación de la bolsa de palabras

○ Transformación  
TF-IDF

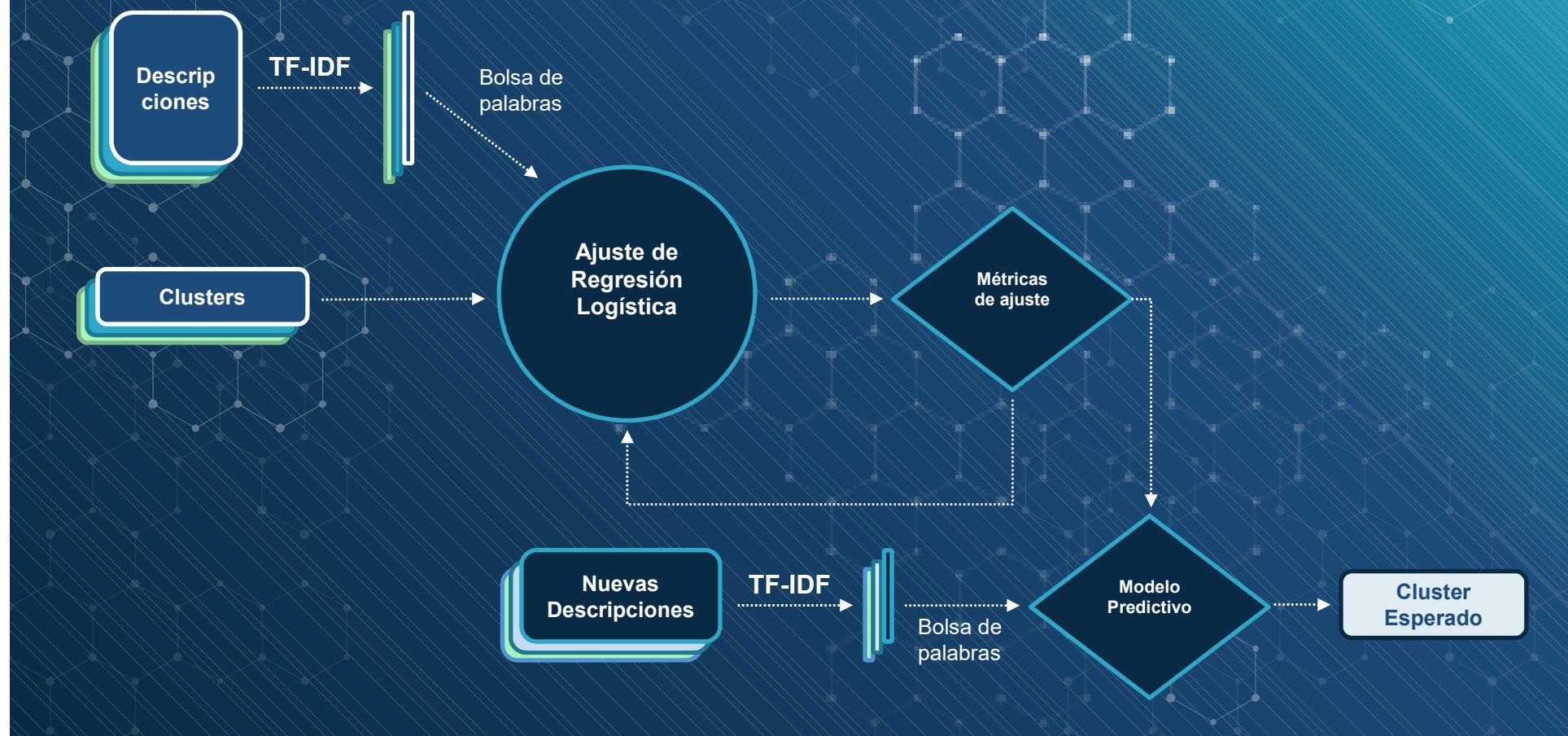
○ Clusterización  
K-Means



**3.5**

## AJUSTE DE MODELOS

# Ajuste de Modelos



**3.6**

## AUTOETIQUETADO

# Autoetiquetado



sents	clean_sents	stopwords_sents	stemms_sents	group_val	cluster_val
COLLAR NUDO MARINO MEDALLA GUADALUPANA GRABADA...	collar nudo marino medalla guadalupana grabada...	collar nudo marino medalla guadalupana grabada...	coll nud marin medall guadalupan grab oro baj ...	oro	2
CRUCIFIO SINTÉTICOS ORO MEDIANO GRAMOS 5.6.....	crucifijo sintéticos oro mediano gramos 5.6 me...	crucifijo sintéticos oro mediano gramos 5.6 me...	crucifij sintet oro median gram 5.6 medi 5.3 f...	oro	1
COLLAR NUDO MARINO ENCONTRADO FALTO DE BROCHE ...	collar nudo marino encontrado falto de broche ...	collar nudo marino encontrado falto broche oro...	coll nud marin encontr fait broch oro baj rot ...	oro	2
ESCLAVA ESLABON COMBINADO BROCHE DE PERICO LIG...	esclava eslabon combinado broche de perico lig...	esclava eslabon combinado broche perico liger...	esclav eslabon combin broch peric liger ray or...	oro	1
TABLETA ACER MOD Y SER NO VIS CABLE USB Y ADAP...	tableta acer mod ser no vis cable usb adapt de...	tableta acer mod ser vis cable usb adapt corr...	tablet acer mod ser vis cabl usb adapt corrien...	otros	0
...	...	...	...	...	...
COMPUTADORA TODO EN UNO HP MODELO PAVILLION TO...	computadora todo en uno hp modelo pavillion to...	computadora hp modelo pavillion touch 23 serie...	comput hp model pavillion touch 23 seri 07 ft ...	otros	2
PAR ARRACADAS Y CRUCIFIO ORO MEDIANO GRAMOS 3...	par arracadas crucifijo oro mediano gramos 3.6	par arracadas crucifijo oro mediano gramos 3.6	par arrac crucifj oro median gram 3.6	oro	3
COLLAR DE ESLAB. PLANCHADO FACET PULSERA DE CA...	collar de eslab planchado facet pulsera de ca...	collar eslabon planchado facet pulsera caucho ...	coll eslab planch facet pulsler cauch aplic tub...	oro	1
LONGINES GRANDE CLASSIQUE QUARTZ DRESS WATCH R...	longines grande classique quartz dress watch r...	longines grande classique quartz dress watch r...	longin grand classiqu quartz dress watch refer...	relojes	1
CRUZ ORO ALTO HECHURA TRIPLE E TNOS ABOLLADO G...	cruz oro alto hechura triple tnos abollado gms...	cruzifijo oro alto hechura triple tnos abollad...	cruz oro alto hechur tripl tnos aboll gms 1.6	oro	1

# Conclusiones

- ◆ Este proyecto necesito de un conocimiento de negocio para concretarse debido al análisis y contraste cualitativo de los resultados para comprobar el correcto funcionamiento
- ◆ Se logro obtener grupos accionables que permitirán una focalización sobre artículos de mayor valor para el negocio
- ◆ Optimización en la toma de decisiones en la cadena de valor comercial
- ◆ La clasificación a primer nivel y segmentación a un nivel de mayor detalle coinciden con los grupos manejados en el negocio en la clasificación realizada en la recepción, lo cual da certeza del resultado esperado
- ◆ Se comprobó en el caso de Relojes su doble clasificación debido a su composición



4



18



# ¡MUCHAS GRACIAS!

