

Predicting signs of problem-gambling from online texts using large language models

Elke Smith, Nils Reiter, Jan Peters

Elke Smith



Department of Psychology
Biological Psychology
University of Cologne
Cologne



INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN



Problem gambling

- Continuation of gambling behaviour despite negative consequences
- Psychiatric disorder, classified as addictive disorder in DSM-5¹
- Prevalence estimated between 0.5% and 3% in Europe²

- Symptoms include
 - Persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress

¹ APA, 2014, *Hogrefe*

² EGBA, 2012

Problem gambling

- Continuation of gambling behaviour despite negative consequences
- Psychiatric disorder, classified as addictive disorder in DSM-5¹
- Prevalence estimated between 0.5% and 3% in Europe²

- Symptoms include
 - Persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress
 - Need to gamble with increasing amounts of money in order to achieve the desired excitement

¹ APA, 2014, *Hogrefe*

² EGBA, 2012

Problem gambling

- Continuation of gambling behaviour despite negative consequences
- Psychiatric disorder, classified as addictive disorder in DSM-5¹
- Prevalence estimated between 0.5% and 3% in Europe²

- Symptoms include
 - Persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress
 - Need to gamble with increasing amounts of money in order to achieve the desired excitement
 - Unsuccessful efforts to control, cut back, or stop gambling

¹ APA, 2014, *Hogrefe*

² EGBA, 2012

Problem gambling

- Continuation of gambling behaviour despite negative consequences
- Psychiatric disorder, classified as addictive disorder in DSM-5¹
- Prevalence estimated between 0.5% and 3% in Europe²

- Symptoms include
 - Persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress
 - Need to gamble with increasing amounts of money in order to achieve the desired excitement
 - Unsuccessful efforts to control, cut back, or stop gambling
 - Returning after losing money gambling ("chasing" losses)

¹ APA, 2014, *Hogrefe*

² EGBA, 2012

Problem gambling

- Continuation of gambling behaviour despite negative consequences
- Psychiatric disorder, classified as addictive disorder in DSM-5¹
- Prevalence estimated between 0.5% and 3% in Europe²

- Symptoms include
 - Persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress
 - Need to gamble with increasing amounts of money in order to achieve the desired excitement
 - Unsuccessful efforts to control, cut back, or stop gambling
 - Returning after losing money gambling ("chasing" losses)
 - Relying on others to provide money to relieve a desperate financial situation caused by gambling

¹ APA, 2014, *Hogrefe*

² EGBA, 2012

Gambling-related cognitive distortions

- Gambling-related cognitive distortions are cognitive biases/maladaptive ways of thinking about gambling ¹
- Persons with problem gambling tend to have more gambling-related cognitive distortions than gamblers without problem gambling behaviour and non-gamblers²
- Distortions manifest as
 - Relating winnings to skill and ability, and losses to bad circumstances

¹ Raylu & Oei, 2004, *Addiction*

² Joukhador et al., 2003, *Psychological Reports*

Gambling-related cognitive distortions

- Gambling-related cognitive distortions are cognitive biases/maladaptive ways of thinking about gambling ¹
- Persons with problem gambling tend to have more gambling-related cognitive distortions than gamblers without problem gambling behaviour and non-gamblers²
- Distortions manifest as
 - Relating winnings to skill and ability, and losses to bad circumstances
 - Assuming that specific numbers and colours increase the chances of winning

¹ Raylu & Oei, 2004, *Addiction*

² Joukhador et al., 2003, *Psychological Reports*

Gambling-related cognitive distortions

- Gambling-related cognitive distortions are cognitive biases/maladaptive ways of thinking about gambling¹
- Persons with problem gambling tend to have more gambling-related cognitive distortions than gamblers without problem gambling behaviour and non-gamblers²
- Distortions manifest as
 - Relating winnings to skill and ability, and losses to bad circumstances
 - Assuming that specific numbers and colours increase the chances of winning
 - Specific rituals and behaviours increase the chances of winning

¹ Raylu & Oei, 2004, *Addiction*

² Joukhador et al., 2003, *Psychological Reports*

Gambling-related cognitive distortions

- Gambling-related cognitive distortions are cognitive biases/maladaptive ways of thinking about gambling¹
- Persons with problem gambling tend to have more gambling-related cognitive distortions than gamblers without problem gambling behaviour and non-gamblers²
- Distortions manifest as
 - Relating winnings to skill and ability, and losses to bad circumstances
 - Assuming that specific numbers and colours increase the chances of winning
 - Specific rituals and behaviours increase the chances of winning
 - Sense of having control over predicting wins

¹ Raylu & Oei, 2004, *Addiction*

² Joukhador et al., 2003, *Psychological Reports*

Online communities on gambling

The screenshot shows the AskGamblers forum homepage. At the top, there is a navigation bar with 'ASKGAMBLERS' logo, a search bar, and menu items: CASINOS, BONUSES, GAMES, COMPLAINTS, NEWS, PARTNERSHIP, FORUM, and ABOUT. Below the navigation bar, there is a 'Popular Contributors' section with a filter for 'Week' and a list of four users: Flamingo97 (127 posts), Althelio (132 posts), jh0000 (121 posts), and Wovvex9013 (121 posts). A 'Show More' link is present. Below this is a 'Forums' section with a 'Start New Topic' button. The main content area is titled 'The AskGamblers' and lists two forum topics: 'Forum Rules' (1 post, by AskGamblers Forum) and 'Exclusive Competitions & Contests' (8,314 posts, by Flamingo97).

The screenshot shows the Wizard Vegas website. The top navigation bar includes 'WIZARD Vegas' logo and menu items: ARTICLES, HOTELS, FORUMS, GAMES, JACKPOTS, SURVEYS, GUIDES, VIDEOS, SHOWS, TRAILS, and BLOGS. Below the navigation bar is a large banner image of a city skyline with the text 'BETTING SYSTEMS' and a navigation path: Home > Forum > Gambling.



[Home](#) [Forum](#) [Chatrooms](#) [Guidelines](#) [MyGamCare](#)

[Forums](#) > [Recovery Diaries](#) > 10k down in 3 months from online casinos

10k down in 3 months from online casinos



@gambler
Posts: 16

Topic starter

I was wondering if anyone can help me justify what I have done the last 3 months.

I gambled most of 2019 had no money the whole year and was majority depressed, then 2020 I pulled my self together gambled once and lost a grand and that was it. then I turn 30 this January and think to my self right you have sorted your self out grow up now and keep saving for that deposit for a home, managed to save 10k and then after turning 30 I've done the 10k every weekend done 1k here 2k there until it was all gone..

The screenshot shows the '10 Fabulous Casino Listings' website. The top navigation bar includes 'CASINOS', 'BONUSES', 'GAMES', 'JACKPOTS', 'NEWS', 'FORUM', and 'MORE'. Below the navigation bar, there is a 'Recommended Casinos' section with three listings: 'bit + Stars' (5 stars), '500Slots Casino' (5 stars), and '500Slots Casino' (5 stars). A 'Strategy of playing slots' section is also visible.

Web scraping

- Method for structured extraction of website data
- Legal for scientific purposes if no access restrictions are circumvented¹
- Packages available for many programming languages (e.g. requests library for *Python*, and rvest package for *R*)

```
96 def scrape(url):
97
98     print("Requesting url ...")
99     response = requests.get(url)
100     soup = BeautifulSoup(response.text, "html.parser")
101
102     if SCRAPE_FORUM:
103         data = soup.find_all("div", class_=forum_div_class)
```

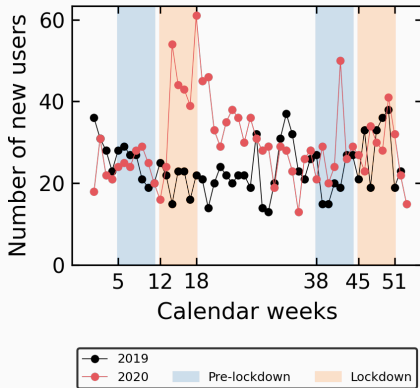
¹ Klawonn, 2019, Available at SSRN 3491192

Database

- Scraped data from a large Germany-speaking online forum on gambling (April 2021)
- 205,385 posts from 4,428 registered user profiles

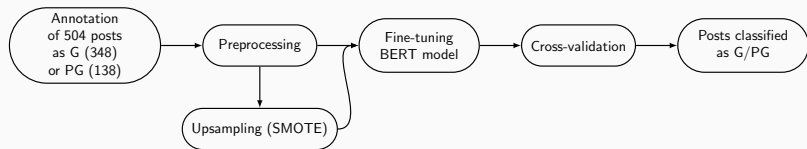
Subforum	N posts
Rules and guidelines	9349
Blackjack	308
Poker	296
Roulette	614
Other games of chance	4701
Slot machines and slot games	6079
Gambling arcades and casinos	3365
Casino complaints	14328
Gambling addiction	4150
Online casinos	140818
Miscellaneous	21377

Changes in posting behaviour in gambling forum during the lockdown in Germany



Can we detect early signs of
problem gambling from online
texts using machine learning
methods?

Workflow



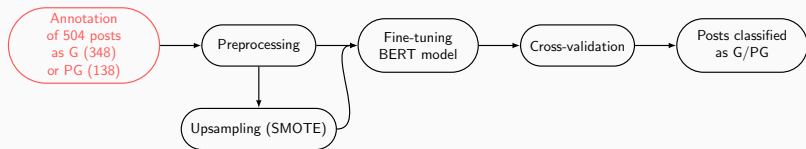
- Annotation based on DSM-5 criteria for gambling disorder and the Gambling Related Cognitions Scale (GRCS)¹
- Fine-tuning of German BERT_{BASE} model²

¹ APA, 2014, *Hogrefe*; Raylu & Oei, 2004, *Addiction*

² Devlin et al., 2019, *Proc Conf Assoc Comput Linguist*; <https://huggingface.co/dbmdz/bert-base-german-uncased>

G: gambling, PG: problem gambling

Workflow



- Annotation based on DSM-5 criteria for gambling disorder and the Gambling Related Cognitions Scale (GRCS)¹
- Fine-tuning of German BERT_{BASE} model²

¹ APA, 2014, *Hogrefe*; Raylu & Oei, 2004, *Addiction*

² Devlin et al., 2019, *Proc Conf Assoc Comput Linguist*; <https://huggingface.co/dbmdz/bert-base-german-uncased>

G: gambling, PG: problem gambling

Annotation

- Gambling post

Annotation

- Gambling post

„...I want to warn you vehemently about this casino!!“

Annotation

- Gambling post
„...I want to warn you vehemently about this casino!!“
- Problem gambling post (problem gambling behaviour)

Annotation

- Gambling post

„...I want to warn you vehemently about this casino!!”

- Problem gambling post (problem gambling behaviour)

„...I kept playing and there was hardly anything left of my €1,400 net income per month. I gambled away hundreds of euros again and again.”

Annotation

- Gambling post

„...I want to warn you vehemently about this casino!!”

- Problem gambling post (problem gambling behaviour)

„...I kept playing and there was hardly anything left of my €1,400 net income per month. I gambled away hundreds of euros again and again.”

- Problem gambling post (cognition distortion)

Annotation

- Gambling post

„...I want to warn you vehemently about this casino!!”

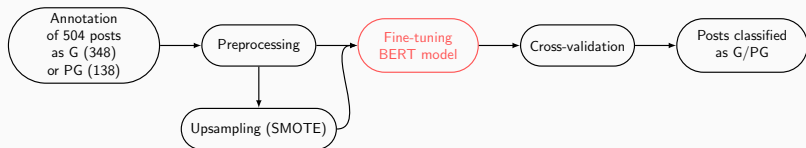
- Problem gambling post (problem gambling behaviour)

„...I kept playing and there was hardly anything left of my €1,400 net income per month. I gambled away hundreds of euros again and again.”

- Problem gambling post (cognition distortion)

„...casino software is absolutely controllable.”

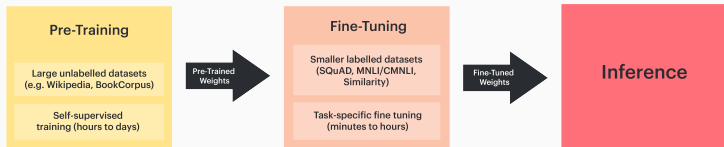
Workflow



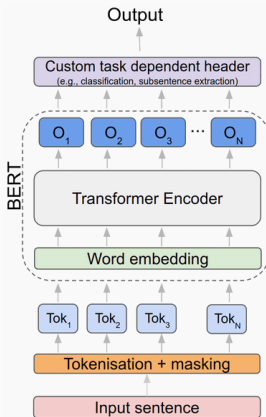
What are large language models (LLMs)?



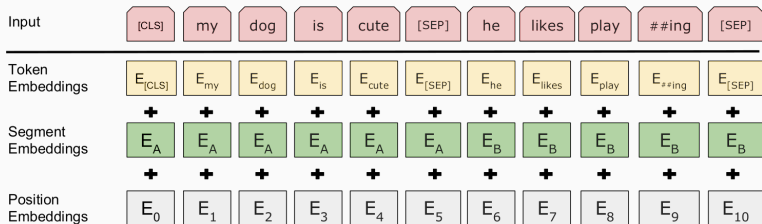
Pre-training and fine-tuning BERT



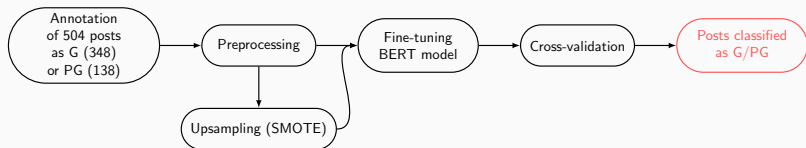
Bidirectional Encoder Representations from Transformers (BERT)



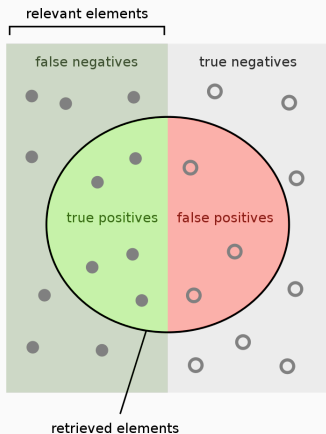
Input representation for BERT



Workflow



Evaluating classification performance



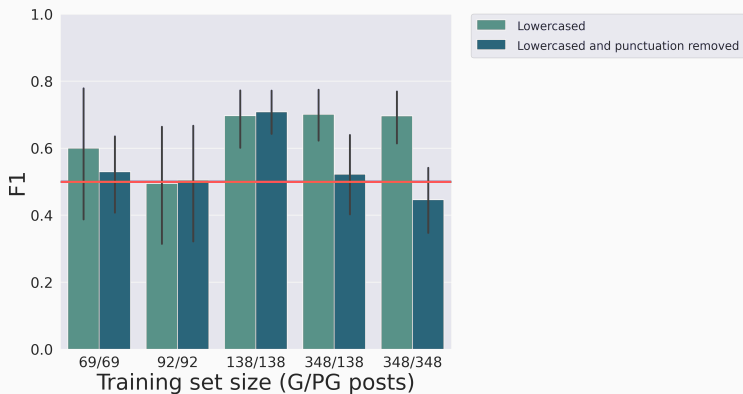
How many retrieved items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are retrieved?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

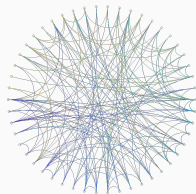
Classification performance



What are the perspectives for
applying LMMs to such data?

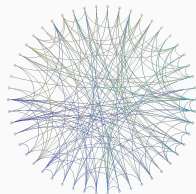
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data



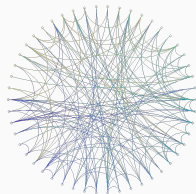
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity



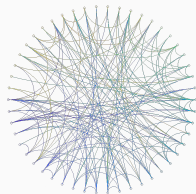
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity
 - Compliance with data protection and privacy



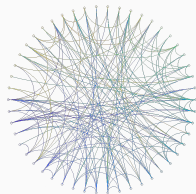
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity
 - Compliance with data protection and privacy
- Signs of problem gambling can be recognised automatically using machine learning methods



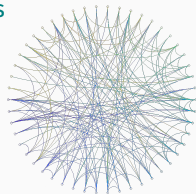
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity
 - Compliance with data protection and privacy
- Signs of problem gambling can be recognised automatically using machine learning methods
 - Use online data for research on behavioural addictions



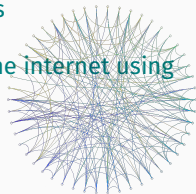
Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity
 - Compliance with data protection and privacy
- Signs of problem gambling can be recognised automatically using machine learning methods
 - Use online data for research on behavioural addictions
 - Validate ML classification through diagnostics



Perspectives

- Increasing exchange on health-related topics in social media generates large amounts of data
 - Data with high ecological validity
 - Compliance with data protection and privacy
- Signs of problem gambling can be recognised automatically using machine learning methods
 - Use online data for research on behavioural addictions
 - Validate ML classification through diagnostics
 - Develop measures for player protection on the internet using computational methods



Thank you!

Elke Smith `e.smith@uni-koeln.de`

Nils Reiter

Jan Peters

Smith, E., Reiter, N., & Peters, J. (2023). Automatic detection of problem-gambling signs from online texts using large language models. *arXiv preprint arXiv:2312.00804*.

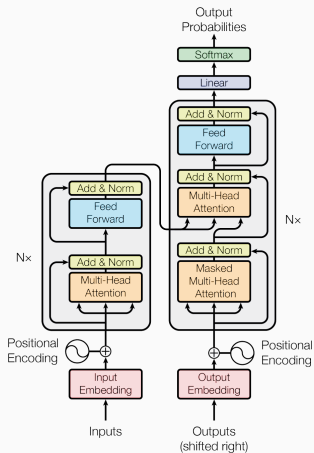


Department of Psychology
Biological Psychology
University of Cologne
Cologne



INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

Transformer architecture



Fine-tuning parameters

MODEL *bert-base-german-uncased*
(BERT_{BASE} has 12 encoder layers, 768 hidden units, 12-self-attention heads, 110 m parameters and is pre-trained on a German *Wikipedia* and *OpenLegalData* dump and news articles)

FINE-TUNING Regression head (linear layer) on top of the output by implementing *BertForSequenceClassification*

SEQ. LENGTH 512 tokens (maximum)

BATCH SIZE 16

OPTIMISER *Adam optimiser* with weight decay ($\epsilon = 1e - 08$)

EPOCHS 2

LEARNING RATE $5e - 5$

Classification performance

