Fake Reviews Detection

NUCES-FAST

March 6, 2019

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

Team Members:

- **Muhammad Moazzam Mushtaq** (P156060@nu.edu.pk)

- **Muhammad Raza** (P156097@nu.edu.pk)

- **Muhammad Ahmad** (P66111@nu.edu.pk)

Supervisor: **Dr. Muhammad Taimoor Khan**
(taimoor.khan@nu.edu.pk)

Table of contents

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

Table of contents

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

Fake Reviews

- **Reviews** refer to any view or opinion made about a product or service by an individual.
- **Fake review** is not an actual customer's honest and impartial opinion or that does not reflect a customer's genuine experience of a product, service or business.
- Fake reviews raises doubts on analysis.
- In general, there can be three types of fake reviews:
    - Positive.
    - Negative.
    - Completely irrelevant.

Table of contents

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

- Research on fake reviews detection is still in the early stages. We review some of the published work as follows:
    - Detecting opinion spams and fake news using text classification, September 2018. [1]
    - Finding Deceptive Opinion Spam by Any Stretch of the Imagination, June 2011 . [2]

Problem Statement

- Fake reviews can negatively affect various customers and companies. [1]
- Tricking customers to buy a product that is entirely different from what was advertised. [2]
- Lack of an efficient way to differentiate between genuine or fake reviews. [1]

- N-gram Based Model.
- Data Pre-processing
    - Stop Word Removal.
    - Stemming.
- Feature Extraction.
- Semantic Similarity Measurement..
    - Word Length Measure.
    - Similarity Between Words.

Dataset

- The dataset was gathered by Myle Ott et al, a research engineer in Facebooks AI Research group (FAIR). [2]
- Information available for every review.
    - Hotel Name.
    - Review Text.
    - Review Sentiment (Positive and Negative).
    - Review Label (Fake and Genuine).
- Dataset contains hotel reviews of 20 Chicago hotels.
- It contains 800 Truthful and 800 fake reviews.

Table of contents

1 Introduction

2 Literature Review

3 Objectives

4 Work Break Down Structure

5 References

6 Suggestions

Objectives

The proposed techniques of Natural Language Processing
(NLP) and Machine Learning using:

- Prediction of text styling on the basis of language model.
- Features extraction from the defined styling procedure.
- Putting the extracted features in classifier models to
  observe the results.

Table of contents

1. **Introduction**

2. **Literature Review**

3. **Objectives**

4. **Work Break Down Structure**

5. **References**

6. **Suggestions**

Work Break Down Structure

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
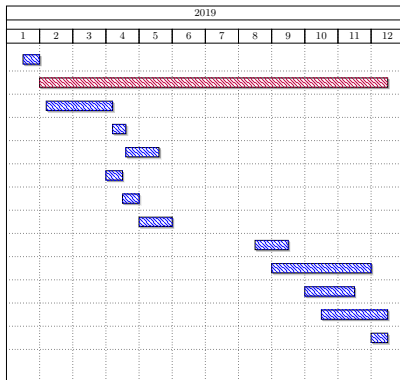Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

References

Fake Reviews
Detection

Introduction
Fake Reviews

Literature
Review
Related Work
Problem
Statement
Models and
Approach
Dataset

Objectives

Work Break
Down
Structure

References

Suggestions

📄 Hadeer Ahmed, Issa Traore, and Sherif Saad.
Detecting opinion spams and fake news using text
classification.
Security and Privacy, 1(1):e9, 2018.

📄 Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T
Hancock.
Finding deceptive opinion spam by any stretch of the
imagination.
In *Proceedings of the 49th Annual Meeting of the
Association for Computational Linguistics: Human
Language Technologies-Volume 1*, pages 309–319.
Association for Computational Linguistics, 2011.

Suggestions?