Replication of de Vel (2012): Mining Email Content for Author Identification Forensics

Marcel Schliebs

Studienstiftung natur- und ingenieurswissenschaftliches Forschungskolleg

Abstract

Lorem ipsum pipapo.

Replication of de Vel (2012): Mining Email Content for Author Identification Forensics

Introduction

In this document, I will try to document the process of replication and re-implementation of the algorithm.

Paper Description

1. Summary, Test Case and Approach
    (a) ...
        i. • xxx
           • yyy
    (b) zzz
        i. • 111
           • 222

Implementation/Operationalization

1. Features and their implementation
2. Other ideas
    (a) Comparative Approach: Test for other countries:
        i. France after Bataclan, other national disasters
            A. CEVIPOF Dataset?
            B. Other Datasets?
    (a) Meta Analysis?
        i. Include all model specifications in the regression as interacting with s.e.
            A. xx
        ii. Could we detect very small but true effects in meta analysis that would not be able to be identified in an individual study due to small effect sizes and large standard errors

(b) Lecture notes Tom Stanley

i. xx

ii. yy

Features

Overview over Style Merker Attributes

Number of blank lines/total number of lines

Sections

Use section and subsection commands to organize your document. LaTeX handles all the formatting and numbering automatically. Use ref and label commands for cross-references.

Comments

You can add inline TODO comments with the todonotes package, like this:

This is an inline comment.

References

LaTeX automatically generates a bibliography in the APA style from your .bib file. The citep command generates a formatted citation in parentheses (**?**). The cite command generates one without parentheses. LaTeX was first discovered by **?**.

Tables and Figures

Use the table and tabular commands for basic tables — see Table 2, for example. You can upload a figure (JPEG, PNG or PDF) using the files menu. To include it in your document, use the includegraphics command as in the code for Figure 1 below.

Mathematics

LaTeX is great at typesetting mathematics. Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i}^{n} X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.

Lists

You can make lists with automatic numbering ...

1. Like this,
2. and like this.

... or bullet points ...

- Like this,
- and like this.

We hope you find writeLaTeX useful, and please let us know if you have any feedback using the help menu above.

Table 1

Caption for the table.

Some	actual	content
prettifies	the	content
as	well	as
using	the	booktabs package

	StyleMarkerAttributeType	Status	VarName	Remarks
1	Number of blank lines/total number of lines	✓		lorem ipsum pipapo
2	Average sentence length	✓		lalala blablabla
3	Average word length (number of characters)	✓		
4	Vocabulary richness i.e., V=M	✓		
5	Total number of function words/M	✓		
6	Function word frequency distribution (122 features)	✗		
7	Total number of short words/M	✗		
8	Count of hapax legomena/M	∼		
9	Count of hapax legomena/V	∼		
10	Total number of characters in words/C	✗		
11	Total number of alphabetic characters in words/C	✓		
12	Total number of upper-case characters in words/C	∼		
13	Total number of digit characters in words/C	✓		
14	Total number of white-space characters/C	✓		
15	Total number of space characters/C	✓		
16	Total number of space characters/number white-space characters	✗		
17	Total number of tab spaces/C	✗		
18	Total number of tab spaces/number white-space	∼		

Item	Quantity
Widgets	42
Gadgets	13

Table 2

An example table.

Figure 1. This is a figure caption.