

Gabriela Cavalcanti de Albuquerque Ayres

Análise de sentimentos de clientes bancários  
em rede social Twitter

Trabalho de Conclusão de Curso apresentado na  
Escola de Economia de São Paulo  
da Fundação Getúlio Vargas, para obtenção do  
certificado de conclusão de curso de especialização  
em Business Analytics e Big Data

São Paulo, Novembro de 2018

Gabriela Cavalcanti de Albuquerque Ayres

Análise de sentimentos de clientes bancários  
em rede social Twitter

**Orientador**

Prof. Gustavo Mirapalheta

**Avaliação**

Satisfatório

☐

Não Satisfatório

☐

## Resumo

Este estudo teve como objetivo de conhecer e explorar o nível de satisfação dos clientes em relação aos produtos oferecidos pelo banco Next através de rede social Twitter. O público em geral do banco é um público jovem e voltado para millenials. Em geral é um público que gosta de receber “mimos” e não há interesse em pagamento de taxas para serviços. Temos uma hipótese que os clientes do Next querem primeiramente um cartão de crédito “Na faixa” e que tentam conseguir o cartão de crédito no Nubank e caso sejam rejeitados acabam por procurar outras opções disponíveis no mercado e acabam vindo para o Next. Eles não sabem que o Next é um banco digital até baixarem o app e comecem a fazer uso da conta e outros recursos.

Para ajudar nessa análise, foram lidos os tweets dos clientes pelo @falanext e criadas nuvens de palavras para identificar o que os clientes falam a respeito do Next e se quando falam, outros bancos também são mencionados. Foi utilizada a técnica de análise de sentimentos para identificar os sentimentos dos clientes ao falarem sobre o Next para saber se estão satisfeitos ou não com o banco e se poderiam ajudar a fazer recomendações para outros colegas e amigos.

Os principais resultados confirmaram que o uso da nuvem de palavras se mostrou útil e alguns tweets realmente mencionam outros bancos, mas muitos agradecem quando recebem o cartão e alguns mimos, como descontos em Uber e cinema. Pudemos perceber que os clientes em sua grande maioria estão satisfeitos com os serviços, pois as palavras utilizadas nos tweets possuem maior característica positiva. Esse trabalho pode ser utilizado para direcionar melhor os produtos e serviços dos bancos para os clientes e ajudar a evitar a evasão de clientes positivos, tentando identificar o que falam a respeito do banco e tomar ações preventivas.

**Palavras chave:** Estratégia de oferta de produtos e serviços, algoritmos de classificação, machine learning, text mining, análise de sentimentos, nuvem de palavras.

## Índice

|   |    |
|---|----|
| Introdução.....                                       | 6  |
| Objetivo .....  | 8  |
| Dados Utilizados.....                                 | 9  |
| Abordagem Metodológica .....                          | 10 |
| Extração de tweets .....                              | 11 |
| Preparação e Limpeza dos Dados .....                  | 12 |
| Retirada de StopWords.....                            | 12 |
| Separação da amostra em learn e test.....             | 13 |
| Algoritmos de classificação .....                     | 13 |
| Acurácia dos algoritmos de classificação .....        | 14 |
| Léxico em português para análise de sentimentos ..... | 14 |
| Classificação de palavras.....                        | 15 |
| Análises .....  | 16 |
| Análise dos tweets positivos .....                    | 18 |
| Análise dos tweets negativos .....                    | 19 |
| Comparação com outros tweets de bancos digitais ..... | 21 |
| Conclusão .....                                       | 24 |
| Referências .....                                     | 24 |
| Próximos Passos .....                                 | 25 |
| Anexos.....   | 26 |
| Requisitos de pacotes .....                           | 26 |

## Índice de Figuras

|  |    |
|--|----|
| Figura 1 - Visão de Arquitetura da Técnica Proposta.....           | 11 |
| Figura 2 - Extração de dados do Twitter .....                      | 11 |
| Figura 3 - Tweets após a preparação e limpeza de dados .....       | 12 |
| Figura 4 - Lista de tweets após a classificação .....              | 15 |
| Figura 5 - Nuvem de Palavras de Tweets @falanext.....              | 16 |
| Figura 6 - Palavras mais utilizadas nos tweets .....               | 17 |
| Figura 7 - Nuvem de palavras de tweets com a palavra 'Conta' ..... | 17 |
| Figura 8 - Percentual de classificação dos tweets Next .....       | 18 |
| Figura 9 - WordCloud de tweets classificados como positivos.....   | 19 |
| Figura 10 - Palavras mais utilizadas nos tweets positivos .....    | 19 |
| Figura 11 - WordCloud de tweets classificados como negativos ..... | 20 |
| Figura 12 - Palavras mais utilizadas nos tweets negativos.....     | 20 |
| Figura 13 - Nuvem de palavras com a palavra 'Atendimento'.....     | 21 |
| Figura 14 - Percentual de classificação dos tweets Nubank.....     | 21 |
| Figura 15 - Nuvem de Palavras de Tweets @nubank .....              | 22 |
| Figura 16 - Amostra de tweets negativos Nubank.....                | 22 |
| Figura 17 - Nuvem de palavras @falanext e @nubank .....            | 23 |
| Figura 18 - Percentual de tweets @falanext e @nubank .....         | 23 |

## Introdução

O uso da internet passou a fazer parte da vida das pessoas. Nas instituições financeiras, as pessoas não querem mais ir ao banco, elas desejam resolver tudo de forma simples, rápida e fácil pela internet, sem muitas burocracias. Os dados estão sendo coletados e armazenados em um ritmo dramático, sendo a maioria dos dados em formato semi-estruturado, e muitos podem conter informações úteis. Cada vez mais as pessoas tendem a compartilhar suas experiências sobre produtos e serviços prestados e fazerem recomendações. Eles usam as redes sociais como Facebook, Twitter, Whatsup, entre outros para expressar seus sentimentos, compartilhar suas experiências e logo perceberam que suas reclamações são rapidamente resolvidas quando expostas nas redes sociais. Para um melhor atendimento ao cliente e a experiência dos serviços, tornou-se uma prática comum revisar os comentários dos clientes nas redes sociais. Entender os comentários dos clientes sobre os produtos e serviços é muito útil tanto para a instituição financeira quanto para as pessoas que desejam se tornar clientes. Ler todos os comentários um por um não é tão eficiente quando o número de comentários é muito grande, uma vez que o conteúdo dos comentários também pode trazer confusões. Algumas análises podem conter frases longas, gírias e expressões idiomáticas, por esse motivo, é bem difícil ler e entender o significado dos comentários.

Se alguém ler apenas alguns números de resenhas e chegar a uma decisão, a decisão poderá ser tendenciosa. Por causa dessas razões, uma melhor técnica de mineração de dados para avaliar os comentários dos clientes que estão em formato semi-estruturado é muito importante. Não apenas os comentários sobre o banco em que são clientes, mas também sobre outros bancos para extrair sua opinião verdadeira. Esse problema vem sendo estudado por muitos pesquisadores nos últimos anos. A área de pesquisa é chamada de mineração de opinião e análise de sentimentos. Existem duas tarefas principais desta área de pesquisa. Eles estão encontrando características dos produtos ou serviços que foram comentadas pelos clientes e decidindo se os comentários são positivos ou negativos. Ambas as tarefas são muito desafiadoras e diferentes pesquisas foram realizadas nesse aspecto. Embora ambas as tarefas sejam cobertas por várias abordagens de pesquisa, há algumas áreas a serem melhoradas. Alguns deles estão identificando verbos, frases verbais e algumas sentenças condicionais. A exploração do uso de “smileys”

na avaliação de opiniões, melhorando as técnicas existentes são alguns dos trabalhos futuros que foram identificados. Emoções são nossos sentimentos e pensamentos subjetivos têm sido estudados em vários campos, pois estão intimamente relacionadas aos sentimentos. A força de um sentimento ou opinião é tipicamente ligada à intensidade de certas emoções. Nas mídias sociais, as pessoas costumam expressar suas emoções usando diferentes "smileys" e se tornou uma tendência hoje. Assim, usar tanto o léxico do sentimento como "smileys" juntos em um algoritmo para avaliar a opinião dos clientes nas redes sociais será mais bem-sucedida. Considerando todos os potenciais de sinergia de mineração de dados, o desenvolvimento de algoritmos de mineração de dados para a extração de opinião pode ser identificado como uma área de pesquisa realmente importante. Essa pesquisa foi conduzida utilizando a análise de opinião baseada em recursos das opiniões e comentários de clientes, utilizando o léxico em português Re-Li, mas acreditamos que poderá ser melhorado no futuro se adicionado o recurso de "smileys" e algumas expressões verbais e gírias, muito usadas nas redes sociais.

## Objetivo

A empresa alvo deste estudo iniciou suas atividades na metade de 2017, com o objetivo de atingir um público jovem, bancado pelo tradicional banco Bradesco. Inicialmente, o Next oferecia apenas planos pagos, ainda que com um período de testes gratuito. O banco lançou uma conta corrente sem tarifas e com cartão de crédito internacional. O Next é um dos bancos digitais mais vantajosos hoje em dia. Por padrão, a conta digital do Next oferece um cartão de débito, que também pode ter a função crédito caso o usuário seja aprovado em uma análise de crédito. Além disso, o Next tem conta poupança e limite de cheque especial compartilhado com empréstimos e cartão de crédito, que é disponibilizado na hora na conta corrente. Ainda que o Next não tenha agência própria, é possível ser atendido em agências do Bradesco e usar os caixas eletrônicos do banco tradicional para sacar dinheiro. Por fim, correntistas do Next têm direito a aplicar dinheiro em um fundo de investimento de renda fixa com taxa de administração anual de 1% que tem títulos atrelados à taxa Selic.

A conta digital do Next é gerenciada pelo aplicativo para iPhone e Android e tem alguns diferenciais interessantes:

- **Flow:** um recurso para controlar como você gasta o dinheiro no Next;
- **Vaquinha:** para quem quer juntar dinheiro com amigos para um objetivo em específico;
- **Objetivo:** é possível definir quanto se quer economizar e o dinheiro é aplicado no fundo de investimento;
- **Mimos:** alguns descontos que vêm com cartões de crédito do Next, como R\$ 20 de desconto no Uber, 50% de desconto no ingresso do Cinemark, entre outros.

O Next possui outros concorrentes como o banco Inter, o Nubank e o Original e cada vez é mais importante entender as características do seu público alvo e suas necessidades, para um atendimento de qualidade e garantia de satisfação, assim como evitar a evasão de clientes positivos.

O público do Next é um público jovem que usa muito as redes sociais e normalmente se utiliza dos recursos digitais para se expressar. Portanto é preciso entender o feedback desses clientes e responder as seguintes perguntas:



- O que trouxe o cliente para o Next? Algum produto específico?
- O que mantém o cliente no Next?
- O que o cliente mais gosta e menos gosta do Next?

Com essas respostas, o banco será capaz de direcionar melhores ofertas e oferecer produtos e serviços de forma diferenciada, conseguir atrair novos clientes e manter sua carteira de clientes de forma crescente.

## Dados Utilizados

Foi criada uma conta de desenvolvedor no [Twitter](#) para possibilitar a coleta dos comentários de clientes através da API Search utilizando a string @falanext.

Após a criação da conta, foi necessário criar alguns dados para viabilizar a autenticação do usuário:

- Consumer Key (API Key)
- Consumer Secret (API Secret)
- Access Token
- Access Token Secret

Um conjunto de tweets foi recuperado de um período até 7 dias, tendo uma média de 500 tweets no total a cada extração realizada.

Os tweets normalmente possuem uma grande quantidade de caracteres especiais contidos em suas mensagens, tornando necessário um processo de limpeza dos dados, antes de iniciar sua análise e interpretação.

## Abordagem Metodológica

Esse trabalho foi realizado em linguagem de programação Python, devido ser esta a linguagem adotada pelo banco Bradesco para projetos voltados a Big Data e Analytics.

A técnica proposta desta pesquisa é uma técnica de mineração de dados e análise de sentimentos baseada em algoritmos de classificação e para que a pesquisa pudesse ser adequadamente realizada e evitar o overfit, uma amostra em torno de 500 tweets foi dividida em learn e test, sendo que a amostra de test foi definida em 0,3 ou 30% do total de tweets. Foram realizados testes com os algoritmos de classificação: Random Forest, Naive Bayes, Boosting e Dummy e posteriormente avaliamos qual dos algoritmos teve um melhor nível de acurácia para então iniciar o processo de classificação dos tweets e as análises propostas neste estudo.

A classificação tem o objetivo de criar categorias positivas e negativas para o conjunto de tweets que foi coletado. Para realizar a classificação foi utilizado o léxico Re-Li [1] em português que separa as palavras por:

- Adjetivos (positivos e negativos)
- Expressões idiomáticas (positivos e negativos)
- Substantivos (positivos e negativos)
- Verbos (positivos e negativos)

Após a classificação dos tweets, foram criadas nuvens de palavras para identificar quais as palavras que foram mais mencionadas nos tweets com o objetivo de identificar os produtos comentados, as maiores satisfações e também os principais problemas referentes às reclamações dos clientes.

Para que as técnicas pudessem ser aplicadas foi necessária a realização de um processo de preparação e limpeza dos dados, uma vez que os tweets possuem caracteres especiais, menções a empresas ou clientes, hashtags, links, entre outros.

Após a preparação e limpeza dos dados foram retiradas palavras consideradas “StopWords” que consistem em artigos, preposições e algumas palavras que não agregam valor na pesquisa e análise que será feita.

A Figura 1 fornece a visão geral da arquitetura de nosso sistema de análise dos tweets.

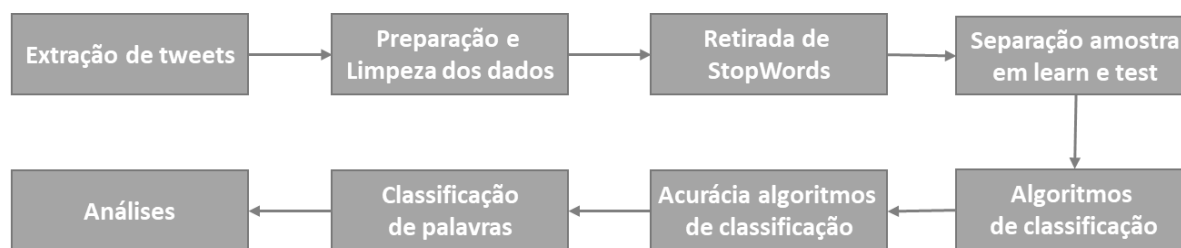


Figura 1 - Visão de Arquitetura da Técnica Proposta

## Extração de tweets

Para a coleta dos tweets foi necessária a criação de uma conta no Twitter e a definição de dados para autenticação conforme abaixo:

- Consumer Key (API Key)
- Consumer Secret (API Secret)
- Access Token
- Access Token Secret

Uma média de 500 tweets foram recuperados de um período de 7 dias. É possível verificar de acordo com a figura 2 um exemplo da extração dos tweets de forma bruta ainda sem os tratamentos de limpeza dos dados aplicados, assim como sua identificação e data de criação.

|   | id                  | created_at          | text   |
|---|---------------------|---------------------|--|
| 0 | 1062088767849328641 | 2018-11-12 21:04:29 | b'desde que cancelei minha conta pessoa flxc3\l... |
| 1 | 1062083613745991682 | 2018-11-12 20:44:00 | b'Toda vez que eu mando um tweet para o @falan...  |
| 2 | 1062083383772221440 | 2018-11-12 20:43:05 | b'@falanext Resolveu, obrigado \xe2\x9d\xa4\xce... |
| 3 | 1062082652809977857 | 2018-11-12 20:40:11 | b'@falanext Ahhh bom, n'xc3\xa3o sabia! A comp...  |
| 4 | 1062082525005275136 | 2018-11-12 20:39:41 | b'@falanext Ainda n'xc3\xa3o verifiquei mas vo...  |
| 5 | 1062081613213982721 | 2018-11-12 20:36:03 | b'RT @AlanMeloU: Ei @veek_oficial. Por qu'xc3\...  |
| 6 | 1062074105585983488 | 2018-11-12 20:06:13 | b'@falanext Quero cart'xc3\xa3o s'xc3\xb3 pra ...  |
| 7 | 1062073801251475462 | 2018-11-12 20:05:01 | b'@falanext De qualquer forma, nunca consegui ...  |
| 8 | 1062073331158052872 | 2018-11-12 20:03:09 | b'@falanext O cart'xc3\xa3o foi recusado 2 vez...  |
| 9 | 1062072064784429057 | 2018-11-12 19:58:07 | b'@Bradesco @falanext Me ajuda kkkkk fui reali...  |

Figura 2 - Extração de dados do Twitter

## Preparação e Limpeza dos Dados

Como se pode observar, os tweets recuperados de forma bruta, possuem muitos caracteres especiais, links, referências, entre outros. Para que a classificação dos tweets pudesse ser realizada, foi necessária a realização de um processo de preparação e limpeza dos dados que consistiu em:

- Retirada de expressões regulares (regex)
- Decodificação HTML
- Retirada de caracteres especiais (UTF-8)
- Retirada de menções a empresas ou clientes com @nome
- Retirada de links URL
- Retirada de hashtags

A Figura 3 demonstra o resultado dos tweets após a realização da preparação e limpeza dos dados.

|   | text  |
|---|---|
| 0 | desde que cancelei minha conta pessoa física n... |
| 1 | toda vez que eu mando um tweet para o eles re...  |
| 2 | resolveu obrigado ❤️                              |
| 3 | ahhh bom não sabia a compra foi realizada sex...  |
| 4 | ainda não verifiquei mas vou tentar e já retorno  |
| 5 | rt ei por quê vocês não fazem pareceria com ...   |
| 6 | quero cartão só pra receber meu salário crédi...  |
| 7 | de qualquer forma nunca consegui contato com ...  |
| 8 | o cartão foi recusado 2 vezes hoje pela manhã...  |
| 9 | me ajuda kkkkk fui realizar um depósito mas ...   |

Figura 3 - Tweets após a preparação e limpeza de dados

## Retirada de StopWords

Para que seja possível avaliar as palavras mais utilizadas nos tweets, é recomendado fazer a retirada de StopWords que consiste em preposições, artigos, alguns verbos como ser e estar, uma vez que essas palavras não terão muito a agregar nas análises desejadas.

Para retirada das StopWords foi utilizado o pacote nltk em português.

```
stopwords = nltk.corpus.stopwords.words('portuguese')
```

Na análise, pudemos perceber que há uma grande quantidade de utilização de gírias e abreviações de palavras como, tô, vc, pq, pra, pro, q, tá, entre outros. Essas palavras foram incluídas nas StopWords para que fosse possível uma análise com um melhor nível de qualidade.

## Separação da amostra em learn e test

Normalmente em técnicas de machine learning, para evitar que o algoritmo utilizado fique “viciado” e aprenda a prever apenas os dados parecidos com a amostra dos dados originais, separamos a amostra, ou seja, os 500 tweets em learn e test. Na prática, costuma-se dividir os dados em 70% para treino e 30% para teste, mas isso não é uma regra, pode ser 90% e 10% ou o que for mais adequado, dependendo muito da característica dos dados a serem analisados e dos requisitos de negócio. No caso desse estudo, para evitar o overfit, a amostra original foi dividida em 70% e 30% e foi utilizado o pacote scikitlearn do Python.

## Algoritmos de classificação

A classificação de dados está presente em diversos problemas reais, tais como: reconhecer padrões em imagens, diferenciar espécies de plantas, classificar tumores benignos e malignos, dentre outros. Este problema é um dos tópicos mais ativos na área de machine learning. O problema de classificação consiste em determinar o rótulo de algum objeto, baseado em um conjunto de atributos extraídos do mesmo. Para que isso ocorra é necessário um conjunto de treinamento com instâncias na qual os rótulos os objetos são conhecidos.

Para esse estudo foram utilizados 4 tipos de algoritmos de classificação do pacote sklearn do Python:

- Random Forest
- Naive Bayes

- Boosting
- Dummy

Como a variável utilizada eram as palavras dos tweets, foi necessário transformar as palavras em números, uma vez que era necessário que a variável independente fosse contínua para conseguirmos ranqueá-la. Para essa transformação foram utilizadas as funções LabelEncoder e fit\_transform do Python para que posteriormente fosse possível rodar os algoritmos de classificação propostos e verificar as variáveis preditivas.

## Acurácia dos algoritmos de classificação

Após rodar os algoritmos de classificação propostos nesse estudo, foi realizada a comparação do nível de acurácia para os algoritmos utilizados e obtivemos os seguintes resultados:

- Random Forest: 0,7
- Naive Bayes: 0,63
- Boosting: 0,66
- Dummy: 0,55

Quanto maior o nível de acurácia do algoritmo, maior é seu nível de acerto para a classificação entre positivos e negativos. Nesse caso, significa que para o algoritmo Random Forest 70% dos tweets foram classificados adequadamente e os outros 30% foram classificados como falso positivos ou falso negativos, mas podemos considerar que 30% é um bom nível de aceitação para essa margem de erro.

Foi utilizada a biblioteca sklearn para os algoritmos de classificação: Random Forest, Naive Bayes, Boosting e Dummy.

## Léxico em português para análise de sentimentos

Para o uso de text mining em português para análise de sentimentos, ainda não há muitos léxicos testados e disponíveis no mercado, nem mesmo nas funções disponíveis nos pacotes do Python. Há muitos léxicos

disponíveis em inglês, mas português, foi encontrado apenas o léxico ReLi, criado no âmbito do projeto Anotadores Semânticos baseados em Aprendizado Ativo, coordenado pelo Departamento de Informática da PUC-Rio. Mais informações sobre o léxico estão disponíveis nas referências desse estudo.

## Classificação de palavras

De acordo com o nível de acurácia obtido entre a comparação dos algoritmos testados, percebemos que os algoritmos não tiveram muita diferença de classificação, entretanto pelo fato do Random Forest ter se destacado como o melhor algoritmo, ou seja, maior taxa de acerto de classificação, foi utilizado esse algoritmo para realizar a classificação dos tweets em positivos e negativos.

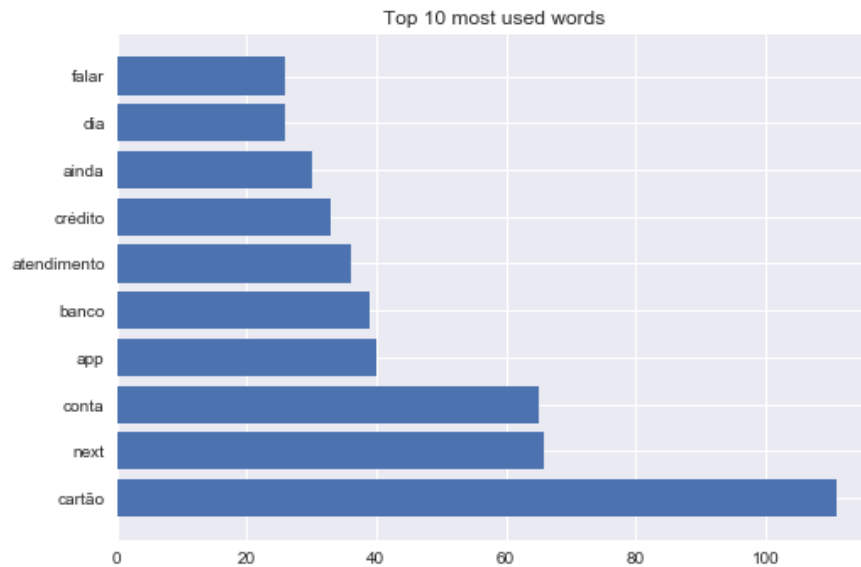
Na figura 4 é possível observar os tweets já classificados em positivos e negativos após rodar o algoritmo de classificação Random Forest e a consulta ao léxico ReLi.

|   | text  | sentiment |
|---|---|-----------|
| 0 | desde que cancelei minha conta pessoa física n... | pos       |
| 1 | toda vez que eu mando um tweet para o eles re...  | pos       |
| 2 | resolveu obrigado ❤️                              | neg       |
| 3 | ahhh bom não sabia a compra foi realizada sex...  | pos       |
| 4 | ainda não verifiquei mas vou tentar e já retorno  | neg       |
| 5 | rt ei por quê vocês não fazem pareceria com ...   | pos       |
| 6 | quero cartão só pra receber meu salário crédi...  | pos       |
| 7 | de qualquer forma nunca consegui contato com ...  | pos       |
| 8 | o cartão foi recusado 2 vezes hoje pela manhã...  | pos       |
| 9 | me ajuda kkkkk fui realizar um depósito mas ...   | pos       |

Figura 4 - Lista de tweets após a classificação







Para uma melhor interpretação do significado da palavra conta nos tweets, rodamos novamente a nuvem de palavras com '@falanext' e a palavra 'conta' e obtivemos a nova nuvem de palavras na figura 7.



Nesse caso, quando os clientes mencionam a palavra ‘Conta’ pudemos perceber que há um interesse geral dos clientes em ‘mimos’ como descontos, uber, ifood, cinemark e ofertas grátis.

De acordo com a análise de sentimentos realizada com os tweets dos clientes Next, foi criado o gráfico da figura 8 e pelo observado parece que os clientes no geral encontram-se satisfeitos, uma vez que 68% dos tweets foram classificados como positivos, entretanto há grande oportunidade de melhorias, pois 32% foram tweets considerados negativos. Por esse motivo, decidimos realizar novas análises de forma a separar os tweets positivos dos negativos a fim de identificar o que está sendo feito que os clientes gostam e que deve continuar a ser feito e o que deve ser feito para melhorar a fim de melhorar os tweets considerados negativos.

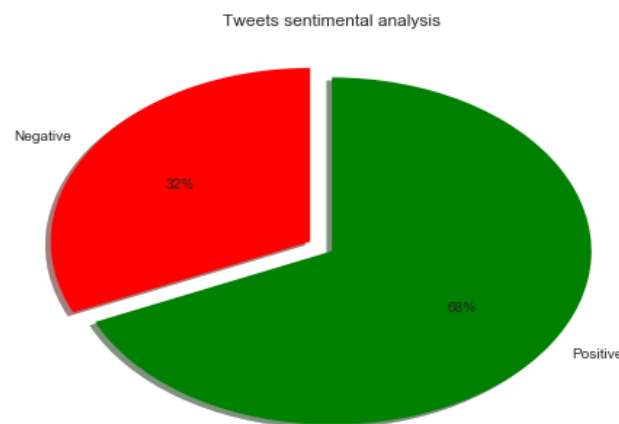


Figura 8 - Percentual de classificação dos tweets Next

### Análise dos tweets positivos

Uma vez que a classificação dos tweets foi feita em positivos e negativos, conseguimos realizar um filtro para recuperar apenas os tweets considerados positivos e montar nova nuvem de palavras a fim de conseguir identificar o que deve continuar a ser oferecido aos clientes.

Como se pode observar, a figura 9 representa a nuvem de palavras de tweets positivos, mas não houve muita diferença da nuvem de palavras apresentada anteriormente, com tweets positivos e negativos.



Figura 9 - WordCloud de tweets classificados como positivos

Fizemos novo gráfico de barras para identificar as palavras mais utilizadas nos tweets positivos, mas também não foi possível identificar diferenças quanto ao gráfico anterior, pois entendemos que pelo fato dos tweets positivos prevalecerem em relação aos negativos, essa separação não se mostrou eficiente.

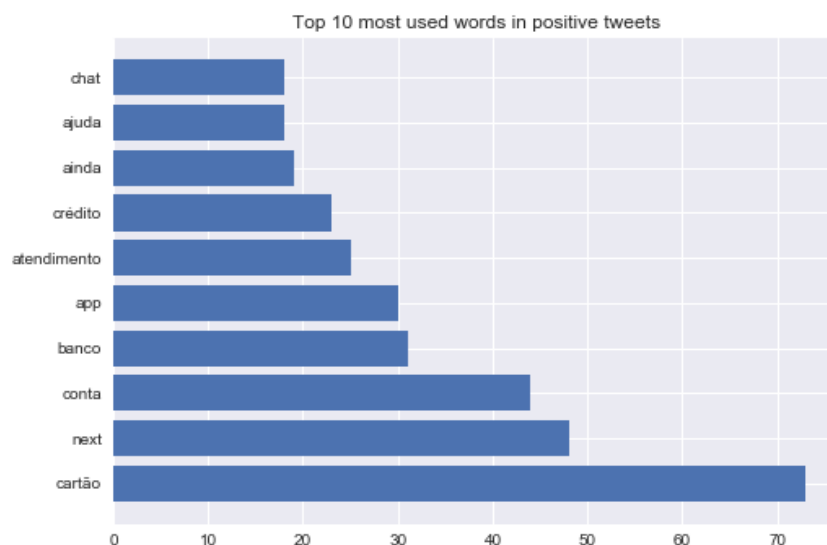


Figura 10 - Palavras mais utilizadas nos tweets positivos

## Análise dos tweets negativos

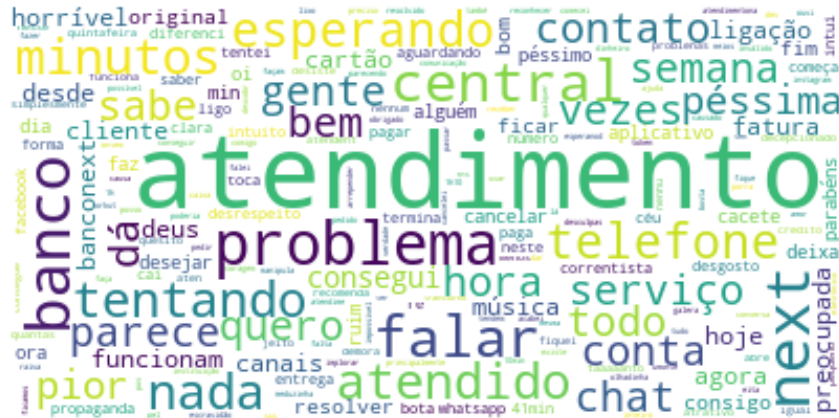
Como a separação dos tweets positivos não se mostrou eficiente, resolvemos fazer a análise dos tweets negativos de forma a identificar os principais problemas onde é possível realizar investimentos para melhorias. De acordo com a nuvem de palavras apresentada na figura 11, continuamos a apresentar as

[illegible]

| Word        | Frequency (approx.) |
|-------------|---------------------|
| cartão      | 38                  |
| conta       | 21                  |
| next        | 18                  |
| ainda       | 11                  |
| atendimento | 11                  |
| crédito     | 10                  |
| app         | 10                  |
| usar        | 10                  |
| fatura      | 9                   |
| bam         | 9                   |

20

Pudemos observar na nuvem de palavras referente a figura 13 que outras palavras aparecem em destaque como: minutos, esperando, tentando, falar, central, telefone. Dá a entender que o cliente está tentando solucionar algum problema e está esperando muito tempo para ser atendido ao telefone.



*Figura 13 - Nuvem de palavras com a palavra 'Atendimento'*

### Comparação com outros tweets de bancos digitais

Resolvemos entender como está o posicionamento dos tweets do Next em relação ao Nubank para avaliar se a classificação de seus tweets possui uma margem semelhante a quantidade de classificações negativas. Pudemos perceber pelo gráfico da figura 14 que mostra uma diferença bem acentuada, a magem de tweets negativos está em apenas 3% e 97% são considerados tweets positivos.

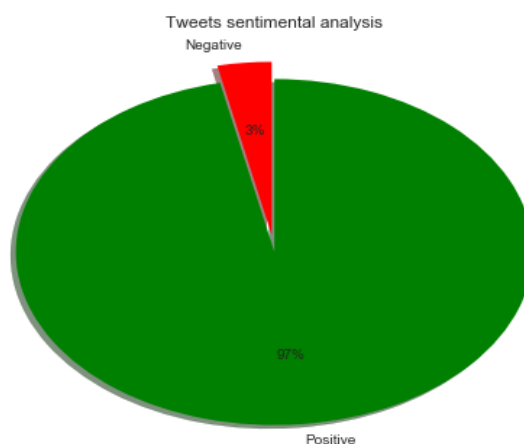


Figura 14 - Percentual de classificação dos tweets Nubank

Percebemos pela análise da nuvem de palavras que os clientes vem mencionando bastante a palavra contactless, que é a nova tecnologia de passar o cartão apenas por contato, com exceção das palavras cartão, nubank e roxinho que dizem muito respeito ao produto e a empresa.

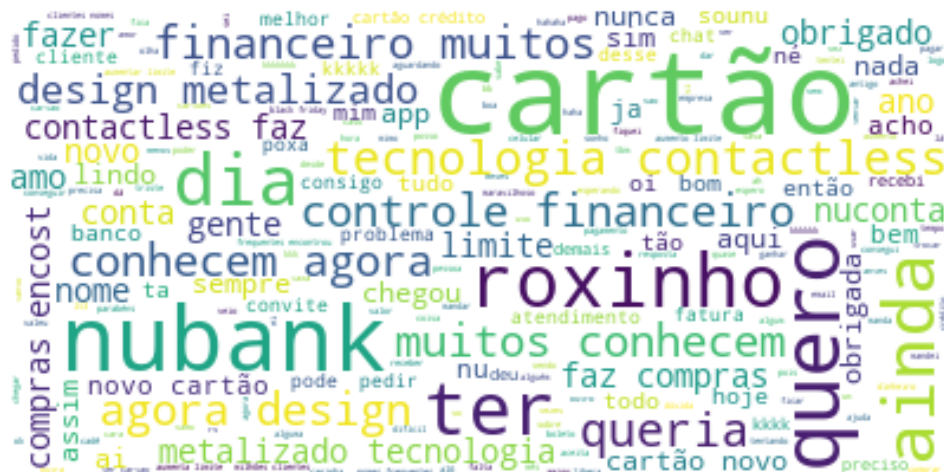


Figura 15 - Nuvem de Palavras de Tweets @nubank

Em relação ao percentual de tweets negativos, não foi possível identificar pela nuvem de palavras e nem pela quantidade das palavras mais utilizadas o que representa um problema para o Nubank, mas pelo filtro de tweets negativos, foi possível identificar que muitos clientes solicitam o cartão e são rejeitados pela análise do Nubank.

|      |  |     |
|------|--|-----|
| 6187 | chocado tenho nubank e não sabia que não ti... | neg |
| 6190 | basta esperar vencer ou seja 2 ou 3 anos pa... | neg |
| 6204 | oi estou te indicando pra pedir o nubank us... | neg |
| 6206 | eu quero uma capinha dessa na minha época s... | neg |
| 6216 | ah eu quero um desses                          | neg |
| 6227 | me enviou um novo cartão recentemente mas n... | neg |
| 6288 | desbloqueei o meu essa semana e estou gosta... | neg |
| 6294 | uso muito como sinônimo de mexase vamos agir   | neg |
| 6298 | oie me manda um convite                        | neg |
| 6303 | alguém me manda um convite                     | neg |

Figura 16 - Amostra de tweets negativos Nubank

Para finalizar a análise, gostaríamos de entender se os clientes que são rejeitados no Nubank, procuram o Next para tentar um cartão de crédito. Para isso, realizamos nova nuvem de palavras, mas dessa vez filtrando as palavras @falanext e @nubank.



Figura 17 - Nuvem de palavras @falanext e @nubank

A hipótese de que os clientes que não são aceitos no Nubank tentam conseguir o cartão do Next posteriormente não é conclusiva, mas pudemos perceber que os clientes quando falam do Nubank juntamente com o Next, o percentual de tweets positivos fica bem maior ao apresentando individualmente pelo Next. Novas análises precisariam ser feitas para que essa hipótese seja confirmada.

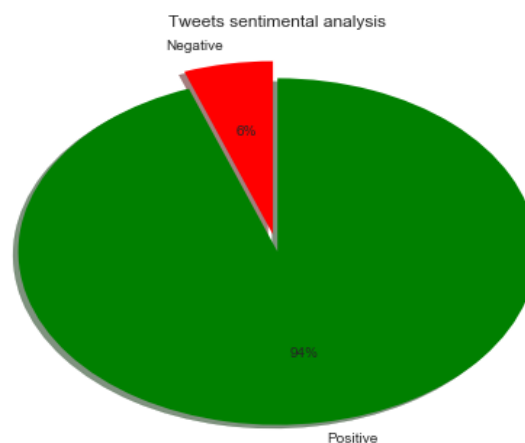


Figura 18 - Percentual de tweets @falanext e @nubank

## Conclusão

Pelo estudo realizado, obtivemos algumas sinalizações para recomendações ao Next, são apenas sinalizações uma vez que só foi realizado o estudo referente a rede social Twitter. Para que esse estudo seja mais conclusivo seria necessário recuperar informações de outras redes sociais e provavelmente dados internos do banco. Como uma grande oportunidade, entendemos que o banco pode direcionar mais esforços no produto cartão de crédito e continuar a ofertar mimos. Outros produtos oferecidos como investimentos, vaquinha, flow não são comentados pelos clientes e não foi percebido tanto interesse nos tweets. Como ponto de melhoria, poderia investir em melhor atendimento aos clientes e investir em tecnologia contactless, uma vez que seu concorrente já parece disponibilizar essa tecnologia e há demonstração de bastante interesse por parte dos clientes nos tweets.

## Referências

**Foundations of Applied Math Labs - Jeffrey Humpherys**

**Re-Li - Léxico em português**

<https://www.linguateca.pt/Repositorio/ReLi/>

**Naive Bayes Classifier**

<https://streamhacker.com/2010/05/10/text-classification-sentiment-analysis-naive-bayes-classifier/>

**Text Classification**

<https://www.nltk.org/book/ch06.html>

<https://textblob.readthedocs.io/en/dev/classifiers.html>



**scikit-learn**

<https://scikit-learn.org/stable/>

**WordCloud in Python**

<https://www.datacamp.com/community/tutorials/wordcloud-python>

**Classifier Algorithms**

<https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>

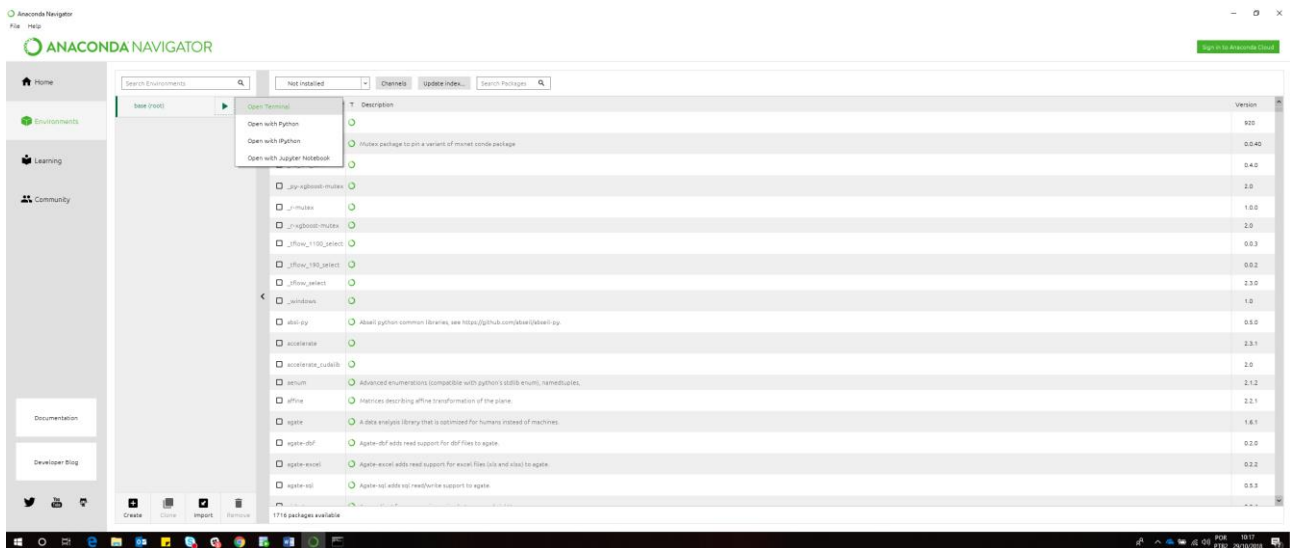
## Próximos Passos

Novas coletas de dados podem ser feitas de outras redes sociais para melhorar as análises realizadas e cruzamento com dados internos do banco podem ser feitas de forma a direcionar um melhor estudo e análise. Uma sugestão para melhorias é fazer análise de voz de forma a investigar o tom de voz dos clientes ao falarem no atendimento, na URA ou com a BIA de forma a entender suas principais queixas e reclamações e classificar seus sentimentos também em voz.

## Anexos

### Requisitos de pacotes

### Instalação dos pacotes Python



### Textblob:

```
conda install -c conda-forge textblob
```

### Tweepy:

```
conda install -c conda-forge tweepy
```

### NLTK:

```
conda install -c conda-forge nltk
```

### WordCloud:

```
conda install -c conda-forge wordcloud
```

### HTMLParser:

```
conda install -c conda-forge HTMLParser
```

### BeautifulSoup

```
conda install -c conda-forge beautifulsoup
```