

# Bag of words with Bayesian Network to detect languages

Santiago E. Bocel  
Universidad Rafael  
Landívar  
Guatemala City, Guatemala  
santiagobocel10@gmail.com

Brenner Hernandez  
Universidad Rafael  
Landívar  
Guatemala City, Guatemala  
velasquezbrenner@gmail.com

Pablo Muralles  
Universidad Rafael  
Landívar  
Guatemala City, Guatemala  
pablomuralles28@gmail.com

Roberto Solares  
Universidad Rafael  
Landívar  
Guatemala City, Guatemala  
betosolaresgar@gmail.com

## ABSTRACT

Something as background

## CCS CONCEPTS

• Computing methodologies → Bayesian network models.

## KEYWORDS

bag of words, bayesian network

### ACM Reference Format:

Santiago E. Bocel, Brenner Hernandez, Pablo Muralles, and Roberto Solares. 2021. Bag of words with Bayesian Network to detect languages. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

Text classification is one of the applications of Machine Learning and consists of cataloging the texts based on their content, that is, performing an analysis of the words to decide what type of text is being identified.

This work is ideal for a machine as they are ideal for processing large amounts of information. However, since the machine does not initially know how to catalog a text based on any criteria, it requires a learning process in advance.

## REFERENCES

- [1] AshishSingh Bhatia and Bostjan Kaluza. 2018. *Machine Learning in Java*. Packt, Birmingham, Uk.
- [2] Tony F. Chan, Gene H. Golub, and Randall J. LeVeque. 1979. Updating the formula and a pairwise algorithm for computing sample variates. Technical report. Stanford University.
- [3] Praveen Dubey. [n. d.] An introduction to bag of words and how to code it in python for nlp. <https://www.freecodecamp.org/news/an-introduction-to-bag-of-words-and-how-to-code-it-in-python-for-nlp-282e87a9da04/>.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*Conference'17, July 2017, Washington, DC, USA*

© 2021 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

- [4] Srinivas Gurralla. [n. d.] Implementation of bag of words using python. <https://www.excelr.com/blog/data-science/natural-language-processing/implementation-of-bag-of-words-using-python>.
- [5] Michael Lanham. 2020. *Practical AI on the Google Cloud Platform*. Jonathan Hassell, editor. (1st. edition). O'Reilly Media, California.
- [6] Rajat Mehta. 2017. *Big data analytics with Java*. Packt, Birmingham, Uk.
- [7] @timleathart. [n. d.] How to handle a zero factor in naive bayes classifier calculation? <https://datascience.stackexchange.com/questions/15526/how-to-handle-a-zero-factor-in-naive-bayes-classifier-calculation>.