



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE  
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA  
UNIDADE ACADÊMICA DE SISTEMAS E COMPUTAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

**IVYNA RAYANY SANTINO ALVES**

**ANÁLISE AUTOMÁTICA DE POSICIONAMENTO DE  
PARLAMENTARES SOBRE A REFORMA DA PREVIDÊNCIA  
POR MEIO DE SEUS DISCURSOS**

**CAMPINA GRANDE - PB**

**2019**

**IVYNA RAYANY SANTINO ALVES**

**ANÁLISE AUTOMÁTICA DE POSICIONAMENTO DE  
PARLAMENTARES SOBRE A REFORMA DA PREVIDÊNCIA  
POR MEIO DE SEUS DISCURSOS**

**Trabalho de Conclusão Curso  
apresentado ao Curso Bacharelado em  
Ciência da Computação do Centro de  
Engenharia Elétrica e Informática da  
Universidade Federal de Campina  
Grande, como requisito parcial para  
obtenção do título de Bacharel em Ciência  
da Computação.**

**Orientador: Professor Dr. Leandro Balby Marinho.**

**CAMPINA GRANDE - PB**

**2019**



A474a Alves, Ivyna Rayany Santino.  
Análise automática do posicionamento de parlamentares sobre a reforma da previdência por meio de seus discursos. / Ivyna Rayany Santino Alves. - 2019.

16 f.

Orientador: Prof. Dr. Leandro Balby Marinho.

Trabalho de Conclusão de Curso - Artigo (Curso de Bacharelado em Ciência da Computação) - Universidade Federal de Campina Grande; Centro de Engenharia Elétrica e Informática.

1. Processamento de linguagem natural. 2. Discursos políticos. 3. Parlamentares - discursos. 4. Análise automática de discursos políticos. 5. Reforma da previdência - discurso. 6. Processamento de textos. 7. Algoritmos semânticos. 8. Word embeddings. 9. Distância semântica entre palavras. 10. Semântica - processamento de textos. I. Marinho, Leandro Balby. II. Título.

CDU:004.6(045)

### **Elaboração da Ficha Catalográfica:**

Johnny Rodrigues Barbosa  
Bibliotecário-Documentalista  
CRB-15/626

**IVYNA RAYANY SANTINO ALVES**

**ANÁLISE AUTOMÁTICA DE POSICIONAMENTO DE  
PARLAMENTARES SOBRE A REFORMA DA PREVIDÊNCIA  
POR MEIO DE SEUS DISCURSOS**

**Trabalho de Conclusão Curso  
apresentado ao Curso Bacharelado em  
Ciência da Computação do Centro de  
Engenharia Elétrica e Informática da  
Universidade Federal de Campina  
Grande, como requisito parcial para  
obtenção do título de Bacharel em Ciência  
da Computação.**

**BANCA EXAMINADORA:**

**Professor Dr. Leandro Balby Marinho  
Orientador – UASC/CEEI/UFCG**

**Professora Dra. Patrícia Duarte de Lima Machado  
Examinadora – UASC/CEEI/UFCG**

**Professor Dr. Tiago Lima Massoni  
Disciplina TCC – UASC/CEEI/UFCG**

**Trabalho aprovado em: 25 de novembro de 2019.**

**CAMPINA GRANDE - PB**

# Análise automática do posicionamento de parlamentares sobre a Reforma da previdência por meio de seus discursos

Trabalho de Conclusão de Curso

Ivyna Rayany Santino Alves

Departamento de Sistemas e  
computação  
Universidade Federal de Campina  
Grande  
Campina Grande, Paraíba, Brasil

ivyna.alves@ccc.ufcg.edu.br

Leandro Balby Marinho

Departamento de Sistemas e  
computação  
Universidade Federal de Campina  
Grande  
Campina Grande, Paraíba, Brasil

lbmarinho@dsc.ufcg.edu.br

## RESUMO

Nos últimos anos, o setor político brasileiro sofreu grandes mudanças, movimentando bastante o Congresso Nacional. Sendo assim, devido ao volume de informação gerado, é difícil para população acompanhar as discussões realizadas em sessões políticas e, conseqüentemente, observar o posicionamento dos parlamentares nelas. Nessa perspectiva, utilizando dados públicos e com o auxílio de técnicas de processamento textual é possível detectar o posicionamento de cada parlamentar a partir das suas falas com a utilização de algoritmos que calcula a distância semântica entre palavras. Dessa forma, com a automatização do processo de verificação de posicionamento permitirá à população visualizar o alinhamento e a contradição de opiniões de parlamentares e de partidos sobre a Reforma da previdência.

## Palavras-chave:

Processamento de linguagem natural. Discursos. Parlamentares. Posicionamento político.

## 1. INTRODUÇÃO

Um grande desafio para se obter uma sociedade politicamente participativa é o acesso à informação. Atualmente, o Congresso Nacional possui ferramentas que possibilitam a transparência de dados sobre parlamentares, partidos, proposições, discursos, entre outros. No entanto, recolher tais informações não é uma tarefa fácil para quem não tem conhecimento técnico, o que constrói uma barreira para a exploração desse dado pela população, sobretudo, no que está sendo discutido em sessões parlamentares.

Diante disso, a maioria das informações sobre assuntos políticos são expostas em jornais, revistas e, principalmente, nos últimos anos, em redes sociais, o que torna mais rápida a compreensão, mas, por outro lado, há uma grande chance de enviar o leitor sobre determinado assunto ou até mesmo, disseminar informação falsa, mais conhecidas como *fake news*.

Sendo assim, essa divulgação de informações opinativas pode ser encontrada no ambiente digital que, segundo Turchi<sup>1</sup> (2009), é no espaço democrático da internet que existe a possibilidade de cada ser humano publicar e disseminar conteúdos, através de postagens em redes sociais, aplicativos e revistas on-line. Com isso, é válido ressaltar que “a web trouxe um poder às mãos dos consumidores nunca antes visto, dada a possibilidade de influenciar outras pessoas” a consumirem e opinarem sobre aquilo que veem (TURCHI, 2009, p. 1).

Diante disso, segundo Rosa (2015), a disseminação de conteúdo é intensa nas redes sociais e os textos podem refletir discursos alinhados ou divergentes, seja expressando opiniões favoráveis, desfavoráveis ou neutros, isto é, “comentários positivos refletem a aceitação e a aprovação sobre um tema em discussão e os comentários negativos à reprovação ou à desconformidade com o tema” (ROSA, 2015, p. 1.)

Nesse contexto, podemos dizer que o campo de Processamento de Linguagem Natural (PLN) tem se tornado uma área de grande relevância tanto em aspectos teóricos quanto práticos. Isso pode ser comprovado “em artigos acadêmicos apresentados em conferências acadêmicas que tinham como foco, mais precisamente, computação e linguística, em centros de estudos associados ao processamento da linguagem natural” (SILVA et al., 2017, p. 1). Essa necessidade de estudo ocorre em virtude da automatização de procedimentos de análise de textos que é um grande desafio e tem um grande valor técnico. Essa repercussão pode ser exemplificada no contexto político, que é o ponto de partida dessa investigação, pois é a partir dessa técnica que será detectado automaticamente o posicionamento de parlamentares e de partidos. Tal ação acontecerá através das falas

---

<sup>1</sup> Os autores retêm os direitos, ao abrigo de uma licença Creative Commons Atribuição CC BY, sobre todo o conteúdo deste artigo (incluindo todos os elementos que possam conter, tais como figuras, desenhos, tabelas), bem como sobre todos os materiais produzidos pelos autores que estejam relacionados ao trabalho relatado e que estejam referenciados no artigo (tais como códigos fonte e bases de dados). Essa licença permite que outros distribuam, adaptem e evoluam seu trabalho, mesmo comercialmente, desde que os autores sejam creditados pela criação original.

coletadas da PEC 06/2019, mais conhecida como Reforma da previdência, auxiliando, assim, cientistas políticos e a sociedade civil na investigação e no acompanhamento do que é discutido no Congresso Nacional.

Diante disso, a interação entre homem e máquina tem se tornado cada vez mais necessária e mais sofisticada. Segundo Rodrigues (2017, p.1):

com o passar dos anos, houve muitas pesquisas e desenvolvimentos nos mais diversos ramos de PLN, destacando-se a tradução automática, considerada pela maioria como o marco inicial na utilização dos computadores para o estudo das línguas naturais.

Essa técnica tem facilitado as soluções computacionais para tais problemas envolvidos nessa temática.

Dessa forma, pode-se afirmar que mediante à ambiguidade inerente à linguagem natural humana e, tomando como objeto de estudo o contexto político brasileiro, através das discussões realizadas no Congresso Nacional, a detecção de posicionamentos de discursos dos parlamentares é subjetiva, o que se faz necessária a objetivação destas informações para maior clareza na sua divulgação. Essa problemática norteou-se por meio da seguinte problematização: Que posicionamento político, diante de um projeto de lei, pode-se observar nos discursos de parlamentares e de partidos políticos a partir do processamento textual? Para responder a esse questionamento, é desejável recorrer a uma solução automatizada, mediante a obtenção do *corpus*, que como caso de estudo da pesquisa, será utilizada as discussões referentes a PEC 06/2019 - Reforma da previdência. Após esse processo, ocorreram a sistematização dos dados, a identificação e a análise dos resultados obtidos de falas, tendo como base modelos de aprendizagem para reconhecer posicionamentos diante dos discursos de deputados sobre a PEC 06/2019. Essa pesquisa cumpre uma função social, pois tem a finalidade de tornar a informação viável e precisa para a sociedade, a partir da disponibilização dos resultados e sua possível aplicação em plataformas de monitoramento político, como a vertente dos sites: [Parlametria](#), [Voz Ativa](#) e [Leggo](#).

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1 Funcionamento das discussões em sessões políticas brasileiras

No Brasil, as sessões políticas são realizadas no Congresso Nacional, mais especificamente na Câmara de deputados e no Senado Federal, com o objetivo de promover discussões sobre pautas para a elaboração e modificações da legislação do país.

O ambiente de discussões é conduzido pelos parlamentares eleitos, embasados na apresentação de documentos, por exemplo, emendas, requerimentos, apresentação de parecer, entre outros, e no estudo de sua viabilidade. Sendo assim, permite-se o compartilhamento de opiniões sobre determinada pauta, a fim de elaborar mudanças que abrangem aspectos econômicos, sociais e de infraestrutura no Brasil.

Diante disso, com base nas ideologias partidárias, os deputados e senadores tomam decisões sobre votações e projetos de leis em reuniões do tipo comissões e do tipo plenário.

#### 2.1.1 Discursos em sessões políticas

Os parlamentares elaboram em conjunto projetos de leis no Congresso Nacional, mas para que esses projetos façam parte da legislação brasileira, o processo inicial é dispor de um ambiente de discussões para estudar a viabilidade da proposta e de exposição do pensamento. Sendo assim, segundo Ribeiro e Mendes (2019),

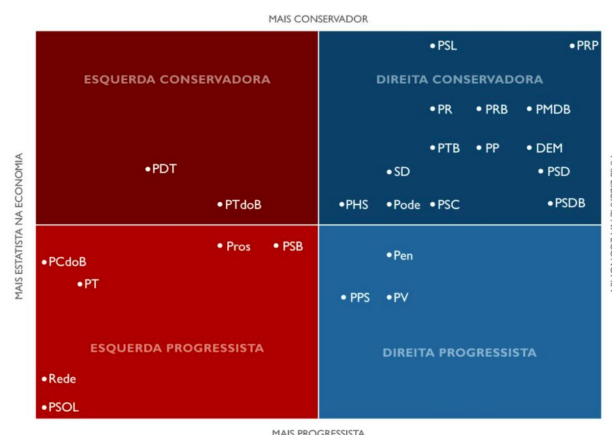
discursos são uma valiosa fonte para a estimação do posicionamento ideológico dos partidos por expressarem o conteúdo substantivo do conflito político. Além disso, as pessoas tendem a se relacionar com aqueles que compartilham as mesmas ideias e opiniões.

Por isso, cada parlamentar escolhe a que partido fazer parte, seguindo as ideologias partidárias que será explicado melhor no próximo tópico.

#### 2.1.2 Ideologias partidárias

O partido político é formado por parlamentares cujos posicionamentos são ancorados em uma ideologia política, que ora é de direita, ora é esquerda, ora é centro.

Segundo Couto (2018), no contexto brasileiro, partidos com ideologia de direita possuem ideais liberais ou neoliberais e tem como principais abordagens a discussão sobre aspectos econômicos, a exemplo do Partido Social Liberal (PSL). Já os partidos de esquerda, geralmente, discutem aspectos políticos-sociais, com o objetivo de promover a maior igualdade social, como por exemplo, o Partido Socialismo e Liberdade (PSOL). Por fim, os partidos de centro oscilam em discutir aspectos econômicos liberais e sociais, como o Movimento Democrático Brasileiro (MDB). A figura, a seguir, demonstra um panorama das ideologias partidárias no Brasil.



**Figura 1: Espectro de ideologia partidária entre partidos brasileiros. Fonte: [Direita ou esquerda? Análise de votações indica posição de partidos brasileiros no espectro ideológico.](#)**

## 2.2 Processamento de linguagem natural

Diante da quantidade de dados dos discursos políticos, podemos dizer que reconhecer padrões em meio a tanta informação é tarefa inviável para o ser humano desenvolver um estudo científico e “mergulhar” nos textos de maneira exploratória sem utilizar nenhuma ferramenta para auxílio de tal atividade. Segundo Andrade (2017, p.1):

processamento de linguagem natural (PNL) é uma área de conhecimento e de pesquisa que facilita essa interação, consiste em uma subárea da inteligência artificial, que desenvolve melhor a forma de interpretação da linguagem humana em diferentes dispositivos.

Dessa forma, a utilização do poder computacional através de PLN torna o processo mais prático e automático para identificar os discursos entre parlamentares que ora são alinhados à ideologia partidária, ora são contraditórios, apresentados nas sessões políticas.

## 2.3 Word Embeddings

A técnica de *word embeddings* é capaz de aprender o contexto semântico e sintático de um texto, sendo que a ideia é transformar cada palavra, ou *token*, do nosso conjunto de frases, denominado de *corpus*, em um vetor numérico que a representa semanticamente. Dessa maneira, segundo Carvalho (2018, p. 9), pode-se dizer que são métodos que fornecem boas representações vetoriais contínuas de baixa dimensão para conjunto de textos não estruturados. Esses vetores são distribuídos, densos, contínuos, contêm tamanho fixo e representam uma palavra, por isso, são muito úteis, pois possuem um forte poder de generalização. Sob essa perspectiva, essas representações contribuem para melhorar o desempenho em diversos métodos e tarefas de PLN. A seguir pode-se observar a construção das *word embeddings*.

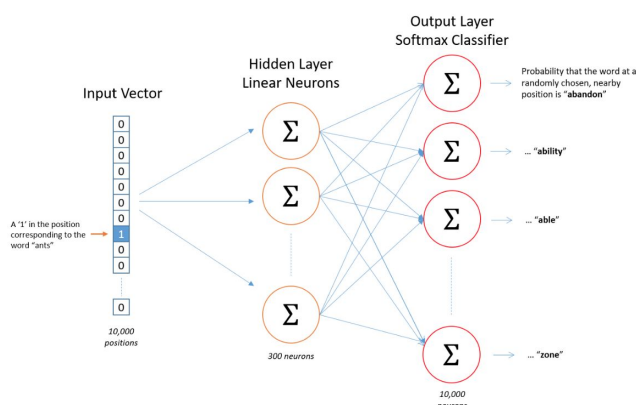


Figura 2: Construção de word embeddings. Fonte: [A simple Word2vec tutorial](#).

Para a figura acima, há a exemplificação de como é a arquitetura da rede neural do modelo *word2vec*, justificado por Ali (2019), que o aprendizado é produzido com vetores de

palavras com o valor de 300 *features* e uma matriz de peso com 10.000 linhas. Para facilitar o entendimento, pode-se pensar da seguinte maneira e, tomando como base da Figura 2. Uma palavra é passada para uma rede neural com uma camada escondida por meio de *one-hot encoding* (técnica que transforma categorias de um *dataframe* em colunas com valores 0 para valores negativos, ou 1 para valores afirmativos). A saída são as próximas palavras que podem aparecer depois da palavra de entrada. A camada intermediária nesse caso seria o *word embedding* da palavra de entrada, que nesse caso, captura o contexto dessa palavra em relação às outras do vocabulário.

Esse procedimento é organizado, a partir de um modelo de aprendizagem, o que torna possível, diante de um contexto comunicativo, reconhecer palavras ou tokens que, semanticamente, são semelhantes. Essa técnica é viável, nesta pesquisa, por meio da utilização de *Word2vec*, cujo funcionamento será detalhado, no próximo tópico.

### 2.3.1 Word2Vec

O *word2vec* é uma das técnicas para o aprendizado de *word embeddings*, que tem por natureza agrupar os vetores de palavras semelhantes no espaço de vetores formados pelo algoritmo e, com isso, detectar semelhanças matemáticas entre eles.

Como exemplo clássico, podemos identificar que a operação a seguir entre os vetores resultantes dos seguintes *tokens*, “rei”, “homem”, “mulher” resultam em “rainha”, ou seja, é capaz de capturar relações entre as palavras como: “homem” está para “mulher” assim como “rei” está para “rainha”. A seguir, a figura demonstra o espaço vetorial, representado na equação somatória entre *word embeddings* dos tokens.

$$\vec{v}(\text{"rei"}) - \vec{v}(\text{"homem"}) + \vec{v}(\text{"mulher"}) = \vec{v}(\text{"rainha"})$$

Equação 1: somatório entre vetores de tokens.

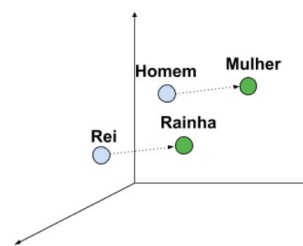
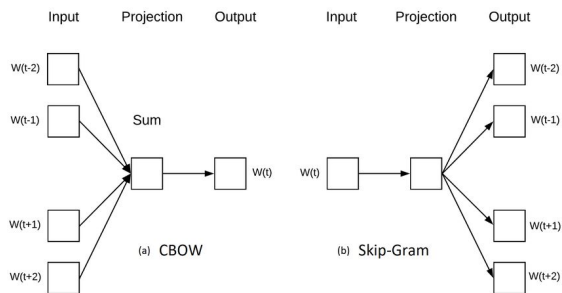


Figura 3: Representação bidimensional simplificada para exemplificar a relação capturada entre palavras. Fonte: [Semelhança entre títulos de produtos com word2vec](#).

Segundo Contrastes (2018), existem dois algoritmos para realizar o treinamento dos dados, sendo elas: CBOW, que prevê qual a palavra que estamos buscando, a partir de um determinado contexto, bem como o Skip-gram que tem como ponto de partida uma determinada palavra, para apontar o contexto que o vocábulo pertence, cujo objetivo é prever o

contexto pelo qual esta palavra veio. Na Figura 4, é possível observar como é o comportamento de cada algoritmo.

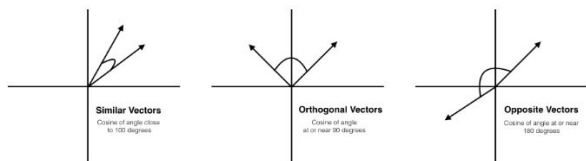


**Figura 4: Representação dos algoritmos CBOV e Skip-Gram de treinamento dos dados.** Fonte: [Similaridade entre títulos de produtos com word2vec.](#)

Após esses procedimentos, é necessário fazer a comparação entre os tokens e, para isso, utilizamos a similaridade dos cossenos. Sabendo que o cosseno de um ângulo  $a$  pode variar entre -1 e 1, quando o ângulo formado entre os vetores se aproxima de  $0^\circ$ , o cosseno se aproxima de 1. Já quando são ortogonais, ou seja, ângulo de  $90^\circ$ , o cosseno é igual a 0. Por fim, quando são opostos, o cosseno se aproxima de -1, isto é, quanto mais próximo o cosseno for de 1, mais semelhante, caso contrário, quanto mais próximo de -1, mais oposto. Tal abordagem matemática está exemplificada, na figura, a seguir:

$$\cos(a, b) = \frac{\langle a, b \rangle}{\|a\| \times \|b\|}$$

**Equação 2: cosseno do ângulo entre os vetores a e b.**



**Figura 5: Representação da similaridade entre cossenos.** Fonte: [Similaridade entre títulos de produtos com word2vec.](#)

A aplicação do *word2vec* resultará em um modelo composto por diversos vetores, normalmente centenas de dimensões, no entanto, para o campo de visão humana, é difícil detectar algum padrão em objetos multidimensionais e, por isso, existe dificuldade para visualizá-los.

Nesse contexto, utilizamos o algoritmo T-SNE, introduzido por Laurens van der Maaten e Geoffrey Hinton em 2008, que através da redução de dimensionalidade, a partir de um conjunto de pontos em um espaço multidimensional, encontra uma representação fiel desses pontos em um espaço de dimensão menor de forma não-linear e se adaptando aos dados, realizando diferentes transformações em diversas regiões desse espaço. Diante disso, o algoritmo é capaz de capturar em grande parte a estrutura local do espaço multidimensional, mas também de revelar a estrutura global do banco de dados como a presença de

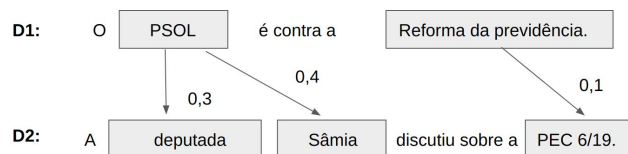
*clusters*, ou seja, agrupar os dados que possuem padrões semelhantes.

Em conjunto com o algoritmo anterior, é aplicada a técnica de Análise de Componentes Principais (ACP), a qual extrai *features* lineares e realiza, assim, um mapeamento linear dos dados para um espaço de menor dimensão, de maneira que a variação dos dados na representação de baixa dimensão seja maximizada, a partir do cálculo dos vetores próprios da matriz de covariância. Outro algoritmo que utiliza esse modelo para verificar a distância e proximidade semântica dos discursos é Word Mover's Distance que será apresentado detalhadamente, no próximo tópico.

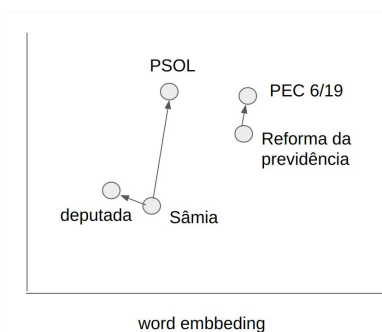
## 2.4 Word Mover's Distance

Diante da utilização da técnica *word2vec* que incorpora o relacionamento semântico entre palavras através da criação de vetores, aplicando o algoritmo *word mover's distance* (*wmd*) para detectar os discursos similares. Segundo Kusner et al. (2015, p. 1), para medir a semelhança semântica e sintática entre esses documentos, tem-se como base o cálculo da distância euclidiana entre dois pontos de vetores. Sendo assim, tal técnica auxilia no entendimento de proximidade entre textos, a partir dos resultados que o algoritmo traz.

As figuras, a seguir, exemplificam o funcionamento da técnica.



**Figura 6: Distância de palavras presentes nos discurso 1 e discurso 2.**



**Figura 7: Representação de proximidade entre palavras do discurso 1 e discurso 2 representados na Figura 6.**

A Figura 6 demonstra o funcionamento do *wmd* a partir dos discursos 1 e 2, sendo eles exemplificados como D1 e D2, respectivamente. Diante deles, foram observados as distâncias entre as palavras 0,3 e 0,4, por exemplo, a palavra PSOL, Sâmia e deputada estão mais próximos devido ao contexto de partido a qual estão inseridos, pois Sâmia é uma deputada do partido PSOL. Já para as palavras PEC 6/19 e Reforma da previdência, a



distância é menor, pelo fato de se tratarem do mesmo “sujeito”, já que se trata de nome formal e de apelido do projeto de lei.

Como continuação da Figura 6, a Figura 7 traz a representação gráfica das distâncias entre as palavras abordadas nos discursos 1 e 2 e as ligações entre suas respectivas abordagens de contexto na qual estão inseridas.

### 3. METODOLOGIA

#### 3.1 Natureza e análise exploratória

A natureza da pesquisa é quali-quantitativa, uma vez que ocorre dois movimentos tanto de quantificar os discursos quanto de interpretá-los. Essa ação ocorre a partir da coleta de dados de discursos referente à PEC 06/2019 - Reforma da previdência e partir deles, em que foi desenvolvida a identificação de posicionamento dos parlamentares, através dos parâmetros de alinhamento e contradição partidária.

A partir do *corpus* coletado, foi realizada a exploração dos dados com o objetivo de visualizar o comportamento e ter uma visão geral dos discursos. Para isso, teve-se como embasamento a seguinte questão: que partidos foram mais ou menos recorrentes nas sessões? Vale ressaltar que será utilizada a linguagem R e a biblioteca ggplot para gerar as visualizações para esses questionamentos.

Além disso, observou-se que palavras são mais citadas em cada comissão e no plenário e se elas são relevantes posteriormente para o modelo. Dessa maneira, utilizamos a linguagem Python3.7 para gerar a nuvem de palavras dessas reuniões, a partir da retirada das *stopwords* (palavras que se repetem muito no vocabulário e que não atribuem valor ao contexto, por exemplo, preposições e artigos) e do algoritmo de *word cloud* que, por sua vez, gera uma visualização com as palavras mais utilizadas na base de dados, os discursos da previdência. Essa visualização demonstra que o tamanho da palavra é proporcional a quantidade de vezes que ela aparece, ou seja, quanto mais a palavra está presente nos discursos, maior ela fica na visualização.

#### 3.2 Contexto de geração de dados

Atualmente, a Câmara de deputados disponibiliza na *API* dos dados abertos informações sobre projetos de leis, parlamentares, partidos, votações, discussões e etc. com o objetivo de expor para população os acontecimentos no Congresso Nacional de forma transparente. No entanto, foi percebido que o dado de discursos de deputados em sessões políticas não é atualizado com frequência e as notas taquigráficas disponibilizadas por sessão não é exposta de forma clara. Portanto, é necessário ter o conhecimento da plataforma para encontrá-las, dificultando, assim, a recuperação e utilização dessas informações para análise e exploração do dado.

Diante dessas necessidades, foi desenvolvido um [raspador de discursos](#), uma ferramenta que extrai textos de pdfs, com o intuito de facilitar a obtenção das falas dos parlamentares, motivado, principalmente, pelo estudo de algoritmos que detectam de forma automática o alinhamento e contradição partidária nos

discursos coletados pela ferramenta utilizando técnicas de coleta de dados que será mencionado na próxima seção.

#### 3.3 Instrumentos de coleta de dados

Para detectar posicionamento dos parlamentares em discursos, antes de tudo, foi realizado a busca por pdfs com os diálogos no site dos dados abertos da Câmara de deputados correspondente a proposição, que teve como escolha para objeto de estudo a PEC 06/2019.

O motivo de ter escolhido a PEC 06/2019, como caso de estudo, foi a repercussão midiática gerada, o envolvimento dos parlamentares na apresentação de documentos e no grande volume de eventos realizados no Congresso causados desde o início de 2019. Tais métricas, foram observadas na plataforma digital Leggo, que por sua vez, permite o acompanhamento de progresso e dos eventos relacionados a um projeto de lei em discussão. Pensando nisso, utilizar os discursos desse projeto de lei, permite observar o comportamento dos parlamentares, através de seus discursos e visualizar o posicionamento de cada um.

De posse dos documentos, o primeiro passo para deixar o dado no formato de tabelas para análises foi extrair os textos dos pdfs e padronizá-los de arquivo de texto para tabelas, a partir de scripts na linguagem de programação Python 3.7. Ao final deste processo, o arquivo contém informações de nome, de partido, de unidade federativa e de texto do discurso do deputado. Essa organização será exemplificada na Tabela 1, a seguir:

**Tabela 1: Descrição do dado de discursos da PEC 06/2019**

Coluna da base de dados	Descrição da coluna
data	Data que o discurso foi falado
autor	Nome do parlamentar
partido	Partido do parlamentar
uf	Estado do parlamentar
discurso	Discurso taquigrafado do parlamentar na sessão política

Além dessas variáveis, foram utilizadas as recomendações de Sales, Balby e Veloso (2019) para treinar o modelo com base de notícias, e para coletar tais dados, foi utilizado um *crawler* dados referentes à política em jornais on-line, como Folha e Estadão do ano de 2014 até 2018, com objetivo de criar uma base de dados de grande volume em seu vocabulário para treinar o modelo *word2vec* descrito no tópico 3.4.

**Tabela 2: Descrição dos dados de notícias das plataformas Folha e Estadão**

Coluna da base de dados	Descrição da coluna
Subfonte	Nome da plataforma de notícias
Título	Título da notícia
Conteúdo	Texto da notícia
URL	Link para notícia
id_notícia	Identificador da notícia
Caderno	Coluna que a notícia foi retirada
Ano	Ano de publicação da notícia

### 3.4 Treinamento

Para a aplicação do wmd, foi realizado o processo de treinamento do modelo de aprendizagem de máquina com os dados de notícias das plataformas Folha e Estadão referentes aos anos 2014 e 2018 da coluna de política, no formato disposto na Tabela 2. Essa sistematização ocorreu com o intuito de que o modelo aprenda, com um grande volume de dados, a se comportar de maneira eficiente e descobrir o posicionamento de deputados posteriormente. No próximo tópico, discussões sobre os mecanismos utilizados serão postos em cena.

#### 3.4.1 Word2Vec, ACP e T-SNE

Com o intuito de criar o modelo de aprendizagem, foram utilizadas técnicas de processamento de linguagem natural, através da biblioteca Gensim da linguagem Python versão 3.7 que utilizou o algoritmo *word2vec* e como entrada, a base de dados das plataformas on-line de notícias descrita na Tabela 2. Como resultado, teremos um modelo formado por vetores de 300 dimensões e para visualizar os resultados é utilizado algoritmos de agrupamento e redução de dimensionalidade como ACP e T-SNE.

A partir dos resultados gerados pelo ACP e T-SNE, foi realizado o processo de validação do modelo. Esta etapa foi envolvido o bom senso humano e conhecimento de causa política brasileira. Tal processo está ligado à observação das palavras geradas como resultado dos algoritmos na visualização e, a verificação do sentido empregado ao contexto, como saber comportamentos de partidos e parlamentares, e constatar que a base de dados de notícias foi suficiente para gerar vetores que fazem sentido no contexto político explorado. Pode-se contar com o auxílio do algoritmo de similaridade de palavras, o *most similar* da biblioteca Gensim que irá retornar as palavras mais próximas a que foi utilizada na pesquisa.

**Tabela 3: Descrição dos parâmetros resultantes do algoritmo most similar da biblioteca Gensim**

Coluna resultantes	Descrição das colunas
Palavra	Palavra similar a passada como parâmetro
Proximidade com o parâmetro	Índice de similaridade entre a palavra resultado e o parâmetro pesquisado.

#### 3.4.2 Sistematização dos dados

Diante da corretude do modelo, foram selecionados discursos favoráveis e contrários a pauta da Reforma da previdência. Para esse procedimento, foi utilizado também o bom senso humano para detectar se a fala era a favor ou contra, selecionando-as do arquivo gerado pelo raspador de discursos. Sendo assim, os discursos escolhidos se encontram no seguinte formato:

**Tabela 4: Descrição dos dados utilizados para análise de posicionamento político**

Coluna da base de dados	Descrição da coluna
id_discurso	Identificador do discurso na base de dados
autor	Nome do parlamentar
partido	Partido do parlamentar
uf	Estado do parlamentar
discurso	Discurso taquigrafado do parlamentar na sessão política

Diante das informações da Tabela 4, foram escolhidos os seguintes parlamentares e seus respectivos discursos:

- Erika Kokay (PT - DF):** “Sr. Presidente, o PT quer demonstrar alinhamento à proposta apresentada. Nós estamos lidando com um tema que vai mexer com um número por volta de 100 milhões de brasileiros e brasileiras, inclusive com a população de baixa renda, que pode não ter mais o direito à aposentadoria.”
- Joice Hasselmann (PSL - SP):** “Presidente, boa tarde ao senhor; boa tarde a todos. Não vou usar os 10 minutos, não, Presidente, vou usar apenas 1 minuto, para ser muito breve e para acelerarmos esta discussão. É de suma importância que nós possamos entender, para que possamos passar a informação verdadeira às pessoas que estão nos acompanhando, de uma maneira muito simples: se a nova Previdência não for aprovada, o Brasil quebra. Essa Reunião de: 16/04/2019 é a argumentação básica, não precisa de mais nada. Só esta está boa: se não aprovarmos a nova Previdência, o

Brasil quebra, para, e acabou, não há mais o que fazer! Para se ter uma ideia, o PIB deste ano ficará abaixo de 1% se nós não tivermos a aprovação desse texto, e o desemprego chegará a 15% em 2023. Isso seria catastrófico, a maior taxa da história! Os Estados estão quebrados e, sem a Nova Previdência, não terão nem como pagar os salários dos servidores. É simples assim! A quebradeira será geral, a começar por aqueles que já estão com a corda no pescoço, como Rio de Janeiro, Rio Grande do Norte, Rio Grande do Sul e Minas Gerais. É uma questão de responsabilidade, Sr. Presidente. Quem de fato quer resolver as questões do nosso País, quem de fato quer cuidar dos mais pobres, quem de fato quer gerar emprego e renda e colocar investimentos neste País vota pela Nova Previdência. Muito obrigada. Vamos em frente!”

O critério de escolha dos discursos acima foi a partir do envolvimento dos partidos nas sessões e no conteúdo ideológico manifestado linguisticamente, pois a seleção vocabular mostra opiniões do parlamentar ou partido, se o discurso é a favor ou contra ao projeto de lei da Reforma da Previdência.

Portanto, com as falas selecionadas, a detecção de alinhamento e contradição partidária foi realizada através do algoritmo *word mover's distance* discutido melhor na próxima seção.

### 3.4.3 Categorias de análise: alinhamento e contradição partidária

Como caso de estudo, foram utilizadas as discussões sobre a Reforma da previdência e explicando um pouco mais sobre a pauta, pode-se dizer que essa proposta tem como maior impacto o setor econômico, logo partidos de direita apoiam a reforma e, como consequência, também há o impacto negativo para aspectos sociais, partidos de esquerda votaram contra a reforma. Dessa maneira, para a pesquisa, parlamentares que falam contrário ao que o partido prega é um caso de contradição partidária e, o alinhamento é dado de forma análoga, o parlamentar fala de forma similar a ideologia do partido. Por exemplo, parlamentares de partidos de esquerda, que falam em seus discursos que são a favor da reforma, estão em contradição partidária já para o alinhamento, seria se um parlamentar de um partido de esquerda fosse contra a reforma.

Diante disso e a partir dos resultados do algoritmo *wmd*, foram observados os três discursos mais similares. No caso do discurso a favor da reforma passado como parâmetro, terá como resultado três discursos similares a ele e indicando o mesmo posicionamento, assim como o caso contrário, para discursos contra a reforma, terá mesmo comportamento só que de forma análoga.

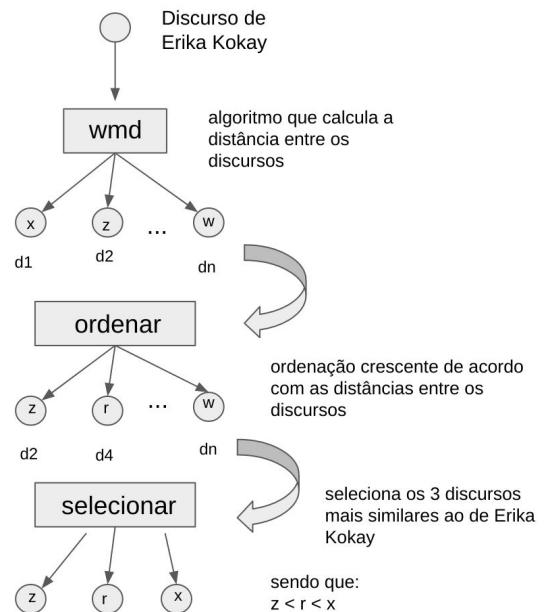


Figura 8: Funcionamento do cálculo de distância entre discursos.

Portanto, a partir dos resultados apresentados pelo procedimento acima, foi analisado o conteúdo dos discursos, o nome e o partido do parlamentar, a fim de observar o alinhamento ou contradição partidária de acordo com os conhecimentos de ideologia entre partidos.

## 4. RESULTADOS

Antes de gerar os resultados do modelo e aplicar o algoritmo de *wmd*, foi necessário conhecer os dados dos discursos da reforma da previdência, dessa forma foi gerado em visualizações por meio de *word cloud* e como resultado são as figuras, a seguir:

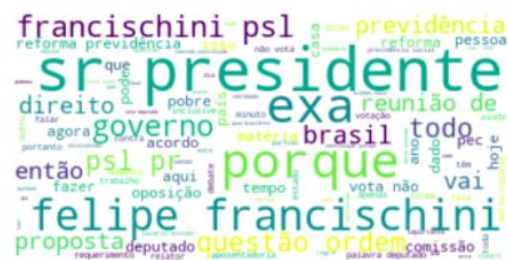
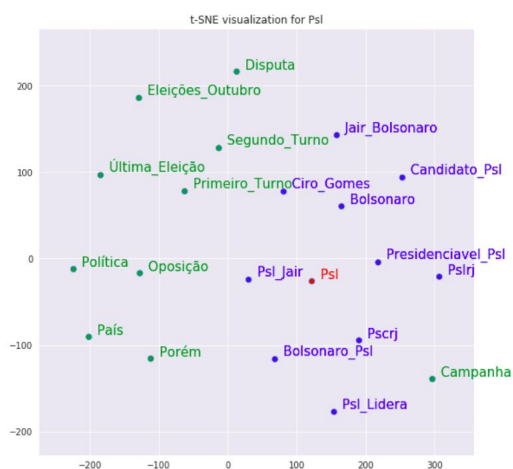


Figura 9: Palavras mais citadas nos discursos da reunião da Comissão de Constituição e Justiça e de Cidadania.





**Figura 15: Aplicação dos algoritmos T-SNE e ACP para a palavra PSL no contexto de eleição.**

Nas Figuras 14 e 15, mostram as palavras mais parecidas com os parâmetros pesquisados, em vermelho, sendo elas PT e PSL. Como resultados mais similares a elas se mostram na cor azul e, para o contexto de eleição, é sinalizado pelas palavras em verde.

Outra forma de validar o modelo foi utilizando o método *most similar*, para calcular as palavras mais similares aos parâmetros pesquisados, PT e PSL combinadas com eleição, expostas nas tabelas seguir. Utilizando mais uma vez conhecimento político, pode-se observar que os vetores pesquisados na função retorna palavras que estão realmente associadas umas com as outras.

**Tabela 5: Resultados de palavras mais similares a PT e eleição**

Palavra	Proximidade com o parâmetro
psdb	0,7169
partido	0,7119
pmdb	0,6996
dilma	0,6925
psb	0,6883

**Tabela 6: Resultados de palavras mais similares a PSL e eleição**

Palavra	Proximidade com o parâmetro
psdb	0,5883
psb	0,5873
primeiro_turno	0,5846
pt	0,5844
eleição_presidencial	0,57119

Vale ressaltar que foi pesquisado pela palavra PT e PSL no contexto de eleições. Para a Figura 14, temos que o partido PT tem como palavras mais próximas partido, sigla, petista e no contexto de eleição parece as palavras campanha, política, disputa. Já para o partido PSL, temos que as palavras mais próximas é Jair Bolsonaro, Bolsonaro PSL e no contexto de eleição assim como o PT, campanha, política, disputa. Tomando como base o conhecimento político, pode-se perceber que o modelo criado está fazendo sentido pelo fato de estar ligado às palavras citadas acima.

Após os processos de validação do modelo, foi aplicado o *wmd* na base de dados dos discursos da reforma da previdência, utilizando o modelo *word2vec* criado, com o objetivo de detectar as falas mais similares e, como parâmetro, foram utilizadas os discursos selecionados no tópico 3.4.2 e, como resultado, detectar o alinhamento e contradição partidária diante dos discursos dos parlamentares. Os resultados para esse procedimento estão evidenciados, a seguir:

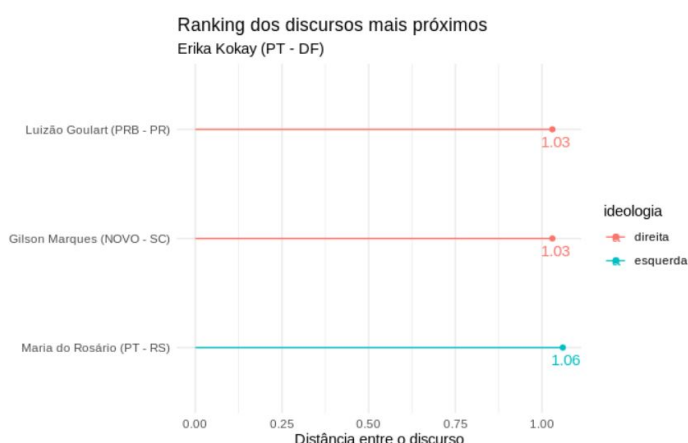
Para o primeiro discurso selecionado, temos a deputada Erika Kokay (PT - DF), demonstrando uma postura favorável à proposta em discussão. Para esse discurso, temos os seguintes resultados:

- Gilson Marques (NOVO - SC):** “Presidente, 56% do Orçamento do Brasil vai para o pagamento da previdência, sendo que só 20% da população recebe esse benefício, ou 40 milhões de pessoas do total de 207 milhões. Existe uma elite que recebe 10% de tudo o que Brasil arrecada. E aí as pessoas dizem: Quem vai pagar o déficit? Esse déficit é histórico. O PT esteve 18 anos no governo e não fez nada para corrigir isso. O déficit não vem de agora. Já atrasamos muito essa discussão e precisamos ter coragem de enfrentá-la. Por que não hoje? Na semana passada, quando esperávamos aqui o Ministro Guedes, e ele não veio, disseram que ele não tinha coragem, que estávamos retardando o processo. Nós não queremos retardar isso. Vamos, sim, discutir a Previdência agora! Em média, 30 milhões de pessoas recebem 1.450 reais. Essa Previdência é a solidariedade do prejuízo. Precisamos enfrentar esse tema. O Partido Novo vota sim.”
- Luizão Goulart (PRB - PR):** “O PRB, Sr. Presidente, tem se aprofundado no tema da Previdência e reconhece que, mais cedo ou mais tarde, teremos que enfrentar o debate, discutir aprofundadamente o tema. Temos divergências em diversos pontos da proposta inicial, mas consideramos que a discussão tem que ser enfrentada. Nós votamos sim.”
- Maria do Rosário (PT - RS):** “Sr. Presidente, eu queria registrar para a sociedade que está nos acompanhando que nós votamos pela obstrução, que é, na verdade, uma obstrução política. Ela representa uma contraposição ativa ao projeto. O sentido dessa obstrução é tentar que



não haja votação. Se V.Exa. mesmo puder explicar... Não, pode deixar que eu explico.”

Como resultado para as distâncias temos na Figura 16:



**Figura 16: Ranking dos discursos mais próximos ao de Erika Kokay selecionado no tópico 3.4.2.**

Para o segundo discurso selecionado, temos a deputada Joice Hasselmann (PSL - SP), que apresenta uma postura favorável a Reforma, argumentando que sem a aprovação do projeto de lei haverá déficit na economia brasileira. Dessa forma, para esse discurso temos os seguintes resultados:

- 1. Zé Neto (PT - BA):** “Sr. Presidente, em verdade, nós já sabemos o resultado do que está acontecendo aqui. Vai haver um embate que não vai se dar no plano jurídico. Há diversas inconstitucionalidades, e nós sabemos que essa disputa vai se dar muito mais no plano político. E o Supremo Tribunal Federal vai ficar aí para fazer o papel dele no que for possível. Ninguém disse que nós da Oposição, principalmente do PT, não fizemos reforma, não fizemos mudança na Previdência. Fizemos, e o próprio Lula reconhece que é necessário, de vez em quando, ajustar as contas da Previdência, porque é uma situação sempre dinâmica. Reconhecemos que os Estados brasileiros passam por momentos difíceis. Estão aí os casos do Rio de Janeiro, do Rio Grande do Sul, de Minas Gerais. Lá na Bahia, temos um déficit na Previdência: uma receita geral de 45 bilhões de reais e um déficit que, neste ano, vai chegar a 4,8 bilhões de reais, mais de 10% do valor de todo o Orçamento. Sabemos disso tudo, mas, em nenhum momento, nós atingimos principalmente o regime geral e os mais pobres deste País. O que vocês querem é mais Brasília. O que vocês querem é tirar o foco da Previdência, colocá-la como uma reforma fiscal. Vamos até mudar um pouco o que está aí. Se queríamos falar de Previdência, com um processo político e econômico tão difícil, devíamos ter atacado a questão tributária e a Lei Kandir; devíamos ter atacado de uma forma em que

podéssemos ter mais velocidade, mais execução e mais especialidade executiva na cobrança de débitos da previdência privada. Deveríamos aqui discutir marcos regulatórios do petróleo. Mas esses que querem mais Brasília e menos Brasil pensam que política é feita com computador. O que eu estou vendo aqui é um bocado de robózinhas que viraram gente que vieram para cá falar com celular na mão, com cartãozinho anotado com quatro, cinco frases, e que não completam nem o tempo que têm, porque não têm o que falar. Parece repetitivo, porque um fala, o outro fala a mesma coisa, e lá vai. Vocês não conhecem o Brasil! Vocês não conhecem o povo brasileiro! Vocês não estão levando em conta os 7 milhões de miseráveis, que voltaram a ser miseráveis em função do golpe que foi dado na Dilma, em função do desastre que aconteceu desde o Aécio, do mesmo timinho. Ficam fantasiando que é isso ou que é aquilo, mas é o mesmo timinho que está lá na cúpula, tentando agora dizer que Não! Nós somos o novo! Novo de quê? O novo propõe que Brasil? A cada dez Municípios brasileiros \$ a cada dez Municípios brasileiros! \$, sete têm receita de Previdência maior do que as receitas do Fundo de Participação dos Municípios \$ FPM. No meu Estado, a Bahia, está muito claro o que está acontecendo: nós temos uma população de 14,8 milhões de habitantes, com uma população urbana de 10,7 milhões e uma população rural de 4 milhões e pouco habitantes; os benefícios da previdência rural são de 1,188 milhões de reais \$ os benefícios da previdência rural! \$, os benefícios da previdência urbana são de 859 milhões de reais. Sabe quanto chega para um Estado que tem uma receita de 45 bilhões de reais? Chegam 28,5 bilhões de reais. Quase 30 bilhões de reais chegam para o Estado pela Previdência. É preciso esclarecer a todos \$ a todos! \$, ao Brasil, aos empresários, que, quando esse dinheiro sai de lá de baixo, de quem ganha menos, da professora, do rural, que não vai mais se aposentar, com as regras que estão aí, todos vão perder: o comércio vai perder, a indústria vai perder, os serviços vão perder. E vão tirar 1 trilhão de reais dos mais pobres, ou 170 bilhões de reais mais ou menos, para botar na mão de quem? De Brasília. Mais Brasília novamente. Esse dinheiro vai voltar para o social? Tenham certeza de que não vai. Estamos aqui dizendo não para essa reforma. Querem debater o Brasil, saiam da Internet, botem o pé no terreiro e vamos conhecer as realidades e as agruras do povo brasileiro.”

- 2. Rubens Bueno (Cidadania - PR):** “Sr. Presidente, desde sempre, nós defendemos a reforma da Previdência, não é de agora, não é do mandato passado, já vem de tempo, por uma simples razão: precisamos garantir a sustentabilidade do sistema. Essa é a razão. Se não garantirmos a sustentabilidade do sistema, daqui a pouco, não vamos ter dinheiro para pagar aposentados e

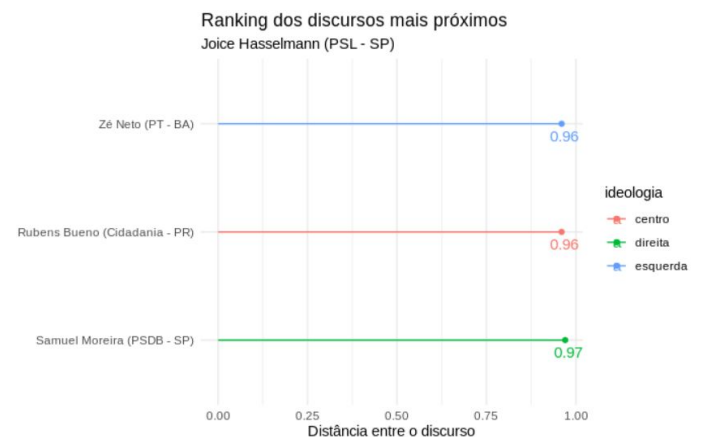
pensionistas. Quem vai pagar, quem vai arcar com as consequências? Exatamente aqueles que mais precisam, porque não há de onde se tirar. Sempre, se nós não fizermos, vai beneficiar aqueles grandes privilegiados que o Brasil tem. O que nós estamos discutindo aqui nesta Comissão é saber se ela é constitucional, se ela pode prosseguir numa Comissão Especial, sabendo que essa é a decisão da Comissão de Constituição e Justiça e de Cidadania. O mérito do projeto S e é válido também que se discuta aqui, nada a opor... Mas é válido também que se discuta lá na Comissão Especial quais as mudanças que se pretendem para a reforma da Previdência, que precisa ser feita. Com o que nós não podemos concordar é que essa reforma venha a atingir os mais pobres. Se estamos tratando do BPC S Benefício de Prestação Continuada, para aqueles que precisam, que têm menos de um quarto de um salário mínimo de renda, nesses não se pode mexer. Mexer aí é atingir aqueles que precisam de proteção social, e a previdência pública também é para isso. Não podemos imaginar que o trabalhador rural venha a sofrer qualquer tipo de ação resultante dessa reforma e que venha a atingir aquele que no campo trabalha de sol a sol. A sua aposentadoria nada mais é do que a contribuição àquele que dedicou a sua vida para o País. Então, não podemos concordar com isso. Quando se trata da capitalização, nós precisamos conhecer melhor o que é. Isso não está claro. A clareza da capitalização é saber até onde isso vai e saber quem vai colocar primeiro o dinheiro S recurso S, que não é pouco, para depois sustentar a capitalização daqueles que vão entrar para o sistema. Sr. Presidente, desde sempre nós defendemos a reforma, mas não uma reforma que vá atingir aqueles que ganham menos. Milhões de brasileiros serão atingidos se isso for aprovado. A reforma da Previdência é para quem ganha muito e se aposenta cedo. Sempre foi a nossa luta estabelecer o teto do INSS para todos. Para quem quer que seja S Deputado, Senador, Ministro, trabalhador S, o teto da Previdência pública seria o mesmo. Previdência pública é para quem ganha menos, não é para quem ganha mais. Os fundos de pensão, seja o FUNPRESP, para o servidor público, seja um fundo de pensão para outro setor da atividade laboral, devem ser feitos de forma que também reinvestam na economia, para que essa dinâmica seja permanente, para que a economia possa funcionar dentro de um círculo virtuoso da economia nacional. Podem saber que nós não vamos de forma alguma deixar aqui a nossa presença. Falo em nome da bancada do Cidadania, em nome do partido Cidadania, quando digo que nós queremos a reforma da Previdência. Sempre a defendemos em todos os documentos do partido, porque sabemos que depende deste Parlamento fazer uma ação corretiva, para que daí possamos ter a tranquilidade

necessária de dizer que há equilíbrio fiscal. Havendo equilíbrio fiscal, teremos condições de investimento na infraestrutura, em saúde, em educação, em segurança pública, que é o dever do Estado para com a sociedade. Nós temos que tratar a questão da reforma da Previdência dentre aqueles que vão ter não só condições de investir. Aí teremos as condições ideais para o Brasil voltar a crescer, se desenvolver. Gerando economia, gerando emprego e renda, poderemos ter até condições melhores para o Governo arrecadar mais. Essa dinâmica é fundamental. Não podemos aqui fechar os olhos para o déficit da Previdência. Se nós juntássemos o dinheiro de 1 ano ou de menos de 2 anos do déficit da Previdência, nós poderíamos ter o saneamento básico zerado no País. Sabemos que iríamos economizar, a cada real investido em saneamento básico, 4 reais em saúde pública. Isso é qualidade de vida, é saber que a partir daí nós teremos as condições ideais para um país muito melhor. Por isso, nós queremos dizer, Sr. Presidente, que vamos votar a favor da reforma da Previdência, com essas colocações que acabei de fazer, até porque já me pronunciei no plenário desta Casa falando sobre isso, na certeza de que estamos Reunião de: 16/04/2019 dando a nossa contribuição para que esta Comissão cumpra o seu papel de admitir esse pedido da reforma da Previdência para a criação da Comissão Especial.”

3. **Samuel Moreira (PSDB - SP):** “Sr. Presidente, Srs. Deputados, quem não conhece alguém com 40 ou 45 anos que já se aposentou aqui no Brasil? Quantas pessoas com 45 ou 50 anos já estão aposentadas aqui no Brasil? Quantas? Tudo isso acontece dentro da lei, rigorosamente dentro da lei. Aliás, como o Deputado que me antecedeu já disse, 65% dos aposentados ganham um salário mínimo, mas quem não conhece aposentados que ganham 30 mil reais por mês? Quantos ganham aposentadoria de 30 mil reais por mês, mas têm emprego, continuam trabalhando e ganhando mais 25 mil reais por mês, enquanto a grande maioria da população está no INSS ganhando um salário mínimo? E nós vamos continuar com essas distorções, achando que não há necessidade de mexer e fazer alterações no nosso sistema de Previdência? Não há unanimidade em torno do mérito da proposta do Governo, mas esse mérito vai ser discutido numa Comissão Especial, a que todos nós vamos levar as nossas divergências. O PSDB fará isso, porque o PSDB está estudando rigorosamente o Benefício de Prestação Continuada. Nós não aceitamos que se retirem direitos e valores daquele que mais precisa. Nós vamos debater a aposentadoria por invalidez e as regras das aposentadorias dos trabalhadores rurais. Nós estamos atentos a essas questões sociais, sobretudo às questões relacionadas à transição e à desconstitucionalidade. Nós vamos debater

isso na Comissão Especial. Agora, não somos daqueles que acham que um projeto de reforma tem que ser enterrado e que tudo deve continuar como está. É inadmissível que se continue tudo como está! Aliás, nós estamos aqui debatendo um projeto de reforma não por grande competência, astúcia ou inteligência, mas porque esse assunto não sai da pauta do Brasil, que vai ter um grave problema fiscal se não debatê-lo com profundidade. O Brasil tem problema fiscal, e a Previdência permite que pessoas mais novas retirem do idoso, permite que cada vez entre mais gente no sistema. Percebe-se nitidamente que nós estamos vivendo mais. Sabe-se que o nosso sistema é deficitário. Então, vamos tirar, mais uma vez e cada vez mais, dos impostos que o cidadão paga, da saúde, da educação, para cobrir o nosso sistema de Previdência. Nós precisamos fazer justiça e ter responsabilidade fiscal com o nosso sistema de Previdência. Nós queremos debater isso na Comissão Especial, item por item, valor por valor, porque nós temos muita responsabilidade para com o sistema fiscal, que é o que está sendo colocado para todos nós debatermos. Ninguém está aqui à toa; nós estamos aqui porque há um problema fiscal. Talvez, se o Brasil estivesse cheio de dinheiro, nós não estaríamos aqui discutindo; ou talvez, se no Brasil houvesse equilíbrio fiscal, nós estaríamos com um sistema deficitário e injusto de Previdência ainda. Nós só estamos colocando o dedo nessa ferida porque há um problema fiscal no Brasil. E o Brasil pode quebrar muito em breve. Portanto, nós vamos debater o mérito. O PSDB debaterá o mérito sempre com duas premissas: justiça social e responsabilidade fiscal sobre o nosso sistema de Previdência. Reunião de: 16/04/2019 Presidente, nós entendemos que há constitucionalidade nesse projeto. Nós vamos apoiar a admissibilidade dele. O capítulo Dos Direitos Sociais da nossa Constituição estabelece a previdência como um direito do cidadão. Nós não podemos assistir à nossa Previdência quebrar, como aconteceu em outros países. E precisamos aproveitar para olhar para aquele que mais precisa, interrompendo essas injustiças, que são tantas! Há pessoas se aposentando mais novas, tudo dentro da lei; há pessoas ganhando 30 mil reais, 35 mil reais por mês, enquanto 65% dos trabalhadores... Aliás, a média do valor recebido por todos que estão no INSS é 1.290 reais, enquanto a média em alguns lugares chega a 24 mil reais, como é o caso desta Casa. Entre funcionários e Deputados, a média do que recebe um aposentado aqui é 24 mil reais, enquanto no INSS é 1.291 reais. E nós vamos continuar sem mexer na nossa Previdência, enterrando esse projeto. Portanto, Presidente, essa é a posição do PSDB. O PSDB continuará firme nesse processo, disposto realmente a mexer no nosso sistema de Previdência. Muito obrigado.”

Como resultados para o cálculo de distância do discurso de Joice Hasselmann, temos na Figura 17:



**Figura 17: Ranking dos discursos mais próximos ao de Joice Hasselmann selecionado no tópico 3.4.2.**

## 5. CONCLUSÃO

A partir dos resultados obtidos pela pesquisa, podemos observar que o método de calcular distância entre discursos foi eficiente para detectar alinhamento e contradição partidária, já que ao observar os discursos de parlamentares, foi possível identificar falas similares ao discurso escolhido para o estudo e junto com o dado de partido do deputado, possibilitou observar se o deputado estava de acordo com a ideologia do partido.

Diante disso, podemos ver que, para o primeiro experimento, o discurso de Erika Kokay, é de contradição partidária, já que grande parte do PT foi contra a reforma e o conteúdo da fala da deputada mostra adesão à uma das propostas apresentadas em reunião. Já para seus resultados, temos que Gilson Marques do Novo e Luizão Goulart do PRB, mostraram alinhamento partidário, já que esses partidos possuem ideologias neoliberais e de direita e são a favor da Reforma, uma vez que para a fala de Maria do Rosário do PT, assim como Erika, mostra um caso de contradição partidária, visto que o PT possui práticas de esquerda.

Já para análise do conteúdo dos discursos, temos que assunto abordado é o alinhamento com a proposta apresentada, sob a justificativa de não ter o direito de aposentadoria, principalmente pessoas de baixa renda. Dessa forma, os resultados mostram-se próximos pelo fato de abordarem assuntos semelhantes, como o alinhamento com a proposta e o déficit na economia e as dificuldades que causará para a população.

Para o segundo caso de estudo, foi selecionado o discurso de Joice Halssemann do PSL, mostrando que a fala é de alinhamento partidário, pois demonstra adesão à proposta e pertence um partido de direita. Diante disso, os resultados de discursos similares, tem-se o discurso do deputado Zé Neto do PT que se mostra contra a reforma, sendo um caso de alinhamento partidário, já que pertence a um partido de esquerda. Para Rubens Bueno do Cidadania, foi a favor da reforma, mas mostra-se como



um caso particular, já que pertence um partido de centro. E, por fim, Samuel Moreira do PSDB, que também foi a favor da reforma, mas se encaixa em um posicionamento de alinhamento partidário, já que pertence a um partido de direita.

E, para a análise do conteúdo do discurso de Joice, temos que o assunto abordado é o alinhamento com a proposta apresentada, tendo como justificativa, a demonstração de do déficit que seria causado se a proposta de lei não fosse implantada. Diante disso, os resultados mais similares abordam também como tema o déficit causado na economia relacionando com o projeto de lei.

Ainda sobre a análise do conteúdo das falas, pode-se observar casos particulares, como os discursos de Zé Neto do PT, sendo o mais próximo ao de Joice, sendo que, são parlamentares de ideologias contrárias e, abordando postura antagônica nos discursos apresentados ao projeto em discussão. Tal afirmação é justificada pelo fato da abordagem do mesmo conteúdo, o déficit econômico, Joice falando que a economia irá ficar defasada se não houver reforma e, por outro lado, Zé Neto falando que a economia será afetada se houver reforma da previdência.

## 6. TRABALHOS FUTUROS

Como sugestão para próximos passos é válido considerar a automatização do processo de captura de pdfs do site da Câmara para um projeto de lei no raspador de discursos, com o objetivo de tornar o processo de aquisição dos dados de discursos mais rápido. Ainda como trabalho futuro para a ferramenta de captura de discursos, outro ponto relevante é expandir a captura de pdfs para o Senado Federal para reconhecer também o alinhamento e contradição partidária entre os senadores brasileiros.

Já para as análises, seria interessante expandir o conhecimento para detectar alinhamento e contradição intrapartidária, ou seja, saber o quão os discursos dos partidos estão próximos ou distantes de um discurso selecionado do mesmo partido e, junto com o experimento da pesquisa, observar se um parlamentar de outro partido está alinhado com a média de distância um partido selecionado.

## 7. REFERÊNCIAS

- [1] TURCHI, S. A internet é democrática, 2009. Disponível em: <<http://sandraturchi.com.br/a-internet-e-democratica/>> Acesso em 07 Junho de 2019.
- [2] ROSA, R. L. Análise de sentimentos e afetividade de textos extraídos das redes sociais, 2015. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/3/3141/tde-1907-2016-115713/publico/RenataLopesRosa2015.pdf>> Acesso em 01 de julho de 2019.
- [3] SILVA, L. K. et al. Análise de Sentimento pela ótica da abordagem multimodal, 2017. Disponível em: <<https://seer.ufrgs.br/renote/article/viewFile/75112/42552>> Acesso em 09 de setembro de 2019.
- [4] RODRIGUES, J. O que é o Processamento de Linguagem Natural?, 2017. Disponível em: <<https://medium.com/botsbrasil/o-que-%C3%A9-o-processamento-de-linguagem-natural-49ece9371cff>> Acesso em 09 de setembro de 2019.
- [5] CÔRTEZ, Sérgio da Costa; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. Mineração de Dados – Funcionalidades, Técnicas e Abordagens. PUCRioInfMCC10/02, 2002. Disponível em: <[ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02\\_10\\_cortes.pdf](ftp://obaluae.inf.puc-rio.br/pub/docs/techreports/02_10_cortes.pdf)>. Acesso em: 01 de out. 2019.
- [6] ANDRADE, Lucas Matheus Santos; BARROS, Rafael Couto; SANTOS, Marcelo Anderson Batista. Processamento de linguagem natural (PLN): Ferramentas e desafios, 2017. Disponível em: <[http://www.editorarealize.com.br/revistas/conidis/trabalhos/TRABALHO\\_EV064\\_MD1\\_SA6\\_ID2260\\_0910201621564\\_9.pdf](http://www.editorarealize.com.br/revistas/conidis/trabalhos/TRABALHO_EV064_MD1_SA6_ID2260_0910201621564_9.pdf)> Acesso em: 11 de novembro de 2019.
- [7] CARVALHO, Matheus Hemínio. Estudo comparativo dos métodos de *Word Embedding* na análise de sentimentos, 2018. Disponível em: <[https://www.cin.ufpe.br/~tg/2018-2/TG\\_CC/tg\\_mhc.pdf](https://www.cin.ufpe.br/~tg/2018-2/TG_CC/tg_mhc.pdf)>. Acesso em: 15 de novembro de 2019.
- [8] CONTRASTES, Felipe. Similaridade entre títulos de produtos com Word2Vec, 2018. Disponível em: <<https://medium.com/luiyalabs/similaridade-entre-t%C3%A9-tulos-de-produtos-com-word2vec-5e26199862f0>> Acesso em: 11 de novembro de 2019.
- [9] KUSNER, Matt; SUN, Yu; KOLKIN, Nicholas I.; WEINBERGER, Killian Q. From Word Embeddings To Document Distances, 2015. Disponível em: <<http://proceedings.mlr.press/v37/kusnerb15.pdf>>. Acesso em: 15 de novembro de 2019.
- [10] RIBEIRO, Pedro Fielú; MENDES, Amanda Silva. Ideologia e Discurso: o posicionamento do Partido dos Trabalhadores na oposição e no governo, 2019. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0103-33522019000100161](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-33522019000100161)>. Acesso em: 15 de novembro de 2019.
- [11] COUTO, Cláudio. O que significa esquerda, direita e centro na política?, 2018. Disponível em: <<https://politica.estadao.com.br/noticias/eleicoes,o-que-significam-direita-esquerda-e-centro-na-politica,70002314116>>. Acesso em: 19 de novembro de 2019.
- [12] SALES, Allan; BALBY, Leandro; VELOSO, Adriano. Media Bias Characterization in Brazilian Presidential Elections, 2019. Disponível em: <<https://dl.acm.org/citation.cfm?id=3343656>>. Acesso em: 20 de novembro de 2019.
- [13] ALI, Zafar. A simple Word2Vec tutorial, 2019. Disponível em: <<https://medium.com/@zafaralibagh6/a-simple-word2vec-tutorial-61e64e38a6a1>>. Acesso em: 20 de novembro de 2019.