LudoTrack: Web Mining, Search Technologies and Natural Language Processing for the Early Detection of Pathological Gambling

David E. Losada[1], Nelly Condori-Fernández[1] and Marcos Fernández-Pichel[1]

[1]*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, 15782, Spain*

Abstract

This document summarises the project entitled "Web Mining, Search Technologies and Natural Language Processing for the Early Detection of Pathological Gambling". This is a research project funded by "Dirección General de Ordenación del Juego - Ministerio de Consumo" (Government of Spain). The project started in early 2024 and will run until the end of 2024. The funding for this project was given in the context of a national call of Ministerio de Consumo ("Convocatoria de subvenciones, durante el ejercicio 2023, para el desarrollo de actividades de investigación relacionadas con la prevención de los trastornos del juego, con los efectos derivados de dichos trastornos o con los riesgos asociados a esta actividad").

Keywords
Pathological Gambling, Web and Text Mining, Search Technologies, Natural Language Processing

1. Introduction

Gambling disorders were incorporated by the World Health Organisation (WHO) into the International Classification of Diseases ICD-11 (published in 2018, [1]), responding to the growing international concern in this area. Already in 2013, Internet gaming disorder had also been included in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [2] as a condition requiring specific study.

Patterns associated with gaming can lead to dysfunction and psychological distress for some players and, in various countries, this problem has generated significant public health concerns [3]. Despite the severity of these disorders, in many cases, individuals do not receive treatment or receive it late. There are well-documented limitations of existing preventive tools and a need for new instruments that distinguish across the spectrum of gaming behaviours, such as "regular and healthy gaming behaviours", "hazardous gaming," or "gaming disorder" [4]. The non-identification or late identification of signs of gaming disorders leads to serious social, health, and economic costs. This also has a significantly worrying impact on the adolescent population.

Language is a powerful indicator of personality traits, emotions, and provides valuable clues about mental health and disorders [5]. We can find distinctive psychological patterns in people not only by analysing the topics they talk about but also by studying the way they use language connectives such as prepositions or pronouns. Our research project aims to develop the necessary computational technologies and models to perform large-scale natural language analysis. It involves designing and implementing new monitoring and analytics tools that, using publicly available information on the web, can mine content to extract traces and evidence related to gaming disorders. More specifically, our main goal is to study the way people use language to reveal early signs of gaming-related disorders. To this end, the most advanced language analysis models would be used, such as deep neural network

✉ david.losada@usc.gal (D. E. Losada); nelly.condori@usc.gal (N. Condori-Fernández); marcos.fernandez.pichel@usc.gal (M. Fernández-Pichel)

🌐 https://citius.gal/ (D. E. Losada); https://citius.gal/ (N. Condori-Fernández); https://citius.gal/ (M. Fernández-Pichel)

🆔 0000-0001-8823-7501 (D. E. Losada); 0000-0002-1044-3871 (N. Condori-Fernández); 0000-0002-6560-9832 (M. Fernández-Pichel)

architectures based on "transformers" and recent large language models like BERT, ChatGPT, or GPT-4 [6, 7, 8, 9].

This project does not aim to develop automatic diagnostic technology. In fact, we believe that diagnostic tasks performed by medical professionals cannot be carried out by completely automated means. Here, we pursue the more realistic goal of, for example, designing methods that detect the emergence of initial signs of gaming disorders and understand the evolution of a person from the initial stages (e.g., mood changes, lack of sleep) to severe stages (e.g., pressing financial problems or suicidal thoughts). This information would be valuable, for example, for public institutions that could receive alerts about growing risks in specific population segments (e.g., to incite preventive measures). These new monitoring tools could extract and present evidence of the emergence and temporal development of gaming-related disorders that could be exploited by clinical professionals. This would enrich their current sources of evidence (usually focused on direct interaction with patients, surveys, and conventional clinical instruments).

The large volume of interactions and publications available on the Internet and social networks allows for massive analysis of psychological traits related to various disorders. It is common for individuals suffering from psychological disorders, such as those related to gaming, to interact with other individuals, express their concerns, share their experiences, and receive online help from specialised professionals. However, the analysis of online users presents challenges in several areas: filtering and searching for information (to find relevant excerpts from users that are pertinent for analysing a given psychological problem), linguistic text analysis and psycholinguistics, estimation of content quality and reputation (for example, with the purpose of recommending reputable information for people suffering from a certain disorder), and massive data processing (distributed computing methods that are scalable and operate in real-time).

2. Objectives of the Project

The main scientific hypothesis is that natural language reveals signs of different psychological disorders, particularly addictive disorders related to gambling, and that we can develop early detection technology based on the analysis of texts published by individuals. The strong relationship between the use of natural language and different psychological conditions has been demonstrated in the past [5], and furthermore, there is multiple evidence that social and web media can provide significant data on various disorders [10, 11]. Our main goal is to carry out this type of analysis on a large scale and incorporate extraction and search methods that are effective for identifying addictive disorders related to gambling. This represents a significant advancement because, in general, research in this area has been limited to small-scale studies (for example, essays written by a small number of already diagnosed patients) and has consistently ignored the temporal component, which is essential for analysing the evolution of disorders and performing early detection.

It is feasible to have data for a project like this, and we already have experience in extracting and analysing information online. There are public and freely available contents (open forums, social networks, etc.) where people openly discuss their problems related to their addiction and tell others that they have been diagnosed with a gambling addictive disorder or that they are beginning to develop it. This includes conventional social networks, with open public groups [12], platforms specialised in gambling addiction (ludopatia.org [13], vidasinjuego [14] and other sources [15, 16]), and personal support platforms [17, 18]. There are also online recovery programs, surveys, and other analytical instruments in that can be exploited for this project. This provides valuable resources to understand the problems these people suffer, categorise them, automatically extract topics (concerns, psychological impact, personality effects, emotions, among others) and obtain reputable contents and recommendations related to these disorders (for example, support programs, opinion surveys, or useful questions/answers). In addition, the established clinical criteria from Diagnostic Manuals for other disorders (for example, for depression: mood changes, loss of interest, etc.) can also be useful for studying the evolution of language use and concerns expressed over time (and how they reflect symptoms of anxiety, depression, etc.). In

this context, we will develop predictive technology demonstrators aimed at relevant stakeholders (for example, health professionals, psychologists, and the Ministry of Consumer Affairs, funder of this project).

The automatic analysis of texts and web pages will focus on public data made freely available online by users or Internet platforms. The developed algorithms will be evaluated with standardised and curated collections. We will not work with personal data, and therefore, the usual guidelines on privacy are not applicable to our project. In any case, we will use appropriate anonymisation strategies to remove proper names in the texts, user account identifiers, and other elements that could reveal any information about the subjects. Moreover, the design of the experiments and, in particular, the construction of the test collections, will carefully follow the recent recommendations on ethical aspects in the design of natural language analysis experiments [19]. It will be necessary to avoid demographic biases, misrepresentation, or exclusion of certain population groups in the training collections (these biases threaten the universality and objectivity of the extracted knowledge); it is also necessary to avoid over-generalisation and overexposure or underexposure (as much as possible avoiding, for example, that the constructed resources are oriented to a single language) and identify possible fraudulent or unethical uses of these technologies.

All activities carried out within the framework of the project will take special care to ensure that the models and solutions do not incorporate any type of gender bias (or other types of biases). The creation of datasets and algorithmic design will follow existing guidelines and recommendations [20] aimed at working with online data and avoiding biases and methodological deficiencies. We will also comply with rigorous ethical practices. We will document in detail the process by which datasets and models are created, and we will critically examine this process. We will extend studies on online data to different platforms, themes, moments, and subpopulations, to determine how results vary across, for example, different cultural, demographic, and behavioural contexts. We will enable transparency mechanisms that allow auditing the developed software and evaluating the biases of the data at the source. It is also relevant to comment that in the activity of this project, there would be no contact or interaction with people suffering from disorders or making publications on the Internet. Therefore, the project is exempt from IRB approval.

The specific objectives of this project are:

- **O1**. Develop new methods and resources that generate useful evidence for the monitoring and prevention of compulsive gambling disorder problems.
  This project addresses an innovative area where there are few test collections and polished open data. Additionally, new metrics for evaluation and early detection measures are needed. It will be necessary to create new test collections, oriented to different use cases related to disorders derived from gambling. For example, by retrieving, processing, and extracting data on the Internet related to the different phases of the problem, ranging from regular and not particularly harmful gambling behaviour to dangerous gambling phases or gambling disorder. To that end, we need to compile on-topic evidence about different issues or related themes (mental health, emotional impact, psychological effects, financial difficulties, academic, work or social problems, legal issues, etc.). The team that leads this project has extensive experience in creating new datasets and reference collections [21, 22, 23, 24, 25, 26, 27]. On the other hand, we will work on defining appropriate evaluation metrics to determine the quality of early detection systems. Here, it is necessary to take into account multiple dimensions, such as the relevance of the extracted information (regarding the area of gambling disorders), the computational efficiency of the extraction methods, scalability, and the validity of these collections to promote the development of intelligent early detection solutions (where the temporal dimension is fundamental).
- **O2**. Define effective methods of text search and filtering and apply them for the identification of high-quality textual sources relevant to the different information needs related to gambling disorders. Define models for analysing themes related to this risk and their temporal evolution.
  It will be necessary to manage large volumes of data and filter out irrelevant information (contents not related to the type of risk to be monitored and studied). We will work here on efficient and

effective methods of searching and filtering relevant information: automatic query generation (on compulsive gambling themes), query expansion, sentence/passage retrieval, relevance feedback and identification of user profiles related to a certain type of risk. Different domain resources (for example, specialised vocabularies and medical terminologies, such as those recently incorporated in the ICD-11 related to gambling) will facilitate the extraction of key terms or expressions for the generation/identification of key passages. We will also address data fusion and topic extraction and analysis. We have extensive experience in these areas: identification of queries on health or nutrition [28], query generation and expansion [29], sentence retrieval [30], and ranking fusion [31].

- **O3**. Develop linguistic resources, train language models and related technologies focused on managing multiple profiles of interest related to gambling disorders and their addictions. Implement advanced natural language processing and linguistic analysis for monitoring content related to these risks.

  We will work on the automatic elaboration of lexicographic resources adapted to the domain of searching for signs of risk of pathological gambling. We will consider some existing multilingual analytical extraction tools, such as LinguaKit [32] or the well-known Stanford NLP Toolkit, whose linguistic modules can be improved and adapted to this project. Also, recent large language models developed by Open AI (ChatGPT, GPT4) and Google (Bard), among others, will also be used to take advantage of their advanced language capabilities which, connected with the appropriate information for this project, can automatically perform tasks such as automatic cataloguing or summary generation.

- **O4**. Develop flexible and efficient solutions for the massive processing of data from multiple online sources, including social media, and implement real-time analysis of online content.

  We have experience in designing and implementing Big Data solutions for early risk detection [33]. We have also developed publicly available tools for real-time processing of social media data [34]. However, there are a number of challenges when designing Big Data solutions in the context of this project. For example, we need to process information in real-time to extract web content and analyse user-generated publications related to gambling.

- **O5**. Define methods for analysing results, generating conclusions, and exploiting expert knowledge (for example, from psychologists, medical specialists, or communication professionals). Determine the ways in which expert knowledge can guide the identification of reputable content and how to adapt communication measures or public preventive strategies to the psychological or risk profile of users.

  The validation of the solutions developed under this project must be carried out by experts in relevant sectors, and the team of the project includes two professionals specialised in Psychology and Communication, respectively. Their participation will allow to inject expert knowledge that can guide the identification of relevant elements (problem phases, psychological impact, prominent themes, etc.), and to validate the results and determine effective exploitation strategies. Likewise, it will also be necessary to determine, when appropriate, timely communication strategies to, for example, configure recommendations for preemptive programs, and suggest reputable and high quality support content for people at risk. In this sense, it will be necessary to advance in understanding how different users react to different preventive or risk communication campaigns and, also, help them to identify toxic, false information or harmful recommendations (for example, commercial sites that incite them to consume and play online). To propose recommendations related to various disorders, it is necessary to study social, personality, and psychological dimensions [35].

3. Expected Results and Exploitation

The project has the potential to produce high-quality results at regional, national, and international levels. Our previous activity in early risk detection and monitoring signs of depression, anorexia,

and other concerning disorders has been very well received internationally and further exploited by psychologists in our environment. For example, some of our Artificial Intelligence methods have led, in collaboration with psychologists from the University of A Coruña and experts from the University of Notre Dame in the USA, to suggestions for improving current monitoring tools for adolescents at risk due to problematic family situations (see [36], a collaborative work between USC, UDC, and the University of Notre Dame aimed at applying Machine Learning to predict risk in adolescents with problematic family situations). Therefore, we have experience in exploiting results with clinical professionals and are in the best position to do so with the results of this new project. On the other hand, the data and resources we have generated in the past (test collections, exploratory experimental challenges, etc.) have had a high global impact and represent a valuable resource to boost research in these areas.

Regarding this project, the new collections, resources, and experimental methods have the potential to produce a global impact and will be distributed publicly and openly so that many other teams can advance in this field. We expect to build large-scale reusable test collections that will become an international reference for studying the interactions between different problems and aspects of addictive gambling disorder and the use of natural language. We also hope to lay the experimental foundations (new performance metrics, new computationally efficient ways to create resources) for the early prediction of addictive gambling risks. These innovative evaluation methods have the potential to be useful not only for this project but also as a way to assess early risk prediction in other domains (for example, identifying sexual predators, cyberbullying, terrorism, etc.).

The new search methods, language technologies, and communication and recommendation models tailored to the case of addictive gambling disorders are also highly innovative developments and can lead to pioneering methods. For example, we hope to propose creative ways to automatically search for signs of disorders and their underlying problems (personal health, emotional and psychological impact, financial, work, academic, social, or legal difficulties, etc.). We will investigate new search methods, based on the automatic construction of queries from expert knowledge, and these information filtering and selection strategies can have an impact beyond this project (for example, to support health-related searches in clinical repositories). The recommendation of content associated with gambling disorders and the related communication strategies will be another highly innovative area of our activity whose results have the potential to contribute beyond the scope of the project. Likewise, the large-scale processing technology aims to serve not only this project but also the international community (for example, in projects and activities requiring real-time processing or analysis on social networks). Moreover, our results have great potential to be published in high-impact international venues and disseminated to society.

In terms of social impact, the participation of experts in Psychology, as a fundamental part of the project, ensures great potential for the exploitation of the results. In this sense, we will propose use cases (individualised analysis of subjects, study of disorders in communities or population groups, etc.) that can help psychologists, educators, and teachers in their daily activity. Indeed, the project can generate new valuable knowledge (for example, providing data on the evolution of different dimensions associated with disorders: disaffection, sleep problems, weight loss, etc.). This result is useful in itself for different professionals interested in addictive gambling disorders.

The project is also promising for sparking new prevention campaigns and communication strategies about gambling addictions. The team of the project incorporates an expert professor in Communication. We expect to obtain substantial evidence about the disorders, their appearance and evolution, and, thus, instigate recommendations to institutions (Ministry of Consumer Affairs and Ministry of Health, in addition to other related agencies and councils at the regional level) on public communication and preemptive policies. This is crucial, as numerous studies warn about the growth of gambling problems in our society. For example, according to 2021 data from the Spanish General Directorate for Gambling Regulation (DGOJ), we had in Spain an estimated number of 1,400,000 online gamblers; and these figures have been growing over the last decade. Although many of these active players do not have a pathological disorder, it is essential to analyse this population and identify the potential emergence of problems. We will work on communicating the project's results to relevant public institutions and, in

fact, we also have experience in this area.

We will also consider, when appropriate, the use of certain project results for signing exploitation agreements with third parties. This project has the potential to transfer results, and we included in the task plan the academic formalisation of problems of interest (for example, extraction, analysis, or search tools) for potential stakeholders. The team has experience in executing R&D contracts with companies. The planned approach for this project consists of, where appropriate, registering the intellectual property of the software so that universities can exploit it through contracts or exploitation agreements. This is compatible with sharing other project results with the scientific community (for example, datasets, linguistic resources, and algorithms to favour reproducibility). This is something we have been doing regularly.

4. The relevance of the project to RCIS

Our project is aligned and relevant to the following key topics of the RCIS conference:

- Information Search and Discovery: By mining publicly available web content, the project aims to discover and extract traces and evidence related to gambling disorders, contributing to the improvement of information search and discovery techniques in this domain.
- Big Data & Business Analytics: Through the application of data science techniques and analytics, the project addresses the challenge of processing vast amounts of web data to identify linguistic patterns indicative of gambling disorders.
- Digital Transformation: The project represents a digital transformation initiative aimed at leveraging advanced technologies to address societal issues such as gambling disorders through innovative approaches in web mining and natural language processing.
- Social Computing: Understanding the way people use language and examining the evolution of language usage patterns in relation to gambling disorders aligns with the principles of social computing, offering insights into multiple user profiles of interest and their interactions in online environments.
- Health Informatics and E-Health: Given the focus on early detection of pathological gambling, the project intersects with the e-health domain, where technological advancements play a crucial role in preventing psychological disorders related to gambling in individuals.

5. Conclusions

In this paper we have presented the project entitled "Web Mining, Search Technologies and Natural Language Processing for the Early Detection of Pathological Gambling", funded by Ministerio de Consumo, Subdirección General de Regulación del Juego (Government of Spain).

This project focuses on technologies and computational models that perform large-scale natural language analysis. The aim is to design and implement new monitoring and analytical tools that, from publicly available information on the web, mine contents and extract traces and evidence related to gambling disorders. More specifically, the main goal is to study the way people use language (and the evolution of the use of language) to reveal early signs of gambling disorders.

Acknowledgments

References

[1] ICD-11: International classification of diseases (11th revision), 2022. URL: https://icd.who.int/.

[2] D. American Psychiatric Association, A. P. Association, et al., Diagnostic and statistical manual of mental disorders: DSM-5, volume 5, American psychiatric association Washington, DC, 2013.

[3] G. Humphreys, Sharpening the focus on gaming disorder, World Health Organization. Bulletin of the World Health Organization 97 (2019) 382–383.

[4] J. Billieux, M. Flayelle, H.-J. Rumpf, D. J. Stein, High involvement versus pathological involvement in video games: A crucial distinction for ensuring the validity and utility of gaming disorder, Current Addiction Reports 6 (2019) 323–330.

[5] J. W. Pennebaker, M. R. Mehl, K. G. Niederhoffer, Psychological aspects of natural language use: Our words, our selves, Annual review of psychology 54 (2003) 547–577.

[6] J. Lin, R. Nogueira, A. Yates, Pretrained transformers for text ranking: Bert and beyond, Springer Nature, 2022.

[7] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[8] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, Deep learning–based text classification: a comprehensive review, ACM computing surveys (CSUR) 54 (2021) 1–40.

[9] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, et al., Sparks of artificial general intelligence: Early experiments with gpt-4, arXiv preprint arXiv:2303.12712 (2023).

[10] E. A. Ríssola, D. E. Losada, F. Crestani, A survey of computational methods for online mental state assessment on social media, ACM Transactions on Computing for Healthcare 2 (2021) 1–31.

[11] F. Crestani, D. E. Losada, J. Parapar, Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the ERisk Project, volume 1018, Springer Nature, 2022.

[12] Ludopatía/adicción al juego. Grupo público de Facebook., 2011. URL: https://www.facebook.com/groups/253782884636115, [accessed February 13, 2024].

[13] Ludopatia.org. Por la rehabilitación de jugadores patológicos y otras adicciones. Foro de discusión., 2010. URL: https://www.ludopatia.org/forum/default.asp, [accessed February 13, 2024].

[14] Vida sin Juego. Foro de Discusión., 2009. URL: https://vidasinjuego.forosactivos.net/, [accessed February 13, 2024].

[15] Ludopatía, adicción y problemas con el juego. Foro de discusión., 2008. URL: http://foroapuestas.forobet.com/ludopatia-adiccion-y-problemas-con-el-juego/, [accessed February 13, 2024].

[16] Ludopatía. Foro de discusión., 2022. URL: https://www.forolinternas.com/viewtopic.php?f=16&t=17505, [accessed February 13, 2024].

[17] Ludopatía online. Plataforma de ayuda personal., 2024. URL: https://ludopatiaonline.com/foro-ludopatia/?amp, [accessed February 13, 2024].

[18] Jugadores anónimos. Plataforma de apoyo., 2024. URL: https://www.jugadoresanonimos.org/, [accessed February 13, 2024].

[19] D. Hovy, S. L. Spruit, The social impact of natural language processing, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2016, pp. 591–598.

[20] A. Olteanu, C. Castillo, F. Diaz, E. Kıcıman, Social data: Biases, methodological pitfalls, and ethical boundaries, Frontiers in big data 2 (2019) 13.

[21] D. E. Losada, F. Crestani, A test collection for research on depression and language use, in:

International conference of the cross-language evaluation forum for European languages, Springer, 2016, pp. 28–39.

[22] D. E. Losada, F. Crestani, J. Parapar, erisk 2017: Clef lab on early risk prediction on the internet: experimental foundations, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017, Proceedings 8, Springer, 2017, pp. 346–360.

[23] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk: early risk prediction on the internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, September 10-14, 2018, Proceedings 9, Springer, 2018, pp. 343–361.

[24] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk 2019 early risk prediction on the internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 10th International Conference of the CLEF Association, CLEF 2019, Lugano, Switzerland, September 9–12, 2019, Proceedings 10, Springer, 2019, pp. 340–357.

[25] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk at clef 2020: Early risk prediction on the internet (extended overview)., CLEF (Working Notes) (2020).

[26] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, Overview of erisk at clef 2021: Early risk prediction on the internet (extended overview)., CLEF (Working Notes) (2021) 864–887.

[27] P. Martın-Rodilla, D. E. Losada, F. Crestani, Overview of erisk 2022: Early risk prediction on the internet, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy, September 5–8, 2022, Proceedings, volume 13390, Springer Nature, 2022, p. 233.

[28] D. E. Losada, M. Herrmann, D. Elsweiler, Cost-effective identification of on-topic search queries using multi-armed bandits, in: Proceedings of the 36th Annual ACM Symposium on Applied Computing, 2021, pp. 645–654.

[29] J. M. Chenlo, J. Parapar, D. E. Losada, Comments-oriented query expansion for opinion retrieval in blogs, in: Advances in Artificial Intelligence: 15th Conference of the Spanish Association for Artificial Intelligence, CAEPIA 2013, Madrid, Spain, September 17-20, 2013. Proceedings 15, Springer, 2013, pp. 32–41.

[30] L. Azzopardi, R. T. Fernández, D. E. Losada, Improving sentence retrieval with an importance prior, in: Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010, pp. 779–780.

[31] D. E. Losada, J. Parapar, A. Barreiro, A rank fusion approach based on score distributions for prioritizing relevance assessments in information retrieval evaluation, Information Fusion 39 (2018) 56–71.

[32] P. Gamallo, M. Garcia, Linguakit: uma ferramenta multilingue para a análise linguística e a extração de informação, Linguamática 9 (2017) 19–28.

[33] R. Martínez-Castaño, J. C. Pichel, D. E. Losada, A big data platform for real time analysis of signs of depression in social media, International journal of environmental research and public health 17 (2020) 4752.

[34] R. Martínez-Castaño, J. C. Pichel, D. E. Losada, Building python-based topologies for massive processing of social media data in real time, in: Proceedings of the 5th Spanish Conference on Information Retrieval, 2018, pp. 1–8.

[35] S. E. Baumgartner, T. Hartmann, The role of health anxiety in online health information search, Cyberpsychology, behavior, and social networking 14 (2011) 613–618.

[36] S. Lopez-Larrosa, V. Sánchez-Souto, D. E. Losada, J. Parapar, A. Barreiro, A. P. Ha, E. M. Cummings, Using machine learning techniques to predict adolescents' involvement in family conflict, Social Science Computer Review 41 (2023) 1581–1607.