UNIVERSITY OF TWENTE

MASTER THESIS

TRUST IN AUTOMATED DECISION MAKING

HOW USER'S TRUST AND PERCEIVED UNDERSTANDING IS
INFLUENCED BY THE QUALITY OF AUTOMATICALLY GENERATED
EXPLANATIONS

Author:
Andrea PAPENMEIER

Supervisor:
Dr. Christin SEIFERT
Dr. Gwenn ENGLEBIENNE

December 13, 2018

UNIVERSITY
OF TWENTE.

Abstract

1   Introduction

asd [1] asd [2]

2   Background

Intro

2.1   Opacity in AI

Intro

2.1.1   Opacity Sources

asd

2.1.2   Application Areas

asd

2.1.3   Examplary Failures

asd

2.1.4   Regulations

asd

2.1.5   Use Case Scenario

asd

2.2   Explanations

Intro

2.2.1   Explanation Types

asd

2.2.2   Social Sciences

asd

2.2.3   Psychology

asd

2.3   Explanations in AI

Intro

2.3.1 Explanation Focus

asd

2.3.2 Related Projects

asd

2.3.3 Explanations for Texts

asd

2.4 Trust in AI

Intro

2.4.1 Persuasiveness

asd

2.4.2 Trust

asd

2.5 Summary

Summary
Hypotheses

3   Dataset

Intro

3.1   Dataset Selection

asd

3.2   Dataset Construction

asd

3.3   Dataset Preprocessing

asd

4   Evaluation

5 Discussion

6   Conclusion

asd

References

[1] Hiva Allahyari and Niklas Lavesson. User-oriented assessment of classification model understandability. In *11th scandinavian conference on Artificial intelligence.* IOS Press, 2011.

[2] Leila Arras, Franziska Horn, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek. " what is relevant in a text document?": An interpretable machine learning approach. *PloS one*, 12(8):e0181142, 2017.