

## PROCESSAMENTO DE LINGUAGEM NATURAL PARA DETECÇÃO DE COMPORTAMENTO DEPRESSIVO

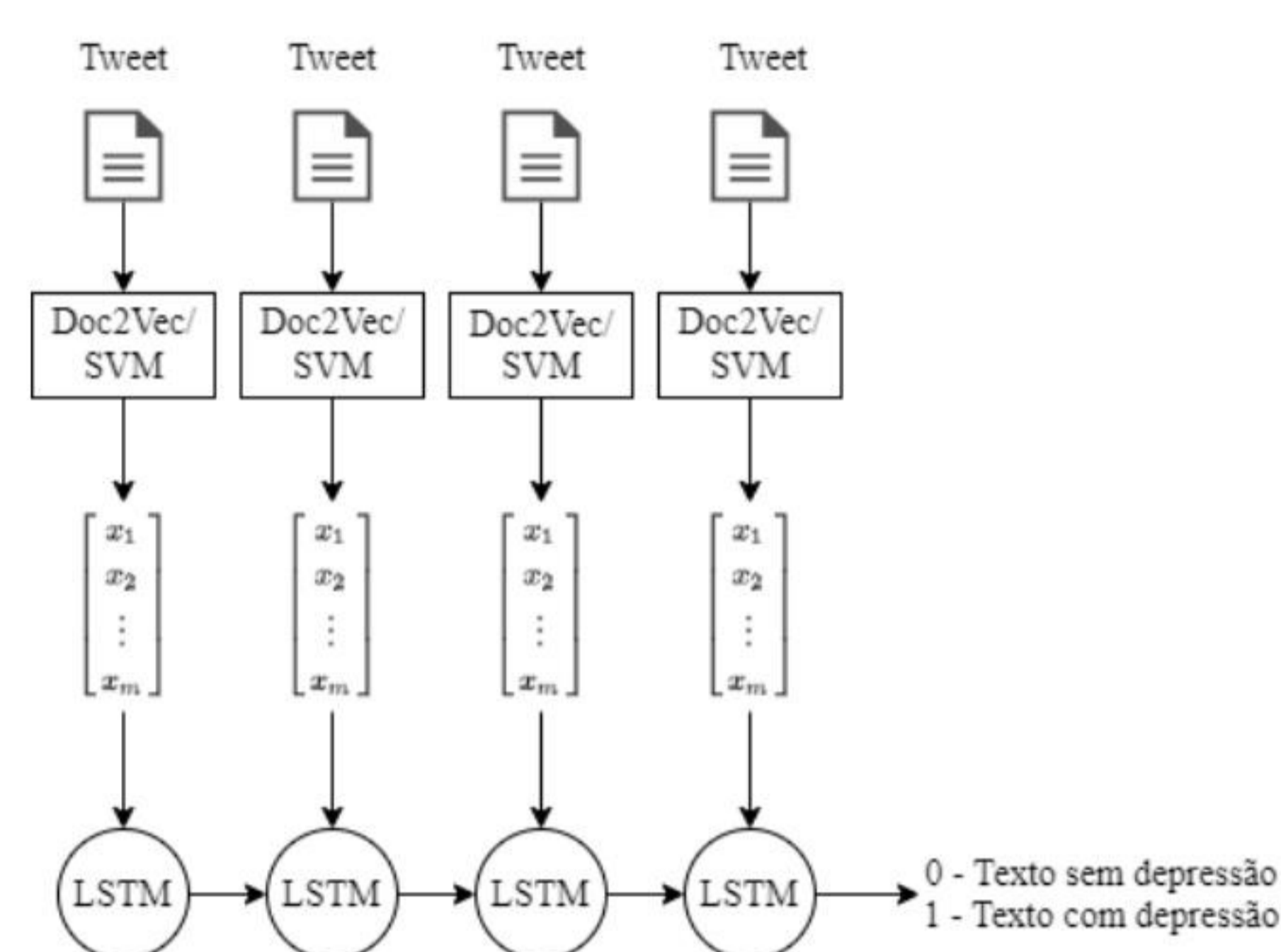
**Aluno:** Gabriel Melo, Youssef Khatib, Eduardo Conte e Kayke Bonafé (dev.gabrielomelo@gmail.com, youssef.ktb@hotmail.com, educontejr@gmail.com, kayke.bonafe98@gmail.com)

**Orientador:** Prof. Dr. Guilherme Alberto Wachs Lopes (gwachs@fei.edu.br)

**Resumo** A depressão é um assunto que vem ganhando destaque nos últimos anos. Um dos tópicos que normalmente é associado à depressão é o uso das redes sociais e o tempo de uso dos seus usuários. Há uma correlação entre a utilização de redes sociais e o desenvolvimento de depressão, dentre outros quadros clínicos. Este trabalho desenvolve, implementa e avalia um modelo computacional baseado em processamento de linguagem natural para classificação de tendências depressivas de usuários do Twitter por meio de suas postagens no decorrer do tempo. O melhor resultado obtido nesse projeto foi de 83,58% de F-measure.

**Descrição** A depressão afeta mais de 294 milhões de pessoas ao redor do mundo. Dessa forma, essa doença tem sido foco de muitas pesquisas científicas, desde seu diagnóstico até mesmo seu tratamento. Um dos tópicos que normalmente é associado à depressão são as redes sociais e o tempo de uso dos seus usuários. Com o avanço da tecnologia e modelos matemático-computacionais, houve um aumento significativo do uso da área processamento de linguagem natural (PLN), desde a classificação de usuários em redes sociais até análise de sentimentos em documentos. Este trabalho desenvolve, implementa e avalia um modelo computacional baseado em processamento de linguagem natural para classificação de tendências depressivas de usuários do Twitter por meio de suas postagens no decorrer do tempo. Como resultado, foi obtido um F-Measure de 83,58% utilizando-se do conteúdo textual acrescido da análise de sentimento dos documentos.

**Metodologia Utilizada** Para a realização deste trabalho, três principais pontos são utilizados: análise textual utilizando o Doc2Vec, análise de sentimentos utilizando SVM e ambas as análises em conjunto.



**Resultados** Como resultado, esse trabalho obteve, na melhor execução, uma precisão de 82,87%, revocação de 84,31%, acurácia de 83,46% e F-Measure 83,58% utilizando os dados concatenados.

	Técnicas	Tamanho da entrada	Tamanho da camada escondida	Precisão	Revocação	Acurácia	F-Measure	Limiar
Execução 1	Sentimento	2	2	0,5230	0,5242	0,5230	0,5236	0,4933
			4	0,5237	0,5197	0,5235	0,5217	0,4933
	Textual	300	300	0,8087	0,8202	0,8134	0,8144	0,5401
			600	<b>0,8248</b>	<b>0,8415</b>	<b>0,8315</b>	<b>0,8331</b>	<b>0,5241</b>
	Concatenado	301	301	0,8225	0,8193	0,8214	0,8209	0,9266
			602	0,8241	0,8404	0,8306	0,8322	0,4246
Execução 2	Sentimento	2	2	0,5227	0,5002	0,5231	0,5112	0,4967
			4	0,5213	0,5298	0,5231	0,5255	0,4935
	Textual	300	300	0,8162	0,8275	0,8212	0,8219	0,6234
			600	0,8265	0,8421	0,8330	0,8342	0,7241
	Concatenado	301	301	0,8189	0,8213	0,8199	0,8201	0,8044
			602	<b>0,8287</b>	<b>0,8431</b>	<b>0,8346</b>	<b>0,8358</b>	<b>0,6092</b>

**Conclusão** Como contribuições deste trabalho, pode-se dizer que a metodologia proposta teve F-measure de 83,58%, obtendo assim um valor próximo ao estado-da-arte. Além disso, é notável que a utilização de conteúdo textual se mostra mais discriminante que o sentimental na classificação de tendências depressivas em textos.