UNIVERSITY OF TWENTE.

KPMG

Marcia Fissette

11-04-2018

Jaarverslag

Verslag van de activiteiten van een bedrijf van het voorgaande jaar.

Verslag bevat

Jaarrekening

- Balans
- Winst- en verliesrekening
- Kasstroomoverzicht

Tekstuele informatie

- Toelichting op de jaarrekening
- Brief aan de aandeelhouders
- Verslag van de accountant
- *Management Analyse & Discussie*

Fraude in jaarverslagen

Financiële effect
Grootste financiële effect van alle fraude types
Voorbeelden:

- Enron: Aandelen gezakt van $90,56 per aandeel naar centen
- Parmalat: schuld van 14 miljard en 300 miljoen euro (8 keer hoger dan gerapporteerd)

Frauduleuze activiteiten
Vervalsing van onderliggende documenten
Vervalste schattingen en oordelen
Achterhouden van belangrijke informatie

Onderzoeksvraag

Kan een **tekst mining** model worden ontwikkeld die **indicaties van fraude** kan detecteren in **jaarverslagen** van bedrijven **wereldwijd?**

Tekst mining

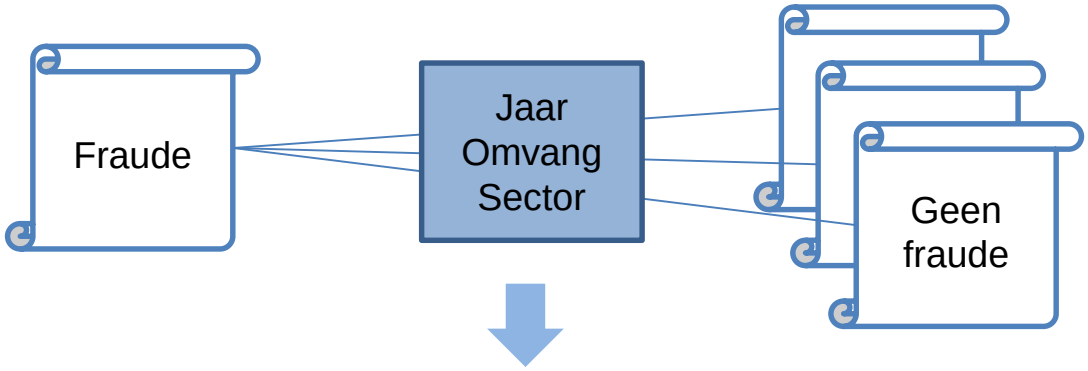
Automatisch identificeren van patronen in tekst

Model bestaat uit

1. Een data set bestaande uit teksten
2. Procedure om tekst te transformeren naar een gestructureerde representatie
3. Machine learning procedure om patronen te leren

Gebruikt om te classificeren

De data



402 frauduleuze jaarverslagen
1.325 niet frauduleuze jaarverslagen

Data voorbereiding

Verwijderen van html tags

Uitsluiten van tabellen

Kleine letters

Tokenization
Zinnen
Woorden

Verwijderen van interpunctie

Features

Unigrams

Bigrams

Beschrijvend

- Tekst lengte
- Lexicale diversiteit

Complexiteit

- Zin complexiteit
- Woord complexiteit

Grammaticaal

- Part-of-speech woord categoriën
- Werkwoordstijd

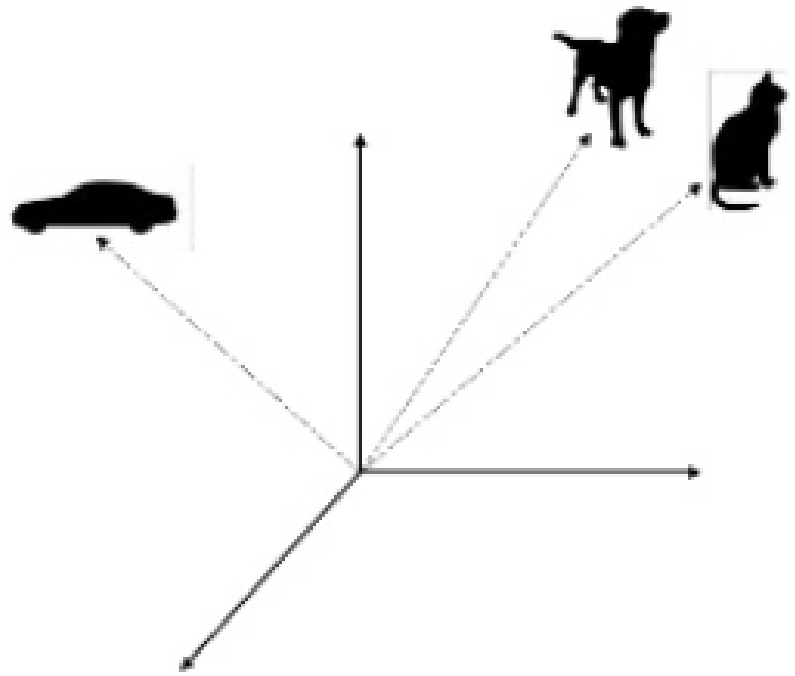
Leesbaarheid

- Jaren onderwijs nodig

Psychologisch

- Positief vs. negatief
- Zekerheid

Woord vectoren



Machine learning

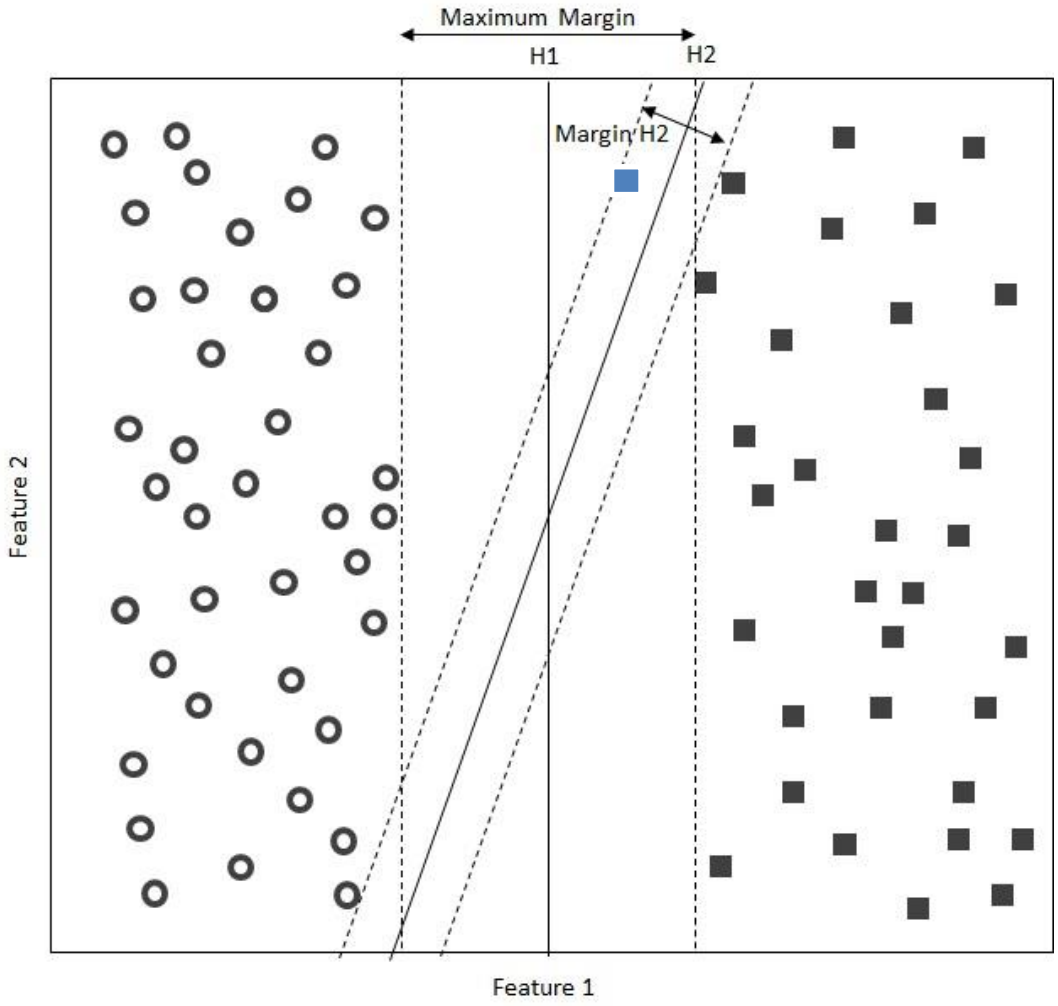
Split data
Development– validation: 70% - 30%

Naïve Bayes

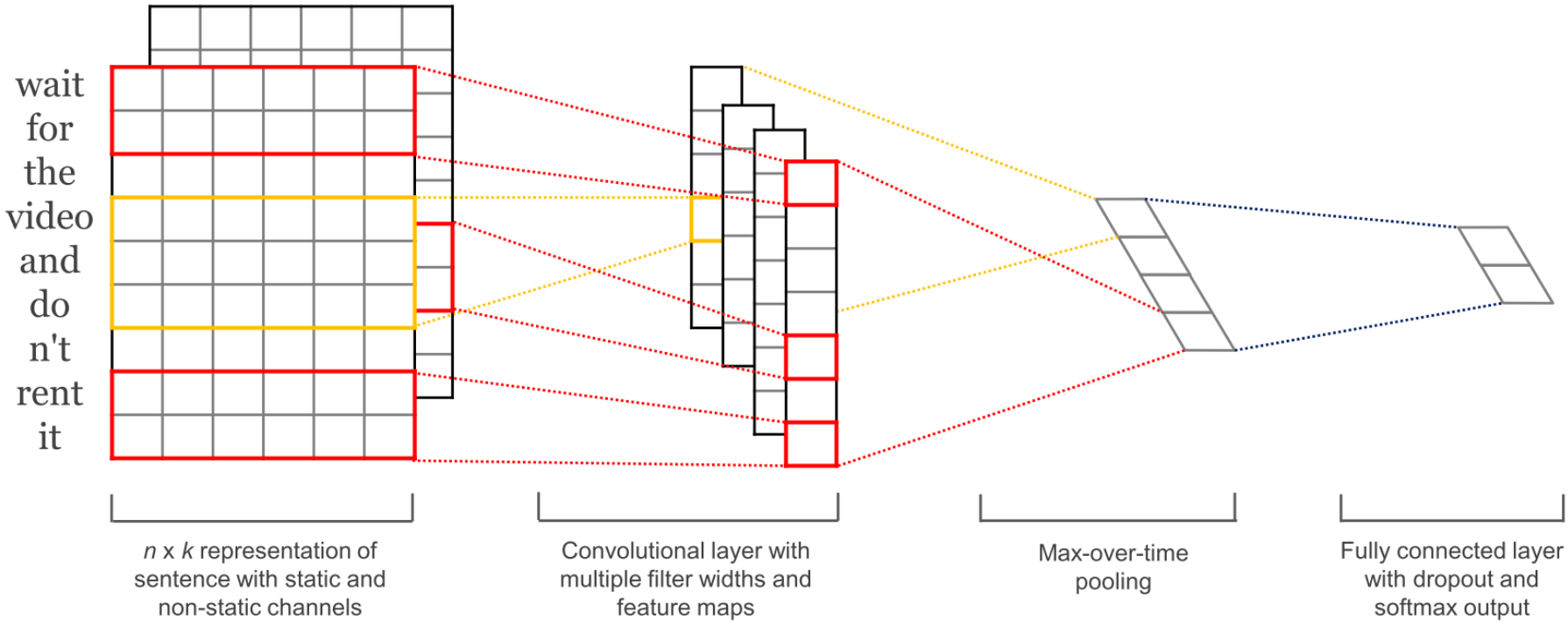
Support Vector Machine

Convolutional Neural Network

Support Vector Machine



Convolutional Neural Network



Resultaat

Kan een **tekst mining** model worden ontwikkeld die **indicaties van fraude** kan detecteren in **jaarverslagen** van bedrijven **wereldwijd?**

Tekst mining kan bijdragen aan fraudedetectie

Naïve Bayes model preseteert beter dan Support Vector Machine in fraude herkenning

Unigrams beste features

Convolutional Neural Network potentie om beter te presteren dan Naïve Bayes

UNIVERSITY OF TWENTE.

KPMG

Marcia Fissette