



## Actividad - Proyecto Final

### Identificación

---

Asignatura: Matemática Aplicada

### Objetivo

---

Implementar un algoritmo de análisis de sentimiento difuso a partir de tweets.

### Actividad

---

Trabajar en grupos de máximo 2 personas, o se puede realizar de forma individual.

El proyecto final consistirá en implementar seis módulos del sistema de análisis de sentimiento propuesto en el siguiente artículo:

Vashishtha, S., & Susan, S. (2019). **Fuzzy rule based unsupervised sentiment analysis from social media posts**. Expert Systems with Applications, 138, 112834. [\[link\]](#)

Los módulos a implementar del artículo son:

1. **Creación de dataset.** Utilizar el dataset [Sentiment140](#). Este dataset también debe de ser pre-procesado de acuerdo al artículo (ver Sección 3.1).
2. **Lexicón de Sentimientos.** Implementar un módulo que utilice el un lexicón de sentimientos de la librería [NLKT](#), y calcula el puntaje positivo y negativo de cada registro en tu dataset siguiendo las instrucciones de la Sección 3.2 del artículo. Al utilizar un analizador de sentimientos de NLKT, solo usar los valores positivos y negativos que retorne. Estos puntajes se deben de incluir al dataset como dos columnas nuevas, puntaje positivo y puntaje negativo.
3. **Fuzzificación.** Implementar la fuzzificación siguiendo las instrucciones de la Sección 3.3.1.
4. **Base de reglas.** Utilizar [skfuzzy](#) para la creación de la base de reglas siguiendo las secciones 3.3.2 y 3.3.3.
5. **Defuzzificación.** Implementar la defuzzificación siguiendo las instrucciones de la Sección 3.3.4. Como resultado agregar una tercera columna al dataset que será el puntaje del sentimiento para cada registro.



6. **Benchmarks.** Crear benchmarks para probar la eficiencia y efectividad de tu sistema. Reporta por el total de tweets positivos, negativos, y neutrales. Por cada uno reporta también el tiempo que toma calcular su sentimiento. Además, reporta el tiempo promedio total.

El resultado final debe de ser un nuevo archivo csv detallando lo siguiente:

- Oración original, label original, puntaje positivo, puntaje negativo, Puntaje Positivo, Puntaje Negativo, Resultado de inferencia, tiempo de ejecución.
- Fuera del archivo csv, reportar el tiempo de ejecución promedio total.

Las condiciones de entrega son las siguientes:

- Todos los proyectos deberán de cargarse en el repositorio github de acceso público: <https://github.com/mdvillagra/matematica-aplicada-2024>.
- En la carpeta <https://github.com/mdvillagra/matematica-aplicada-2023/tree/main/proyectos> crear un directorio con los nombres de los integrantes del grupo, y dentro de ese directorio creado levantar el código fuente. Haz un fork del repositorio y luego un pull request al origen.
- La fecha y hora de entrega del proyecto se considerará como la hora y fecha del último *pull request* realizado en este repositorio. Se permitirán entregas hasta la fecha y hora del 2do. examen parcial.
- Los grupos pueden decidir el lenguaje de programación que deseen utilizar. En cualquier caso, se debe de agregar las instrucciones de compilación en un archivo *markdown* junto al código fuente. Si el proyecto posee muchas dependencias, se debe agregar un resolvidor de dependencias, por ejemplo, *makefile*, *requirements.txt*, etc.

En el día y hora de examen establecidos, cada grupo deberá de presentar y defender su código levantado en el repositorio. Si el proyecto se realiza en grupo, ambos integrantes deben de presentarse al examen.

El examen será realizado mediante videoconferencia en la fecha y horario establecido para el examen parcial.

## Plazo

---

El plazo de entrega del proyecto es hasta el día de la fecha del examen parcial **miércoles 6 de Noviembre hasta las 18:00hs.**

## Indicadores de Evaluación

---



La

actividad podrá realizarse de forma individual o en grupos de hasta 2 personas. No se admitirán grupos de 3 o más personas. Un trabajo entregado con 3 o más integrantes será evaluado con nota 0 sin excepciones. El *pull request* realizado en el repositorio será rechazado para grupos con 3 o más personas.

El código del proyecto solo se recibirá utilizando el repositorio github indicado. No se aceptarán códigos fuente por otros medios.

Los grupos que se presentan al examen deberán de defender su código mostrando cómo se ejecuta y mostrar ejemplos de ejecución. También deberán de responder preguntas del instructor del curso.

Si en el caso en que el código fue copiado de otro trabajo, o uno de los integrantes del equipo no puede explicar su código, el grupo llevará 0 puntos sin la necesidad de evaluar los módulos implementados. Si solo se presenta un integrante del grupo, solo ese integrante llevará nota.

Durante su presentación el grupo deberá compilar el código fuente y mostrar su ejecución. El código no debe de mostrar errores de compilación ni errores de ejecución. En caso contrario, la nota será de 0 puntos.

Si todas las condiciones mencionadas anteriormente se cumplen, la rúbrica de abajo se utilizará para corregir el código presentado.

RÚBRICA PARA EL CÓDIGO				
Criterio	0 puntos	1 punto	2 puntos	3 puntos
<b>Módulo 1</b> <b>Creación de dataset</b>	No implementa el módulo o no funciona.	Lee el dataset en una estructura de datos adecuada.	Preprocesa el texto extraído.	
<b>Módulo 2</b> <b>Lexicón de sentimientos</b>	No implementa el módulo o no funciona o no siguió el artículo para el cálculo de los puntajes.	Calcula correctamente uno de los puntajes positivos o negativos siguiendo el artículo.	Calcula correctamente ambos puntajes siguiendo el artículo.	Integra los puntajes al dataset como dos columnas nuevas.
<b>Módulo 3</b> <b>Fuzzificación</b>	No implementa el módulo o no funciona o no siguió el artículo.	Implementa correctamente 3 o 4 variables difusas.	Implementa correctamente 5 variables difusas.	Implementa correctamente las 6 variables difusas.
<b>Módulo 4</b>	No implementa el módulo o no		Implementa correctamente	Implementa correctamente la



<b>Base de reglas</b>	funciona.		todas las reglas.	agregación de reglas.
<b>Módulo 5</b> <b>Defuzzificador</b>	No implementa el módulo o no funciona.	Implementa correctamente el centro de área.	Implementa correctamente la salida del sistema de inferencia.	
<b>Módulo 6</b> <b>Benchmarks</b>	No implementa el módulo o no funciona.	Imprime el total de tweets positivos, neutrales y negativos.	Además de lo anterior, calcula el tiempo promedio de ejecución.	Además de lo anterior, calcula el tiempo promedio total.