Term Paper Data Science 1

Docent: Prof. Dr. Lena Wiese
Semester: Summer Term 2021

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Institute of Computer Science
Goethe-Universität Frankfurt a. M.

Authors: FRANZISKA HICKING

your Student ID

your.email@ddre.ss

branch of study (Bachelor/Master, semester count)

JONAS ELPELT

your Student ID

your.email@ddre.ss

branch of study (Bachelor/Master, semester count)

JULIAN RUMMEL

your Student ID

your.email@ddre.ss

branch of study (Bachelor/Master, semester count)

NIKLAS CONEN

your Student ID

your.email@ddre.ss

branch of study (Bachelor/Master, semester count)

Date: April 27, 2021

Chosen Project Topic:
T4 - DISTANCE MEASURES AND CLUSTERING

Abstract

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet,

Contents

1 Problem Description

- What general problem is addressed?

- What is the general methodology that is used?

Some Latex-specific hints:

- You can use abbreviations, like Active Directory (AD), Microsoft (MS) or Compact Disc (CD).

- There are also symbols like for example π, φ and λ.

- Last but not least, use glossary entries like Active Directory and File.

- Do not forget to cite related work, like [1] or [2].

2 Description of Specific Methods and Algorithms

- What specific problem is addressed?

- What is the specific methodology that is used?

- What improvement is shown?

3 Data Set Description

- What benchmark data sets are used?

4 Description of Python libraries used

- Describe in detail what technology/library is used.

5 Description of Evaluation Module

- What are the results and how are they measured?

6 Web Frontend and User Manual

- Describe the implementation and write a brief user manual with screenshots.

7 Conclusion

- Summarize the main points and achievements

- Add your own assessment/criticism on the topic

References

[1] Lior Okman, Nurit Gal-Oz, Yaron Gonen, Ehud Gudes, and Jenny Abramov. Security issues in nosql databases. In *10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 541–547. IEEE, 2011.

[2] Dhruba Borthakur, Jonathan Gray, Joydeep Sen Sarma, Kannan Muthukkaruppan, Nicolas Spiegelberg, Hairong Kuang, Karthik Ranganathan, Dmytro Molkov, Aravind Menon, Samuel Rash, et al. Apache hadoop goes realtime at facebook. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 1071–1080. ACM, 2011.