

Analisis de sentimiento de textos financieros

Pablo Franco

Universidad Tecnologica Nacional

Introduccion

En el marco de la cursada de la materia Procesamiento del Lenguaje Natural de la Universidad Tecnologica Nacional, se realiza esta actividad para fijar conocimientos mediante la aplicacion practica de tecnicas de analisis de sentimiento de textos y tecnicas de recuperacion de informacion.

Objetivo

Clasificar un texto para identificar si el sentimiento que un lector interpreta es positivo, negativo o neutro dentro de un contexto financiero. Esta informacion tiene como finalidad ayudar a analizar el comportamiento de activos financieros. Se utilizo como muestra activos financieros del mundo de las criptomonedas.

Tecnologias

Se utilizo para la realizacion del proyecto el lenguaje de programacion Python3[1] por ser esta tecnologia la mas extendida en los campos academicos de la ciencia de datos. La implementacion utilizada en esta practica es CPython version 3.10. Esta implementacion se puede utilizar y descargar libremente de internet. Python3 es compatible con gran variedad de arquitecturas y sistemas operativos, en este ensayo se utilizaron arquitecturas amd64/x86 sobre Windows 10 y Debian. Vale mencionar la utilizacion de los modulos de Python fundamentales para la realizacion de esta practica, `happytransformer`[2] (modulo de python para aplicar modelos predictivos), `newspaper`[3] (modulo de python para extraer y normalizar el texto de una pagina web) y `newsapi`[4] (modulo de python

que permite consultar articulos actualizados en cadenas de noticias de todo el mundo). Tambien se utilizo un modelo predictivo de tipo BERT llamado `finbert`[5] el cual esta preentrenado con valoraciones de sentimiento sobre noticias del mundo financiero.

Metodo

Se escogen diferentes cadenas de noticias para utilizar como fuente para el analisis de sentimiento de textos. Ademas de las fuentes se escogen nombres de criptomonedas de las cuales recuperar las noticias. Esta informacion se utiliza para obtener las noticias por medio de un concentrador de noticias internacionales[6] tomando todas las noticias relacionadas a partir del dia anterior hasta el momento actual. Estas noticias se normalizan por medio del modulo `newspaper`[7] el cual utiliza modelos de `pln` para corregir la puntuacion y eliminar los simbolos o definiciones de `html` que no corresponde encontrarlos dentro del texto. Al texto normalizado se le aplica un modelo predictivo tipo BERT[8] denominado `finbert` para clasificarlos en positivo, negativo, neutro. En base a estos resultados se hace un conteo de las probabilidades para obtener un estimado del sentimiento aproximado de cada criptomoneda.

Conclusion

EL modelo de analisis de sentimientos utilizado en esta practica resulto util para clasificar el sentimiento acerca de un activo financiero a partir de noticias actuales y asi aportar informacion en la toma de decision de una inversion.

References

- [1] *Python3* - <https://www.python.org/>
- [2] *Pypi module happytransformer* - <https://pypi.org/project/newspaper3k/>
- [3] *Pypi module newspaper3k* - <https://pypi.org/project/newspaper3k/>
- [4] *Pypi module newsapi-python* - <https://pypi.org/project/newsapi-python/>
- [5] *ProsusAi/finbert* - <https://huggingface.co/ProsusAI/finbert>
- [6] *NewsApi usage* - <https://newsapi.org/docs/client-libraries/python>
- [7] *Newspaper usage* - <https://newspaper.readthedocs.io/en/latest/>
- [8] *HappyTransformer usage* - <https://happytransformer.com/text-classification/usage/>