

Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Bruna Luzzi

**Análise de avaliações de consumidores sobre
aplicativos bancários utilizando mineração de
texto**

**Curitiba
2023**

Bruna Luzzi

Análise de avaliações de consumidores sobre aplicativos bancários utilizando mineração de texto

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Walmes M. Zeviani

Curitiba
2023

Análise de avaliações de consumidores sobre aplicativos bancários utilizando mineração de texto

An analysis of mobile banking apps reviews using text mining

Bruna Luzzi¹

¹Departamento de Estatística, Universidade Federal do Paraná Av. Cel. Francisco H. dos Santos 100, Jardim das Américas, 81530-000, Curitiba, PR, Brasil*

A utilização dos aplicativos bancários para realização de transações financeiras vem substituindo os canais de atendimento presencial. Dentro deste contexto, este artigo propõe a análise de avaliações de consumidores através da utilização de técnicas de mineração de texto para entender a experiência dos clientes na utilização do mobile banking. Os métodos utilizados foram a modelagem de tópicos Latent Dirichlet Allocation, POS tagging e word embedding. Enquanto o modelo LDA não retornou resultados satisfatórios, a combinação dos métodos de POS tagging e word embedding mostrou que as principais funcionalidades mencionadas pelos usuários são relacionadas à eficiência, facilidade de utilização, segurança e produtos ofertados pelos bancos. **Palavras-chave:** aplicativos bancários, experiência do cliente, reviews, mineração de texto

The use of banking apps for financial transactions is replacing in-person service channels. Within this context, this article proposes the analysis of customer reviews through the use of text mining techniques to understand customers experience with mobile banking. The methods employed were Latent Dirichlet Allocation (LDA) topic modeling, POS tagging, and word embedding. While the LDA model did not return satisfactory results, the combination of POS tagging and word embedding showed that the main functionalities mentioned by users are related to efficiency, usability, security, and financial products.

Keywords: mobile banking, customer experience, reviews, text mining

1. Introdução

O avanço tecnológico teve um efeito significativo nos serviços bancários, proporcionando maior agilidade e flexibilidade aos clientes. Dentre os diversos fatores que têm colaborado para a migração dos clientes do atendimento presencial para os aplicativos bancários, destacam-se a conveniência e a praticidade oferecidas pelos aplicativos, permitindo realizar transações financeiras a qualquer momento e em qualquer lugar, a crescente familiaridade e adoção de smartphones pela população e, mais recentemente, a pandemia da COVID-19 que acelerou significativamente o aumento da utilização de aplicativos bancários, à medida que as restrições de distanciamento social incentivaram os clientes a buscar alternativas seguras e convenientes para realizar transações financeiras sem sair de casa.

Segundo estudos disponibilizados pela Federação Brasileira de Bancos (Febraban), as operações realiza-

das no mobile banking no ano de 2022 tiveram alta de 54% em comparação com 2021 e este canal já é o primeiro no ranking em volume de transações realizadas, com aproximadamente 107 bilhões de transações em 2022, em segundo lugar está o canal internet banking com 17 bilhões de transações [1]. Estes dados evidenciam a importância de compreender a experiência dos clientes na utilização dos aplicativos bancários.

Dentro deste contexto, este artigo propõe a análise de avaliações de consumidores de aplicativos bancários através da utilização de técnicas de mineração de texto a fim de identificar quais são as funcionalidades mais citadas pelos seus usuários. O objetivo é extrair informações relevantes que sirvam de insumo para que as instituições financeiras possam identificar oportunidades de melhorias na experiência dos clientes.

Dois bancos brasileiros foram selecionados para compor o estudo, um banco originalmente digital e um banco tradicional. Esta escolha tem como objetivo indagar a existência de diferenças significativas nas funci-

*luzziBruna@hotmail.com

onalidades mais citadas pelo público de consumidores de cada um destes distintos modelos de bancos.

O presente artigo está estruturado da seguinte forma: A seção 2 aborda a descrição da extração dos dados e das ferramentas utilizadas nas análises; a seção 3 apresenta os resultados obtidos e, por fim, a seção 4 contém as considerações finais relacionadas ao estudo.

2. Materiais e Métodos

2.1. Dados

Os reviews dos aplicativos dos bancos brasileiros Bradesco e Nubank foram extraídos do serviço de distribuição digital *Google Play Store* através do pacote *Google-Play-Scraper* disponível para a linguagem *Python*. Na Fig. 1, podemos observar alguns exemplos dos comentários extraídos.

1. Aplicativo não está acessando está dando sistema indisponível.
2. Muito instável. Já desinstalei e reinstalei e app não funciona. Muitos erros.
3. Ruim muito ruim
4. Aplicativo funcionava perfeitamente, mas de uns tempos pra cá trava toda hora e fica dando mensagens de erro
5. Fica pedindo atualização a todos os acessos, mesmo aceitando recentemente precisa resolver
6. Demora pra instalar o app

Figura 1: Exemplos de avaliações.

Além do conteúdo textual das avaliações, outras informações também são disponibilizadas por esta API, são elas: *reviewId*, *userName*, *userImage*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt* e *appVersion*.

As informações utilizadas neste estudo foram o “*reviewId*” para identificação única de cada avaliação, o “*content*” que é o texto comentado pelo usuário, o “*score*” fornecido pelo usuário, um número de 1 a 5 que indica o grau de satisfação dos usuários das aplicações e o campo “*at*” que indica a data em que o comentário foi publicado.

2.2. Mineração de Texto

A mineração de texto é uma ferramenta computacional que permite extrair informações relevantes a partir de um extenso volume de texto não estruturado. Essa área de estudo se dedica à conversão de dados não estruturados em conjuntos de dados estruturados, resultando

em economia de tempo e esforço. Além disso, a mineração de texto tem como objetivo reduzir a dimensionalidade dos dados, tornando-os mais gerenciáveis, significativos e relevantes [2].

Neste estudo, três métodos muito relevantes no campo da mineração de texto foram utilizados. O primeiro método empregado foi a modelagem de tópicos, metodologia muito útil que tem como objetivo descobrir padrões no uso de palavras e conectar documentos que compartilham padrões [3]. No âmbito da análise sintática, utilizaram-se duas importantes ferramentas: Part Of Speech Tagging e Word Embedding.

A Fig. 2 abaixo apresenta um workflow simplificado do experimento bem como as ferramentas utilizadas em cada etapa.

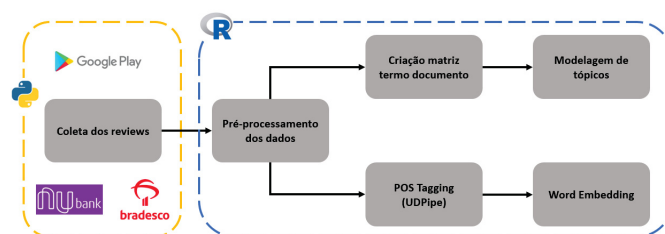


Figura 2: Workflow do experimento.

2.3. Latent Dirichlet Allocation

Uma das técnicas mais utilizadas na modelagem de tópicos é o Latent Dirichlet Allocation, um modelo probabilístico proposto por Blei et al. [4], no qual os documentos são representados como misturas de tópicos, onde cada tópico é caracterizado por uma distribuição de probabilidade dos termos.

O LDA se vale de dois hiper-parâmetros, que são oriundos da distribuição de Dirichlet: α que representa a densidade de tópicos por documento e β que representa a densidade de termos por tópico [5]. A Fig. 3 abaixo ilustra o fluxo de informações dentro deste modelo.

2.4. Part Of Speech Tagging

A análise lexical morfológica, frequentemente referenciada como Part of Speech (POS) Tagging é um campo da mineração de texto que utiliza o contexto para atribuir categorias lexicais às palavras que compõem um texto [6].

Existem diversos algoritmos que podem ser usados para a realização do POS Tagging, neste trabalho utilizou-se o algoritmo treinado UDPipe, disponível para a lin-

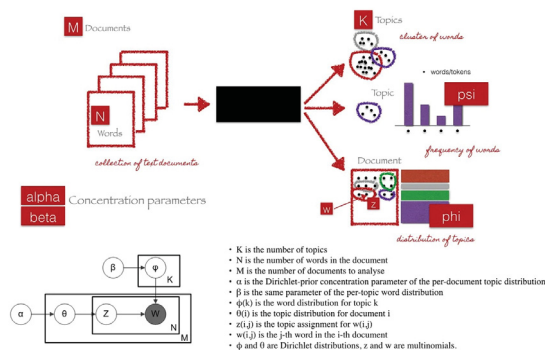


Figura 3: Pipeline LDA. (Fonte <http://chdoig.github.io/pytexas2015-topic-modeling/\protect#3/4>)

guagem de programação R [7]. O UDPipe é um pipeline treinável que oferece uma solução prática para realizar tarefas de tokenização independente de linguagem, pós-marcação, lematização e análise de dependência de texto bruto [8].

A tokenização consiste em segmentar um texto em unidades linguísticas menores com um significado único, a fim de facilitar o processamento de textos e o processo de lematização consiste em reduzir todas as inflexões possíveis de cada palavra em sua forma base ou lema, reduzindo a complexidade do vocabulário [9].

2.5. Word Embedding

O Word Embedding é um modelo autosupervisionado proposto por Pennington et al. [10], que consiste na representação vetorial das palavras a partir de um corpus baseada em redes neurais para capturar relações semânticas entre as palavras e entender o contexto em que uma palavra é usada.

Um dos métodos mais conhecidos para produzir o word embedding, e que será utilizado neste artigo, é o Global Vectors GloVe. Este modelo está baseado em duas etapas principais, onde a primeira etapa consiste na construção de uma matriz de coocorrência de termos a partir de um corpus de treinamento, na segunda etapa a fatoração da matriz é realizada para obter os vetores que representam os termos [11].

3. Resultados

3.1. Conjunto de Dados

Os aplicativos bancários estão em constante atualização, por esse motivo somente as avaliações realizadas nos últimos 6 meses, referente a data de desenvolvi-

mento deste projeto, foram selecionadas para compor a análise, a fim de evitar a identificação de oportunidades de melhorias que já foram contempladas pelas instituições financeiras.

Dado o objetivo de identificar oportunidades de melhorias nos aplicativos bancários, somente os comentários de teor negativo foram selecionados para compor o conjunto de dados. Para isso, define-se que comentários aos quais foram atribuídos score 1 ou 2 são negativos.

Com a finalidade de selecionar comentários que tenham valor informacional para a análise, somente reviews com 30 ou mais caracteres foram mantidos no dataset, eliminando comentários similares a: “Muito bom!”, “Muito ruim.”, “péssimo”, “ótimo”, entre outros.

3.2. Processamento dos Dados

A partir do conjunto de dados selecionado conforme descrito na seção anterior, utilizou-se o pacote **tm**, disponível para a linguagem de programação de código aberto R, para limpar e padronizar os dados. Os processos realizados estão descritos abaixo.

- **Conversão do texto para caixa baixa:** Esse processo é necessário pois as técnicas utilizadas neste estudo são *case sensitive*, isto é, sem a realização da conversão para caixa baixa palavras como “Banco” e “banco” seriam consideradas como dois termos distintos.
- **Remoção de acentuação:** Assim como descrito acima, a remoção de acentuação é necessária para que palavras digitadas erroneamente, sem acentuação ou com acentuação incorreta não sejam considerados dois termos diferentes durante a aplicação dos métodos.
- **Remoção de pontuação e numerais:** As técnicas utilizadas não distinguem pontuações e dígitos numéricos de outros caracteres, além disso, na grande maioria das vezes numerais não influenciam no significado dos textos.
- **Remoção de stopwords:** *Stopwords* são palavras que não possuem valor analítico, isto é, não adicionam significado ao texto, normalmente compostas por artigos, preposições e pronomes. Outras palavras podem ser adicionadas a relação de *stopwords* dependendo do contexto que está sendo analisado.
- **Remoção de espaços em branco:** Remoção dos espaços em branco que são gerados a partir das limpezas descritas acima.

Limite	Fazer	Cartão
Consigo	Limite	Nunca
Nada	Agora	Crédito
Dinheiro	Não	Nada
Tudo	Consigo	Dinheiro
Entrar	Cartão	Pix
Senha	Crédito	Limite
Ruim	Pix	Consigo
Abrir	Dias	Saldo
Porque	Vou	Dia
Assim	Fica	Sempre

Tabela 2: Palavras mais frequentes determinadas pela modelagem de tópicos utilizando $k = 3$ grupos para o Banco Nubank.

Chave	Abre	Consigo	Horrível	Erro
Trava	Péssimo	Abrir	Lento	Horrível
Ruim	Pix	Fica	Pessimo	Seguranca
Pix	Acessar	Fazer	Segurança	Pior
Entrar	Senha	Pix	Consigo	Péssimo
Consigo	Toda	Agora	Abrir	Fazer
Sempre	Fazer	Péssimo	Fazer	Abrir
Agencia	Atualização	Acessar	Agencia	Agencia
Nunca	Vou	Pagar	Ruim	Consigo
Fazer	Celular	Abre	Erro	Trava
Horrível	Erro	Senha	Agora	Veze

Tabela 3: Palavras mais frequentes determinadas pela modelagem de tópicos utilizando $k = 5$ grupos para o Banco Bradesco.

Limite	Cartão	Dinheiro	Limite	Cartão
Crédito	Crédito	Consigo	Fazer	Consigo
Pix	Fazer	Fazer	Nada	Nada
Dinheiro	Vou	Limite	Vou	Pix
Nunca	Agora	Dias	Senha	Crédito
Agora	Nunca	Cartao	Credito	Saldo
Fatura	Saldo	Nada	Cartão	Simplemente
Valor	Não	E-mail	Nunca	Problema
Acessar	Consigo	Usar	Ter	Fica
Cliente	Fica	Péssimo	Dia	Sempre
Tudo	Dinheiro	Vai	Hoje	Ter

Tabela 4: Palavras mais frequentes determinadas pela modelagem de tópicos utilizando $k = 5$ grupos para o Banco Nubank.

3.4. POS Tagging

Dentro do contexto das análises sintáticas, utilizou-se o POS tagging para identificar a classe gramatical dos termos que compõem os documentos do corpus.

Visto que as técnicas de análise sintática levam em consideração o contexto em que os termos estão inseridos, o dataset utilizado nesta etapa não passou pelo

processo de limpeza e padronização de dados listadas na subseção b de processamento dos dados.

Os pacotes **text2vec** e **udpipe** disponíveis na linguagem de programação R foram utilizados para categorizar e filtrar os tokens em substantivos e adjetivos, uma vez que essas classes gramaticais tendem a estar mais relacionadas à funcionalidades e características do que outras classes como, pronomes, advérbios, conjunções, entre outros.

As nuvens de palavras representadas nas figuras abaixo foram geradas a partir dos 50 termos mais frequentes classificados como substantivos e adjetivos para cada um dos bancos.

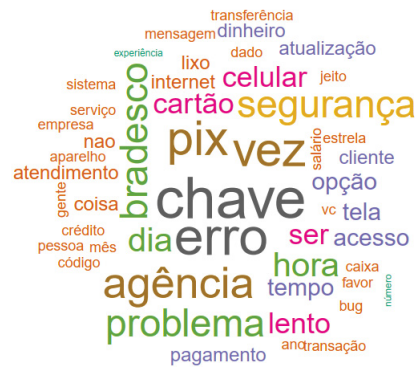


Figura 6: Nuvem de palavras para os substantivos identificados com a aplicação do pos-tagging para o Banco Bradesco.



Figura 7: Nuvem de palavras para os adjetivos identificados com a aplicação do pos-tagging para o Banco Bradesco.

Analisando os resultados, nota-se que o algoritmo não possui 100% de acurácia, uma vez que palavras como "aplicativo", "nubank", "app" e "bradesco" são classificadas hora como adjetivo hora como substantivo. Todavia, a identificação das categorias léxicas de adjetivos e substantivos realizados com a aplicação do método, permite a seleção dos termos que refletem funcionalidades e características dos aplicativos e, con-

Termo	Distância
nova	0.66345
nova_atualizacao	0.66741
pode	0.68272
sistema	0.70751
precisa	0.70832
atendimento	0.71027
falha	0.71871
poderia	0.71982
score	0.72607
dessa	0.72843
pessima	0.73101
clientes	0.73235
comprar	0.738
forma	0.74072

Tabela 6: Palavras com maior proximidade da palavra “segurança” no espaço vetor criado com uso do word embedding para o Banco Nubank.

4. Conclusão

Este artigo explorou a utilização de técnicas de mineração de texto com o objetivo de identificar quais as funcionalidades que mais impactam a experiência dos usuários de aplicativos bancários. Os métodos selecionados são baseados em aprendizado de máquina e análise sintática.

Os resultados da aplicação da modelagem de tópicos não foram satisfatórios pois os tópicos identificados eram muito similares entre si e não permitiram identificar as principais funcionalidades mencionadas pelos usuários.

Por outro lado, as aplicações das técnicas de análise sintática retornaram resultados mais satisfatórios. Com o POS tagging foi possível identificar que as principais funcionalidades mencionadas pelos usuários são relacionadas à eficiência, facilidade de utilização, segurança e produtos ofertados pelos bancos.

Por fim, a utilização do word embeddings possibilitou a identificação do contexto nos quais os termos identificados a partir do POS tagging estavam alocados. Comparando os resultados obtidos para as duas instituições financeiras utilizadas nesta última etapa, pode-se identificar que os reviews do banco Bradesco estão mais relacionados à termos que refletem limitações técnicas do aplicativo, já os reviews do banco Nubank apesar de possuírem uma volumetria menor de referências a limitações técnicas como lentidão, etc. acabam tendo mais menções aos serviços do banco,

como liberação de crédito, entre outras operações financeiras.

Agradecimentos

Gostaria de agradecer ao meu orientador Walmes Marques Zeviani e a todos os professores do curso DSBD - UFPR pelo suporte no desenvolvimento deste projeto.

Referências

- [1] Federação Brasileira de Bancos. Pesquisa Febraban de Tecnologia Bancária 2023 - Volume 2. *Technical report*, FEBRABAN, 2023. Retirado de <https://portal.febraban.org.br/pagina/3106/48/pt-br/pesquisa>
- [2] Ristova Maglovska, C. *What do hotel guests really want? An analysis of online reviews using text mining - Hotel and Tourism Management*, 2020, Vol. 8, No. 1: 37-48.
- [3] Gomes C., Viegas F., Luiz W., Rocha L. CluWords: Explorando Clusters Semânticos entre Palavras para Aprimorar Modelagem de Tópicos. DCOMP/UFESJ - São João del-Rei, MG, Brasil.
- [4] Blei, D.; Ng, A.; Jordan, M. *Latent dirichlet allocation*. *Adv. Neural Inf. Process. Syst.* 2001, 14, 601–608.
- [5] Amorim A., Murrugarra-Llerena N., Silva V., Oliveira D., Paes Al. Modelagem de Tópicos em Textos Curtos: uma Avaliação Experimental (2022).
- [6] Cutting, D., Kupiec, J., Pedersen, J. and Sibun, P. *A practical part-of-speech tagger*. In: *Proceedings of the Third Conference on Applied Natural Language Processing*, pages 133–140. Association for Computational Linguistics, Trento (1992).
- [7] Wijffels, J. Udpipes: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the 'Udpipes' 'Nlp' Toolkit (2021).
- [8] Straka M., Straková J. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. Charles University.
- [9] Indurkha, N. and Damerou, F. J. *Handbook of natural language processing*. Chapman and Hall/CRC (2010).
- [10] Pennington J, Socher R, Manning CD. *Glove: Global Vectors for Word Representation*. In *EMNLP*. 2014; 14:1532-1543.
- [11] Naili M., Chaibi H. A., Ghezala B. H. H. Comparative study of word embedding methods in topic segmentation. University of Mannouba (2010).

Apêndice 1

Termo	Distância
chave_seguranca	0.14571
seguranca	0.14977
cadastrar	0.35753
cadastrar_chave	0.36822
criar	0.4418
ativar_chave	0.48256
nova	0.49841
ativar	0.49958
consigo	0.56818
chave_pix	0.5884

Tabela 7: Palavras com maior proximidade da palavra “chave” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
chave_pix	0.14526
cadastrar	0.27212
cadastrar_chave	0.36289
fecha	0.51212
sozinho	0.61711
pix	0.6208
area_pix	0.66989
tento	0.67292
fazer_pix	0.67919
area	0.709

Tabela 8: Palavras com maior proximidade da palavra “chave” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
fazer	0.30978
fazer_pix	0.35116
transferencia	0.40336
chave_pix	0.46975
consigo_fazer	0.52283
pagamento	0.5511
consigo	0.59794
chave	0.61099
erro	0.61574
usar	0.63702

Tabela 9: Palavras com maior proximidade da palavra “pix” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
fazer_pix	0.27869
fazer	0.34064
transferencia	0.39853
saldo	0.51456
pagamento	0.54164
chave_pix	0.59014
fecha	0.59388
area	0.59776
via	0.61409
sumiu	0.61472

Tabela 10: Palavras com maior proximidade da palavra “pix” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
credito	0.25244
cartao_credito	0.34278
chegou	0.48105
debito	0.49582
fatura	0.54413
consegui	0.61807
nunca	0.66879
quero	0.67048
cancelar	0.68785
pedir	0.70184

Tabela 11: Palavras com maior proximidade da palavra “cartão” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
cartao_credito	0.21553
credito	0.23909
limite	0.42605
limite_cartao	0.49647
uso	0.49945
nunca	0.52924
fatura	0.53218
anos	0.55697
debito	0.5836
uso_cartao	0.59276

Tabela 12: Palavras com maior proximidade da palavra “cartão” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
ultima_atualizacao	0.26258
ultima	0.37564
ficou	0.47091
nova	0.62476
apos	0.64472
cada	0.68895
urgente	0.69237
alguma	0.72243
ainda	0.73522
simplesmente	0.74385

Tabela 13: Palavras com maior proximidade da palavra “atualização” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
ultima_atualizacao	0.31883
nova	0.36425
ultima	0.36549
nova_atualizacao	0.45992
ficou	0.55667
dessa	0.5662
horrivel	0.60839
apos	0.61917
pior	0.69841
fica	0.70151

Tabela 14: Palavras com maior proximidade da palavra “atualização” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
ver	0.32858
extrato	0.58505
indisponivel	0.65129
falha	0.66326
ano	0.70417
sistema	0.70672
pago	0.72666
aparecendo	0.73038
apresenta	0.73072
mostra	0.73548
tirar	0.73618
dando_erro	0.74474
fazendo	0.74716
facio	0.74861
vem	0.75281
serve	0.76164
saber	0.76452

Tabela 15: Palavras com maior proximidade da palavra “saldo” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
ver	0.31021
ver_saldo	0.36508
aparece	0.43427
pix	0.51456
consigo	0.52265
sumiu	0.54317
opcao	0.56245
mostra	0.57385
aparecendo	0.57575
transferencia	0.5936
fazer	0.60354
fazer_pix	0.6176
consigo_ver	0.62054
simplesmente	0.66082
consigo_fazer	0.69007
conseguindo	0.69223
fazer_transferencia	0.69621

Tabela 16: Palavras com maior proximidade da palavra “saldo” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
tirar	0.61744
usar	0.63016
pagar	0.66745
proprio	0.67662
complicado	0.67878
posso	0.68555
transferir	0.69087
vou	0.71497
acontecendo	0.71847
absurdo	0.73386

Tabela 17: Palavras com maior proximidade da palavra “dinheiro” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
la	0.44641
usar	0.55289
pagar	0.56489
todo	0.58253
preso	0.60972
tirar	0.62321
transferir	0.62527
agora	0.62887
sumiu	0.6429
consigo	0.65408

Tabela 18: Palavras com maior proximidade da palavra “dinheiro” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
acesso	0.60201
complicado	0.6136
assim	0.62141
horrível	0.67776
cada	0.68394
usar	0.69622
fone	0.69996
fica	0.71896
conseguir	0.74897
ainda	0.75513

Tabela 19: Palavras com maior proximidade da palavra “difícil” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
assim	0.55514
conseguir	0.62006
usar	0.67736
tendo	0.69007
ficando	0.69815
horrível	0.70573
pessimo	0.71316
utilizar	0.74287
atendimento	0.75073
limite	0.76335

Tabela 20: Palavras com maior proximidade da palavra “difícil” no espaço vetor criado com uso do word embedding para o Banco Nubank.

Termo	Distância
trava	0.41803
demais	0.43186
demora	0.51793
lento_trava	0.52699
lento	0.53043
pesado	0.53282
horrível	0.53737
e_lento	0.54199
ruim	0.57743
pessimo	0.58776

Tabela 21: Palavras com maior proximidade da palavra “lento” no espaço vetor criado com uso do word embedding para o Banco Bradesco.

Termo	Distância
cara	0.58727
demais	0.59541
horrrivel	0.67235
super	0.69106
travando	0.69624
qualquer	0.70774
bugs	0.71273
complicado	0.71641
varios	0.72615
abri	0.73078

Tabela 22: Palavras com maior proximidade da palavra “lento” no espaço vetor criado com uso do word embedding para o Banco Nubank.