UNIVERSITÀ DI PISA

**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**Dottorato di Ricerca in Ingegneria dell'Informazione**

Doctoral Course

# "Web usage mining on e-commerce websites"

Prof. Grażyna Suchacka, PhD

*University of Opole, Poland*

*gsuchacka@uni.opole.pl*

**Short Abstract:** The course deals with the application of machine learning methods to Web data, in particular in the context of two research problems: detecting Web bots and predicting purchases in online stores. The first problem is due to the presence of artificial agents on the Web which pose a threat to the website security, privacy, and performance. Continuous development of artificial agents' technology makes detection of Web bot, both in the offline and real-time settings, harder and harder. The second problem is connected with discovering various user profiles on e-commerce websites and identifying user sessions with high probability of making a purchase. The problems under consideration are key issues in the era of the rapid development of e-commerce, advanced Web-based technologies, and big data.

**Course Contents in brief:**

1. Introduction to Web usage mining (1 hour)
   Web usage data. Web usage mining in the context of online stores. Data pre-processing for Web usage mining. Reconstruction of user sessions.

2. Web bot detection (2 hours)
   Problems of offline and online bot detection. Characteristics and differences in bot and human Web traffic. Feature selection and feature extraction for bot detection. Offline bot detection with supervised and unsupervised learning methods. Online bot detection.

3. Online purchase prediction (1 hour)
   Problem of predicting online purchases. Feature selection for online purchase prediction. Purchase prediction with machine learning methods.

**Total # of hours of lecture**: 4 hours

**References:**

[1] 2023 Bad bot report. Imperva, https://www.imperva.com/resources/reports/2023-Imperva-Bad-Bot-Report.pdf.

[2] Cabri A., Masulli F., Rovetta S., Suchacka G.: A quantum-inspired classifier for early Web bot detection. IEEE Transactions on Information Forensics and Security, Vol. 17, IEEE, April 2022, 2022, pp. 1684-1697.

[3] Suchacka G., Cabri A., Rovetta S., Masulli F.: Efficient on-the-fly Web bot detection. Knowledge-Based Systems, Vol. 223, Elsevier, July 2021, 107074.

[4] Rovetta S., Suchacka G., Masulli F.: Bot recognition in a Web store: An approach based on unsupervised learning. Journal of Network and Computer Applications, Vol. 157, Elsevier, May 2020, 102577.

[5] Suchacka G., Iwański J.: Identifying legitimate Web users and bots with different traffic profiles – an Information Bottleneck approach. Knowledge-Based Systems, Vol. 197, Elsevier, June 2020, 105875.

[6] Iwański J., Suchacka G., Chodak G.: Application of the Information Bottleneck method to discover user profiles in a Web store. Journal of Organizational Computing and Electronic Commerce, Vol. 28, No. 2, Taylor & Francis, 2018, pp. 98-121.

[7] Suchacka G., Chodak G.: Using association rules to assess purchase probability in online stores. Information Systems and e-Business Management, Vol. 15, Issue 3, Springer, August 2017, pp. 751-780.

---

**CV of the Teacher**

Grażyna Suchacka received the M.Sc. degrees in Computer Science and in Management, as well as the Ph.D. degree in Computer Science (with distinction), from Wroclaw University of Science and Technology (Poland). Now she is an Assistant Professor in the Institute of Informatics at the University of Opole (Poland). Her research interests include data analysis and modelling, data mining, and Quality of Web Service with special regard to bot detection and electronic commerce support.

**Final Exam:** Yes (an online questionnaire)

**Room and Schedule**

Room: *Aula Riunioni del Dipartimento di Ingegneria dell'Informazione, Via G. Caruso 16, Pisa – Ground Floor*

Schedule:

Day1 – 21/05/2024, 9:00-13:00