

# Data Science Game

Benjamin DONNOT, Antoine MIECH, Thibault LAUGEL, Antoine LY

October 2016

The Data Science Game is the first international datascience university competition, gathering students from all around the world. This competition is open to students : from first year of master up until last year of PhD. It aims both at promoting machine learning and benchmarking the level of students from the best datascience programs worldwide.

The event takes place each year; as for now, two sessions have been organized in 2014-2015 and 2015-2016.

One of the particularities of the Data Science Game is its format, as it is divided in two phases: a first online eliminatory phase where students, assembled in teams of 4, have around 5 weeks to solve a machine learning prediction problem. Past online phases included a natural language processing and a computer vision problem. It aims at assessing students broad machine learning skills. During this first phase of the competition, participants are encouraged to communicate with each other, with their professors or even with professionals. The 20 teams that manage to have the best prediction are then qualified for the final phase. However, in order to keep the competition as international as possible, only one team per university can take part in the final phase.

The format of the final phase is a hackathon: qualified students are invited in one place for an entire week-end to solve a final problem, usually given by one of the main sponsors of the event.

In order to help the students tackling this real-world problem they may not be used to, the Data Science Game also invite mentors. Mentors are usually datascientists working in companies, or coming from the academic world. Their diverse profiles, as well as their expertise help the students improving their results a lot in a short period of time. For the students taking part in this final phase, this is a unique opportunity to strengthen their academical skills competing on an almost real world problematic with real data.

The final phase of the two previous editions took part in a castle near Paris provided by one of our sponsors. Gathering 80 students from all around the world and a dozen mentors coming from both academic and big international companies is highly profitable for everyone and help people know each other and share expertise.

At the end of the competition, teams are asked to present their work. After only a few hours of sleeping, students must clear their minds enough to be able to sum up in a very short time (3 minutes) what they have done for the previous 30 hours: the methods they

used, the difficulties they faced, etc. This exercise allows the teams to expose their communication skills, which are also very important in the real world.

**Previous Edition.** Last year eliminatory phase gathered 143 teams from more than 50 different prestigious universities (such as Stanford, National University of Singapore, Columbia University, Cambridge University, Ecole Polytechnique or Moscow State University...), from 28 different countries from 5 different continents. The level was very high as some top ranked kagglers were participating to the event. It has been supported by various sponsors such as: Google, Microsoft, AXA, CapGemini but also some other institutions such as ChaLearn or Etalab.

**Next Edition.** Considering the spectacular growth witnessed between the two first editions, the Data Science Game is still expected to grow a lot in terms of number of participants, universities and countries involved. The association organizing the event

is welcoming any organization or company that would like to sponsor the event and / or any suggestion of interesting data set for both phases of the challenge.

The organizing team hope that they will be able to continue this wonderful experience for the years to come : both students and sponsors seem to have enjoyed this event, and gave great and positive feedback. A lot will be done to continue improving the quality of the next editions. Many thanks to our sponsors especially CodaLab and Chalearn for making it possible.

**Acknowledgement** This work would not have been possible without the help of all the members of the Data Science Game association. We would like to thank them for their work. Without the help of our sponsors : AXA, Capgemini, Microsoft, Google and many others this challenge could not exist. We would also like to thank them. For more information, please visit "[www.datasciencegame.com](http://www.datasciencegame.com)".