

Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



ANALYZING THE POTENTIAL OF FEATURE GROUPS FOR MISINFORMATION DETECTION IN WHATSAPP

Patrick de Angeli Almeida - Departamento de Informática (DPI) - UFV - patrick.angeli@ufv.br

Julio Cesar Soares dos Reis - Departamento de Informática (DPI) - UFV - jreis@ufv.br

Desinformação, Fake News, Aprendizado de Máquina, Plataformas Digitais, WhatsApp

Introdução

As plataformas digitais, incluindo redes sociais e aplicativos de mensagem instantânea, estão presentes em mais de 70% das telas iniciais dos smartphones dos usuários, e, neste contexto, o WhatsApp, pode ser considerado o aplicativo de comunicação mais utilizado no Brasil. Dentro deste assunto, um dos maiores desafios relacionados ao combate da desinformação em plataformas digitais, está relacionado à proposição de soluções que sejam úteis para mitigação do problema.

Resultados e Discussão

A Tabela abaixo apresenta os resultados experimentais obtidos neste trabalho, utilizando o classificador XGBoost e os atributos explorados. Considerando todas as combinações possíveis de grupos analisados (i.e., Conteúdo, Fonte e Ambiente), podemos observar que os melhores resultados foram obtidos explorando atributos dos grupos de Conteúdo + Fonte, com 0,82 e 0,95 em termos de AUC e Macro-F1, respectivamente. Em termos de AUC, atributos extraídos da fonte, de forma isolada ou combinada (e.g., + atributos extraídos do Ambiente), também mostraram resultados bastante promissores.

Conjuntos de Atributos	AUC	Macro-F1
Conteúdo	0,68 ($\pm 0,31$)	0,80 ($\pm 0,31$)
Fonte	0,80 ($\pm 0,19$)	0,84 ($\pm 0,19$)
Ambiente	0,79 ($\pm 0,39$)	0,86 ($\pm 0,39$)
Conteúdo + Fonte	0,82 ($\pm 0,18$)	0,95 ($\pm 0,18$)
Conteúdo + Ambiente	0,77 ($\pm 0,22$)	0,84 ($\pm 0,22$)
Fonte + Ambiente	0,79 ($\pm 0,20$)	0,86 ($\pm 0,20$)
Conteúdo + Fonte + Ambiente	0,78 ($\pm 0,21$)	0,84 ($\pm 0,21$)

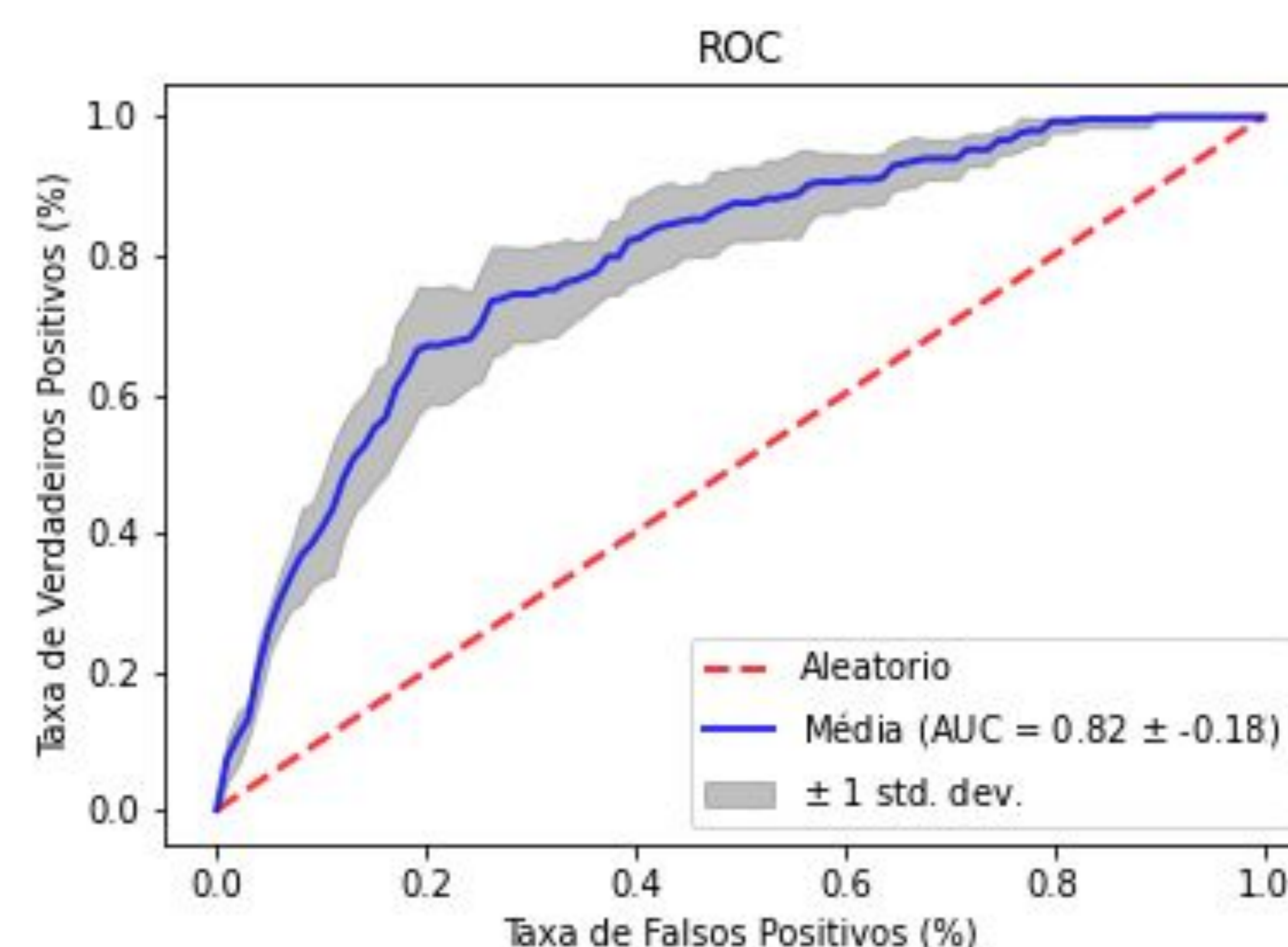
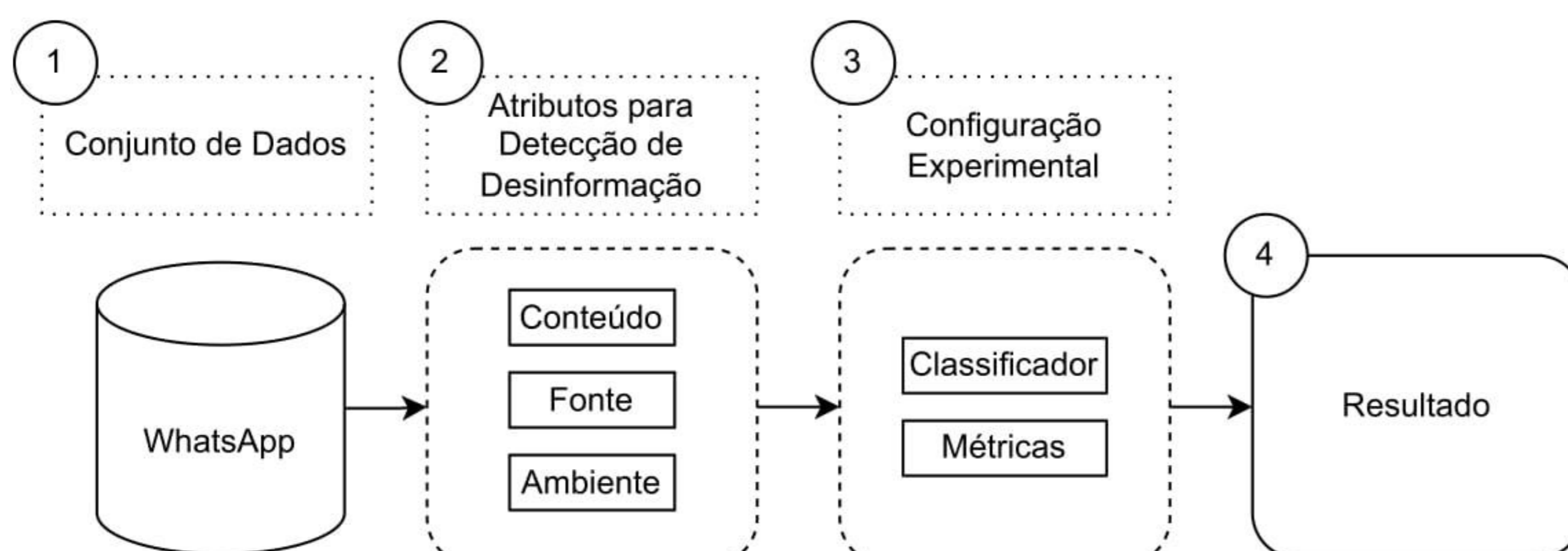
Objetivos

Este trabalho investiga o desempenho de modelos gerados a partir de grupos de atributos propostos e implementados em esforços anteriores. Para isso, exploramos uma base de dados do WhatsApp, com informações da última eleição presidencial brasileira, ou seja 2018, que notavelmente foi marcada pela disseminação de desinformação. Em seguida, avaliamos o desempenho de uma abordagem de aprendizado de máquina (i.e., XGBoost) na realização da tarefa de identificar desinformação.

Conclusões

De forma geral, os resultados mostram que é possível gerar bons modelos com um número reduzido de (grupos de) atributos, usando o XGBoost. Os melhores resultados de classificação podem detectar corretamente quase todas as mensagens contendo desinformação, enquanto classificam incorretamente aproximadamente 60%, o que acreditamos já seja suficiente para auxiliar as agências de checagem de fatos no processo de identificação do conteúdo a ser verificado.

Material e Métodos



Apoio Financeiro e Agradecimentos

Este trabalho foi parcialmente financiado pela FAPEMIG e CNPq (Edital UFV PIBIC/CNPq 2021-2022).

Bibliografia

Julio C. S. Reis et al. 2020. Towards automatic fake news detection in digital platforms: properties, limitations, and applications. (2020).