A MISSING VALUE APPROACH ON FACEBOOK BIG DATA: LIKE, DISLIKE OR NOTHING?

Andrea Marletta [1], Paolo Mariani[1] and Erika Grammatica [2]

[1] Department of Economics, Management and Statistics, University of Milano-Bicocca, (e-mail: `andrea.marletta@unimib.it`, `paolo.mariani@unimib.it`)

[2] Crea MC, Milan, (e-mail: `e.grammatica@campus.unimib.it`)

In the last years, the quantity of socio-economic data has grown steeply thanks to the diffusion of Internet and the raise of electronic devices. The Internet diffusion lead to a paradigm change based on new tools. Data extracted from these tools could represent strategic sources to help companies to reach competitive positions for leading the markets. Social media represent an important instrument to mine data and raise the consumers' knowledge. They are able to declare their preferences just giving a "like". This study aims to inspect the mechanism behind users' behaviour. In particular, the attention will be focused on missing expression of the "likes". A statistical model has been proposed to discern a "Like" from a "Dislike" from a "Nothing". The proposed approach could help companies to measure how much a brand is known and how is good to influence users' choices.

KEYWORDS: Big Data, forecasting, Facebook, missing values

References

Little R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association: 1198-1202.

Huisman M. (2009). Imputation of missing network data: some simple procedures. Department of Psychology, University of Groningen, Grote Kruisstraat, the Netherlands.

Rubin D.B. (1976). "Inference and Missing Data." Biometrika n. 63: 581-592.