Big data analytics for dummies
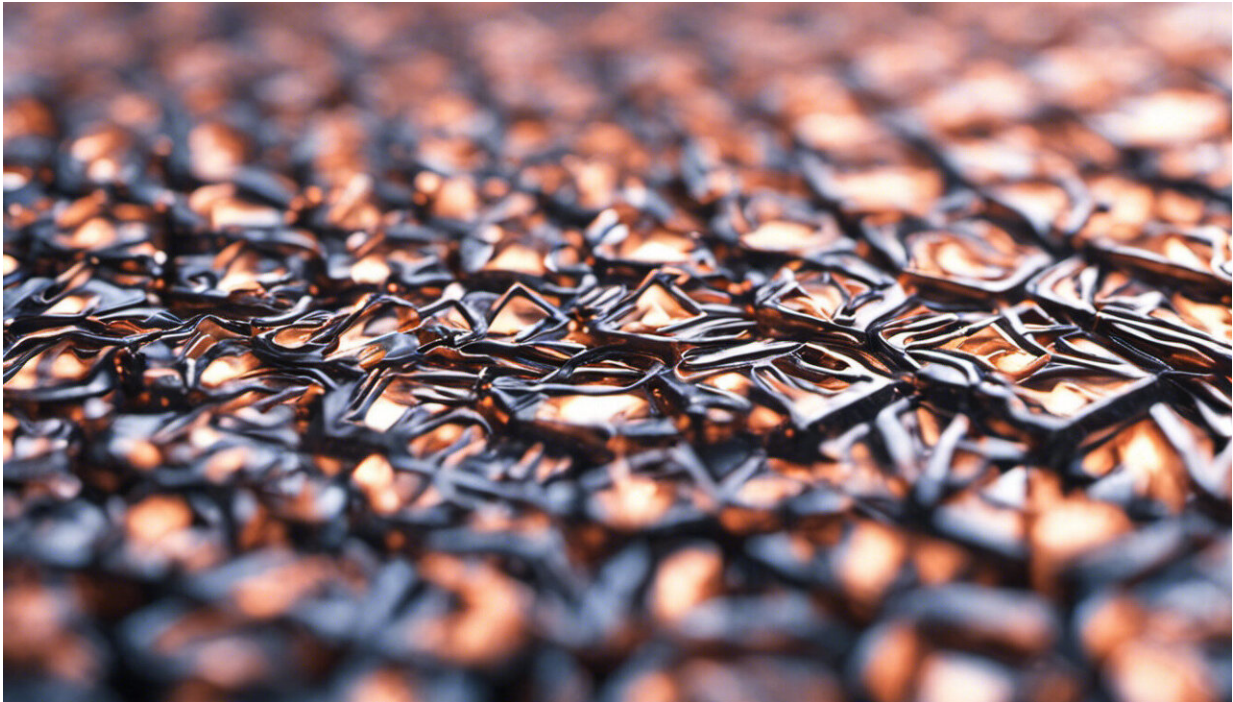
June 28 2018

Credit: AI-generated image (disclaimer)

Big Data is still very much an elite thing: only the most IT-savvy and wealthy businesses have a shot at scratching the surface of its potential. All this could be about to change thanks to a Big Data analytics platform developed under the TOREADOR project, which will automatically handle all major problems related to on-demand data preparation.

"Expectations of Big Data are very high, but the gap between ambition

and execution is still large, especially for SMEs," Dr. Ernesto Damiani sighs. And he should know: since early 2016, Dr. Damiani has been leading a 10-strong consortium looking into the reasons for these mixed fortunes and the possible solutions.

If relatively few SMEs have incorporated Big Data analytics into their offerings or internal processes, it's mainly for two reasons. The first is a lack of competence in Big Data analytics, as Dr. Damiani explains. A company willing, for instance, to tailor its offerings to customer behaviour using a free app would have to resort to very expensive consultancy. It's currently the only way to map business goals to a class of data science and technology solutions.

"Concretely, the project brief could be something along the lines of 'collect the events generated by core customers' apps and use them to train a scalable random-forest multi-category classifier of their behaviour to be deployed on a public cloud service'," he says.

The second reason is the long roll-out time and, again, the prohibitive cost of Big Data campaigns even when the data science approach has already been identified. Together, these problems have been keeping SMEs and non-ICT-savvy businesses away from Big Data analytics, although they account for a substantial share of the EU manufacturing backbone.

The TOREADOR (TrustwOrthy model-awaRE Analytics Data platfORm) methodology and toolkit offer a solution to both problems: they automate and commoditise Big Data analytics, while making its tailoring to domain-specific customer requirements much easier than before.

The TOREADOR framework supports two automated transformations. The first one starts from a machine-readable declarative model which

collects the data owner goals, and ends in a technology independent semantics-aware procedural model describing the computation to be carried out. Then, the second transformation builds upon the procedural model to compute a technology dependent deployment model. The latter can be executed on an Apache platform, at the customer's premises, on commercial cloud services like AWS, as Python code executable on the Azure platform or as a Docker container.

"Our declarative models can interactively collect the business goals of Big Data campaigns and allow the TOREADOR toolkit to provide automatic advice on the feasibility of solutions. Our procedural models then provide an innovative description of the Big Data analytics computation in the OWL/S semantics-aware standards, and our compilers translate these procedural models into fully executable workflows or even into natively parallelised Python code. We're looking at an iterative development process, where non-IT-savvy users can quickly set up a campaign by generating a workflow executable on a public cloud service, and then – if needed – call in developers for generating self-contained Python code," Dr. Damiani explains.

Project partners have already identified four industrial pilots in the fields of predictive aircraft engine maintenance, predictive management of solar power plants, business application logs analysis, and clickstreams analysis for e-commerce applications.

"The TOREADOR platform is available and has been deployed at the four pilot sites. It has also been made available as a free pre-release to selected members of the TOREADOR community, which is composed of European companies (several of them SMEs) recruited with the help of TAIGER (Spain), an innovative SME in the TOREADOR consortium. Details on these early adopters are available on our website. Besides, the TOREADOR methodology has been released to other European projects using Big Data campaigns like EVOTION," Dr.

Damiani says.

The project is scheduled for completion at the end of 2018. Until then, the consortium intends to keep enlarging the catalogue of services available in the platform and provide examples of TOREADOR-enabled Big Data campaigns, including training and deployment of advanced machine learning models.

Provided by CORDIS

Citation: Big data analytics for dummies (2018, June 28) retrieved 28 June 2024 from https://phys.org/news/2018-06-big-analytics-dummies.html