

ANEXO I

Modelo de Resumo

Resumo de Projeto de Iniciação Científica e Tecnológica

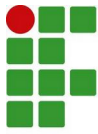
Os resumos dos projetos de iniciação e tecnológica são publicações com o objetivo de divulgar sinteticamente todos os projetos desenvolvidos por bolsistas do CNPq e pela comunidade acadêmica do IFSC. Caso os resultados do projeto tenham sido publicados, o título e resumo do mesmo devem ser alterados para se evitar plágio.

1. Identificação do Trabalho	
Título original do projeto	Aplicação de Técnicas de Aprendizagem de Máquina para Detecção de Fake News
Editais do projeto de pesquisa	02/2020/PROPPI/UNIVERSAL
Título para caderno de resumos	Desenvolvimento de Modelos Inteligentes para Detecção de Fake News
Coordenador do projeto de pesquisa	Cristiano Mesquita Garcia
E-mail do Coordenador	cristiano.garcia@ifsc.edu.br
Autores	Eduardo C. R. Borges, Samuel S. Feitosa, Cristiano M. Garcia
Palavras-chave	Aprendizado de máquina; Fake news; Classificação

2. Resumo do Trabalho *(entre 200 e 400 palavras, apenas texto, sem imagens, quadros ou tabelas. O resumo deve apresentar, necessariamente: objetivos, metodologia e resultados do projeto de pesquisa. O texto deve ser escrito de forma corrida, fonte Arial, tamanho 10, alinhamento justificado.)*

Introdução

Nos últimos anos, devido ao desenvolvimento e à democratização tecnológica, o acesso e produção de informações disponíveis na Internet estão cada vez mais facilitados. Embora isso seja claramente um ponto positivo, informações falsas podem ser transmitidas a diversas pessoas sem que haja qualquer controle. Fake News, que se caracterizam por serem fraudes, rumores ou desinformações, vêm se tornando um problema muito grande, especialmente em momentos decisivos como eleições e, mais recentemente, na vacinação contra a COVID-19. Técnicas de aprendizado de máquina podem ser utilizadas para detecção de padrões em mensagens com conteúdo potencialmente falso pela utilização de processamento de linguagem natural (PLN). O objetivo final deste trabalho, que ainda está em andamento, é a construção de uma extensão para navegadores que alerte o usuário em caso de revisão sistemática; obtenção de notícias classificadas em verdadeiras ou falsas; desenvolvimento de modelos de aprendizado de máquina para detecção de conteúdo potencialmente falso e fornecimento do modelo em forma de web service para consumo a partir de extensão para navegadores. Com exceção da última, todas as tarefas foram finalizadas. É esperado ainda a escrita de artigo científico com comparação com outros trabalhos da literatura. Foram utilizadas a linguagem de programação Python, com as bibliotecas Sci-kit Learn (PEDREGOSA F, et al., 2011), NLTK (BIRD et al, 2009) e Pandas (MCKINNEY et al, 2010) para uso de algoritmos de aprendizado de máquina, processamento de linguagem natural e manipulação de dados, respectivamente. Como base de dados, foi utilizado o Fake.br Corpus (SANTOS et al, 2018). Os experimentos foram realizados com diversos modelos de Machine Learning, que utilizam



técnicas de estatística e probabilidade, para avaliá-las de acordo com as métricas adequadas (Acurácia, precisão, revocação e F1-score). Entre os melhores resultados, estão Random Forest Classifier, Redes Neurais (MLPClassifier) e Light Gradient Boosting (LightGBM), esse obtendo o melhor resultado com uma acurácia de 92% outros modelos ficaram com uma porcentagem inferior, como o Adaboost: 88% e Logistic Regression: 79% . Também foram realizados testes com análise PCA para redução de dimensionalidade, o que ajudaria o modelo, reduzindo a complexidade no treinamento. Os resultados com esse processamento não obtiveram valores com diferença estatística significativa, se comparado aos modelos mencionados anteriormente, que não utilizaram PCA.

3. Referências Utilizadas no Trabalho *(seguir as normas da ABNT)*

BIRD, Steven; KLEIN, Ewan; LOPER, Edward. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.", 2009.

MCKINNEY, Wes et al. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010. p. 51-56.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. the Journal of machine Learning research, v. 12, p. 2825-2830, 2011.

SANTOS, Roney et al. The Fake.Br Corpus - A corpus of fake news for Brazilian Portuguese. In: Latin American and Iberian Languages Open Corpora Forum (OpenCor). 2018. p. 1-2.

4. Agradecimentos

A equipe do projeto agradece ao Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina – IFSC, pelo apoio recebido na forma de bolsas para discentes, viabilizando a execução das atividades do projeto de pesquisa.