1SC4891 – E-reputation

Instructors: Bich-Lien Doan
Department: DOMINANTE - INFORMATIQUE ET NUMÉRIQUE
Language of instruction: FRANCAIS
Campus: CAMPUS DE PARIS - SACLAY
Workload (HEE): 40
On-site hours (HPE): 27,00

Description

EI-1: E-reputation: sentiment analysis and NLP

Information management is strategic for Octopeek. Its importance is measurable by the volume of information processed, by the speed of information evolution as well as by the time spent in the activity of information research. This issue has led to the emergence in recent years of new businesses around Data Science and Big Data.

Context and objective :

Octopeek SAS proposes to share its business experience on the development of a Data Science model for customer needs. The aim is to qualify the e-reputation of a company, a brand, a personality by analyzing feelings on data from social networks or the Web.

The objective is to make an analysis of feelings about a company, a group or simply a phenomenon in order to determine the causality of these feelings on the future of the entity concerned. Example: analysis of feelings on a television channel "HBO" which proposes a trend series currently "Game Of Thrones". At the business level, the channel's managers are trying to find out whether it would be interesting to launch a spin-off of this series soon or not. To do so, they need to conduct a study on how fans feel about their series. The best approach is to collect the reaction of fans on a source that ensures we have access to so-called "hot" or real-time data. The data collections used for testing will be either public collections available on the web (Twitter) or collections provided by Octopeek.

Quarter number
ST4

Prerequisites (in terms of CS courses)
none

Syllabus
In this project, students choose a topic on which to conduct a study on users' feelings, analysing the feelings related to their comments. The first step is to collect tweets from real-time data sources. Then, after cleaning up the collected data, the students experiment with algorithms to get the best recommendation and use the precision measures to evaluate the chosen model. The implementation of the project will be in the form of a challenge between the teams, which will allow them to confront the best algorithm at the end of the project.

Class components (lecture, labs, etc.)
project work
real-life setting

Grading
continuous checking
oral defense

Learning outcomes covered on the course

Understand the customer's problem and translate the need
Identify Data Sets needed for the model
Collect, cleanse and qualify data
Elaborate a model using the arsenal of datascience algorithms
Validate the model