The University of Southampton (UoS) have delivered various Massive Open Online Courses (MOOCs) through the FutureLearn platform. The number of participants vary in the range of few thousands up to fifteen thousands. This makes interaction, follow up, grading and other learning process activities particularly challenging, not to mention the diversity of participants’ cultures, countries, languages and backgrounds. The data belonging to users’ activity is generated daily by FutureLearn, anonymized and sent in CSV format to the partner institutions, with a basic report.

To discover patterns and previously unknown facts using data science techniques, to inform stakeholders’ actions to improve courses’ content, engagement and satisfaction.

**Visual Analytics**

We engineered some features, defined after a detailed analysis of the raw data such as activity duration times (derived from timestamps).

Participants may take longer than estimated in various types of learning activity. What factors could be affecting their learning experience?

Uncovering these could inform the course design.

**Prediction Model**

- Highly unbalanced classes (Figure 4)
- How many participants are likely to leave? (Predicting behaviour one and two weeks after the week of interest).
- Prediction of participants eligible for a certificate? (eligibility is determined by completion rate higher than 50%)

**Conclusions**

- Data collection processes can be improved to gather better demographics and course related details.
- Visual Analytics of transformed data may offer insights that can inform course design by highlighting potential areas of improvement.
- Prediction modelling requires several iterations with different engineered features to improve its accuracy and consistency. Due to data collection differences, predictors may differ from those in other MOOC platform projects.

**Bibliography and Acknowledgments**


Some images are taken from Wikimedia Commons. The FutureLearn logo is property of FutureLearn.

This work has been partially supported by the EPSRC ERDF Partnership Project “The MOOC Observatory: Dashboard Management, analysis and visualization of MOOC data”, Southampton, 2016.