

# **Bayesian latent class model: A framework to study the prevalence and the performance of algorithms in bot detection**

*Dr. Bekir Cetintav,*

*Burdur Mehmet Akif Ersoy University, Veterinary Faculty, Department of Biostatistics, TURKEY*

*Hakki Besim Bardakci,*

*Arabam.com, Research&Development Center, Istanbul, TURKEY*

## ***Abstract***

Bot detection is the process of analyzing all traffic to a website, mobile app or API to detect and block malicious bots while allowing access to human visitors and partner bots like google, bing, etc. Most studies address the problem of bot detection based on blacklisted IP information and available knowledge of user behavior. Labeling is often done manually or using heuristic methods, especially for advanced bots, and algorithms work based on these labeled data. There is no gold standard method used in this regard. Therefore, it is very difficult to determine the exact amount of bot traffic (true prevalence) and performances of algorithms used in classification with standard metrics used in ML. In such cases, when only imperfect reference tests or insufficient numbers of appropriate reference samples of known status are available, Frequentist and Bayesian Latent Class Models (LCMs) can be used to draw inferences on test accuracy and true prevalence. In this study, a real-time bot detection application made on one of Turkey's leading ad listing websites is included. The results show that BLCM, which is used successfully in the field of epidemiology to determine the performance of methods for the detection of real viruses and prevalence in the absence of a gold standard, can also be used in the field of web bot detection.

## ***Keywords***

Bayesian Latent Class Models, web bot detection, machine learning, performance metrics

## ***Acknowledgements***

This study is supported by COST Action CA18208 / Novel tools for test evaluation and disease prevalence estimation (HARMONY).