

INFORME DE REVISIÓN

DATOS DEL PROYECTO DE TITULACIÓN Y DEL ESTUDIANTE			
Nombre y apellidos	Dulce María Reyes Lucas		
Nombre del Proyecto	Análisis de sentimiento y extracción de tópicos en comentarios de aplicaciones móviles del sector bancario		
DATOS DEL REVISOR			
Apellidos	Tellez Avila		
Nombre	Eric Sadit		
Institución de origen	INFOTEC		
Correo electrónico	eric.tellez@infotec.mx	Teléfono	4494646909

	VALORACIÓN DEL PROYECTO			
	Muy buena (10-9)	Buena (8-7)	Suficiente (6-5)	Deficiente (<5)
Originalidad/ Interés del tema			X	
Definición de los objetivos	X			
Marco teórico y antecedentes			X	
Metodología		X		
Relevancia de los resultados				
Discusión/ Conclusiones				
Bibliografía, otras fuentes y recursos			X	
Aspectos Formales				

Puede utilizar los siguientes espacios para efectuar observaciones o recomendaciones de mejora.

Definición de los objetivos
Queda claro el objetivo en general, pero es poco novedoso. Por otra parte, no es clara la motivación. Me imagino que la alumna trabaja para el banco en cuestión pero no se especifica. En cuanto, en cuanto a la parte de minería de opinión, la metodología que se usará para verificar el desempeño de los modelos es poco clara.
Marco teórico y antecedentes
Se necesita ahondar en la revisión del estado del arte, así como el marco teórico (procesamiento de lenguaje natural, minería de opinión, aprendizaje computacional, ¿deeplearning?). Me parece que se debe ampliar la revisión a literatura científica relacionada. Sobre los proyectos relacionados que se mencionan, habría que explicarlos mejor y citarlos de manera adecuada.
Metodología y su adecuación a los objetivos propuestos
Es necesario especificar una metodología para la parte de minería de opinión y evaluación de modelos. Lo que esta ahora es muy general sobre el proceso.
Relevancia y calidad de los resultados obtenidos
Me parece que aun no es posible evaluar esta parte.
Discusión / Conclusiones
Me parece que es necesario mejorar la parte metodológica, sobre todo del tratamiento de los datos y la evaluación de los modelos.
Bibliografía, otras fuentes y recursos

Tal vez podría servir el siguiente artículo para el uso de un vocabulario y recursos para el español

[Regionalized models for Spanish language variations based on Twitter](#)

ES Tellez, D Moctezuma, S Miranda, M Graff, G Ruiz - Language Resources and Evaluation, 2023

Ojo, no es un diccionario curado como sería el de la RAE o algo así, es más bien un vocabulario que sale de analizar el español de redes sociales. Si no te funciona como recurso, igual y si como idea para procesar tus datos y obtener un vocabulario.

Dicho esto, no creo que sea buena idea hacer corrección ortográfica. De hecho ningún tipo de corrección, pero si eso es lo que quieres hacer, también puedes revisar

<https://github.com/sadit/SimilaritySearch.jl/> (ahi hay varios artículos que explican como funciona) y en especial tiene implementada la función de distancia de Levenshtein

<https://github.com/sadit/SimilaritySearch.jl/blob/main/src/distances/strings.jl>

Aunque esta para Julia, puedes hacer una tabla hash que conecte vocabularios (vocabulario del dataset con vocabulario de un diccionario). También puedes usar los word-embeddings del artículo LRE23 para esta parte, no solo Levenshtein (semántico vs léxico).

Para la clasificación podrías usar el

<https://github.com/INGEOTEC/EvoMSA>

<https://ieeexplore.ieee.org/abstract/document/8956106>

Tiene una buena documentación y puede funcionar con modelos externos

Si quieres usar modelos basados en deeplearning, te recomendaría usar los BILMA

Aspectos formales

Valoración global

El tema es interesante para casos de uso pero poco novedoso, tal y como se describe. Es posible encontrar problemas interesantes sobre el análisis de tópicos y clasificación.

En cuanto a los reportes, son difíciles de leer pero la presentación es buena y clara. Me parece importante que la alumna ahonde en la metodología de procesamiento de lenguaje y aprendizaje computacional, más que en la parte de la extracción (aunque coincido que es necesario saber el origen y las características de los datasets para entender de mejor manera los datos)