

IMDB Reviews

Text Mining and Search

Agazzi Ruben 844736, Cominetti Fabrizio 882737

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Keywords

Keyword1 — Keyword2 — Keyword3

University: University of Milano-Bicocca

Contents

| | | |
|-----|--------------------------|---|
| 1 | Introduction | 1 |
| 2 | Data | 1 |
| 2.1 | Data Pre-Processing | 1 |
| 2.2 | Dimensionality Reduction | 1 |
| 3 | Text Representation | 1 |
| 4 | Text Classification | 1 |
| 5 | Text Clustering | 1 |
| 6 | Summary | 1 |

1. Introduction

...

2. Data

Large Movie Review Dataset.// This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training, and 25,000 for testing. There is additional unlabeled data for use as well. Raw text and already processed bag of words formats are provided.

2.1 Data Pre-Processing

...

2.2 Dimensionality Reduction

...

3. Text Representation

...

4. Text Classification

...

5. Text Clustering

...

6. Summary

...