



Arquitectura de Computadoras (UADE)

Trabajo de Investigación: Parallel Computing

FABIÁN ALEJANDRO GÓMEZ

JUNIO-2019



Objetivos

- ▶ Realizar una actividad práctica de paralelismo utilizando **Multiprocessing en lenguaje Python**.
- ▶ Aplicar la técnica de computación paralela en una actividad concreta de procesamiento de **lenguaje natural proveniente de opiniones reales de clientes de compañías de servicio móvil**.
- ▶ Realizar un **análisis comparativo** entre la técnica de multiprocesamiento y el procesamiento simple y diferentes técnicas de programación, para observar en que situaciones es más conveniente su utilización y comprobarlas.

Metodología

Se utilizan dos tamaños de Base de datos para realizar comparaciones. Sobre las mismas se aplican procesamiento de lenguaje natural específicamente en la variable string, mediante multiprocessing (4 procesos) y simpleprocessing. Se compara el tiempo en segundos que demora cada procesamiento.

Base de Datos: Opiniones sobre Servicio Tel. Móvil

- Gran volumen: >1.000.000 registros
- Pequeño volumen: 6.000 registros



Procesamiento

VARIABLES

- Id
- Recomendación (NPS)
 - Causa 1
 - Causa 2
 - Causa 3
- **Motivo Espontaneo (verbatim, texto largo)**
 - Empresa
 - Región
 - Producto

Respuestas Estructuradas
(Escala y Opciones)

**Objetivo: Análisis de
Lenguaje Natural**

Variables independientes
(aperturas de análisis)

NLTK:
Tokenizer

PANDAS:
Sentiment /Motives
Analysis

Librería de Python que divide los strings e identifica sus unidades (palabras).

Permite aplicar luego análisis de distribución,taggeo gramatical, digrams, trigrams etc

Librería de Python para manejar estructuras de datos y tablas.

Permite aplicar luego selección de palabras y expresiones y realizar tags específicos

NLTK Tokenizer

Texto en Lenguaje Natural

id	empresa	region	prod	nps	npstipo	causa1	causa2	causa3	verbatim
2000014150	Movistar	AMBA	Mixto	8	Neutros	Los precios y variedad de planespacks	Precio del plan	Precio del planBeneficios en las recargas	TODO PERO LA COMPAÑIA EN LA PLATA QUE ME DE UN MEJOR PORQUE ES LO UNICO QUE HAY NO TENGO COBERTURA LA SEÑAL DEL INTERNET LA SEÑAL 4G NO EXISTELAS ANTENAS NO HAY Y NO HAY SEÑAL EN
2000008075	Movistar	AMBA	Postpago	7	Neutros	La (variedad disponibilidad precio) de l	Variedad de equipos (marcas y mo	Oferta de accesorios	BAJAR LAS TARIFAS EN CUANTO AL COSTO DEL ABONO MENSUAL
2000004673	Movistar	AMBA	Mixto	9	Promotores	Los precios y variedad de planespacks	Precio del plan	Precio del planBeneficios en las recargas	PORQUE NO HE TENIDO PROBLEMA CON EL SERVICIO DE ESTA LINEA EL INTERNET ES BUENO ES DECIR ANTES TENIA SOMBAHORA ME DAN 3GB CUANDO CAMBIE DE PLAN Y PAGO 500 PESO
2000177223	Claro	SUR	Mixto	8	Neutros	La calidad y cobertura de la señal	Cobertura de la señal (para Solo vo	Cobertura de la señal	LOS USUARIOS SUFRIMOIS MUCHO PORQUE SE CAE EL SISTEMA MUY SEGUIDO Y SE CAEN LAS LLAMADAS
2000177207	Claro	SUR	Prepago	5	Detractores	La calidad y cobertura de la señal	Calidad de llamadas	Calidad de llamadas	EL INTERNET ES UNA PORQUERIA SE CORTA NO ANDA BIEN CUANDO NO TENGO WIFI
2000152979	Movistar	LITORAL	Mixto	9	Promotores	La calidad y cobertura de la señal	Calidad de llamadas	Calidad de llamadas	PORQUE ES LA UNICA EMPRESA QUE HE UTIIZADO Y NO HE TENIDO NINGUN INCOVENIENTE CUANDO HA TENIDO ALGUN INCOVENIENTE SIEMPRE ME HAN INFORMADO DE MANERA CORRECTA COMO SOLUCIONARLO
2000006149	Movistar	AMBA	Mixto	5	Detractores	La modalidad de pago de la factura y cc	Facilidad de compra (pago de facti	Email	QUE TRATE DE HACER UNA COMPRA TELEFONICAMENTE Y FUE UN DESASTRE PORQUE DESPUES DE UN MES Y DESPUES HABERME HECHO EL DEBITO FUE QUE ME INFORMARON QUE HICIERON LA REVERSION DEL PAGO
2000001578	Movistar	AMBA	Prepago	7	Neutros	Los precios y variedad de planespacks	InternetDatos a través de celular	Plan todo en uno (combina llamados men	EL INTERNET SE CUELGA Y NO ME PERMITE CONVERSAR POR WHATAPP
2000009568	Movistar	AMBA	Mixto	5	Detractores	La modalidad de pago de la factura y cc	Claridad y simplicidad de la facturaz	Facilidad para encontrar puntos de venta	PORQUE HAY BASTANTES FALLAS COMO MEGABAYCUANDO VOY Y RECLAMO Y NO RESPONDEN Y ME DICEN QUE NO HAY A BONOS
2000011999	Movistar	AMBA	Mixto	8	Neutros	La calidad y cobertura de la señal	Calidad de llamadas	Calidad de llamadas	TIENE QUE MEJORAR LA ATENCION AL CLIENTE ALGUNAS VECES NO SE PUEDE LLEGAR A CONTACTAR CON EL ASESOR
2000007759	Movistar	AMBA	Mixto	10	Promotores	La modalidad de pago de la factura y cc	Facilidad para encontrar puntos de	Descarga de juegos y aplicaciones	PORQUE NUNCA ME AFALLADO ME FUI AFUERA TAMBIEN ME ANDUBO BIEN NO TUBE PROBLEMAS EN EL SENTIDO DE QUE TUBE UN COMPAÑIA Y NO PODIA HABLAR A LARGA DISTANCIA Y CON MOVISTAR NO TENGO
2000177921	Claro	SUR	Postpago	8	Neutros	La atención al cliente	Otros	Otros	en el tema aveces de la atencion me compre un telefono me vendieron usados y fallado y reclamado varias veces
2000002001	Movistar	AMBA	Mixto	10	Promotores	La calidad y cobertura de la señal	Calidad de envío de mensajes (SM	Calidad de envío de mensaje de texto (Sv	BUEN FUNCIONAMIENTO EFICAZ EN TODO
2000003237	Movistar	AMBA	Prepago	9	Promotores	La calidad y cobertura de la señal	Calidad de llamadas	Calidad de llamadas	PORQUE TIENE SEÑAL EN TODOS LADOS PARA RECIBIR Y REALIZAR LLAMADAS
2000155310	Claro	LITORAL	Prepago	6	Detractores	La calidad y cobertura de la señal	Calidad de llamadas	Calidad de llamadas	PORQUE CUANDO ESTOY EN MI CASA NO TENGO SEÑAL G DENTRO NO SIRVE EL INTERNET DENTRO DE LA CASAFUERA DE LA CASA SI HAY SEÑAL
2000004784	Movistar	AMBA	Postpago	5	Detractores	La calidad y cobertura de la señal	Otros	Otros	POR QUE DESPUES DE 10 AÑOS PASE DE TODO CON MOVISTAR LOS SERVICIOS NUNCA FUERON BUENOS MUCHAS VECES SE CORTAN LAS COMUNICACIONES CUANDO ESTOY HABLANDO SE CORTA DE LA NADA TIENE PR
2000015148	Claro	AMBA	Mixto	10	Promotores	Los precios y variedad de planespacks	Precio del plan	Precio del planBeneficios en las recargas	ES ACCESIBLE Y TENGO COBERTURA EN TODOS LADOS
2000155244	Personal	LITORAL	Mixto	8	Neutros	Los precios y variedad de planespacks	InternetDatos a través de celular	Plan todo en uno (combina llamados men	DEBERIA MEJORAR EL INTERNET YA QUE SE CORTA Y NO PUEDO NAVEGAR CUANDO INGRESO A YOUTUBE NO ME CARGAN LOS VIDEOS Y ME INDICA QUE YA A CULMINADO EL CREDITO
2000155273	Claro	LITORAL	Prepago	9	Promotores	La calidad y cobertura de la señal	Velocidad de conexión a internet d	Velocidad de conexión a internet desde el	TIENE BUEN SERVICIO A VECES ME DEMORO EN PAGAR Y ME PERMITEN 2 O 3 PARA PAGAR EL SERVICIO ADEMAS NO HE TENIDO PROBLEMA
2000014685	Claro	AMBA	Mixto	7	Neutros	La calidad y cobertura de la señal	Velocidad de conexión a internet d	Velocidad de conexión a internet desde el	LA ATENCION AL CLIENTE TENGO QUE LLAMAR MAS DE 5 VECES PARA QUE ME SOLUCIONAR LOS PLANES A LA HORA DEL AUMENTO ME AUMENTARON SIN AVISARME Y PAGUE MAS
2000177972	Claro	SUR	Mixto	9	Promotores	La atención al cliente	Facilidad de compra (nuevas líneas	Facilidad de compra (nuevas líneas equip	PORQUE NUNCA HE TENIDO PROBLEMAS CON EL SERVICIO EN MI LINEAMI INTERNET ES MUY RAPIDO Y NO SE PARAN LOS VIDEOS
2000016927	Personal	AMBA	Mixto	5	Detractores	Los precios y variedad de planespacks	Beneficios en las recargas (Multipl	InternetDatos a través de celular	PORQUE ES CARO HAY OTRAS EMPRESAS QUE SON BARATAS CON RELACION A LOS MINUTOS PERSONAL ES POCAS LO QUE ME DAN
2000165641	Movistar	SUR	Mixto	10	Promotores	La calidad y cobertura de la señal	Velocidad de conexión a internet d	Velocidad de conexión a internet desde el	PORQUE ACA EN LA ZONA ANDA PERFECTO Y BIEN POR EJEMPLO ACA IGUAL TENEMOS CLARO Y PERSONA PERO PERSONAL ANDA MUY MAL Y MOVISTAR RESPONDE BIEN NUNCA E TENIDO PROBLEMAS EL INTERNET ES

NLTK Tokenizer.

El procesamiento para identificar los Tockens de las frases de NL, conforma la etapa inicial de un proyecto de Text Analytics. En esta etapa se determinan las unidades léxicas básicas, que luego en otras etapas podría realizarse el Filtrado de Palabras Neutras o stopwords (artículos, preposiciones, etc), y el análisis de combinaciones léxicas (Digrams, Trigrams, etc)

```
Console 1/A
[llamada , perdida ]
['porque', 'tiene', 'buena', 'señal', 'donde', 'voy', 'me', 'da', 'los', 'packs', 'que', 'me', 'dicen', 'que', 'me', 'dan']
['porque', 'el', 'alcance', 'de', 'la', 'línea', 'en', 'distintos', 'lugares', 'de', 'la', 'provincia', 'mas', 'precisamente', 'en', 'los', 'barrios', '4g', 'no', 'levanta', 'la', 'línea']
['el', 'abono', 'yo', 'tengo', 'abono', 'fijo', 'le', 'dan', 'mensaje', 'en', 'el', 'plan', 'y', 'yo', 'no', 'lo', 'uso']
['porque', 'tiene', 'muy', 'buena', 'señal', 'y', 'tiene', 'muy', 'buenas', 'promociones']
['el', 'costo', 'por', 'ejemplo', 'del', 'importe', 'de', 'la', 'factura']
['recomiendo', 'a', 'claro', 'porque', 'nunca', 'tuve', 'problemas', 'me', 'dan', 'buena', 'atencion', 'los', 'celulares', 'son', 'buenos', 'no', 'tienes']
['porque', 'nunca', 'he', 'tenido', 'problemas', 'con', 'las', 'boletas', 'siempre', 'andubo', 'bien', 'y', 'no', 'le', 'cobraron', 'nunca', 'demas']
['tiene', 'que', 'mejorar', 'el', 'servicio', 'de', 'internet', 'en', 'la', 'sierra', 'porque', 'falla', 'bastante', 'los', 'datos', 'moviles', 'no', 'la', 'señal', 'que', 'tiene', 'en', 'el', 'campo', 'en', 'cualquier', 'lado', 'y', 'tengo', 'un', 'buen', 'servicio']
['porque', 'es', 'mas', 'rapido', 'para', 'navegar', 'por', 'ejemplo', 'porque', 'antes', 'tenia', 'personal', 'y', 'el', 'internet', 'era', 'lento', 'entra', 'rapido', 'es', 'mas', 'facil', 'descargar', 'las', 'aplicaciones', 'y', 'tienes', 'mas', 'beneficios', 'tambien', 'es', 'mas', 'economico', 'internet', 'es', 'mejor', 'la', 'comunicacion', 'y', 'tiene', 'mejor', 'señal']
['porque', 'nunca', 'he', 'tenido', 'problemas', 'y', 'si', 'no', 'me', 'llagaba', 'la', 'boleta', 'llamaba', 'y', 'lo', 'solucionaban', 'y', 'llamaba', 'creo', 'que', 'las', 'otras', 'empresas', 'tienen', '4g', 'y', 'claro', 'no', 'me', 'brinda', 'esa', 'señal', 'y', 'a', '4', 'km', 'no', 'haya', 'tan', 'de', 'compañia', 'quiera', 'que', 'la', 'empresa', 'mejore', 'el', 'servicio']
```

PANDAS: Sentiment Analysis

Sentimientos

Opiniones Espontaneas

negativo	positivo	verbatim
1 True	False	la compañía en la plata que me de un mejor porque es lo unico que hay no tengo señal se corta la señal del internet la señal 4g
2 False	False	bajar las tarifas en cuanto al costo del abono mensual
3 False	True	porque no he tenido problema con el servicio de esta linea el internet es bueno es decir antes tenia 50mbahora me dan 3gb cu
4 True	False	los usuarios sufrimois mucho porque se cae el sistema muy seguido y se caen las llamadas
5 True	False	el internet es una porqueria se corta no anda bien cuando no tengo wifi
6 False	True	porque es la unica empresa que he utiizado y no he tenido ningun incoveniente cuando ha tenido algun incoveniente siempre r
7 True	False	que trate de hacer una compra telefonicamente y fue un desastre porque despues de un mes y despues haberme hecho el deb
8 True	False	el internet se cuelga y no me permite conversar por whatsapp
9 True	False	porque hay bastantes fallas como megabaycuando voy y reclamo y no responder y me dicen que no hay a bonos
10 True	False	tiene que mejorar la atencion al cliente algunas veces no se puede llegar a contactar con el asesor
11 False	True	porque nunca me afallado me fui afuera tambien me andubo bien no tube problemas en el sentido de que tube un compañía y
12 True	False	en el tema aveces de la atencion me compre un telefono me vendieron usados y fallado y reclamado varias veces
13 False	True	buen funcionamiento eficaz en todo
14 False	True	porque tiene señal en todos lados para recibir y realizar llamadas
15 True	False	porque cuando estoy en mi casa no tengo señal g dentro no sirve el internet dentro de la casafuera de la casa si hay señal
16 True	True	por que despues de 10 años pase de todo con movistar los servicios nunca fueron buenos muchas veces se cortan las comunica
17 True	False	es accesible y tengo cobertura en todos lados
18 True	False	deberia mejorar el internet ya que se corta y no puedo navegar cuando ingreso a youtube no me cargan los videos y me indica c
19 True	False	tiene buen servicio a veces me demoro en pagar y me permiten 2 o 3 para pagar el servicio ademas no he tenido problema
20 True	False	la atencion al cliente tengo que llamar mas de 5 veces para que me solucionar los planes a la hora del aumento me aumentaron
21 False	True	porque nunca he tenido problemas con el servicio en mi lineami internet es muy rapido y no se paran los videos
22 True	False	porque es caro hay otras empresas que son baratas con relacion a los minutos personal es pocas lo que me dan
23 True	True	porque aca en la zona anda perfecto y bien por ejemplo aca igual tenemos claro y persona pero personal anda muy mal y movis
24 False	True	porque tiene señal en todos lados y esta bueno

Sentiment Analysis

Se procesa mediante la búsqueda de palabras clave y combinaciones de palabras, para etiquetar a cada frase.

Si la Frase contiene elementos considerados dentro de la lista de **POSITIVOS**, se generará un nuevo campo con un valor True.

Se realiza lo mismo para Sentimientos **NEGATIVOS**

PANDAS: Motive Analysis

Motivos

Opiniones Espontaneas

servicio	conveniencia	relacionamiento	verbatim
1 True	False	False	todo pero la compavtia en la plata que me de un mejor porque es lo unico que hay no tengo sevta se corta la sevta del inte
2 False	True	False	bajar las tarifas en cuanto al costo del abono mensual
3 True	True	False	porque no he tenido problema con el servicio de esta linea el internet es bueno es decir antes tenia 50mbahora me dan 3gb d
4 True	False	False	los usuarios sufrimois mucho porque se cae el sistema muy seguido y se caen las llamadas
5 True	False	False	el internet es una porqueria se corta no anda bien cuando no tengo wifi
6 False	False	True	porque es la unica empresa que he utiizado y no he tenido ningun inconveniente cuando ha tenido algun inconveniente siempre
7 False	True	True	que trate de hacer una compra telefonicamente y fue un desastre porque despues de un mes y despues haberme hecho el del
8 True	False	False	el internet se cuelga y no me permite conversar por whatapp
9 False	False	True	porque hay bastantes fallas como megabaycuando voy y reclamo y no responden y me dicen que no hay a bonos
10 False	False	True	tiene que mejorar la atencion al cliente algunas veces no se puede llegar a contactar con el asesor
11 True	False	False	porque nunca me afallado me fui afuera tambien me andubo bien no tube problemas en el sentido de que tube un compavtia
12 False	False	True	en el tema aveces de la atencion me compre un telefono me vendieron usados y fallado y reclamado varias veces
13 True	False	False	buen funcionamiento eficaz en todo
14 True	False	False	porque tiene sevta en todos lados para recibir y realiza llamadas
15 True	False	False	porque cuando estoy en mi casa no tengo sevta g dentro no sirve el internet dentro de la casa fuera de la casa si hay sevta
16 True	False	False	por que despues de 10 avtos pase de todo con movistar los servicios nunca fueron buenos muchas veces se cortan las comun
17 True	True	False	es accesible y tengo cobertura en todos lados
18 True	True	False	deberia mejorar el internet ya que se corta y no puedo navegar cuando ingreso a youtube no me cargan los videos y me indica
19 False	True	False	tiene buen servicio a veces me demoro en pagar y me permiten 2 o 3 para pagar el servicio ademas no he tenido problema
20 False	True	True	la atencion al cliente tengo que llamar mas de 5 veces para que me solucionen los planes a la hora del aumento me aumentar
21 True	False	False	porque nunca he tenido problemas con el servicio en mi linea el internet es muy rapido y no se paran los videos
22 False	True	False	porque es caro hay otras empresas que son baratas con relacion a los minutos personal es pocas lo que me dan
23 True	False	False	porque aca en la zona anda perfecto y bien por ejemplo aca igual tenemos claro y persona pero personal anda muy mal y mo
24 True	False	False	porque tiene sevta en todos lados y esta bueno
25 True	False	False	la del internet no me permite enviar videos ni mensajes y de las llamada porque escucho interferencia
26 True	False	False	el servicio la cobertura en algunas zonas y la sevta 4g de la capital
27 True	True	False	mejorar los precios de los abonos me dura poco las pocas que hago una recarga me llaman mucho del centro de movistar par
28 True	False	False	consume mucho el internet en gogle cuando voy a plicacion
29 True	False	False	la sevta porque se va

Motive Analysis

Se procesa mediante la búsqueda de palabras clave y combinaciones de palabras para etiquetar cada frase. Si la Frase contiene elementos incluidos dentro de la lista de temáticas de, por ejemplo, **SERVICIO**, se generará un nuevo campo con un valor True como flag.

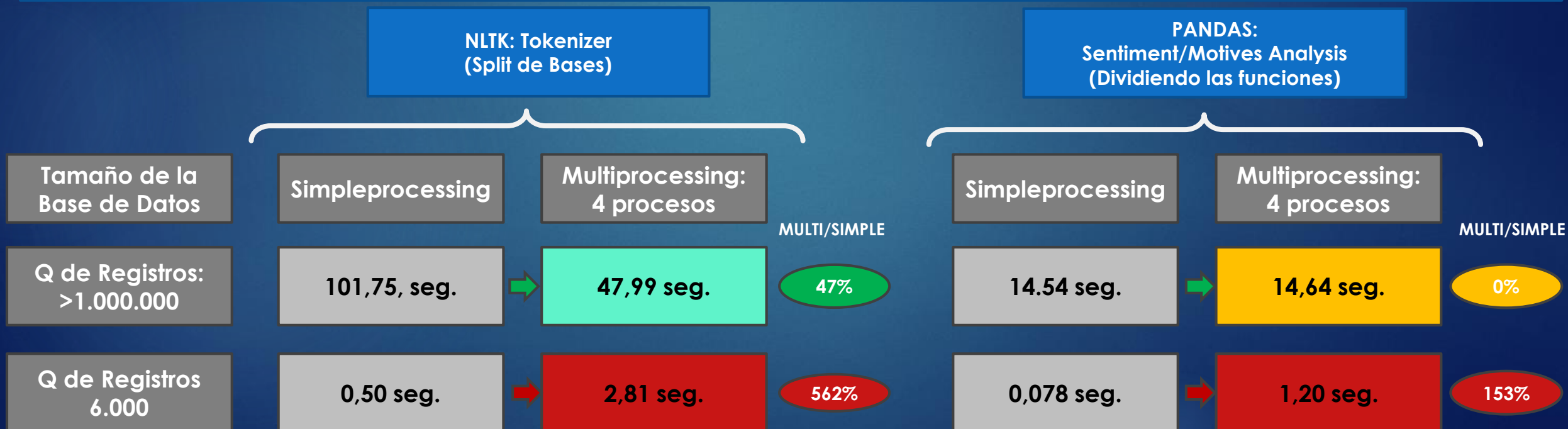
Se realiza el mismo procedimiento para motivos de Económicos (**CONVENIENCIA**) o de Atención al Cliente (**RELACIONAMIENTO**).

Análisis y Conclusiones



INTEL CORE i7 – 8650U CPU 1,90 Ghz

- Se observó que el rendimiento de Multiprocessing, es más notorio en el procesamiento el archivo de gran volumen, donde tuvo mejor performance que el procesamiento simple
- El decrecimiento de la demora en el procesamiento por multiprocessing varia en función del codigo utilizado(Comparativamente tuvo reducción en Tokenizer realizando Split de base de datos, mientras que no lo hizo en la clasificación realizada mediante Pandas, dividiendo las funciones)
- Al procesar archivos pequeños la diferencia entre Simple y Multiprocessing no es significativa, e inclusive aparentemente podria ser contraproducente

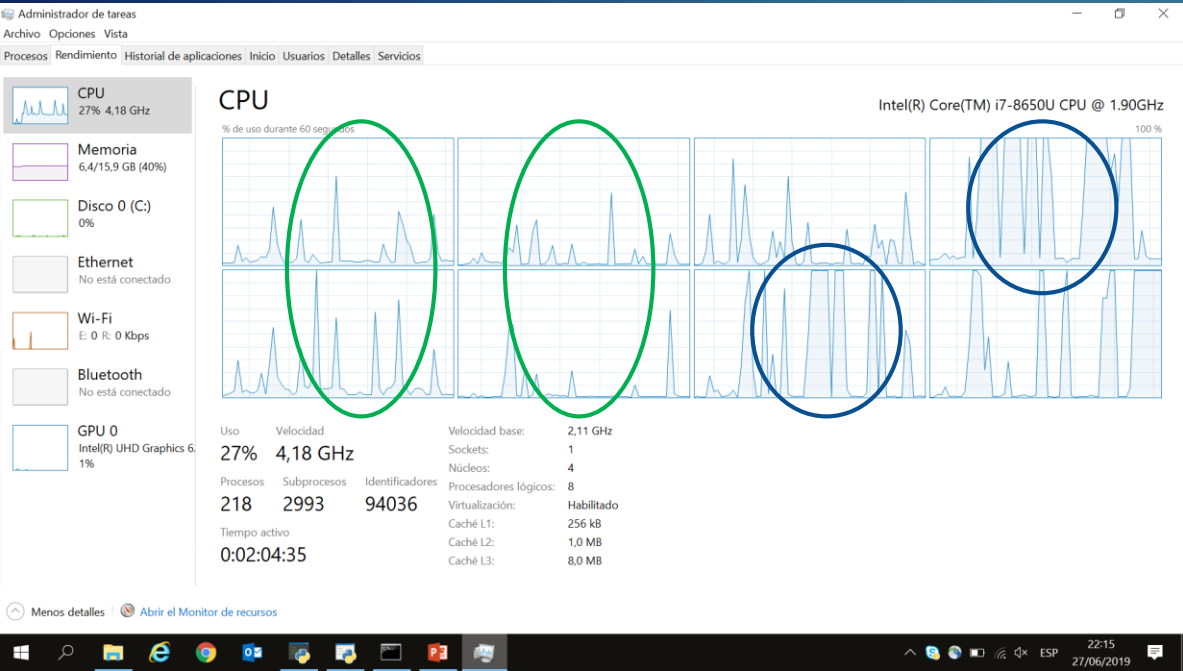


NLTK Tokenizer

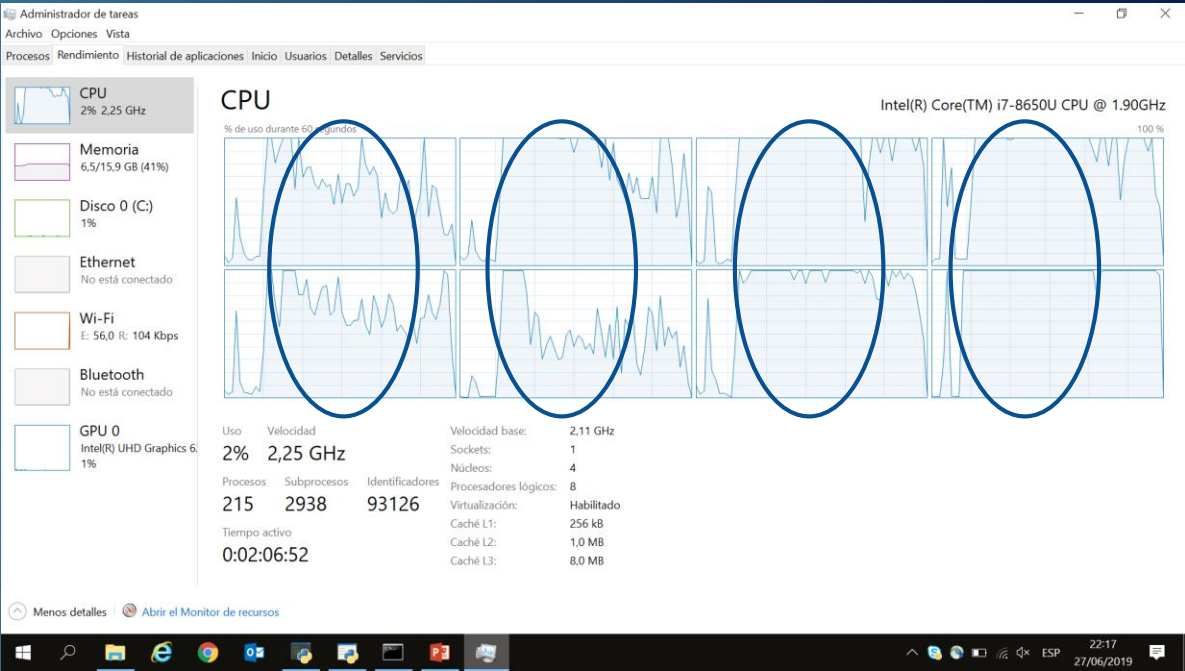


INTEL CORE i7 – 8650U CPU 1,90 Ghz

Simpleprocessing



Multiprocessing:
4 procesos

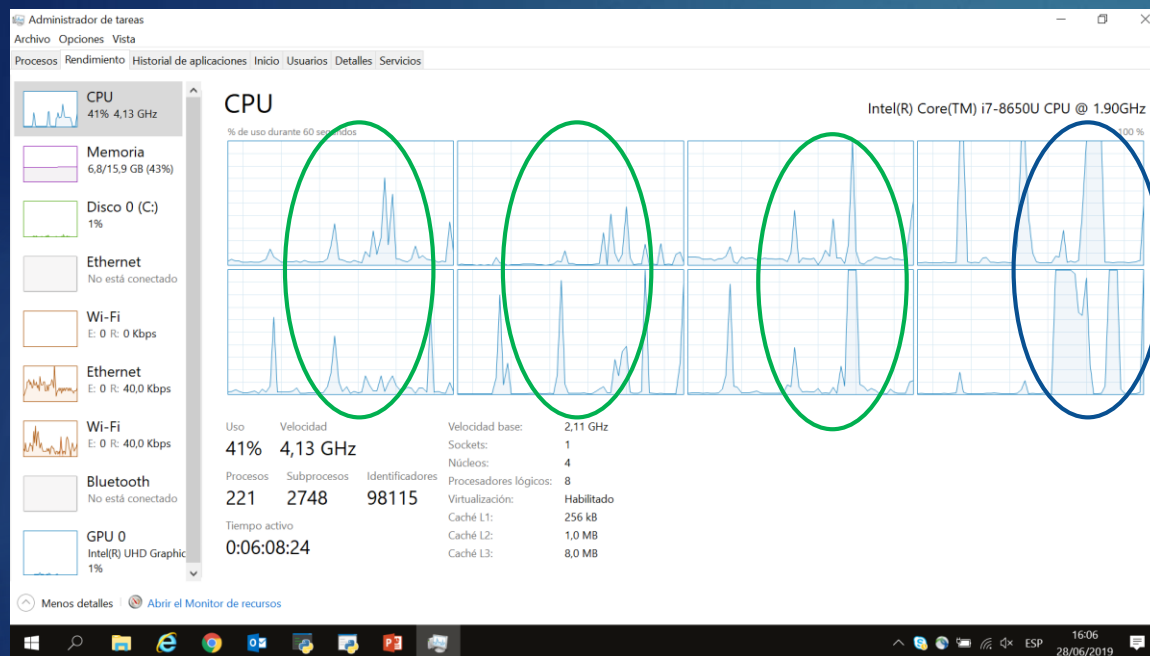


PANDAS: Sentiment Analysis

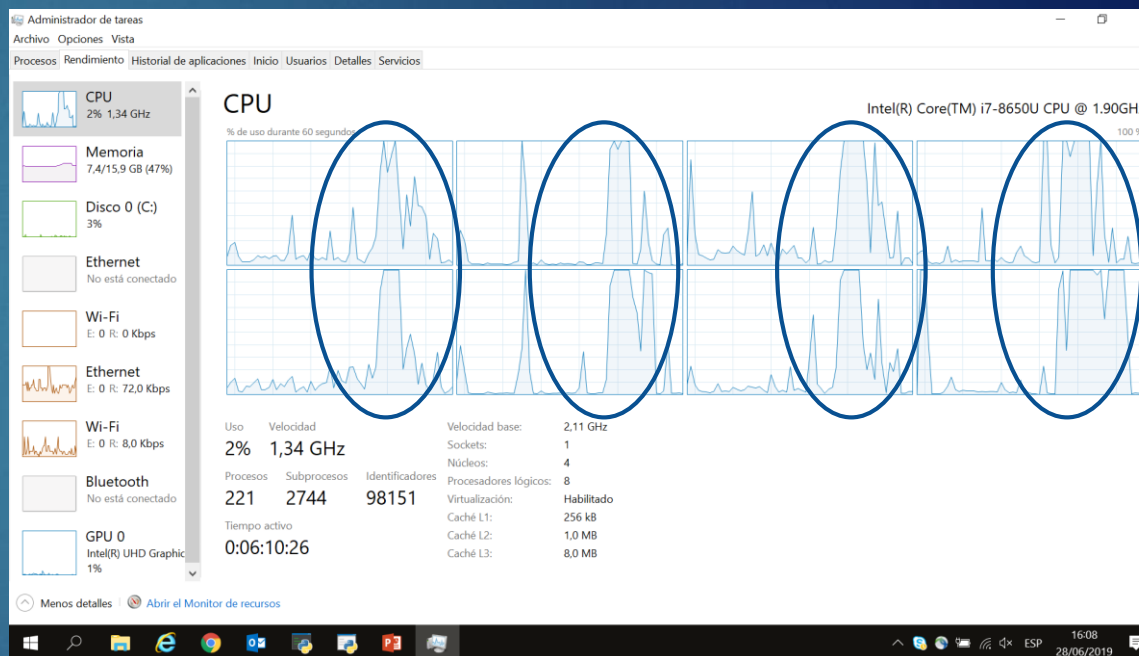


INTEL CORE i7 – 8650U CPU 1,90 Ghz

Simpleprocessing



Multiprocessing:
4 procesos





Código

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\NLTK-Tokenizer-Multipleprocessing.py

1from multiprocessing import Pool

2import time

3import nltk

4import pandas as pd

5import numpy as np

6

7# Set some pandas options controlling output format

8pd.set_option('display.notebook_repr_html', False)

9pd.set_option('display.max_columns', 99)

10pd.set_option('display.max_rows', 99999)

11pd.set_option('display.width', 9999)

12

13

14def naturalang1():

15 for sent in (verbatim1):

16 nltk.word_tokenize(sent, language='spanish')

17

18def naturalang2():

19 for sent in (verbatim2):

20 nltk.word_tokenize(sent, language='spanish')

21

22def naturalang3():

23 for sent in (verbatim3):

24 nltk.word_tokenize(sent, language='spanish')

25

26def naturalang4():

27 for sent in (verbatim4):

28 nltk.word_tokenize(sent, language='spanish')

29

Explorador de archivos

Nombre

> Borrador

2018NPS-Stor.csv

2018NPS.csv

2019-06-Gomez Fabian Alejandro - Arquietectura de Computadoras - Multip

NLTK-Tokenizer-Multipleprocessing-Micro.py

NLTK-Tokenizer-Multipleprocessing.py

NLTK-Tokenizer-Simpleprocessing-Micro.py

NLTK-Tokenizer-Simpleprocessing.py

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

In [3]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Multipleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

PARTICIONES EN BLOQUES: 4

Tiempo en segundos (4 Processes) - 48.013792753219604

In [4]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Simpleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

Index(['empresa', 'nps', 'npstipo', 'causa1', 'causa2', 'causa3', 'region', 'prod', 'verbatim'], dtype='object')

Tiempo en segundos (W/Processes) - 98.01481604576111

In [5]:

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: ASCIILínea: 11Columna: 37Memoria: 37 %

20:41

28/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\NLTK-Tokenizer-Multipleprocessing.py

30

31 dfm = pd.read_csv("2018NPS-Stor.csv", index_col='id', encoding='utf-8', sep=';')

32 verbatim = dfm['verbatim']

33

34

35 ndivisiones = 4

36 print("PARTICIONES EN BLOQUES: ", ndivisiones)

37 verbatim1 = (np.array_split(dfm['verbatim'],ndivisiones)[0])

38 verbatim2 = (np.array_split(dfm['verbatim'],ndivisiones)[1])

39 verbatim3 = (np.array_split(dfm['verbatim'],ndivisiones)[2])

40 verbatim4 = (np.array_split(dfm['verbatim'],ndivisiones)[3])

41

42

43 if __name__ == '__main__':

44 processes = 4

45 pool = Pool(processes)

46

47 start2 = time.time()

48

49 r1 = pool.apply_async(naturalang1)

50

51 r2 = pool.apply_async(naturalang2)

52

53 r3 = pool.apply_async(naturalang3)

54

55 r4 = pool.apply_async(naturalang4)

56

57

58 pool.close()

59 pool.join()

60 end2 = time.time()

61

62 print('Tiempo en segundos (4 Processes) -', end2 - start2)

63

64 #

Explorador de archivos

Nombre

> Borrador

2018NPS-Stor.csv

2018NPS.csv

2019-06-Gomez Fabian Alejandro - Arquietectura de Computadoras - Multip

NLTK-Tokenizer-Multipleprocessing-Micro.py

NLTK-Tokenizer-Multipleprocessing.py

NLTK-Tokenizer-Simpleprocessing-Micro.py

NLTK-Tokenizer-Simpleprocessing.py

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

In [3]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Multipleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

PARTICIONES EN BLOQUES: 4

Tiempo en segundos (4 Processes) - 48.013792753219604

In [4]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Simpleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

Index(['empresa', 'nps', 'npstipo', 'causa1', 'causa2', 'causa3', 'region', 'prod', 'verbatim'], dtype='object')

Tiempo en segundos (W/Processes) - 98.01481604576111

In [5]:

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: ASCII

Línea: 11Columna: 37Memoria: 37 %

20:4128/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\NLTK-Tokenizer-Simpleprocessing.py

Explorador de archivos

NLTK-Tokenizer-Multipleprocessing.pyNLTK-Tokenizer-Simpleprocessing.py*Sentiment-Panda-Multipleprocessing2.pySentiment-P

```
1 import time
2 import nltk
3 import pandas as pd
4
5
6 # Set some pandas options controlling output format
7 pd.set_option('display.notebook_repr_html', False)
8 pd.set_option('display.max_columns', 99)
9 pd.set_option('display.max_rows', 99999)
10 pd.set_option('display.width', 9999)
11
12
13 def naturalang():
14     for sent in (verbatim):
15         nltk.word_tokenize(sent, language='spanish')
16
17
18 dfm = pd.read_csv("2018NPS-Stor.csv", index_col='id', encoding='utf-8', sep=';')
19 verbatim = dfm['verbatim']
20
21 dfm['verbatim'] = dfm['verbatim'].str.lower()
22
23 print(dfm.columns)
24
25
26 start = time.time()
27 naturalang()
28 end = time.time()
29 print('Tiempo en segundos (W/Processes) -', end - start)
30
31
```

Nombre

> Borrador

2018NPS-Stor.csv

2018NPS.csv

2019-06-Gomez Fabian Alejandro - Arquietectura de Computadoras - Multip

NLTK-Tokenizer-Multipleprocessing-Micro.py

NLTK-Tokenizer-Multipleprocessing.py

NLTK-Tokenizer-Simpleprocessing-Micro.py

NLTK-Tokenizer-Simpleprocessing.py

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

In [3]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Multipleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')
PARTICIONES EN BLOQUES: 4
Tiempo en segundos (4 Processes) - 48.013792753219604

In [4]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/NLTK-Tokenizer-Simpleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')
Index(['empresa', 'nps', 'npstipo', 'causa1', 'causa2', 'causa3', 'region', 'prod', 'verbatim'], dtype='object')
Tiempo en segundos (W/Processes) - 98.01481604576111

In [5]:

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: ASCII

Línea: 24Columna: 1Memoria: 37 %

20:42

28/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\Sentiment-Panda-Multiprocessing2.py

Explorador de archivos

Terminal de IPython

1from multiprocessing import Pool

2import time

3import pandas as pd

4import numpy as np

5

6

7# Set some pandas options controlling output format

8pd.set_option('display.notebook_repr_html', False)

9pd.set_option('display.max_columns', 99)

10pd.set_option('display.max_rows', 99999)

11pd.set_option('display.width', 9999)

12

13

14def sentiment(x):

15 dfm['positivo'] = dfm['verbatim'].str.contains(positivo)

16 dfm['negativo'] = dfm['verbatim'].str.contains(negativo)

17 dfm1 = dfm[['negativo','positivo','verbatim']]

18

19

20def sentiment_analysis(x):

21 positivo_sum = dfm.groupby(['empresa', 'prod'])['positivo'].sum().rename("count")

22 positivo_porc = positivo_sum / positivo_sum.groupby(level=0).sum()

23 print("_____")

24 print("Cantidad de Respuestas Positivas")

25 print("_____")

26 print(positivo_sum)

27 print("_____")

28 print("Porcentaje de Respuestas Positivas")

29 print("_____")

30 print(positivo_porc)

31

Nombre

> Borrador

> 2018NPS-Stor.csv

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

In [6]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/Sentiment-Panda-Multiprocessing2.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

Index(['empresa', 'nps', 'npstipo', 'causa1', 'causa2', 'causa3', 'region', 'prod', 'verbatim'], dtype='object')

Cantidad de Respuestas Positivas

empresaprod

ClaroMixto37428.0

Postpago10882.0

Prepago36229.0

MovistarMixto28203.0

Postpago9236.0

Prepago22982.0

PersonalMixto21561.0

Postpago17056.0

Prepago31256.0

Name: count, dtype: float64

Porcentaje de Respuestas Positivas

empresaprod

ClaroMixto0.442731

Permisos: RWFin de línea: LFCodificación: UTF-8-GUESSEDLínea: 6Columna: 1Memoria: 38 %

20:4428/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\Sentiment-Panda-Multiprocessing2.py

Explorador de archivos

def motive(x):
dfm['servicio'] = dfm['verbatim'].str.contains(servicio)
dfm['conveniencia'] = dfm['verbatim'].str.contains(conveniencia)
dfm['relacionamiento'] = dfm['verbatim'].str.contains(relacionamiento)
dfm2 = dfm[['servicio', 'conveniencia', 'relacionamiento', 'verbatim']]

def motive_analysis(x):
servicio_sum = dfm.groupby(['empresa', 'prod'])['servicio'].sum().rename("count")
servicio_porcentaje = servicio_sum / servicio_sum.groupby(level=0).sum()
print("Cantidad de Respuestas sobre Servicio")
print(servicio_sum)
print("Porcentaje de Respuestas sobre Servicio")
print(servicio_porcentaje)

dfm = pd.read_csv("2018NPS-Stor.csv", index_col='id', encoding='utf-8', sep=';')
dfm['verbatim'] = dfm['verbatim'].str.lower()
verbatim = dfm['verbatim']
print(dfm.columns)

positivo = ('bueno|genial|barbaro|buenas|bien|perfecto|satisfecho|recomiendo|buenisimo|lujo|en
negativo = ('malo|pesimo|malisimo|peor|caro|problema|problemas|fallas|corta|mal|mejorar|pierde
servicio = ('cobertura|señal|llamo|llama|llamadas|llamados|mensajes|internet|navegar|navegacion

Nombre

Borrador

2018NPS-Stor.csv

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

Prepago 119154.0
Movistar Mixto 90803.0
Postpago 33646.0
Prepago 74649.0
Personal Mixto 78448.0
Postpago 58303.0
Prepago 110839.0
Name: count, dtype: float64

Porcentaje de Respuestas sobre Servicio

empresa	prod	
Claro	Mixto	0.422525
	Postpago	0.112685
Movistar	Prepago	0.464790
	Mixto	0.456072
Personal	Postpago	0.168992
	Prepago	0.374936
	Mixto	0.316846
	Postpago	0.235482
	Prepago	0.447672

Name: count, dtype: float64
Tiempo en segundos (4 Processes) - 14.936394453048706

In [7]:

Permisos: RWFin de línea: LFCodificación: UTF-8-GUESSEDLínea: 6Columna: 1Memoria: 37 %

20:44

28/06/2019

Python 3.7

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

File Explorer Icons

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\Sentiment-Panda-Multiprocessing2.py

Explorador de archivos

NLTK-Tokenizer-Multipleprocessing.pyNLTK-Tokenizer-Simpleprocessing.pySentiment-Panda-Multiprocessing2.pySentiment-Par

```
56 verbatim = dfm['verbatim']
57
58 print(dfm.columns)
59
60 positivo = ('bueno|genial|barbaro|buenas|bien|perfecto|satisfecho|recomiendo|buenisimo|lujo|encantado|m
61 negativo = ('malo|pesimo|malisimo|peor|caro|problema|problemas|fallas|corta|mal|mejorar|pierde|cortes')
62 servicio = ('cobertura|señal|llamo|llama|llamadas|llamados|mensajes|internet|navegar|navegacion|navego|
63 relacionamiento = ('atencion|atención|oficina|representante|atienden|resuelven|solucion|resolucion|espe
64 conveniencia = ('precio|precios|coto|costoso|caro|barato|carisimo|baratisimo|aumento|aumenta|aumentos|t
65
66
67 if __name__ == '__main__':
68     processes = 4
69     pool = Pool(processes)
70
71     start = time.time()
72
73     r1 = pool.apply_async(sentiment(verbatim))
74
75
76     r2 = pool.apply_async(motive(verbatim))
77
78
79     r3 = pool.apply_async(sentiment_analysis(verbatim))
80
81
82     r4 = pool.apply_async(motive_analysis(verbatim))
83
84
85     pool.close()
86     pool.join()
87
88     end = time.time()
89
90     print('Tiempo en segundos (4 Processes) -'. end - start)
```

Nombre

> Borrador

2018NPS-Stor.csv

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

Prepago 119154.0
Movistar Mixto 90803.0
Postpago 33646.0
Prepago 74649.0
Personal Mixto 78448.0
Postpago 58303.0
Prepago 110839.0
Name: count, dtype: float64
Porcentaje de Respuestas sobre Servicio
empresa prod
Claro Mixto 0.422525
Postpago 0.112685
Prepago 0.464790
Movistar Mixto 0.456072
Postpago 0.168992
Prepago 0.374936
Personal Mixto 0.316846
Postpago 0.235482
Prepago 0.447672
Name: count, dtype: float64
Tiempo en segundos (4 Processes) - 14.936394453048706
In [7]:

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: UTF-8-GUESSEDLínea: 6Columna: 1Memoria: 38 %

Windows Taskbar

20:45 28/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\Sentiment-Panda-Simpleprocessing.py

Explorador de archivos

1import time

2import pandas as pd

3

4# Set some pandas options controlling output format

5pd.set_option('display.notebook_repr_html', False)

6pd.set_option('display.max_columns', 99)

7pd.set_option('display.max_rows', 99999)

8pd.set_option('display.width', 9999)

9

10def sentiment(x):

11dfm['positivo'] = dfm['verbatim'].str.contains(positivo)

12dfm['negativo'] = dfm['verbatim'].str.contains(negativo)

13dfm1 = dfm[['negativo', 'positivo', 'verbatim']]

14

15

16def sentiment_analysis(x):

17positivo_sum = dfm.groupby(['empresa', 'prod'])['positivo'].sum().rename("count")

18positivo_porcentaje = positivo_sum / positivo_sum.groupby(level=0).sum()

19print("_____")

20print("Cantidad de Respuestas Positivas")

21print("_____")

22print(positivo_sum)

23print("_____")

24print("Porcentaje de Respuestas Positivas")

25print("_____")

26print(positivo_porcentaje)

27

28def motive(x):

29dfm['servicio'] = dfm['verbatim'].str.contains(servicio)

30dfm['conveniencia'] = dfm['verbatim'].str.contains(conveniencia)

31dfm['relacionamiento'] = dfm['verbatim'].str.contains(relacionamiento)

32dfm2 = dfm[['servicio', 'conveniencia', 'relacionamiento', 'verbatim']]

33

34

35def motive_analysis(x):

Nombre

> Borrador

> 2018NPS-Stor.csv

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

In [8]: runfile('C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico/Sentiment-Panda-Simpleprocessing.py', wdir='C:/Users/U200820/Documents/Fabian/Arquitectura/Trabajo Practico')

Index(['empresa', 'nps', 'npstipo', 'causa1', 'causa2', 'causa3', 'region', 'prod', 'verbatim'], dtype='object')

Cantidad de Respuestas Positivas

empresa	prod	
Claro	Mixto	37428.0
	Postpago	10882.0
	Prepago	36229.0
Movistar	Mixto	28203.0
	Postpago	9236.0
	Prepago	22982.0
Personal	Mixto	21561.0
	Postpago	17056.0
	Prepago	31256.0

Name: count, dtype: float64

Porcentaje de Respuestas Positivas

empresa	prod	
---------	------	--

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: UTF-8-GUESSEDLínea: 4Columna: 1Memoria: 39 %

20:45

28/06/2019

ArchivoEditarBuscarCódigo fuenteEjecutarDepurarTerminalesProyectosHerramientasVerAyuda

C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico

Editor - C:\Users\U200820\Documents\Fabian\Arquitectura\Trabajo Practico\Sentiment-Panda-Simpleprocessing.py

Explorador de archivos

34

35 def motive_analysis(x):

36 dfm['servicio'] = dfm['verbatim'].str.contains(servicio)

37 servicio_sum = dfm.groupby(['empresa', 'prod'])['servicio'].sum().rename("count")

38 servicio_por = servicio_sum / servicio_sum.groupby(level=0).sum()

39 print("_____")

40 print("Cantidad de Respuestas sobre Servicio")

41 print("_____")

42 print(servicio_sum)

43 print("_____")

44 print("Porcentaje de Respuestas sobre Servicio")

45 print("_____")

46 print(servicio_por)

47

48 dfm = pd.read_csv("2018NPS-Stor.csv", index_col='id', encoding='utf-8', sep=';')

49 dfm['verbatim'] = dfm['verbatim'].str.lower()

50 verbatim = dfm['verbatim']

51 print(dfm.columns)

52

53 positivo = ('bueno|genial|barbaro|buenas|bien|perfecto|satisfecho|recomiendo|buenisimo|lujo|encantado|m

54 negativo = ('malo|pesimo|malisimo|peor|caro|problema|problemas|fallas|corta|mal|mejorar|pierde|cortes')

55 servicio = ('cobertura|señal|llamo|llama|llamadas|llamados|mensajes|internet|navegar|navegacion|navego|

56 relacionamiento = ('atencion|atención|oficina|representante|atienden|resuelven|solucion|resolucion|espe

57 conveniencia = ('precio|precios|coto|costoso|caro|barato|carisimo|baratisimo|aumento|aumenta|aumentos|t

58

59 start = time.time()

60

61 sentiment(verbatim)

62 sentiment_analysis(verbatim)

63 motive(verbatim)

64 motive_analysis(verbatim)

65

66 end = time.time()

67 print('Tiempo en segundos (W/Processes) -', end - start)

Nombre

> Borrador

> 2018NPS-Stor.csv

AyudaExplorador de variablesExplorador de archivos

Terminal de IPython

Terminal 1/A

Claro Mixto 108319.0

Postpago 28888.0

Prepago 119154.0

Movistar Mixto 90803.0

Postpago 33646.0

Prepago 74649.0

Personal Mixto 78448.0

Postpago 58303.0

Prepago 110839.0

Name: count, dtype: float64

Porcentaje de Respuestas sobre Servicio

empresa prod

Claro Mixto 0.422525

Postpago 0.112685

Prepago 0.464790

Movistar Mixto 0.456072

Postpago 0.168992

Prepago 0.374936

Personal Mixto 0.316846

Postpago 0.235482

Prepago 0.447672

Name: count, dtype: float64

Tiempo en segundos (W/Processes) - 14.257910966873169

Terminal de IPythonHistorial de comandos

Permisos: RWFin de línea: LFCodificación: UTF-8-GUESSEDLínea: 4Columna: 1Memoria: 39 %

Windows Explorer Edge Outlook Mail

20:46

28/06/2019



Muchas Gracias!