

A

Resultados Detalhados

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	82.4	81.7	83.2	79.6
500	82.3	82.5	84.9	82.6
1000	82.0	81.1	86.1	83.2
2000	81.3	80.8	86.3	82.0
3000	80.1	79.6	86.3	81.9
5000	81.8	81.0	87.2	81.7
8000	81.8	80.8	87.3	81.0
10000	82.1	81.4	86.8	81.5
12000	81.3	80.6	86.6	80.2
14000	82.3	80.8	86.7	78.8
16000	82.5	81.3	86.7	79.2
18000	82.1	81.7	86.2	79.6
20000	82.5	81.3	86.5	79.2
22000	82.3	80.9	86.3	79.3
24000	82.2	81.4	86.4	78.9
26000	82.1	81.1	86.2	78.3
30000	82.1	81.1	86.4	78.4

Tabela A.1: Resultados para unigramas do corpus original

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	73.2	73.9	72.9	72.9
500	79.2	77.1	77.6	77.6
1000	79.2	77.9	78.8	78.7
2000	81.0	79.6	80.3	79.8
3000	81.1	79.2	81.4	80.9
5000	81.8	80.6	82.4	82.1
10000	83.1	80.4	82.7	82.4
15000	83.7	81.4	83.5	82.1
20000	81.1	79.3	83.4	82.5
25000	78.6	76.6	82.5	82.0
30000	79.5	77.5	82.8	82.1
35000	83.2	81.0	83.1	82.7
40000	84.9	82.5	83.8	83.1
50000	81.8	80.1	82.9	82.3
60000	86.0	84.5	83.8	83.2
70000	85.3	84.3	83.0	82.7
80000	85.2	84.1	82.6	82.6
90000	83.8	83.3	82.5	81.3
100000	83.4	83.4	82.4	81.2

Tabela A.2: Resultados para bigramas do corpus original

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	81.8	82.0	82.0	78.9
500	82.7	81.1	84.5	80.7
1000	80.7	79.6	84.3	81.7
2000	80.4	79.1	85.2	80.9
3000	80.2	79.6	85.9	80.8
5000	82.1	81.2	86.3	81.4
8000	80.5	79.6	86.3	81.0
10000	82.6	81.1	86.2	80.8
12000	82.7	81.6	86.4	80.6
15000	82.7	81.5	86.6	78.9
18000	83.1	81.9	86.5	79.4
21000	83.6	81.5	86.1	79.2
24000	83.8	81.8	86.4	79.1
27000	83.1	81.4	85.8	78.9
30000	83.2	81.7	85.8	78.0
33000	83.1	81.5	86.0	78.0
36000	82.9	81.3	86.0	78.0

Tabela A.3: Resultados para unigramas com classe gramatical (*part of speech*)

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	74.2	74.6	74.4	74.1
500	77.7	74.4	76.7	77.9
1000	78.6	76.2	78.6	78.1
2000	79.7	77.5	79.2	79.2
3000	76.9	75.7	79.7	79.3
5000	79.6	77.9	80.4	81.6
10000	80.6	79.6	80.9	81.5
15000	80.4	79.0	82.2	82.3
20000	77.9	77.1	81.7	81.8
25000	76.9	76.0	81.5	82.2
30000	79.9	78.6	81.9	82.7
35000	82.9	81.0	82.5	82.7
40000	84.3	82.1	82.7	82.5
50000	82.3	80.5	82.2	82.5
60000	86.3	84.7	82.7	82.6
70000	85.9	84.6	83.0	82.4
80000	85.2	84.3	82.8	82.2
90000	84.5	83.6	82.3	81.8
100000	84.3	83.5	82.4	81.5

Tabela A.4: Resultados para bigramas com classe gramatical (*part of speech*)

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	79.2	80.9	80.8	79.3
500	83.8	83.3	85.6	82.7
1000	84.2	83.0	85.6	84.1
2000	85.2	83.8	85.3	83.1
3000	84.6	83.4	86.5	82.9
5000	84.4	82.8	86.2	82.4
8000	83.9	83.4	85.7	82.1
10000	85.2	84.2	85.5	81.0
12000	84.2	83.6	85.6	80.6
14000	84.1	84.2	85.8	80.7
16000	83.4	84.3	85.4	80.7
18000	83.2	84.1	85.4	80.6
20000	82.9	83.9	85.2	80.6

Tabela A.5: Resultados para unigramas com subjetividade filtrada

Features	NB - Binário	NB - Freq	SVM - Binário	SVM - Freq
100	69.7	72.0	74.0	75.1
500	78.5	79.1	77.3	77.4
1000	81.9	82.4	79.6	80.8
2000	83.3	83.6	81.7	82.2
3000	84.9	84.9	81.6	82.3
5000	84.6	84.2	81.8	82.5
10000	85.6	86.0	82.5	83.3
15000	84.1	83.5	82.5	83.3
20000	84.8	84.2	82.1	83.6
25000	87.6	86.8	83.0	83.9
30000	86.3	85.8	83.2	83.8
35000	85.7	85.3	83.1	84.4
40000	87.8	87.2	84.4	84.7
45000	87.3	86.7	82.9	83.5
50000	86.9	86.6	82.7	83.5
60000	85.7	86.0	82.8	82.8
70000	85.5	85.8	82.6	82.9
80000	84.8	85.1	82.5	82.7

Tabela A.6: Resultados para bigramas com subjetividade filtrada

B

Fluxograma do Ambiente de Experimentação

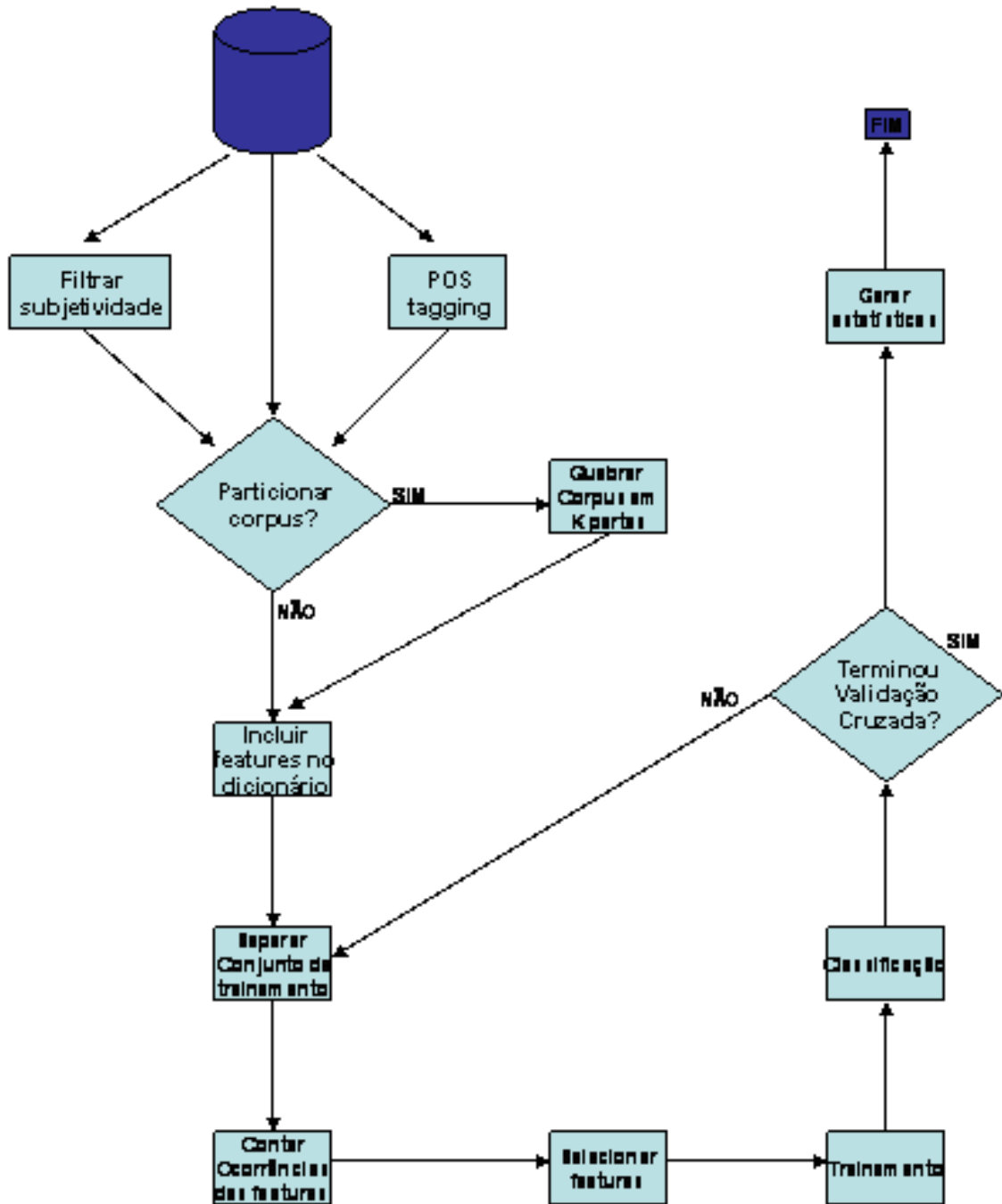


Figura B.1: Fluxograma da Solução

Referências Bibliográficas

- [01] T. Bayes, **An Essay Towards Solving a Problem in the Doctrine of Chances**, Philos. Trans. R. Soc. London, 1763 4
- [02] P. Bernstein, **Against the Gods: The Remarkable Story of Risk**, John Wiley & Sons, New York, US, 1996. 4
- [03] B. Boser, I. Guyon, V. Vapnik, **A Training Algorithm for Optimal Margin Classifiers**, Proc. of the 5th Annual ACM Workshop on Computational Learning Theory, pp.144-152, ACM Press, 1992. 5
- [04] S. Chakrabarti, **Mining the Web: Discovering Knowledge from Hypertext Data**, Morgan Kaufmann, 2002. 3.5, 4, 7
- [05] N. Cristianini and J. Shawe-Taylor, **An Introduction to Support Vector Machines**, Cambridge University Press, Cambridge, UK, 2000. 5.1.1, 5.2.1, 5.2.1
- [06] K. Dave, S. Lawrence and D. Pennock, **Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews**, International World Wide Web Conference, Budapest, Hungary, pp. 519-528, 2003. 2.1
- [07] M. Gamon and A. Anthony, **Automatic identification of sentiment vocabulary: exploiting low association with known sentiment terms**, In Proceedings of the ACL, 2005.
- [08] S. Geisser, **The predictive sample reuse method with applications**, Journal of the American Statistical Association, 70(350), 1975. 6.2
- [09] V. Hatzivassiloglou and K.R. McKeown, **Predicting the semantic orientation of adjectives**, Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics and the 8th Conference of the European Chapter of the ACL (pp. 174-181), New Brunswick, NJ: Association of Computational Linguistics, 1997 2.2
- [10] T. Joachims, **Text Categorization with Support Vector Machines: Learning with Many Relevant Features**, Proc. of the 10th European Conference on Machine Learning, 1998. 6.4

- [11] C.D. Manning and H. Schütze, **Foundations on Statistical Natural Language Processing**. The MIT Press, Cambridge, Massachusetts, 1999.
- [12] S. Matsumoto, H. Takamura and M. Okumura, **Sentiment Classification using Word Sub-Sequences and Dependency Sub-Trees**. Advances in Knowledge Discovery and Data Mining: 9th Pacific-Asia Conference, PAKDD 2005. 2.2, 6.1
- [13] A. McCallum and K. Nigam, **A comparison of event models for Naive bayes text classification**, In Proc. of the AAAI-98 Workshop on learning for text categorization, pp. 41-48. 2.2, 4
- [14] T.M. Mitchell, **Machile Learning**, McGraw Hill, New York, US, 1996.
- [15] R. Mukras, **A Comparison Of Machine Learning Techniques Applied to Sentiment Classification**, M.Sc. Thesis, University of Sussex, UK, 2004. 6.1
- [16] T. Mullen and N. Collier, **Sentiment Analysis using Support Vector Machines with Diverse Information Sources**, Proc. of the 9th EMNLP, pp.412-418, 2004.
- [17] B. Pang, L. Lee, **A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts**, Proc. of the ACL, 2004. 1.1, 2.2, 2.2, 3.4, 6.3, 6.3.1, 6.4.1
- [18] B. Pang and L. Lee.. **Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rating scales**. In Proceedings of the ACL, 2005. 6.1
- [19] B. Pang, L. Lee and S. Vaithyanathan, **Thumbs up? Sentiment classification using machine learning techniques**, In Proc. of the EMNLP, 2002. (document), 1.1, 2.2, 2.1, 6.1, 6.4.1
- [20] F. Peng, D. Shuurmans and S. Wang, **Augumenting naive bayes text classifiers with statistical language models**, Information Retrieval, 7 (3-4), pp. 317-345, 2004.
- [21] E. Rilloy, J. Wieb and W. Philips, **Exploiting subjectivity classification to improve information extraction**, In AAAI-2005. 3.4
- [22] S. Russell and P. Norvig, **Artificial Intelligence: A Modern Approach**, 4, 4.1.2, 6
- [23] K. Toutanova, D. Klein, C. Manning, and Y. Singer, **Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network**, In Proc. of HLT-NAACL, pp. 252-259, 2003 3.3

- [24] P. Turney, **Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification reviews**, In Proc. of the ACL, 2002. Prentice Hall, Englewood Cliffs, NJ, 1995 2.2
- [25] P. Turney and M. Littman, **Un-supervised learning of semantic orientation from a hundred-billion-word corpus**, Technical Report EGB-1094, National Research Council Canada, 2002.
- [26] C. Whitelaw, N. Garg and S. Argamon, **Using appraisal taxonomies for sentiment analysis**, In Proc. Second Midwest Computational Linguistic Colloquium, 2005. 2.2, 6.1
- [27] I. Witten, A. Moffat and T.C. Bell, **Managing Gigabytes: Compressing and Indexing Documents and Images 2nd ed.**, Morgan Kaufmann Publishers Inc., San Francisco, US, 1999.
- [28] J. Yi, T. Nasukawa, R. Bunescu and W. Niblac, **Sentiment analyser: Extracting sentiments about a given topic using natural language processing techniques**, In IEEE Intl. Conf. on Data Mining (ICDM), 2003.
- [29] H. Yu and V. Hatzivassiloglou, **Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences**, In Proceedings of EMNLP, 2003. 3.4