Dirección General Adjunta de Innovación Financiera e Infraestructuras de Mercado

Departamento de Operaciones

**Analizador de sentimiento de la estabilidad financiera**

1. **Introducción**

La herramienta “Analizador de Sentimiento de la Estabilidad Financiera” está programada en Python y procesa una serie de textos de entrada ordenados en el tiempo, analizando el sentimiento de la estabilidad financiera contenido en ellos mediante unos diccionarios configurables, dando como resultado una serie de gráficas, nubes de palabras y ficheros Excel que ayudan a la interpretación de los resultados. El código de esta herramienta está basado en el empleado para el artículo: [“Análisis de sentimiento del Informe de Estabilidad Financiera” Documento de Trabajo 2011 Banco de España](https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/20/Fich/dt2011.pdf).

1. **Requisitos**

Para la ejecución de la herramienta es necesario tener Python 3.7 instalado con las siguientes librerías adicionales: spacy, pandas, numpy, nltk, plotly, matplotlib, sklearn y wordcloud.

Para tener el entorno más parecido al de desarrollo, se recomienda usar Windows y tener Anaconda instalado en C:\Program Files\Anaconda\. La herramienta no ha sido probada en otros entornos.

1. **Diccionarios**

La herramienta incorpora 4 diccionarios. El usuario experto puede especificar el diccionario a usar por la herramienta como un parámetro en la línea de comandos. Los diccionarios están situados en la carpeta dictionary:

* finalBdE\_dict.xlsx. Diccionario por defecto. Este diccionario contiene palabras con signos diacríticos (tildes) e incluye información de la distribución de opiniones de los diferentes anotadores que participaron en la creación del diccionario, con lo que se incluyen palabras que su categorización final fue neutra, pero fue valorada con sentimiento por algún anotador. Esta distribución es la que se usa para realizar los gráficos boxplot y de bandas.
* BdE\_dict.xlsx. Diccionario original sin signos diacríticos y sin incluir las palabras que finalmente se categorizaron como neutras. Contiene dos palabras menos que el original dado que *contagió* y *contagio* se consideran la misma, así como *deterioro* y *deterioró*. Los gráficos de boxplot y de bandas se realizan eliminando aleatoriamente un 5% de las palabras del diccionario en cada iteración.
* BCCh\_dict.xlsx. Diccionario confeccionado por el Banco Central de Chile. No contiene signos diacríticos y tampoco información de la distribución de anotadores, por lo que los gráficos de boxplot y de bandas se realizan por eliminación de un 5% de las palabras en cada iteración.
* BCCH\_BdE\_dict.xlsx. Diccionario consolidado del Banco de España y Banco Central de Chile. El diccionario no contiene diacríticos ni valoración de anotadores.

1. **Ejecución**

La ejecución se puede realizar desde un IDE (por ejemplo, *Spyder*) o directamente mediante los archivos .bat que se proveen. En el caso de que se desee ejecutar desde un IDE, el archivo principal es isef\_textos.py y se encuentra situado en la carpeta python.

Por comodidad, se proveen 4 archivos .bat configurados con los correspondientes parámetros de línea de comandos para el procesado de los textos según el diccionario especificado: isef\_textos BCCH.bat, isef\_textos BCCH\_BdE.bat, isef\_textos BdE.bat y isef\_textos.bat. Estos archivos asumen que Anaconda está instalado en C:\Program Files\Anaconda\.

1. **Configuración**

Los archivos .bat que se proporcionan podrían requerir configuración en función de la instalación específica del entorno Python.

La herramienta espera que los ficheros de entrada se sitúen en las subcarpetas body, intro y newsp bajo la carpeta input. Estos ficheros tienen que estar en formato texto plano y cada fichero corresponde a una fecha. La fecha se especifica en el nombre del fichero:

* Carpeta body.
  + Patrón a seguir: IEF%Y%m%d\_all.txt.
  + Ejemplo: IEF20200201\_all.txt
* Carpeta intro.
  + Patrón a seguir: IEF%Y%m%d.txt.
  + Ejemplo: IEF20200201.txt
* Carpeta newsp.
  + Patrón a seguir: %Y%m%d.txt.
  + Ejemplo: 20200201.txt

Los eventos que aparecen en los gráficos se configuran en la variable annotations del fichero isef\_charts.py.

Como ejemplo, la línea:

["2012-05-07",' Rescate <br \> Bankia',0],

Indica un evento el 7 de mayo del 2012 con una nota en dos líneas “Rescate” en la superior y “Bankia” en la inferior. El desplazamiento horizontal del texto respecto la vertical es de 0.

1. **Ficheros de salida**

En la carpeta output se sitúan 4 subcarpetas que, si el proceso finaliza correctamente, contendrán el resultado del análisis:

* Carpeta charts. En esta carpeta se generarán tres gráficas por cada una de las carpetas de entrada con textos. Se usa como sufijo el nombre de la carpeta de entrada correspondiente:
  + Gráfica boxplot. Muestra la evolución del ISEF con bandas de percentiles en función del diccionario seleccionado. Por defecto las cajas con los rangos de los valores de la pendiente aparecen ocultas. Haciendo clic en la leyenda se muestran.
  + Gráfica isef. Muestra la evolución del ISEF.
  + Gráfica de negatividades y positividades. Muestra la evolución de las negatividades y positividades.
* Carpeta data. En esta carpeta se generarán cuatro archivos por cada una de las carpetas de entrada con datos para revisar el proceso. En función del sufijo:
  + Sin sufijo: fichero csv con información del número de palabras positivas y negativas, así como el índice resultante para cada fichero.
  + Con sufijo boxplot: fichero csv con información del valor de la pendiente en cada punto.
  + Con sufijo range: fichero csv con información del rango de variación del índice.
  + Con sufijo score: fichero Excel con cada una de las frases identificadas con sentimiento y las palabras que aportaban sentimiento, así como indicador de si estaban negadas o no.
  + Con sufijo topics: fichero csv con los valores tfidf de las palabras que configurarán la nube de palabras de temas.
* Carpeta wc\_tone. Contiene las nubes de palabras de las palabras con connotación de cada uno de los ficheros.
* Carpeta wordcloud. Contiene las nubes de palabras lematizadas con mayor valor tfidf de cada uno de los ficheros.

1. **Contacto**

Para cualquier duda contactar con:

[angel-ivan.moreno@bde.es](mailto:angel-ivan.moreno@bde.es)

[carlos.gpedraz@bde.es](mailto:carlos.gpedraz@bde.es)