

## **Curriculum vitae di SERGIO BOLASCO**

Già docente, nel Master di “Data Science” presso la Università di Roma2 “Tor Vergata”, del corso di *Text Mining* negli anni accademici 2014-2015, 2015-2016 e 2016-2017.

Già professore ordinario di Statistica presso la Facoltà di Economia dell’Università di Roma “La Sapienza”, afferente al Dipartimento MEMOTEF, collocato a riposo dal 1 marzo 2012.

Già docente, nel Master di "Fonti Strumenti e Metodi per la Ricerca sociale" (FSeM) presso la Facoltà di Statistica dell'Università degli studi di Roma "La Sapienza", di un corso di "*Text mining e analisi dei dati testuali*" (fino al 2013).

### **CURRICULUM ESSENZIALE**

Laureato in Scienze Statistiche e Demografiche (1971).

Nell’ultimo a.a. di insegnamento (2014-2015), titolare dell’insegnamento di "*Metodi esplorativi per l'analisi dei dati e laboratorio di data mining*".

Nel periodo 2003-2009 nel corso di laurea specialistica “Metodi quantitativi per l’economia e l’impresa (METEI) presso la Facoltà di Economia della SAPIENZA, di cui è stato fra i promotori e coordinatore, ha tenuto i corsi di “*Laboratorio avanzato di analisi dei dati e data mining*” e di “*Casi di studio in Data mining e Text mining*”.

Già professore associato nei corsi di “*Statistica*” e di “*Matematica per le scienze sociali*” presso l’Università di Salerno, nel corso di laurea in Sociologia (1987-1995).

Già ricercatore presso la Facoltà di Scienze Statistiche, Dipartimento di Statistica, Probabilità e Statistiche Applicate dell’Università di Roma "La Sapienza" (1980-1987).

Già professore incaricato presso la Facoltà di Scienze Sociali della Pontificia Università Gregoriana dal 1973 al 2003. In quegli anni ha tenuto corsi di *Matematica*, di *Statistica*, di *Analisi quantitativa e qualitativa*, di *Informatica* e di *Metodi di analisi automatica dei testi*.

Dal 1978 al 2012 membro della Società Italiana di Statistica; già redattore della rivista internazionale METRON; dal 1993 al 1997 rappresentante per l'Italia nel consiglio dell'International Federation of Classification Societies (IFCS).

Dal 1993 è membro del comitato scientifico internazionale del JADT (*Journées internationales d'Analyse statistique des Données Textuelles*), nel cui ambito ha organizzato a Roma, nel dicembre 1995 le Terze Giornate Internazionali JADT2005 e nel 2010 la loro decima edizione (<http://JADT2010.uniroma1.it>).

Nell’arco degli ultimi venti anni ha promosso, partecipato e diretto numerose ricerche finanziate di interesse nazionale e internazionale sui temi dell’analisi dei dati e sul *text mining* con particolare riferimento alle tecniche esplorative di tipo multidimensionale e ai metodi di analisi automatica dei testi. Fra le altre, ha coordinato l’unità operativa italiana nel progetto europeo NEMIS (*Network of Excellence in Text Mining and its Applications in Statistics*) negli anni dal 2002 al 2004 (<http://nemis.cti.gr>).

Fin dall'anno 2000 ha ideato il pacchetto TaLTaC, software per l'analisi automatica dei testi e per il text mining, distribuito nella versione T1 (2000-2005) e T2 (2006-2010) e successiva attuale versione T3 specificamente orientata ad applicazioni di Big Data, sviluppata con tecnologia web.

Nel periodo 2004-2005 ha collaborato nell'ambito di un protocollo di ricerca fra Istat e cinque università italiane (Bologna, Torino, Roma-Sapienza, Milano-Bicocca, Trieste) relativamente alla Indagine sull'uso del tempo *TUS 2002-2003* al fine di incentivarne lo sfruttamento dei risultati, secondo un approccio multidisciplinare. In tale contesto il sottoscritto ha sviluppato proposte di analisi lessicale e testuale dei dati non strutturati contenuti nei diari delle attività quotidiane, proposte pubblicate in [14].

Ha svolto periodi di studio e ricerca all'estero nelle Università di Parigi, di Montpellier e di Lille in Francia, di Dubrovnic in Jugoslavia e di Trier in Germania.

Ha svolto seminari e scuole estive in varie Università italiane ed estere, fra cui: Besançon, Leon, Lille, Montreal, Paris, Toulouse, Bologna, Bolzano, Firenze, Modena, Napoli, Padova, Parma, Pavia, Perugia, Pisa, Pollenzo, Roma1, Roma2, Roma3, Salerno, Siena, Torino, Urbino, Viterbo.

Dal 2005 al 2009, in qualità di esperto straniero è stato membro del Conseil Scientifique de l'UMS 2913 Maison de Science de l'Homme di Besançon.

Negli anni 2011, 2012, 2013 e 2015 ha organizzato quattro edizioni di una Scuola internazionale sul tema "*Introduzione all'analisi automatica dei testi e strumenti di text mining*" di tipo interdisciplinare, che ha visto la partecipazione di linguisti, statistici, informatici, sociologi, psicologi e scienziati della comunicazione.

E' membro dell' Editorial Board delle riviste:

*International Review of Sociology*, edita da Routledge, Taylor and Francis Group

(<http://www.tandfonline.com/action/aboutThisJournal?show=editorialBoard&journalCode=cirs20>)

*Lexicometrica*, rivista on-line dell'Università di Paris 3, fondata da André Salem,

(<http://lexicometrica.univ-paris3.fr/comite.html>), .

*Linguisticae Investigationes*, Revue internationale de linguistique générale et française, fondata da Maurice Gross (<http://infolingu.univ-mlv.fr/LI/LINGVISTICAEINVESTIGATIONES.html>),

## COLLABORAZIONI CON ENTI NON UNIVERSITARI E ATTIVITÀ PROFESSIONALI

Dal 1975, ha svolto attività di promozione e/o direzione di numerose ricerche con funzioni di coordinamento scientifico, in Enti e Istituti di ricerca d'interesse nazionale o in Aziende, fra le quali: ABI - Associazione Bancaria Italiana; AGCM - Autorità Garante della Concorrenza e del Mercato; AGT Engineering; CENSIS - Centro Studi Investimenti Sociali; Culture - Ricerca e Comunicazione; Dipartimento della Funzione Pubblica della Presidenza del Consiglio dei Ministri; Federmeccanica; Fivol - Federazione Italiana per il Volontariato; Formstat; GEPI s.p.a.; INAIL; INDEC, Instituto Nacional de Estadística y Censos, Argentina; IRRE del Lazio; ISFOL; ISFORT - Istituto superiore di formazione e ricerca per i trasporti; INVALSI; Istat; Istituto Tagliacarne; MIPA - Consorzio per lo sviluppo delle Metodologie e delle Innovazioni nella Pubblica Amministrazione; Provincia di Trento; RAI Radio Televisione Italiana; Regione Lazio; Regione Puglia; Regione Autonoma Valle d'Aosta; Scuola Superiore della Pubblica Amministrazione; SOMEA - Società per la Matematica e l'Economia Applicate; Telecom Italia.

Ha collaborato con i seguenti Editori: Aracne, Carocci Editore, Cisu, Editori Riuniti, Electa Editrice, Eri - Edizioni RAI, Franco Angeli, Il Mulino, LED Edizioni Universitarie, Manifesto

Libri, Marzorati Editore, North Holland, Nuova Italia Scientifica, Sage, Casa Editrice Università “la Sapienza”, Springer-Verlag.

Ha svolto attività di referee per: CLADAG, COMPSTAT, IFCS, JADT, LED, Lexicometrica, Mots, Rassegna Italiana di Psicologia, Statistica Applicata - Italian Journal of Applied Statistics, SIS – Società Italiana di Statistica, Springer.

## SETTORI DI INTERESSE SCIENTIFICO

Metodi e tecniche di Text Mining; metodologie per il trattamento automatico dei dati testuali; analisi delle domande aperte in survey qualitative, focus-group, interviste non direttive, messaggi/forum in Internet; analisi del contenuto e del discorso. Software per l'analisi automatica dei testi e il text mining.

Analisi dei dati: tecniche statistiche multidimensionali, analisi fattoriali, metodi di classificazione automatica e di segmentazione, metodi multiway. Analisi socio-economiche sul territorio. Studi di customer satisfaction. Semiologia grafica e trattamento grafico dell'informazione.

### *Selezione delle pubblicazioni più recenti*

- [1] Bolasco, S., Baiocchi, F., Canzonetti, A., DeGasperis, G. (2016). TaLTaC3.0, un software multi-lessicale e uni-testuale ad architettura web , in D. Mayaffre, C. Poudat, L. Vanni, V. Magri, P. Follette (eds.), Proceedings of JADT 2016, CNRS University Nice Sophia Antipolis, Volume I, pp. 225-235.
- [2] Bolasco S. (2015). Sulla costruzione di un corpus per l'analisi automatica del linguaggio parlamentare dei leader, in L. Giuliano e P. Villani (eds.), "Il linguaggio della leadership politica tra la prima e la seconda Repubblica. Problemi di metodo e linee di ricerca", Camera dei Deputati, Roma, pp. 44-50.
- [3] Bolasco S., Di Pietro S., Giuliano L. (2015). Metafore concettuali e locuzioni verbali nel linguaggio dei leader, in L. Giuliano e P. Villani (eds.), "Il linguaggio della leadership politica tra la prima e la seconda Repubblica. Problemi di metodo e linee di ricerca", Camera dei Deputati, Roma, pp. 51-63.
- [4] Bolasco S. (1999-2014). *Analisi multidimensionale dei dati. Metodi, strategie e criteri di interpretazione*. Carocci, Roma, pp. 358 (sesta ristampa 2014).
- [5] Bolasco S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Carocci, Roma, pp. 410.
- [6] Bolasco S., Pavone P. (2012). *How the text mining measures complex phenomena in official statistics*, Proceedings of the XLVI Scientific Meeting of the Italian Statistical Society, Roma, 20-22 giugno 2012
- [7] Bolasco S. (2012). Introduction to the Automatic Analysis of Textual Data via a Case Study. in *Statistica Applicata - Italian Journal of Applied Statistics* (Special Issue: Advances in Textual Data Analysis and Text Mining), vol. 22(1), pp. 5-19.
- [8] Bolasco S. (2010). Taltac2.10 Sviluppo, esperienze ed elementi essenziali di analisi automatica dei testi, LED, Milano, pp. 55. (<http://www.ledonline.it/Taltac/allegati/459-7-TALTAC-COL.pdf>)
- [9] Bolasco S., Chiari I., Giuliano L. (eds.) (2010). *JADT2010. Statistical Analysis of Textual Data*, Proceedings of 10th International Conference JADT, LED, Milano 2 voll., pp. 1330.
- [10] Bolasco S. (2010). Il riconoscimento automatico di locuzioni verbali con l'ausilio del software Taltac2, *Rassegna Italiana di Linguistica Applicata*, XLII 1-2 : 39-56. (ISSN 0033-9725).
- [11] Bolasco S., Pavone P. (2010). Automatic Dictionary and Rule-Based Systems for Extracting Information from Text, in Francesco Palumbo, Carlo Natale Lauro, Michael Greenacre (eds.) "Data

- Analysis and Classification” Proceedings of the 6th Conference of the Classification and Data Analysis Group of the Società Italiana di Statistica, Springer, Berlin-Heidelberg, pp. 189-198.
- [12] Bolasco S., Pavone P. (2008). *Multi-class categorization based on cluster analysis and TFIDF*, in S. Heiden & B. Pincemin (eds.) } *JADT2008*, Presses Universitaires de Lyon, vol. 1, pp. 209-218.
  - [13] Bolasco S., Pavone P. (2007). *Automatic dictionary and rule-based systems for extracting information from text*, in Classification and Data Analysis 2007. Book of short papers CLADAG2007. EUM - Edizioni Università di Macerata, pp. 255-258.
  - [14] Bolasco S., D’Avino E., Pavone P. (2007) *Analisi dei diari giornalieri con strumenti di statistica testuale e text mining*, in Maria Clelia Romano (a cura di) “I tempi della vita quotidiana. Un approccio multidisciplinare all’analisi dell’uso del tempo.” Roma, ISTAT, pp. 309-340.
  - [15] La Noce M., Bolasco S., Allegra E., Ruocco V., Capo F.M. (2006). Merger control in Italy 1995-2003: a statistical study of the enforcement practice by mining the text of Authority resolutions, in *International Journal of the Economics of Business* Vol. 13, 2, pp. 307-334.
  - [16] Bolasco S. (2005). Statistica testuale e text mining: alcuni paradigmi applicativi, *Quaderni di Statistica*, Liguori Ed., 7, p. 17-53.
  - [17] Baiocchi F., Bolasco S., Canzonetti A., Capo F. M. (2005). Estrazione di informazione da testi per la classificazione automatica di una base documentale: la soluzione di Text Mining per l’Authority della Concorrenza, in S. Bolasco, et al. (eds.) *Text mining: uno strumento strategico per imprese e istituzioni*, CISU, Roma, 2005, p. 45-54.
  - [18] Bolasco S., Canzonetti A., Capo F. M., della Ratta-Rinaldi F., Singh B. K. (2005) "Understanding Text Mining: a Pragmatic Approach" in S. Sirmakessis (ed.) *Knowledge Mining*, Series: Studies in Fuzziness and Soft Computing, Springer Verlag, Heidelberg, p. 31-51.
  - [19] Bolasco S., Canzonetti A. (2005). Some insights into the evolution of 1990s' standard Italian using Text Mining techniques and automatic categorisation, in M. Vichi, P. Monari, S. Mignani e A. Montanari (eds.) "New developments in classification and data analysis" Serie *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag, Berlin, p. 293-302.
  - [20] Bolasco S. (2005). La reperibilità statistica di tendenze diacroniche nell’uso delle parole, in I. Chiari e T. De Mauro (eds.) *Parole e Numeri. Analisi quantitativa dei fatti di lingua*, Aracne, Roma, p. 335-354.
  - .....
  - Bolasco S. (2000). Déclarations et répliques gouvernementales dans le discours parlementaire italien, deux genres discursifs. *Mots*, 64, p. 97-112.
  - Bolasco, S. (1996) *Il lessico del discorso programmatico di governo* in Villone M. Zuliani A. (a cura di) *L’attività dei governi della repubblica italiana (1948-1994)*, Bologna: Il Mulino, p. 163-349.

Autorizzo la pubblicazione ai sensi del D.Lgs. n. 33/2013 “Riordino della disciplina riguardante gli obblighi di pubblicità, trasparenza e diffusione di informazioni da parte delle pubbliche amministrazioni” e acconsento all’utilizzo delle informazioni ivi contenute ai sensi del D.L. n. 196/2003 “Codice in materia di protezione dei dati personali”.

Roma, 31 luglio 2017