PREDICTING STUDENT PERFORMANCE FROM THEIR BEHAVIOR IN LEARNING MANAGEMENT SYSTEMS

TILBURG UNIVERSITY

Parisa Shayan   Menno van Zaanen

p.shayan@uvt.nl   m.m.vanzaanen@uvt.nl

TiCC
Tilburg Centre for Cognition and Communication

Introduction

- Throughout the curriculum, students' performance is measured
  Formative During courses (How are students doing?)
  Summative After courses (Have students attained the required level?)
- Testing is quite labor intensive (for lecturer and student)
- Can we test (or cluster) students based on their online behavior?
- LMS provides accurate insight into students' online behavior
- Goal: Non-intrusive formative testing
  - More personalized education
    - Strong students may require additional challenges
    - Weak students may require extra help
  - Course improvements

Background

- Student performance prediction is major focus of LA and EDM
  (Romero and Ventura, 2013)
- Most LA and EDM research used LMS data for summative assessment
  (Minaei-Bidgoli and Punch, 2003; Morris et al., 2005; Zafra and Ventura, 2009; Macfadyen
  and Dawson, 2010; Romero and Ventura, 2013; Zacharis, 2015)
- Student characteristics and past performance have higher predictive value than LMS data
  (Tempelaar et al., 2015; Conijn et al., 2017)
- Important: include student characteristics and past performance in predictive model

Research questions

On-going research:
- Which properties describe student performance *during* the course?
- Do these properties provide insight in an educational context?
- How can we give the properties in a meaningful and understandable way to lecturers?
- How generally applicable are the results?

Dataset

- Online educational behavior of 426 Eindhoven University of Technology students
- Data from 5 courses: 888 students (some students participate in multiple courses)
- Fall and winter quarters of 2014–2015
- Information from Moodle LMS
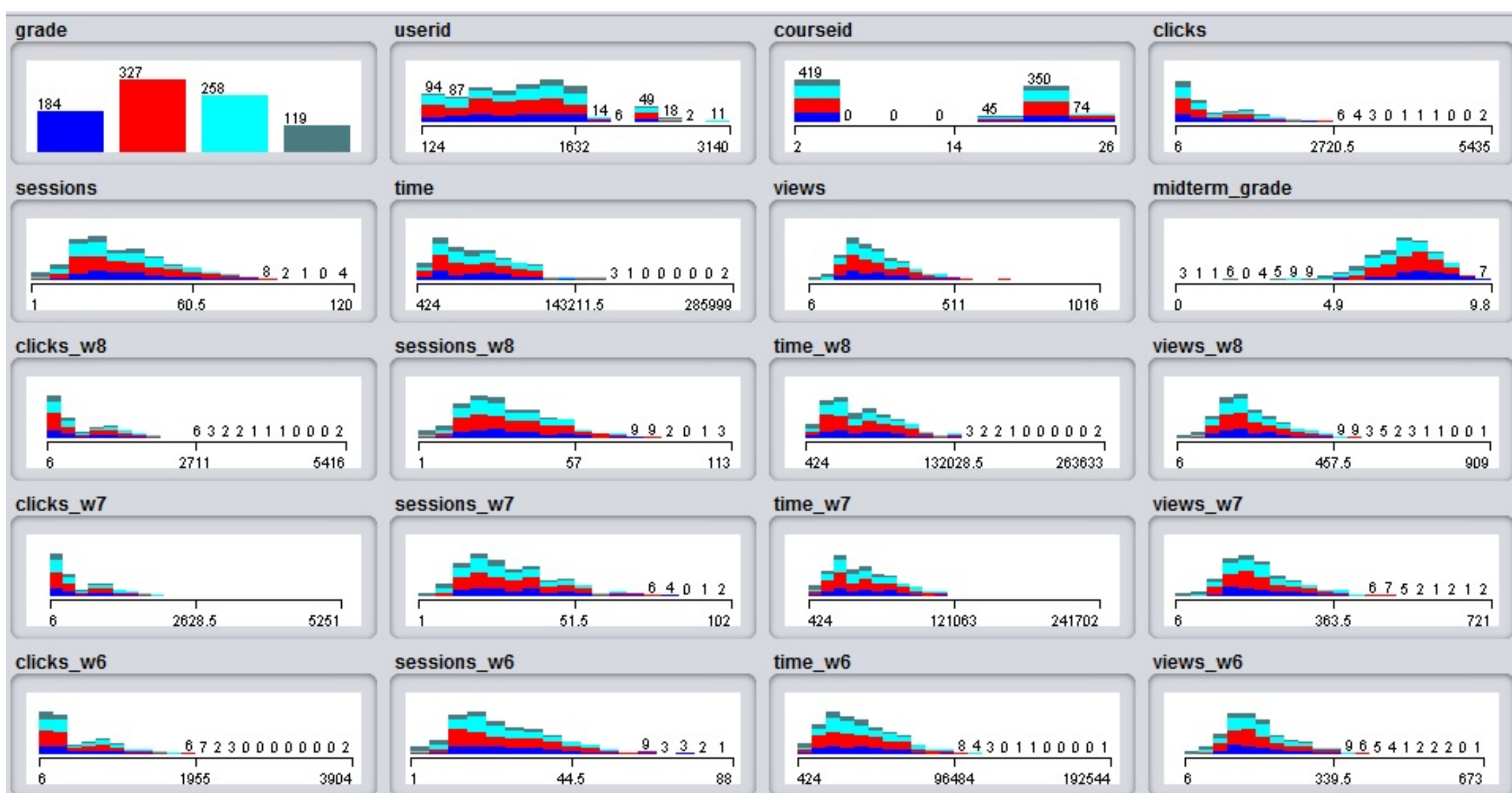
- Strong and weak students are found throughout dataset

Issues

- Most previous work uses all data (i.e., after the course has finished)
  - Intervention requires knowing student performance *during* the course
- Predictive model should provide *understandable* information
  - Lecturers may use this information to help students or improve course

Predictive model

- Decision tree using data early during course
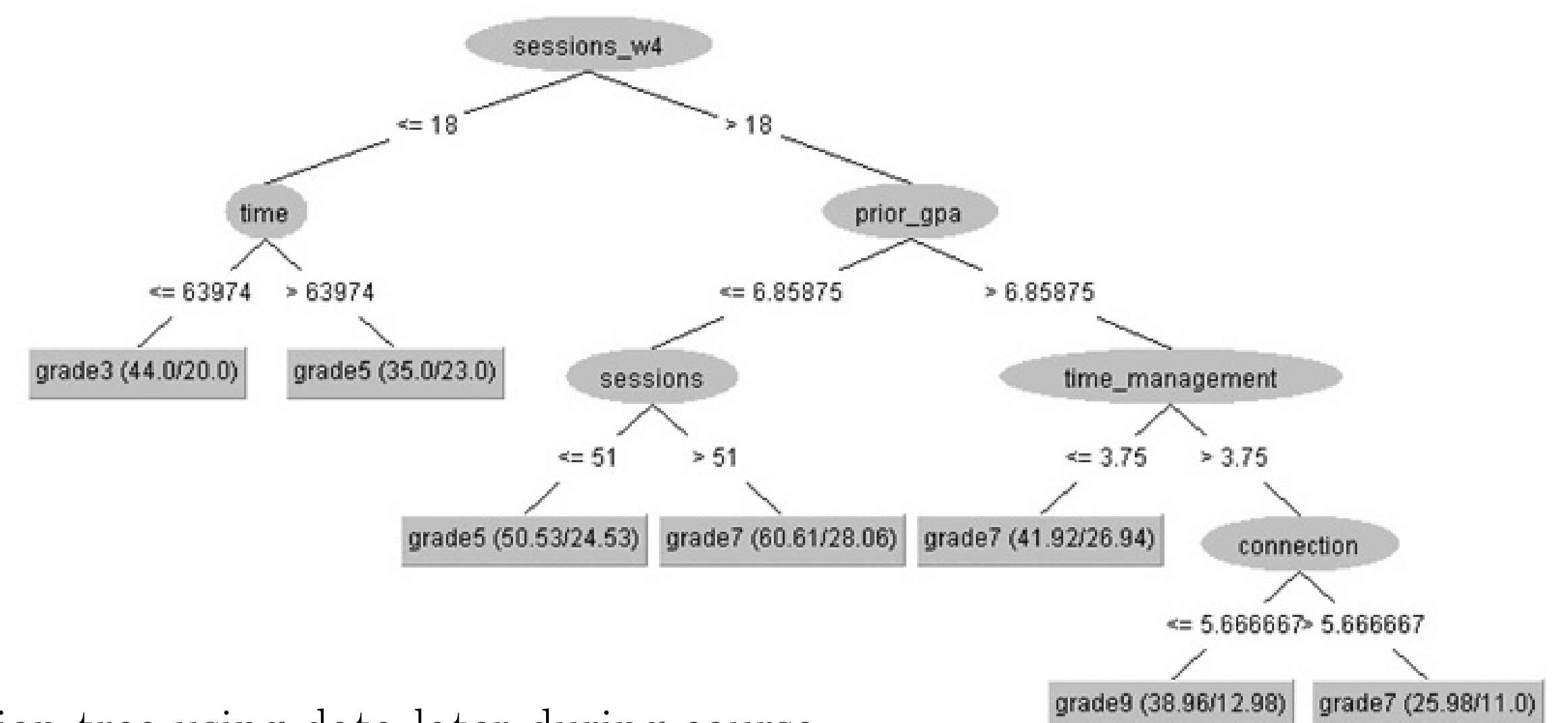
- Decision tree using data later during course
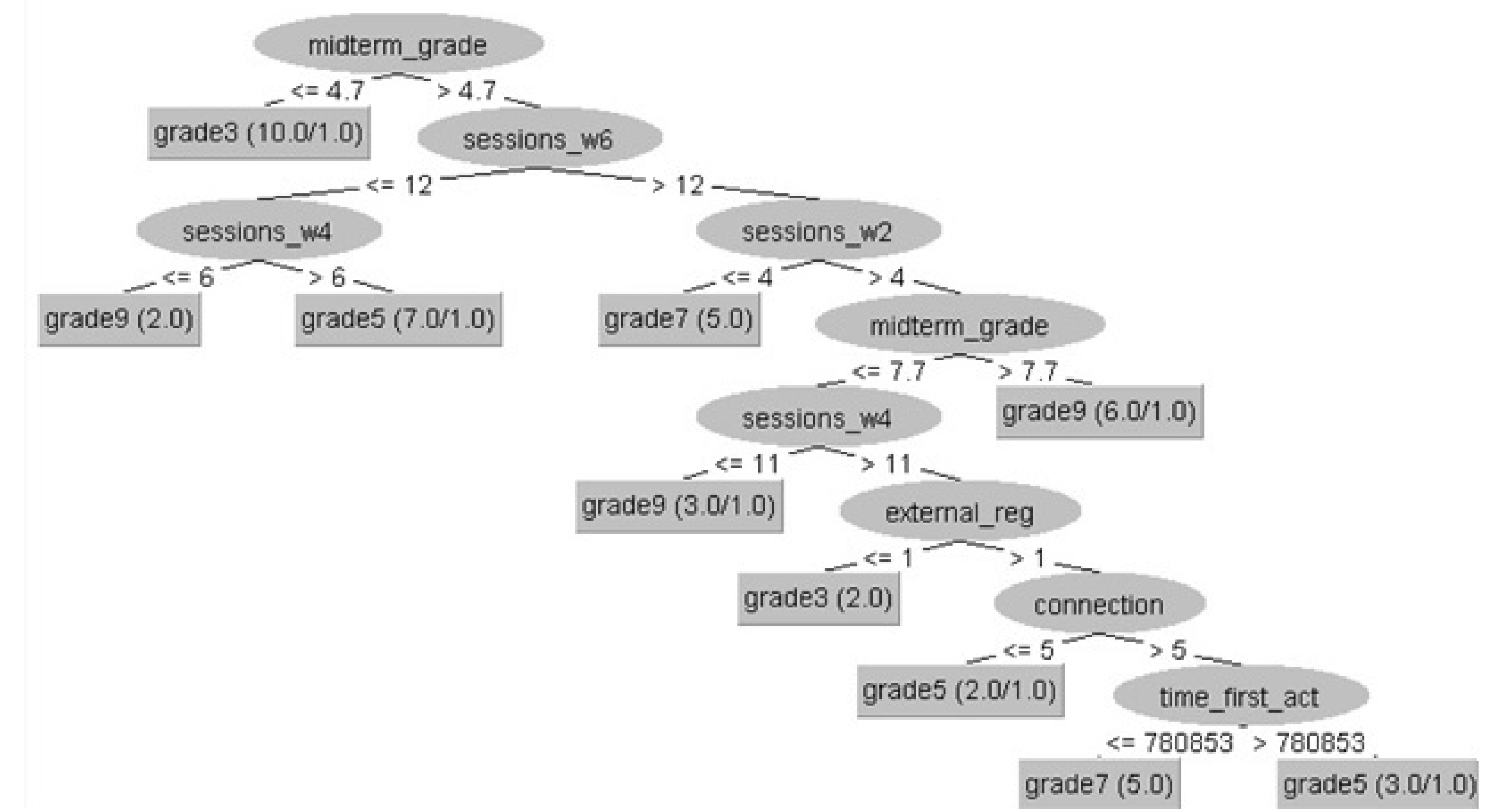
Conclusion

- Online student behavior is accessible data
  to investigate student performance
- Learning analytics and educational data mining
  can be used to extract understandable and useful patterns
- Results can
  - identify weak/strong students
  - identify course properties that may be improved
- Extracted patterns may help educational institutes and lecturers
  - improve decision making about educational courses
  - improve course quality (informed student differentiation)

Future work

- Generalization of results is not (yet) known
- Unclear exactly which features
  - have a large impact on prediction performance
  - make sense in an educational setting
  - are useful for lecturers
- Different machine learning methods
  - have different classification performance
  - result in different types of human readable information

References

Conijn, R., Snijders, C., Kleingeld, A., and Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1):17–29.

Macfadyen, L. P. and Dawson, S. D. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & education*, 54(2):588–599.

Minaei-Bidgoli, B. and Punch, W. F. (2003). Using genetic algorithms for data mining optimization in an educational web-based system. In *Genetic and evolutionary computation conference*, pages 2252–2263, Berlin, Heidelberg. Springer.

Morris, L. V., Finnegan, C., and Wu, S.-S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3):221–231.

Romero, C. and Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1):12–27.

Tempelaar, D. T., Rienties, B., and Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47:157–167.

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27:44–53.

Zafra, A. and Ventura, S. (2009). Predicting student grades in learning management systems with multiple instance genetic programming. *International Working Group on Educational Data Mining*.