



Programa de Pós-Graduação em Modelagem Computacional



UM MODELO COMPUTACIONAL PARA MINERAÇÃO DE DADOS NO FACEBOOK APLICADO A INFERÊNCIA DOS INDICADORES DE CIDADES INTELIGENTES

Gabriel da Silva Almeida



SUMÁRIO

1 OBJETIVOS

- GERAIS
- ESPECÍFICOS

2 MOTIVAÇÃO

3 JUSTIFICATIVA

4 FUNDAMENTAÇÃO TEÓRICA

5 METODOLOGIA

6 RESULTADOS

7 CONSIDERAÇÕES FINAIS

8 REFERÊNCIAS

OBJETIVOS GERAIS

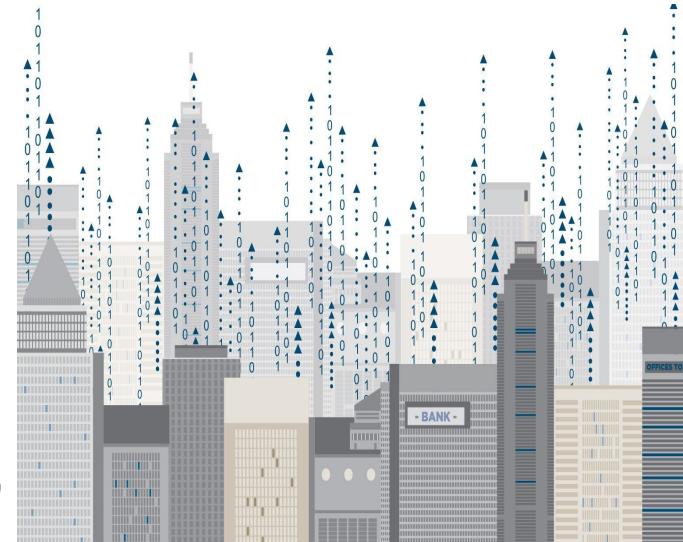
- ❑ Propor uma metodologia baseada em classificação de textos a ser aplicada em postagens de redes sociais;
- ❑ Obter indícios sobre a percepção coletiva da população a partir de termos representativos, relacionados aos indicadores de qualidade de vida no âmbito de cidades inteligentes.

OBJETIVOS ESPECÍFICOS

- ❑ Realizar um estudo bibliográfico referente aos indicadores de cidades inteligentes e mineração de textos;
- ❑ Propor um modelo computacional para obtenção de termos representativos relacionados a aspectos e indicadores de cidades inteligentes;
- ❑ Aplicar o modelo a um estudo de caso envolvendo:
 - 40 páginas e grupos do *Facebook*;
 - 53579 postagens (*posts*) relacionadas às informações da cidade do Rio Grande - RS.

MOTIVAÇÃO

- ❑ Enfoque na Computação Social;
- ❑ Aumento de dados na *web*;
- ❑ Aplicação de técnicas de mineração de dados em documentos não estruturados;
- ❑ Geração de informações
(Fry, 2014; Woodall et al., 2017);
- ❑ Redes sociais como fonte de dados
(McCormick et al., 2017).



JUSTIFICATIVA

- ❑ Contexto de *Smart Cities*;
 - ❑ Analisar dados expressos pela população;
 - ❑ Redes sociais:
 - Contêm publicações sobre assuntos das mais diversas naturezas e informações que, muitas vezes, podem expressar percepções relevantes a serem consideradas para a tomada de decisão, por exemplo, na gestão pública, para o entendimento das manifestações da população;
 - Tratam-se de espaços onde indivíduos manifestam comentários relacionados a experiências do seu cotidiano.



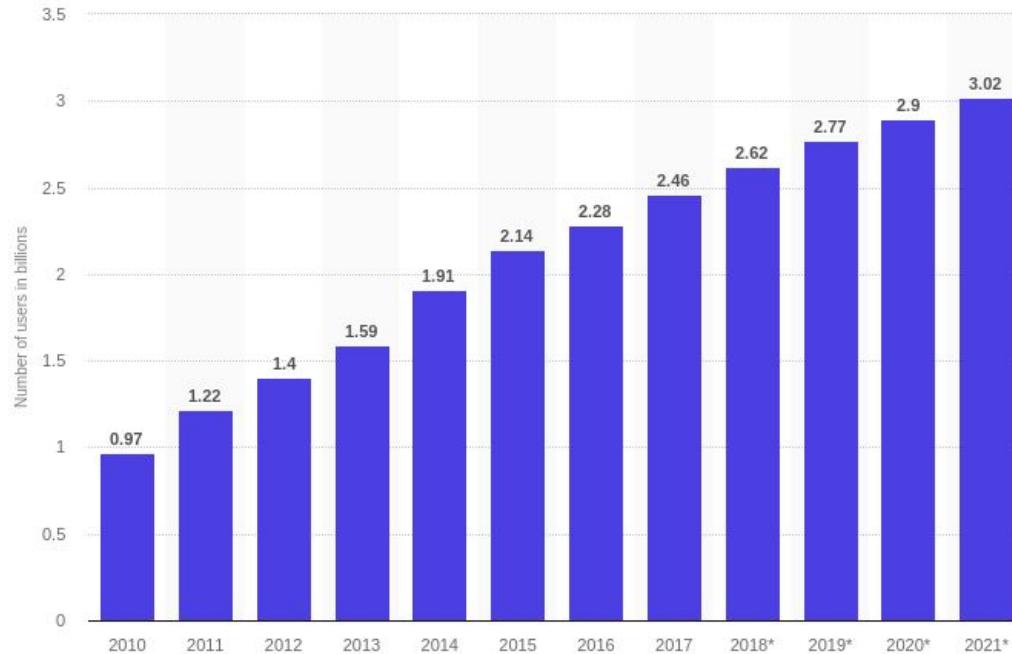
JUSTIFICATIVA

- Quantidade de pessoas que usam redes sociais (STATISTA, 2017a);
 - Quantidade de pessoas que usam *Facebook* no Brasil (STATISTA, 2017b);
 - Popularidade do *Facebook* no Brasil (STATISTA, 2017c);
 - Ausência de limite de caracteres em publicações do *Facebook*.



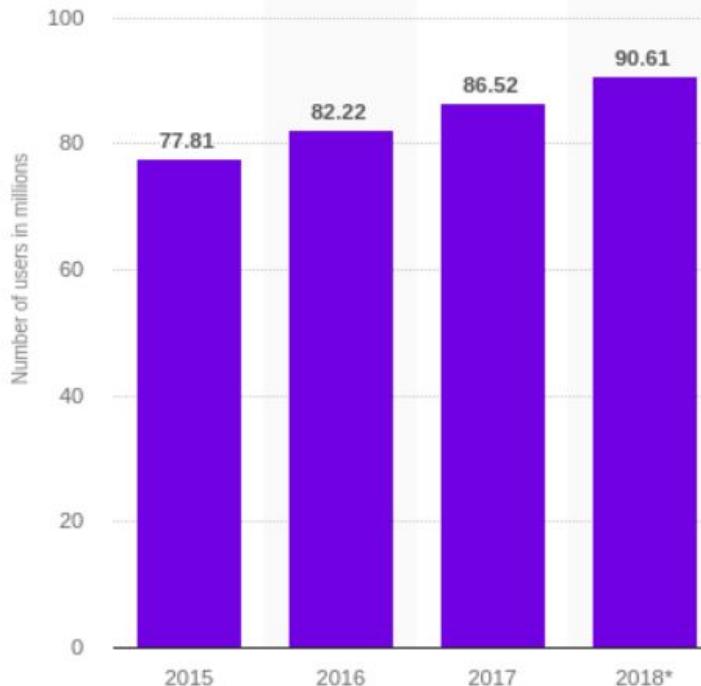
JUSTIFICATIVA

- Quantidade de pessoas que usam redes sociais (STATISTA, 2017a)



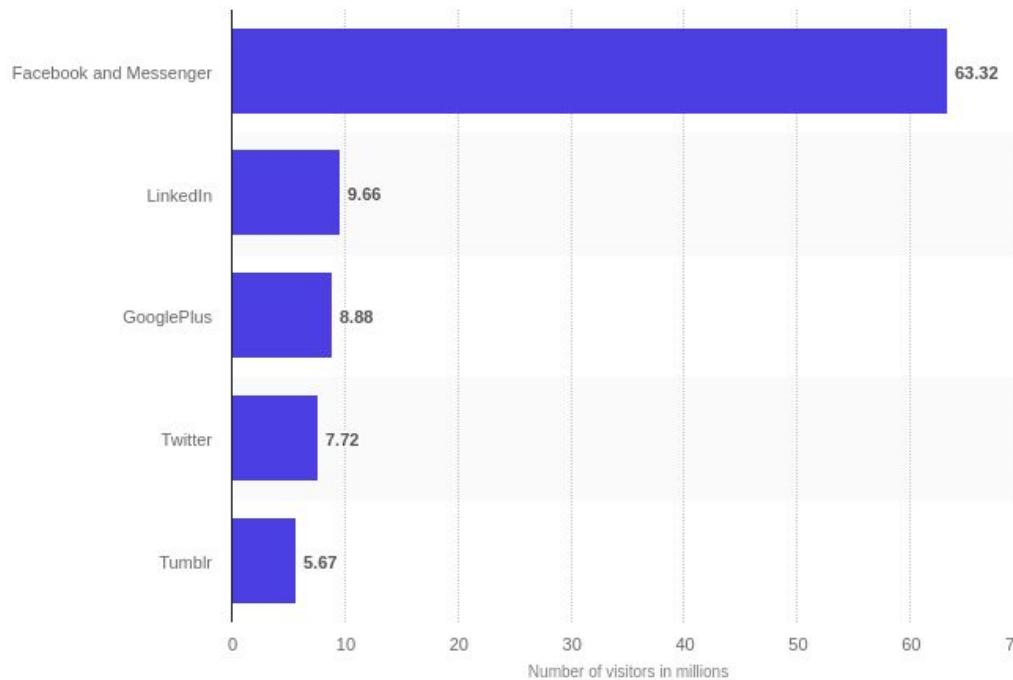
JUSTIFICATIVA

- Quantidade de pessoas que usam *Facebook* no Brasil (STATISTA, 2017b)



JUSTIFICATIVA

- Popularidade do *Facebook* no Brasil (STATISTA, 2017c)



FUNDAMENTAÇÃO TEÓRICA

- Cidades Inteligentes

- ❑ Segundo Hall et al. (2000), cidades inteligentes são aquelas que monitoram e integram as condições de operações críticas da cidade, atuando de forma preventiva para a continuidade de suas atividades fundamentais.
- ❑ Para Harrison e Donnelly (2011), são cidades que fazem uso sistemático das TICs para promover a eficiência no planejamento, execução e manutenção dos serviços e infraestruturas urbanos.
- ❑ De acordo com Komninos (2013), consiste em um ambiente que engloba um espaço digital, através das TICs, associado ao processamento de informações, a transferência de conhecimento e a instrumentos vinculados à tecnologia.

FUNDAMENTAÇÃO TEÓRICA

- Indicadores de Cidades Inteligentes

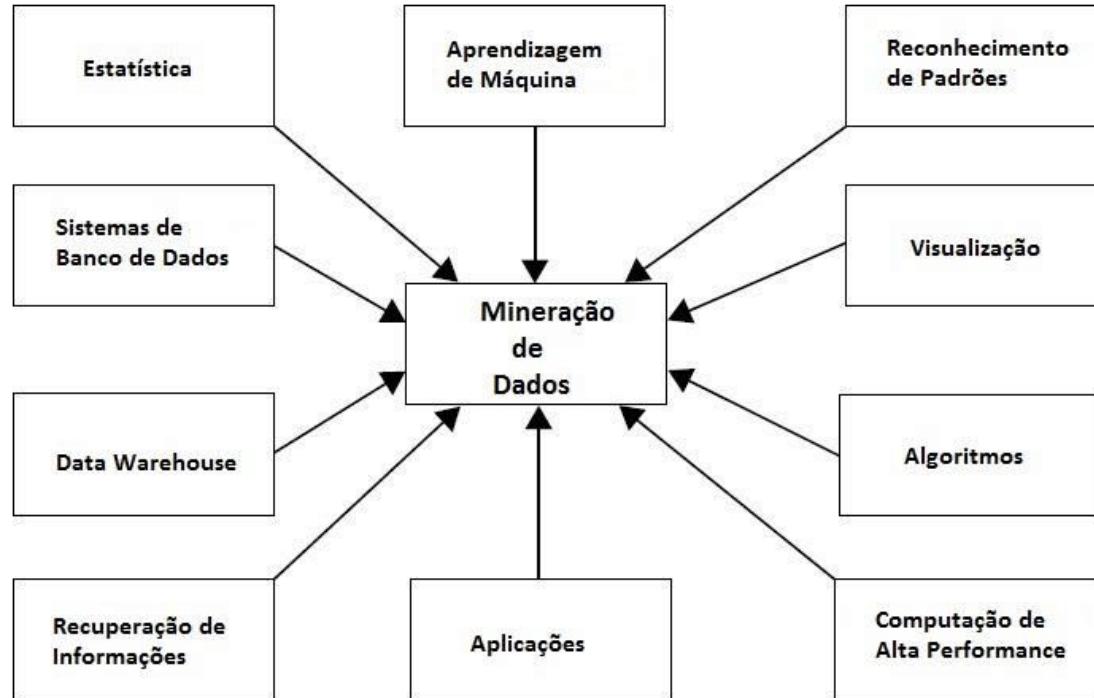
	Indicador	Mobilidade e Acessibilidade	Urbanismo	Meio Ambiente	Energia	Tecnologia e Inovação	Saúde	Segurança	Educação	Empreendedorismo	Governança	Economia
Mobilidade e Acessibilidade	Proporção de ônibus / auto.	X	X									
	Idade Média da Frota	X	X	X								
	Ônibus / Habitantes	X	X									
	Outros modais de transporte coletivo	X	X	X								
	Ciclovias	X							X			
	Rampa para Cadeirante (acessibilidade)	X										X
	Nº de voos semanais (conectividade)	X										
	Transporte Rodoviário (conectividade)	X										
Urbanismo	Lei zoneamento ou uso e ocupação do solo		X									X
	Lei operação urbana consorciada		X									X
	Plano Diretor Estratégico Municipal		X									
	Emissão de certidão negativa de débito e alvará online		X									
	Vias Pavimentadas	X	X									
	Despesa Municipal com Urbanismo		X									X
Meio Ambiente	Atendimento urbano de água		X	X								
	Perdas na distribuição			X								
	Atendimento urbano de esgoto		X	X					X			
	Recuperação de materiais recicláveis			X								
	Cobertura do serviço de coleta de resíduos			X				X				
	Arborização		X	X								
	Monitoramento de Áreas de Risco			X					X			X
Energia	Tarifa Média				X							
	Domicílios com energia de fonte diferente da distribuidora				X							
	Produção de Energia em Usinas de Energia Eólica			X	X							
	Produção de Energia em Usinas de UFV			X	X							
	Produção de Energia em Usinas de Biomassa			X	X							
	Iluminação Pública			X						X		X
	Domicílios com existência de energia elétrica			X								
Tecnologia e Inovação	Conexões de Banda Larga com + de 34 mb					X						
	Municípios com Backhaul de Fibra Ótica					X						
	Cobertura 4G					X						
	Trabalhadores com ensino superior				X					X		
	Acessos do Serviço de Comunicação Multimídia				X							
	Patentes				X						X	
	Bolsa CNPQ				X					X	X	

FUNDAMENTAÇÃO TEÓRICA

- Indicadores de Cidades Inteligentes

FUNDAMENTAÇÃO TEÓRICA

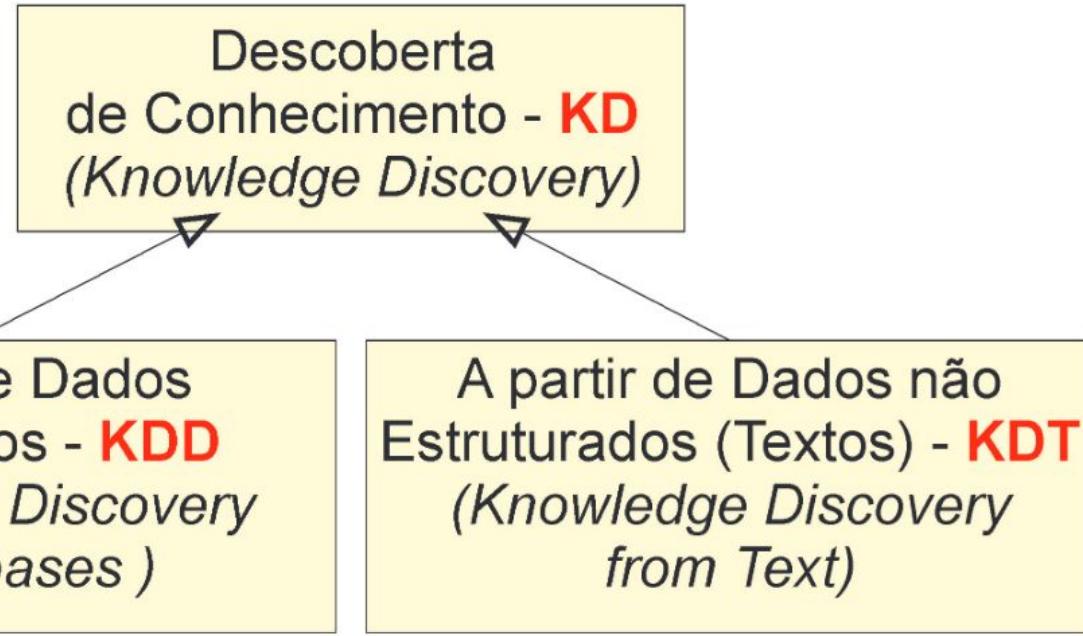
- Mineração de dados



Fonte adaptada de Han, Pei e Kamber (2011).

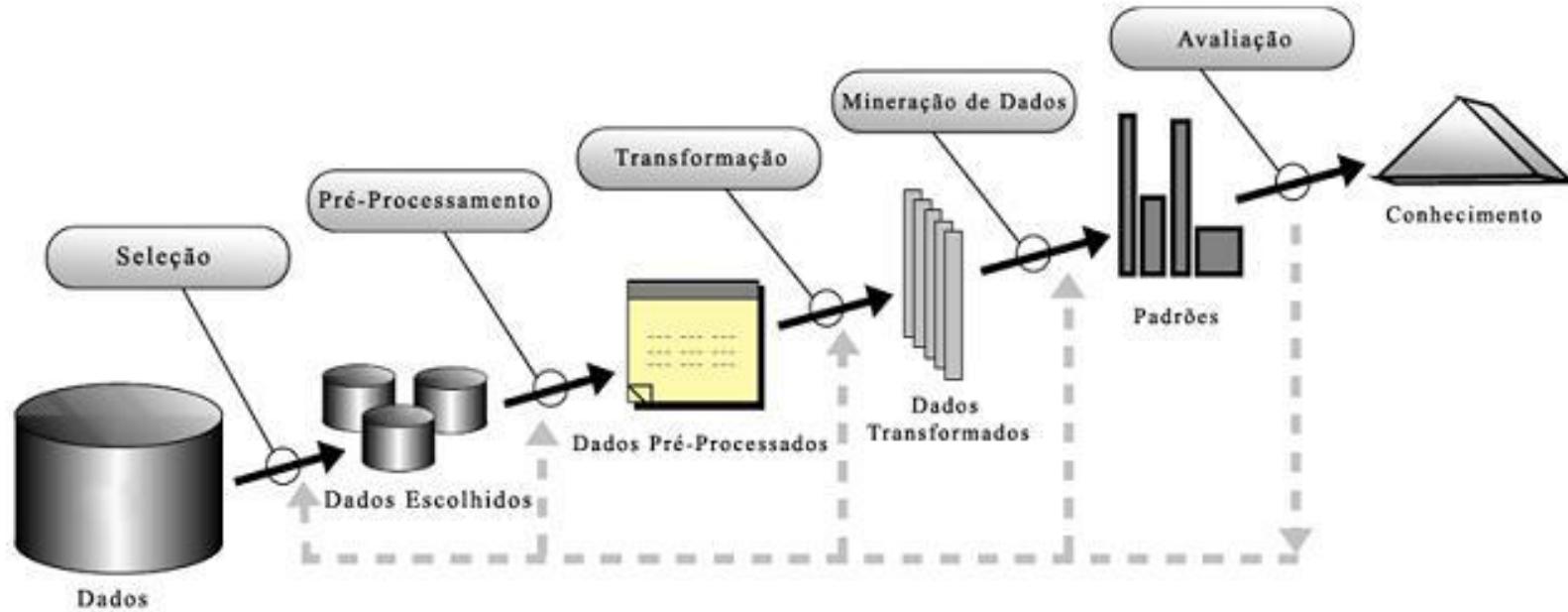
FUNDAMENTAÇÃO TEÓRICA

- Mineração de dados



FUNDAMENTAÇÃO TEÓRICA

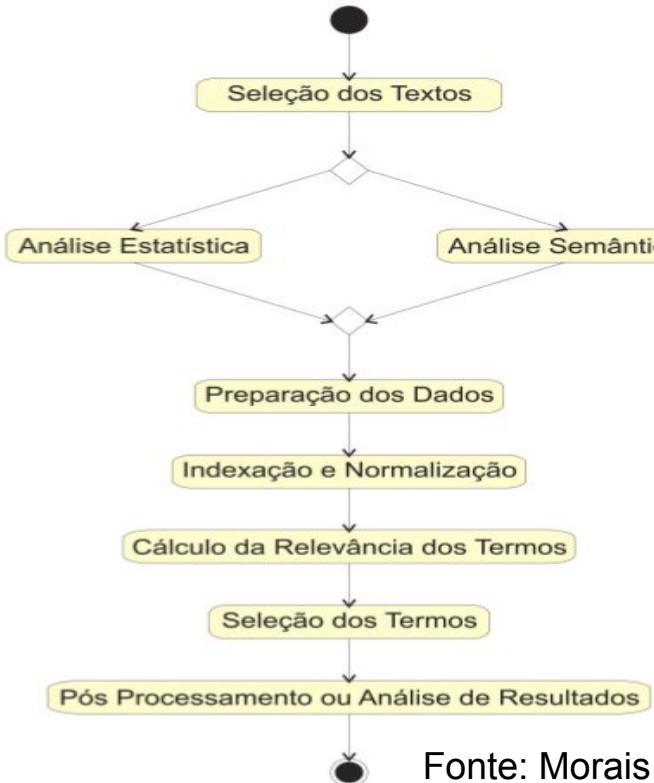
- Mineração de dados (KDD)



Fonte adaptada de Goldschmidt e Passos (2005).

FUNDAMENTAÇÃO TEÓRICA

- Mineração de dados (KDT)



Fonte: Morais e Ambrósio (2007).

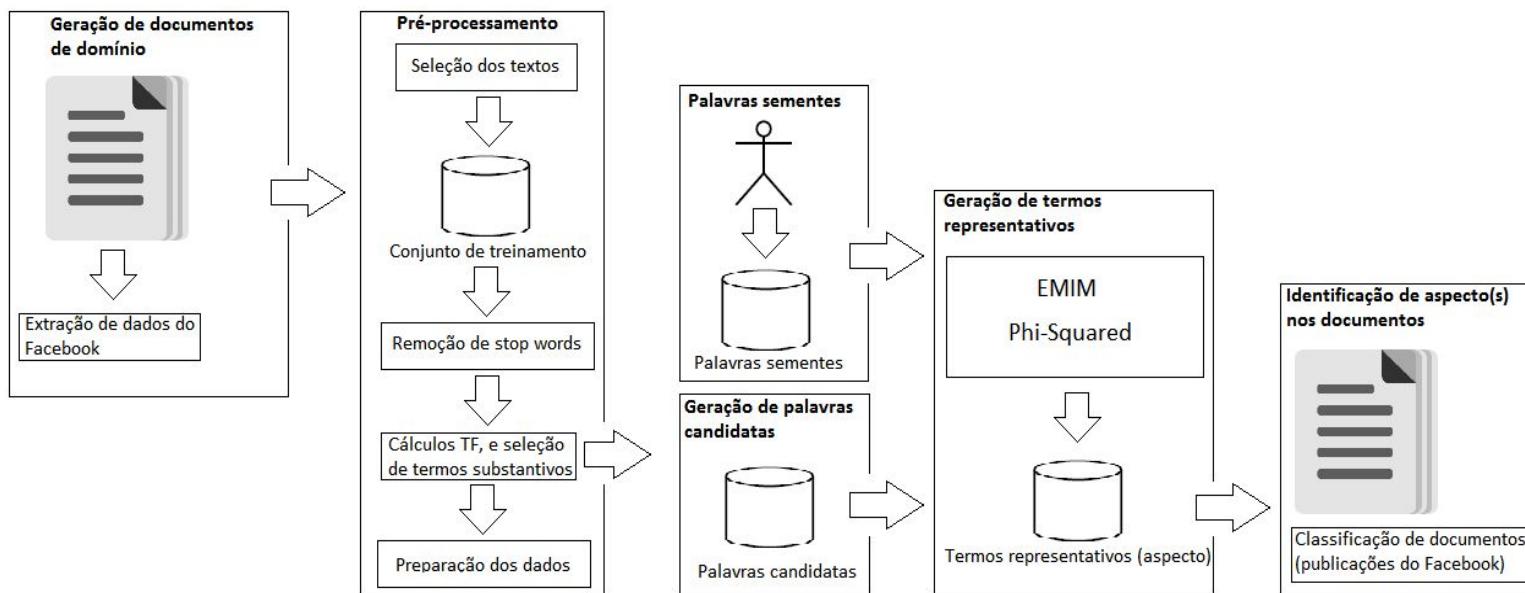
METODOLOGIA

- ❑ Pesquisa exploratória onde busca-se estudar técnicas e ferramentas de mineração de dados capazes de permitir inferir o grau de inteligência de um território, a partir de postagens do *Facebook*.
- ❑ Apresenta-se um modelo e suas ferramentas capazes de pré-processar estas postagens, bem como fornecer termos representativos associados aos diferentes aspectos identificados na literatura como importantes na determinação do nível de inteligência de uma cidade.
- ❑ A partir dos termos representativos, propõe-se também prover um método capaz de classificar uma postagem quanto a um aspecto.

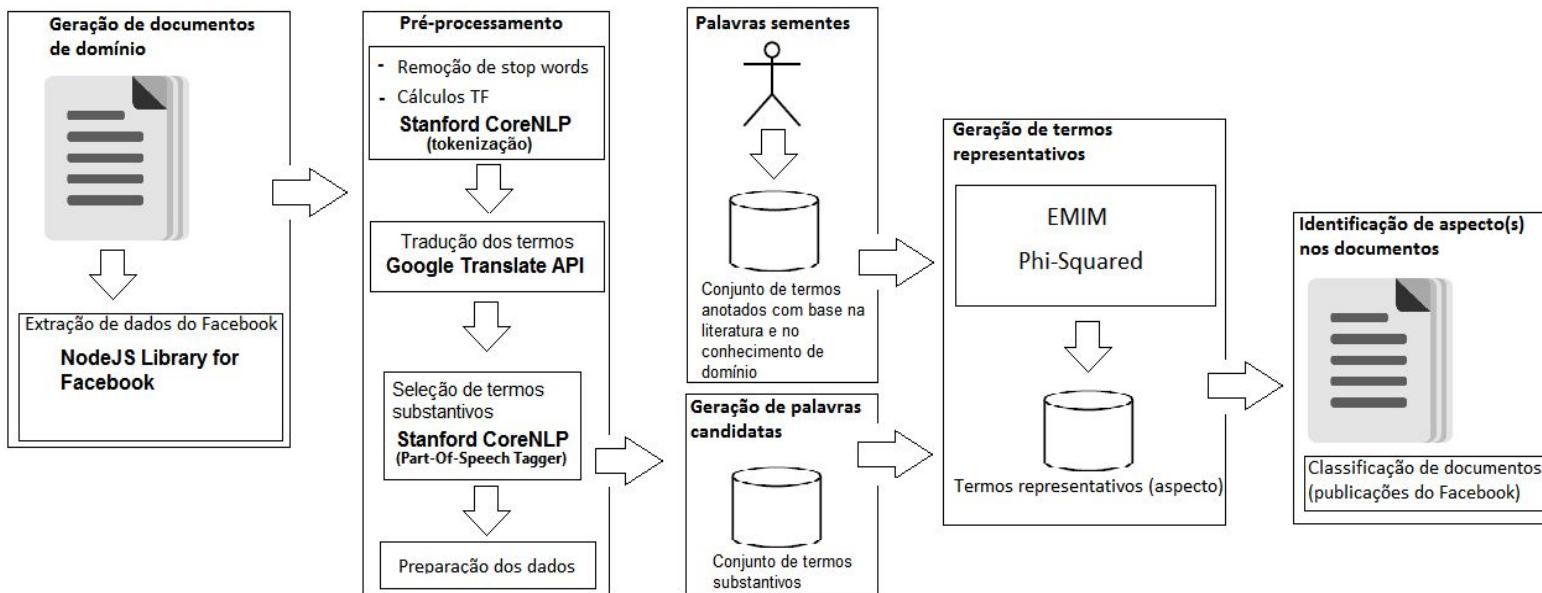
METODOLOGIA - ASPECTOS

- Economia
- Educação
- Empreendedorismo
- Energia
- Governança
- Meio Ambiente
- Mobilidade
- Saúde
- Segurança
- Tecnologia & Inovação
- Urbanismo

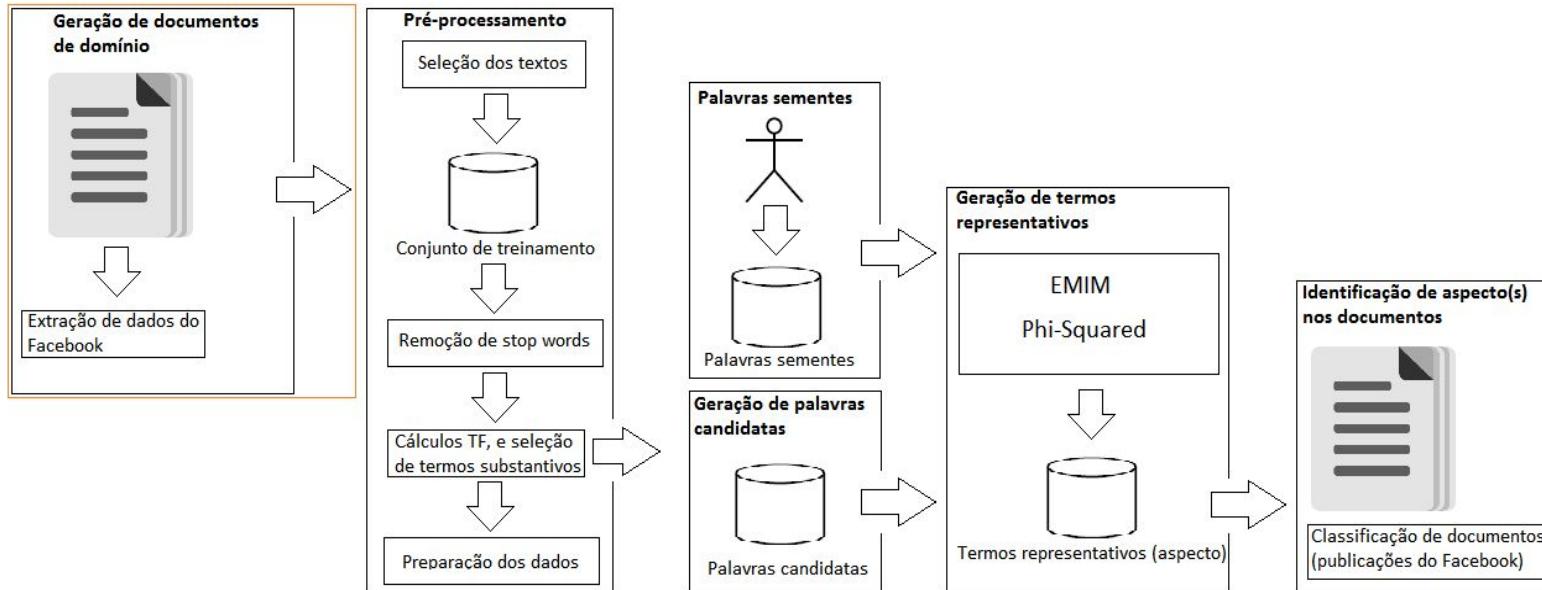
MODELO COMPUTACIONAL



MODELO COMPUTACIONAL

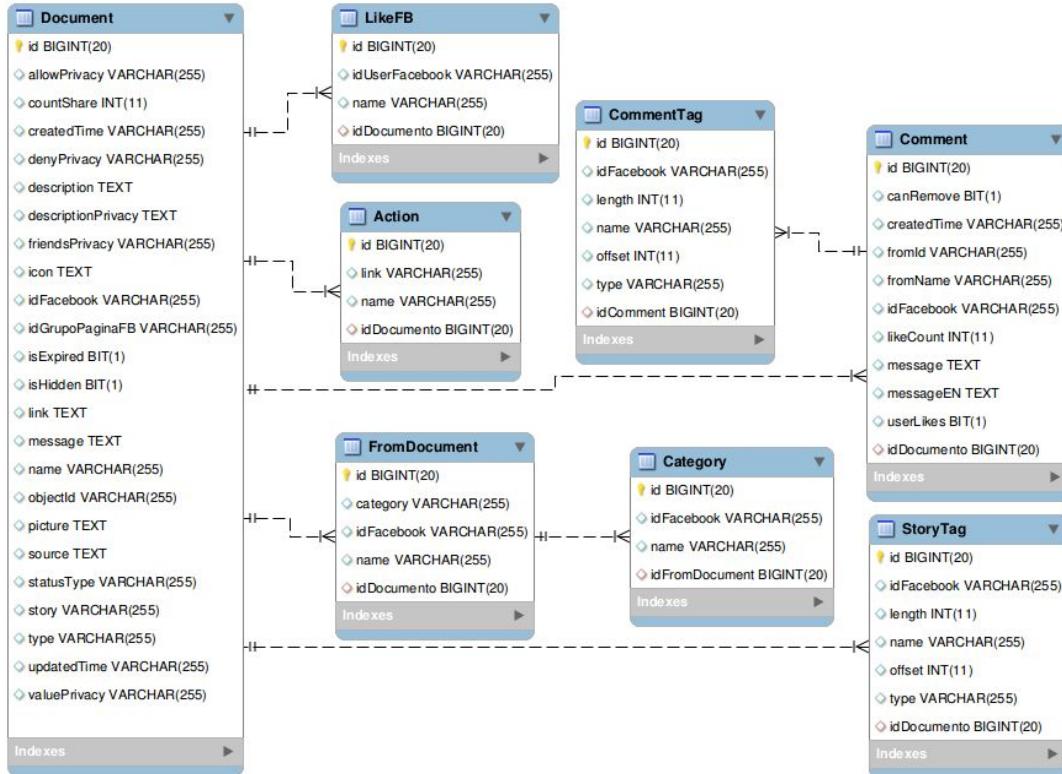


MODELO COMPUTACIONAL

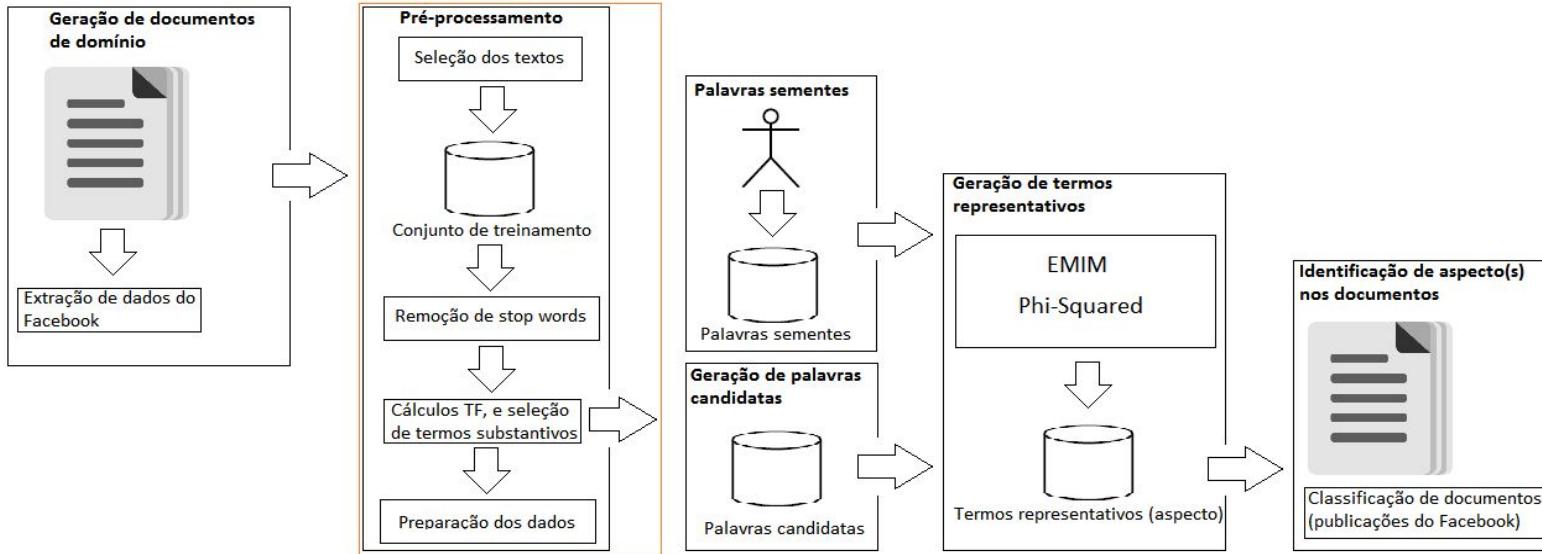


MODELO COMPUTACIONAL

- Geração de documentos de domínio



MODELO COMPUTACIONAL



Classificação de documentos do Facebook

CLASSIFICADOS NÃO CLASSIFICADOS

ID:129856

DATA:14/02/2017

NAME:

Rio Grande On Line.

DESCRIPTION:

Rio Grande On Line :.....Registrados dois acidentes com ciclistas nas últimas horas.....

Acidente na ERS 734 próximo ao shopping

Uma ciclista de 33 anos foi atropelado por um carro na Ers 734 na tarde desta Segunda Feira,foi socorrido e encaminhado a santa casa.

Acidente na Avenida no Cassino

Uma ciclista de 20 anos foi atropelada por volta das 21:00 hs na Avenida Rio Grande no Cassino por um carro. A vitima teve diversos ferimentos e foi encaminhada a santa casa de Rio Grande.

Fonte : Brigada Militar

Foto : Ilustrativa

Texto e Postagem : Fabiano Correa

MESSAGE:

Alimentação

Cultura

Economia

Educação

Empreendedorismo

Alimentação
Cultura
Economia
Educação
Empreendedorismo
Energia
Esporte
Governança (Administração pública)
Meio Ambiente
Mobilidade
Política
Saúde
Segurança
Serviço
Tecnologia e Inovação
Trabalho
Urbanismo

De acordo com a sua opinião, como você define este texto em relação ao conteúdo da publicação? O texto expressa um sentimento:

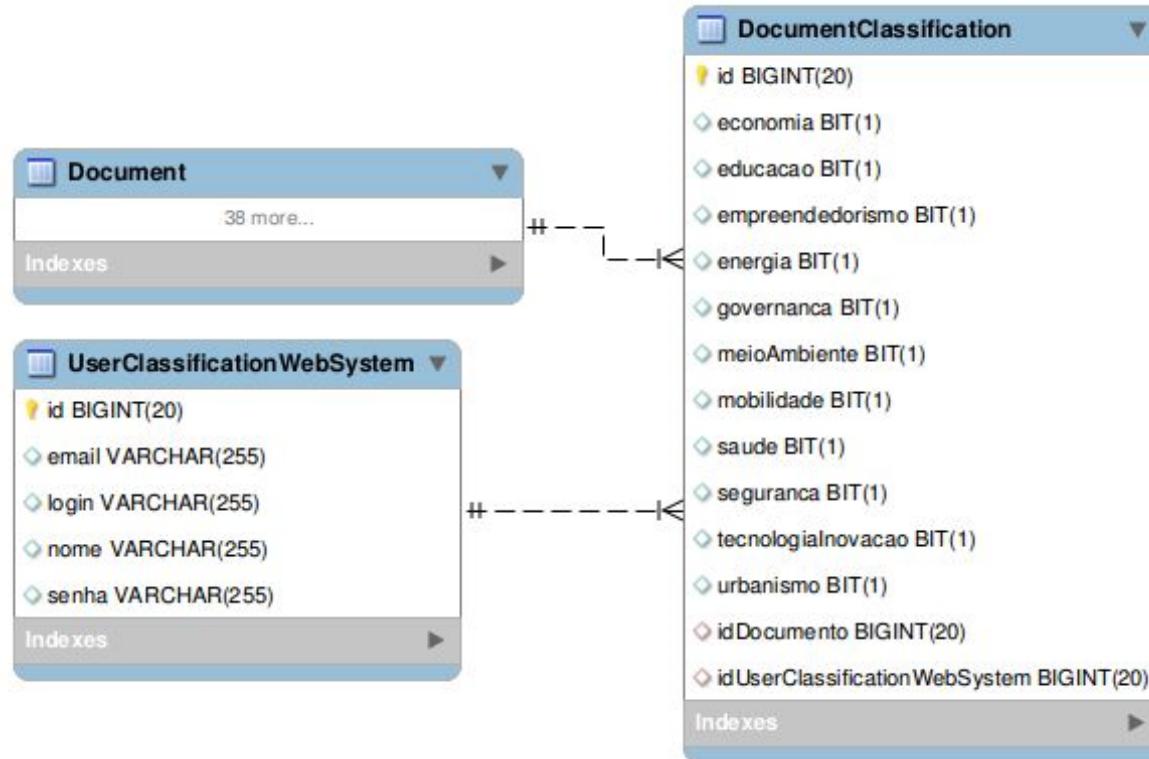
- Negativo
- Neutro
- Positivo

Enviar / Próximo



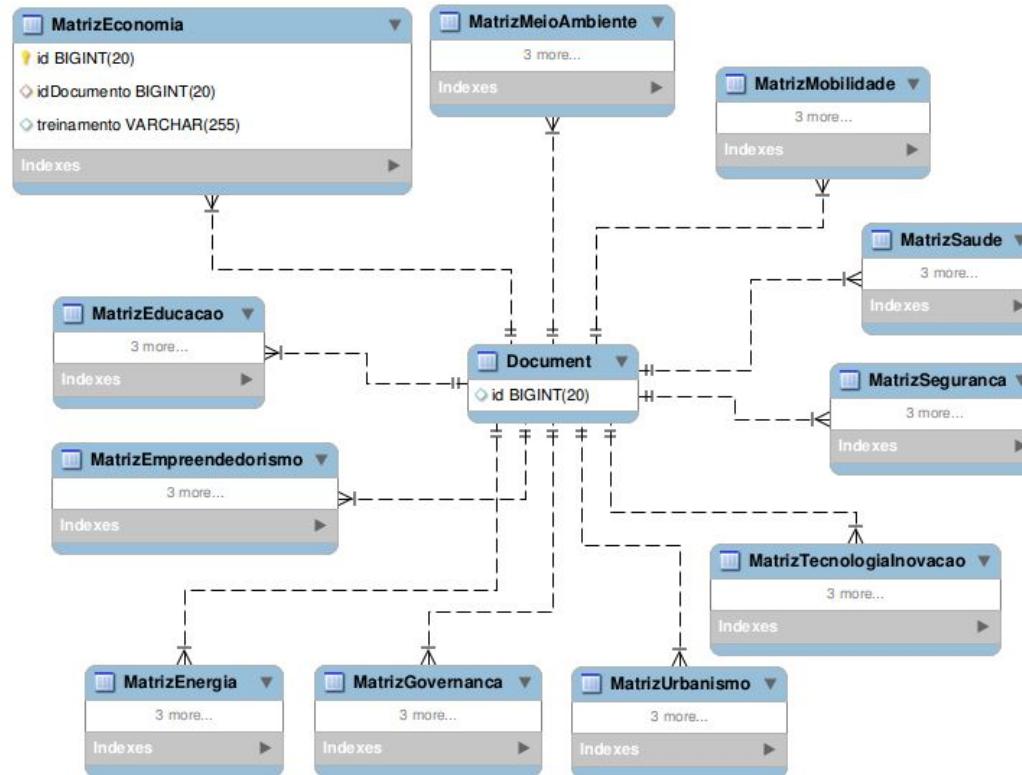
MODELO COMPUTACIONAL

- Pré-processamento (*dataset* anotado)



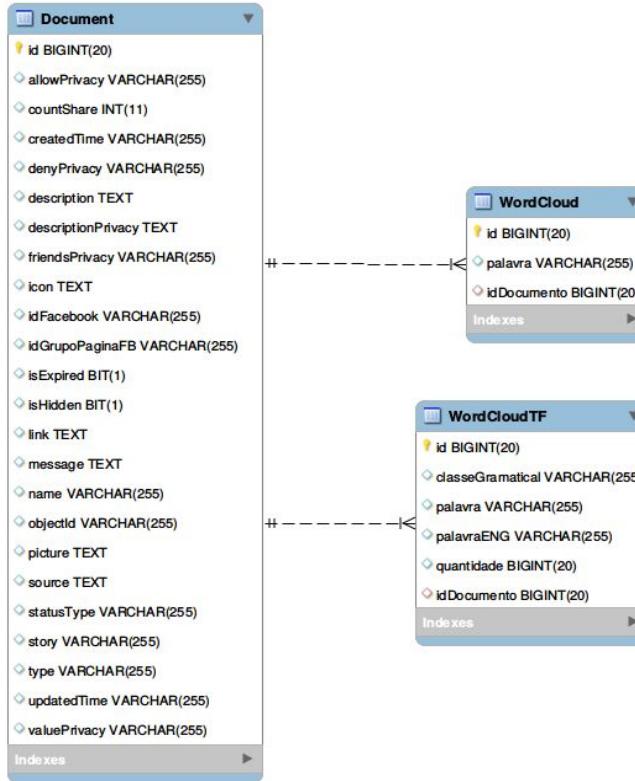
MODELO COMPUTACIONAL

- Pré-processamento (*dataset de treinamento e validação*)



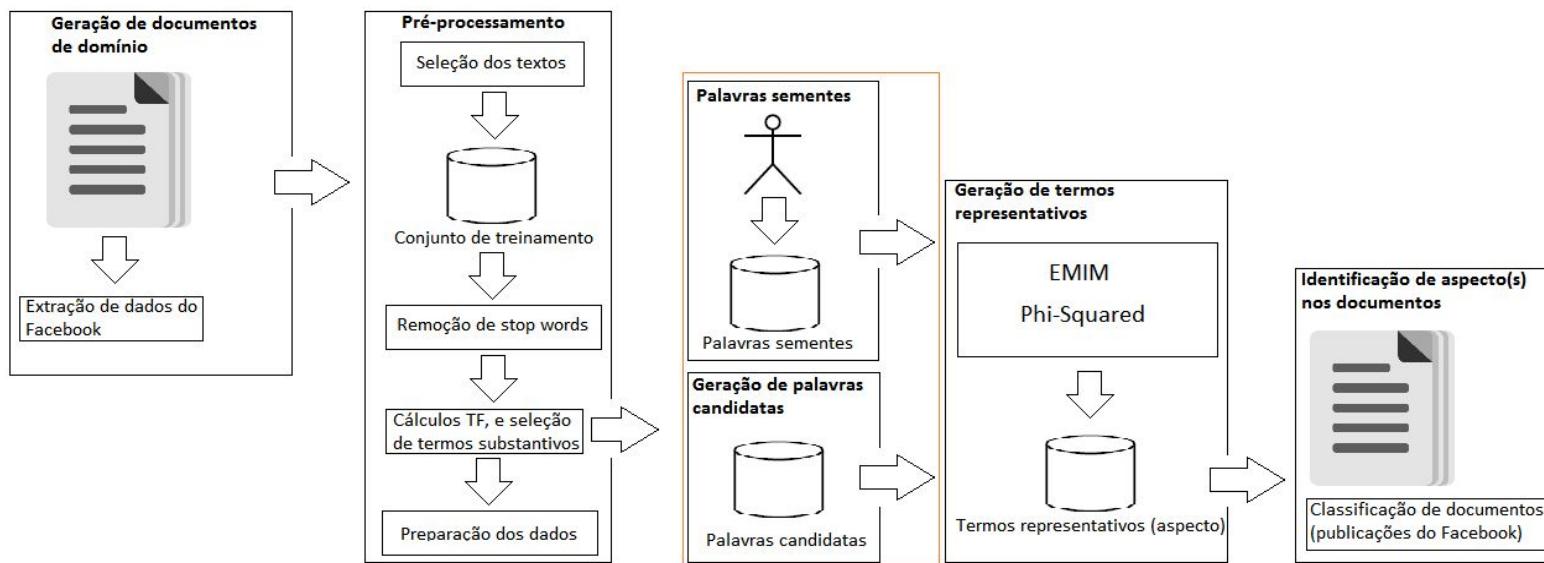
MODELO COMPUTACIONAL

- Pré-processamento (preparação dos dados)



- **Remoção de stop words;**
- **Cálculos Term Frequency (TF);**
- **Tradução dos termos;**
- **Anotação de termos substantivos (Part-Of-Speech Tagger (POS Tagger)).**

MODELO COMPUTACIONAL



MODELO COMPUTACIONAL

- Palavras sementes

Avaliação de qualidade das palavras sementes anotadas

Aspecto	Experimentos									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Economia	85 / 97	103 / 126	107 / 126	107 / 126	108 / 126	108 / 126	106 / 126	105 / 126	107 / 126	3 / 126
Educação	181 / 194	233 / 255	236 / 255	234 / 255	235 / 255	236 / 255	235 / 255	233 / 255	236 / 255	235 / 255
Empreendedorismo	3 / 4	6 / 8	6 / 8	5 / 8	6 / 8	6 / 8	7 / 8	6 / 8	6 / 8	5 / 8
Energia	14 / 16	23 / 26	24 / 26	23 / 26	24 / 26	25 / 26	23 / 26	24 / 26	23 / 26	23 / 26
Governança	80 / 85	108 / 115	106 / 115	103 / 115	103 / 115	106 / 115	105 / 115	104 / 115	103 / 115	103 / 115
Meio Ambiente	69 / 84	94 / 115	96 / 115	100 / 115	96 / 115	100 / 115	99 / 115	94 / 115	98 / 115	94 / 115
Mobilidade	53 / 61	64 / 73	64 / 73	62 / 73	64 / 73	62 / 73	63 / 73	63 / 73	64 / 73	64 / 73
Saúde	46 / 53	55 / 67	55 / 67	54 / 67	56 / 67	58 / 67	53 / 67	56 / 67	55 / 67	54 / 67
Segurança	235 / 260	327 / 364	329 / 364	319 / 364	328 / 364	325 / 364	326 / 364	329 / 364	324 / 364	321 / 364
Tecnologia & Inovação	4 / 5	6 / 7	5 / 7	5 / 7	5 / 7	6 / 7	5 / 7	6 / 7	5 / 7	6 / 7
Urbanismo	35 / 37	39 / 43	40 / 43	40 / 43	39 / 43	40 / 43	41 / 43	41 / 43	42 / 43	40 / 43

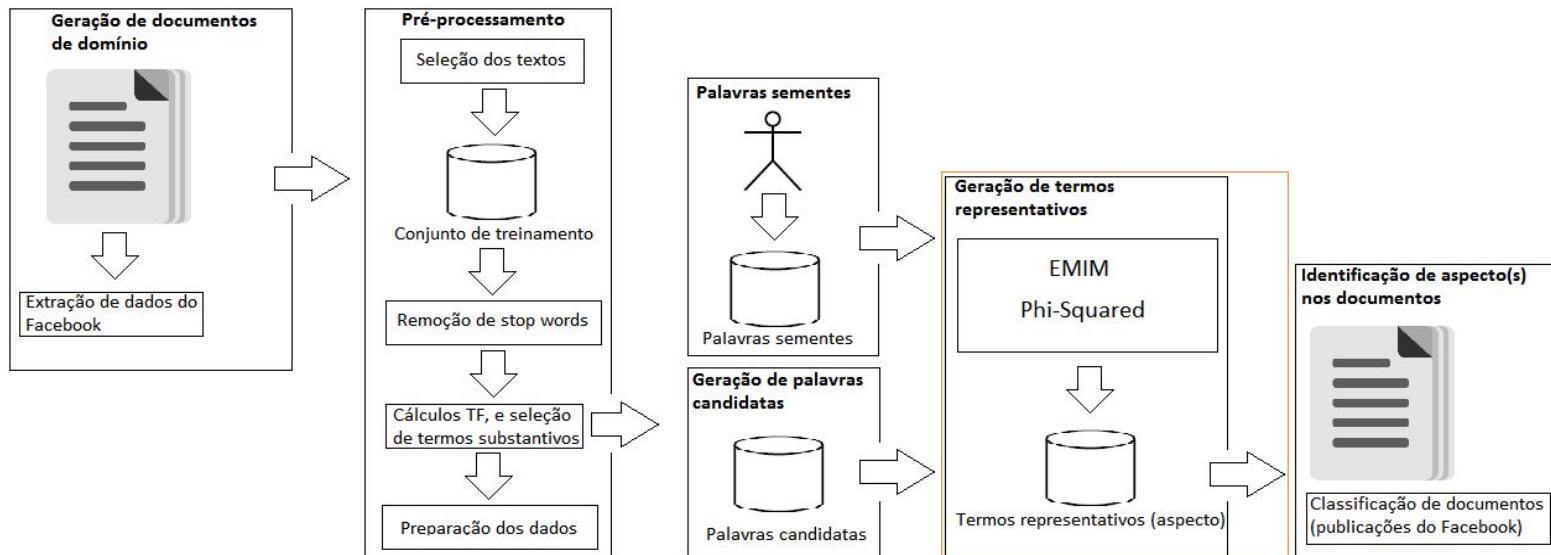
MODELO COMPUTACIONAL

- Palavras candidatas

Quantidade de palavras candidatas geradas por experimento

Aspecto	Experimentos									
	I	II	III	IV	V	VI	VII	VIII	IX	X
Economia	1220	1322	1312	1383	1284	1443	1472	1430	1338	1367
Educação	1955	2079	2210	2236	2303	2202	2198	2269	2228	2247
Empreendedorismo	75	278	272	244	260	260	148	265	140	124
Energia	294	343	347	355	372	363	192	360	324	258
Governança	1755	1888	1860	1936	1868	1995	1902	2021	2031	2044
Meio Ambiente	953	1147	1075	1048	1137	1143	1030	1131	1202	1170
Mobilidade	543	673	682	739	693	721	727	716	684	732
Saúde	868	1161	1115	1069	1076	1022	1062	1098	1106	976
Segurança	1941	2500	2505	2434	2478	2470	2498	2430	2413	2478
Tecnologia & Inovação	183	226	238	130	230	219	140	223	145	243
Urbanismo	487	548	597	554	593	579	583	465	550	554

MODELO COMPUTACIONAL



MODELO COMPUTACIONAL

- Geração de termos representativos

$$EMIM_{(x,y)} = \log_2 \frac{a(a + b + c + d)}{(a + b)(a + c)} > 0$$

$$\phi^2 = \frac{(ad - bc)^2}{(a + b)(a + c)(b + d)(c + d)}, \text{ onde: } 0 \leq \phi^2 \leq 1$$

a = quantidade de vezes que as palavras x e y co-ocorrem em um documento;

b = quantidade de vezes em que x ocorre em um documento e y não ocorre;

c = quantidade de vezes em que y ocorre em um documento e x não ocorre;

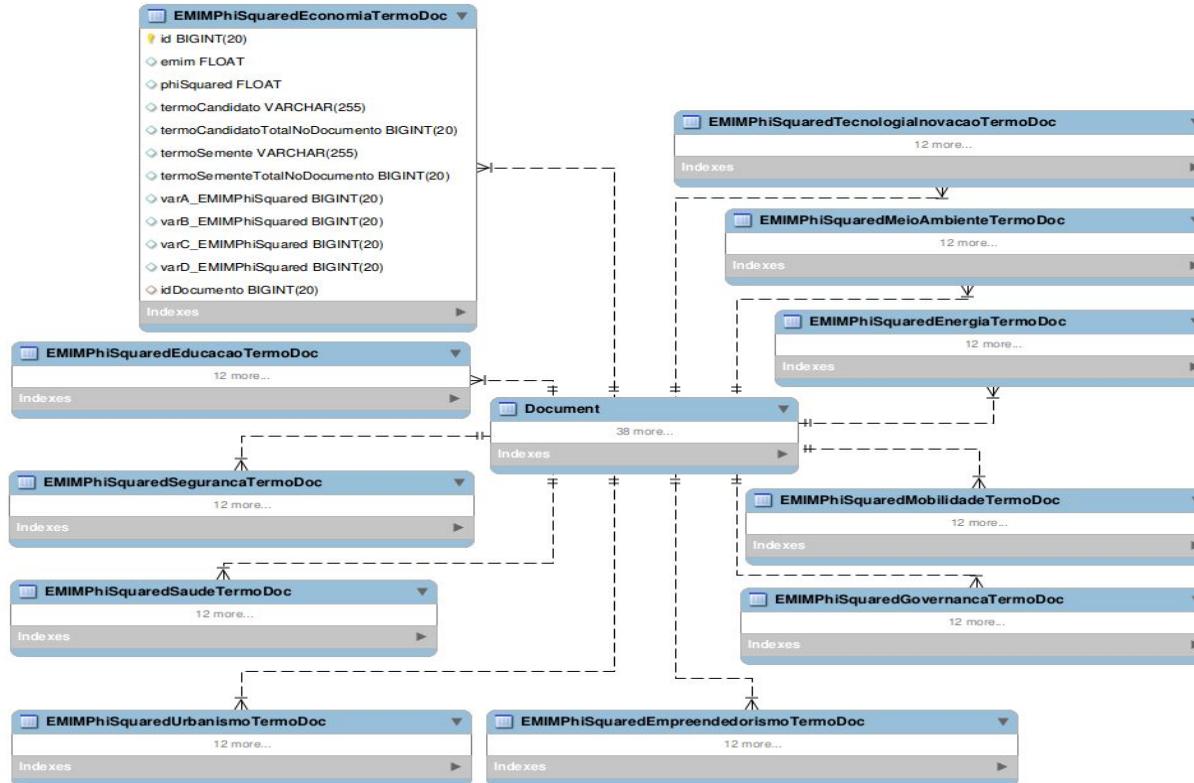
d = tamanho da coleção de documentos menos o número de documentos que não contenham x e/ou y, ou seja, a quantidade de documentos que contenham x e/ou y.

x -> palavra candidata

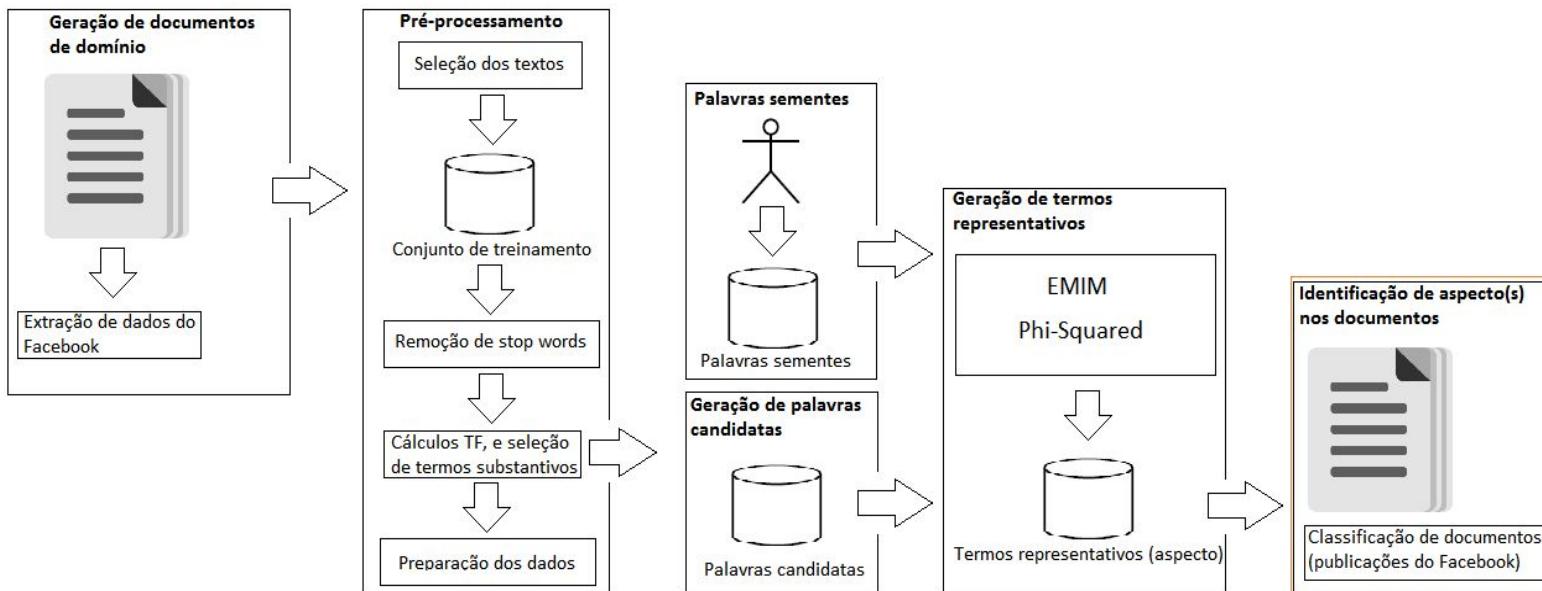
y -> palavra semente anotada

MODELO COMPUTACIONAL

- Geração de termos representativos



MODELO COMPUTACIONAL



RESULTADOS - ESTUDO DE CASO

- ❑ Publicações coletadas de 40 grupos e páginas relacionadas às informações do município do Rio Grande (53579)
- ❑ Total de 1496 anotações

Aspecto	Quantidade
Economia	157
Educação	318
Empreendedorismo	11
Energia	32
Governança	143
Meio Ambiente	143
Mobilidade	91
Saúde	83
Segurança	455
Tecnologia & Inovação	10
Urbanismo	53

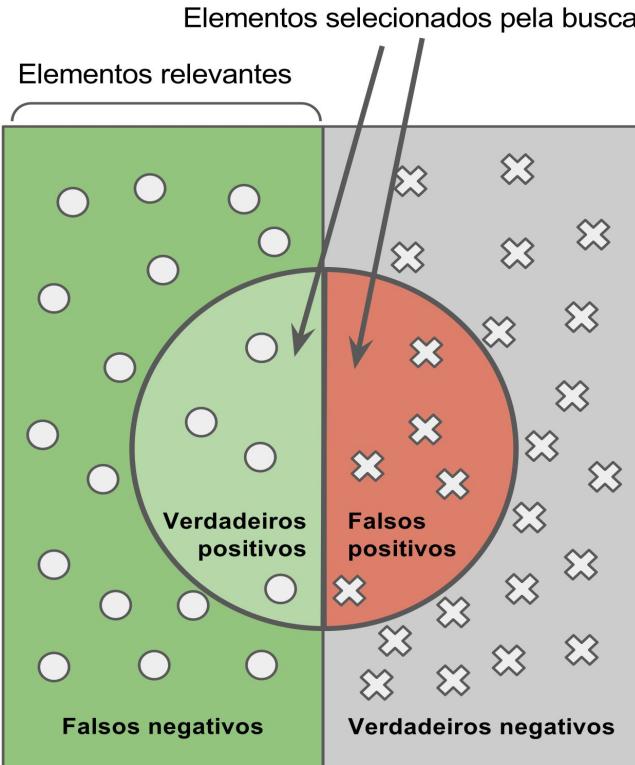
ESTUDO DE CASO

- Métricas para avaliação de um classificador

- ❑ **Acurácia:** proporção de resultados classificados corretamente sobre o total de registros analisados pelo classificador.
- ❑ **Precisão:** métrica para analisar a quantidade de erros falsos positivos cometidos pelo classificador.
- ❑ **Revocação:** mede a fração de exemplos verdadeiros positivos, ou seja, a fração de documentos classificados corretamente.
- ❑ **F-Score:** representa a média harmônica entre a precisão e a revocação.

ESTUDO DE CASO

- Avaliação do classificador (matriz de confusão)



$$\text{Precisão} = \frac{\text{Verdadeiros positivos}}{\text{Falsos positivos} + \text{Verdadeiros positivos}}$$



$$\text{Revocação} = \frac{\text{Verdadeiros positivos}}{\text{Falsos negativos} + \text{Verdadeiros positivos}}$$



"Quantos elementos selecionados são relevantes?"

"Quantos elementos relevantes foram selecionados?"

RESULTADOS - ESTUDO DE CASO

- Avaliação do classificador no corpus de validação

Aspecto	Experimento	Termos	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-score	Acurácia
Economia	IV	SR ⁷	31	266	22	3	263	9	8.800	7.097	7.857	9.596
Educação	VIII	R ⁸	63	234	53	5	229	10	9.138	8.413	8.760	9.495
Empreendedorismo	III	SR ⁷	3	294	2	0	294	1	10.0	6.667	8.0	9.966
Energia	V	R ⁸	6	291	5	0	291	1	10.0	8.333	9.091	9.966
Governança	VIII	R ⁸	28	269	23	11	258	5	6.765	8.214	7.419	9.461
Meio Ambiente	III	SR ⁷	28	269	24	1	268	4	9.600	8.571	9.057	9.832
Mobilidade	IX	SR ⁷	18	279	12	2	277	6	8.571	6.667	7.500	9.731
Saúde	IV	SR ⁷	16	281	11	0	281	5	10.0	6.875	8.148	9.832
Segurança	IX	R ⁸	91	206	80	7	199	11	9.195	8.791	8.989	9.394
Tecnologia & Inovação	III	S ⁹	3	294	2	1	293	1	6.667	6.667	6.667	9.933
Urbanismo	VII	R ⁸	10	287	6	0	287	4	10.0	6.0	7.5	9.865

¹ Documentos Relevantes

² Documentos Não Relevantes

³ Verdadeiros Positivos

⁴ Falsos Positivos

⁵ Verdadeiros Negativos

⁶ Falsos Negativos

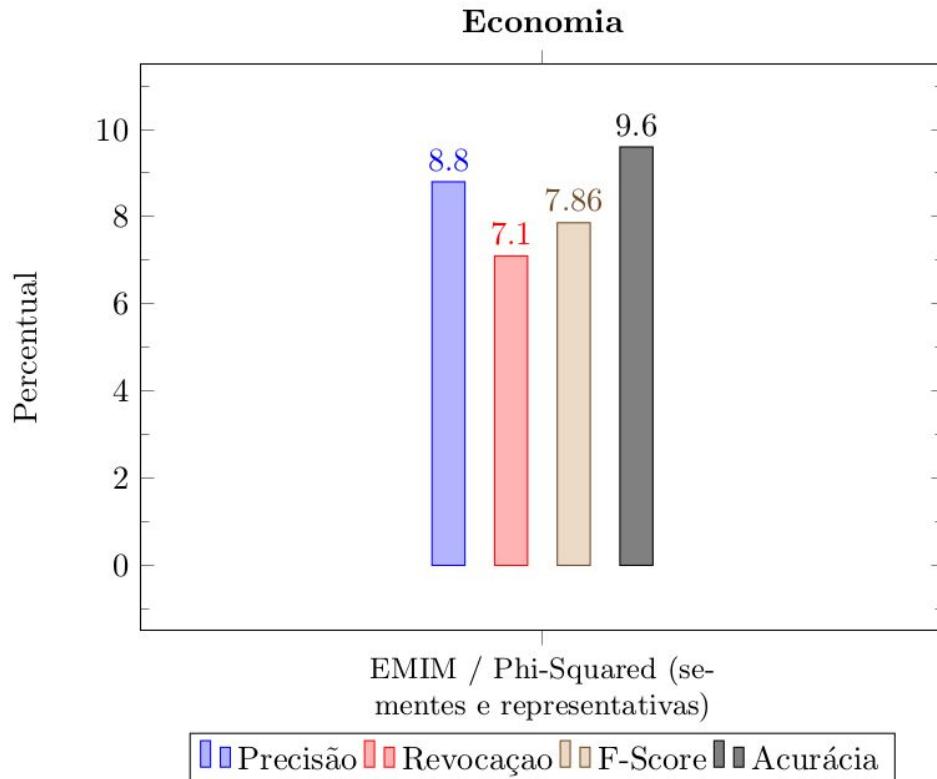
⁷ Sementes e representativas

⁸ Representativas

⁹ Sementes

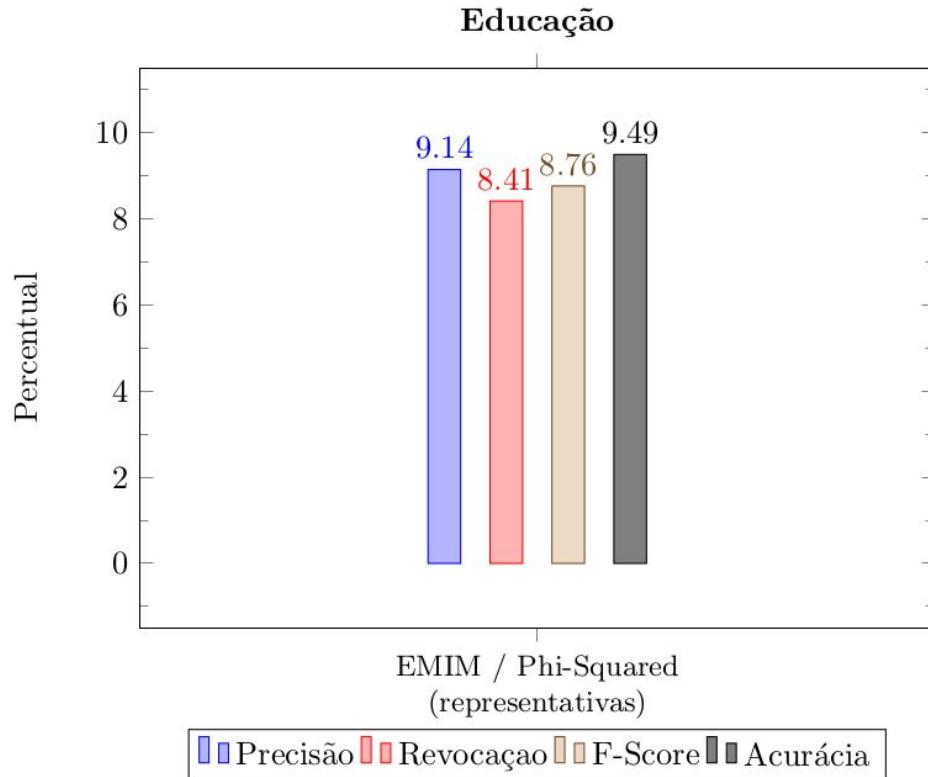
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



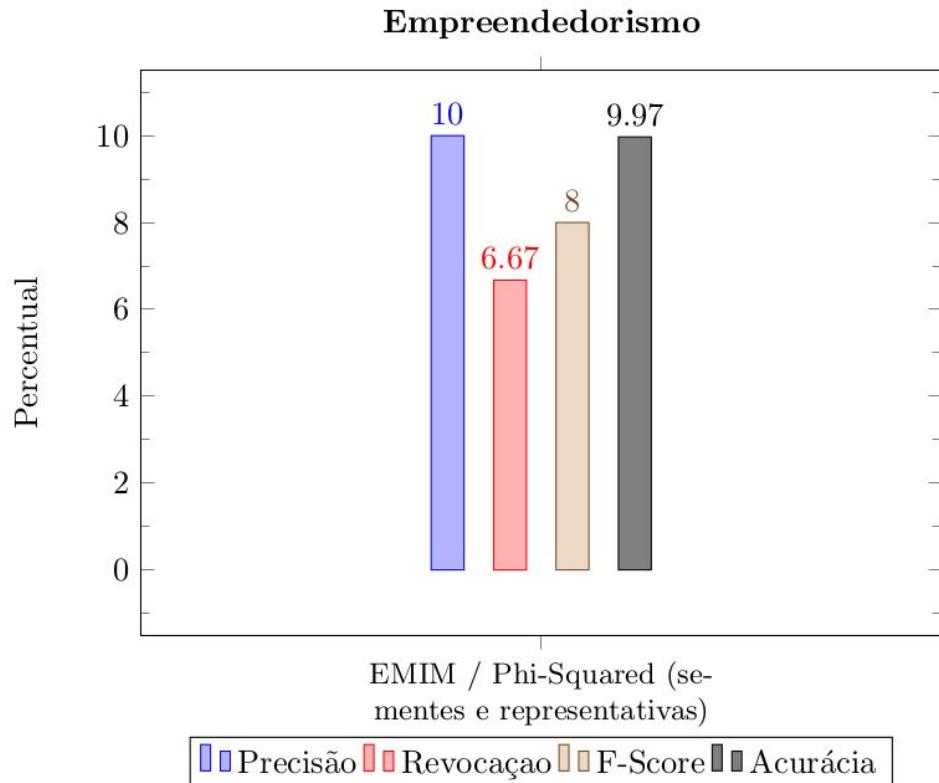
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



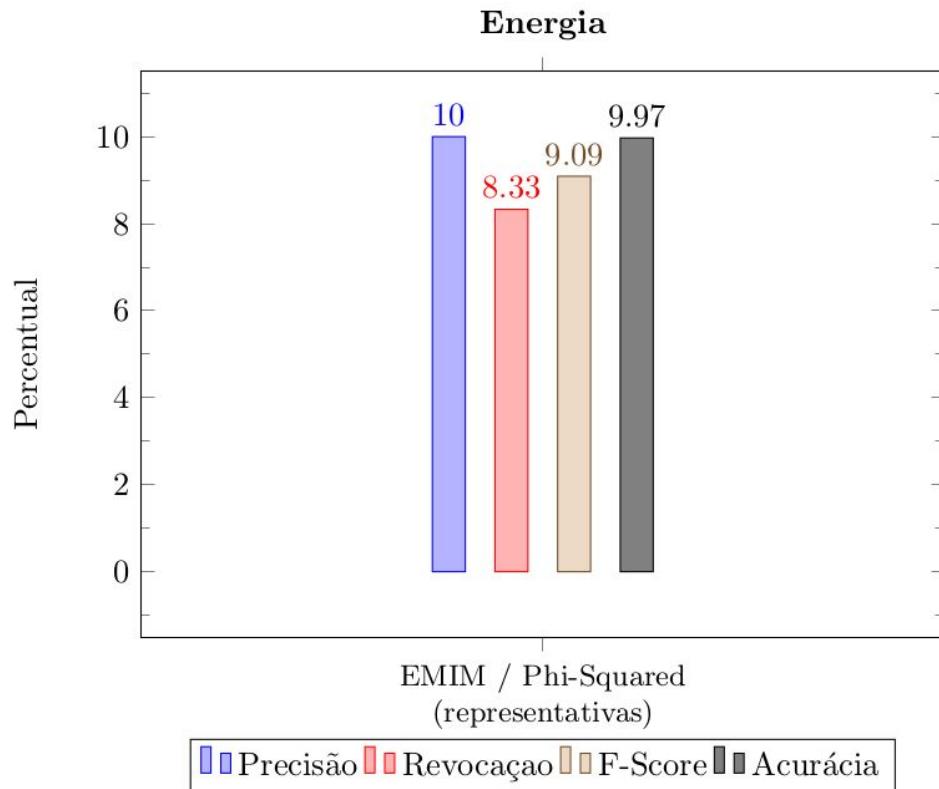
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



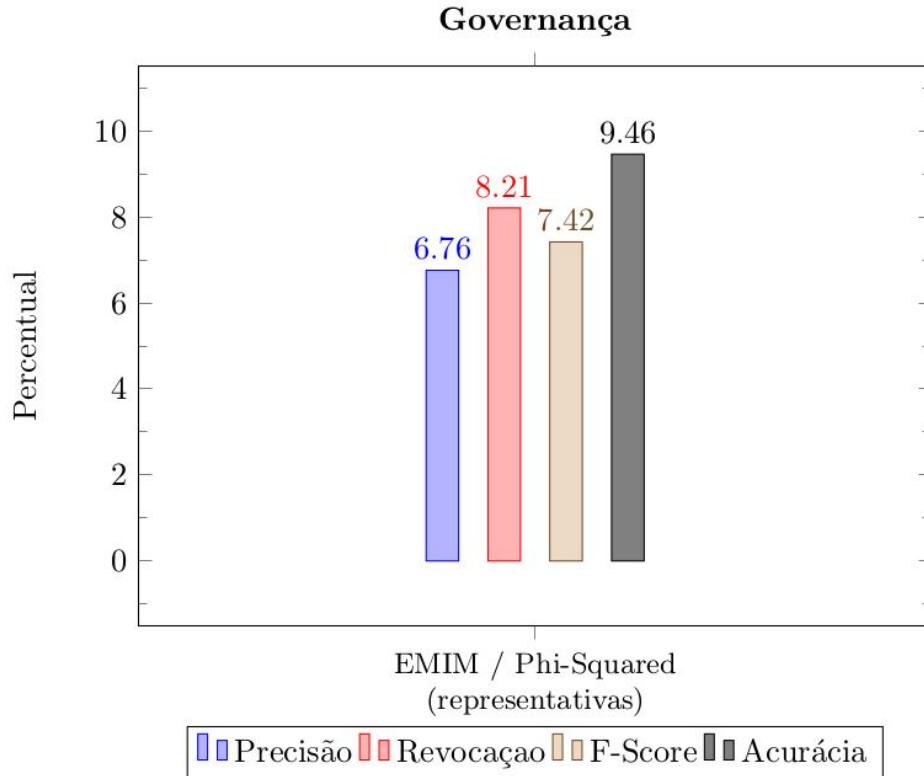
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



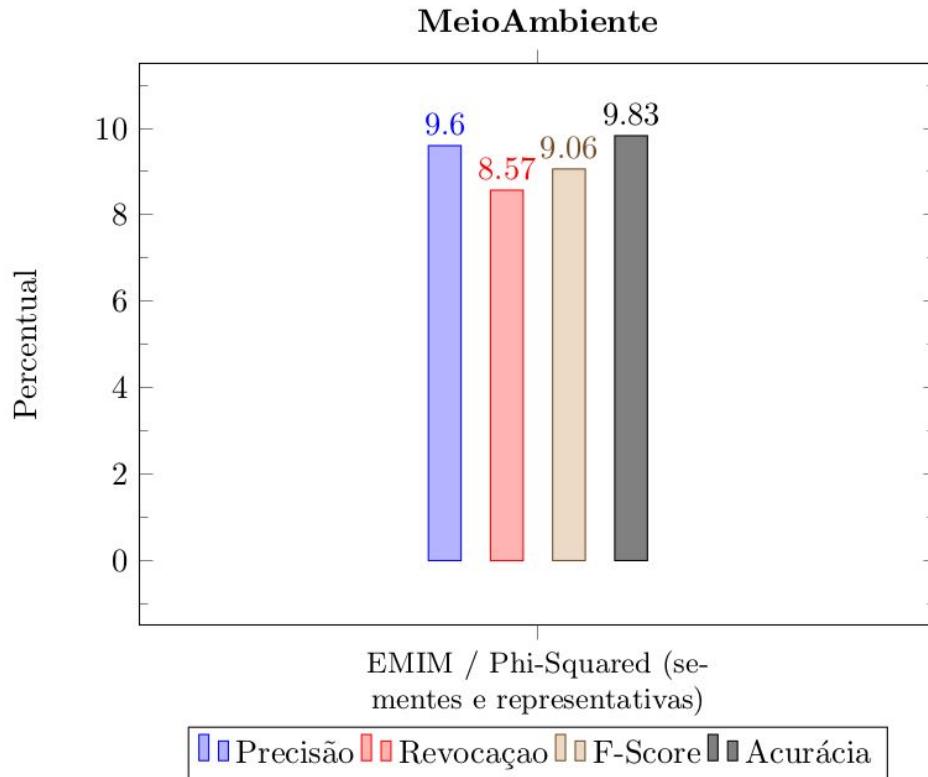
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



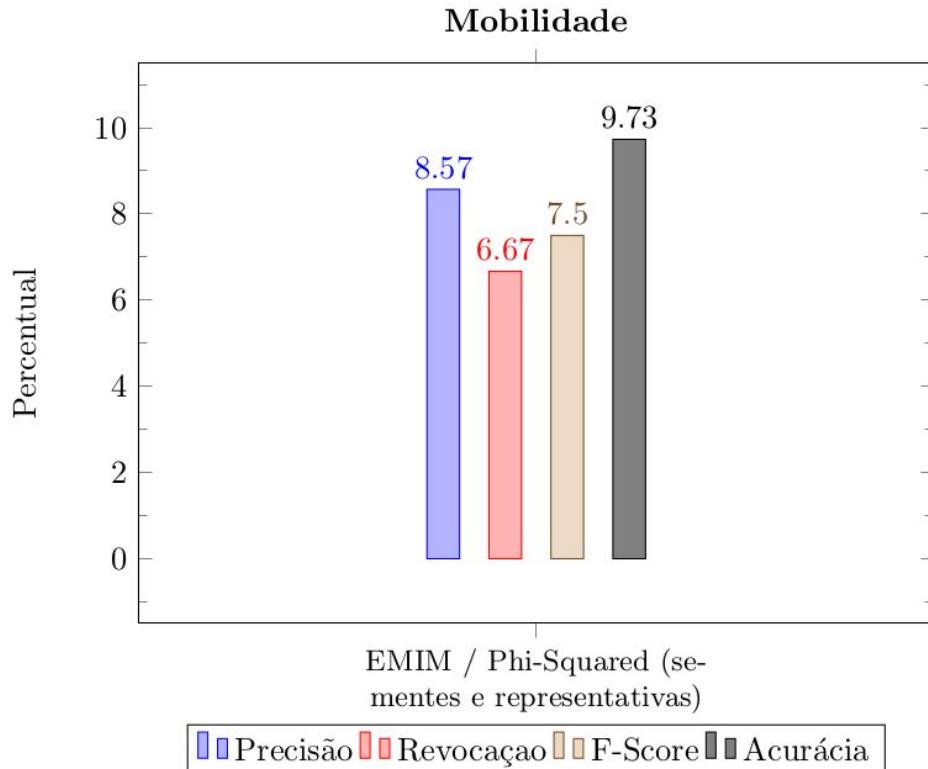
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



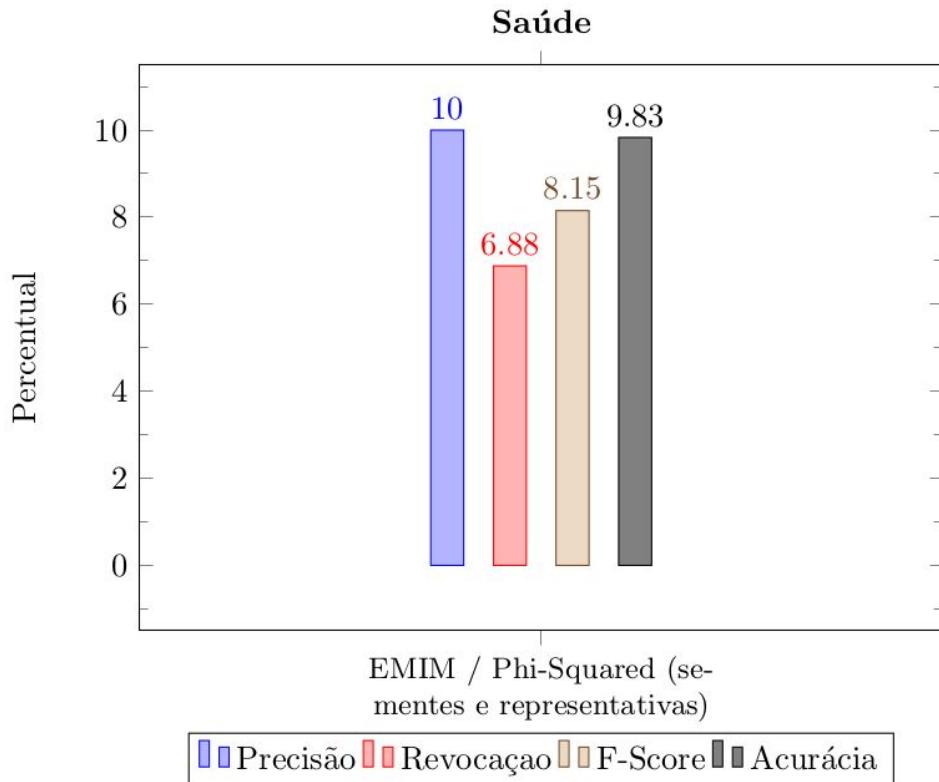
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



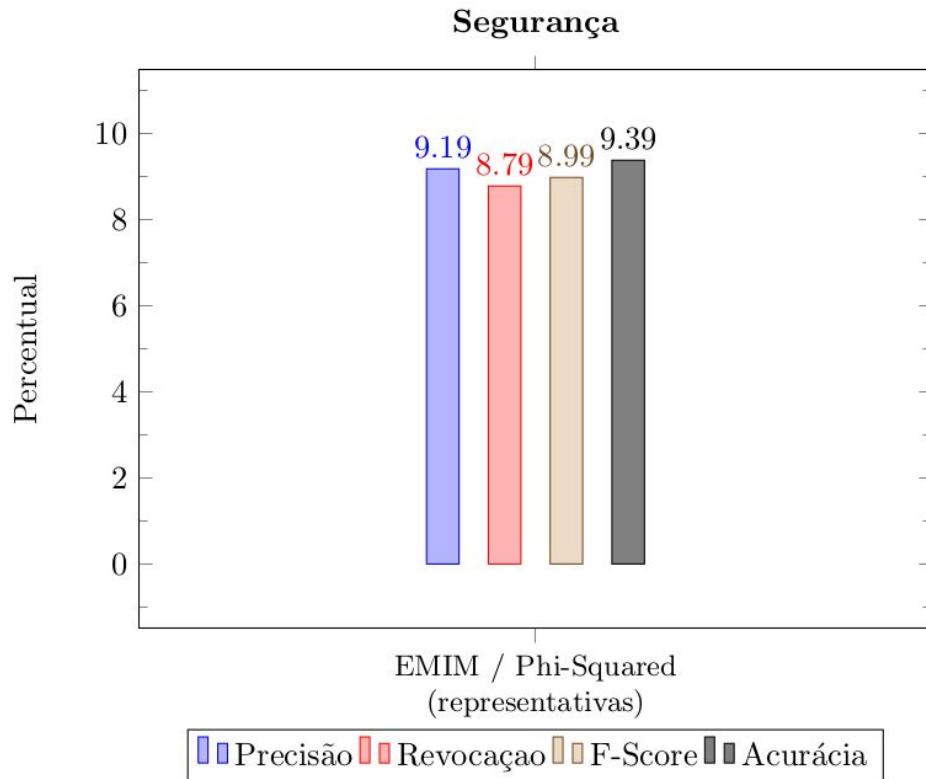
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



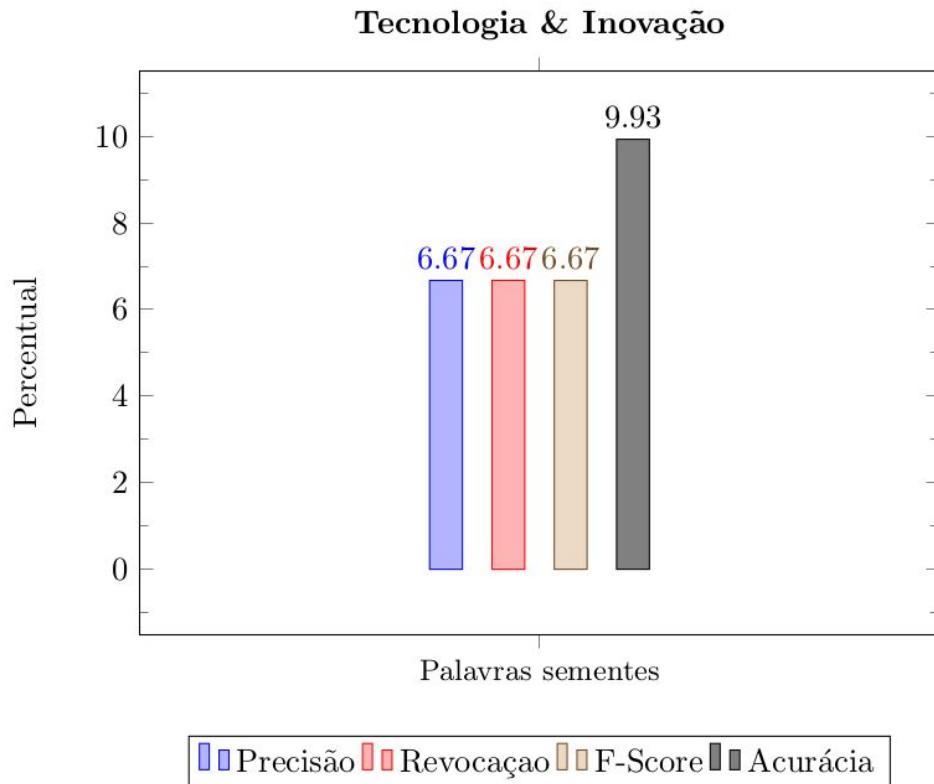
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



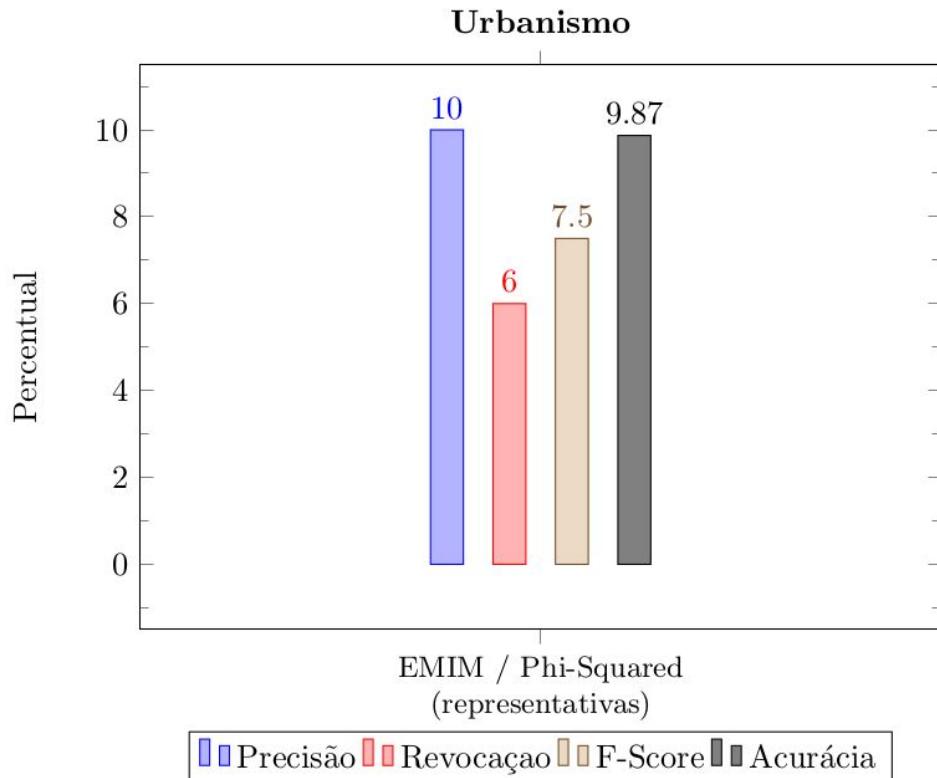
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



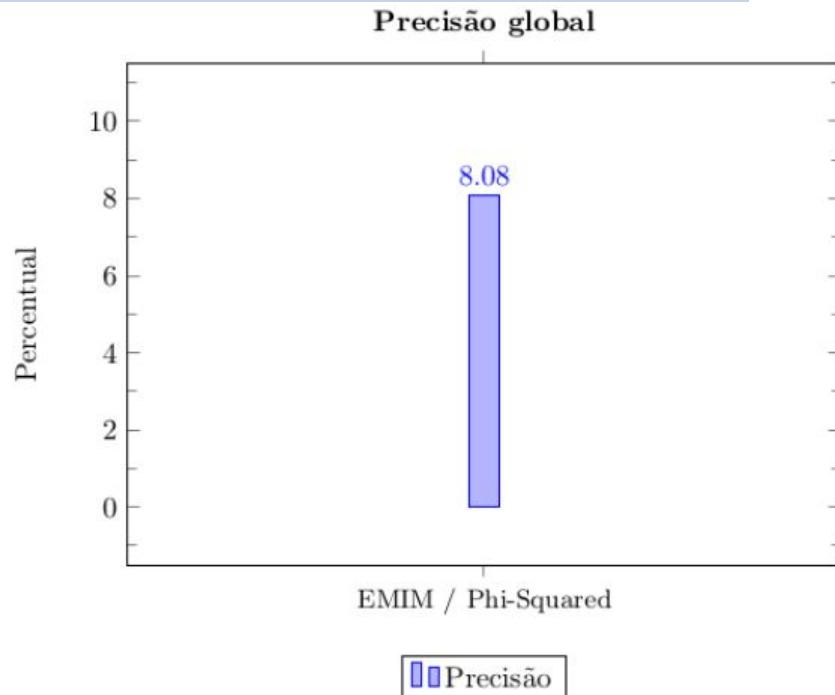
RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



RESULTADOS - ESTUDO DE CASO

- Métricas de avaliação do classificador no corpus de validação



Quantidade de documentos: 297
Verdadeiros positivos: 240

RESULTADOS - ESTUDO DE CASO

- Avaliação do classificador no corpus anotado

Aspecto	DR ¹	DNR ²	VP ³	FP ⁴	VN ⁵	FN ⁶	Precisão	Revocação	F-scrore	Acurácia
Economia	157	1339	136	37	1302	21	7.86	8.66	8.24	9.61
Educação	318	1178	258	24	1154	60	9.15	8.11	8.6	9.44
Empreendedorismo	11	1485	6	4	1481	5	6.0	5.45	5.71	9.94
Energia	32	1464	23	0	1464	9	10.0	7.19	8.36	9.94
Governança	143	1353	119	50	1303	24	7.04	8.32	7.63	9.51
Meio Ambiente	143	1353	117	12	1341	26	9.07	8.18	8.6	9.75
Mobilidade	91	1405	65	30	1375	26	6.84	7.14	6.99	9.63
Saúde	83	1413	57	9	1404	26	8.64	6.87	7.65	9.77
Segurança	455	1041	362	24	1017	93	9.38	7.96	8.61	9.22
Tecnologia & Inovação	10	1486	0	0	1486	10	0	0	0	9.93
Urbanismo	53	1443	24	4	1439	29	8.57	4.53	5.93	9.78

¹ Documentos Relevantes

² Documentos Não Relevantes

³ Verdadeiros Positivos

⁴ Falsos Positivos

⁵ Verdadeiros Negativos

⁶ Falsos Negativos

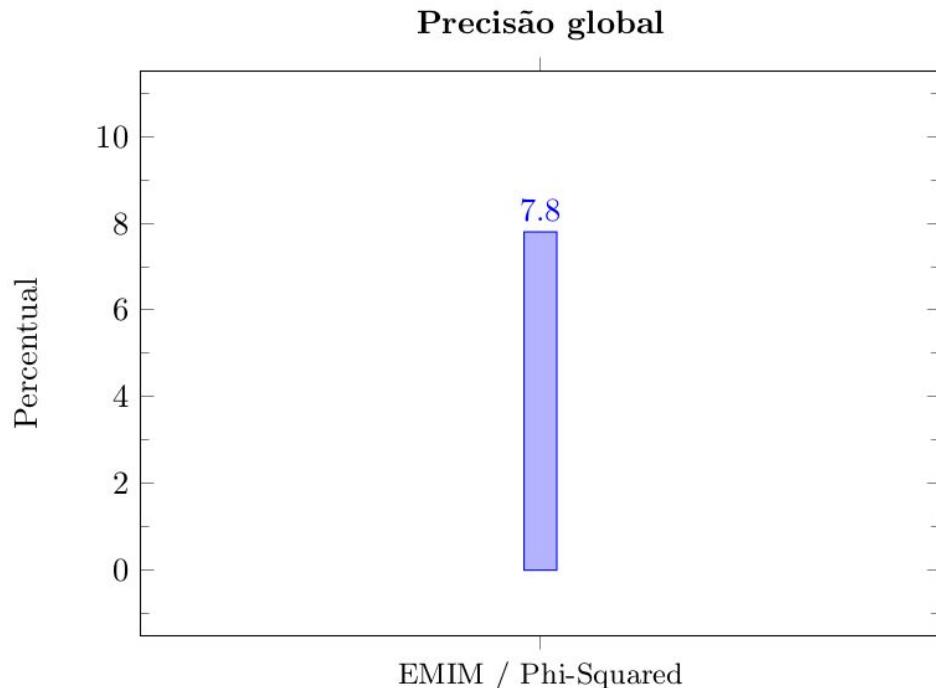
⁷ Sementes e representativas

⁸ Representativas

⁹ Sementes

RESULTADOS - ESTUDO DE CASO

- Avaliação do classificador no corpus anotado



RESULTADOS - ESTUDO DE CASO

- Classificação de documentos

Aspecto	Quantidade
Economia	8462
Educação	9098
Empreendedorismo	341
Energia	705
Governança	6217
Meio Ambiente	2577
Mobilidade	2754
Saúde	1710
Segurança	9726
Tecnologia & Inovação	78
Urbanismo	394
Não classificados	11517
Total de documentos	53579

RESULTADOS - ESTUDO DE CASO

- Nuvem de palavras do aspecto Educação a partir do corpus anotado



RESULTADOS - ESTUDO DE CASO

- Nuvem de palavras do aspecto Mobilidade a partir do corpus anotado



RESULTADOS - ESTUDO DE CASO

- Nuvem de palavras do aspecto Mobilidade a partir do corpus classificado



CONSIDERAÇÕES FINAIS

- ❑ Os resultados obtidos apresentam indícios que o método pode ser aplicado, uma vez que as métricas de avaliação foram adequadas;
- ❑ Melhores resultados foram encontrados para os aspectos com maior número de anotações;
- ❑ Como trabalhos futuros propõe-se a expansão do *dataset* e mineração de opiniões nos documentos classificados;
- ❑ Os termos representativos e as classificações obtidas neste estudo, podem ser utilizados em etapa de pós-processamento para a obtenção de um discurso coletivo sobre a percepção do cidadão em relação ao nível de inteligência da sua cidade.

REFERÊNCIAS

AGGARWAL, C. C.; ZHAI, C. *Mining text data*. [S.I.]: Springer Science & Business Media, 2012. Citado na página 47.

ANDRADE, D. C. de. Priorização de nós de uma rede de sensores sem fio com base na análise de dados de redes sociais. *Anais Seminário de Iniciação Científica*, n. 21, 2017. Citado 4 vezes nas páginas 17, 58, 59 e 60.

BARALIS, E.; CAGLIERO, L.; MAHOTO, N.; FIORI, A. Graphsum: Discovering correlations among multiple terms for graph-based summarization. *Information Sciences*, Elsevier, v. 249, p. 96–109, 2013. Citado na página 47.

BERRONE, P.; RICART, J. E.; CARRASO, C.; RICART, R. *Iese: Cities in motion index 2016. IESE, University of Navarra Business School, New York, USA*, 2016. Citado na página 34.

CABENA, P.; HADJINIAN, P.; STADLER, R.; VERHEES, J.; ZANASI, A.; JOSE, C. I. B. M. C. S.; JOSE, C. I. T. S. O. S. *Discovering data mining: from concept to implementation*. [S.I.]: Prentice Hall PTR New Jersey, 1997. Citado na página 27.

CARAGLIU, A.; BO, C. D.; NIJKAMP, P. Smart cities in europe. *Journal of urban technology*, Taylor & Francis, v. 18, n. 2, p. 65–82, 2011. Citado na página 33.

CHURCH, K. W.; GALE, W. A. Concordances for parallel text. In: *Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research*. [S.I.: s.n.], 1991. p. 40–62. Citado 2 vezes nas páginas 48 e 49.

CHURCH, K. W.; HANKS, P. Word association norms, mutual information, and lexicography. *Computational linguistics*, MIT Press, v. 16, n. 1, p. 22–29, 1990. Citado na página 48.

CHURCH, K. W.; MERCER, R. L. Introduction to the special issue on computational linguistics using large corpora. *Computational linguistics*, MIT Press, v. 19, n. 1, p. 1–24, 1993. Citado na página 48.

CORDEIRO, A. D. et al. Gerador inteligente de sistemas com auto-aprendizagem para gestão de informações e conhecimento. Florianópolis, SC, 2005. Citado na página 41.

DIAS, L. de C. Análise de redes sociais com uso de aprendizado de máquina para prever o tráfego de veículos em zonas urbanas. 2017. Citado 6 vezes nas páginas 17, 53, 54, 55, 56 e 57.

EBECKEN, N. F.; LOPES, M. C. S.; COSTA, M. C. et al. Mineração de textos. *Sistemas inteligentes: fundamentos e aplicações*. São Carlos: Manole, p. 337–370, 2003. Citado na página 43.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado na página 38.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, ACM, v. 39, n. 11, p. 27–34, 1996. Citado na página 39.

REFERÊNCIAS

FIGUEREDO, M.; SOUSA, A. Utilizando redes sociais para detectar crimes em tempo real. 2016. Citado 2 vezes nas páginas 17 e 52.

GIFFINGER, R.; GUDRUN, H. Smart cities ranking: an effective instrument for the positioning of the cities? *ACE: architecture, city and environment*, Centre de Política del Sòl i Valoracions-Universitat Politècnica de Catalunya, v. 4, n. 12, p. 7–26, 2010. Citado na página 33.

HALL, R. E.; BOWERMAN, B.; BRAVERMAN, J.; TAYLOR, J.; TODOSOW, H.; WIMMERSPERG, U. V. *The vision of a smart city*. [S.I.], 2000. Citado 2 vezes nas páginas 28 e 33.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.I.]: Elsevier, 2011. Citado 2 vezes nas páginas 17 e 36.

HARRISON, C.; DONNELLY, I. A. A theory of smart cities. In: *Proceedings of the 55th Annual Meeting of the ISSS-2011*, Hull, UK. [S.I.: s.n.], 2011. v. 55, n. 1. Citado na página 33.

JUNIOR, J. R. C. Desenvolvimento de uma metodologia para mineração de textos. *Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro*, 2007. Citado 2 vezes nas páginas 41 e 43.

KHATOUN, R.; ZEADALLY, S. Smart cities: concepts, architectures, research opportunities. *Communications of the ACM*, ACM, v. 59, n. 8, p. 46–57, 2016. Citado na página 34.

KOMINIOS, N. *Intelligent cities: innovation, knowledge systems and digital spaces*. [S.I.]: Routledge, 2013. Citado na página 33.

LEVENSHTEIN, V. I. Binary codes capable of correcting deletions, insertions, and reversals. In: *Soviet physics doklady*. [S.I.: s.n.], 1966. v. 10, n. 8, p. 707–710. Citado na página 54.

LOPES, M. C. S. Mineração de dados textuais utilizando técnicas de clustering para o idioma português. *Rio de Janeiro: sm*, 2004. Citado na página 40.

MADEIRA, R. d O. C. *Aplicação de técnicas de mineração de texto na detecção de discrepâncias em documentos fiscais*. Tese (Doutorado), 2015. Citado na página 37.

MANNING, C. D.; MANNING, C. D.; SCHÜTZE, H. *Foundations of statistical natural language processing*. [S.I.]: MIT press, 1999. Citado na página 48.

MORAIS, E. A. M.; AMBRÓSIO, A. P. L. Mineração de textos. *Relatório Técnico-Instituto de Informática (UFG)*, 2007. Citado 8 vezes nas páginas 17, 37, 41, 42, 43, 44, 45 e 46.

MOURA, M. Proposta de utilização de mineração de textos para seleção, classificação e qualificação de documentos. *Embrapa Informática Agropecuária-Documentos (INFOTECA-E)*, Campinas: Embrapa Informática Agropecuária, 2004., 2004. Citado na página 40.

PALAZZO, M.; LOH, S.; AMARAL, L.; WIVES, L. Descoberta de conhecimento em textos através da análise de seqüências temporais. In: *Workshop em Algoritmos e Aplicações de Mineração de Dados-WAAMD, SBD: Sociedade Brasileira de Computação*. [S.I.: s.n.], v. 2, p. 49–56. Citado na página 39.

PASSOS, E.; GOLDSCHMIDT, R. Data mining: um guia prático. *Editora Campus*, Rio de Janeiro, 2005. Citado 2 vezes nas páginas 17 e 38.

REFERÊNCIAS

RAMOS, H. d. S. C.; BRÄSCHER, M. Aplicação da descoberta de conhecimento em textos para apoio à construção de indicadores infométricos para a área de c&t. *Ciência da Informação*, v. 38, n. 2, 2009. Citado na página 27.

SALTON, G.; MCGILL, M. J. *Introduction to modern information retrieval*. McGraw-Hill, Inc., 1986. Citado na página 41.

SÁPIRAS, L. A. Mineração de opiniões em aspectos em fontes de opiniões fracamente estruturadas. 2015. Citado na página 48.

SCHAEFFER, C. L. Monitoramento de mídias sociais no processo de análise da percepção dos usuários sobre a segurança de porto alegre. 2016. Citado na página 51.

SÉRGIO, M. C.; SILVA, T. do Nascimento da; GONÇALVES, A. L. Descoberta de conhecimento a partir de informações não estruturadas por meio de técnicas de correlação e associação. *Em Questão*, Universidade Federal do Rio Grande do Sul, v. 22, n. 2, 2016. Citado na página 47.

STATISTA, A. The statistics portal. *Web site: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/*, 2017. Citado 2 vezes nas páginas 17 e 30.

STATISTA, A. The statistics portal. *Web site: https://www.statista.com/statistics/244936/number-of-facebook-users-in-brazil/*, 2017. Citado 3 vezes nas páginas 17, 30 e 62.

STATISTA, A. The statistics portal. *Web site: https://www.statista.com/statistics/254734/most-popular-social-networking-sites-in-brazil/*, 2017. Citado 2 vezes nas páginas 17 e 31.

SYSTEM, U. *Ranking Connected Smart Cities*. 2017. [Accessado 31-08-2018]. Disponível em: <<http://www.connectedsmartcities.com.br/resultados-do-ranking-connected-smart-cities/?lang=en>>. Citado 2 vezes nas páginas 28 e 34.

SYSTEMS, A. U. Inteligencia de mercado | urban systems brasil. *Web site: https://www.urbansystems.com.br*, 2018. Citado na página 34.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to data mining*. 1st. [S.I.]: Boston: Pearson Addison Wesley. xxi, 2005. Citado 2 vezes nas páginas 37 e 47.

VELOSO, R. A. Geopantin: Uma ferramenta de rede social baseada em localização inspirada em soluções de cidades inteligentes. 2012. Citado na página 57.

WEISS, M. C. Cidades inteligentes como nova prática para o gerenciamento dos serviços e infraestruturas urbanas: estudo de caso da cidade de porto alegre/marcos cesar weissl; orientação: Profº drº roberto carlos bernardes. 2013. Citado na página 28.

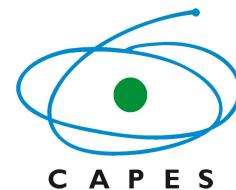
WIVES, L. Recursos de text mining. Disponível por www.inf.ufrgs.br/~wives/portugues/textmining.html (15 de maio de 2001), 2005. Citado na página 39.

WIVES, L. K. Tecnologias de descoberta de conhecimento em textos aplicadas à inteligência competitiva. *Exame de Qualificação EQ-069, PPGC-UFRGS*, 2002. Citado 4 vezes nas páginas 37, 41, 44 e 46.

REFERÊNCIAS

ZHANG, C.; ZHOU, G.; YUAN, Q.; ZHUANG, H.; ZHENG, Y.; KAPLAN, L.; WANG, S.; HAN, J. Geoburst: Real-time local event detection in geo-tagged tweet streams. In: ACM. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. [S.I.], 2016. p. 513–522. Citado na página 59.

AGRADECIMENTOS





Obrigado!





Perguntas?

