

Text mining of customer service interactions as a tool for early detection of problem gambling

What this research is about

Problem gambling (PG) can negatively affect many areas of life, including mental health and relationships. Identifying early signs of PG can prevent its onset or worsening. Text mining is an innovative method for identifying people at risk for PG. Previous research has examined messages posted on online forums or social media by people who may be experiencing PG. The findings suggest that these people tend to express negative emotions such as frustration, guilt, and anger. This study aimed to identify linguistic markers that signal risk for PG (e.g., specific words and tones) by analyzing online interactions between users of an Italian gambling operator and its customer service representatives. An interactive system and digital interface was developed to assist in early risk detection based on such linguistic markers.

What the researchers did

In the first phase, the researchers analyzed a dataset consisting of 19,000 online customer service chats provided by the Italian gambling operator Sisal. The data were linked to PGSI scores obtained in a brief survey conducted by Sisal as part of its responsible gambling and service evaluation. The survey was administered prior to the collection of the chat data.

The researchers first created subsamples of 150 conversations with people at low risk (PGSI scores lower than 8) and 170 conversations with people with high risk (PGSI scores higher than 15) for maximum contrast. They then manually analyzed these conversations for (1) interaction patterns; (2) elements related to tone, punctuation, and

What you need to know

In this study, the researchers identified linguistic markers that could be used to identify people at risk of problem gambling (PG). They analyzed 19,000 customer service chats sourced from an Italian gambling operator. The researchers found that people at higher risk of PG had longer and more complex conversations. They more frequently discussed financial losses and used heightened emotional and threatening language compared to those with lower risk. A model was developed to predict risk of PG based on such linguistic markers. The model was able to correctly identify the risk level of users 78% of the time.

In the second phase of the study, the model was further refined using a new dataset of 5,900 conversations. A four-level risk classification was applied (no risk, low/moderate risk, high risk, and severe problem gambling). Finally, an interactive system and digital interface was developed to help with early risk detection.

paralinguistic cues (e.g., repeating phrases, capitalizations); and (3) frequently discussed topics. From there, they identified linguistic markers associated with PG and assigned them higher weights. A machine learning model for predicting risk of PG was developed and trained on the dataset.

In the second phase, the model was further refined using a new dataset of 5,900 conversations. The new dataset was manually coded, and a four-level risk classification aligning with the PGSI's classification was applied: no risk (scores of 0), low/moderate risk

(scores of 1–7), high risk (scores of 8–14), and severe PG (scores of 15 or higher). Each conversation was assigned to a risk level based on the linguistic markers identified. Finally, an interactive system and digital interface was developed to support gambling operators in early risk detection.

What the researchers found

In phase 1, most conversations came from people at low risk (PGSI scores lower than 8). However, people at high risk (scores higher than 15) typically had longer and more complex conversations with customer service representatives. People who met the threshold for PG (scores of 8 or higher) more frequently expressed topics related to financial losses, bonus requests, and service complaints. They tended to use more demanding and urgent language, aggressive tone, and more intense emotional expressions. They were also more likely to use threats, express frustrations, or repeat requests. The model developed in phase 1 was able to correctly identify the risk level of users 78% of the time.

In phase 2, the linguistic markers identified were grouped into macro-categories: behavioural indicators (e.g., references to spending patterns, cancelled withdrawals), emotional markers of desperation (e.g., expressions of frustration), interactional markers (e.g., urgency, aggressiveness), and cognitive markers associated with gambling-related thinking biases. Behavioural indicators were the most common, followed by emotional markers.

An interactive system and digital interface was implemented using the Chat Checker platform. This platform displays the conversation between a user and customer service, highlights linguistic markers that signal risk for PG, and assigns a risk level. The developed system can be used as a decision-support system to help gambling operators identify users at risk of PG.

How you can use this research

The findings suggest that computational linguistics and text mining can be combined with psychological insights for early detection of PG.

About the researchers

Chiara Barone, Michele Paleologo, Giulia Sesini, and Guendalina Graffigna are affiliated with the Department of Psychology at the Università Cattolica del Sacro Cuore in Milan, Italy. Michele Paleologo and Guendalina Graffigna are also affiliated with the EngageMinds HUB – Consumer, Food and Health Engagement Research Center of Università Cattolica del Sacro Cuore in Milan, Italy. Guendalina Graffigna is also affiliated with the Faculty of Agriculture, Food and Environmental Science at the Università Cattolica del Sacro Cuore in Cremona, Italy. **Giovanni Moretti and Marco Passarotti** are affiliated with the Department of Philology, Literature and Linguistics at the Università Cattolica del Sacro Cuore in Milan, Italy. **Benedetta Cesti, Virginia Attisani, Rebecca Losi, and Stefano De Vita** are affiliated with the Responsible Gaming Department of Sisal S.p.A. based in Milan, Italy. For more information about this study, please contact Chiara Barone at chiara.barone2@unicatt.it

Citation

Barone, C., Paleologo, M., Moretti, G., Sesini, G., Cesti, B., Attisani, V., Losi, R., De Vita, S., Passarotti, M., & Graffigna, G. (2026). Detecting gambling addiction: How text mining of customer–operator online chats reveals behavioral risks. *Journal of Gambling Studies*. Advance online publication.

<https://doi.org/10.1007/s10899-026-10500-8>

Study Funding

This study was funded by the gambling operator Sisal S.p.A within a collaborative research framework. Open access funding provided by Università Cattolica del Sacro Cuore within the CRUI-CARE Agreement.

About Greo Evidence Insights

Greo Evidence Insights is an independent, non-profit organization specializing in research, knowledge mobilization, and evaluation in the health and well-being sectors. Greo transforms research and insights into actionable strategies to prevent and reduce harm related to gambling, gaming, technology use, and substance use. To learn more about funding for Research Snapshots visit [our website](#).

