



Università
Bocconi
MILANO



POLITECNICO
MILANO 1863



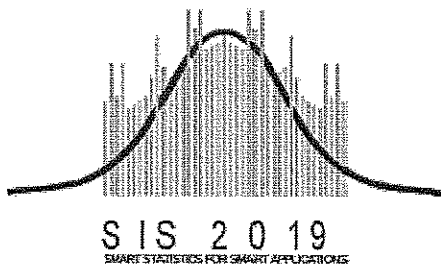
UNIVERSITÀ DEGLI STUDI DI MILANO
UNIVERSITÀ DEGLI STUDI DI MILANO



mathesia

Smart Statistics for Smart Applications

Book of Short Papers SIS2019



Società
Italiana di
Statistica



Editors: Giuseppe Arbia, Stefano Peluso,
Alessia Pini and Giulia Rivellini

Copyright © 2019

PUBLISHED BY PEARSON

WWW.PEARSON.COM

Giugno 2019 ISBN 9788891915108

Preface

Section 1. Plenary Sessions and Round Table

Preface	3
Shallow Learning for Data Science	7
<i>Antonio Canale</i>	
Smart Statistics: concept, technology and service	17
<i>David John Hand, Maurizio Vichi</i>	
Tavola rotonda "Smart ageing: lunga vita attiva, salute e nuove tecnologie"	19

Section 2. Invited Papers

Demography in the Digital Era: New Data Sources for Population Research	23
Demografia nell'era digitale: nuovi fonti di dati per gli studi di popolazione.....	23
<i>Diego Alburez-Gutierrez, Samin Aref, Sofia Gil-Clavel, André Grow, Daniela V. Negraia, Emilio Zagheni</i>	
Stationarity of a general class of observation driven models for discrete valued processes..	31
Stazionarietà di una classe generale di modelli observation-driven per processi a valori discreti	
<i>Mirko Armillotta, Alessandra Luati and Monia Lupparelli</i>	
An extension of the censored gaussian lasso estimator	39
Un'estensione dello stimatore cglasso	
<i>Luigi Augugliaro and Gianluca Sottile and Veronica Vinciotti</i>	
A formal approach to data swapping and disclosure limitation techniques.....	47
Un approccio formale per tecniche di trasformazione dei dati in problemi di privacy	
<i>F. Ayed, M. Battiston and F. Camerlenghi</i>	
A new ordinary kriging predictor for histogram data in L2-Wasserstein space	55
Un nuovo predittore kriging per istogrammi nello spazio L2-Wasserstein	
<i>Antonio Balzanella and Antonio Irpino and Rosanna Verde</i>	
Keywords dynamics in online social networks: a case-study from Twitter.....	63
La dinamica delle parole chiave nelle reti sociali online: un esempio tratto da Twitter	
<i>Carolina Becatti, Irene Crimaldi and Fabio Saracco</i>	
Statistical Matching of HBS and ADL to analyse living conditions, poverty and happiness.....	71
Statistical Matching of HBS e ADL per l'analisi di condizioni di vita, povertà e felicità	
<i>Cristina Bernini, Silvia Emili, Maria Rosaria Ferrante</i>	
Statistical sources for cybersecurity and measurement issues.....	79
Fonti statistiche per la sicurezza cibernetica e problemi di misurazione	
<i>Claudia Biancotti, Riccardo Cristadoro, Raffaele Tartaglia Polcini</i>	
Use of GPS-enabled devices data to analyse commuting flows between Tuscan municipalities.....	89
Un'analisi dei flussi di pendolarismo sistemati tra i comuni toscani tramite l'utilizzo di dati GPS	
<i>Chiara Bocci, Leonardo Piccini and Emilia Rocco</i>	
Statistical calibration of the digital twin of a connected health object	97
Inversione statistica dei parametri di ingresso per il gemello digitale di un oggetto sanitario collegato	
<i>Nicolas Bousquet and Walid Dabachine</i>	
Time Series Forecasting: Is there a role for neural networks?	103
Le Reti Neurali nella Previsione di Serie Storiche	
<i>Giuseppe Bruno, Sabina Marchetti, Juri Marcucci, Diana Nicoletti</i>	

Modelling weighted signed networks.....	111
Modellazione di reti segnate pesate	
<i>Alberto Caimo and Isabella Gollini</i>	
Issues on Bayesian nonparametric measures of disclosure risk	119
Questioni su misure Bayesiane nonparametriche di rischio di "disclosure"	
<i>Federico Camerlenghi, Cirzia Carota and Stefano Favaro</i>	
Hierarchies of nonparametric priors.....	125
Gerarchie di distribuzioni iniziali nonparametriche	
<i>Federico Camerlenghi, Stefano Favaro and Lorenzo Masoero</i>	
Issues with Nonparametric Disclosure Risk Assessment.....	133
Questioni sull'Analisi Nonparametrica del Rischio di "Disclosure"	
<i>Federico Camerlenghi, Stefano Favaro, Zacharie Naulet and Francesca Panero</i>	
Technologies and data science for a better health both at individual and population level. ..	141
Two practical research cases.	
Tecnologie e data science per una salute migliore sia a livello individuale che di popolazione.	
<i>Stefano Campostrini and Lucia Zanutto</i>	
Temporal sentiment analysis with distributed lag models	149
Analisi temporale del "sentiment" con modelli a lag distribuiti	
<i>Carrannante M., Mattered R., Misuraca M., Scepi G., Spano M.</i>	
A statistical investigation on the relationships among financial disclosure, sociodemographic variables, financial literacy and retail investors' risk assessment ability	157
Indagine empirica sulle relazioni tra prospetti per la diffusione di informazioni finanziarie, variabili sociodemografiche, educazione finanziaria e abilità di valutazione del rischio	
<i>Rosella Castellano, Marco Mancinelli and Pasquale Samacchiaro</i>	
Bayesian Model Comparison based on Wasserstein Distances.....	167
Confronto di Modelli Bayesiani tramite Distanze di Wasserstein	
<i>Marta Catalano, Antonio Lijoi and Igor Prünster</i>	
Hierarchical Clustering and Dimensionality Reduction for Big Data	173
Clustering e Riduzione Dimensionale Gerarchici per Dati di Grandi Dimensioni	
<i>Carlo Cavicchia, Maurizio Vichi and Giorgia Zaccaria</i>	
ICOs success drivers: a textual and statistical analysis.....	181
Fattori di successo nelle ICOs: un'analisi testuale e statistica	
<i>Paola Cerchiello and Anca Mirela Toma</i>	
Small area estimators with linked data.....	189
Stimatori per piccole aree nel caso di dati ottenuti attraverso il record linkage	
<i>Chambers Raymond and Fabrizi Enrico and Salvati Nicola</i>	
Optimal Portfolio Selection via network theory in banking and insurance sector	197
<i>Gian Paolo Clemente, Rosanna Grassi and Asmerilda Hitaj</i>	
Matching error(s) and quality of statistical matching in complex surveys.....	205
Errori di matching e qualità del matching statistico in indagini complesse	
<i>Pier Luigi Contii and Daniela Marella</i>	
Hotel search engine architecture based on online reviews' content.....	213
Un motore di ricerca per gli hotel basato sulle recensioni online	
<i>Claudio Conversano, Maurizio Romano and Francesco Mola</i>	
Economic Crisis and Earnings Management: a Statistical Analysis	219
Crisi Economica e Gestione degli Utili: un'Analisi Statistica	
<i>C. Cusatelli, A.M. D'Ugento, M. Giacalone, F. Grimaldi</i>	
A Comparison of Nonparametric Bivariate Survival Functions.....	227
Confronto tra stimatori non-parametrici della funzione di sopravvivenza bivariata	
<i>Hongsheng Dai and Marialisa Restaino</i>	
Predictive Algorithms in Criminal Justice.....	237
Algoritmi predittivi e giustizia penale	
<i>Francesco D'Alessandro</i>	

A proposal for an integrated approach between sentiment analysis and social network analysis.....	247
Una proposta per un approccio integrato tra analisi del sentimento e analisi delle reti sociali	
<i>Domenico De Stefano and Francesco Santelli</i>	
A meta-tissue non-parametric factor analysis model for gene co-expression	255
Meta-analisi fattoriale non parametrica per lo studio di espressioni genetiche in diversi tessuti	
<i>Roberta De Vito and Barbara Engelhardt</i>	
Bayesian estimate of population count with false captures: a latent class approach.....	261
Stima Bayesiana della popolazione con false catture: un approccio basato sulle classi latenti	
<i>Davide Di Cecco, Marco Di Zio and Brunero Liseo</i>	
Spherical regression with local rotations and implementation in R.....	269
Regressione sferica con rotazioni locali ed implementazione in R	
<i>Marco Di Marzio, Stefania Fensore, Agnese Panzera, Charles C. Taylor</i>	
A clustering method for network data to analyse association football playing styles	277
Un metodo di raggruppamento per dati di rete finalizzato all'analisi degli schemi di gioco nel calcio	
<i>Jacopo Diqigiovanni</i>	
Big data in longitudinal observational studies: how to deal with non-probability samples and technological changes.....	285
I Big data negli studi longitudinali: come trattare campioni non probabilistici e cambi di tecnologia	
<i>Clelia Di Serio, Luca Del Core, Eugenio Montini and Andrea Calabria</i>	
Smart Data For Smart Health.....	293
Smart Data Per Smart Health	
<i>Clelia Di Serio, Ernst C. Wit, Elena Bottinelli and Roberto Buccione</i>	
Detecting and classifying moments in basketball matches using sensor tracked data.....	297
Una procedura per identificare e classificare momenti di gioco in pallacanestro con l'uso di dati sensori.	
<i>Tullio Facchinetti and Rodolfo Metulini and Paola Zuccolotto</i>	
Ordered response models for cyber risk	305
Modelli a risposta ordinale per la valutazione del cyber risk	
<i>Silvia Facchinetti and Claudia Tarantola</i>	
Functional data analysis-based sensitivity analysis of integrated assessment Models for climate change modelling	313
Analisi di sensitività basata sull'analisi di dati funzionali per modelli di valutazione integrata dei cambiamenti climatici	
<i>Matteo Fontana, Massimo Tavoni and Simone Vantini</i>	
Coupled Gaussian Processes for Functional Data Analysis.....	319
Processi gaussiani per l'analisi dei dati funzionali	
<i>L. Fontanella, S. Fontanella, R. Ignaccolo, L. Ippoliti, P. Valentini</i>	
Two-fold data streams dimensionality reduction approach via FDA	323
Un approccio a due fasi per la riduzione di dimensionalità di data streams via FDA	
<i>F. Fortuna, T. Di Battista and S.A. Gattone</i>	
Statistical analysis of Sylt's coastal profiles using a spatiotemporal functional model	331
<i>Rik Gijssman, Philipp Otto, Torsten Schlurmann, Jan Visscher</i>	
Bootstrap prediction intervals for weighted TAR predictors	339
Intervalli di previsione bootstrap per previsori ponderati per modelli TAR	
<i>Francesco Giordano and Marcella Niglio</i>	
A rank graduation index to prioritise cyber risks.....	347
Un indice di graduazione per assegnare livelli di priorità ai rischi informatici	
<i>Paolo Giudici and Emanuela Raffinetti</i>	
Vector Error Correction models to measure connectedness of bitcoin exchange markets	355
Modelli di Vector Error Correction per misurare la connessione delle piattaforme di scambio di bitcoin	
<i>Paolo Giudici and Paolo Pagnottoni</i>	
Estimation of lineup efficiency effects in Basketball using play-by-play data.....	363
L'uso dei dati del play-by-play per la stima degli effetti di quintetto nella pallacanestro	
<i>Luca Grassetti, Ruggero Bellio, Giovanni Fonseca and Paolo Vidoni</i>	
Trajectory clustering using adaptive squared distances.....	371
Clustering di traiettorie attraverso distanze adattative quadratiche	
<i>Antonio Ippino</i>	

Bayesian Analysis of Privacy Attacks on GPS Trajectories	379
Analisi Bayesiana degli Attacchi alla Privacy su Traiettorie GPS	
<i>Sirio Legramanti</i>	
Data Analytics in the Insurance Industry: Market trends and lessons from a use case customer predictive modelling	387
Data Analytics nel settore assicurativo: principali trend e considerazioni da un caso d'uso applicato alla predizione del comportamento degli assicurati	
<i>Cristian Losito and Francesco Pantisano</i>	
BasketballAnalyzeR: the R package for basketball analytics	395
BasketballAnalyzeR: il pacchetto R per l'analisi dei dati nella pallacanestro	
<i>Marica Manisera, Marco Sandri and Paola Zuccolotto</i>	
Data Integration by Graphical Models	403
Utilizzo dei modelli grafici per l'integrazione dei dati	
<i>Daniela Marella and Paola Vicard and Vincenzina Vitale</i>	
A two-part finite mixture quantile regression model for semi-continuous longitudinal data	409
<i>Maruotti Antonello, Merlo Luca and Petrella Lea</i>	
Multivariate change-point analysis for climate time series	415
Analisi di change-point multivariate per serie storiche climatiche	
<i>Gianluca Mastrantonio, Giovanna Jona Lasinio, Alessio Pollice, Giulia Capotorti, Lorenzo Teodonio and Carlo Blasi</i>	
A divide-et-impera approach for the spatial prediction of object data over complex regions	423
Un approccio divide-et-impera per la previsione spaziale di dati oggetto su regioni complesse	
<i>Alessandra Menafoglio e Piercesare Secchi</i>	
A strategy for the matching of mobile phone signals with census data.....	427
Una strategia per l'abbinamento di segnali di telefonia mobile con dati censuari	
<i>Rodolfo Melulini and Maurizio Carpita</i>	
Risk-based analyses for non-proportional reinsurance pricing	435
Analisi Risk-based per il pricing nella riassicurazione di trattati non proporzionali	
<i>Fabio Moraldi and Nino Savelli</i>	
A Simplified Efficient and Direct Unequal Probability Resampling	441
Un semplice Ricampionamento, efficiente e diretto per campioni a probabilità variabili	
<i>Federica Nicolussi, Fulvia Mecatti and Pier Luigi Conti</i>	
Labour Law: Machine vs. Employer Powers Diritto del lavoro: Macchina vs. Poteri datoriali	449
<i>Antonella Occhino – Michele Faloli</i>	
Domain knowledge based priors for clustering.....	455
Distribuzioni a priori per l'analisi di raggruppamento basate sulla conoscenza di settore	
<i>Sally Paganin</i>	
Clustering of Behavioral Spatial Trajectories in Neuropsychological Assessment	463
Analisi dei gruppi di traiettorie spaziali nella valutazione neuropsicologica	
<i>Francesco Palumbo, Antonio Cerrato, Michela Ponticorvo, Onofrio Gigliotta, Paolo Bartolomeo, Orazio Miglino</i>	
What is wrong in the debate about smart contracts.....	471
Smart contract e diritto: riflessioni critiche su un dualismo fuorviante	
<i>Roberto Pardolesi and Antonio Davola</i>	
Financial Transaction Data for the Nowcasting in Official Statistics.....	485
Transazioni elettroniche di pagamento per le previsioni a breve nella Statistica ufficiale	
<i>Righi A., Ardizzi G., Gambini A., Iannaccone R., Moauro F., Renzi N. and Zurlo D.</i>	
On the examination of a criticality measure for a complex system in a forecasting perspective	493
Esame di una misura di criticità per un sistema complesso in una prospettiva previsiva	
<i>Renata Rotondi and Elisa Varini</i>	
Knowledge discovery for dynamic textual data: temporal patterns of topics and word clusters in corpora of scientific literature	501
Estrazione della conoscenza da dati testuali dinamici: evoluzione temporale di argomenti e gruppi di parole in corpora di letteratura scientifica	
<i>Stefano Sbalchiero, Matilde Trevisani and Arjuna Tuzzi</i>	

Classifying the Willingness to Act in Social Media Data: Supervised Machine Learning for U.N. 2030 Agenda	509
Classificare la volontà di agire nei dati dei Social Media: Supervised Machine Learning per l'Agenda 2030 delle Nazioni Unite	
<i>Andrea Sciandra, Alessio Surian and Livio Finos</i>	
Classification of spatio-temporal point pattern in the presence of clutter using K-th nearest neighbour distances.....	517
Classificazione dei processi puntuali spazio-temporali basata sulla distanza dal K-mo vicino più vicino	
<i>Siino Marianna, Francisco J. Rodriguez-Cortés, Jorge Mateu, Giada Adelfio</i>	
Modelling properties of high-dimensional molecular systems	525
La modellazione di sistemi molecolari ad alta dimensionalità	
<i>Debora Slanzi, Valentina Mamei and Irene Poli</i>	
Non-crossing parametric quantile functions: an application to extreme temperatures	533
Il problema del crossing con funzioni quantiliche parametriche: un'applicazione alle temperature estreme	
<i>Gianluca Sottile and Paolo Frumento</i>	
A new tuning parameter selector in lasso regression.....	541
Un nuovo criterio di selezione per il parametro di penalizzazione nella regressione lasso	
<i>Gianluca Sottile and Vito MR Muggeo</i>	
Similarity patterns, topological information and credit scoring models	549
Strutture di similarità, informazioni topologiche e modelli di credit scoring	
<i>Alessandro Spelta, Branka Hadji-Misheva and Paolo Giudici</i>	
Between hawks and doves: measuring central bank communication	557
Fra falchi e colombe: valutazione delle comunicazioni di Banca Centrale	
<i>Elien Tobback, Stefano Nardelli, David Martens</i>	
New methods and data sources for the population census	561
Nuovi metodi e fonti per il censimento della popolazione	
<i>Paolo Valente</i>	
FinTech and the Search for "Smart" Regulation	569
Fintech e la ricerca di una regolamentazione "smart"	
<i>Silvia Vanon</i>	
An anisotropic model for global climate data	577
Un modello anisotropico per i dati climatici globali	
<i>Nil Venet and Alessandro Fassò</i>	
Analysis of the financial performance in Italian football championship clubs via GEE and diagnostic measures.....	585
Analisi delle performance finanziarie delle squadre di calcio di serie A via GEE e misure di diagnostica	
<i>Maria Kelly Venezuela, Anna Crisci, Luigi D'Ambra, D'Ambra Antonello</i>	
A statistical space-time functional model for air quality analysis and mapping.....	593
Un modello statistico spazio-tempo funzionale per l'analisi e la mappatura della qualità dell'aria	
<i>Yaqiong Wang, Alessandro Fassò and Francesco Finazzi</i>	
Tempering and computational efficiency of Bayesian variable selection.....	599
Tempering e l'efficienza computazionale della selezione bayesiana delle variabili	
<i>Giacomo Zanella and Gareth O. Roberts</i>	
Dimensions and links for Hate Speech in the social media	607
Dimensioni e legami per i discorsi di odio nei social media	
<i>Emma Zavaronne, Guido Ferilli</i>	

Section 3. Contributed Papers

Density-based Algorithm and Network Analysis for GPS Data.....	617
Algoritmi di Cluster e Reti per lo studio di dati GPS	
<i>Antonino Abbruzzo, Mauro Ferante, Stefano De Cantis</i>	
Local inference on functional data based on the control of the family-wise error rate	623
Inferenza locale per dati funzionali basata sul controllo del family-wise error rate	
<i>Konrad Abramowicz, Alessia Pini, Lina Schellin, Sara Sjöstedt de Luna, Aymeric Stamm, and Simone Vantini</i>	

Application and validation of dynamic Poisson models to measure credit contagion	629
Applicazione e validazione di modelli di Poisson dinamici per misurare il contagio nel credito	
<i>Arianna Agosto and Emanuela Raffinetti</i>	
Monitoring SDGs at territorial level: the case of Lombardy.....	637
Il monitoraggio degli SDGs a livello territoriale: il caso della Lombardia	
<i>Leonardo Alaimo, Livia Celardo, Filomena Maggino, Adolfo Morrone, Federico Olivieri</i>	
The Experts Method for the prediction of periodic multivariate time series of high dimension.....	643
Il Metodo degli Esperti per la previsione di serie temporali multivariate e periodiche, di dimensione elevata	
<i>Giacomo Aletti, Marco Bellan and Alessandra Micheletti</i>	
Regression with time-dependent PDE regularization for the analysis of spatio-temporal data	649
Regressione con regolarizzazione di PDE tempo dipendenti per modellizzare dati spatio-temporali	
<i>Eleonora Arnone, Laura Azzimonti, Fabio Nobile, Laura M. Sangalli</i>	
A network analysis of museum preferences: the Firenzecard experience.....	653
Un'analisi di rete delle preferenze museali: l'esperienza della Firenzecard	
<i>Silvia Bacci, Bruno Bertaccini, Roberto Dinelli, Antonio Giusti, and Alessandra Petrucci</i>	
A statistical learning approach to group response categories in questionnaires.....	659
Un approccio basato sull'apprendimento statistico per raggruppare le categorie di risposta nei questionari	
<i>Michela Battauz</i>	
Tree-based Functional Data Analysis for Classification and Regression.....	665
Alberi di Classificazione e Regressione per dati Funzionali	
<i>Edoardo Belli, Enrico Ragaini, Simone Vantini</i>	
PDE-regularized regression for anisotropic	669
spatial fields Regressione con regolarizzazione differenziale per campi spaziali anisotropi	
<i>Mara S. Bernardi, Michelle Carey, James O. Ramsay and Laura M. Sangalli</i>	
A Bayesian model for network flow data: an application to BikeMi trips	673
<i>Giulia Bissoli, Celeste Principi, Gian Matteo Rinaldi, Mario Beraha and Alessandra Guglielmi</i>	
Statistical classics in the big data era. When (astro-physical) models are nonregular.....	679
Statistica classica nell'era dei big data. Verosimiglianza e modelli non regolari	
<i>Alessandra R. Brazzale and Valentina Mameli</i>	
Bayesian Variable Selection for High Dimensional Logistic Regression	685
Selezione bayesiana delle variabili nel modello di regressione logistica ad alta dimensionalità	
<i>Claudio Busatto, Andrea Sottosanti and Mauro Bernardi</i>	
Bayesian modeling for large spatio-temporal data: an application to mobile networks.....	691
Modelli bayesiani per grandi dataset spatio-temporali: un'applicazione a dati di telefonia mobile	
<i>Annalisa Cadonna, Andrea Cremaschi, Alessandra Guglielmi</i>	
A Mathematical Framework for Population of Networks: Comparing Public Transport of Different Cities.	697
Un approccio matematico all'analisi di una popolazione di networks: come confrontare il sistema di trasporto pubblico di diverse città.	
<i>Anna Calissano, Aasa Feragen, Simone Vantini</i>	
How Important Discrimination is for the Job Satisfaction of Immigrants in Italy: A Counterfactual Approach	703
Quanto influisce la discriminazione sulla soddisfazione lavorativa degli immigrati in Italia: un approccio controfattuale	
<i>Maria Gabriella Campolo, Antonino Di Pino and Michele Limosani</i>	
Unfolding the SEcrets of LongEvity: Current Trends and future prospects (SELECT)	709
A path through morbidity, disability and mortality in Italy and Europe	
<i>Stefano Campostrini, Daniele Durante, Fabrizio Faggiano and Stefano Mazzucco</i>	
Galaxy color distribution estimation via dependent nonparametric mixtures	713
Stima della distribuzione del colore delle galassie via misture nonparametriche dipendenti	
<i>Antonio Canale, Riccardo Corradin and Bernardo Nipoti</i>	
A case for order optimal matching: a salary gap study.....	719
Un algoritmo di matching ottimale ordinato per un studio sulle differenze salariali	
<i>Massimo Cannas</i>	

A Prediction Method for Ordinal Consistent Partial Least Squares	725
Un Metodo di Previsione per l'Algoritmo Ordinal Consistent Partial Least Squares	
<i>Gabriele Cantaluppi and Florian Schubert</i>	
Functional control charts for monitoring ship operating conditions and CO2 emissions based on scalar-on-function linear model	731
Carte di controllo funzionali per il monitoraggio delle condizioni operative e delle emissioni di CO2 di navi da carico e passeggeri mediante modello di regressione funzionale con risposta scalare	
<i>Christian Capezza, Antonio Lepore, Alessandra Menafoglio, Biagio Palumbo, and Simone Vantini</i>	
Predicting and improving smart mobility: a robust model-based approach to the BikeMi BSS	737
Prevedere e migliorare la mobilità smart: un approccio robusto di classificazione applicato a BikeMi	
<i>Andrea Cappozzo, Francesca Greselin and Giancarlo Manzi</i>	
Public support for an EU-wide social benefit scheme: evidence from Round 8 of the European Social Survey (ESS)	743
Sostegno pubblico a un sistema di prestazioni sociali a livello dell'Unione Europea: i risultati del Round 8 della European Social Survey (ESS)	
<i>Paolo Emilio Cardone</i>	
Revenue management strategies and Booking.com ghost rates: a statistical analysis	751
Strategie di revenue management e Booking.com ghost rates: un'analisi statistica	
<i>Cinzia Carota, Consuelo R. Nava, Marco Alderighi</i>	
Analysing international migration flows: a Bayesian network approach	757
Analisi dei flussi migratori internazionali attraverso l'impiego di modelli grafici	
<i>Federico Castelletti and Emanuela Furfaro</i>	
A sparse estimator for the function-on-function linear regression model	763
Uno stimatore sparso per il modello di regressione lineare con regressore e risposta funzionali	
<i>Fabio Centofanti, Matteo Fontana, Antonio Lepore, and Simone Vantini</i>	
Robustness and fuzzy multidimensional poverty indicators: a simulation study.....	769
Robustezza ed indicatori fuzzy multidimensionali della povertà: uno studio di simulazione	
<i>Michele Costa</i>	
Text Based Pricing Modelling: an Application to the Fashion Industry	775
Modellazione dei prezzi basata su dati testuali: un'applicazione all'industria fashion	
<i>Federico Crescenzi, Marzia Freo and Alessandra Luati</i>	
Model based clustering in group life insurance via Bayesian nonparametric mixtures	781
Raggruppamento basato sul modello nel settore assicurativo: un approccio bayesiano nonparametrico	
<i>Laura D'Angelo</i>	
Smart Tools for Academic Submission Decisions: Waiting Times Modeling	787
Strumenti "Smart" per sottoporre i manoscritti accademici: modelli per i tempi di attesa	
<i>Francesca De Battisti - Giancarlo Manzi</i>	
On the Use of Control Variables in PLS-SEM	793
Sull'Uso delle Variabili di Controllo nei PLS-SEM	
<i>Francesca De Battisti and Elena Siletti</i>	
Partial dependence with copula and financial applications	799
Dipendenza parziale con funzioni copula e applicazioni finanziarie	
<i>Giovanni De Luca, Marta Nai Ruscone and Giorgia Riveccio</i>	
Exploring the relationship between fertility and well-being: What is smart?.....	805
Esplorando la relazione tra fecondità e benessere: cosa c'è di smart?	
<i>Alessandra De Rose, Filomena Racioppi, Maria Rita Sebastiani</i>	
Web-Based Data Collection and Quality Issues in Co-Authorship Network Analysis	811
Qualità dei dati bibliografici raccolti via web per l'analisi di reti di collaborazione scientifica	
<i>Domenico De Stefano, Vittorio Fuccella, Susanna Zaccarin</i>	
A new regression model for bounded multivariate responses.....	817
Un nuovo modello di regressione per risposte multivariate limitate	
<i>Agnese Maria Di Brisco, Roberto Ascari, Sonia Migliorati and Andrea Ongaro</i>	
Turning big data into smart data: two examples based on the analysis of the Mappa dei Rischi dei Comuni Italiani.....	823
Trasformare i big data in smart data: due esempi di analisi della Mappa dei Rischi dei Comuni Italiani	
<i>Oleksandr Didkovskiy, Alessandra Menafoglio, Piercesare Secchi, Giovanni Azzone</i>	

Hidden Markov Model estimation via Particle Gibbs	829
Stima di Hidden Markov Model tramite Particle Gibbs	
<i>Pierfrancesco Alaimo Di Loro, Enrico Ciminello and Luca Tardella</i>	
A note on marginal effects in logistic regression with independent covariates	837
Una nota sugli effetti marginali nella regressione logistica con covariate indipendenti	
<i>Marco Doretti</i>	
DNA mixtures: a case study involving a Romani reference population	843
Misure di DNA: un caso di studio riguardante una popolazione di riferimento dei Rom	
<i>Francesco Dotto, Julia Mortera and Vincenzo Pascali</i>	
Pivotal seeding for K-means based on clustering ensembles	849
Inizializzazione pivotale dell'algoritmo delle K-medie tramite raggruppamento con metodi di insieme	
<i>Leonardo Egidi, Roberta Pappadà, Francesco Pauli, Nicola Torelli</i>	
Optimal scoring of partially ordered data, with an application to the ranking of smart cities	855
Scoring ottimale di dati parzialmente ordinati, con un'applicazione al ranking delle smart city	
<i>Marco Fattore, Alberto Arcagni, Filomena Maggino</i>	
Bounded Domain Density Estimation	861
Stima della densità non-parametrica su domini bidimensionali limitati	
<i>Federico Ferraccioli, Laura M. Sangalli and Livio Finos</i>	
Polarization and long-run mobility: yearly wages comparison in three southern European countries	867
Polarizzazione e mobilità sul lungo periodo: un confronto fra salari annuali in tre Paesi sud-Europei	
<i>Ferretti C., Crosato L., Cipollini F., Ganugi P.</i>	
Design of Experiments, aberration and Market Basket Analysis	873
Pianificazione degli esperimenti, aberrazione e Market Basket Analysis	
<i>Roberto Fontana and Fabio Rapali</i>	
Generalized Procrustes Analysis for Multilingual Studies	879
Analisi Procrustiana Generalizzata per studi Multilingue	
<i>Alessia Forciniti, Michelangelo Misuraca, Germana Scepti, Maria Spano</i>	
Prior specification in flexible models	885
Specificazione delle prior in modelli flessibili	
<i>Maria Franco-Villoria, Massimo Ventrucchi and Haavard Rue</i>	
Modeling Cyclists' Itinerary Choices: Evidence from a Docking Station-Based Bike-Sharing System	889
Un modello per gli itinerari dei ciclisti: risultati da un bike-sharing a stazioni fisse	
<i>S. T. Gaito - G. Manzi - G. Saibene - S. Salini - M. Zignani</i>	
A PARAFAC-ALS variant for fitting large data sets	895
Una variante del PARAFAC-ALS per approssimare data set di grandi dimensioni	
<i>Michele Gallo, Violetta Simonacci and Massimo Guarino</i>	
A Convex Mixture Model for Binomial Regression	901
Un modello mistura convessa per la Regressione Binomiale	
<i>Luisa Gallarossa and Antonio Canale</i>	
Blockchain as a universal tool for business improvement	907
Blockchain come strumento universale per il miglioramento del business	
<i>Massimiliano Giacalone, Diego Carmine Sinitò, Emilio Massa, Federica Oddo, Enrico Medda, Vito Santarcangelo</i>	
Seasonality in tourist flows: a decomposition of the change in seasonal concentration	913
La stagionalità nei flussi turistici: una scomposizione della variazione nella concentrazione stagionale	
<i>Luigi Grossi and Mauro Mussini</i>	
Are Real World Data the smart way of doing Health Analytics?	919
Real World Data: la base di una nuova ricerca clinica?	
<i>Francesca Ieva</i>	
Internet use and leisure activities: are all young people equal?	925
Internet e tempo libero: i giovani sono uguali tra loro?	
<i>Giuseppe Lamberti, Jordi Lopez Sintas and Pilar Lopez Belbeze</i>	
On a Family of Transformed Stochastic Orders	931
Su una famiglia di ordinamenti stocastici trasformati	
<i>Tommaso Lando and Lucio Bertoli-Barsotti</i>	

Bayesian stochastic search for Ising chain graph models.....	935
<i>Ricerca stocastica Bayesiana per modelli grafici a catena Ising</i>	
<i>Andrea Lazerini · Monia Lupparelli · Francesco C. Stingo</i>	
On the statistical design of parameters for variables sampling plans based on process capability index Cpk	941
<i>Progettazione statistica dei parametri per il piano di campionamento per variabili basato sull'indice di capacità di processo Cpk</i>	
<i>Antonio Lepore, Biagio Palumbo and Philippe Castagliola</i>	
Nowcasting foreign tourist arrivals using Google Trends: an application to the city of Florence, Italy.....	947
<i>Nowcasting degli arrivi turistici stranieri usando Google Trends: un'applicazione nella città di Firenze, Italia</i>	
<i>Alessandro Magrini</i>	
Inclusive growth in European countries: a cointegration analysis	953
<i>La crescita inclusiva nei paesi europei: un'analisi di cointegrazione</i>	
<i>Paolo Mariani, Andrea Marletta, Alessandra Michelangeli</i>	
ESCO- the European Labour Language: a conceptual and operational asset in support of labour governance in complex environments	959
<i>ESCO il linguaggio europeo del lavoro: uno strumento concettuale ed operativo per le politiche del lavoro in contesti complessi</i>	
<i>Cristilla Martelli, Laura Grassini, Adham Kahlawi, Maria Flora Salvatori, Lucia Buzzigoli</i>	
Hidden Markov Models for High Dimensional Data	965
<i>Hidden Markov Models per dati ad alta dimensionalità</i>	
<i>Martino, A., Guatteri, G., Paganoni, A.M.</i>	
Classification of Italian classes via bivariate semi parametric multilevel models	971
<i>Classificazione delle classi italiane per mezzo di modelli bivariati a effetti misti semi parametrici</i>	
<i>Chiara Masci, Francesca Ieva, Tommaso Agasisti and Anna Maria Paganoni</i>	
Data Mining Application to Healthcare Fraud Detection: Two-Step Unsupervised Clustering Method for Outlier Detection with Administrative Databases.....	977
<i>Data Mining Applicato al Riconoscimento Frodi in Sanità: Algoritmo a Due Step per l'Identificazione di Outliers con Database Amministrativi</i>	
<i>Massi Michela C., Ieva Francesca, Lettieri Emanuele</i>	
Multivariate analysis and biodiversity partitioning of a demersal fish community: an application to Lazio coast	985
<i>Analisi multivariata e partizione della biodiversità di una comunità di specie demersali: un'applicazione alla costa laziale</i>	
<i>M. Mingione, G. Jona Lasinio, S. Martino, F. Colloca</i>	
Latent Markov models with discrete separate cluster random effects on initial and transition probabilities.....	991
<i>Modelli Latent Markov ad effetti casuali discreti e separati per le probabilità iniziali e di transizione</i>	
<i>Giorgio E. Montanari and Marco Doretti</i>	
Unsuitability of likelihood-based asymptotic confidence intervals for Response-Adaptive designs in normal homoscedastic trials	997
<i>Inadeguatezza degli intervalli di confidenza asintotici basati sulla verosimiglianza per disegni Response-Adaptive in caso di risposte normali omoschedastiche</i>	
<i>Marco Novelli and Maroussa Zagoraiou</i>	
Local Hypothesis Testing for Functional Data: Extending False Discovery Rate to the Functional Framework.....	1003
<i>Verifica locale delle ipotesi nell'ambito dei dati funzionali: estensione della nozione di False Discovery Rate al contesto funzionale</i>	
<i>Niels Asken Lundtorp Olsen, Alessia Pini, and Simone Vantini</i>	
Educational mismatch and attitudes towards migration in Europe.....	1009
<i>Disallineamento fra formazione e lavoro e atteggiamenti verso le migrazioni in Europa</i>	
<i>Marco Guido Palladino and Emiliano Sironi</i>	
Soft thresholding Bayesian variable selection for compositional data analysis.....	1015
<i>Selezione di Variabili Bayesiana con funzioni di soglia per l'analisi di dati di composizione</i>	
<i>Matteo Pedone, Francesco C. Stingo</i>	
Sentiment-driven investment strategies: a practical example of AI-powered engines in a corporate setting	1021
<i>Strategie d'investimento guidate dal sentiment: un esempio pratico di Intelligenza Artificiale in contesto aziendale</i>	
<i>Mattia Pedrini, Sebastian Donoso, Enrico Deusebio, Nicola Donelli, Gabriele Arici, Andrea Cosentini, Paola Mosconi, Diego Ostinelli and Claudio Cocchis</i>	

Betting on football: a model to predict match outcomes	1027
Scommettere sul calcio: un nuovo modello per prevedere l'esito delle partite	
<i>Marco Petretta, Lorenzo Schiavon and Jacopo Diqiugiovanni</i>	
Estimation of dynamic quantile models via the MM algorithm	1033
Stima di modelli Quantilici Dinamici con algoritmo MM	
<i>Fabrizio Poggioni, Mauro Bernardi, Lea Petrella</i>	
The decomposition by subpopulations of the Pietra index: an application to the professional football teams in Italy	1039
La scomposizione per sottopopolazioni dell'indice di Pietra: un'applicazione alle squadre professionistiche di calcio in Italia	
<i>Francesco Porro and Mariangela Zenga</i>	
An Object Oriented Data Analysis of Tweets: the Case of Queen Elizabeth Olympic Park.	1045
Object Oriented Data Analysis di Tweet: il caso del Queen Elizabeth Olympic Park	
<i>Paola Riva, Paola Sturla, Anna Calissano and Simone Vantini</i>	
Bias reduced estimation of a fixed effects model for Expected Goals in association football	1051
Stima non distorta di un modello Expected Goal con effetti fissi nel calcio	
<i>Lorenzo Schiavon and Nicola Sartori</i>	
Looking for Efficient Methods to Collect and Geolocalise Tweets	1057
Alla ricerca di metodi efficienti per raccogliere e geolocalizzare tweet	
<i>Stephan Schlosser, Daniele Toninelli and Silvia Fabris</i>	
Principal ranking profiles.....	1063
Principal ranking profiles	
<i>Mariangela Sciandra, Antonella Plaia</i>	
A statistical model for voting probabilities	1069
Un modello statistico per le probabilità di voto	
<i>Rosaria Simone, Stefania Capecechi</i>	
How Citizen Science and smartphones can help to produce timely and reliable information? Evidence from the "Food Price Crowdsourcing in Africa" (FPCA) project in Nigeria.....	1075
Citizen Science e smartphone posso aiutare nella raccolta di dati tempestivi e affidabili? Testimonianze del progetto "Food Price Crowdsourcing in Africa" (FPCA) condotto in Nigeria	
<i>Gloria Solano-Hermosilla, Fabio Micale, Vincenzo Nardelli, Julius Adewopo, Celso Gorrín González</i>	
Dealing with uncertainty in automated test assembly problems.....	1083
La gestione dell'incertezza nei problemi di assemblaggio automatizzato dei test	
<i>Giada Spaccapanico Proietti, Mariagiulia Matteucci and Stefania Mignani</i>	
Joint Models: a smart way to include functional data in healthcare analytics	1089
Modelli congiunti: un metodo per includere i dati funzionali nelle analisi in ambito sanitario	
<i>Marta Spreafico, Francesca Ieva</i>	
Bayesian multiscale mixture of Gaussian kernels for density estimation	1095
Stima di densità tramite misture bayesiane multiscala di kernel gaussiani	
<i>Marco Stefanucci and Antonio Canale</i>	
Dynamic Bayesian clustering of running activities.....	1101
Clustering Bayesiano dinamico di attività di corsa	
<i>Mattia Sival and Mauro Bernardi</i>	
Employment and fertility in couples: whose employment uncertainty matter most?	1107
Lavoro e fecondità in coppia: il ruolo dell'incertezza lavorativa secondo una prospettiva di genere	
<i>Valentina Tocchioni, Daniele Vignoli, Alessandra Mattei, Bruno Arpino</i>	
A Functional Data Analysis Approach to Study a Bike Sharing Mobility Network in the City of Milan	1113
<i>Agostino Torti, Alessia Pini and Simone Vantini</i>	
Multiresolution Topological Data Analysis for Robust Activity Tracking	1119
<i>Giovanni Trappolini, Tullia Padellini, and Pierpaolo Brutti</i>	
Semilinear regression trees.....	1125
Alberi di regressione semilineari	
<i>Giulia Vannucci and Anna Gottard</i>	

A models selection criterion for evaluation of heat wave hazard: a case study of the city of Prato.....1131
Un criterio di selezione dei modelli per la valutazione della pericolosità delle ondate di calore: un caso studio della città di Prato
Veronica Villani, Giuliana Barbato, Elvira Romano and Paola Mercogliano

Digital Inequalities and ICT Devices: The ambiguous Role of Smartphones.....1139
Laura Zannella, Marina Zannella

Section 4. Posters

Modelling Hedonic Price using semiparametric M-quantile regression1147
Regressione m-quantilica semiparametrica per la modellizzazione dei prezzi edonici
Riccardo Borgoni, Antonella Carcagni, Alessandra Michelangeli, Nicola Salvati

Bayesian mixed latent factor model for multi-response marine litter data with multi-source auxiliary information1153
Modello bayesiano misto a fattori latenti per l'abbondanza di rifiuti marini con informazioni ausiliarie di diversa provenienza
Crescenza Calculli, Alessio Pollice, Marco V. Guglielmi and Porzia Maiorano

Official statistics to support the projects of A Scuola di OpenCoesione1159
L'esperienza di monitoraggio civico in Lombardia nell'anno scolastico 2018-19
del Vicario G. and Di Gennaro L. and Ferrazza D. and Spinella V. and Viviano L.

Spatial Logistic Regression for Events Lying on a Network: Car Crashes in Milan1165
Regressione logistica per eventi su network: gli incidenti automobilistici nel comune di Milano
Andrea Gilardi, Riccardo Borgoni and Diego Zappa

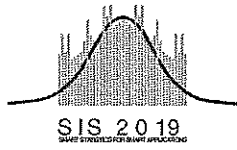
Variable selection and classification by the GRID procedure1171
Selezione e classificazione delle variabili attraverso il metodo GRID
Francesco Giordano, Soumendra Nath Lahiri and Maria Lucia Parrella

Joint VaR and ES forecasting in a multiple quantile regression framework.....1177
Stima congiunta del VaR e dell'ES attraverso la regressione quantilica multipla
Merlo Luca, Petrella Lea and Raponi Valentina

Approximate Bayesian Computation methods to model Multistage Carcinogenesis1183
Metodi di Approximate Bayesian Computation per modellare la Cancerogenesi Multistadiale
Consuelo R. Nava, Cinzia Carota, Jordy Bolton, Corrado Magnani, Francesco Barone-Adesi

Co-clustering TripAdvisor data for personalized recommendations1189
Co-clustering di dati TripAdvisor per un sistema di raccomandazioni personalizzato
Giulia Pascali, Alessandro Casa and Giovanna Menardi

Latent class analysis of endoreduplicated nuclei in confocal microscopy.....1195
Analisi di classi latenti per dati di nuclei endoreduplicati tramite microscopia confocale
Ivan Sciascia ivan.sciascia@unito.it, Gennaro Carotenuto gennaro.carotenuto@unito.it, Andrea Genre andrea.genre@unito.it, Università di Torino Dipartimento di Scienze della vita e biologia dei sistemi, viale Mattioli 25, 10125 Torino



Preface

Preface

This book includes the scientific contributions presented at the Intermediate Meeting of Italian Statistical Society (SIS) held in Milan at the Università Cattolica del Sacro Cuore, from June 18th to 21th of 2019. Following a long tradition (and a statutory indication of the Society), the intermediate meetings are held bi-annually on specific themes. This year, aiming at bridging the gap between statistics and the world of Big Data and Data Science, the conference was entirely devoted to the theme of “Smart Statistics for Smart Applications”. In this way the Italian Statistical Society had the explicit intention to answer the high and rapidly increasing demand on the subject, by providing academics, researchers and practitioners with a forum where new ideas and new methods could meet with new needs, new research questions and new applications.

The Conference could not have been organized without the joint effort of the Milanese network of Università Cattolica del Sacro Cuore, Università degli Studi di Milano Bicocca, Università Bocconi, Università “Vita e salute” San Raffaele, Politecnico di Milano and Università Statale di Milano. Members of all these universities took part actively to the Local Organizing Committee. The Conference has also greatly benefited from the contribution of the strategic partner Mathesia, which contributed to the various aspects of the organization, with special focus to the active involvement of private firms and companies and of the non-academic components.

The conference has registered more than 200 scientific contributions, including papers presented in plenary invited sessions, papers collected in specialized and solicited sessions on specific themes, about 100 contributions spontaneously submitted to the Program Committee and a poster session. All contributions were focused on the conference theme and provided a good overview of the state-of-the-art of the subject, from methodological and theoretical contributions, to applied works and case studies. The two plenary lectures were devoted to the (provocative) idea of “shallow learning”, as opposed to the more in-vogue idea of deep learning”, and to the problems linked with Big Data veridicity and reliability. A plenary round table draw the participants attention on the concept of smart ageing.

A distinctive feature of this conference, relative to previous analogous experiences, was the presence of many round tables and activities focused on topics of interest for a wider audience, freely open to external participation. These activities were termed *Fuoricongvegn*” and included a special session on “Data skills: Statistics and education for future jobs” organized jointly with Pearson Italia Publisher, a round table on “How to Close the Gap Between the Practice and Theory in Digital Transformation Era” organized joint with Mathesia, a colloquium on “Big Data and Big Responsibility”, a round table on “Political polls in the Big Data era”, a round table on “Big Data and Public Administration”, a round table on the changing role of the statistical scientific societies in a new interconnected world and the fifth edition of the statistical competition “Stats Under the Stars (SUS5)” organized by the Bocconi University, a whole-night hackathon on real-world business analytic problems for young Data Scientists.

More information about the *fuoricongvegn* activities may be found on the website of the meeting¹. We offer this book to all members of the Italian Statistical Society, to all participants of the conference and to all interested people, in the hope that this will provide them with a good snapshot of the on-going research in this exciting new area of statistical studies. We deeply thank all contributors for having submitted their work to the conference and all the researchers who did an outstanding job in acting as referees accurately and timely. Finally we wish to express our gratitude to the publisher Pearson Italia for all the support received.

Giuseppe Arbia
Stefano Peluso
Alessia Pini
Giulia Rivellini

¹ URL: <http://meetings3.sis-statistica.org/index.php/SIS2019/sis2019/schedConf/overview>

Blockchain as a universal tool for business improvement

Massimiliano Giacalone, Diego Carmine Sinitò, Emilio Massa, Federica Oddo, Enrico Medda, Vito Santarcangelo

Abstract: The aim of this work is to present the characteristics of the blockchain technology and its potential in corporate case study applications. The paper presents in detail an example of the implementation of permissioned blockchain and other examples of blockchain (also of semantic type) applied to the temporal certification of business processes of some brilliant southern Italy realities.

Key words: blockchain, bitcoin, safety for business

1 What is the Blockchain and how it works

Thanks to the fame of the bitcoin cryptocurrency, lately also blockchain technology has become object of great attention. Blockchains, thanks to their properties, are becoming an increasingly popular tool for companies that need to certify and keep their data safe.

1.1 *Blockchain and Bitcoin*

If we would ask someone what a blockchain is, they probably would link the answer to the word bitcoin. But how are bitcoin and blockchain related? This two entities were born simultaneously, however the blockchain is independent from bitcoin and the demonstration of this is in the evolution and the birth of new types of blockchain. Satoshi Nakamoto, pseudonym of the creator of cryptocurrency bitcoin, in his paper "Bitcoin: A Peer-to-Peer Electronic Cash System" introduces a system for the exchange of a virtual currency that does not require a financial institution that certifies the transactions. In this new scenario the problem of double spending was solved by using a distributed register where all the transactions are kept, which bases its security on a system of signing transactions with public key - private key, hashes and a proof encryption. The register obtained would be immutable unless after an

alteration the attacker had sufficient computing power to perform the cryptographic test again for all the blocks that make up the register before the other participants in the network.

1.2 Blockchain: structure and types

When a user makes a transaction on the bitcoin network, this is announced publicly to all the nodes that are part of it, in this way it is possible to certify that the transaction took place in a specific time (timestamp). Each node collects all new transactions by forming a block. When a block is complete, it can take part of the chain after passing a cryptographic test. The proof-of-work at the base of bitcoin is the increment of a numerical value, called *nonce*, which causes the block hash to start with a certain number of zeros. Once the cryptographic problem is solved, the node communicates to all the other nodes the solution found. Nodes accept blocking only if all transactions within are valid and they express their acceptance by creating the next block which includes the hash of the block that has just been accepted. Thanks to its distributed structure and the large number of nodes, the blockchain guarantees the immutability of data and business continuity.

This kind of network is also called public, this means that anyone with a device connected to the internet can become a network node and can take part at the creation of blocks and the validation of new blocks or have access to the entire transaction register. Various types of blockchain which provide for a form of centralized authority have been recently developed. In permissioned blockchain, unlike public blockchain, access is restricted to some users only. Furthermore, the central authority defines the role of a user within the network and to which information he has access. There are different levels of access to the network, each with different functionalities:

- Reading the ledger, which may be subject to certain constraints or may be accessible for all nodes.
- The possibility of making transactions within the network, which must then be validated and inserted into the blockchain.
- Perform block mining and validation operations.

This more performing kind of blockchain, is therefore preferred by companies that want to maintain a high level of confidentiality of their data while benefiting from all the characteristics of the blockchain.

2 Safety for business

The blockchain is designed to keep safe and unaltered the data inside. The data are inserted into the blocks in way to create a chain held together by the integrity of the same data. If a single data is altered the whole chain is invalidated.

2.1 *Block composition*

This security is guaranteed by a particular string of 64 characters called *hash* which represents all the information encrypted within a block and must meet previously agreed criteria. This *hash* is an alphanumeric string of characters generated by the SHA-256 function. The block is made up of different fields based on the use that you want to make of the blockchain and the data you want to protect. There are some fields of the block that are essentials for the blockchain to be well structured and capable of being efficiently constructed and these fields are:

An **index**, which allows to uniquely identify each block; the **timestamp**, which allows to identify the exact moment in which the block is generated; the **hash** of the previous block. The last data allows to create the famous "chain", each hash of a block will become a fundamental component of the next one. *Nonce* is a really important value for the respect of all the above mentioned criteria and will be further analyzed later; to all this basic information we must add all the ones we want to actually store safe and intact. For the realization of a new block, it was achieved a function which allows to obtain the index of the last block inserted.

2.2 *"Chain" creation*

It was created a function that permit, after obtaining this information, all of this data to be acquired. This function (*push ()*) takes care of acquiring information such as index, timestamp and all the data we want to preserve. Within this function, once the hash of the previous block has been obtained (through its index), the block index is incremented so as to become the index of the block that we are currently inserting. Next, the heart of the insertion: the function that deals with "undermining" and creating a hash that allows to respect the criterion.

This criterion is useful because it allows us to create a non-trivial and independent hash of the data we actually want to memorize, for this task is used the *nonce*. The *nonce* is initially set to zero. The hash is generated with all the data, index, timestamp, information to be stored, previous block hash and the *nonce*, then it is consulted a function that generates the hash through SHA-256 encryption. If the criterion is respected, the block is added to the chain, otherwise it enters in a *while* loop. The while condition is repeated until the hash is composed in such a way as to meet the previously decided requirement. For each iteration the nonce is

incremented by one, this allows to generate a completely different hash for each iteration, up to the requirement. The more difficult is the requirement to achieve, more iterations will be needed, greater will be potentially the nonce value. Once the wanted hash is obtained, the block is completed and the information obtained can be safely stored.

2.3 Validation of the process

All the information are stored in a database. For this reason there is the possibility that somehow they can be modified, so it is necessary to use an algorithm that allows to understand if the blockchain has been compromised and is invalidated. This is made possible thanks to a function that runs through the entire blockchain and recalculates the hashes of all the blocks: through a *for* loop, a specially created matrix crosses each block and verifies its integrity. At each iteration through the data present in the array, the hash of the block is recalculated and compared to the stored hash. The hash of the previous block must coincide with the previous hash of the current block, if one of these two conditions is not respected then the blockchain has been invalidated. If successful, the iteration will be repeated until the whole blockchain is verified.

3 Cases study analysis

We will show below several case studies in which the blockchain was used with various kind of implementation. This will explain the potential, the possible uses and developments of this new technology that holds great potential for all the multiplicity of services or processes which need information integrity, time validations and business continuity.

3.1 L'Antincendio Srl

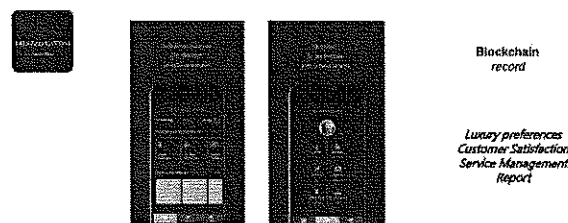
When we talk about services or products designed to guarantee people safety, the temporal certification and the integrity of the information about them becomes a fundamental factor. In this context the company L'Antincendio Srl has found an answer in the blockchain technology that has been used to map and certify all the maintenance and customer reports on each individual fire extinguisher. In this way you will not only have the possibility to reconstruct the time-certified history of each individual fire extinguisher, but you can also analyze the information in order to obtain the security level of the product with the certainty of data integrity. This is a clear example of how this technology has been used by a company to improve the level of safety of its products.

3.2 *Capurso Azienda casearia Srl*

The case study about Capurso Azienda casearia Srl is different, in this example the company, in order to protect the quality of its product and remain compliant with the DOP hygiene standards also in key 231, required an effective tool for mapping all the processes concerning the production of the raw material (cow's milk), in this case the blockchain was used to guarantee the integrity of the data during the audit of the entire production process of the raw material and to certify the history of every single dairy cow allowing the company to pursue increasingly higher quality standards.

3.3 *Palazzo Gattini*

A very special case study concerns the blockchain implemented for the Palazzo Gattini structure, which offers luxury services with high quality standards and interfaces with a very demanding clientele. In this case, it was decided to map not only the entire process concerning the service offer, but also the preferences of each individual user (luxury preference) and the customer satisfaction within the ledger. Given the strong semantic link between the customer and the information saved on the ledger, it has been used a semantic blockchain, that is to say that within each individual block there will also be the hash of the previous block concerning the client in question, in this way we can analyze in a better way the information collected to obtain the level of luxury compliance of the structure.



3.4 *Erreffe Srl*

We can find another example of the use of blockchain in the project manager of Erreffe srl, in this case all the tasks within the company have been mapped in the ledger with the related processes according to the individual tasks, it has been fundamental to certify all the progress in order of time in way to be able to view and analyze the evolution and of the progress with complete and certified data.

3.5 *Caldarola Srl*

In the last case study the blockchain technology is used by the company Caldarola Srl for the vehicle rental software. Here, all the information (generic or about maintenance) related to the individual vehicles present in the company are saved in the blockchain including picking and releasing. In this case the time certification has a great value in order to allow the company to know the current status of all vehicles and to be able to view the chronology of all the maintenances working on integral data and with certified time values without using particularly complex processes for this kind of certification.

4 Conclusions and further developments

This paper presented interesting application developments of permissioned blockchain applied to practical case studies of small companies in southern Italy. Incorporating the potential of this technology in the processes of southern SMEs is an indication of how much this technology is transversal and of great importance for business improvement and for meeting regulatory compliance requirements.

An example of application may concern the certification of machinery 4.0, as the interconnection of the same can be proved by blockchain recording of the data of the interconnection of individual machines, which would have probative validity. From this point of view, the process of processing certified log via blockchain, which would allow to compare the certified "real process" with the company procedures in order to determine a certified "process gap", can also be applied.

References

1. Nakamoto, S. (2009). Bitcoin: A Peer-to-Peer Electronic Cash System. <https://bitcoin.org/bitcoin.pdf>.
2. M.Colucci, UIBM, Patent, 27/12/2018 - 102018000021061, Sistema ad alta innovazione per la certificazione dei processi di manutenzione antincendio
3. F.Fanari, V.Santarcangelo,D.C. Sinitò, Esperienze di Ricerca e Sviluppo applicate alle brillanti realtà del nostro sud,RCE Multimedia, 2018
4. A. Brandonisio, UIBM, 28/12/2018 n. 102018000021313, Sistema basato su blockchain per la certificazione di filiera di prodotti caseari
5. E. Grassano, UIBM, 02/01/2019 rif n. 102019000000025, Sistema blockchain oriented per la certificazione dei servizi di una struttura luxury
6. R. Festa, C. Dell'Acqua, S. Burgi, UIBM, 2019, "Sistema intelligente per la gestione di commessa mediante blockchain in ottica BIM"
7. M.Masi, UIBM, Rif. 102018000007495, Sistema basato su blockchain per il controllo della flotta in compliance
8. M.Masi, SIAE 21/08/2018,Rif. 012763, Sagest Fleet Gdpr Blockchain