Uso de Conceitos Básicos de Machine Learning Na Classificação de E-mails Como Spam

Lucas Lima da Cunha

lucas.lcunha@hotmail.com

Introdução

Mensagens de spams são consideradas um grande problema na internet, tanto por conta do gasto em processamento e armazenamento de um e-mail inútil, quanto pelos riscos que muitos apresentam a privacidade e segurança dos receptores, sendo carregados de conteúdo maliciosos.

Cientes deste problema empresas como a Google empenham grande investimento e esforço em melhorar suas ferramentas de detecção de spam através de técnicas como machine learning. Segundo um artigo online, recentemente ao passarem a utilizar um framework de ML de código aberto chamado TensorFlow, desenvolvido pela própria Google, eles passaram a bloquear por volta de 100 milhões de mensagens de spam a mais por dia e alegam uma precisão de mais de 99,9%ⁱ.

Analisando essa demanda de tecnologia para resolução deste problema foi desenvolvido um estudo com as aplicações de machine learning na classificação de emails.

Metodologia

Foram utilizadas a plataforma Colab, da Google, para desenvolvimento dos notebooks (códigos de ML), bibliotecas básicas e um dataset (conjunto de dados) contendo mais de 5 mil mensagens de e-mails, a quantidade que cada uma das 149 palavras mais comuns aparecem na mensagem, contagem de quantas vezes aparecem palavras comuns ao todo, a data da mensagem e um campo informando se é spam ou não.

O processo de estudo deste dataset foi dividido em dois. Na primeira parte foram extraídas estatísticas destes dados e na segunda foi desenvolvido um script que aprenderia com este dataset e seria capaz de identificar padrões de spam e classificar mensagens.

- Primeira Parte

Utilizando então a plataforma Colab foi importado o dataset e dele foi possível extrair estatísticas interessantes a partir de alguns comandos, como por exemplo uma nuvem de palavras evidenciando as palavras mais comuns, os dias de cada mês com maior movimentação de mensagens, gráficos comparando a quantidade de mensagens spam e não spam por mês entre outros.

- Segunda Parte

Na segunda parte, também utilizando o mesmo dataset e a plataforma Colab foi possível "ensinar" o código, apenas com os padrões de quais palavras e suas quantidades, que aparecem no conjunto de dados, a classificar mensagens de spam com uma alta taxa de acerto.

Para realizar esse teste o dataset foi divido em dois, a primeira parte contendo 80% dos registros foi utilizada para treinar o código e os outros 20% foram extraídos o campo de classificação do e-mail e então foi deixado que a máquina já treinada classificasse-os como spam ou não, tudo isso a partir de poucas linhas de código. Após a classificação pela máquina os resultados foram comparados com a coluna extraída anteriormente.

Resultados

Como resultados da primeira parte podemos observar dados e estatísticas interessantes, com poucas linhas de código e poucas bibliotecas, já na segunda parte, também em poucas linhas e baixa complexidade foi possível desenvolver um sistema de classificação com uma taxa de acerto de 97.11%, o que poderia ainda ser melhorado com mais treinos e tempo e se tornado uma aplicação real de classificação de e-mail.

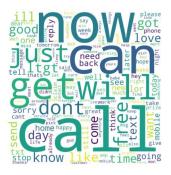


Figura 1 Palavras mais comuns

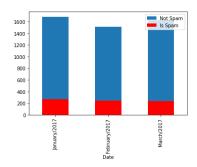


Figura 2 Spam e total por mês

Conclusão

O uso de machine learning para esta atividade permitiu o desenvolvimento de uma ferramenta simples e escalável com alta taxa de acerto e potencial para otimização, mostrando a capacidade de resolução de problemas através de ML e, em casos como este, o acessível desenvolvimento, testes e aplicações.

Referências

ⁱ KUMARAN, Neil. Spam does not bring us joy—ridding Gmail of 100 million more spam messages with TensorFlow. **Google Cloud**, 6 de fev. de 2019. Disponível em: < https://cloud.google.com/blog/products/g-suite/ridding-gmail-of-100-million-more-spam-messages-with-tensorflow>. Acesso em: 05 de fev. de 2021.