Summary Post

by Constantinos Kyriacou - Thursday, 10 November 2022, 1:27 AM

As our society and our knowledge evolve, so does the way we can use data. Nowadays , the decisions of organizations and companies are fully guided by large masses of data which are imported from different data sources ,in order to be processed and analyzed for this purpose. The completion of this procedure , can be achieved much more efficiently since technology and IoT (internet of things) are progressing rapidly.

However, even with the rapid evolution of technology sometimes the processing and analysis of big data can be very time-consuming and costly. Additionally, while storing our data we need to be sure that our data sources are reliable and will lead us to the right conclusions and decisions, otherwise erroneous decisions may lead to disastrous consequences. Furthermore, it is required that the General Data Protection Regulations (GDPR) are fully respected during the data collection, data storage and data analysis procedures. More precisely, data coming from various IoT devices, may contain some personal details which only certain people will be able to access. Missing and incomplete data ,make the task of the specialist even more difficult ,since they have to keep in mind the appropriate configuration of the data as well.

Therefore, it becomes necessary for a data scientist to take into consideration these challenges. Data needs to be accessed timely, accurately, and efficiently. Data cleaning is a fundamental procedure which contributes in preparing our database for processing and analysis. During this procedure, bad and irrelative data are being removed to simplify our analysis process. At the same time, handling missing and incomplete data is a fundamental part of this procedure which a data scientist should be aware.

Data is all around us in our everyday life. Hence, the correct harvesting and processing of data is of utmost importance for modern companies. On the other hand, irresponsible data management can be disastrous and affect the smooth operation of the business.

References:

1)Forgó, N., Hänold, S. and Schütze, B., 2017. The principle of purpose limitation and big data. In *New technology, big data and the law* (pp. 17-42). Springer, Singapore.

2)Ji, C., Li, Y., Qiu, W., Jin, Y., Xu, Y., Awada, U., Li, K. and Qu, W., 2012. Big data processing: Big challenges and opportunities. *Journal of Interconnection Networks*, *13*(03n04), p.1250009.

3) Huxley, E 2020, *Big data architectures,* viewed 26 September 2022, <**https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/** > .

4)Krasnow Waterman, K. and Bruening, P.J., 2014. Big Data analytics: risks and responsibilities. *International Data Privacy Law*, *4*(2), pp.89-95.

5)Agrawal, D., Bernstein, P., Bertino, E., Davidson, S., Dayal, U., Franklin, M., Gehrke, J., Haas, L., Halevy, A., Han, J. and Jagadish, H.V., 2011. Challenges and opportunities with Big Data 2011-1.