

Introdução

Nos dias de hoje, com as pessoas muito mais envolvidas no mundo digital, os meios de comunicação também seguiram se adaptando. Hoje temos diversas formas de sermos abordados por um anúncio ou propaganda, seja por meio de aplicativos de redes sociais ou até mesmo por meio de SMSs (mensagens de texto no celular). As pessoas diariamente dão seus números de celular para sites buscando acesso a serviços ou produtos, sem notar a cláusula que diz que seu número pode ser utilizado para receber ofertas e anúncios.

Desta maneira, existem formas de classificar as mensagens recebidas como um spam, ou seja, uma mensagem indesejada que foi enviada ao grande número de destinatários. Uma dessas maneiras é utilizar o aprendizado de máquina, onde mensagens já classificadas são submetidas a um processo de treinamento e teste. Assim que obtidos bons resultados o algoritmo pode ser aplicado para classificação das mensagens.

Um dos métodos existentes para classificação é a técnica de Naive Bayes. O método de Naive Bayes assume que as características de um conjunto de dados são independentes (Naive em tradução livre significa inocente). (DSILVA, 2018) Este modelo funciona bem com documentos de texto desde que as palavras no texto sejam independentes entre si e a localização de uma palavra não depende de outra palavra.

Metodologia

Para desenvolvimento do trabalho foi utilizada um conjunto de dados já minerado, contendo as mensagens de texto já classificadas entre spam ou não spam, as palavras presentes em cada mensagem de texto, assim como a quantidade de palavras por mensagem, quantidade de palavras comuns por mensagem e data da mensagem.

Estes dados foram submetidos a alguns cálculos, obtendo assim algumas informações. Foi desenvolvido um gráfico de nuvem de palavras que apresenta as palavras mais frequentes encontradas no conjunto de dados de acordo com o tamanho da palavra no gráfico, conforme Figura 1.

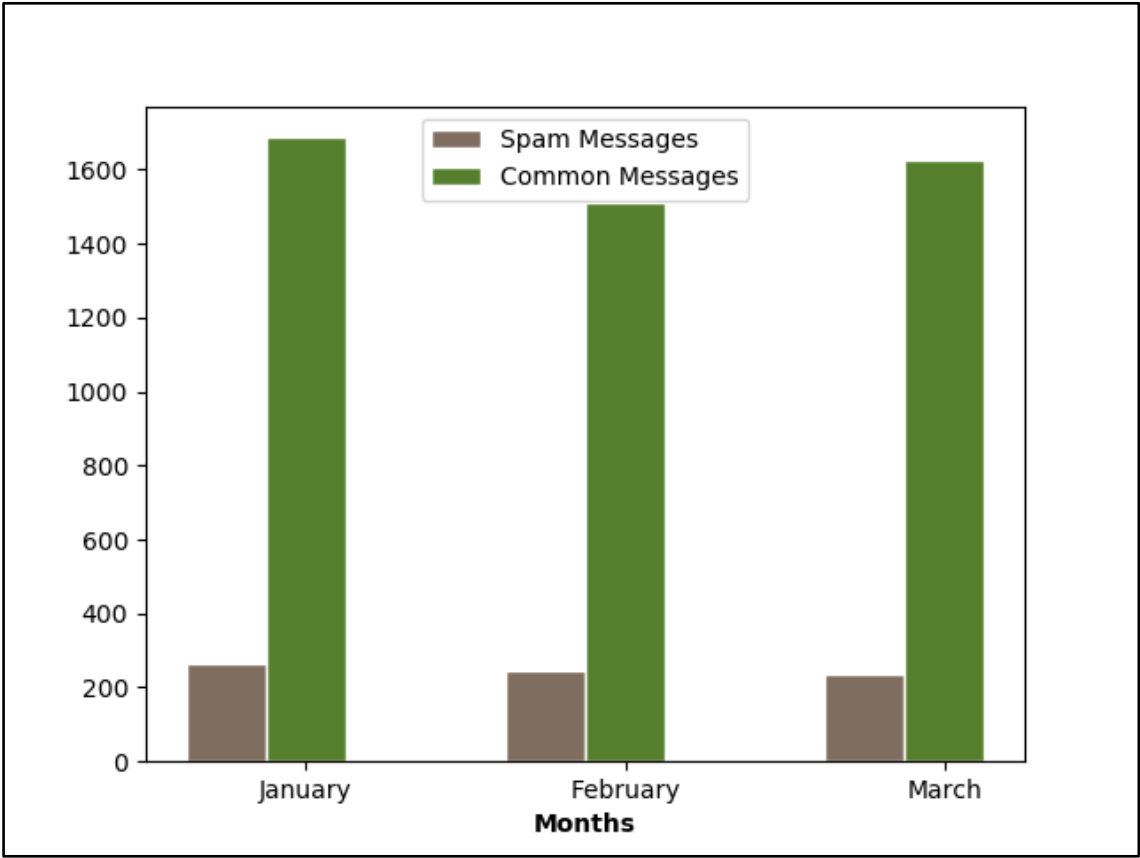
Figura 1 – Nuvem de palavras



Fonte: elaborado pelo autor

A partir deste gráfico é possível perceber quais são as palavras mais frequentes nas mensagens. Além disso, também foi desenvolvido um gráfico que demonstra um comparativo entre mensagens comuns e mensagens de spam recebidas por mês, conforme Figura 2.

Figura 2 - Mensagens recebidas por mês



Fonte: elaborado pelo autor

Com o gráfico acima é possível perceber a quantidade de mensagens de spam recebidas por mês, assim como a quantidade de mensagens comuns.

Durante a aplicação do algoritmo de Naive Bayes, foram utilizados 20% das mensagens para testes e os outros 80% para treinamento. Também foi optado pelo método multinominal Naive Bayes por considerar a contagem de palavras como entrada. (SIDDARTHA, 2018).

Resultados

Sobre os resultados obtidos, após a aplicação do método de Naive Bayes sobre o conjunto de dados de teste (20% de todo conjunto), foi possível atingir os valores abaixo, conforme Quadro 1.

Quadro 1 - Resultados do teste

Métrica	Valor
Acurácia	98,91%
Precisão	98,72%
Sensibilidade	93,90%
F1-Score	0.96

Fonte: elaborado pelo autor.

Conforme Rodrigues (2019), os dados obtidos podem ser explicados da seguinte maneira:

- Acurácia: indica uma performance geral do modelo, dentre todas as classificações, quantas o modelo classificou corretamente.
- Precisão: considerando as classificações que o modelo fez como positivo, quantas estão corretas. Dada pela fórmula: Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Positivos).
- Sensibilidade: dentre as situações de valor positivo como esperado, quantas estão corretas. Dada pela fórmula: Verdadeiros Positivos / (Verdadeiros Positivos + Falsos Negativos).
- F1-Score: Média harmônica entre precisão e sensibilidade, obtendo um valor entre 0 e 1, onde 1 é o melhor resultado possível.

Utilizando as mesmas métricas de avaliação explicadas acima, o método também foi executado sobre todo o conjunto de dados, a fim de reclassificá-lo. O Quadro 2 apresenta os dados obtidos.

Quadro 2 - Resultados sobre conjunto de dados

Métrica	Valor
Acurácia	99,28%
Precisão	98,36%
Sensibilidade	96,25%
F1-Score	0,97

Fonte: elaborado pelo autor.

Conforme os valores obtidos, é possível notar que os resultados foram positivos em relação a classificação das mensagens de texto. Com base no valor de precisão obtido, consegue-se perceber qual foi o percentual de mensagens que foram classificadas como spam quando na verdade são mensagens comuns, calcula-se a diferença de 100% e o valor de precisão, atingindo assim o valor de 1,64% de mensagens comuns classificadas como spam.

Ao avaliar os tempos de execução das etapas de treinamento, teste e classificação, são obtidos os resultados no Quadro 3 abaixo

Quadro 3 - Tempos de execução

Etapas	Tempo (segundos)
Treinamento	0.0059986114501953125
Teste	0.0009999275207519531
Classificação	0.001001119613647461

Fonte: elaborado pelo autor.

As medições de tempo foram obtidas na ordem apresentada, portanto o tempo de classificação foi medido após a etapa de teste ser executada.

Conclusão

Com o desenvolvimento deste trabalho conclui-se que o algoritmo de Naive Bayes com o método multinominal tem um bom resultado para classificação de documentos de textos, como SMSs e e-mails. Devido ao fato de que o algoritmo com este modelo funciona com características discretas, são utilizadas as contagens de palavras dos documentos de texto para classificação, considerando que as palavras não necessariamente estão relacionadas.

No cenário estudado os métodos utilizados tiveram o comportamento dentro do esperado, atingindo boa acurácia, precisão e F1-Score, ou seja, em poucos casos as mensagens foram classificadas incorretamente, fazendo com que uma mensagem comum válida fosse classificada como spam.

Por fim, os tempos de execução do algoritmo também são satisfatórios considerando o volume de dados (mais de 5000 mensagens de texto).

Referências

DSILVA, Deepal. **Ham or Spam? SMS Text Classification with Machine Learning**. 2018. Disponível em: <https://towardsdatascience.com/sms-text-classification-a51defc2361c>. Acesso em: 01 mar. 2020.

RODRIGUES, Vitor. **Métricas de Avaliação**: acurácia, precisão, recall... quais as diferenças?. 2019. Disponível em: <https://medium.com/@vitorborbarodrigues/m%C3%A9tricas-de-avalia%C3%A7%C3%A3o-acur%C3%A1cia-precis%C3%A3o-recall-quais-as-diferen%C3%A7as-c8f05e0a513c>. Acesso em: 03 mar. 2020.

SIDDARTHA, Manu. **SMS Spam Classifier**: Naive Bayes ML Algo. 2018. Disponível em: <https://www.kaggle.com/sid321axn/sms-spam-classifier-naive-bayes-ml-algo>. Acesso em: 01 mar. 2020.